TECHNICAL REPORT
R-276
January 2001

Parameter Identification: A New Perspective (Second Draft)

Judea Pearl
Cognitive Systems Laboratory
Computer Science Department
University of California, Los Angeles, CA 90024
judea@cs.ucla.edu

1 Introduction and Preliminary Terminology:

A model M is a set of structural equations with (zero or more) free parameters, $p, q, r, \ldots$, that is, unknown parameters whose values are to be estimated from a combination of assumptions and data. The assumptions embedded in such a model are of several kinds: (1) zero (or fixed) coefficients in some equations, (2) equality or inequality constraints among some of the parameters and (3) zero covariance relations among error terms (also called disturbances). Some of these assumptions are encoded implicitly in the equations (e.g., the absence of certain variables in an equation), while others are specified explicitly, using expressions such as: $p = q$ or $cov(e_i, e_j) = 0$.

An instantiation of a model M is an assignment of values to the model's parameters; such instantiations will be denoted as m_1, m_2 etc. The value of parameter p in instantiation m_1 of M will be denoted as $p(m_1)$. Every instantiation m_i of model M gives rise to a unique covariance matrix $\sigma(m_i)$, where σ is the population covariance matrix of the observed variables.

Definition 1 *(Parameter identification)*
A parameter p in model M is identified if for any two instantiations of M, m_1 and m_2, we have:

$$p(m_1) = p(m_2) \text{ whenever } \sigma(m_1) = \sigma(m_2)$$

Definition 2 *(Model identification)*
A model M is identified iff all parameters of M are identified

Definition 3 *(Model overidentification and justidentification)*
A model M is overidentified *if (1) M is identified and (2) M imposes some constraints on σ, that is, there exists a covariance matrix σ' such that $\sigma(m_i) \neq \sigma'$ for every instantiation m_i of M. M is* justidentified *if it is identified and not overidentified, that is, for every σ' we can find an instantiation m_i such that $\sigma(m_i) = \sigma'$.*

Definition 3 highlights the desirable aspect of overidentification – *testability*. It is only by violating its implied constraints that we can falsify a model, and it is only by escaping the threat of such violation that a model attains our confidence.

Traditionally, however, model overidentification has rarely been determined by direct examination of the model's constraints[1] but, rather indirectly, by attempting to solve for the model parameters and discovering parameters that can be expressed as two or more distinct[2] functions of σ, for example, $p = f_1(\sigma)$ and $p = f_2(\sigma)$. This immediately leads to a constraint $f_1(\sigma) = f_2(\sigma)$ which, according to Definition 3, renders the model overidentified, since every σ' for which $f_1(\sigma') \neq f_2(\sigma')$ must be excluded by the model.

This indirect method of determining model overidentification (hence model testability) has led to a tradition of labeling the *parameters* themselves as overidentified or justidentified; parameters that were found to have more than one solution were labeled overidentified, those that were not found to have more than one solution were labeled justidentified, and the model as a whole was classified according to its parameters. In the words of Bollen (1989, p. 90) "A model is overidentified when each parameter is identified and at least one parameter is overidentified. A model is exactly identified when each parameter is identified but none is overidentified."

Although no formal definition of parameter overidentification has been formulated, save for the informal requirement of having "more than one solution" [MacCallum, 1995, p. 28] or of being "determined from σ in different ways" [Joreskog, 1979, p. 108], the idea that parameters themselves carry the desirable feature of being overidentified, and that this desirable feature may vary from parameter to parameter became deeply entrenched in the literature. Paralleling the desirability of overidentified models, most researchers expect overidentified parameters to be *more testable* than justidentified parameters. Typical of this expectation is the economists' search for two or more instrumental variables for a given parameter, and the development of the Wu-Hausman test for deciding if the estimates induced by two instruments are the same [Bowden and Turkington, 1984].

Unfortunately, the standard conception of overidentification does not support these expectations. If we take literally the criterion that a parameter is overidentified when it can be expressed as two or more distinct functions of the covariance matrix σ, we get the untenable conclusion that, if one parameter is overidentified, then every other (identified) parameter in the model must also be overidentified. Indeed, whenever an overidentified model induces a constraint $g(\sigma) = 0$, it also yields (at least) two solutions for any identified parameter $p = f(\sigma)$, because we can always obtain a second, distinct solution for p by writing $p = f(\sigma) - g(\sigma)t(\sigma)$, with arbitrary $t(\sigma)$. Thus, to capture the expectations and practice of

[1] This is unfortunate, but understandable, given the meager mathematical tools available to researchers in this area. Current techniques permit the direct reading of model constraints from the model's graph [Pearl, 2000, p. 140–149] and direct testing of those constraints [Shipley, 2000].

[2] Two functions $f_1(\sigma)$ and $f_2(\sigma)$ are distinct if there exists a σ' such that $f_1(\sigma') \neq f_2(\sigma')$.

most researchers, additional qualifications must be formulated to supplement and refine the traditional notion of overidentification. Such qualifications were probably assumed by the original definers of this notion, but were kept implicit for various historical reasons.

The next definition explicates those dormant qualifications and makes them operational.

Definition 4 *(Parameter overidentification)*
A parameter p is overidentified if there are two or more distinct sets of logically independent assumptions in M such that:

1. *each set is* sufficient *for deriving the value of p as a function of σ, $p = f(\sigma)$*
2. *each set induces a* distinct *function $p = f(\sigma)$,*
3. *each assumption set is* minimal, *that is, no proper subset of those assumptions is sufficient for the derivation of p.*

Definition 4 differs from the standard criterion in two important aspects. First, it interprets multiplicity of solutions in terms of distinct sets of assumptions underlying those solutions, rather than distinct functions from σ to p. Second, Definition 4 insists on the sets of assumptions being minimal, thus ruling out redundant assumptions that do not contribute to the derivation of p.

Definition 5 *(Degree of overidentification)*
A parameter p (of model M) is identified to degree k (read: k-identified) if there are k distinct sets of assumptions in M that satisfy the conditions of Definition 4. p is said to be m-corroborated if there are m distinct sets of assumptions in M that satisfy conditions (1) and (3) of Definition 4, possibly yielding $k < m$ distinct estimands for p.

Definition 6 *A parameter p (of model M) is said to be* justidentified *if it is identified to the degree 1 (see Definition 5) that is, there is only one set of assumptions in M that meets the conditions of Definition 4.*

Generalization to non-linear, non-Gaussian systems is straightforward. Parameters are replaced with "queries" and σ is replaced with the density function over the observed variables.

2 Properties of Assumption-Based Overidentification

Theorem 1 *Every identified model containing an overidentified parameter must be overidentified in the sense of Definition 3.*

The proof is immediate from the discussion after Definition 3.

Note that the converse to Theorem 1 does not hold – a model may be overidentified while containing no overidentified parameter. This is illustrated in the following example.

Example 1 *Consider the model*

$$x = e_x,\ \ y = e_y,\ \ cov(e_x, e_y) = 0$$

This model comprises two free parameters, $Var(e_x)$ and $Var(e_y)$, both are justidentified, and, yet, imposing the constraint $cov(x, y) = 0$ on σ, the model is classified as overidentified. To see that $Var(e_x)$ is justidentified, note that assumption $x = e_x$ is sufficient and minimal for deriving $Var(e_x) = Var(x)$; any other set of assumptions would be nonminimal. The same goes for $Var(y)$.

Let us examine now how Definition 4 resolves difficulties in the traditional definitions of overidentification.

Example 2 *Consider the model in Figure 1.*

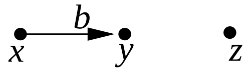

Figure 1:

It stands for the equations:

$$\begin{aligned} x &= e_x \\ y &= bx + e_y \\ z &= e_z \end{aligned}$$

together with the assumptions $cov(e_i, e_j) = 0,\ i \neq j$. The constraints induced by this model are:

$$cov(x, z) = cov(z, y) = 0$$

and the best estimate of b is the regression coefficient $R_{yx} = cov(x, y)/var(x)$. Parameter b is clearly justidentified (and so are all other parameters of M), because the assumption set

$$\begin{aligned} x &= e_x \\ y &= bx + e_y \\ cov(e_x, e_y) &= 0 \end{aligned}$$

is both minimal and sufficient for deriving $b = R_{yx}$, and no other set has these properties.[3]

This classification matches our intuition that a parameter (b) should not turn from justidentified to overidentified merely by adding another variable (z) that is totally independent of all other variables. Note, however, that the introduction of such a variable renders the model overidentified (per Definition 3) since it constrains σ with the restrictions $cov(x, z) = cov(z, y) = 0$. The importance of using minimal and sufficient sets of assumptions rests with recognizing these added constraints as superfluous for the derivation of b, thus retaining the just-identified status of b.

[3]This assumes that we have compelling theoretical knowledge to rule out any causal influence of z on x or y. For example, z may be the outcome of a coin flip.

The standard definitions of parameter overidentification are quite ambivalent as to the classification of b in this example. If we take literally the criterion that a parameter is overidentified when it "can be determined from σ *in different ways*" [Joreskog, 1979, p. 108] or when "there is more than one distinct such solution" [MacCallum, 1995, p. 28] then we face the untenable conclusion that b is overidentified, because it can be expressed as several distinct functions of σ. For example

$$\begin{aligned} b &= R_{yx} \\ b &= R_{yx} - cov(x,z) \\ b &= R_{yx} - cov(x,z) - cov(z,y) \end{aligned}$$

and many more...

The same difficulty prevails when we attempt to refine the standard definition by insisting that only functions involving distinct element of σ be considered "distinct." Such refinement was informally expressed in James et al. (1982): "... because the solutions for the other parameters in each of these equations are obtainable *using other elements of* σ [my italics]. Parameter α_{21} is therefore very much overidentified" (page 133). Example 2 demonstrates how parameter b can be expressed in terms of several distinct functions of σ, each involving different elements of σ and, still, we do not wish to characterize b as overidentified.

One way of exposing the irrelevance of z is to insist on deriving b from a minimal subset of the *reduced equations*, that is, the unique set of equations that result when we solve each element of σ as a function of the model parameters (i.e., the path coefficients and the variance-covariance elements of the error terms). Given this set of $n(n+1)/2$ reduced equations, we seek minimal subsets of equations that are sufficient for solving b in terms of σ. If more than one solution can be found this way, we classify b as overidentified, else, it is justidentified.

In our example, the reduced equations read:

$$\begin{aligned} var(x) &= var(e_x) \\ var(y) &= b^2 \cdot var(e_x) + var(e_x) \\ var(z) &= var(e_z) \\ cov(x,y) &= b \cdot var(x) \\ cov(x,z) &= cov(e_x, e_z) = 0 \\ cov(z,y) &= cov(e_y, e_z) = 0 \end{aligned}$$

Since the fourth equation is sufficient (and minimal) for deriving b in terms of σ, and since no other set of equation enjoys these properties, we proclaim b just-identified.

This criterion of parameter overidentification, which can be traced back to Werts and Linn (1970)[4], can be viewed as an approximation of Definition 4; it also uses the principle of minimal sufficiency, but it applies the principle to the set of reduced equations, instead

[4] Werts and Linn illustrated this criterion by example: "By deleting, one at a time, Equations 7, 8, 9, and 10 from the set of Equations 5 to 10, four distinct subsets of five equations in five unknowns are obtained, each of which leads to a solution..." [ibid, p. 195]. Stan Mulaik (email message posted August 14) has formulated this criterion in general form, as expressed here.

of the set of basic assumptions. A related, and possibly equivalent criterion applies the minimality principle directly to the solutions of the parameter in question, insisting that distinct solutions be based on minimal sets of elements (of σ). In our example, the solutions $b = R_{yx} - cov(x, z)$ and $b = R_{yx} - cov(x, z) - cov(z, y)$ would be discarded, because each comprises a superset of elements (of σ) that appear in the solution $b = R_{yx}$.[5] The next two examples show why both approximations are too coarse to serve as a basis for determining the identification status of parameters.

Example 3 *Consider the model in Figure 2.*

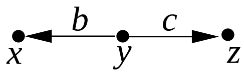

Figure 2:

This model stands for the equations:

$$\begin{aligned} y &= e_y \\ x &= by + e_x \\ z &= cy + e_z \end{aligned}$$

together with the assumptions $cov(e_i, e_j) = 0,\ i \neq j$. *The constraints induced by this model are:*

$$R_{zx} = R_{yx} R_{zy}$$

In this example, both parameters (b and c) would be classified as overidentified by the conventional SEM definition, but are rendered justidentified by Definition 4. Indeed, if we attempt to derive expressions for b and c, we obtain the equations

$$\begin{aligned} R_{xy} &= b \\ R_{xz} &= bc \\ R_{zy} &= c \end{aligned}$$

From these equations, b can be derived in two distinct ways:

$$\begin{aligned} b &= R_{yx} \\ b &= R_{zx}/R_{zy} \end{aligned}$$

and c, likewise, can be derived in two distinct ways:

$$\begin{aligned} c &= R_{zy} \\ c &= R_{zx}/R_{yx} \end{aligned}$$

[5]This criterion was proposed by William Rozeboom (e-mail posting, June 2000) and endorsed by Ken Bollen (e-mail posting, August 17, 2000). It can be shown that overidentification based on this criterion implies overidentification based on reduced equations; the converse is conjectured but not yet proven.

Thus, it appears as though both b and c are entitled to an overidentified status.

However, this conclusion clashes violently with basic intuition; variable z can be regarded as a noisy measurement of y, and we can not allow a parameter (b) to turn overidentified by simply adding a noisy measurement (z) to a precise measurement of y. The same holds for parameter c, once we regard x as a noisy measurement of y.

Definition 4 sides with our intuition and classifies both parameters as justidentified. It points out to us that, although the constraint $R_{zx} = R_{yx}R_{zy}$ leads to new expressions for b (and c), this constraint is redundant, as it plays no role whatsoever in the derivation of b (or c). In other words, we can easily violate this constraint, e.g., by relaxing the assumption $cov(e_x, e_y) = 0$, without spoiling the derivation of $b = R_{yx}$ and $c = R_{zy}$. To filter out such spurious constraints, Definition 4 calls for examining the basic assumptions behind the model, and for selecting those assumptions that are absolutely necessary for the derivation of b (or c).

In our example, the basic assumptions can be enumerated as follows:

(1) $y = e_y$

(2) $x = by + e_x$

(3) $z = cy + e_z$

(4) $cov(e_y, e_x) = 0$

(5) $cov(e_z, e_y) = 0$

(6) $cov(e_y, e_z) = 0$

Examining these assumptions, we note that the estimands for b and c are not overdetermined; for b we have precisely one sufficient (and minimal) subset:

(1) $y = e_y$

(2) $x = by + e_x$

(4) $cov(e_y, e_x) = 0$

and for c we have, likewise, one sufficient (and minimal) subset:

(1) $y = e_y$

(3) $z = cy + e_z$

(5) $cov(e_z, e_y) = 0$

Thus, the correct estimands for b and c are the usual regression coefficients $b = R_{xy}$ and $c = R_{zy}$, and the constraint $R_{xz} = bc$ can be ignored, because it does not offer an independent way of estimating b (or c); this constraint can be violated without perturbing the validity of the estimands $b = R_{xy}$ and $c = R_{zy}$.[6] Indeed, by relaxing assumption (6), $cov(e_x, e_z) = 0$, we violate the constraint $R_{xz} = bc$ without affecting the derivation of $b = R_{xy}$ $c = R_{zy}$. Moreover, any derivation of b that rests on the assumption $cov(e_x, e_z) = 0$ also rests on assumptions (1), (2), and (4), which in themselves are sufficient for deriving $b = R_{xy}$. This means that any failure of the estimate $b = R_{xy}$ due to violations of assumptions (1), (2), and (4) would also result in the failure of the estimate $b = R_{xz}/R_{xy}$, because it rests on assumption (6).

What are the practical implications of these differences? While traditional practice behooves us to regard the estimate of b as ambiguous, and to choose amongst (or average over) the contenders $b = R_{xy}, b = R_{xz}/R_{zy}$ (and perhaps also $b = R_{xy \cdot z}$), the new definition now classifies b as unambiguous, properly estimated as $\hat{b} = \hat{R}_{xy}$. This means that in testing the model above for fit, we should not fix b and c at their ML estimates, but rather at their regression estimates: $\hat{b} = \hat{R}_{xy}$, $\hat{c} = \hat{R}_{zy}$. The reason is that ML estimates may attempt to compensate for possible violation of assumption (6), when in fact this assumption has nothing to do with the proper estimate of b (or c).

Another way to explicate why Definition 4 prefers the estimate $b = R_{xy}$ over the estimate $b = R_{xz}/R_{xy}$ is as follows: Any misspecification error that violates $b = R_{xy}$ will also violate $b = R_{xz}/R_{xy}$ but not the other way around; there are misspecification errors that violate $b = R_{xz}/R_{xy}$ but leave $b = R_{xy}$ valid.

Is there an objective way of deciding between the traditional practice and the one proposed by Definition 4? In other words, is there an objective way of quantifying the damage caused to SEM researchers by following the traditional practice vis a vis the one suggested by Definition 4. The answer clearly depends on what we possibly hope to gain by choosing the *right* parameter estimates before submitting a model to a test of global fit. The natural answer is of course to gain higher rejection rates of misspecified models, and higher acceptance rates of properly specified models. I predict that simulation experiments will prove these advantages of the new proposal.

Global model testing is only one part of SEM research (and in my opinion a disproportionally emphasized part), another aspect of SEM studies is that of model construction and interpretation. The next example compares the traditional and new criteria along this dimension.

Example 4 *Consider a structural model M given by the chain in Figure 3,*

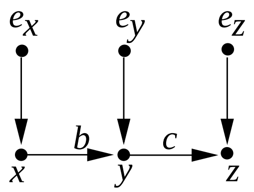

Figure 3:

[6]If the model is correctly specified, the two estimators of b, $b(1) = \sum_i X_i Y_i / \sum_i (Y_i)^2$ and $b(2) = \sum_i (X_i Z_i) / \sum_i Y_i Z_i$ would usually differ, due to sampling variations. However, the analysis of Goldberger (1973b) shows that the former estimator is sufficient while the latter is redundant; taking a weighted average of the two cannot increase estimation power. I thank David Kenney for bringing this result to my attention.

which stands for the equations:

$$\begin{aligned} x &= e_x \\ y &= bx + e_y \\ z &= cy + e_z \end{aligned}$$

together with the assumptions $cov(e_i, e_j) = 0,\ i \neq j$. *The constraints induced by this model are again:*

$$R_{zx} = R_{yx} R_{zy}$$

On the surface it appears that b and c have equal status. If we express the elements of σ in terms of the structural parameters, we obtain:

$$\begin{aligned} R_{yx} &= b \\ R_{zx} &= bc \\ R_{zy} &= c \end{aligned}$$

Thus, there is nothing here to warn us of any asymmetry between b and c and, as in Example 3, b and c can each be derived in two different ways:

$$b = R_{yx} \qquad b = R_{zx}/R_{zy}$$

and

$$c = R_{zy} \qquad c = R_{zx}/R_{yx}$$

Naturally, because of this apparent symmetry, b and c are classified as overidentified in most SEM texts.[7], which stands contrary to basic understanding of modeling and measurements. We should not allow a parameter to turn overidentified (hence more testable) by merely adding to the model a noisy measurement of an observed variable. Because z is a noisy measurement of y, we cannot gain any information (about b) from such measurement, once we have a precise measurement of y.

We shall now see that Definition 4 classifies b as justidentified and c as overidentified. The analysis of b is identical to that of Example 3, but the analysis of c reveals three distinct minimal sets of assumptions that are sufficient for deriving c, two of the three yield the same estimand.

We have previously accounted for zero-coefficient assumptions using the structural equations themselves. Here, we will find it more insightful to list such assumptions explicitly. Accordingly, instead of expressing the absence of direct effect between x and z using the equation

$$z = cy + e_z,$$

[7]In e-mail exchange, most discussants (including Stan Mulaik, William Rozeboom and Kenneth Bollen) indeed classified b and c as overidentified, with the exception of Bill Shipley who voted with me for classifying c alone as overidentified and b as justidentified. Mueller (1996, p. 50) also classifies b a justidentified, albeit for a different reason.

we will now write two separate assumptions:

$$z = cy + dx + e_z, \text{ and } d = 0;$$

the former defines the role of the coefficients c and d relative to variables z, y, x and e_z, the latter places d at zero.

Thus, the complete list of assumptions in this model reads:

(1) $x = e_x$

(2) $y = bx + e_y$

(3) $z = cy + dx + e_z$

(4) $cov(e_z, e_x) = 0$

(5) $cov(e_z, e_y) = 0$

(6) $cov(e_x, e_y) = 0$

(7) $d = 0$

There are three distinct minimal sets of assumptions capable of yielding a solution for c; we will denote them by A_1, A_2, and A_3.

Assumption set A_1

(1) $x = e_x$

(2) $y = bx + e_y$

(3) $z = cy + dx + e_z$

(5) $cov(e_z, e_y) = 0$

(6) $cov(e_1, e_y) = 0$

This set yields the estimand: $c = R_{zy \cdot x} = (R_{zy} - R_{zx}R_{yx})/(1 - R^2_{yx})$,

Assumption set A_2

(1) $x = e_x$

(2) $y = bx + e_y$

(3) $z = cy + dx + e_z$

(4) $cov(e_z, e_x) = 0$

(5) $cov(e_z, e_y) = 0$

also yielding the estimand: $c = R_{zy \cdot x} = (R_{zy} - R_{zx}R_{yx})/(1 - R^2_{yx})$,

Assumption set A_3

(1) $x = e_x$

(2) $y = bx + e_y$

(3) $z = cy + dx + e_z$

(4) $cov(e_z, e_x) = 0$

(7) $d = 0$

This set yields the instrumental-variable (IV) estimand: $c = R_{zx}/R_{yx}$.

Figure 4 provides a graphic illustration of these assumption sets, where each missing edge represents an assumption and each edge (i.e., an arrow or a bi-directed arc) represents a relaxation of an assumption (since it permits the corresponding parameter to remain free). We see that c is corroborated by three distinct set of assumptions, yielding two distinct estimands; the first two sets are degenerate, leading to the same estimand, hence c is classified as 2-identified and 3-corroborated (see Definition 5).

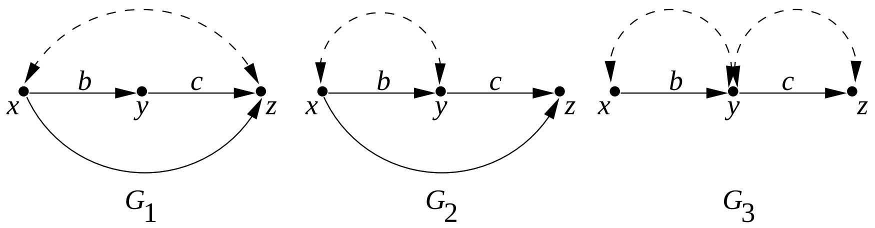

Figure 4: Graphs representing assumption sets A_1, A_2, and A_3, respectively.

Note that assumption (7), $d = 0$, is not needed for deriving $c = R_{zy \cdot x}$. Moreover, we cannot relax both assumption (4) and (6), as this would render c non-identifiable. Finally, had we not separated (7) from (3), we would not be able to detect that A_2 is minimal, because it would appear as a superset of A_3.

It is also interesting to note that the natural estimand $c = R_{zy}$ is not selected as appropriate for c, because its derivation rests on the assumptions $\{(1), (2), (3), (4), (6), (7)\}$, which is a superset of each of A_1, A_2 and A_3. The implication is that R_{zy} is not as robust to misspecification errors as the conditional regression coefficient $R_{zy \cdot x}$ or the instrumental variable estimand R_{zx}/R_{yx}. The conditional regression coefficient $R_{zy \cdot x}$ is robust to violation of assumptions (4) and (7) (see G_1 in Fig. 4) or assumptions (6) and (7) (see G_2 in Fig. 4), while the instrumental variable estimand R_{zx}/R_{yx} is robust to violations of assumption (5) and (6), (see G_3, Fig. 4). The estimand $c = R_{zy}$, on the other hand, is robust to violation of assumption (6) alone, hence it is "dominated" by each of the other two estimands; there exists no data generating model that would render $c = R_{zy}$ unbiased and the $c = R_{zx}/R_{yx}$ (or $c = R_{zy \cdot x}$) biased. In contrast, there exist models in which $c = R_{zx}/R_{yx}$ (or $c = R_{zy \cdot x}$) is unbiased and $c = R_{zy}$ is biased; the graphs depicted in Fig. 4 represent in fact such models.

We now attend to the analysis of b. If we restrict the model to be recursive (i.e., feedback-less, as we did in Example 3) and examine the set of assumptions embodied in the model of Fig. 3, we find that parameter b is corroborated by only one minimal set of assumptions, given by:

(1) $x = e_x$

(2) $y = bx + e_y$

(6) $cov(e_x, e_y) = 0$

These assumptions yield the regression estimand, $b = R_{yx}$. Since any other derivation of b must rest on these three assumptions, we conclude that no other set of assumptions can satisfy the minimality condition of Definition 4. Therefore, using Definition 6, b is classified as justidentified.

Popular attempts to attribute to b a second estimand, $b = R_{zx}/R_{zy}$, fail to recognize the fact that the second estimand is merely a noisy version of the first, for it relies on the same assumptions as the first, plus more. Therefore, if the two estimates of b happen to disagree in a specific study, we can conclude that the disagreement must originates with violation of those extra assumptions that are needed for the second, and we can safely discard the second in favor of the first. Not so with c. If the two estimates of c disagree, we have no reason to discard one in favor of the other, because the two rest on two distinct sets of assumptions, and it is always possible that either one of the two sets is valid. Conversely, if the two estimates of c happen to coincide in a specific study, c obtains a greater conformation from the data since, for c to be false, the coincidence of the two estimates can only be explained by an unlikely miracle. Not so with b. The coincidence of its two estimates might well be attributed to the validity of only those extra assumptions needed for the second estimate, but the basic common assumption needed for deriving b (namely, assumption (6)) may well be violated.

Example 5 *(Contributed by Kenneth Bollen)*
This example involves a single factor model with two parallel measurements:

$$\begin{aligned} x_1 &= F + d_1 \\ x_2 &= F + d_2 \end{aligned}$$

F is a latent variable, d_1 and d_2 are two errors that each have means of zero, are uncorrelated with F, and are uncorrelated with each other. In addition, we assume $V(d_1) = V(d_2) = V(d)$ where $V(d)$ is the error variance for the two measurements. The parameters of interest are $V(F)$ and $V(d)$.

The following is an explicit list of the assumptions in the model:

(0) $F = e$

(1) $x_1 = F + d_1$

(2) $x_2 = F + d_2$

(3) $cov(d_1, d_2) = 0$

(4) $V(d_1) = V(d_2)$

(5) $cov(d_1, e) = 0$

(6) $cov(d_2, e) = 0$

This leads to:

$$\begin{array}{rcll} V(x_1) & = & V(F) + V(d_1); & \{0,1,5\} \\ V(x_2) & = & V(F) + V(d_2); & \{0,2,6\} \\ C(x_1, x_2) & = & V(F); & \{0,1,2,3,5,6\} \end{array}$$

where the numbers on the right indicate the assumptions needed for the derivation of the corresponding expressions.

We can now express the model parameters, $V(F)$ and $V(d)$, in terms of estimable quantities, $V(x_1), V(x_2)$ and $C(x_1, x_2)$,

$$\begin{array}{rcll} V(F) & = & C(x_1, x_2) & \{0,1,2,3,5,6\} \\ V(d) & = & V(x_1) - C(x_1, x_2) & \{0,1,4,5\} \\ V(d) & = & V(x_2) - C(x_1, x_2) & \{0,2,4,6\} \end{array}$$

and we see that $V(F)$ is supported by one (minimal) set of assumptions, $\{0,1,2,3,5,6\}$, while $V(d)$ is supported by two distinct (and minimal) sets of assumptions: $\{0,1,4,5\}$ and $\{0,2,4,6\}$. This gives us the license to entitle V(F) "justidentified" and $V(d)$ "overidentified".

In an e-mail discussion on this issue, I have concocted another expression for $V(F)$,

$$V(F) = C(x_1, x_2) + V(x_1) - V(x_2),$$

and claimed that, if we decide overidentification by counting expressions, then $V(F)$ is overidentified because it is determined by two distinct expressions, each involving different elements of σ. Furthermore, one cannot dismiss the new expression for $V(F)$ simply because it differs from the first expression by a quantity $V(x_1) - V(x_2)$ that is immediately recognized as zero; the two expressions for $V(d)$ also differ by that same quantity, $V(x_1) - V(x_2)$, yet $V(d)$ is classified as overidentified.

This argument is wrong of course, but it demonstrates clearly that overidentification cannot be decided by counting distinct expressions for a parameter. The license for dismissing the concocted expression as redundant rests on the principle of minimality; the concocted expression cannot be derived without using assumption (4) $V(d_1) = V(d_2)$, hence the list of assumptions supporting the new expression would be $\{0,1,2,3,4,5,6\}$ which is a superset of the assumption set we originally had for $V(F)$. So, the new expression is not minimal, and should not be counted as overidentifying.

3 Graphical tests for overidentification

In this section we restrict our attention to parameters in the form of path coefficients, excluding variances and covariances of unmeasured variables, and we devise a graphical test for the overidentification of such parameters. The test rests on the following lemma, which generalizes Theorem 5.3.5 in [Pearl, 2000, p. 150], and embraces both instrumental variables and regression methods in one graphical criterion. (See also ibid, Definition 7.4.1, p. 248).

Lemma 1 *(Graphical identification of direct effects)*
Let c stand for the path coefficient assigned to the arrow $X \to Y$ in a causal graph G. Parameter c is identified if there exists a pair (W, Z), where W is a node in G and Z is a (possibly empty) set of nodes in G, such that:

1. *Z consists of nondescendants of Y,*
2. *Z d-separates W from Y in the graph G_c formed by removing $X \to Y$ from G.*
3. *W and X are d-connected, given Z, in G_c.*

Moreover, the estimand induced by the pair (W, Z) is given by:

$$c = \frac{cov(Y, W|Z)}{cov(X, W|Z)}.$$

The graphical test offered by Lemma 1 is sufficient but not necessary, that is, some parameters are identifiable, though no identifying (W, Z) pair can be found in G (see ibid, Fig. 5.11, p. 154). The test applies nevertheless to a large set of identification problems, and it can be improved to include several instrumental variables W. We now apply Lemma 1 to Definition 4, and associate the absence of a link with an "assumption."

Definition 7 *(Maximal IV-pairs)*[8]
A pair (W, Z) is said to be an **IV-pair** *for $X \to Y$, if it satisfies conditions (1–3) of Lemma 1. (IV connotes "instrumental variable.") An IV-pair (W, Z) for $X \to Y$ is said to be* maximal *in G, if it is an IV-pair for $X \to Y$ in some graph G' that contains G, and any edge-supergraph of G' admits no IV-pair (for $X \to Y$), not even collectively.*[9]

Theorem 2 *(Graphical test for overidentification)*
A path parameter c on arrow $X \to Y$ is overidentified if there exist two or more distinct maximal IV-pairs for $X \to Y$.

Corollary 1 *(Test for k-identifiability)*
A path parameter c on arrow $X \to Y$ is k-identified if there exist k distinct maximal IV-pairs for $X \to Y$.

Example 6 *Consider the chain in Fig. 5(a). In this example, c is 2-identified, because the pairs $(W = X_2, Z = X_1)$ and $(W = X_1, Z = 0)$ are maximal IV-pairs for $X_2 \to X_3$. The former yields the estimand $c = R_{32 \cdot 1}$, the latter yields $c = R_{31}/R_{21}$.*

[8] Carlos Brito was instrumental in formulating this definition.
[9] The qualification "not even collectively" aims to exclude graphs that admit no IV-pair for $X \to Y$, yet permit, nevertheless, the identification of c through the collective action of k IV-pairs for k parameters (e.g., [Pearl, 2000, Fig. 5.11]). The precise graphical characterization of this class of graphs is currently under formulation, but will not be needed for the examples discussed in this paper.

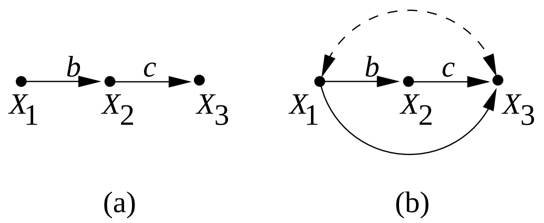

Figure 5:

Note that the robust estimand of c is $R_{32\cdot1}$, not R_{32}. This is because the pair $(W = X_2, Z = \emptyset)$, which yields R_{32}, is not maximal; there exists an edge-supergraph of G (shown in Fig. 6(b)) in which $Z = \emptyset$ fails to d-separate X_2 from X_3, while $Z = X_1$ does d-separate X_2 from X_3. The latter separation qualifies $(W = X_2, Z = X_1)$ as an IV-pair for $X_2 \to X_3$, and yields $c = R_{32\cdot1}$.

The question remains how we can perform the test without constructing all possible supergraphs.

Every (W, Z) pair has a set $S(W, Z)$ of maximally filled graphs, namely supergraphs of G to which we cannot add any edge without spoiling condition (2) of Lemma 1. To test whether (W, Z) leads to robust estimand, we need to test each member of $S(W, Z)$ so that no edge can be added without spoiling the identification of c. Thus, the complexity of the test rests on the size of $S(W, Z)$.

Graphs G_1 and G_2 in Fig. 4 constitute two maximally filled graphs for the IV-pair $(W = y, Z = x)$; G_3 is maximally filled for $(W = x, Z = \emptyset)$.

4 Assumption-based Parameter Estimation

The preceding analysis shows ways of overcoming two major deficiencies in current methods of parameter estimation. The first, illustrated in Example 2, is the problem of *irrelevant over-identification*; certain assumptions in a model may render the model over-identified, and contribute to improved overall fit, while playing no role whatsoever in the estimation of the parameters of interest. Researchers relying on global fit measures may thus obtain a false sense of confidence in the estimates of those parameters. It is often the case that only selected portions of a model gather support through confrontation with the data, while others do not, and it is important to separate the formers from the latter. The second is the problem of *irrelevant misspecifications.* If one or two of the model assumptions are incorrect, the model as a whole would be rejected as misspecified, though the incorrect assumptions may be totally irrelevant to the parameters of interest. For instance, if the assumption $cov(e_y, e_z)$ in Example 4 (Figure 3) was incorrect, the constraint $R_{zx} = R_{yx}R_{zy}$ would clash with the data, and the model would be rejected, though the regression estimate $b = R_{yx}$ remains perfectly valid. Thus, by relying on overall criteria of fit or misfit, researchers may miss opportunities to recover unbiased estimates of those parameters that are not affected by errors of misspecification.

One can of course diagnose the offending assumptions by systematically relaxing certain sets of assumptions and observing their impact on the model's fit. When the offending assumptions are diagnosed, they can be modified to yield a more realistic model, and more reliable estimates of the parameters, but only of those parameters that can be identified in the relaxed model. In this section we explore ways of identifying parameters that are insensitive to misspecifications without modifying the offending assumptions.

If the target of analysis is a parameter p (or a set of parameters), and if we wish to assess the degree of support that the estimation of p earns through confrontation with the data, we need to assess the disparity between the data and the model assumptions, but we need to consider only those assumptions that are relevant to the identification of p, all other assumptions should be ignored. Thus, the basic notion needed for our analysis is that of "irrelevance"; when can we declare a certain assumption irrelevant to a given parameter p?

One simplistic definition would be to classify as relevant assumptions that are absolutely necessary for the identification of p. In the model of Figure 2, since b can be identified even if we violate the assumptions $cov(e_z, e_y) = cov(e_z, e_x) = 0$, we declare these assumptions irrelevant to b, and we can ignore variable z altogether. However, this definition would not work in general, because no assumption is absolutely necessary; any assumption can be disposed with if we enforce the model with additional assumptions. Take, for example, the model in Figure 3; the assumption $cov(e_y, e_z) = 0$ is not absolutely necessary for the identification of c, because c can be identified even when e_y and e_z are correlated (see G_3 in Figure 4), yet we cannot label this assumption irrelevant to c.

The following definition provides a more refined characterization of irrelevance.

Definition 8 *Let A be an assumption embodied in model M, and p a parameter in M. A is said to be* relevant *to p if and only if there exists a set of assumptions S in M such that S and A sustain the identification of p but S alone does not sustain such identification.*

Theorem 3 *An assumption A is relevant to p if and only if A is a member of a minimal set of assumptions sufficient for identifying p.*

Proof:
Let *msa* abbreviate "minimal set of assumptions sufficient for identifying p" and let the symbol "$\models p$" denote the relation "sufficient for identifying p" ($\not\models p$, its negation). If A is a member of some *msa* then, by definition, it is relevant. Conversely, if A is relevant, we will construct a *msa* of which A is a member. If A is relevant to p, then there exists a set S such that $S + A \models p$ and $S \not\models p$. Consider any minimal subset S' of S that satisfies the properties above, namely

$$S' + A \models p \text{ and } S' \not\models p,$$

and, for every proper subset S'' of S', we have (from minimality)

$$S'' + A \not\models p \text{ and } S'' \not\models p,$$

(we use monotonicity here; removing assumptions cannot entail any conclusion that is not entailed before removal). The three properties: $S' + A \models p$, $S' \not\models p$, and $S'' + A \not\models p$ (for all $S'' \subset S'$) qualify $S' + A$ as *msa*, and completes the proof of Theorem 3. QED.

Thus, if we wish to prune from M all assumptions that are irrelevant to p, we ought to retain only the *union* of all minimal sets of assumptions sufficient for identifying p. This union constitutes another model, in which all assumptions are relevant to p. We call this new model the *p-relevant* submodel of M, M_p which we formalize by a definition.

Definition 9 *Let A_M be the set of assumptions embodied in model M, and let p be an identifiable parameter in M. The p-relevant submodel of M, denoted M_p is a model consisting of the union of all minimal subsets of A_M sufficient for identifying p.*

We can naturally generalize this definition to any quantity of interest, say q, which is identifiable in M.

Definition 10 *Let A_M be the set of assumptions embodied in model M, and let q be any quantity identifiable in M. The q-relevant submodel of M, denoted M_q is a model consisting of the union of all minimal subsets of A_M sufficient for identifying q.*

We can now associate with any parameter in a model properties that are normally associated with models, for example, fit indices, degree of fitness, degrees of freedom (df) and so on; we simply compute these properties for M_p, and attribute the results to p. For example, if D_q measures the fitness of M_q to a body of data, we can say that quantity q has disparity D_q with $df(q)$ degrees of freedom.

Consider the model of Figure 3. If $q = b, M_b$ would consist of one assumption, $cov(e_x, e_y) = 0$, since this assumption is minimally sufficient for the identification of b. Discarding all other assumptions of A is equivalent to considering the arrow $x \to y$ alone, while discarding the portions of the model associated with z. Since M_b is saturated (that is, just identified) it has zero degrees of freedom, and we can say that b has zero degrees of freedom, or $df(b) = 0$. If $q = c$, M_c would be the entire model M, because the union of assumption sets A_1, A_2 and A_3 span all the seven assumptions of M. We can therefore say that c has one degree of freedom, or $df(c) = 1$.

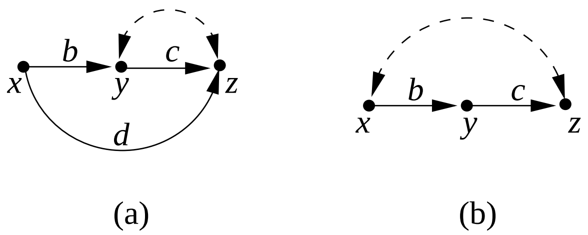

Figure 6:

Now assume that the quantity of interest, q, stands for the total effects of x on z, denoted $TE(x, z)$. There are two minimal subsets of assumptions in M that are sufficient for identifying q. Figure 6 represents these subsets through their respective (maximal) subgraphs; model 6(a) yields the estimand $TE(x, z) = R_{zx}$, while 6(b) yields $TE(x, z) = R_{yx}R_{zy \cdot x}$. Note that although c is not identified in the model of Figure 6(a), the total effect of x of on z, $TE(x, z) = d + bc$, is nevertheless identified. The union of the two assumption sets coincides with the original model M (as can be seen by taking the intersection of the corresponding arcs in the two subgraphs. Thus, $M = M_q$, and we conclude that $TE(x, z)$ is 2-identified, and has one degree of freedom.

For all three quantities, b, c and $TE(x, z)$ we obtained degrees of freedom that are one less than the corresponding degrees of identification, $k(q) = bf(q)$. This is a general relationship, as shown in the next Theorem.

Theorem 4 *The degrees of freedom associated with any quantity q computable from model M is given by $bf(q) = k(q) - 1$, where $k(q)$ stands for the degree of identifiability (Definition 5).*

Proof
$bf(q)$ is given by the number of independent equality constraints that model M_q imposes on the covariance matrix. M_q consists of m distinct msa's, which yield m estimands for q, $q = q_i(\sigma)$, $i = 1, \ldots, m$, k of which are distinct. Since all these k functions must yield the same value for q, they induce $k - 1$ independent equality constraints:

$$q_i(\sigma) = q_{i+1}(\sigma), \; i = 1, 2, \ldots, k - 1$$

This amounts to $k - 1$ degrees of freedom for M_q, hence, $bf(q) = k(q) - 1$. QED.

We thus obtain another interpretation of k, the degree of identifiability; k equals one plus the degrees of freedom associated with the q-relevant submodel of M.

5 Concluding Remarks

I envision that current method of model testing will eventually be replaced by a more meaningful style of tests. Global fitness tests will be forgotten and software programs will output the list of robust estimates for each parameter of interest. The coincidence or disparity of those estimates will be taken as the true indicators of how well the model conclusions are supported by the data.

References

[Bollen, 1989] K.A. Bollen. *Structural Equations with Latent Variables.* John Wiley, New York, 1989.

[Bowden and Turkington, 1984] R.J. Bowden and D.A. Turkington. *Instrumental Variables.* Cambridge University Press, Cambridge, England, 1984.

[Goldberger, 1973] A.S Goldberger. Efficient estimation in overidentified models: An interpretive analysis. In A.S. Goldberger and O.D. Duncan, editors, *Structural Equation Modeling in the Social Sciences*, pages 131–152. Seminar Press, 1973.

[James *et al.*, 1982] L.R. James, S.A. Mulaik, and J.M. Brett. *Causal Analysis: Assumptions, Models, and Data.* Studying Organizations, 1. Sage, Beverly Hills, 1982.

[Joreskog, 1979] K.G. Joreskog. *Advances in Factor Analysis and Structural Equation Models*, chapter Structural equation models in the social sciences: Specification, estimation and testing, pages 105–127. Abt Books, Cambridge, MA, 1979.

[MacCallum, 1995] R.C. MacCallum. Model specification, procedures, strategies, and related issues (chapter 2). In R.H. Hoyle, editor, *Structural Equation Modeling*, pages 16–36. Sage Publications, Thousand Oaks, CA, 1995.

[Mueller, 1996] R.O. Mueller. *Basic Principles of Structural Equation Modeling.* Springer, New York, 1996.

[Pearl, 2000] J. Pearl. *Causality: Models, Reasoning, and Inference.* Cambridge University Press, New York, 2000.

[Shipley, 2000] B. Shipley. A new inferential test for path models based on directed acyclic graphs. *Structural Equation Modeling*, 7(2):206–218, 2000.

[Werts and Linn, 1970] C.E. Werts and R.L. Linn. Path analysis: Psychological examples. *Psychological Bulletin*, 74(3):193–212, 1970. jp stacks.