3 Model Specification

SEM starts with the model specification. A structural equation model is essentially a set of structural equations that puts constraints on the model-implied variance-covariance matrix of observed variables. It can be subdivided in the inner model and the outer model. The inner model captures the relationships among the constructs. The outer model entails the relationships between the constructs and the observed variables. The employed auxiliary theory dictates a certain type of outer model: composite model vs. reflective measurement model. The chapter also explains causal-formative measurement. Single-indicator constructs can be modeled as well as categorical exogenous variables. Finally, various forms of inner models are discussed.

3.1 What Is a Structural Equation Model?

All elements of SEM circulate around the structural equation model: model specification, model identification, model estimation, and model testing and assessment. Without exaggeration one could thus say that the structural equation model forms the core of SEM. It deserves thus a deeper look.

Structural equation models are a special case of statistical or mathematical models. They entail a set of mathematical equations that describe a section of the observable world, thereby putting constraints on the model-implied variance-covariance matrix of observed variables. The mathematical equations are functions of the form $\{\mathcal{V}\} = f(\{\mathcal{V}\}, \{\mathcal{R}\}, \{p\})$. $\{\mathcal{V}\}$ is a set of model variables, $\{\mathcal{R}\}$ is a set of relationships, and $\{p\}$ denotes a set of model parameters.

Structural equation models can not only be expressed by means of equations, but also graphically. Graphically representing structural equation models has intuitive appeal, because it often facilitates the understanding and interpretation of the model. To illustrate the variables, relationships,

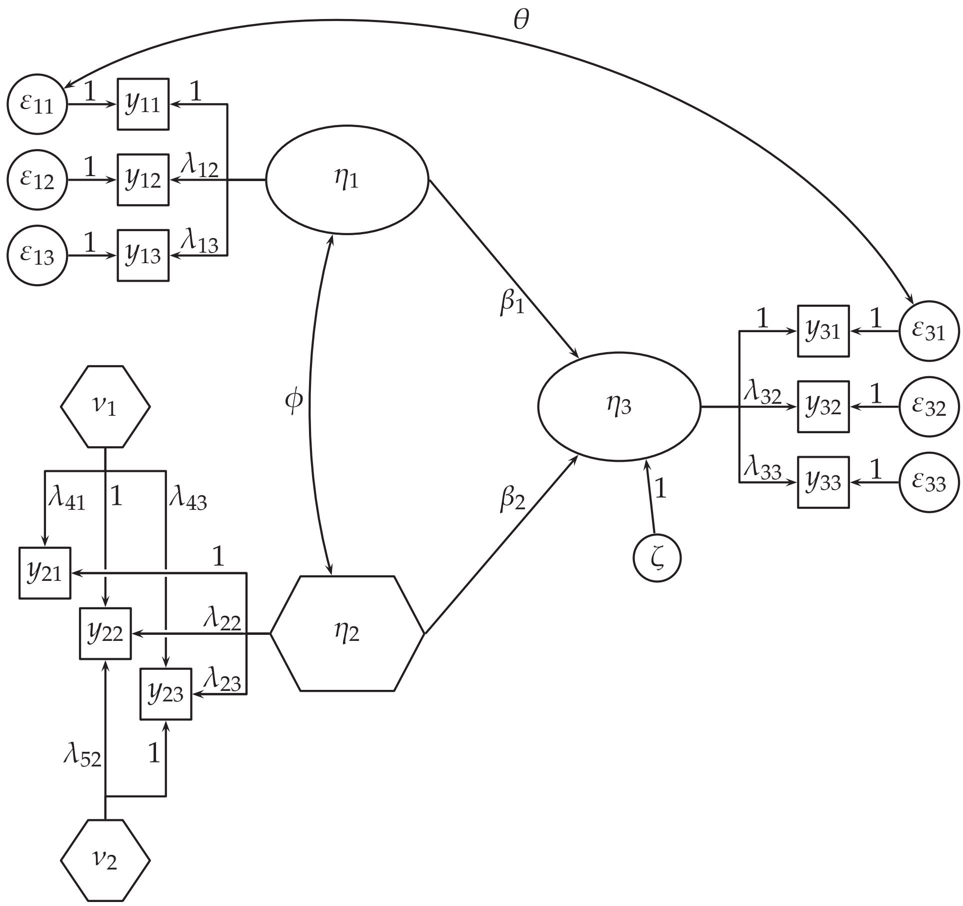

FIGURE 3.1. An exemplary structural equation model with three constructs.

and parameters that are typical for structural equation models, we make use of the small structural equation model depicted in Figure 3.1.

In SEM, we can distinguish between three types of variables: observed variables, unobserved variables, and synthetic variables. All three types of variables can be found in Figure 3.1.

We speak of an *observed variable* if in our data we have concrete values of this variable for the individual observations. In graphical representations of structural equation models, observed variables are denoted by squares or rectangles. There are nine observed variables in the exemplary model of Figure 3.1: y_{11} to y_{13}, y_{21} to y_{23}, and y_{31} to y_{33}.

In contrast, *unobserved variables* do not have a sample realization; i.e., we do not have concrete values of an unobserved variable for the individual observations. Unobserved variables are visualized using round shapes, particularly ovals and circles. There are three types of unobserved variables: latent variables, measurement errors, and disturbance terms.

Latent variables are constructs in the sense of measurement theory; i.e., the axiom of local independence applies (see Section 2.3). They are usually employed to represent theoretical concepts of behavioral research. Latent variables are typically represented by ovals. In Figure 3.1, η_1 and η_3 are latent variables.

Another type of unobserved variables is *measurement error*. Measurement error captures that portion of an observed variable's variance that cannot be explained by the corresponding latent variable. Measurement errors are typically depicted as small circles. Figure 3.1 contains six measurement errors: ε_{11} to ε_{13} and ε_{31} to ε_{33}. Latent variables and measurement errors are discussed more thoroughly in Subsection 3.2.2.

If a construct in a structural equation model is explained by one or more other variables, usually a part of its variance remains unexplained. This unexplained variance is captured by a *disturbance term*. Disturbance terms are visualized as circles. In Figure 3.1, ζ is a disturbance term.

Synthetic variables are not obtained through data collection but created by means of data transformation. The dominant way of creating synthetic variables is by forming linear combinations of other variables. Depending on whether these other variables are observed or unobserved, the resulting synthetic variable will be observed or unobserved. Two types of synthetic variables can be distinguished: emergent variables and excrescent variables.

Emergent variables are constructs in the sense of synthesis theory; i.e., the axiom of unity applies (see Section 2.4). They are usually employed to represent forged concepts of design research. The axiom of unity requires that an emergent variable is related to at least one other variable in a structural equation model. Emergent variables are typically represented by hexagons. In Figure 3.1, η_2 is an emergent variable.

Excrescent variables capture the remaining variance of a set of variables after an emergent variable has been extracted. Excrescent variables are unrelated with all variables they are not formed of. They are visualized by means of small hexagons. In Figure 3.1, ν_1 and ν_2 are excrescent variables. Emergent and excrescent variables are the building blocks of composite models as presented in Subsection 3.2.1.

There is another important distinction of variables in a structural equation model, namely by the role that they play. Variables that do not depend on any other variable in the model are called *exogenous*. They have a cause external to the structural equation model. In a graphical representation of a structural equation model, exogenous variables can be recognized by the fact that no straight arrow points at them. For instance, the variables η_1, ε_{11}, ν_1, and ζ in Figure 3.1 are endogenous. In contrast, *endogenous* variables have an internal cause, which means that there is at least one other variable in the structural equation model causing them. In a graphical representation of a structural equation model, endogenous variables can be recognized by the fact that at least one straight arrow points at them. For instance, the variables η_3 and y_{11}in Figure 3.1 are exogenous.[1]

Predominantly, two types of relationships between variables are studied in SEM: direct linear relationships and covariances. A *direct linear relationship* expresses a dependence relationship; it is characterized by a straight arrow pointing from an independent variable to a dependent variable. The parameters quantifying direct linear relationships between constructs are called path coefficients. In Figure 3.1, the parameters β_1 and β_2 are path coefficients. Direct linear relationships pointing from constructs to observed variables are called (indicator) loadings. For loadings, usually the Greek letter λ is employed.

Covariances mainly occur at three locations within a structural equation model: between constructs, between error terms, and between disturbance terms. Figure 3.1 contains a covariance between the constructs η_1 and η_2 denoted by ϕ, and a covariance between the error terms ε_{11} and ε_{31} denoted by θ. A covariance between a disturbance term and another disturbance term or a construct are denoted by ψ.

Finally, the *variances* of all introduced variables except for the observed indicators are model parameters as well. The variances make use of the same symbol as the according covariances, because they can all be expressed as elements of a variance-covariance matrix. Hence, the variances of the emergent, latent, and excrescent variables are denoted by ϕ, the variances of the disturbance terms by ψ, and the error variances by θ. Figure 3.2 summarizes the elements of structural equation models that have a visual representation.[2]

Structural equation models are typically composed of two submodels: the outer model and the inner model. The outer model contains the equations postulated by the auxiliary theory. In contrast, the inner model captures the effects among constructs; it models thus the substantial theory. In the remainder, we discuss each of these submodels in more depth.

[1] In order to distinguish between exogenous and endogenous constructs, some scholars prefer to use the Greek letter ξ for exogenous constructs and the Greek letter η for endogenous constructs only.

[2] Some SEM software with graphical user interface does not offer the possibility to specify synthetic variables. In such a case, one can usually employ the workaround of specifying unobserved variables instead of synthetic variables.

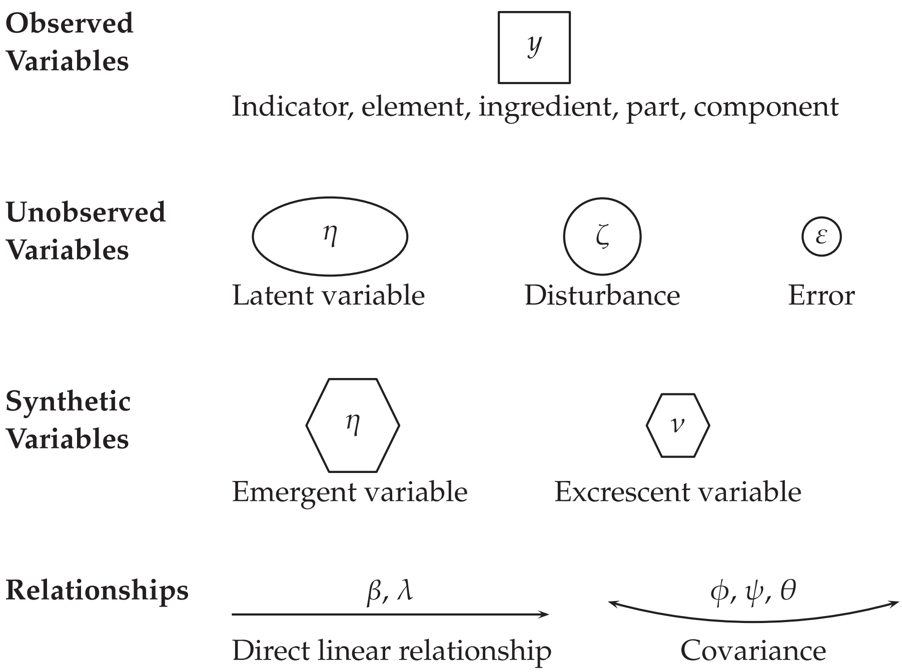

FIGURE 3.2. Symbols employed in graphical representations of structural equation models.

3.2 The Outer Model

The outer model expresses the relationships between the observed variables and the constructs. The observed variables are also known under the terms "indicators," "manifest variables," or "observable variables." Each observed variable corresponds to a column in the empirical dataset.

The outer model implements the constructs' auxiliary theories. Since there are two auxiliary theories, there are also two main outer models with multiple indicators: The *composite model* (see Subsection 3.2.1) should be used to operationalize forged concepts if synthesis theory is used. The *reflective measurement model* (see Subsection 3.2.2) is the model of choice to operationalize theoretical concepts of behavioral research if measurement theory is applied. Reflective measurement models can be complemented with causal-formative measurement (see Subsection 3.2.3).

For each concept, analysts must decide whether they intend to model it as an emergent variable or a latent variable. As depicted in Figure 3.3, this decision should be based on the nature of the concept.

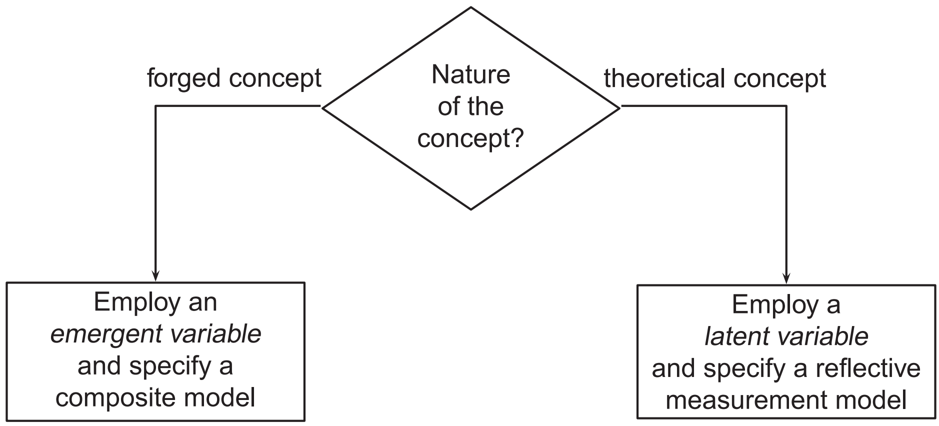

FIGURE 3.3. Decision tree for specifying outer models.

For forged concepts, researchers should employ the synthesis theory as auxiliary theory and hence rely on emergent variables. This means that phenomena such as capabilities, indices, interventions, norms, plans, policies, portfolios, processes, recipes, strategies, and values are best modeled as emergent variables. In contrast, for theoretical concepts, researchers should make use of the measurement theory and rely on latent variables. This means that phenomena like attitudes, diseases, emotions, feelings, perceptions, and traits should be modeled as latent variables.

Since outer models connect constructs with observed variables, they must not only suit the type of concept but also the role of the observed variables. The roles of observed variables are defined through the relationship with the corresponding concept. If observed variables have a material relationship with a concept, i.e., they form or make up a forged concept, then they play a role as components of an emergent variable. In the methodological literature they are also referred to as composite-formative indicators. If observed variables have a causal relationship with a theoretical concept of behavioral research, i.e., they cause or are caused by it, then they can play the role of an indicator of a latent variable. At least some of a latent variable's indicators must be consequences, the so-called effect indicators. Additionally, there might be some indicators that are antecedents, the so-called cause indicators or causal-formative indicators.

There is some confusion in the literature about what is meant by formative measurement. Authors referring to formative measurement sometimes discuss the characteristics of composite models and sometimes those of causal-formative measurement models (see in particular early contributions on formative measurement, such as Diamantopoulos & Winklhofer, 2001; Jarvis, MacKenzie, & Podsakoff, 2003). This confusion can be traced back to Edwards and Bagozzi (2000), who deliberately sought a term that characterizes both causal and definitorial relationships. The confusion has culminated in statements like "When an endogenous latent variable relies on formative indicators for measurement, empirical studies can say nothing about the relationship between exogenous variables and the endogenous formative latent variable" (Cadogan & Lee, 2013, p. 233; for a rejoinder see Rigdon, 2014) or variance-based SEM "is not an adequate approach to modeling scenarios where a latent variable of interest is endogenous to other latent variables in the research model in addition to its own observed formative indicators" (Aguirre-Urreta & Marakas, 2013, p. 776; for a rejoinder, see Rigdon et al., 2014). The confusion can be cleared up if one carefully distinguishes between composite models of emergent variables and causal-formative measurement of latent variables. Emergent variables clearly distinguish between antecedents and indicators. They *can* play the role of an endogenous variable in a larger model. In contrast, latent variables measured in a causal-formative way make no distinction between causal antecedents and cause indicators. They automatically play the role of endogenous variables.

For researchers, it is important to correctly classify their observed variables, because components (composite-formative indicators), effect indicators (reflective indicators), and cause indicators (causal-formative indicators) differ with regard to several characteristics. Table 3.1 summarizes the differences between the three roles of observed variables.

While it is recommended to operationalize concepts by means of multiple observed variables, for various reasons there is sometimes only one observed variable available per concept (Diamantopoulos, Sarstedt, Fuchs, Wilczynski, & Kaiser, 2012). The resulting single-indicator measurement is treated in Subsection 3.2.4. Particular care is required if observed variables are categorical (see Subsection 3.2.5).

Sometimes concepts are operationalized even without observed variables. In the realm of covariance-based SEM, such operationalizations are discussed under the term "phantom variables" (Rindskopf, 1984). In most of these instances, auxiliary theories are applied to the relationships among constructs. Technically, such phantom variables are realized as second-order constructs. This advanced form of SEM is explained in Chapter 10.

3.2.1 Composite Models

The composite model, also referred to as the composite factor model (Henseler et al., 2014) or the composite-formative model (Bollen & Diamantopoulos, 2017), assumes a definitorial relationship between a construct and its indicators. This means that the construct is made up of its indicators or

TABLE 3.1. Different Roles of Observed Variables

Characteristic	Component	Effect indicator	Cause indicator
Observed variable's role	Ingredient, part, element	Consequence	Antecedent, cause
Corresponding construct	Emergent variable	Latent variable	Latent variable
Correlations among observed variables	High correlations are common, but not required	High correlations are expected	No reason to expect the measures are correlated
Proneness to measurement error	Can contain measurement error	Contains measurement error	Can contain measurement error
Informative about measurement error	Not informative about measurement error	Jointly informative about measurement error	Not informative about measurement error
Consequences of dropping an indicator	Dropping an indicator alters the construct and may change its meaning	Dropping an indicator does not alter the meaning of the construct	Dropping an indicator increases the error on construct level

elements. In composite models, a composite serves as a proxy of the concept under investigation (Ketterlinus, Bookstein, Sampson, & Lamb, 1989; Maraun & Halpin, 2008; Rigdon, 2012; Tenenhaus 2008). The composite model is the tool of the trade for forged concepts, when researchers employ the synthesis theory as auxiliary theory and hence rely on emergent variables. The composite model is thus suitable for a plethora of phenomena: activities, approaches, availability, baskets, capabilities, classifications, configurations, compilations, consumption, convenience, decisions, designs, developments, efficacy, equity, expenditures, frameworks, indices, instruments, interventions, inventories, justice, manipulations, maps, methods, mixes, models, modifications, norms, operations, orientations, patterns, plans, policies, portfolios, practices, prestige, procedures, processes, production, propositions, qualifications, quality, power, recipes, resources, skills, solutions, sources, standards, status, strategies, structures, support, systems, tactics, technologies, tools, treatments, typologies, and values are best modeled as emergent variables. All the aforementioned phenomena are not naturally occurring, but artifacts that have been created by humans.[3]

[3]Note that my judgment refers to the mentioned phenomena, and not perceptions of them. For instance, when I list quality, I refer to factual or delivered quality, not perceived quality.

Composites are facing an increasing popularity in business and social science. Guidelines for index construction (see, for instance, Diamantopoulos & Winklhofer, 2001) have had substantial influence on researchers who proposed new indices. In the social sciences, composites are used more and more for measuring complex phenomena such as poverty, progress, and well-being (Lauro, Grassia, & Cataldo, 2018). Probably the most well-known index in sociology is socio-economic status, which is defined by income and education (Nunnally & Bernstein, 1994). Moreover, the composite model "will often be appropriate for ecological studies because of the multifaceted nature of [their] theoretical concepts" (Grace, Anderson, Olff, & Scheiner, 2010, p. 67). In general, the composite model turns out to be a formidable instrument for modeling part-whole relationships. Nelson and Stolterman (2003, p. 119) remind us that "[a]lthough it's true that 'the whole is greater than the sum its parts,' we must also acknowledge that the whole is of these parts."

In many instances when the term "formative construct" is used in literature, authors actually mean emergent variables, not latent variables measured in a formative way. Statements about formative constructs like that their "[i]ndicators are defining characteristics of the construct" (Jarvis et al., 2003, p. 203) or that "[f]ormative constructs occur when the items describe and define the construct" (Petter, Straub, & Rai, 2007, p. 623) clearly point to a definitorial relationship between a construct and its observed variables and thus a composite model. Even a statement like "formative constructs are inextricably tied to their measures (i.e. they do not exist independently of measurement [. . .])" (Diamantopoulos, 2006, p. 14) fits more to a composite model than to a latent variable measured through causal-formative indicators. To conclude, composite models can be considered the model of choice for formative constructs if these are understood as being formed or defined (not caused!) by their components. Another application of composite models is so-called formed attributes (Rossiter, 2002).

Components as the observed variables in composite models differ from cause indicators (causal-formative indicators) in a crucial point: the type of causality. Whereas cause indicators are causal antecedents of their construct, components are material ingredients and play a definitorial role. As has become customary for artifacts (cf. Gregor, 2009), we engage in a wider understanding of causality, namely as originally proposed by Aristotle. "Aristotle distinguished four causes . . . material, formal, efficient, and final. Respectively, they indicate that from which something was made (material cause), the pattern by which something was made (formal cause), that from which comes the immediate origin of movement or rest (efficient

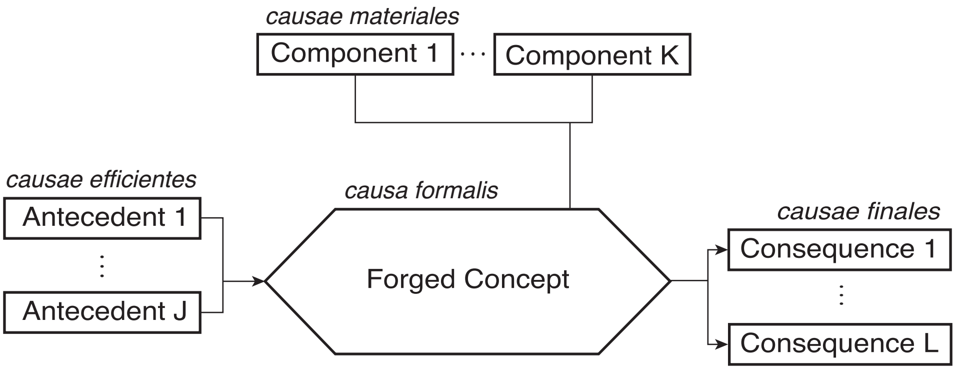

FIGURE 3.4. Aristotle's four causes within a composite model.

cause), and the end for which it is made (final cause)... " (Poole, 2000, p. 42). Müller-Merbach (2005) sees ubiquitous value in Aristotle's four causes: for a more effective design of artifacts or systems of any kind, for better understanding reality, and for structuring our knowledge. Already Heidegger (1954) noted that all four causes play a role in design research. Concretely, Aristotle's four causes can help to better understand how a forged concept is embedded in a nomological net. Figure 3.4 shows such a prototypical research framework: The components serve as material causes; antecedent variables are efficient causes; the purposes for which the concept was forged are the final causes; and the forged concept itself represents the formal cause.

In formal terms, the composite model regards the emergent variable η_j as a linear combination of its components y_{jk}, each weighted by a component weight w_{jk}:

$$\eta_j = \sum_{k=1}^{K} w_{jk} \cdot y_{jk} \tag{3.1}$$

This equation underlines the definitorial role of the components, because they fully produce the composite.

Noteworthily, there is no unique solution for the weights of single emergent variables as specified by Equation 3.1. The weights could have any value without rendering the model wrong, i.e., the model is not identified (see also Chapter 4). Consequently, single composites are not accessible to model testing. Since they do not impose any restrictions on the model-implied variance-covariance matrix of observed variables, some purists would not regard them as models at all. However, as soon as a composite

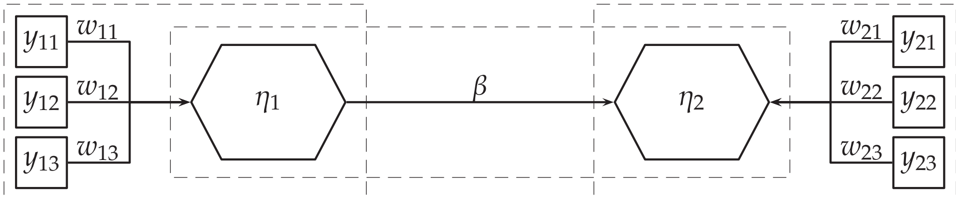

FIGURE 3.5. Specifying composite models as partial models in variance-based SEM.

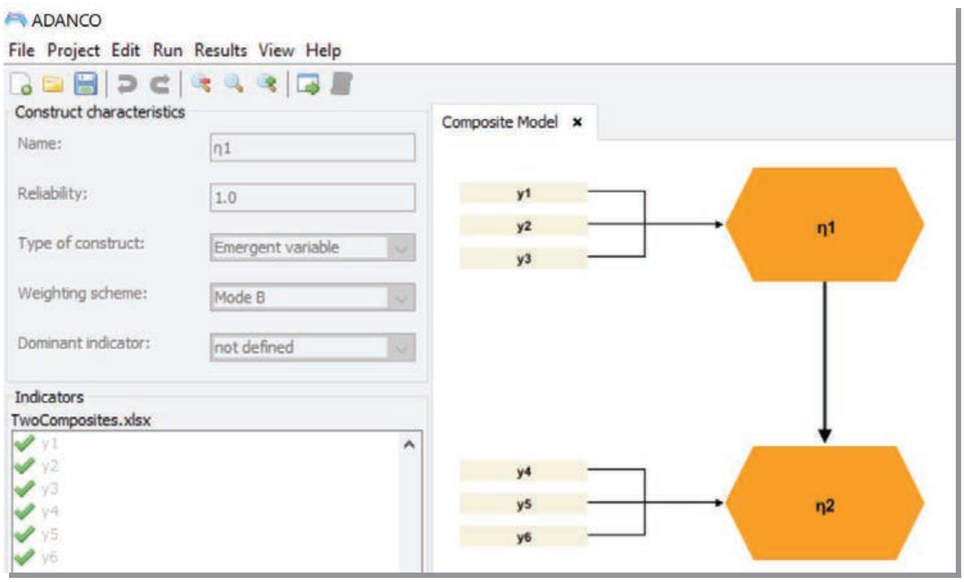

FIGURE 3.6. Specifying emergent variables in ADANCO.

is linked to at least one other variable in a structural equation model, we do have a testable composite model. Figure 3.5 depicts a small composite model of two related composites, η_1 and η_2. This visualization is typical for variance-based SEM. Figure 3.6 shows the graphical implementation of this composite model in ADANCO 2.3.1.

The graphical representation of a composite model as depicted in Figure 3.5 has intuitive appeal. The arrows of the observed variables point to the construct, emphasizing the definitorial relationship. Moreover, the graphical model contains the weights as model parameters, and thereby facilitates the interpretation of the emergent variable. However, this way of graphically representing composite models has a major disadvantage, namely that the graphical model in fact consists of two separate submodels that are not fully integrated. On the one hand, there is a submodel specifying the relationships between the construct and its observed variables.

On the other hand, there is a submodel specifying the construct's interrelatedness with other variables of the structural equation model. As a consequence, there is no straightforward way for deriving the model-implied variance-covariance matrix – a fact that has already led some researchers to postulate that components must be uncorrelated (see, e.g., MacCallum & Browne, 1993), although this is an unnecessary limitation.

It has become clear that in composite models, the relationships between the components and the construct are not cause-effect relationships, but rather a prescription of how the ingredients should be arranged to form a new entity. More precisely, one can speak of a prescription for dimension reduction (Dijkstra & Henseler, 2011). The idea of dimension reduction permits an alternative view on composite models.

The K components $\boldsymbol{y}$ of a composite span a K-dimensional space (assuming that there is no perfect collinearity among components). Any non-trivial linear combination of components is a synthetic variable that represents a single dimension of this space. Two types of synthetic variables can be distinguished: Some synthetic variables are purposefully formed in such a way that they have as strong as possible relations with other variables of the structural equation model. These synthetic variables are called "emergent variables." They are denoted as $\boldsymbol{\eta}$. Next to them, there can be synthetic variables that are unrelated to all other variables of the structural equation model. Since these synthetic variables are in some sense superfluous, I call them "excrescent variables." They are denoted as $\boldsymbol{\nu}$. Jointly, the emergent variables and the excrescent variables span the K-dimensional space of the components $\boldsymbol{y}$ (where $\boldsymbol{\Lambda}$ is a matrix of loadings). In matrix notation, this is expressed by Equation 3.2:

$$\boldsymbol{y} = \boldsymbol{\Lambda} \begin{pmatrix} \boldsymbol{\eta} \\ \boldsymbol{\nu} \end{pmatrix} \tag{3.2}$$

Synthesis theory's axiom of unity states that if a set of components make up one forged concept, then there is one and only one emergent variable. Thus, if synthesis theory holds (see Section 2.4), then there is exactly one dimension that is related to other variables of the structural equation model, whereas the remaining dimensions are entirely unrelated to other variables of the structural equation model. Figure 3.7 illustrates the composite model of an emergent variable η with three components y_1 to y_3. This specification of a composite model is one integrative model in contrast to the previous specification of a composite model as shown in Figure 3.5.

On the one hand, the representation of a composite model in terms of synthetic variables as depicted in Figure 3.7 is less intuitive than the one shown in Figure 3.5, because it does not display in a straightforward

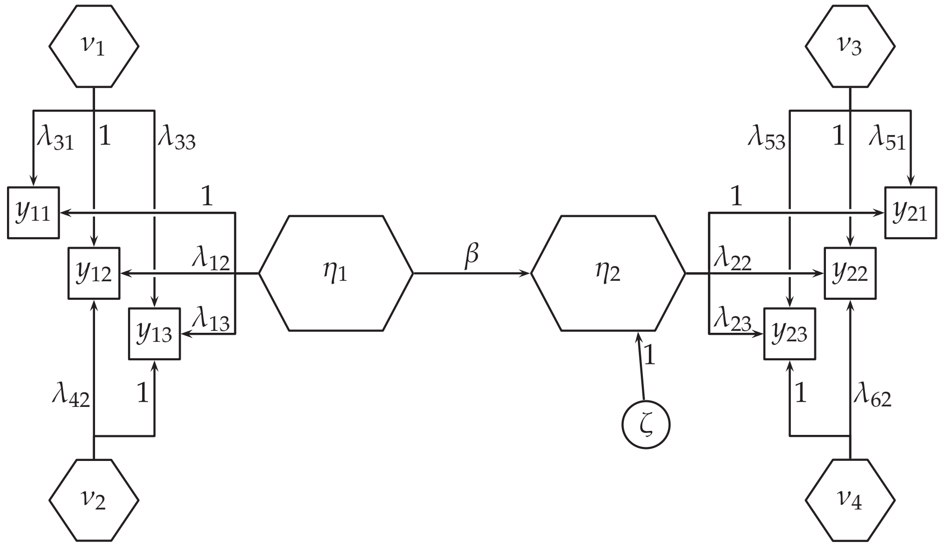

FIGURE 3.7. Specifying composite models in covariance-based SEM.

manner that the emergent variable is formed as a linear combination of its components. On the other hand, it demonstrates that the model can be expressed using component loadings instead of component weights as model parameters. It also has the advantage that the model-implied variance-covariance matrix of observed variables can be determined in the conventional way (see Equation 6.1 on page 119).

The composite model imposes fewer restrictions on the covariances between indicators of the same construct than a reflective measurement model. Since composite models are less restrictive than reflective measurement models, they typically have a higher overall model fit (Landis, Beal, & Tesluk, 2000).

Composite models as such do not explicitly take measurement error into account. In their standard form, they assume that the components are free from measurement error. In many practical situations of empirical research, this assumption is certainly untenable. Fortunately, there are two options available of how measurement error can be taken into account at the stage of model specification.

Firstly, researchers can rely on extant knowledge about the reliability of the observed variables or the composite. Using variance-based estimators, it is possible to manually predefine a composite's reliability so that a correction for attenuation can be employed. In particular, ADANCO 2.3.1 permits manually defining the reliability of emergent variables with one or more components. In covariance-based SEM, one should include indicator errors with predefined error variances. In both instances, it is not possible to determine the amount of measurement error based on the available data, but one has to rely on information external to the analysis.

Secondly, instead of employing error-prone observed variables as components of a composite, one could use error-free latent variables as components. The resulting model will entail a Type-II second-order construct (see Section 10.3), i.e., an emergent variable made up of latent variables.

3.2.2 Reflective Measurement Models

Reflective measurement models form the backbone of behavioral research. They can be used to model phenomena like attitudes, diseases, emotions, experiences, feelings, intentions, needs, perceptions, and traits. The reflective measurement model is so strongly established in the social and business sciences that hardly anyone questions its applicability, despite the fact that empirical evidence almost always speaks against it (see, e.g., Henseler et al., 2014; Rigdon, 2012; Schönemann & Steiger, 1976).

The reflective measurement model has its roots in classical test theory and psychometrics (Nunnally & Bernstein, 1994); it is a realization of measurement theory as presented in Section 2.3. The observed variables are assumed to reflect variation in a latent variable and, thereby, changes in the construct are expected to be manifested in changes in all indicators comprising the multi-item scale. Thus, the direction of causality is from the construct to the indicators. Reflective measurement models are essentially common factor models, which postulate that there is a latent variable underlying a set of observed variables or indicators. In turn, each observed variable is regarded as an error-afflicted manifestation of a latent variable, as expressed by the following equation (see for instance Matsueda, 2012):

$$\boldsymbol{y} = \boldsymbol{\lambda}\eta + \boldsymbol{\varepsilon} \tag{3.3}$$

In this equation, $\boldsymbol{y}$ is a vector of the observed variables, $\boldsymbol{\lambda}$ is a vector of loadings, η is a latent variable, and $\boldsymbol{\varepsilon}$ is a vector of measurement errors. The measurement errors are assumed to be centered around zero and uncorrelated with other variables in the model (observed variables, latent variables, errors, etc.). Figure 3.8 depicts a typical reflective measurement model. It consists of three observed variables y_1 to y_3, three measurement errors ε_1 to ε_3 with the according variances θ_1 to θ_3, and a latent variable η with a variance of ϕ. Figure 3.9 on page 53 shows what the according specification in ADANCO looks like. It differs in several aspects from Figure 3.8: Firstly, the measurement errors are not drawn. Since ADANCO 2.3.1 does not allow to constrain or free variances or covariances of measurement errors, these model elements are omitted for the sake of a simpler visual model.

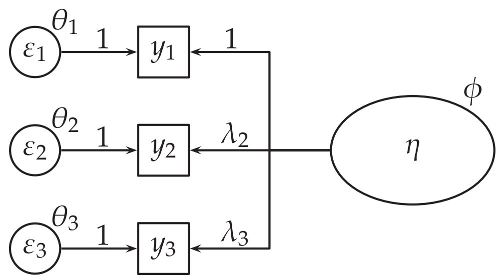

FIGURE 3.8. Reflective measurement.

Secondly, there is a difference in the parameterization with regard to the variance of the latent variable η. In Figure 3.8, the first indicator's loading λ_1 is fixed to 1 so that the variance ϕ of the latent variable η will equal the useful portion of the variance of the observed variable y_1. In contrast, variance-based SEM and thus ADANCO makes use of standardized constructs, which means that the variance of the latent variable η is fixed to one. In exchange, the first indicator's loading λ_1 becomes a free parameter.

The latent variable is not directly observable, but only the correlational pattern of its indicators provides indirect support for its existence. The reflective measurement model is somewhat peculiar in the sense that the number of dimensions underlying a reflective measurement model's variables is one higher than the rank of the empirical correlation matrix. That means that as an outcome of the model specification somehow an additional dimension emerges. For instance, whereas three indicators y_1 to y_3 span a three-dimensional space, the new orthogonal variables of the factor model, η and ε_1 to ε_3, will span a four-dimensional space. Where does this fourth dimension come from? While one could argue that this additional dimension captures the transcendence of the theoretical concept, one could evenly well argue that it is mere imagination. An analogy would be to draw a three-dimensional graph on a sheet of paper. The graph itself will never leave the two-dimensional space, but it provides a glimpse on an additional dimension. Technically speaking, factor models are subject to factor indeterminacy, which means that the relationship between a common factor and variables outside the model can have an arbitrary strength (Rigdon, Becker, & Sarstedt, 2019). Depending on someone's view on this, one can complain about factor indeterminacy or appraise the transcendence of the theoretical concept (Schönemann & Steiger, 1976).

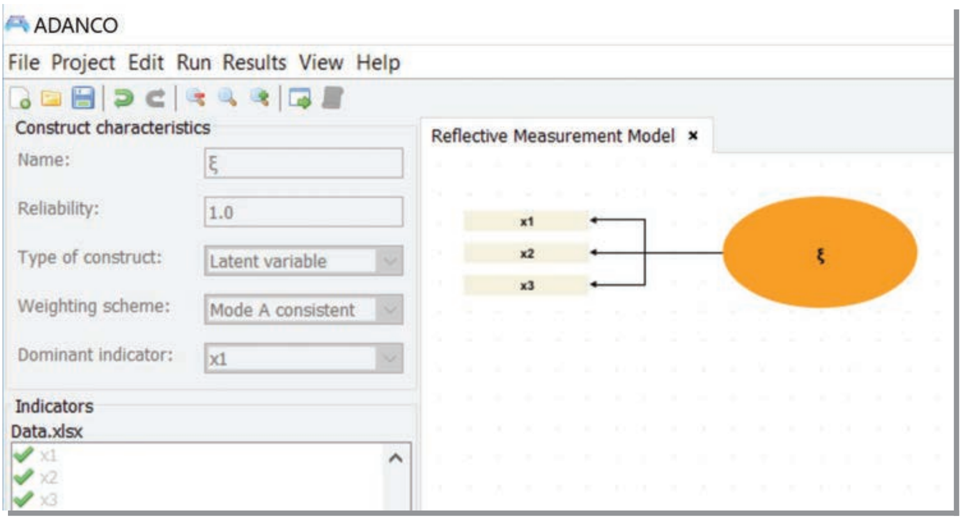

FIGURE 3.9. Reflective measurement in ADANCO.

3.2.3 Causal-Formative Measurement Models

Like the reflective measurement model, the causal-formative measurement model (often just referred to as the formative measurement model) also relies on measurement theory as the underlying auxiliary theory. However, it assumes a different epistemic relationship between the latent variable and its indicators: The indicators are considered as immediate causes of the focal latent variable (Fassott & Henseler, 2015). The following equation represents a causal-formative measurement model, where β_{ji} indicates each formative indicator's contribution to η_j, and ζ_j is an error term:

$$\eta_j = \sum_{i=1}^{I} \beta_{ji} \cdot y_{ji} + \zeta_j \tag{3.4}$$

On first sight, this equation strongly resembles the one for a composite model; only the measurement error on the construct level makes it distinct. However, there is a fundamental difference between the two: Whereas composite models assume a definitorial relationship between the composite and its components, causal-formative measurement models assume a causal relationship between the latent variable and its indicators. In particular, cause indicators are assumed to cause the latent variable.

The measurement error on the construct level implies that the latent variable of interest has not been perfectly measured by its formative indicators. Except for rare cases when all causes can be measured (see, e.g., Diamantopoulos, 2006), it is indispensable to also have a reflective measurement model. Otherwise it is not possible to capture the entire content of

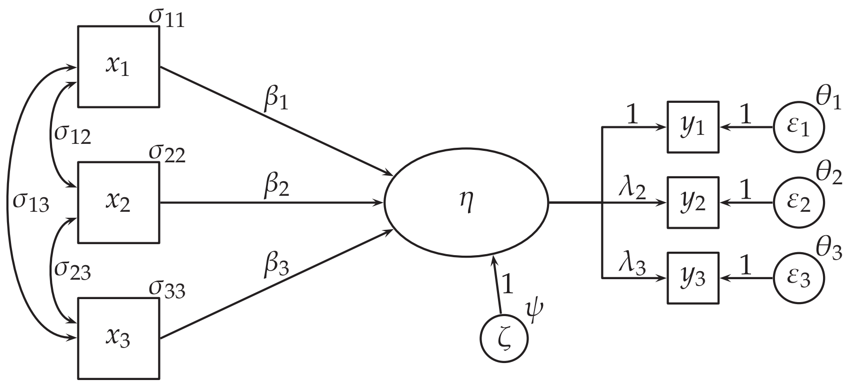

FIGURE 3.10. Causal-formative measurement (Henseler, 2017).

the theoretical concept (Aguirre-Urreta, Rönkkö, & Marakas, 2016). The reflective indicators can be observed or latent, as long as there are at least two reflective indicators whose correlation is fully attributable to the construct as a common cause. Figure 3.10 depicts a causal-formative measurement model, and Figure 3.11 shows how to specify this model in ADANCO.

Whereas the older literature on variance-based SEM tended to equate formative measurement models with composite models (see, e.g., Chin, 1998; Hwang & Takane, 2004), it is only recently that scholars started recommending the multiple-indicators, multiple-causes (MIMIC) model specification for causal-formative measurement in variance-based SEM, as depicted in Figure 3.10 (Rigdon et al., 2014). For covariance-based SEM, such types of models have been the standard for decades (see, e.g., Bagozzi, 1980).

Particular care is required if a construct with a causal-formative measurement model is meant to be explained by other constructs in the model. Researchers should then apply the litmus test of whether these other constructs are theorized to directly or indirectly cause the construct. In the case of a direct causal relationship, the other constructs should be added as additional formative indicators. In the case of an indirect causal relationship, the extant formative indicators mediate the effects of the other constructs. Consequently, the researcher should include effects from the other constructs on the formative indicators in the model.

3.2.4 Single-Indicator Measurement Models

If only one observed variable is available to operationalize a concept, this is referred to as a single-indicator measurement (Diamantopoulos et al., 2012). Single-indicator measurement has the disadvantage that if an indicator is

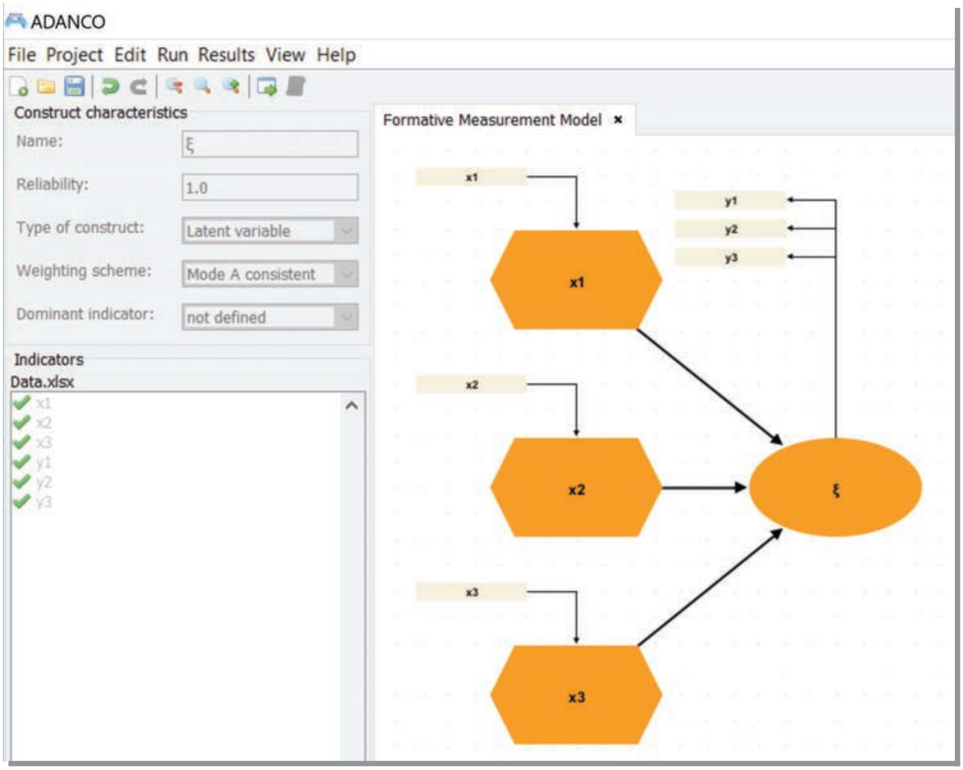

FIGURE 3.11. Causal-formative measurement in ADANCO.

regarded to contain measurement error, it is not possible to determine the amount of random measurement error in the indicator based on the available data. Instead, one has to rely on external knowledge about the observed variable's reliability and to predefine it in the model specification. Therefore, multi-indicator measurement is typically preferred over single-indicator measurement.

Although the use of a single observed variable may sound straightforward, it has some intricacies. Firstly, the researcher may have a particular model in mind. Secondly, the specification should anticipate whether a covariance-based or a variance-based estimator will be employed, because the observed variable's potential measurement error is treated differently. Figure 3.12 lists the various options available depending on the intended model and estimator.

Drawing from Figure 3.12, the following guidelines can be formulated depending on the employed estimator. If covariance-based estimators are used, then analysts should select that specification among the six shown in Figure 3.12 that optimally implements his or her intended model. In

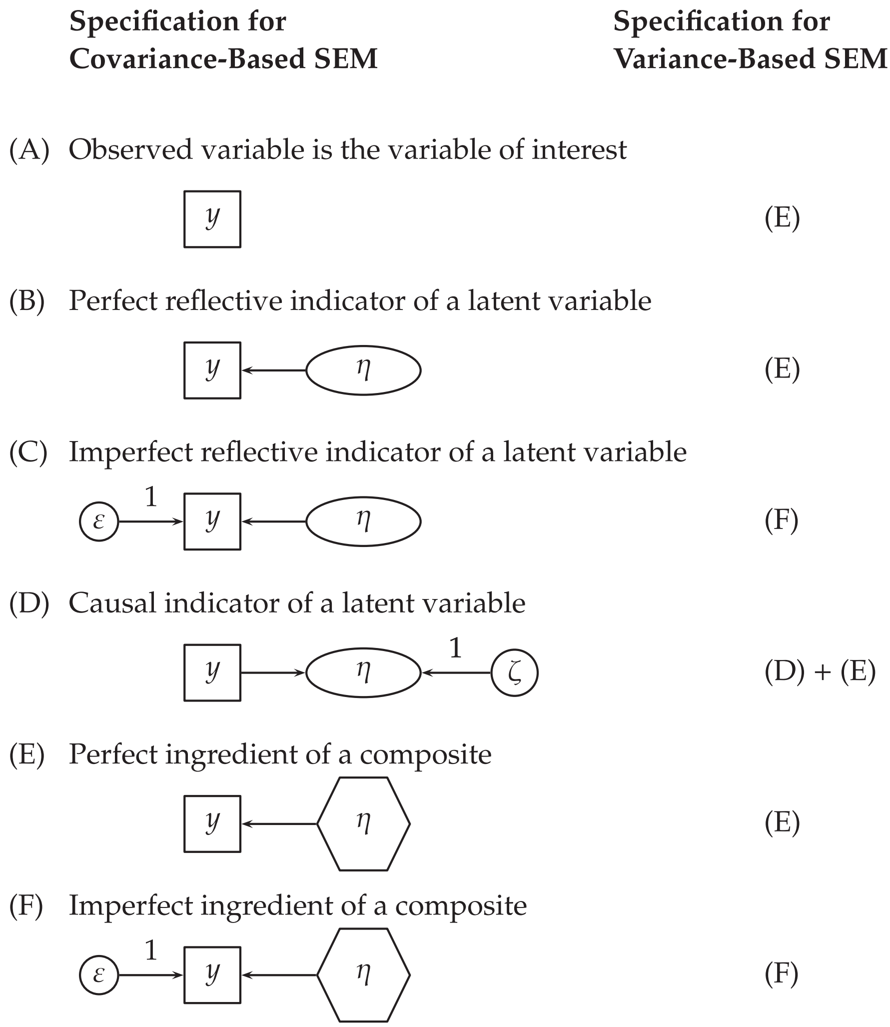

FIGURE 3.12. Possible specifications with single indicators depending on the intended model and use of estimator.

contrast, if variance-based estimators are used, then analysts can always treat the observed variable as the component of a composite. Figure 3.13 shows how single-indicator measurement can be specified in ADANCO. In particular, note the manually set value for the construct's reliability.

3.2.5 Categorical Variables

Like multiple regression, composite-based SEM requires metric data for the dependent variables. Dependent variables are the indicators of the re-

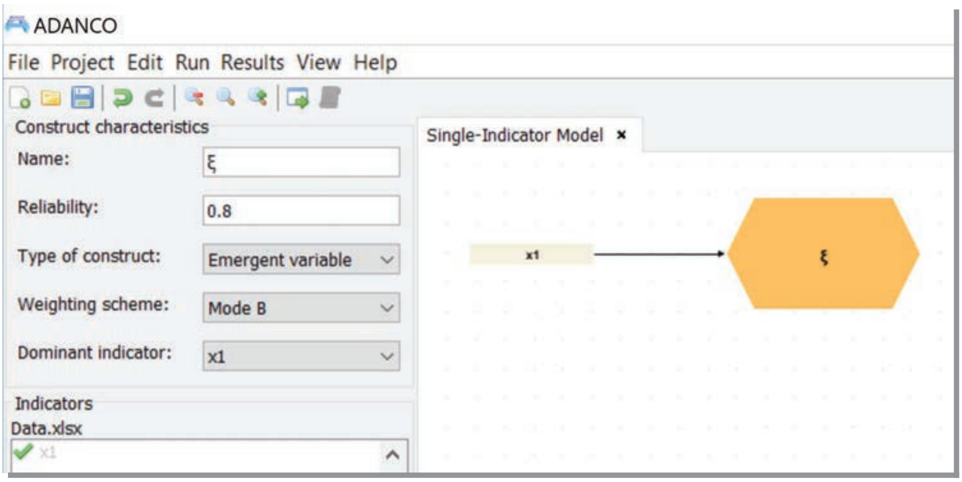

FIGURE 3.13. Single-indicator measurement in ADANCO.

flective measurement models as well as the endogenous constructs. Quasi-metric data stemming from multi-point scales such as Likert scales or semantic differential scales is also acceptable as long as the scale points can be assumed to be equidistant. Particularly if the multi-point scales have five or more scale points, the information loss compared to continuous variables is not substantial (Rhemtulla, Brosseau-Liard, & Savalei, 2012).

To some extent it is also possible to include categorical variables in a structural equation model. Categorical variables are particularly relevant for analyzing experiments (cf. Streukens, Wetzels, Daryanto, & de Ruyter, 2010) or for control variables such as industry (cf. Braojos, Benitez, & Llorens Montes, 2015) or ownership structure (cf. Chen, Wang, Nevo, Benitez, & Kou, 2017).

The estimation of structural equation models with categorical variables requires either a modification of the estimation routine (cf. Betzin & Henseler, 2005; Cantaluppi & Boari, 2016; Schuberth, Henseler, & Dijkstra, 2018b) or a special treatment of the observed categorical variables. Here, we will focus on the latter. If a categorical variable has only two levels (i.e., it is dichotomous), it can serve immediately as a construct indicator. If a categorical variable has more than two levels, it should be transformed into as many dummy variables as there are levels. A composite model is then formed out of all but one dummy variable. The remaining dummy variable characterizes the reference level. Figure 3.14 illustrates how a categorical variable "Industry" with six different levels (construction, manufacturing, chemicals, transport, finance, others) would be included in a structural

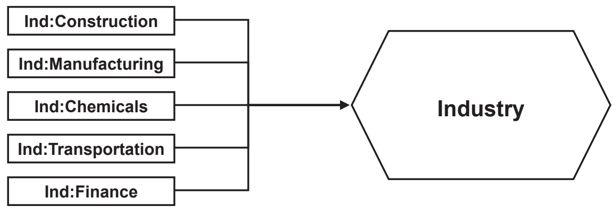

FIGURE 3.14. Specifying a categorical variable with six levels in a structural equation model using five dummy variables (the sixth level, others, serves as the reference level).

equation model. Preferably, categorical variables should only play the role of exogenous variables in a structural equation model.

3.3 The Inner Model

The inner model (also called structural model) specifies the relationships between the constructs. The size and significance of path relationships are typically the focus points of the scientific endeavors pursued in empirical research. The inner model consists of endogenous and exogenous constructs as well as the (typically linear) relationships between them. Endogenous constructs are those constructs that are at least partially explained by other constructs in the inner model. They are typically denoted by the Greek letter η. In contrast, exogenous constructs are those whose values are considered as given for the structural equation model. Their values are determined by variables that are outside of the model's scope. In order to emphasize this difference, exogenous constructs are sometimes denoted by the Greek letter ξ. In this book, they will be denoted by η as well. This allows for a simpler mathematical notation of the inner model. In matrix notation, the inner model takes the form of Equation 3.5 (see, for instance, Matsueda, 2012):

$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\zeta} \tag{3.5}$$

In this equation, $\boldsymbol{\eta}$ denotes the vector of constructs, $\boldsymbol{B}$ is the matrix of path coefficients, and $\boldsymbol{\zeta}$ is a vector of disturbance terms. It is also possible to express the inner model equation by equation. For an endogenous construct η_j, the inner model equation could then look like this:

$$\eta_j = \beta_1\eta_1 + \beta_2\eta_2 + \cdots + \beta_{j-1}\eta_{j-1} + \zeta_j \tag{3.6}$$

Next to the path coefficients, there may be covariances between those constructs that play an exogenous role within the inner model and/or the disturbance terms of the endogenous constructs. They are contained in $\boldsymbol{\Psi}$, the variance-covariance matrix of the inner model residuals and the exogenous constructs.

The inner model implies a variance-covariance matrix of constructs, $\mathbf{\Phi}$. With the help of $\boldsymbol{I}$, the identity matrix, it can be determined as follows (Bollen, 1989b; Matsueda, 2012):

$$\hat{\mathbf{\Phi}} = (\boldsymbol{I} - \boldsymbol{B})^{-1}\, \mathbf{\Psi}\, (\boldsymbol{I} - \boldsymbol{B}')^{-1} \tag{3.7}$$

Variance-based SEM is usually limited to recursive inner models (however, see Dijkstra & Henseler, 2015a, for an exception). Recursivity means that the inner model does not contain any feedback loop, and $\mathbf{\Psi}$ contains only values in the main diagonal plus the covariances between exogenous constructs. The structural equations of a recursive inner model can be sorted in such a manner that the matrix of path coefficients $\boldsymbol{B}$ is a strictly triangular matrix. If $\mathbf{\Psi}$ contains off-diagonal elements other than covariances between exogenous constructs or if an inner model contains causal loops, the inner model is non-recursive.

In order to facilitate a separate inspection of a structural equation model's outer and inner model, a two-step approach recommended by Anderson and Gerbing (1988) has become customary. In a first step, an inner model with all possible linear relationships between constructs is specified, the so-called *saturated model* (Gefen, Straub, & Rigdon, 2011). For the sake of ease, the relationships are modeled as covariances. Depending on the outer model employed, an analysis of the saturated model constitutes a CFA, a CCA (see Chapter 8), or a combination of both, i.e., a confirmatory composite/factor analysis (CCFA). The saturated model allows an undisturbed assessment of the outer model, because the inner model does not impose any constraints and hence does not induce any misfit. Once the outer model is deemed satisfactory by the analyst, in a second step, the actual inner model is specified. This model is called the *theoretical model* (Gefen et al., 2011) or the *estimated model* (Henseler, Hubona, & Ray, 2016). From a theory-testing perspective, only those estimated models are of interest that have a lower number of estimated parameters than the saturated model. Otherwise, the two models would be equivalent (they would yield the same model-implied construct variance-covariance matrix), and model fit tests are not informative about the inner model. Figure 3.15 illustrates the difference between the saturated model and the estimated model at the hand of an example.

Many different forms of relationships between constructs can be specified in the inner model: direct relationships, indirect relationships, unanalyzed relationships, no relationships, spurious relationships, bidirectional relationships, nonlinear relationships, and moderated relationships. Figure 3.16 provides a graphical overview of these common forms of inner models, enumerated from ❶ to ❽.

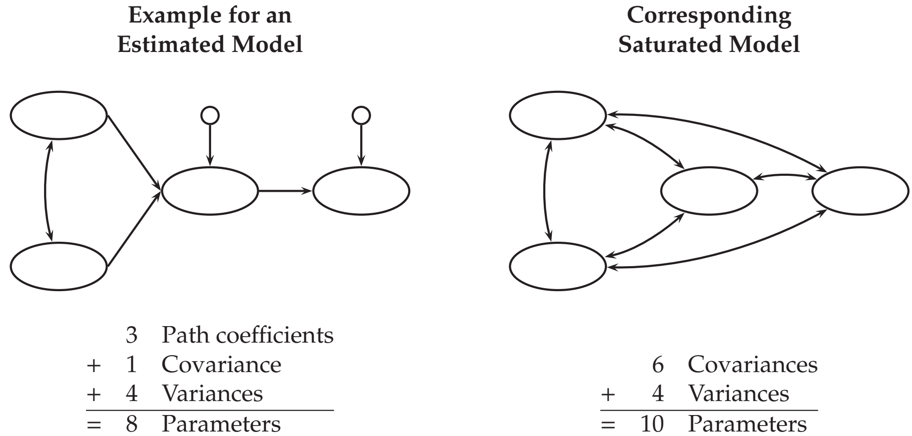

FIGURE 3.15. Estimated model vs. saturated model.

❶ Direct Relationships

Direct relationships are specified if a construct η_1 is hypothesized to have a direct linear effect on another construct η_2 if all other constructs in the model that are hypothesized to impact η_2 are kept constant (i.e., ceteris paribus). A positive direct relationship captures a hypothesis of the type "The higher η_1, the higher η_2 (ceteris paribus)," and a negative direct relationship captures a hypothesis of the type "The higher η_1, the lower η_2 (ceteris paribus)." Direct relationships correspond to a single element of the path coefficient matrix $\boldsymbol{B}$.

All constructs affected by direct relationships, i.e., all endogenous constructs, are assumed to have disturbance terms. In variance-based SEM, these disturbance terms are taken into account automatically and therefore do not have to be explicitly specified. In contrast, many implementations of covariance-based SEM require the explicit specification of disturbance terms.

❷ Indirect Relationship

Indirect relationships entail that a construct η_1 does not directly influence an endogenous construct η_3, but only indirectly via another construct η_2. The latter plays the role of a mediator variable (or simply mediator). Mediation analysis covers this type of inner model. It is explained in Chapter 9. Indirect relationships encompass a set of at least two elements of the path coefficient matrix $\boldsymbol{B}$.

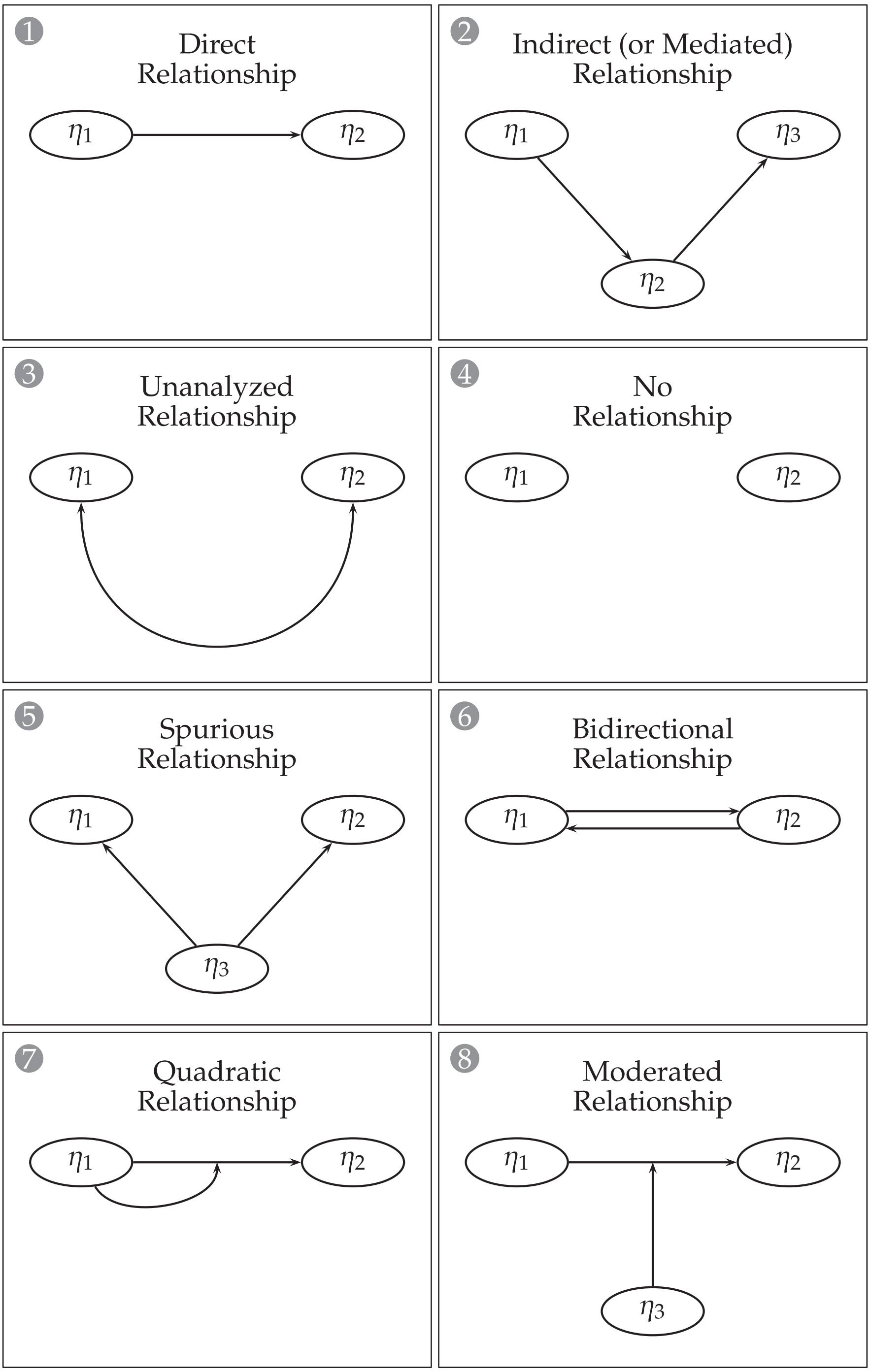

FIGURE 3.16. Examples of relationships between constructs.

❸ Unanalyzed Relationship

In the case of unanalyzed relationships, a model allows for concomitant variation between two constructs η_1 and η_2, but makes no effort to explain the underlying reason why the two constructs covary. Unanalyzed relationships correspond to a single element of $\boldsymbol{\Psi}$, the variance-covariance matrix of the inner model disturbance terms and the exogenous constructs. That means that unanalyzed relationships can only be specified between exogenous constructs and/or disturbance terms. An endogenous variable cannot have unanalyzed relationships; instead, its disturbance term should be specified to have the unanalyzed relationship.

Most software implementations of variance-based SEM do not allow to explicitly specify unanalyzed relationships. Only the covariances between exogenous constructs are automatically specified. In contrast, most software implementations of covariance-based SEM allow one to specify other unanalyzed relationships in an inner model.

❹ No Relationship

If two constructs η_1 and η_2 are specified to have no relationship and at least one of them is endogenous, then the corresponding elements of the path coefficient matrix $\boldsymbol{B}$ are zero. If both constructs are exogenous, the graphical model ❹ depicted in Figure 3.16 is decoded differently depending on the type of SEM employed. If a model is specified for covariance-based SEM, "no relationship" between two exogenous constructs means that these two constructs are not allowed to covary, and the corresponding element of $\boldsymbol{\Psi}$ is fixed to zero. In contrast, if a model is specified for variance-based SEM, "no relationship" between two exogenous constructs means that this relation is unanalyzed. Consequently, the corresponding element of $\boldsymbol{\Psi}$ is freely estimated.

❺ Spurious Relationship

Two constructs η_1 and η_2 have a spurious relationship if they exhibit concomitant variation although they do not cause each other. Instead, they share a common cause η_3, which is responsible for the shared variance of η_1 and η_2. If a researcher specified a direct effect from η_1 to η_2 without accounting for η_3, he or she would observe a spurious effect of η_1 on η_2.

❻ Bidirectional Relationship

A bidirectional relationship means that a construct η_1 simultaneously affects and is affected by a construct η_2. By definition, both η_1 and η_2 are endogenous constructs. Although the visual representation of a bidirectional relationship may resemble an unanalyzed relationship (both have two arrowheads), their meaning is thus fundamentally different.

In addition to the two path coefficients, which constitute elements of the path coefficient matrix $\boldsymbol{B}$, it is indispensable to free the covariance between the two endogenous variable's disturbance terms. This corresponds to a single element of $\boldsymbol{\Psi}$. Inner models with one or more bidirectional relationships are non-recursive. In most implementations of variance-based SEM it is not possible to specify inner models with a bidirectional relationship; the only exception is `cSEM`. Instead, it may be advantageous to rely on covariance-based SEM (see, e.g., Benitez, Ray, & Henseler, 2018, for a combined use of variance- and covariance-based SEM).

❼ Quadratic Relationship

In principle, it is possible for inner models to leave the comfortable realm of linear relations. In many cases, simple polynomial extensions can help model several forms of nonlinearity (see Dijkstra & Henseler, 2011; Henseler et al., 2012, and Section 11.5). Already the simplest polynomial extension, a quadratic term, permits to model various nonlinear patterns. Section 11.5 provides a deeper discussion of nonlinear effects.

❽ Moderated Relationship

A particular form of nonlinearity is moderation (also called "interaction"). One refers to an interaction effect if a focal effect is not constant, but depends on the level of another construct in the model. Several approaches for modeling interaction effects using composite-based SEM have been proposed (e.g., Dijkstra & Schermelleh-Engel, 2014; Fassott et al., 2016; Henseler & Chin, 2010; Henseler & Fassott, 2010). Chapter 11 explains how to model moderated relationships.

3.4 Software Tutorial: Model Specification

The task of this tutorial is to specify a model that is meant to answer the question to what extent agricultural inequality and industrial development impact political stability. The dataset was initially compiled by Russett (1964), discussed and reprinted by Gifi (1990), and partially transformed by Tenenhaus and Tenenhaus (2011).[4] For 47 countries, it contains the following variables[5]:

- Indicators of agricultural inequality:
  - *gini*: The Gini index of concentration.
  - *farm*: The percentage of landholders who collectively occupy one-half of all the agricultural land (starting with the farmers with the smallest plots of land and working toward the largest).
  - *rent*: The percentage of the total number of farms that rent all their land. Transformation: $\ln(\# + 1)$.
- Indicators of industrial development:
  - *gnpr*: The 1955 gross national product per capita in U.S. dollars. Transformation: $\ln(\#)$.
  - *labo*: The percentage of the labor force employed in agriculture. Transformation: $\ln(\#)$.
- Indicators of political stability:
  - *inst*: Instability of personnel based on the term of office of the chief executive. Transformation: $\exp(\# - 16.3)$.
  - *ecks*: The total number of politically motivated violent incidents, from plots to protracted guerrilla warfare. Transformation: $\ln(\# + 1)$.
  - *deat*: The number of people killed as a result of internal group violence per 1,000,000 people. Transformation: $\ln(\# + 1)$.
  - *stab*: One if the country has a stable democracy, and zero otherwise.
  - *dict*: One if the country experiences a dictatorship, and zero otherwise.

The raw data as well as the ADANCO project file can be downloaded from the companion website:

`https://www.guilford.com/henseler-materials`

[4] For more details on the genesis of the data see Russett (1964); for more details on the data imputation see Tenenhaus and Tenenhaus (2011); for more details on the data transformations see Tenenhaus and Tenenhaus (2011) and Gifi (1990).

[5] In the transformation, # denotes the original variable. Three missing values were imputed with values suggested by Tenenhaus and Tenenhaus (2011).

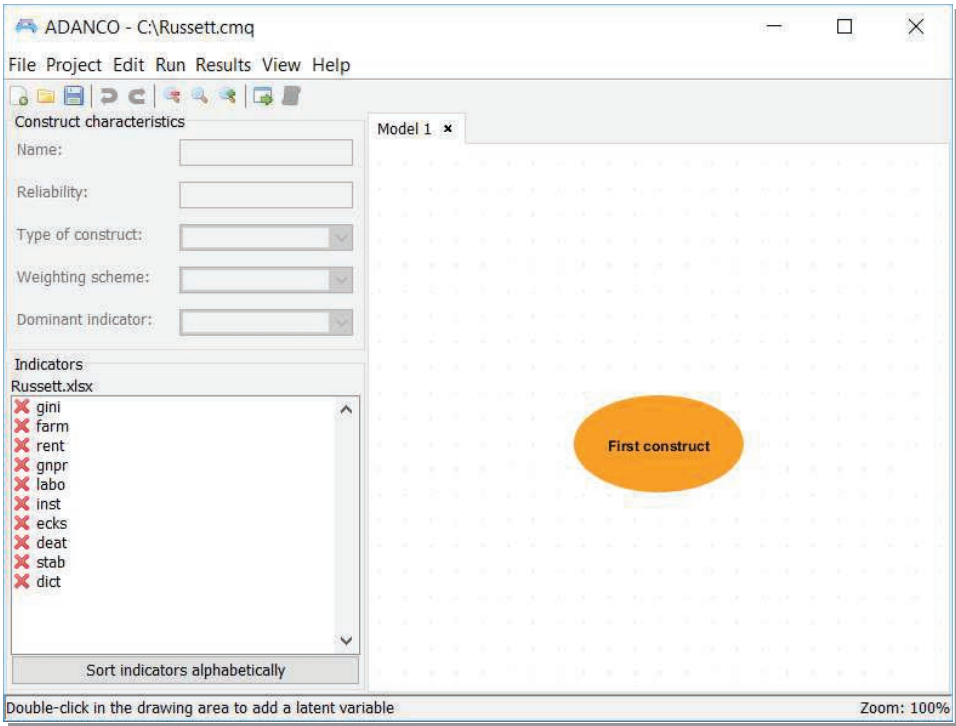

FIGURE 3.17. Specifying structural equation models in ADANCO: The model panel.

3.4.1 Specifying Structural Equation Models in ADANCO

The data file `russett.xlsx` can be imported as described in Section 1.4 (page 19). Once ADANCO has recognized the data, it shows the model panel as depicted in Figure 3.17.

When ADANCO creates a new project, the first construct will always be drawn automatically. By double-clicking on constructs one can **rename constructs**. Alternatively, one can **mark constructs** by a single mouse-click and press F2. We can give this construct the name *Political instability*.

In order to **assign indicators** to a construct, one selects one or more indicators from the list of indicators (eventually pressing the SHIFT or CTRL key to select more than one indicator at once) and moves them through drag and drop onto the intended construct. To **move indicators**, mark one indicator of a construct and move it keeping the mouse button pressed. It will move together with all other indicators of that construct. In the current case, we should select the indicators *inst*, *ecks*, *deat*, *stab*, and *dict* and assign them to the construct *Political instability*. Once all indicators of a construct are assigned, one should define one **dominant indicator**. The dominant indicator should be a variable of which the researcher knows with sufficient certainty that the correlation between the indicator and the construct should be positive (see also Chapter 4).

Researchers should make a conscious choice of the auxiliary theory employed for each construct. Shall a construct be modeled as a latent variable or as an emergent variable? Per default, constructs will be modeled as latent variables, the dominant type of constructs in social sciences. If one would like to model a construct as an emergent variable, one should mark this construct by a single mouse click and change the type of construct in the upper left of the shell to "emergent variable." The properties of a marked construct can also be changed in the upper left of the shell. Alternatively, right-clicking on a construct opens a **context menu**, which permits one to change the **construct properties**. In line with Tenenhaus and Tenenhaus (2011), we propose to specify *Political instability* as an emergent variable. As a consequence of this specification, the construct will obtain the shape of a hexagon and the arrows of this construct's outer model will point from the indicators to the construct.

An even quicker way to simultaneously specify constructs and assign indicators to them is to select a block of indicators and move them through drag and drop onto a free place in the model panel. ADANCO will then create a new latent variable, assign the selected indicator to it, and name the construct according to those initial letters that all of its indicators have in common. We use this mechanism twice. Firstly, we select the indicators *gini*, *farm*, and *rent* and drop them onto the model panel. We rename the *New Construct* to *Agricultural inequality*, and change its type to "emergent variable." Secondly, we proceed in the same way with the remaining indicators, i.e., *gnpr* and *labo*. We rename the *New Construct* to *Industrial development*, and change its type to "emergent variable," too.

In order to **draw paths**, i.e., arrows of the inner model, one must click on the source of the arrow (i.e., the independent variable), keep the mouse button pressed, and release it on the target of the arrow (i.e., the dependent variable). Please note that to draw arrows of the inner model, all constructs must be unmarked.

From Version 2.2 on, ADANCO supports labels for structural paths. If one right-clicks on a structural path, a context menu provides the opportunity to assign a label. For instance, if a certain effect corresponds to the first hypothesis of a scientific paper, it would make sense to call this effect "H1."

It is a good practice of conducting statistical analyses to document what has been done. ADANCO facilitates this provision of transparency in two ways. Firstly, analysts can specify a **model title** by right-clicking into the model panel and choosing the option "Add title." Secondly, analysts can

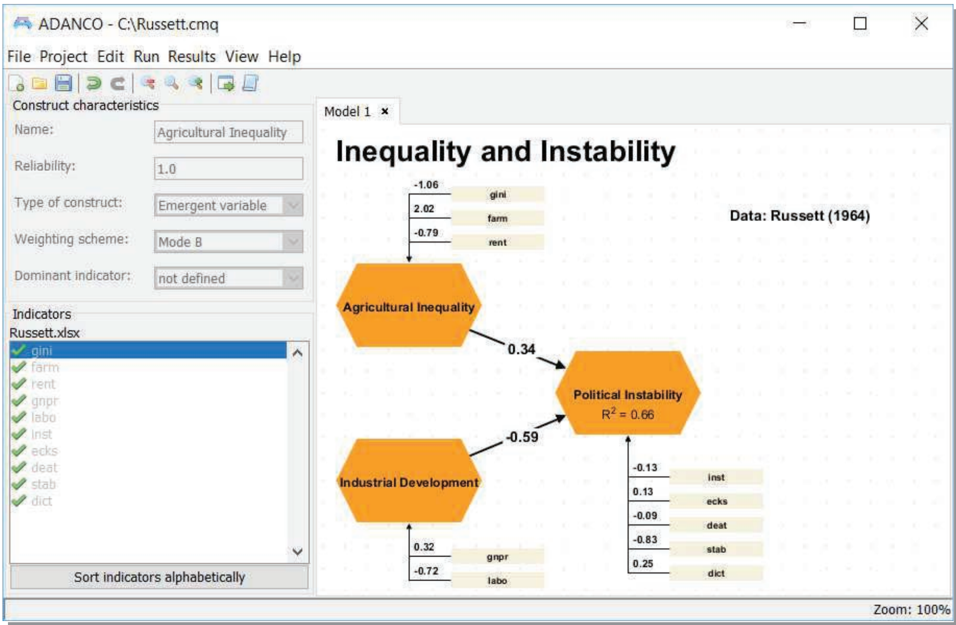

FIGURE 3.18. Specifying structural equation models in ADANCO: A model based on Russett (1964).

add comments by right-clicking into the model panel and choosing the option "Add comment." While in the heat of the moment adding titles and comments may appear like a waste of time, analysts will usually embrace titles and comments once they reopen a past analysis after a long time again, and this additional information helps to recapture what the analysis was all about.

To **delete objects** of the model such as constructs, indicators, paths, title, or comments, simply mark the respective object and press DELETE. Alternatively, one can right-click on the object and select the "Delete" option in the context menu.

As soon as a valid structural equation model has been specified, ADANCO will show some selected estimation results: loadings for the reflective measurement models, weights for the composite models, coefficients of determination (R^2 values) for all endogenous constructs, and the path coefficients. The final model is shown in Figure 3.18.

Defining Parameters

ADANCO 2.3.1 allows one to define additional parameters. These parameters are derived from a pair of existing parameters that are exposed to an arithmetic function such as summation, subtraction, multiplication, or di-

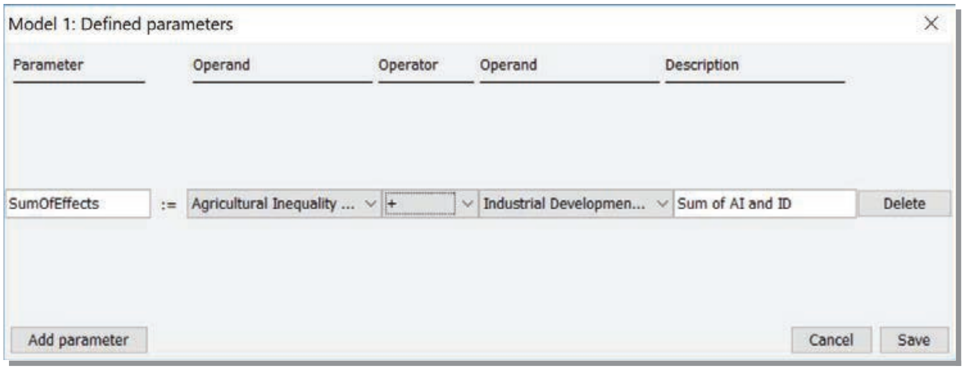

FIGURE 3.19. ADANCO provides the option to define parameters.

vision. The existing parameters can be path coefficients or already defined parameters. This permits a cascaded definition of new parameters. For instance, to define a new parameter $Q := \beta_1 \cdot \beta_3 + \beta_2 \cdot \beta_4$, one could define two auxiliary parameters $Q_1 := \beta_1 \cdot \beta_3$ and $Q_2 := \beta_2 \cdot \beta_4$ and use them to finally define $Q := Q_1 + Q_2$. Figure 3.19 shows how a parameter sum is defined using the corresponding ADANCO dialogue window.

Derived parameters can be useful to answer specific research questions. Particularly four applications are worth mentioning: parameter sums, parameter differences, indirect relationships, and parameter ratios.

A parameter sum is a relevant piece of information when a researcher expects two effects to cancel each other out. The hypothesis that two effects have the same magnitude but the opposite sign is equivalent to the hypothesis that the sum of two effects is equal to zero. For instance, Eggert, Henseler, and Hollmann (2012) applied a parameter sum in order to test a structural equation model derived from balance theory (Heider, 1958). They formulated the hypothesis that the difference in loyalty toward two parties determines to which party someone will remain loyal if the circumstances no longer allow him or her to stay loyal to both parties simultaneously.

Parameter differences are relevant if an analyst wants to compare parameters, e.g., as part of a ranking. According to Rodríguez-Entrena, Schuberth, and Gelhard (2018, p. 59), the "difference between two parameter estimates might be particularly valuable when model estimates are proposed to guide decision makers in handling budget constraints (e.g., selection of marketing strategies, success factors or investment in alternative instruments of innovation, process, and product, etc.)." Sometimes, the finding that one effect is significantly different from zero whereas another one is not may inspire analysts to conclude that the two effects differ from each other. However, such a conclusion would be premature. As Gelman and Stern (2006, p. 328) emphasize, “the difference between ‘significant’ and ‘not significant’ is not itself statistically significant.” Instead, analysts should test the difference between the two effects. If the difference between the two effects is statistically different from zero, there is empirical support that one effect has a different magnitude than the other one.

Parameter products play an important role as parameters for indirect relationships. Indirect relationships form the core of mediation analysis, which is explained in Chapter 9.

Parameter ratios are useful if an analyst would like to express a parameter in relative terms to some other parameter(s). An example of a parameter ratio is the variance accounted for (VAF), a coefficient to quantify complementary mediation (see Chapter 9).

Limitations and Implicit Specifications

Some model specifications are made automatically and cannot be manually changed: Measurement errors are assumed to be uncorrelated with all other variables and errors in the model; disturbance terms are assumed to be orthogonal to their predictor variables and to each other; correlations between exogenous variables are free. Because these specifications hold across models, it has become customary not to draw measurement errors and their correlations in PLS path models. As a consequence, measurement models in variance-based SEM may appear less detailed than those of covariance-based SEM; however, some specifications are implicit and are simply not visualized. Since ADANCO 2.3.1 does not allow either constraining or freeing factor models’ error correlations, these model elements are not drawn. Inner models must be recursive, i.e., there should be no loop. In ADANCO 2.3.1, inner models can consist of several unconnected pieces, i.e., constructs need not be connected with other constructs. However, in order to facilitate identification, it is strongly recommended to have each construct be connected at least with one other construct (see Chapter 4). Finally, construct names must be unique.

3.4.2 Specifying Structural Equation Models in `cSEM`

Model specification in `cSEM` makes use of a syntax that was first introduced by the `lavaan` package (Rosseel, 2012). It employs four operators:

=~ serves to specify reflective measurement models; it assigns observed variables to a latent variable.

<~ serves to specify composite models; it assigns observed variables

to an emergent variable.

~ serves to specify the inner model; it defines on which independent variables a construct shall be regressed.

~~ serves to specify correlations among measurement errors within a block of indicators.

The model of Russett can then be specified as follows:

```
model_Russett = ' # Specify the composite models
                AgrIneq <~ gini + farm + rent
                IndDev  <~ gnpr + labo
                PolInst <~ inst + ecks + deat + stab + dict

                # Specify the relation among the
                # emergent variables
                PolInst ~ AgrIneq + IndDev
                '
```

To load the dataset, the following code snippet can be used (eventually, the path needs to be adjusted):

```
Russett <- as.data.frame(readxl::read_excel("C:/Russett.
   xlsx"))
```

To estimate the model, we load the `cSEM` package and employ its `csem()` function:

```
library(cSEM)
```

```
## Attaching package: 'cSEM'
## The following object is masked from 'package:stats':
##
##     predict
```

```
out <- csem(.data = Russett, .model = model_Russett,
            # To reproduce the ADANCO results
              .PLS_weight_scheme_inner = 'factorial',
              .tolerance = 1e-06
           )
```

Finally, we retrieve the results:

```
summarize(out)
```

```
## --------------------------------------------------------------------------
## ------------------------------- Overview ---------------------------------
##
## General information:
## ------------------------
## Estimation status                  = Ok
## Number of observations             = 47
## Weight estimator                   = PLS-PM
## Inner weighting scheme             = factorial
## Type of indicator correlation      = Pearson
## Path model estimator               = OLS
## Second order approach              = NA
## Type of path model                 = Linear
## Disattenuated                      = No
##
## Construct details:
## ------------------
## Name     Modeled as  Order        Mode
##
## AgrIneq  Composite   First order  modeB
## IndDev   Composite   First order  modeB
## PolInst  Composite   First order  modeB
##
## ------------------------------- Estimates --------------------------------
##
## Estimated path coefficients:
## ============================
## Path               Estimate  Std. error  t-stat.  p-value
## PolInst ~ AgrIneq    0.3379          NA       NA       NA
## PolInst ~ IndDev     0.5926          NA       NA       NA
##
## Estimated loadings:
## ===================
## Loading            Estimate  Std. error  t-stat.  p-value
## AgrIneq =~ gini      0.5365          NA       NA       NA
## AgrIneq =~ farm      0.6715          NA       NA       NA
## AgrIneq =~ rent     -0.2644          NA       NA       NA
## IndDev  =~ gnpr     -0.9094          NA       NA       NA
## IndDev  =~ labo      0.9824          NA       NA       NA
## PolInst =~ inst      0.1923          NA       NA       NA
## PolInst =~ ecks      0.6310          NA       NA       NA
## PolInst =~ deat      0.5240          NA       NA       NA
## PolInst =~ stab     -0.9685          NA       NA       NA
## PolInst =~ dict      0.7395          NA       NA       NA
##
## Estimated weights:
```

```
## ==================
## Weights            Estimate  Std. error  t-stat.  p-value
## AgrIneq <~ gini     -1.0570          NA       NA       NA
## AgrIneq <~ farm      2.0241          NA       NA       NA
## AgrIneq <~ rent     -0.7859          NA       NA       NA
## IndDev  <~ gnpr     -0.3228          NA       NA       NA
## IndDev  <~ labo      0.7191          NA       NA       NA
## PolInst <~ inst     -0.1337          NA       NA       NA
## PolInst <~ ecks      0.1287          NA       NA       NA
## PolInst <~ deat     -0.0854          NA       NA       NA
## PolInst <~ stab     -0.8337          NA       NA       NA
## PolInst <~ dict      0.2459          NA       NA       NA
##
## Estimated construct correlations:
## =================================
## Correlation        Estimate  Std. error  t-stat.  p-value
## AgrIneq ~~ IndDev    0.4875          NA       NA       NA
##
## Estimated indicator correlations:
## =================================
## Correlation        Estimate  Std. error  t-stat.  p-value
## gini ~~ farm         0.9376          NA       NA       NA
## gini ~~ rent         0.3873          NA       NA       NA
## farm ~~ rent         0.4599          NA       NA       NA
## gnpr ~~ labo        -0.8156          NA       NA       NA
## inst ~~ ecks         0.3261          NA       NA       NA
## inst ~~ deat         0.0835          NA       NA       NA
## inst ~~ stab        -0.3434          NA       NA       NA
## inst ~~ dict         0.0198          NA       NA       NA
## ecks ~~ deat         0.6277          NA       NA       NA
## ecks ~~ stab        -0.6034          NA       NA       NA
## ecks ~~ dict         0.3920          NA       NA       NA
## deat ~~ stab        -0.4905          NA       NA       NA
## deat ~~ dict         0.5321          NA       NA       NA
## stab ~~ dict        -0.5893          NA       NA       NA
##
## -------------------------------- Effects --------------------------------
##
## Estimated total effects:
## ========================
## Total effect       Estimate  Std. error  t-stat.  p-value
## PolInst ~ AgrIneq    0.3379          NA       NA       NA
## PolInst ~ IndDev     0.5926          NA       NA       NA
##
## Estimated indirect effects:
## ===========================
## Indirect effect    Estimate  Std. error  t-stat.  p-value
## NA                       NA          NA       NA       NA
##
## -------------------------------------------------------------------------
```