Estimating causal effects in linear regression models with observational data: The Instrumental Variables Regression Model

Alberto Maydeu-Olivares

University of South Carolina

University of Barcelona

Dexin Shi

Amanda J. Fairchild

University of South Carolina

Psychological Methods (forthcoming)

Authors note: Correspondence concerning this article should be addressed to Alberto Maydeu-Olivares, Dept. of Psychology, University of South Carolina. Barnwell College. 1512 Pendleton St. Columbia, SC 29208. E-mail: amaydeu@sc.edu. This research was supported by the National Science Foundation under Grant No. SES-1659936.

Abstract

Instrumental variable methods are an underutilized tool to enhance causal inference in Psychology. By way of incorporating predictors of the predictors (called "instruments" in the Econometrics literature) into the model, instrumental variable regression (IVR) is able to draw causal inferences of a predictor on an outcome. We show that by regressing the outcome y on the predictors $\mathbf{x}$ and the predictors on the instruments, and modeling correlated disturbance terms between the predictor and outcome, causal inferences can be drawn on y on $\mathbf{x}$ if the IVR model cannot be rejected in a structural equation framework. We provide a tutorial on how to apply this model using ML estimation as implemented in structural equation modeling (SEM) software. We additionally provide code to identify instruments given a theoretical model, to select the best subset of instruments when more than necessary are available, and we guide researchers on how to apply this model using SEM. Finally, we demonstrate how the IVR model can be estimated using a number of estimators developed in Econometrics (e.g., two-stage least squares regression) and point out that the latter is simply a multi-stage SEM estimator of the IVR model.

Keywords: observational data, causal inference, regression

Estimating causal effects in linear regression models with observational data: The Instrumental Variables Regression Model

Early on, students are often taught that experiments are the gold standard for scientific research, as randomization of subjects to conditions ensures that groups will be equal on all non-treatment related dimensions (probabilistically, over the long run). This ensures protection against possible confounders of the relationships under study. Despite the advantages of true experimental designs, successful randomization of subjects to groups is not always possible. It may be unethical to assign subjects to groups, or a factor of interest may be an immutable, naturally occurring phenomenon. Even if random assignment is possible, a researcher may face problems of treatment compliance that renders randomization moot. Indeed, developing methods to estimate causal effects in the face of non-compliance is a burgeoning area of study (e.g., Peugh, Strotman, McGrady, Rausch, & Kashikar-Zuck, 2017; Schochet & Chiang, 2011; Stuart & Jo, 2015; Yau & Little, 2001). Moreover, random assignment is not practical when researchers are concerned with relationships that involve continuous independent variables. This latter point is manifest in several areas of study within psychology, such as life course development. In domains that are heavily reliant on observational data, statistical methods to enhance causal inference of observed relationships are paramount. A variety of techniques have been proposed in the literature to bolster causal inferences in observational data(Morgan & Winship, 2007; Rubin, 2006; Vanderweele, 2015). The preponderance of these methods that are used in psychology research seek to yield causal inferences by way of estimating effects in the potential outcomes framework (e.g., Rubin, 1974), which defines a unit-level causal effect as the difference between a subject's *potential* outcomes. That is, this approach considers the outcomes associated with all possible conditions to which a subject could be assigned rather than only the condition in which that subject was actually observed. By doing so, the researcher is able to estimate a counterfactual observation for each individual (i.e., the value of the outcome that would have been observed had a subject been in a different condition) such that an average causal effect can be estimated across observations in the data. An underutilized method to draw causal inferences in Psychology is the use of instrumental variable methods.

Instrumental variable methods have been used widely in areas outside of Psychology, such as Economics or Epidemiology. There is clear benefit in instrumental variable methods, as they focus on magnitude of effects sizes in a study (rather than exclusively focusing on statistical significance), and they are easy to implement. Though this set of tools has potential to help improve causal inferences in our field, barriers have precluded psychology researchers from integrating these methods routinely in their work. The purpose of the present study is to remove those barriers.

Consider a simple regression model, y on x. Empirical support for the inference that x causes y can be obtained by bringing in predictors of x into the model ($\mathbf{z}$, called "instruments" in Econometrics terminology), and fitting the *instrumental variables regression* (IVR) model (y on x on z with correlated errors between the disturbance terms of y and x). Historically, the most widely used estimator for this model has been the two stage least squares estimator (2SLS; Theil, 1953). The IVR model can be estimated effortlessly within a structural equation modeling (SEM) framework using any SEM estimator, such as maximum likelihood (ML), however, as whether causal inferences can be drawn depends on the fit of the model rather than the estimator used. Researchers familiar with the SEM framework are likely to find fitting the IVR model using this framework more intuitive than using the estimators developed in Econometrics. Econometric estimators such as 2SLS simply take advantage of the specific structure of the IVR model to estimate its parameters using a multi-stage mean and covariance structure approach with non-iterative solutions (Maydeu-Olivares, Shi, & Rosseel, 2018)[1].

In this paper, we provide a non-technical introduction to instrumental variable regression methods for a SEM audience[2]. We begin by discussing causality in regression models within a SEM framework. In particular, we address the question of how the IVR model enables a researcher to draw causal inferences in regression models with observational data. We then discuss how to estimate this model using SEM software and the new modeling avenues this provides, as well as demystify competing terminology across disciplines (e.g., "test of exclusion constraints" in Econometrics and "chi-square test" in Psychology). Next, we discuss how to identify potential instruments (i.e., variables that may serve as instruments in an IVR model) under a theoretical model. Finally, we discuss how to select the best subset of potential instruments when there are more potential instruments than necessary. We use example data throughout the article to illustrate the different concepts discussed. We additionally provide supplementary material to this article, which includes software scripts in STATA, SAS, R, and Mplus that will enable applied researchers to reproduce the example results reported in this article.

Predicting a company's market share from its degree of innovation performance

To focus concepts, throughout this article we use a small example involving the prediction of a company's market share from its degree of innovation performance. Maydeu-Olivares and Lado (2003) gathered data on 122 European companies from the insurance sector. They recorded their national market share from public records and gathered information from the companies' degree of innovation performance by administering a four-item questionnaire to the company's senior executive and computing the sum of their responses to these items. The items were on a 7-point Likert format and inquired about the success of new products/services (defined as improved products, product extensions, or new product lines) introduced by the company. We choose this example because an economic variable (in this case measured with a small error) is predicted from a psychological variable (in this case with, possibly, considerable measurement error).

Letting *i* denote company, a structural model can be put forth in the form of a simple regression model, $y_i = \alpha + \beta x_i + \varepsilon_i$ involving the log of market share, *y*, and the self-reported innovation performance, *x* (see Figure 1a). In this equation, ε (an error or disturbance term) includes all determinants of market share not included in the structural model and measurement error of *y*, that is, the measurement error of market share. In estimating the regression model, we obtain $\hat{\beta}$ = .11 (.02), R^2 = 18.3%. Thus, in European insurance companies a one-point increase in self-reported innovation performance leads to an 11% increase in market share[3]. This is a tacit causal statement; if indeed *x* causes *y*, a 95% confidence interval for β is $C[.07 \le \beta \le .15]=.95$.

In this example, we cannot claim empirical support for our causal statement. When only *y* and *x* are available for analysis, a *y* on *x* model cannot be distinguished empirically from an *x* on *y* model, nor from a *y* with *x* model; all three models are equivalent in a path analytic framework. However, we can provide empirical support for the statement *x* causes *y* when predictors of *x*, denoted here as **z**, are brought into the model, and some other model conditions that we will discuss, are met. Neither experimental manipulation nor temporal sequence is needed to formulate the causal claim in this context: *y*, *x*, and **z** can be observational variables.

Insert Figure 1 about here

The instrumental variables regression (IVR) model

Consider a target structural model (y on x) and its corresponding instrumental variables regression model (y on x on $\mathbf{z}$ with correlated errors between the disturbances of y and x) depicted in Figures 1a and 1b. By fitting the IVR model, causal inferences can be drawn for our target model. To formulate a causal inference, we need to show that by fitting the IVR model we can overcome any threats to the validity to the causal relationship of interest in the target model. There are a number of threats to the validity of causal inferences for y on x including reciprocal causation, the presence of omitted variables, measurement error in the predictors $\mathbf{x}$, autoregressive errors, spatial effects, etc. Bollen (2012) provides a good overview of these threats to validity in regression models. For brevity, we shall focus on just three of the threats here, which we depict graphically in Figure 2: reverse causation (x on y¸ Figure 2a), reciprocal causation (y on x, and x on y¸ Figure 2b), and the presence of omitted variables (y on z and x, and x on z¸ Figure 2c).

Insert Figure 2 about here

These models are a threat to the validity of inferences in our target model y on x because if any of these models is the data generating mechanism (DGM) and we fit our target model, we will not consistently estimate the true parameter of interest, β_{yx} [4]. The general idea behind the use of the IVR model is that these threats to the validity of y on x inferences can be addressed if we fit a y on x regression with correlated errors between the disturbances of y and x (Figure 2d). In so doing, we are able to consistently capture the true value of the β_{yx} parameter.

The model depicted in Figure 2d is not identified (i.e., cannot be estimated) if only data on *y* and *x* is available to the researcher, however. To be able to estimate the model in Figure 2d additional variables need to be brought into the model. For instance, predictors of *x*, labeled **z**, may be incorporated, leading to the IVR model depicted in Figure 1b. It is important to emphasize that the role of **z** in the IVR model is secondary. Our interest lies in drawing inferences on the causal effect of *x* on *y*. We are not interested in drawing inferences regarding *x* on **z**. Rather, we simply use **z** to *predict x*. Because of their instrumental role, the predictors of *x* are aptly named instrumental variables in the Econometrics literature (Cameron & Trivedi, 2005; Greene, 2007; Wooldridge, 2013). We now show that the use of the IVR model overcomes each of the three threats to the validity of causal inferences in our target model.

Suppose the data generating mechanism is the reverse causation model depicted in Figure 2a. When instruments for *x* are brought into the model, we obtain the DGM depicted in Figure 3a. If we fit the IVR model depicted in Figure 1b to this DGM, we will estimate the population parameters presented in Table 1[5]. We see in this table that the IVR model will consistently estimate the true slope, $\beta_{yx} = 0$, if the DGM is the reverse causation model of Figure 3a. Moreover, we see that a researcher fitting the IVR model will also consistently estimate the true slope no matter whether the DGM is the reciprocal causation model (Figure 3b), the omitted variables model (Figure 3c) when only *y*, *x*, and z_1 and z_2 are available for analysis, or a reciprocal causation model with omitted variables (Figure 3d)[6]. For an alternative intuition on how instrumental variable models provide consistent estimates see Foster and McLanahan (1996).

\-------------------------------------

Insert Figure 3 and Table 1 about here

What can we make causal inferences on? The role of model testing

The IVR model can be straightforwardly estimated using any structural equations modeling program (AMOS: Arbuckle, 2014; EQS: Bentler, 2004; LISREL: Jöreskog & Sörbom, 2017; Mplus: Muthén & Muthén, 2017; Lavaan: Rosseel, 2012) and its fit evaluated using the chi-square test. Provided there is sufficient power, failing to reject the null hypothesis of model fit using the chi-square test provides empirical support for inferences drawn on the parameters of the IVR model. Moreover, we will be able to make causal inferences on any population parameter within the model for which we are able to estimate its true value regardless of what the data generating mechanism is. Table 1 reveals that regardless of the data generating mechanism, by fitting the IVR model one

(a) Always recovers the parameters involving the regression of *y* on *x* (intercepts and slopes)

(b) Always recovers the parameters involving the regression of *x* on **z** (intercepts and slopes)

(c) Does not always recover the error variance parameters of *y* and *x*.

Result (a) means that when we fail to reject the IVR model, we can draw causal inferences on the intercepts and slopes in a *y* on *x* regression. Result (c) indicates that we cannot draw causal inferences on the percentage of variance of *y* accounted for by *x* (R^2_{yx}). That is, we can estimate R^2_{yx} under the IVR model, but we cannot be certain that the estimate is the actual percentage of variance of *y* explained by *x* since the value will depend on the unknown data generating mechanism. Finally, although result (b) indicates the IVR model will always recover parameters involving the regression of *x* on **z**, we will not be able to draw causal inferences on the *x* on **z** intercepts and slopes as the *x* on **z** submodel is saturated (i.e., *df* = 0), such that the chi-square statistic of model fit only assesses the validity of inferences between *y* and **z**.

Model assumptions and identification

As with any overall test statistic, the chi-square statistic tests *all* model assumptions simultaneously. In addition to the assumptions graphically displayed by the path diagram of Figure 1b, the IVR model assumes that: (1) All relationships are linear, and (2) All relationships are homoscedastic (i.e., constant variance around the regression lines)[7]. These assumptions can be assessed graphically by examining the plots of individual observations' standardized residuals (estimated errors) vs. predicted values for the equations in the IVR model: $\mathbf{x}$ on $\mathbf{z}$, and y on $\hat{\mathbf{x}}$. These residuals plots should always be inspected, regardless of the outcome of the chi-square statistic. If we fail to reject the model, the plots should be examined to determine if the chi-square statistic lacked power due to small sample size to detect violations of linearity or homoscedasticity. If the chi-square statistic rejects the model, the plots should be examined to determine whether lack of linearity or homoscedasticity is the source of misfit.

For the IVR model in Figure 1b to be identified, we must incorporate at least as many instrumental variables, $\mathbf{z}$, as there are predictors ($\mathbf{x}$) in our target model (Bollen, 1989). If there are as many instrumental variables as predictors, the model is just identified (i.e., $df = 0$), and it cannot be tested using the chi-square statistic. To test the model, we recommend bringing in at least one more instrumental variable than predictors are in the model.

Identifying potential instrumental variables

Whether a variable is a suitable instrument is established by assessing the fit of the IVR model. However, in applications it is necessary to identify variables that potentially may serve as instruments. Simply using trial and error to identify variables to serve as instruments is akin to failing to account for multiple testing.

We can identify variables that may serve as potential instruments by using our theoretical knowledge and a full structural model for the relations between y, x, and $\mathbf{z}$ (Bollen, 1996b; Cameron & Trivedi, 2005). Let us return to our market share and innovation performance example. Maydeu-Olivares and Lado (2003) provided a full structural model of the effect of innovation performance on market share that is reproduced here as Figure 4, involving a company's market orientation, degree of innovation, and customer loyalty[8].

Insert Figure 4 about here

The two conditions for a variable z to qualify as potential instrument of the y on $\mathbf{x}$ relationship are that it is (a) correlated with variables in $\mathbf{x}$, and (b) uncorrelated with the disturbance of y, ε_y (Bollen, 1996b, p. 133). The theoretical model displayed in Figure 4a reveals immediately two potential instruments: the companies' market orientation (z_1), and degree of innovation (z_2). There is a direct path between each of these variables and our predictor x, innovation performance, such that they satisfy condition (a) to be instruments of innovation performance. Also, neither has a correlation with the disturbance of our outcome y, market share (ε_y). Therefore, they satisfy condition (b) to be instruments. Though market orientation has an indirect path to market share through innovation and customer loyalty, it is not correlated with the disturbance term of market share as the indirect effect is completely mediated through customer loyalty. Customer loyalty is also a potential instrument under the theoretical model displayed in Figure 4a. It satisfies condition (a) as customer loyalty and innovation performance are correlated due to their common dependence on market orientation, and it satisfies condition (b) because it is uncorrelated with the disturbance of market share. Note that if the model in 4a is correct, but you are not able to collect data on customer loyalty, the theoretical model needs to take into account the effect of customer loyalty on the remaining variables, as depicted in Figure 4b, and market orientation will no longer meet the condition for a valid instrument, as it would be correlated with the disturbance term of the outcome.

For simple models such as that portrayed in Figure 1, it need not be difficult to identify potential instruments. To check condition (b), possible correlations between a potential instrument and the outcome's disturbance should be examined. To check condition (a), the model-implied covariance or correlation matrix should be examined. The latter is straightforward if data are available for all variables involved. It may not be straightforward to determine potential instruments, however, in more complex theoretical models or when data are not available for all variables. To facilitate determination, in the Appendix we provide two algebraic conditions that may be used to verify whether a variable is a potential instrumental variable for a target model within a structural equations framework. We also provide R code in section two of the supplementary materials to this article that enables researchers to determine which variables are potential instruments for their target model in the absence of data. This code provides invaluable information at the planning stages of a research project as it enables applied researchers to determine what variables they should gather data on to be used as instruments.

Estimation of the instrumental variables regression model

As indicated earlier, IVR models have historically been estimated using 2SLS[9]. Estimation of the model in a SEM framework can be applied in a straightforward manner, however, using estimators with which the SEM researcher is familiar (e.g., ML). The IVR model is simply y on $\mathbf{x}$ on $\mathbf{z}$, with correlated errors between y and $\mathbf{x}$. We let p_x be the number of predictors $\mathbf{x}$, p_z be the number of instrumental variables $\mathbf{z}$, and consider the case with a single dependent variable y. The total number of variables is therefore $p = p_x + p_z + 1$. Letting $\mathbf{w}' = \left(y, \mathbf{x}', \mathbf{z}'\right)'$, the covariance structure implied by the IVR model can be partitioned according to the partitioning of **w** as

$$\boldsymbol{\Sigma}_0 = \begin{pmatrix} \sigma_{yy} & & \\ \boldsymbol{\sigma}_{xy} & \boldsymbol{\Sigma}_{xx} & \\ \boldsymbol{\sigma}_{zy} & \boldsymbol{\Sigma}_{zx} & \boldsymbol{\Sigma}_{zz} \end{pmatrix}. \tag{1}$$

SEM estimation of the IVR model proceeds as for any other structural model. The most popular estimation method in SEM applications is maximum likelihood (ML). For normally distributed mean centered data, ML estimates are obtained by minimizing

$$F_{ML} = \ln\left|\boldsymbol{\Sigma}_0\right| + \mathrm{tr}\left(\boldsymbol{\Sigma}_0^{-1}\mathbf{S}\right) - \ln\left|\mathbf{S}\right| - p, \tag{2}$$

where **S** denotes the ML sample covariance matrix[10], and $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ is the population covariance structure (1), a function of $\boldsymbol{\theta}$, the model parameters to be estimated from the data.

Standard errors and goodness of fit tests can be obtained under normality assumptions, or robust to non-normality (Maydeu-Olivares, 2017; Satorra & Bentler, 1994). The most commonly used statistic to test a model under normality assumptions is the likelihood ratio (*LR*) statistic. Because the LR test statistic follows a chi-square distribution in large samples, it is most often referred to as the chi-square test in the SEM literature. The number of degrees of freedom when fitting the IVR model is always $p_z - p_x$. When data are not normally distributed, robust standard errors and goodness of fit tests are obtained using the asymptotically distribution free (ADF) assumptions[11] described in Browne (1982, 1984). Browne also described a goodness of fit statistic based on residual covariances that can be used with any estimator under normality or ADF assumptions[12]. This statistic follows a chi-square distribution in large samples with the same degrees of freedom as the LR statistic, and under normality assumptions, it is asymptotically equivalent to the LR statistic (Satorra & Bentler, 1994). When testing the IVR model, Browne's test statistic is also algebraically equivalent to test statistics used in 2SLS estimation (Maydeu-Olivares, Rosseel, & Shi, 2019)[13]. Note, however, that in SEM applications the most popular overall goodness of fit statistic reported under ADF assumptions is not Browne's test statistic but rather the LR statistic adjusted by its asymptotic mean, or by its asymptotic mean and variance (Satorra & Bentler, 1994)[14].

Some remarks on estimating the IVR model

After suitable potential instrumental variables have been identified and the IVR model has been fitted, applied researchers can verify whether a potential instrumental variable satisfies conditions to serve as a suitable instrument in the IVR model. Within a SEM framework, the first condition, i.e., a lack of correlation between the instrument and the outcome's disturbance, can be verified using the overall chi-square test of model fit. Support for the condition obtained by failing to reject the model. A more focused test is obtained by examining the modification index (aka as Lagrange or score test statistics, Sörbom, 1989) for the correlated error between the outcome's disturbance and the instrument.

The second condition, whether the potential instrument z significantly predicts the predictor(s) $\mathbf{x}$, can be checked by performing a test for nested models, where the fit of the IVR model is compared to a model where the effects of z on $\mathbf{x}$ are set to zero. Under normality, such model comparisons can be performed using the likelihood ratio statistic (Steiger, Shapiro, & Browne, 1985): the difference between the overall goodness of fit test statistics follows a chi-square distribution with degrees of freedom equal to the difference of degrees of freedom. A test for nested models robust to non-normality can be obtained when mean corrections are applied to the overall test statistics of each model (Satorra & Bentler, 2001, 2010), and also when mean and variance corrections are employed (Asparouhov & Muthén, 2006). In the special case of a single predictor x, the condition can be examined using the z statistic (a Wald test) for the slope x on z.

Threats to the validity and accuracy of inferences drawn using instrumental variable models

For inferences drawn using instrumental variables to be valid and accurate, the IVR must be correctly specified and have adequate power to identify violations of model assumptions. Use of the IVR model in the presence of model misspecification or violation of assumptions will lead to invalid inferences. Though overall goodness of fit tests will test all assumptions of the IVR model simultaneously, they will have differential power to detect different violations of the model assumptions.

Arguably, the main violation of the IVR model assumptions involves the presence of correlations between the y and $\mathbf{z}$ error terms. In this case, inferences drawn on the y on $\mathbf{x}$ relationship using the IVR are biased and invalid. The SEM approach to estimating IVR empowers applied researchers to estimate IVR models with correlated residuals between y and z, something that is unfeasible using the 2SLS regression approach. To that end, it is arguable that fitting IVR models with ML in an SEM framework may be more desirable.

Even when the IVR model is correctly specified, statistical theory used to draw causal inferences in the model is based on large sample theory. In finite samples, estimates may be biased and/or show large sampling variability. A large body of research has emerged in the Econometrics literature focusing on the drivers of the performance of 2SLS in correctly specified models. This research has focused on the impact of the strength of the relationship $\mathbf{x}$ on $\mathbf{z}$ (i.e., whether instrumental variables are 'weak' predictors of $\mathbf{x}$) has on the performance of instrumental variable methods (Bound, Jaeger, & Baker, 1995; Staiger & Stock, 1997; Stock &

Yogo, 2005), indicating that it does not suffice that the effect of **z** on **x** be non-significant; it needs to be 'strong'. These findings are, in principle, applicable to any estimator (e.g., ML) used to draw causal inferences using the IVR model, as the validity of inferences depends on the model and not on the estimator.

Most of these results are based on normality assumptions, reflecting that “much more is known about weak instruments in this case” (Stock, Wright, & Yogo, 2002, p. 518). The use of these results is twofold: a) to assess whether the solution obtained using instrumental variables is reasonably unbiased; b) to select the best subset of instruments when there is a set of valid instruments from which to choose. In this regard, extant research (e.g., Bollen, Kirby, Curran, Paxton, & Chen, 2007) suggests that in small samples ($N < 200$) instrumental variable methods are most accurate (less biased) when the number of instruments exceeds by one the number of predictors, that is, when $p_z = p_x + 1$.

Weak instruments

Standard practice in the Econometrics literature is to check for weak instruments in IVR models, and psychology researchers should follow suit. For a single predictor (x) model, the F statistic to predict x from **z** is used to check the weakness of instruments[15]. Stock and Yogo (2005) tabulated critical values for the F statistic to rule out weak instruments. A minimum value of $F = 10$ is recommended to consider the instruments as strong (Stock, Wright, & Yogo, 2002), and by extension for instrumental variable inferences to be reliable. When there are multiple predictors **x**, inspection of the individual F statistics is no longer sufficient (e.g., Sanderson & Windmeijer, 2016). In this case, under normality assumptions, the strength of the instruments is assessed by computing the smallest eigenvalue of the Cragg and Donald (1993) statistic (T_{CD}, see the Appendix). The minimum eigenvalue obtained in an application should be compared to the critical values of Table 1 in Stock and Yogo (2005).

We provide R code in section three of our supplementary materials to compute the F statistic when there is a single predictor x, or the minimum eigenvalue of the T_{CD} statistic when there are multiple predictors $\mathbf{x}$, for every combination of instruments equal or larger than the number of predictors in the model (i.e., $p_z \geq p_x$). The optimal subset of instruments is the subset of instruments with largest F or largest minimum eigenvalue of the T_{CD} statistic. In addition, when the IVR model is fitted by ML using a SEM approach, the model with all potential predictors against the model with the subset of predictors with largest F or T_{CD} statistic can be compared using a test statistic for nested models. To do so, the slopes for the unselected instruments in the optimal model are set to zero.

Numerical examples

We turn now to two empirical examples to further exemplify the use of IVR models in applied work. Our first example considers the case of a carefully laid out theoretical model that focuses on a single relationship under study, and allows us to describe how one would go about choosing potential instruments. Our second example changes focus to a more realistic scenario, where a researcher uses a broader theoretical backdrop against which to consider multiple relationships under study (that are conditional on other variables in the model). We also used standardized parameters in this latter example to demonstrate that inferences can be drawn on standardized parameters under the IVR model.

ML vs. 2SLS estimation of the market share and innovation performance example

In this example, it is of interest to determine the causal effect of innovation performance on the log of market share. To do so, we fitted an IVR model according to the theoretical model underlying the relationships between the observed variables involved (see Figure 4), involving a company's market orientation, degree of innovation, and customer loyalty. Details on sample characteristics, and motivation for this study were provided at the beginning of this article. Results obtained assuming normality or under ADF assumptions were very similar, and therefore we report here only the results under normality. The plot of innovation performance residuals vs. expected (fitted) values reveals a linear and homoscedastic relation. The residual plot of ln(market share) also suggests a linear and homoscedastic relation (see Figure 5a).

Insert Figure 5 about here

The likelihood ratio test statistic (commonly referred to as the chi-square test in the SEM literature) when market orientation, innovation degree, and customer loyalty (z_1 to z_3) are used as instruments is $X^2(2) = .97$, $p = .62$. The F statistic when all three potential instruments are used is $F = 34.16$. The effect of customer loyalty on innovation performance was not statistically significant however, $\hat{\beta}_{xz_3} = .06$ (SE = .10), $p = .59$. Therefore, we removed this instrument from the model. The F statistic associated with market orientation and degree of innovation as a subset of two potential instruments was $F = 51.59$. These results suggest that the two instruments are strong. Though we can compute an F statistic for single instruments (highest F statistic for a single instrument was $F = 75.27$ for innovation degree), recall that only including one instrument would prevent us from being able to test the model fit (i.e., $df = 0$). Full results for the model with market orientation and degree of innovation as instruments are provided in Table 2[16]. As we can see in this table, ML vs. 2SLS estimates are often equal (with two decimal precision).

Insert Table 2 about here

ML results were obtained using Mplus (Muthén & Muthén, 2017); for estimation under ADF assumptions we used choice MLMV (for details, see Maydeu-Olivares, 2017). We provide the data and Mplus code in section one of the supplementary materials to this article. We also provide code in the Lavaan (Rosseel, 2012) R package that matches the Mplus results, as well as code to obtain the two-stage least squares (2SLS) estimates using Stata (StataCorp, 2017) -*ivregress* command, SAS (SAS Institute, 2014) –*proc* syslin command, and R(R Development Core Team, 2015)- *ivreg* and *IVR* commands.

Returning to the results shown on Table 2, we see that after removing customer loyalty as an instrument, we obtain $X^2(1) = .01$, $p = .91$. The estimated effect of innovation performance on market share is $\hat{\beta}_{yx} = .20$ (.03). Since the IVR cannot be rejected, this is the causal effect of innovation performance on ln(market share) with a 95% confidence interval of: $C[.13 \leq \beta \leq .26] = .95$[17]. This estimate is substantially larger and less precise than the result obtained using ordinary regression methods reported earlier, $C[.07 \leq \beta \leq .15] =. 95$. The percentage of variance of market share accounted for by innovation performance according to the IVR model is R^2 = 6.1%., which is substantially smaller than the R^2 obtained by ordinary regression methods, 18%.We cannot make an empirically supported causal statement about the R^2 using the IVR model as it would require knowing the DGM. A causal statement about the IVR R^2 can only be supported on theoretical grounds.

These data provide an excellent example of the need to examine model assumptions using residual plots. We fitted the IVR model to these data without taking the natural log of market share, obtaining $X^2(1) = .30$, $p = .59$. Hence, we cannot reject the model using this test statistic.

However, the model is misspecified as can be readily observed by examining a plot of market share residuals vs. expected values (Figure 5b). We see in this plot that we if we do not take the log of market share, the relationship is clearly heteroscedastic with large outliers. Inferences drawn using the IVR model if a log transformation is not applied to market share are considerably less accurate than those obtained using ordinary regression methods to the logged variable.

Causal inferences on the standardized effects of problem orientation on catastrophizing

Catastrophizing, an irrational belief characterized by evaluation of negative events as worse than they should actually be (Ellis, 1962), has proved to be a key variable in the prediction of pain-related outcomes (Suso-Ribera et al., 2016). In this example, we use the subset of chronic pain patients described in Suso-Ribera et al. (2016) that was measured over time. Only time 1 data is used. Sample size is 175. We are interested in determining the causal effect (if any) of problem orientation (D'Zurilla & Goldfried, 1971) on catastrophizing. Problem orientation is defined as the schemas one holds about problems in everyday life and ones assessment of their ability to solve them. Problem orientation may be positive (PPO: e.g., seeing problems as challenges and opportunities for growth) or negative (NPO: e.g., seeing problems as threats, barriers, or obstacles) (D'Zurilla, Nezu, & Maydeu-Olivares, 2004; Maydeu-Olivares & D'Zurilla, 1996).

We hypothesize that NPO has a causal effect on catastrophizing, and we are unsure as to whether PPO also has a causal effect (holding NPO constant) on catastrophizing. Research by D'Zurilla et al. (2011) suggests that the Big Five personality factors (e.g., Goldberg, 1993) may be suitable instruments (i.e., they may be strong predictors of positive and negative problem orientation). Consequently, to find out the causal effect of PPO and NPO on catastrophizing we fitted an IVR model using the Big Five personality factors as instruments: Neuroticism (N), Extraversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C).

Catastrophizing was measured using the short version of the General Attitudes and Beliefs Scale (GABS-SV: Kirkby, Wertheim, & Birch, 2007), problem orientation was measured using the short version of the Social Problem Solving Inventory-Revised (SPSI-R: D'Zurilla, Nezu, & Maydeu-Olivares, 2002). Finally, the Big Five was measured using the NEO Five Factor Inventory (NEO-FFI: Costa & McCrae, 1992). Because the variables being modeled are sums of item scores, the unstandardized intercepts and slopes are not of particular interest. In this instance, it is common to make inferences on standardized parameters.

The minimum eigenvalue of the T_{CD} statistic when all five instruments are used is 23.59. This is larger than the critical value at the 5% level reported in Stock and Yogo's (2005) Table 1 for $p_x = 2$ and $p_z = 5$, 13.97. Thus, we can infer that our instruments are strong. The largest minimum eigenvalue of T_{CD} across all combinations of four instruments was 23.59 for {N, O, A, C}, across all combinations of three instruments was 20.02. These results suggest using only {N, O, A, C} as instruments. For ML estimation, when all five instruments are used, we obtained $X^2 = 4.04$ (3), $p = .26$. We used the likelihood ratio test statistic for nested models to compare the fit of a IVR model with five instruments against a model in which all slopes for E were set to zero, obtaining $X^2_{dif} = 1.93$, $df_{dif} = 1$, $p = .17$. We conclude that E can be dropped as an instrument in this example. A path diagram of the resulting model is shown in Figure 6. We see in this figure that path diagrams of IVR models when there is more than one predictor x appear rather clogged and they are not very informative. IVR models are best described in words. The fitted model is catastrophizing on {PPO, NPO} on {N, O, A, C}, with correlated errors between catastrophizing and {PPO, NPO}, and correlated errors between PPO and NPO.

Insert Figure 6 about here

We provide in Table 3 ML and 2SLS estimates for selected standardized parameters of this model, as well as goodness of fit results. Results are reported under normality assumptions, as differences with ADF results were negligible. We see in this Table that results do not differ much across estimators (ML vs. 2SLS).

Insert Table 3 about here

Plots of individual residuals did not suggest any violation of the linearity and homoscedastic assumptions; also, the chi square test failed to reject the model (see Table 3). Therefore, we can estimate the causal effects of problem orientation on catastrophizing in chronic pain patients. The interpretation of these causal effects is similar to that of standardized regression coefficients. For instance, we see in Table 3 that the ML standardized effect of Negative Problem Orientation (NPO) on catastrophizing (CATAS) is β =1.07 (SE = .17). In words, for one standard deviation increase in NPO we expect a 1.07 standard deviation increase in catastrophizing, holding Positive Problem Orientation fixed. If we add or remove predictors of catastrophizing, we should expect our estimate of the effect of NPO on CATAS to change, just as in regression. In addition, if we drop or add instruments, we should expect our estimate of the effect of NPO on CATAS to change, hopefully, not by much. Note that we will be able to recover our intercept and slopes parameters of interest even in the face of one of the described threats to internal validity illustrated in Figure 2, as the IVR model is robust against these influences (see Bollen, 2012 for a complete discussion).

It is of interest to compare the results obtained using the IVR model with those obtained from an ordinary regression analysis (see Table 3). As with the previous example, we see that IVR results are larger and less precise than the regression results. For instance, a 95% confidence interval for the standardized causal effect of CATAS on NPO is $C[.74 \leq \beta \leq 1.41] = .95$. In contrast, the interval for this effect under the basic regression model is $C[.47 \leq \beta \leq .68] = .95$. Since we have no evidence to reject the IVR model, we are reasonably confident in the causal inference drawn using this model. This is twice the size of the effect estimated using ordinary regression analysis, in standardized units. The wider confidence interval for the IVR results is due to the difference in magnitude of the standard errors estimated by regression (.06) and by the IVR model (.17). In our experience, standard errors of standardized parameters are often around $1/\sqrt{N}$, where N denotes sample size (.08 in this case). The SEs from the ordinary regression model in our example are indeed around this magnitude. The SEs from the IVR model for the estimates of the instruments on predictors are around this magnitude. However, the SEs for the causal effects from the IVR model, as well as for the SEs for the covariances between predictors and the disturbance of the outcome, are twice the expected magnitude. The degree of inflation of SEs for the IVR slope estimates depends on the correlations between the instruments and the predictors: the larger the correlations, the smaller the SE inflation relative to regression results.

Insert Figure 6 and Table 3 about here

In applications, it may be of interest to test whether the estimates obtained using the IVR model equal those obtained using regression. With ML estimation, this can be accomplished by fitting the IVR model setting the *unstandardized* parameters of interest (CATAS on PPO and NPO) equal to the regression estimates and performing a test for nested models. Setting CATAS on PPO equal to the regression value we obtain $X^2 = 2.79$ (1), $p = .10$, and we conclude that we cannot reject the hypothesis that the regression and IVR estimates of CATAS on PPO are equal. For NPO, we obtain $X^2 = 9.41$ (1), $p < .001$ and we conclude that the causal estimate of NPO on CATAS is larger than the regression estimate. We note that the correlation between NPO and catastrophizing disturbances under the IVR model is -.52 (SE = .10).

The estimated R^2 when predicting catastrophizing is .09 under the IVR model and .33 using regression. Recall that we cannot make a causal inference on R^2_{yx} under the IVR model. In fact, the R^2_{yx} may be negative when the IVR model is used and should be reported as ‘undefined’. The estimated R^2_{yx} may be negative because this parameter is simply one minus the estimated standardized error variance of the outcome, which may be larger than one.

Smaller standard errors, and therefore narrower confidence intervals, for the parameters of interest may be obtained by using a model more parsimonious than the IVR: correlated errors between y and $\mathbf{x}$ may be set to zero for some $\mathbf{x}$, or some $\mathbf{x}$ on $\mathbf{z}$ paths may be set to zero. Fitting such models is straightforward using a SEM approach. However, imposing such exclusion constraints should always be based on substantive theory. Simply setting to zero non-significant parameters capitalizes on chance and should be avoided, unless a cross-validation sample of new observations is available to verify the resulting model. For this example, we lack theory to suggest which parameters of the IVR model may be set to zero. For illustrative purposes only, we set the correlated errors between CATAS and PPO to zero. Using ML, a similar model fit is obtained as this parameter is not statistically significant. However, we obtain a narrower confidence interval for the standardized effect CATAS on NPO, C[$.74 \le \beta \le 1.16$] = .95, and a much narrower effect CATAS on PPO C[$-.03 \le \beta \le .28$] = .95 vs. C[$-0.10 \le \beta \le 0.73$] = .95 under the IVR model. Similarly, setting to zero instead PPO on A, PPO on O, and NPO on O results in a model with somewhat narrower confidence intervals for the parameters of interest (CATAS on PPO and NPO).

Discussion

Regression models are used to predict a quantitative variable from one or more predictors.[18] They lie at the core of much research across all disciplines. Most importantly, regression models are often used to represent researchers' causal assumptions about phenomena; that is, they are used as structural models. These assumptions arise from prior studies, the research design, and theory. Often, researchers do not explicitly lay out their causal assumptions; much less provide empirical support for them. Consider a regression model involving a single continuous predictor, x, of an outcome, y. The regression model is generally described by providing a justification for x having an impact on y, along with the sign (positive or negative) of the effect. Provided the justification is strong enough, this statement can be taken to be a causal assumption, even though, using the terminology of Bollen and Pearl (2013), it is a weak causal assumption, as it only puts forth a range of values for such an effect. For a regression model to be considered a structural model, all causal assumptions must be presented and defended, theoretically, empirically, or preferably both.

In this paper, we have described a model that enables researchers to draw empirical support for causal inferences in regression models (y on $\mathbf{x}$), even when data are observational. The model involves bringing in predictors ($\mathbf{z}$, aka instruments) of the predictors in the target regression model and fitting y on **x**, **x** on **z**, and y with **x** (i.e., the errors of y and **x** are correlated). This is the *instrumental variables regression* (IVR) model. When the IVR model cannot be rejected by the overall chi-square test of model fit, empirical support is obtained to draw causal inferences on y on **x** parameters. However, failure to reject the IVR model does not provide empirical support to draw causal inferences on **x** on **z** parameters, nor on y with **x** parameters, nor on the variance parameters of y nor **x**. The latter implies that causal inferences cannot be drawn on R^2_{yx} , the percentage of variance of the outcome of interest explained by our predictors.

In technical terms, causal inferences can be drawn on a parameter if it can be consistently estimated when the fitted model is misspecified in certain ways. In other words, within a class of models, causal inferences can be drawn on a parameter θ_i if its value can be estimated in large samples regardless of the data generating mechanism. Whether or not causal inferences can be drawn depends on the data generating mechanism and the fitted model; not on the estimation method used. To draw causal inferences for the class of models considered in this paper, applied researchers can use either estimation methods that predominate in Econometrics (e.g., 2SLS) or simply use familiar estimators in structural equations modeling software programs; see Maydeu-Olivares, Shi and Rosseel (2019) for a Monte-Carlo comparison of both approaches.

The ML estimator described in the SEM literature is often referred to as full information ML (FIML) estimator in the Econometrics literature. In the case of complete data, and when the IVR model involves a single outcome variable y (as in the models considered in the paper), the ML/FIML estimator of the IVR model has a closed form solution (Anderson & Rubin, 1949, 1950) and it is referred to in the Econometrics literature as limited information ML (LIML estimator) –see Davidson and MacKinnon(2004) for technical details. Thus, for the models considered here, the ML estimator used in SEM is referred to as LIML in Econometrics. Also, in the Econometrics literature, standard errors and goodness of fit statistics obtained under ADF assumptions are referred to as robust to heteroscedasticity. In the SEM literature, they are referred to as robust to non-normality. ADF standard errors and goodness of fit tests are robust to departures from normality and homoscedasticity. Finally, what is referred to in the Econometrics literature as 'tests for exclusion constraints', or 'test of overidentifying restrictions', is the overall goodness of fit test statistic in SEM (i.e., the chi-square test).

An advantage of using SEM software is that it enables fitting the IVR model with correlated errors between y and $\mathbf{z}$ (i.e., a model with endogenous instruments) as depicted in Figure 7. The correlations between instruments and the errors of the outcome can be fixed to different values to perform a sensitivity analysis of the causal inferences. Alternatively, some of them can be estimated from data. Causal inferences can be drawn on y on $\mathbf{x}$ parameters when the IVR model with endogenous predictors is correctly specified. In addition, SEM software provides modification indices (e.g., Lagrange or score test statistics) for the correlated errors between y and $\mathbf{z}$ when the IVR model is fitted. When the IVR model is rejected by the overall chi-square test of model fit, these modification indices may be used to determine the source of misfit and correlated errors between y and $\mathbf{z}$ may be added to the IVR model. However, modification indices are uninformative when there is only one degree of freedom available for testing[19]. Finally, the use of a SEM framework enables using instrumental variable methods for structural equation models including latent variables (Bollen, 1996a, 1996b, 2001, 2018; Bollen, Kolenikov, & Bauldry, 2014; Kirby & Bollen, 2009), even when the observed variables are discrete (Bollen & Maydeu-Olivares, 2007; Jin & Cao, 2018; Nestler, 2013).

Insert Figure 7 about here

The correlation between the errors of *y* and **x** plays an important role in the IVR model. If its value is zero the IVR parameter estimates will be equal to those obtained using two separate regressions (*y* on **x**, and **x** on **z**, Land, 1973). As the magnitude of the correlations between the residuals of *y* and **x** increases when fitting an IVR model, the *y* on **x** intercepts and slopes will depart from those obtained using a *y* on **x** regression at the expense of larger SEs. The inflation of the IVR SEs relative to those obtained in regression is governed by the magnitude of the correlations between the regression predictors, **x**, and the instruments, **z** (the larger the better).

In closing, we believe that instrumental variable methods may be more valuable in psychological research that in economic research because relationships between variables in psychological research are generally strong. However, these methods have been underused in psychological research, perhaps because they were not well understood. We hope that our integration of instrumental variable regression methods within a SEM framework will enhance their use in psychological research and related fields. We expect and look forward to many more applications of these methods in psychological research.

References

Anderson, T. W., & Rubin, H. (1949). Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations. *The Annals of Mathematical Statistics*, *20*(1), 46–63. https://doi.org/10.1214/aoms/1177730090

Anderson, T. W., & Rubin, H. (1950). The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equations. *The Annals of Mathematical Statistics*, *21*(4), 570–582. https://doi.org/10.1214/aoms/1177729752

Arbuckle, J. L. (2014). Amos (version 23) [Computer program]. Chicago, IL: IBM SPSS.

Asparouhov, T., & Muthén, B. (2006). Robust Chi Square Difference Testing with Mean and Variance Adjusted Test Statistics. *Mplus Web Notes: No 10*. Retrieved from http://www.statmodel.com/download/webnotes/webnote10.pdf

Bentler, P. M. (2004). EQS 6 [Computer Program]. Encino, CA: Multivariate Software Inc.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Bollen, K. A. (1996a). A limited-information estimator for LISREL models with and without heteroscedastic errors. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced Structural Equation Modeling* (pp. 227–241). Mahwah, NJ: Lawrence Erlbaum.

Bollen, K. A. (1996b). An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika*, *61*(1), 109–121. https://doi.org/10.1007/BF02296961

Bollen, K. A. (2001). Two-stage least squares and latent variable models: Simultaneous estimation and robustness to misspecifications. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural Equation Modeling: Present and Future: A Festschrift in honor of Karl Jöreskog* (pp. 119–138). Lincolnwood, IL: Scientific Software International.

Bollen, K. A. (2012). Instrumental Variables in Sociology and the Social Sciences. *Annual*

Review of Sociology, *38*(1), 37–72. https://doi.org/10.1146/annurev-soc-081309-150141

Bollen, K. A. (2018). Model Implied Instrumental Variables (MIIVs): An Alternative Orientation to Structural Equation Modeling. *Multivariate Behavioral Research*, *0*(0), 1–16. https://doi.org/10.1080/00273171.2018.1483224

Bollen, K. A., Kirby, J. B., Curran, P. J., Paxton, P. M., & Chen, F. (2007). Latent Variable Models Under Misspecification: Two-Stage Least Squares (2SLS) and Maximum Likelihood (ML) Estimators. *Sociological Methods & Research*, *36*(1), 48–86. https://doi.org/10.1177/0049124107301947

Bollen, K. A., Kolenikov, S., & Bauldry, S. (2014). Model-Implied Instrumental Variable—Generalized Method of Moments (MIIV-GMM) Estimators for Latent Variable Models. *Psychometrika*, *79*(1), 20–50. https://doi.org/10.1007/s11336-013-9335-3

Bollen, K. A., & Maydeu-Olivares, A. (2007). A polychoric instrumental variable (PIV) estimator for Structural Equation Models with categorical variables. *Psychometrika*, *72*(3), 309–326. https://doi.org/10.1007/s11336-007-9006-3

Bollen, K. A., & Pearl, J. (2013). Eight myths about causality and structural equation models. In S. L. Morgan (Ed.), *Handbook of Causal Analysis for Social Research* (pp. 301–328). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-6094-3

Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak. *Journal of the American Statistical Association*, *90*(430), 443–450. https://doi.org/10.1080/01621459.1995.10476536

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72–141). Cambridge, UK: Cambridge University Press.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*(1), 62–83. https://doi.org/10.1111/j.2044-8317.1984.tb00789.x

Cameron, A., & Trivedi, P. (2005). *Microeconometrics: methods and applications*. New York: Cambridge University Press.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual.* Odessa, FL: Psychological Assessment Resources.

Cragg, J. G., & Donald, S. G. (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory*, *9*(2), 222–240. https://doi.org/10.1017/S0266466600007519

D'Zurilla, T. J., & Goldfried, M. R. (1971). Problem solving and behavior modification. *Journal of Abnormal Psychology*, *78*(1), 107–126. https://doi.org/10.1037/h0031360

D'Zurilla, T. J., Maydeu-Olivares, A., Gallardo-Pujol, D., D'Zurilla, T. J., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2011). Predicting social problem solving using personality traits. *Personality and Individual Differences*, *50*(2), 142–147. https://doi.org/10.1016/j.paid.2010.09.015

D'Zurilla, T. J., Nezu, A. M., & Maydeu-Olivares, A. (2002). Manual of the Social Problem-Solving Inventory-Revised. North Tonawanda, NY: Multi-Health Systems, Inc.

D'Zurilla, T. J., Nezu, A. M., & Maydeu-Olivares, A. (2004). Social problem solving: Theory and assessment. In E. C. Chang, T. J. D'Zurilla, & L. J. Sanna (Eds.), *Social problem solving: Theory, research, and training* (pp. 11–27). Washington, DC, US: American Psychological Association. https://doi.org/10.1037/10805-000

Davidson, R., & MacKinnon, J. G. (2004). *Econometric Theory and Methods*. Oxford, England: Oxford University Press.

Ellis, A. (1962). *Reason and emotion in psychotherapy*. New York: Lyle Stuart.

Foster, E. M., & McLanahan, S. (1996). An illustration of the use of instrumental variables: Do neighborhood conditions affect a young person's chance of finishing high school? *Psychological Methods*, *1*(3), 249–260. https://doi.org/10.1037//1082-989X.1.3.249

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*(1), 26–34. https://doi.org/10.1037/0003-066X.48.1.26

Greene, W. H. (2007). *Econometric analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, *50*(4), 1029. https://doi.org/10.2307/1912775

Jin, S., & Cao, C. (2018). Selecting polychoric instrumental variables in confirmatory factor analysis: An alternative specification test and effects of instrumental variables. *British Journal of Mathematical and Statistical Psychology*, *71*(2), 387–413. https://doi.org/10.1111/bmsp.12128

Jöreskog, K. G., Olsson, U. H., & Wallentin, F. J. (2016). *Multivariate analysis with LISREL*. New York: Springer. https://doi.org/10.1007/978-0-387-98135-2

Jöreskog, K. G., & Sörbom, D. (2017). LISREL (Version 9.3) [Computer program]. Chicago, IL: Scientific Software International.

Kirby, J. B., & Bollen, K. A. (2009). Using Instrumental Variable Tests To Evaluate Model Specification in Latent Variable Structural Equation Models. *Sociological Methodology*, *39*(1), 327–355. https://doi.org/10.1111/j.1467-9531.2009.01217.x

Land, K. (1973). Identification, parameter estimation, and hypothesis testing in recursive sociological models. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural Equation Models in the Social Sciences* (pp. 19–49). New York: Seminar Press.

Lindner, H., Kirkby, R., Wertheim, E., & Birch, P. (2007). Scala de atitudini si convingeri generale, forma scurta – GABS-SV. In D. David (Ed.), *Sistem de evaluare clinica*. Cluj-Napoca: RTC.

Maydeu-Olivares, A. (2006). Limited information estimation and testing of discretized multivariate normal structural models. *Psychometrika*, *71*(1), 57–77. https://doi.org/10.1007/s11336-005-0773-4

Maydeu-Olivares, A. (2017). Maximum likelihood estimation of structural equation models for continuous data: Standard errors and goodness of fit. *Structural Equation Modeling*, *24*(3), 383–394. https://doi.org/10.1080/10705511.2016.1269606

Maydeu-Olivares, A., & D'Zurilla, T. J. (1996). A factor-analytic study of the Social Problem-Solving Inventory: An integration of theory and data. *Cognitive Therapy and Research*, *20*(2), 115–133. https://doi.org/10.1007/BF02228030

Maydeu-Olivares, A., & Lado, N. (2003). Market orientation and business economic performance: A mediated model. *International Journal of Service Industry Management*, *14*(3), 284–309. https://doi.org/10.1108/09564230310478837

Maydeu-Olivares, A., Rosseel, Y., & Shi, D. (2019). Instrumental variable regression: A mean and covariance structure approach. *Manuscript under Review*.

Maydeu-Olivares, A., & Shi, D. (2017). Effect Sizes of Model Misfit in Structural Equation Models: Standardized Residual Covariances and Residual Correlations. *Methodology*, *13*(Supplement 1), 23–30. https://doi.org/10.1027/1614-2241/a000129

Maydeu-Olivares, A., Shi, D., & Rosseel, Y. (2019). Instrumental variables two-stage least squares (2SLS) vs. maximum likelihood structural equation modeling of causal effects in linear regression models. *Structural Equation Modeling*.

Morgan, S. L., & Winship, C. (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge, UK: Cambridge University Press.

Muthén, L. K., & Muthén, B. (2017). MPLUS 8 [Computer program]. Los Angeles, CA: Muthén & Muthén.

Nestler, S. (2013). A Monte Carlo study comparing PIV, ULS and DWLS in the estimation of dichotomous confirmatory factor analysis. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 127–43. https://doi.org/10.1111/j.2044-8317.2012.02044.x

Peugh, J. L., Strotman, D., McGrady, M., Rausch, J., & Kashikar-Zuck, S. (2017). Beyond intent to treat (ITT): A complier average causal effect (CACE) estimation primer. *Journal of School Psychology*, *60*, 7–24. https://doi.org/10.1016/j.jsp.2015.12.006

R Development Core Team. (2015). R: A Language and Environment for Statistical Computing. Vienna, Austria.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701. https://doi.org/10.1037/h0037350

Rubin, D. B. (2006). *Matched sampling for causal effects*. *Matched Sampling for Causal Effects*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511810725

Sanderson, E., & Windmeijer, F. (2016). A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics*, *190*(2), 212–221.

https://doi.org/10.1016/j.jeconom.2015.06.004

Sargan, J. D. (1958). The Estimation of Economic Relationships using Instrumental Variables. *Econometrica*, *26*(3), 393. https://doi.org/10.2307/1907619

SAS Institute. (2014). *SAS/ETS® 13.2 User's Guide. The SYSLIN Procedure*. Cary, NC: SAS Institute Inc.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. Von Eye & C. C. Clogg (Eds.), *Latent variable analysis. Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*(4), 507–514. https://doi.org/10.1007/BF02296192

Satorra, A., & Bentler, P. M. (2010). Ensuring Positiveness of the Scaled Difference Chi-square Test Statistic. *Psychometrika*, *75*(2), 243–248. https://doi.org/10.1007/s11336-009-9135-y

Schochet, P. Z., & Chiang, H. S. (2011). Estimation and Identification of the Complier Average Causal Effect Parameter in Education RCTs. *Journal of Educational and Behavioral Statistics*, *36*(3), 307–345. https://doi.org/10.3102/1076998610375837

Sörbom, D. (1989). Model modification. *Psychometrika*, *54*(3), 371–384. https://doi.org/10.1007/BF02294623

Staiger, D., & Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, *65*(3), 557. https://doi.org/10.2307/2171753

StataCorp. (2017). Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC.

Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential Chi-square statistics. *Psychometrika*, *50*(3), 253–263.

Stock, J. H., Wright, J. H., & Yogo, M. (2002). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business & Economic Statistics*, *20*(4), 518–529. https://doi.org/10.1198/073500102288618658

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In J. H. Stock & D. W. K. Andrews (Eds.), *Identification and inference for econometric models: Essays in honor of Thomas J. Rothenberg* (pp. 80–108). Cambridge, UK: Cambridge University Press.

Stuart, E. A., & Jo, B. (2015). Assessing the sensitivity of methods for estimating principal causal effects. *Statistical Methods in Medical Research*, *24*(6), 657–674. https://doi.org/10.1177/0962280211421840

Suso-Ribera, C., Jornet-Gibert, M., Ribera Canudas, M. V., McCracken, L. M., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2016). There's More Than Catastrophizing in Chronic Pain: Low Frustration Tolerance and Self-Downing Also Predict Mental Health in Chronic Pain Patients. *Journal of Clinical Psychology in Medical Settings*, *23*(2), 192–206. https://doi.org/10.1007/s10880-016-9454-y

Theil, H. (1953). *Estimation and simultaneous correlation in complete equation systems*. The Hague: Central Planning Bureau.

Vanderweele, T. J. (2015). *Explanation in causal inference. Methods for mediation and interaction*. Oxford, England: Oxford University Press.

Wooldridge, J. M. (2013). *Introductory Econometrics. A modern approach* (5th ed.). Mason, OH: South-Western.

Yau, L. H. Y., & Little, R. J. (2001). Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *Journal of the American Statistical Association*, *96*(456), 1232–1244. https://doi.org/10.1198/016214501753381887

Appendix

Identifying potential instruments in path analysis models

Consider a path analysis model with p variables, $\mathbf{y} = \mathbf{B}\mathbf{y} + \boldsymbol{\varepsilon}$. We let $\mathrm{cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}$. The covariance structure implied by the model is

$$\boldsymbol{\Sigma}_0 = \left(\mathbf{I} - \mathbf{B}\right)^{-1} \boldsymbol{\Psi} \left(\mathbf{I} - \mathbf{B}\right)^{-1\prime}. \qquad (1)$$

Also, since $\mathbf{y} = \left(\mathbf{I} - \mathbf{B}\right)^{-1} \boldsymbol{\varepsilon}$

$$\mathbf{E} \equiv \mathrm{cov}(\mathbf{y}\boldsymbol{\varepsilon}') = \left(\mathbf{I} - \mathbf{B}\right)^{-1} \boldsymbol{\Psi}. \qquad (2)$$

As a result, given a structural model, potential instruments **z** can be identified as variables that a) have non-zero covariances with **x** variables in (1), and b) have zero covariances with the disturbance of *y* in (2).

For the market share example, let **y** = {market share, innovation performance, innovation degree, customer loyalty, and market orientation}. According to Figure 4,

$$\mathbf{B} = \begin{pmatrix} 0 & \beta_{12} & 0 & \beta_{14} & 0 \\ 0 & 0 & \beta_{23} & 0 & \beta_{25} \\ 0 & 0 & 0 & 0 & \beta_{35} \\ 0 & 0 & 0 & 0 & \beta_{45} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \ \boldsymbol{\Psi} = \begin{pmatrix} \psi_{11} & 0 & 0 & 0 & 0 \\ 0 & \psi_{22} & 0 & 0 & 0 \\ 0 & 0 & \psi_{33} & 0 & 0 \\ 0 & 0 & 0 & \psi_{44} & 0 \\ 0 & 0 & 0 & 0 & \psi_{55} \end{pmatrix}. \qquad (3)$$

Potential instruments correspond to a) rows with non-zero entries in column 2 (the predictor *x*) in $\boldsymbol{\Sigma}_0$, and b) rows with zero entries in column 1 (the error term for the outcome *y*) in $\mathbf{E}$. Under the model, all variables are correlated (there are no zero entries in $\boldsymbol{\Sigma}_0$). Therefore, innovation degree, customer loyalty, and market orientation meet condition a). Also, for this model,

$$\mathbf{E}=\begin{pmatrix}\psi_{11} & \beta_{12}\psi_{22} & \beta_{12}\left(\beta_{23}\psi_{33}+0\right) & \beta_{14}\psi_{44} & \left(\beta_{12}\left(\beta_{25}+\beta_{23}\beta_{35}\right)+\beta_{14}\beta_{45}\right)\psi_{55}\\ 0 & \psi_{22} & \beta_{23}\psi_{33}+0 & 0 & \left(\beta_{25}+\beta_{23}\beta_{35}\right)\psi_{55}\\ 0 & 0 & \psi_{33} & 0 & \beta_{35}\psi_{55}\\ 0 & 0 & 0 & \psi_{44} & \beta_{45}\psi_{55}\\ 0 & 0 & 0 & 0 & \psi_{55}\end{pmatrix}, \qquad (4)$$

and innovation degree, customer loyalty, and market orientation (variables 3 to 5) have zero entries in column 1. Therefore, these variables also meet condition b) and they are potential instrumental variables under the model.

The instrumental variables regression (IVR) model

For ease of exposition, we assume that all variables are mean centered so that there is no mean structure (all means and intercepts are zero). Letting $\mathbf{w}'=\left(y,\mathbf{x}',\mathbf{z}'\right)'$, the IVR model can be written in matrix form as $\mathbf{w}=\mathbf{B}\mathbf{w}+\boldsymbol{\varepsilon}=\left(\mathbf{I}-\mathbf{B}\right)^{-1}\boldsymbol{\varepsilon}$, and we denote the covariance matrix of the disturbances $\boldsymbol{\varepsilon}$ by $\boldsymbol{\Psi}$. $\mathbf{B}$ and $\boldsymbol{\Psi}$ can be partitioned according to the partitioning of $\mathbf{w}$ as

$$\mathbf{B}=\begin{pmatrix}\mathbf{0} & \boldsymbol{\beta}_{yx} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{B}_{xz}\\ \mathbf{0} & \mathbf{0} & \mathbf{0}\end{pmatrix}, \qquad \boldsymbol{\Psi}=\begin{pmatrix}\psi_{yy} & & \\ \boldsymbol{\psi}_{xy} & \boldsymbol{\Psi}_{xx} & \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Psi}_{zz}\end{pmatrix}, \qquad (5)$$

with implied covariance structure $\boldsymbol{\Sigma}_0=\left(\mathbf{I}-\mathbf{B}\right)^{-1}\boldsymbol{\Psi}\left(\mathbf{I}-\mathbf{B}\right)^{-1\prime}$, which can be partitioned according to the partitioning of $\mathbf{w}$ as (1). There are no exclusion constraints (parameters set to zero) in any of the submatrices in (5); that is, the model assumes that all $\mathbf{x}$ and y errors are correlated, every $\mathbf{x}$ is determined by every $\mathbf{z}$, and that all $\mathbf{x}$ determine y.

Weak instruments

The F statistic to predict x from $\mathbf{z}$ to assess the weakness of instruments can be written as

$$F=\frac{R_{xz}^2}{1-R_{xz}^2}\frac{N-p_z-1}{p_z}, \qquad (6)$$

and it follows an $F(p_z, N - p_z - 1)$ distribution. Notice that equation (6) reveals that whether the instruments are strong or weak depends on combination of the strength of the relationship between x and $\mathbf{z}$ (i.e., R^2_{xz}), and sample size (N).

The Cragg and Donald (1993) statistic is

$$T_{CD} = N\hat{\mathbf{\Psi}}_{xx}^{-1/2}\hat{\mathbf{B}}_{xz}\hat{\mathbf{\Sigma}}_{zz}\hat{\mathbf{B}}_{xz}{}'\hat{\mathbf{\Psi}}_{xx}^{-1/2} \quad . \tag{7}$$

Under normality assumptions, the trace of the Cragg and Donald statistic is the Wald statistic for testing that there is no relationship between $\mathbf{x}$ and $\mathbf{z}$, i.e., for testing $\mathbf{B}_{xz} = \mathbf{0}$. The Wald statistic follows in large samples a chi-square distribution with degrees of freedom equal to the parameters in $\mathbf{B}_{zv}$. When there is only a single predictor x, the Wald statistic (7) is related to the F statistic (6) as follows:

$$F = \frac{(N - p_z - 1)}{p_z}\frac{T_{CD}}{N} \quad . \tag{8}$$

Standardized parameters

The standardized slopes and standardized variance/covariance parameters of the IVR model, $\mathbf{B}^*$ and $\mathbf{\Psi}^*$, are obtained from the unstandardized parameters using

$$\mathbf{B}^* = \mathbf{DBD}^{-1} \ , \qquad\qquad \mathbf{\Psi}^* = \mathbf{D\Psi D} , \tag{9}$$

where $\mathbf{D} = \left(\mathrm{diag}(\mathbf{\Sigma}_0)\right)^{1/2}$. Using (9), we can verify as we did with the unstandardized parameters presented in Table 1 that by fitting the IVR model we obtain the standardized slopes y on $\mathbf{x}$ when the data mechanism is reverse causation, reciprocal causation, or includes omitted variables. Therefore, we can make causal inferences on standardized slopes as well.

Footnotes

[1] The 2SLS estimation procedure and statistical theory to obtain SEs for parameter estimates is in many ways analogous to procedures used to estimate polychoric correlations and their SEs (see Maydeu-Olivares, 2006).

[2] All technical materials are provided in the Appendix. Also, the mean structure is not discussed: We assume the regression intercepts are not of interest as it improves readability.

[3] Since $e^{.11} = 1.11$, recall that market share is in logs. The log transformation converts a multiplicative model into a linear one. Further details on the rationale, use and interpretation of logarithmic transformations in regression models can be found in Wooldridge (2013, pp. 43-44) and Jöreskog, Olsson and Wallentin (2016, p. 55)

[4] Consider for instance that the reverse causation model in Figure 2a is the data generating mechanism. Under this DGM, the true parameter of interest, β_{yx}, equals zero. However, if we fit our target model to data generated using the reverse causation DGM we will not estimate $\beta_{yx} = 0$. Rather, we will be estimating $\beta_{yx} = \dfrac{\tilde{\beta}_{xy}\tilde{\psi}_{yy}}{\tilde{\beta}_{xy}^2\tilde{\psi}_{yy} + \tilde{\psi}_{xx}}$, with $\tilde{\psi}_{yy} = \mathrm{var}(\varepsilon_y)$ and $\tilde{\psi}_{xx} = \mathrm{var}(\varepsilon_{xx})$, where tildes denote the parameters of the DGM. To see this, we recall that the target model (*y* on *x*) and the reverse causation model (*x* on *y*) are equivalent. Then, we gather the parameters in the reverse causation model in the vector $\tilde{\boldsymbol{\theta}}' = (\tilde{\beta}_{xy}, \tilde{\psi}_{yy}, \tilde{\psi}_{xx})$, and the parameters of our target model in the vector $\boldsymbol{\theta}' = (\beta_{yx}, \psi_{yy}, \psi_{xx})$. We write the covariance structures implied by each model, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\tilde{\boldsymbol{\theta}})$, set them equal, $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\tilde{\boldsymbol{\theta}})$, and solve for $\boldsymbol{\theta}$ as a function of $\tilde{\boldsymbol{\theta}}$.

[5] As before, we write the covariance structures implied by each model, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\tilde{\boldsymbol{\theta}})$, set them equal, $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\tilde{\boldsymbol{\theta}})$, and solve for $\boldsymbol{\theta}$ as a function of $\tilde{\boldsymbol{\theta}}$.

[6] The model in Figure 3a is a special case of the model in Figure 3b. Models in Figures 3a to 3c are a special case of the model in Figure 3d.

[7] We also assume that the *y* on *x* relationship is consistent across levels of *x* (β_{yx} does not change for different levels of *z*), and that the *x* on **z** relationships are consistent across levels of **z**. Within a frequentist framework, these assumptions are not testable and can only be supported theoretically.

[8] These variables were obtained by summing the responses by each company's senior executive to sets of items measuring these attributes. Therefore, they are also self-reported measures on these attributes.

[9] 2SLS estimation of the IVR model is a misnomer, as estimating all the parameters of the model requires three stages. In the first stage, one estimates the parameters involved in the **x** on **z** regression using only the **x** and **z** variables. In the second stage, the slopes of *y* on **x** are estimated from the covariances between **z** and *y*, holding the parameters estimated in the first stage fixed. Finally, in the third stage, the error variance of *y* and the covariances between the errors of *y* and **x** are estimated from the sample variance of *y* and the sample covariances between **x** and *y*, respectively, holding the first and second stage estimates fixed. Software implementation of 2SLS focuses on the second stage estimation results: the *y* on **x** estimates. Often, third stage results are not reported. Some programs do not report first stage estimates either. Standard errors for 2SLS can be obtained under normality assumptions or robust to non-normality (under ADF assumptions). The overall goodness of fit of the model can be tested using either Sargan's (1958)

test statistic or the *J* test proposed by Hansen (1982), under normality or ADF assumptions, respectively.

[10] This is $(N - 1)/N$ times the sample covariance matrix, where *N* denotes sample size.

[11] The ADF assumptions are that the observations are independently and identically distributed with finite first eight order moments.

[12] This test statistic is sometimes referred to the literature as T_B and should not be confused with another test statistic proposed by Browne (1982, 1984) obtained as the minimum of of the weighted least squares under ADF assumptions. The latter is only asymptotically chi-square for that specific fit function.

[13] When Browne's test statistic is used to assess the fit of the IVR model under normality assumptions using the 2SLS estimates, it reduces to Sargan's statistic. When Browne's statistic is computed under ADF assumptions using the 2SLS estimates, it reduces to the *J* statistic.

[14] In small samples, the mean and variance correction provides more accurate *p*-values (Maydeu-Olivares, 2017).

[15] Degrees of freedom are p_z and $N - p_z - 1$.

[16] Although the fit of the model when only one instrument is used cannot be tested, since in this case we could not reject the IVR model when two instruments were used, we could safely use a model with just one instrument. In fact, when only the best instrument (innovation degree) is used (and its SE) equal the values reported in Table 2 (with 2 digit precision).

[17] Strictly speaking, and as in any other SEM model, if the overall test statistic fails to reject the model, and if the residual graphs fail to reveal any model misspecification, all that can be

concluded is that we do not have empirical evidence to reject the conclusions drawn using the model. This does not necessarily mean that the inferences are correct.

[18] Dummy variables can also be included as predictors in a regression model.

[19] When there is only one degree of freedom, modification indices for all free parameters are of the same magnitude and similar to the magnitude of the chi-square statistic to assess overall model fit. In this case, standardized residual covariances (Maydeu-Olivares & Shi, 2017) are more useful than modification indices. Further research is needed on this topic.

Table 1

IVR population parameters obtained for different data generating mechanisms

IVR parameters (Figure 1b)	Data generating mechanisms			
	Reverse causation (Figure 3a)	Reciprocal causation (Figure 3b)	Omitted variables (Figure 3c)	Omitted variables and reciprocal causation (Figure 3d)
β_{yx}	$\tilde{\beta}_{yx} = 0$	$\tilde{\beta}_{yx}$	$\tilde{\beta}_{yx}$	$\tilde{\beta}_{yx}$
ψ_{yy}	$\tilde{\psi}_{yy}$	$\tilde{\psi}_{yy}$	$\tilde{\beta}^2_{yz}\tilde{\psi}_{zz} + \tilde{\psi}_{yy}$	$\tilde{\beta}^2_{yz}\tilde{\psi}_{zz} + \tilde{\psi}_{yy}$
β_{xz_1}	$\tilde{\beta}_{xz_1}$	$\dfrac{\tilde{\beta}_{zv_1}}{1-\tilde{\beta}_{yx}\tilde{\beta}_{xy}}$	$\tilde{\beta}_{xz_1}$	$\dfrac{\tilde{\beta}_{zv_1}}{1-\tilde{\beta}_{yx}\tilde{\beta}_{xy}}$
β_{xz_2}	$\tilde{\beta}_{xz_2}$	$\dfrac{\tilde{\beta}_{zv_2}}{1-\tilde{\beta}_{yx}\tilde{\beta}_{xy}}$	$\tilde{\beta}_{xz_2}$	$\dfrac{\tilde{\beta}_{zv_2}}{1-\tilde{\beta}_{yx}\tilde{\beta}_{xy}}$
ψ_{yx}	$\tilde{\beta}_{xy}\tilde{\psi}_{yy}$	$\dfrac{\tilde{\beta}_{xy}\tilde{\psi}_{yy}}{1-\tilde{\beta}_{yx}\tilde{\beta}_{xy}}$	$\tilde{\beta}_{yv}\tilde{\beta}_{zv}\tilde{\psi}_{vv}$	$\dfrac{\tilde{\beta}_{xy}\tilde{\beta}^2_{yv}\tilde{\psi}_{vv} + \tilde{\beta}_{yv}\tilde{\beta}_{zv}\tilde{\psi}_{vv} + \tilde{\beta}_{xy}\tilde{\psi}_{yy}}{1-\tilde{\beta}_{yx}\tilde{\beta}_{xy}}$
ψ_{xx}	$\tilde{\beta}^2_{xy}\tilde{\psi}_{yy} + \tilde{\psi}_{xx}$	$\dfrac{\tilde{\beta}^2_{xy}\tilde{\psi}_{yy} + \tilde{\psi}_{xx}}{\left(1-\tilde{\beta}_{yx}\tilde{\beta}_{xy}\right)^2}$	$\tilde{\beta}^2_{zv}\tilde{\psi}_{vv} + \tilde{\psi}_{xx}$	$\dfrac{\tilde{\beta}^2_{zv}\tilde{\psi}_{vv} + \tilde{\psi}_{xx} + \tilde{\beta}^2_{xy}\left(\tilde{\beta}^2_{yv}\tilde{\psi}_{vv} + \tilde{\psi}_{yy}\right) + 2\tilde{\beta}_{xy}\tilde{\beta}_{yv}\tilde{\beta}_{zv}\tilde{\psi}_{vv}}{\left(1-\tilde{\beta}_{yx}\tilde{\beta}_{xy}\right)^2}$
$\psi_{z_1z_1}$	$\tilde{\psi}_{z_1z_1}$	$\tilde{\psi}_{z_1z_1}$	$\tilde{\psi}_{z_1z_1}$	$\tilde{\psi}_{z_1z_1}$
$\psi_{z_2z_1}$	$\tilde{\psi}_{z_2z_1}$	$\tilde{\psi}_{z_2z_1}$	$\tilde{\psi}_{z_2z_1}$	$\tilde{\psi}_{z_2z_1}$
$\psi_{z_2z_2}$	$\tilde{\psi}_{z_2z_2}$	$\tilde{\psi}_{z_2z_2}$	$\tilde{\psi}_{z_2z_2}$	$\tilde{\psi}_{z_2z_2}$

Note: the parameters of the data generating mechanisms of Figure 3 are denoted in this table by $\tilde{\theta}$ and the estimated population IVR parameters by θ. θ is obtained by setting $\sigma(\theta) = \sigma(\tilde{\theta})$ and solving for θ.

Table 2

ML and 2SLS estimates, standard errors, and goodness of fit test for the IVR model applied to the market share and innovation performance example

IVR parameters (Figure 1b)	ML			2SLS		
	est	NT SE	ADF SE	est	NT SE	ADF SE
β_{yx}	0.20	0.03	0.04	0.20	0.03	0.04
β_{yz_1}	0.13	0.03	0.04	0.13	0.03	0.04
β_{yz_2}	0.77	0.13	0.13	0.78	0.14	0.15
ψ_{yy}	1.62	0.26	0.37	1.62	0.26	0.37
ψ_{yx}	-2.38	0.68	0.93	-2.38	0.68	0.93
ψ_{xx}	14.36	1.84	2.59	14.36	1.84	2.58
$\psi_{z_1z_1}$	175.66	22.49	22.39	175.66	22.49	22.32
$\psi_{z_2z_1}$	21.27	3.98	3.64	21.27	3.98	3.62
$\psi_{z_2z_2}$	8.43	1.08	1.00	8.43	1.08	0.99
R^2_{yx}	.06	-	-	.06	-	-
R^2_{xv}	.46	-	-	.46	-	-

Goodness of fit test (df = 1)

ML				2SLS			
NT		ADF		NT		ADF	
X^2	p	X^2	p	X^2	p	X^2	p
.01	.91	.01	.92	.01	.91	.01	.92

Notes: NT = results under normality, ADF = results robust to non-normality and heteroscedasticity. The ML X^2 is the likelihood ratio (LR) statistic (NT) and the mean and variance corrected LR statistic (ADF). The 2SLS X^2 is Sargan’s statistic (NT), and the *J* statistic (ADF). Both are algebraically equal to Browne’s test statistic based on residual covariances.

Table 3

Standardized ML and 2SLS estimates, standard errors, and goodness of fit tests for the IVR model applied to the catastrophizing and problem orientation example. Regression estimates also provided for comparison

	ML		2SLS		regression	
	est	SE	est	SE	est	SE
CATAS ← PPO	0.32	0.21	0.28	0.20	<0.01	0.07
CATAS ← NPO	1.07	0.17	1.03	0.16	0.58	0.06
PPO ← N	-0.05	0.07	-0.05	0.07		
PPO ← O	0.19	0.06	0.18	0.06		
PPO ← A	-0.01	0.06	-0.01	0.06		
PPO ← C	0.48	0.07	0.48	0.07		
NPO ← N	0.46	0.06	0.47	0.06		
NPO ← O	-0.01	0.05	-0.05	0.06		
NPO ← A	0.16	0.05	0.17	0.06		
NPO ← C	-0.27	0.06	-0.24	0.07		
CATAS ↔ PPO	-0.19	0.17	-0.13	0.14		
CATAS ↔ NPO	-0.52	0.10	-0.36	0.12		
PPO ↔ NPO	-0.11	0.08	-0.07	0.05		
R^2 (CATAS)	0.09	-	0.14	-	0.33	-
R^2 (PPO)	0.31	-	0.31	-		
R^2 (NPO)	0.39	-	0.39	-		

Goodness of fit test (df = 2)

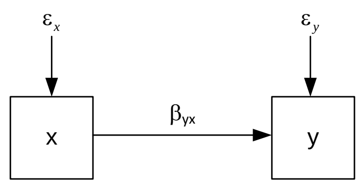
ML		2SLS	
X^2	p	X^2	p
2.11	0.35	2.15	0.34

Note: Results under normality. CATAS = catastrophizing, PPO = positive problem orientation, NPO = negative problem orientation, N = Neuroticism, O = Openness to experience, A = Agreeableness, C = Conscientiousness. The ML X^2 is the likelihood ratio (LR) statistic (NT). The 2SLS X^2 is Sargan's statistic (NT), algebraically equal to Browne's test statistic based on residual covariances.

Figure 1

Instrumental variables regression (IVR) model that enables drawing causal inferences on the target model

a) Target model



b) Instrumental variables regression (IVR) model

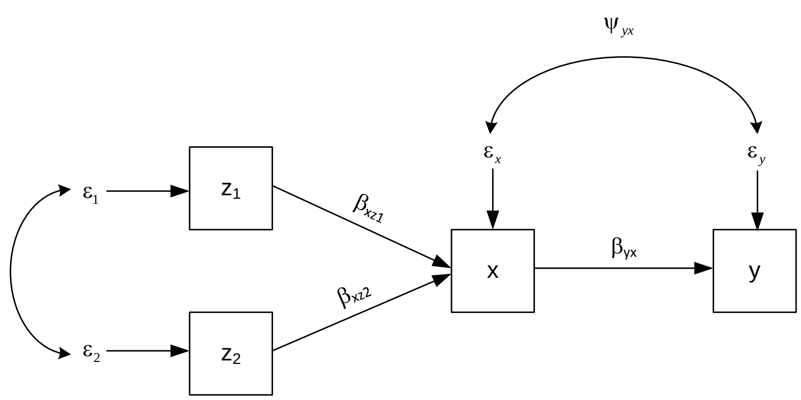
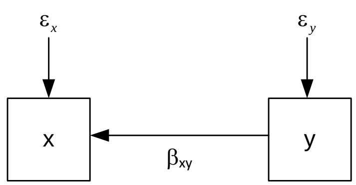


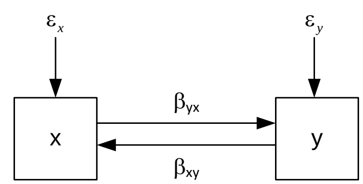
Figure 2

Threats to the validity of causal inferences drawn on the target model (competing data generating mechanisms)

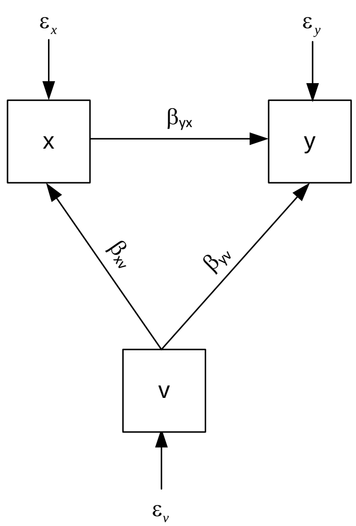
a) Reverse causation



b) Reciprocal causation



c) Omitted variables



d) Correlated errors

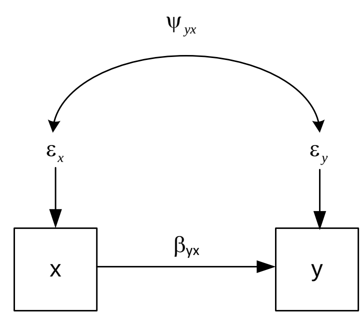
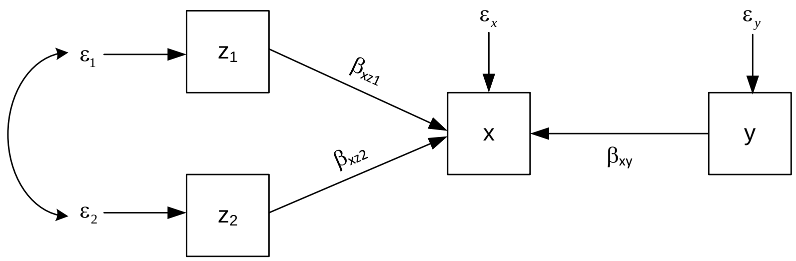


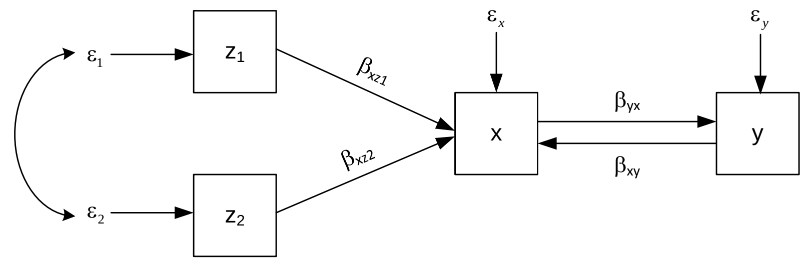
Figure 3

Competing data generating mechanisms to the IVR model obtained by incorporating predictors of x to the models of Figure 2

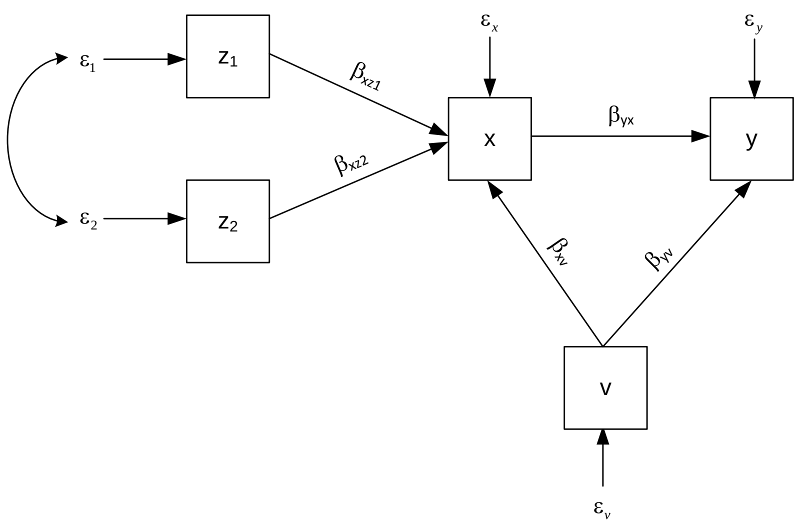
a) Reverse causation



b) Reciprocal causation



c) Omitted variables



d) Reciprocal causation and omitted variables

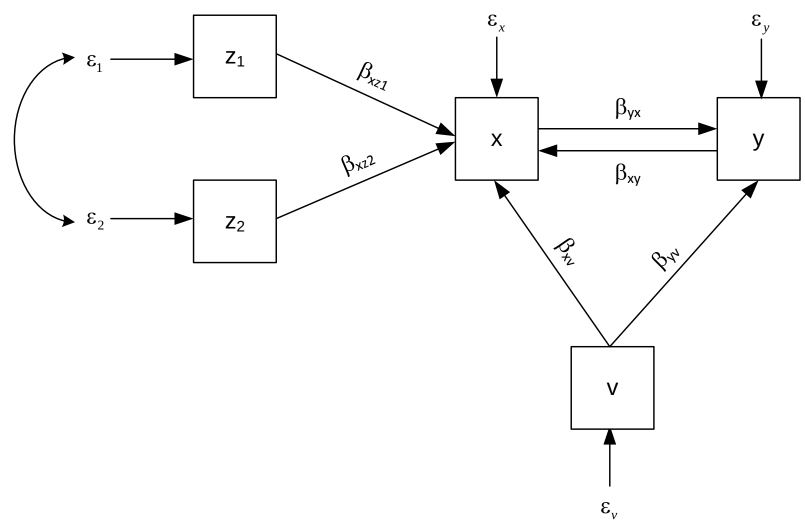
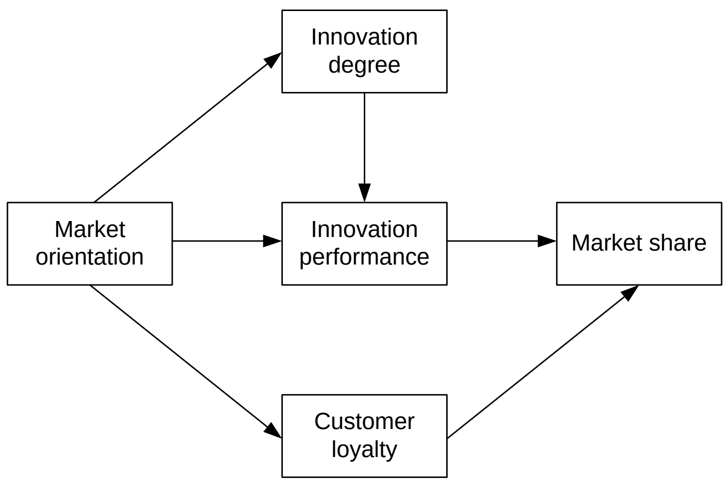


Figure 4

A theoretical model of the relationship between market share and innovation performance

a) Original model



b) Model if customer loyalty is not available for analysis

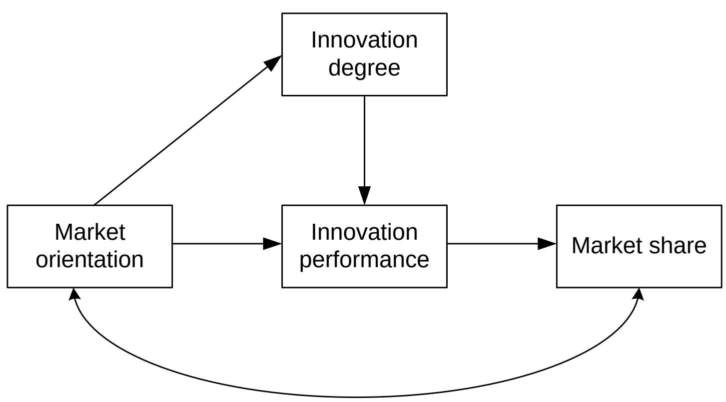


Figure 5

Residual scatterplots to assess linearity and homoscedasticity

(a) residual of $\ln(y)$ *vs.* $\widehat{\ln}(y)$ *(b) residual of* y *vs.* $\hat{y}$

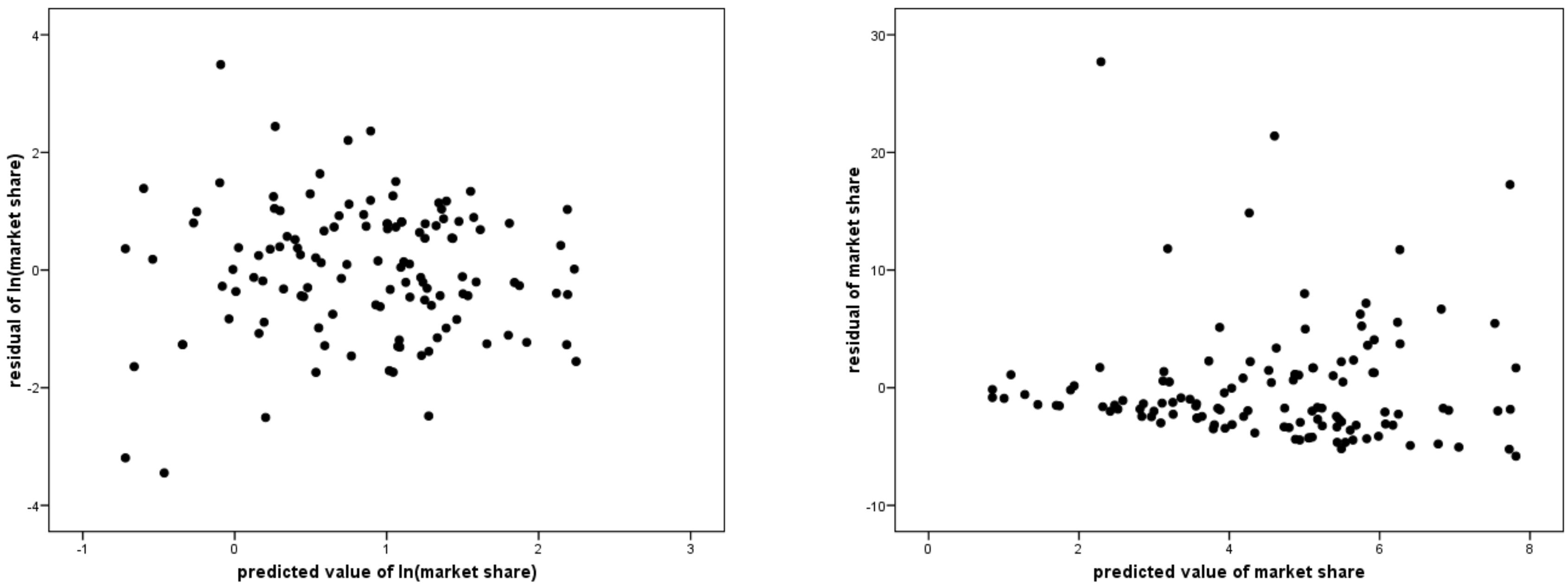
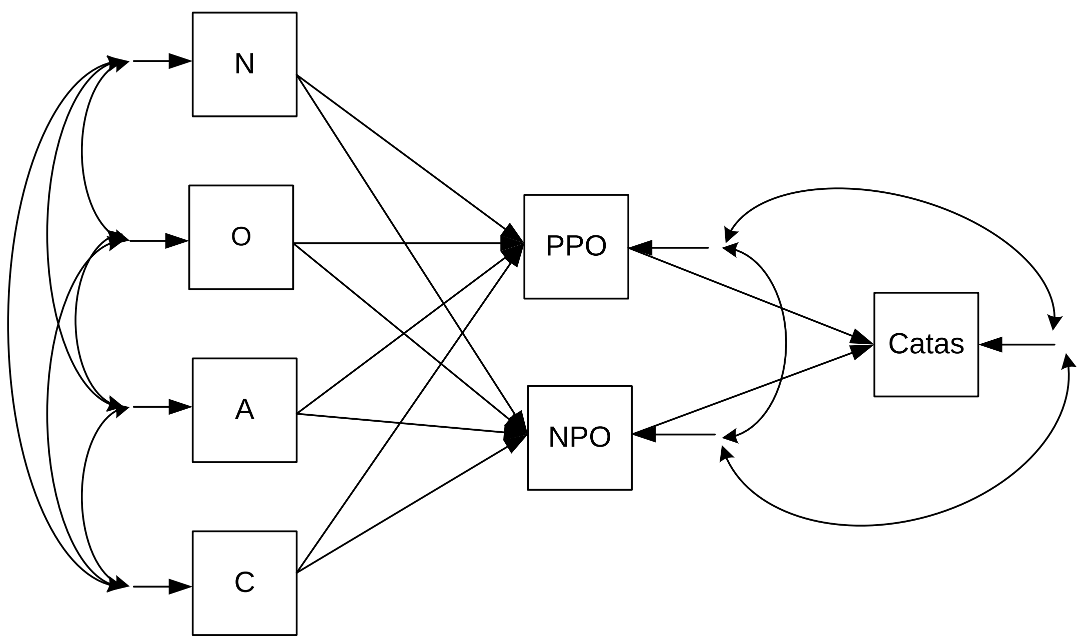


Figure 6

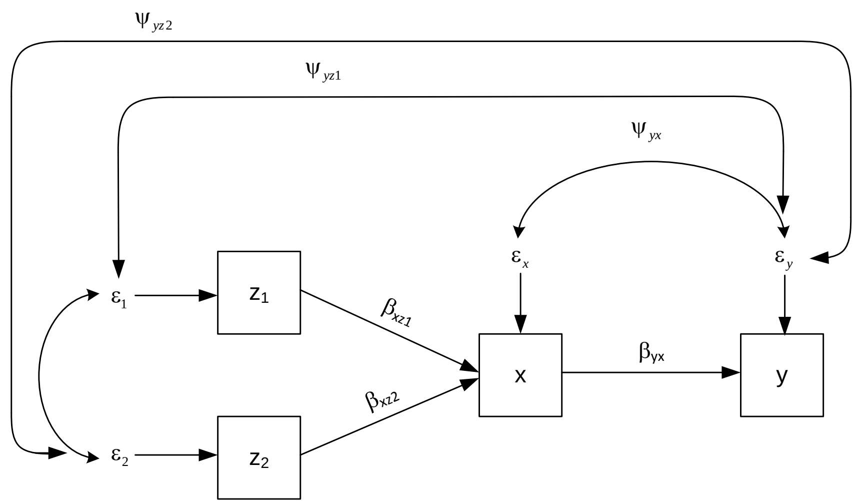
An instrumental variables regression model of causal effects of problem solving orientation on catastrophizing



Note: Catas = catastrophizing, PPO = positive problem orientation, NPO = negative problem orientation, N = Neuroticism, O = Openness to experience, A = Agreeableness, C = Conscientiousness

Figure 7

The instrumental variables regression (IVR) model with endogenous instruments



Note: This model is not identified as *df* = –1. In general, with p_x = number of predictors **x**, p_z = number of instrumental variables **z**, at most $p_z - p_x$ correlations between **z** and *y* disturbances can be estimated.