CAUSATION AND COMPONENTS IN MARKET SHARE-PERFORMANCE MODELS: THE ROLE OF IDENTITIES

by

Kusum L. Ailawadi*

and

Paul W. Farris**

Revised
March 14, 2003

* Amos Tuck School of Business Administration
  Dartmouth College
  Tuck Hall
  Hanover, NH 03755

** Darden Graduate School of Business Administration
  University of Virginia
  P.O. Box 6550
  Charlottesville, VA 22906-6550

Acknowledgements:
We thank William Comanor for providing the data used in this paper, and Richard Bagozzi, Jeffrey Edwards, Hubert Gatignon, Caroline Henderson, Scott Neslin, Len Parsons, Phillip Pfeifer, and participants of the PIMS In Retrospect and Prospect conference at the University of Virginia for helpful comments on this work.
CAUSATION AND THE COMPONENTS OF MARKET SHARE-PERFORMANCE MODELS: THE ROLE OF IDENTITIES

Abstract

The marketing literature contains several structural models, many of them based on the PIMS database, in which one variable is a definitional component of another, related to it through an identity. These definitional relationships have the potential for providing important insights into marketing phenomena, if they are appropriately modeled. On the other hand, they result in inconsistent parameter estimates if they are not separated from other non-definitional relationships in the model that need to be empirically estimated. This paper first discusses the substantive information that can be obtained by studying the definitional components of a composite variable instead of the variable alone. Then, it examines each of the ways in which definitional relationships appear in marketing models, identifies those that are misspecified, analyzes the impact of the misspecification, and provides the correct specification. It also disproves a commonly held belief that using instrumental variable estimation in a simultaneous equation system resolves the problems caused by mixing definitional relationships with structural ones. Thus, it provides a comprehensive view of both the potential benefits and pitfalls of definitional relationships in structural models. Much of the work reviewed in this paper was inspired and enabled by the PIMS research database, which provides data not only on profitability, but also each of its cost and revenue components for a variety of strategic business units over multiple years.
1. INTRODUCTION

The PIMS research program, through the unique data it gathered, has made possible the development of a large body of substantive knowledge on the determinants of financial performance. In the three decades since it was initiated, the program has also inspired and enabled the use of rigorous econometric methods to study models of performance, particularly models of the association between ROI and market share, and refine some of the early findings on this subject. Of particular value in these advances is the availability of longitudinal data from a large cross-section of strategic business units, on not only profitability but all its revenue and cost components. The purpose of this paper is to highlight some of the methodological lessons from this research, particularly as they relate to the appropriate handling of definitional relationships in structural models.

Several of the constructs that are of interest to marketing researchers can be decomposed into two or more definitional components, and are called composite variables. The definitional components of a composite variable are an integral part of its conceptual identity. Their effect on the composite is known without error and does not require estimation with observational data. Thus, they are different from non-definitional determinants whose identity is conceptually distinct from the composite and whose relationship with it does require empirical estimation. A ubiquitous example in marketing and strategy research is the composite variable profit, and its definition components, revenue and various costs. The distinction between definitional components and other determinants has several important implications for the specification of structural models containing either the composite variable alone or the composite variable and one or more of its components. The relationship between a composite variable and its components presents the
potential for improved understanding of the phenomenon being studied, if it is appropriately specified, and the danger of econometric bias and misinterpretation if it is not.

This paper evaluates each of the ways in which definitional relationships are and should be incorporated in marketing models, thus presenting a comprehensive analysis of the application and misapplication of composite-component relationships. Our objective is threefold. One, we discuss the usefulness of the empirical information that can be obtained if the definitional relationship between a composite variable and its components is appropriately modeled. Two, we identify and resolve the problems that arise when it is not modeled appropriately, thus reducing the likelihood of such model misspecification in future research. Three, we aid researchers in correctly interpreting previous empirical research that suffers from misspecification, but offers other valuable insights that should not be dismissed.

The paper is organized as follows. Section 2 identifies the various ways in which definitional relationships appear in the literature, providing examples of each type of model. It reviews research that has noted the problems associated with these models or attempted to address them, and provides closure on some debated issues. Section 3 examines an approach that is perceived as a solution to the problems associated with mixing definitional and structural relationships -- the use of instrumental variable based estimation procedures in simultaneous systems where a composite variable is specified to be a function of one of its components and vice versa. We first interpret the estimated coefficients in such systems, showing that simultaneous system estimation is not a solution to the problem, and then provide an appropriate specification for the system. Section 4 illustrates this process using Comanor and Wilson’s (1974) widely cited model of the Advertising-Profit relationship. Finally, Section 5 concludes the paper with a
discussion of the role of PIMS and composite-component relationships in furthering the understanding of marketing phenomena, their limitations, and some implications for future research.

2. A REVIEW OF DEFINITIONAL RELATIONSHIPS IN PERFORMANCE MODELS

Duncan (1966) was perhaps the first researcher to note that the relationship of a composite variable with its definitional components is fundamentally different from its relationship with other determinants that may affect it but are not a definitional part of it. He pointed out the benefits of examining the components of a composite variable:

"where such decomposition is available, it is of interest (1) to compute the relative contributions of the components to variation in the composite variable and (2) to ascertain how causes affecting the composite variable are transmitted via the respective components".

At the same time he also cautioned against treating components like other causes:

"By this route one can arrive at the meaningless result that net migration is a more important "cause" of population growth than is change in manufacturing output".

Philosophers of science draw a similar distinction between analytical and synthetic statements, as is clear from the following excerpt from Bagozzi (1980):

"An analytic statement is one in which its predicate is "contained in" or is "part of" the concept of its subject.....[It] can only be logically true or false, not factually so. A synthetic statement, in contrast, is one in which the meaning of the predicate is not contained in the concept of the subject.....It is possible to subject the [synthetic] statement to empirical validation."

Bagozzi also notes that causality is synthetic in that the concept of a cause is independent of the concept of an event - any particular cause and effect contain distinct factual information. By this yardstick, therefore, the relationship of a composite variable with its components is analytic and not causal in the same way as is its empirical relationship with a non-definitional determinant.
Despite this early distinction between components and non-definitional determinants of a composite variable, we, as researchers, continue to mix the two and treat them on the same footing in our models of marketing phenomena. Table 1 summarizes the various ways, both appropriate and inappropriate, in which composite variables and/or their definitional components appear in the literature, along with examples of each type of model. We briefly discuss them in this section, reviewing studies that have either used the correct specification or identified problems with the incorrect ones, and identifying issues that still remain unresolved.

<Insert Table 1 About Here>

2.1 The Composite-Component Identity

There is no uncertainty associated with the coefficients of a composite-component relationship in a sample of empirical observations. Denoting the composite by "Z" and its components by \( z_i \), we can write the identity for a two-component additive composite as:

\[
Z = a_1 z_1 + a_2 z_2
\]  

(2.1)

and the expression for its variance as:

\[
\sigma_Z^2 = (a_1^2 \sigma_{z_1}^2) + (a_2^2 \sigma_{z_2}^2) + (2 a_1 a_2 \sigma_{z_1 z_2})
\]  

(2.2)

There is no error in either of these two equations no unknown parameters that need to be empirically estimated. However, the degree to which variation in the composite variable (Z) is due to variation in a given component (\( z_1 \) or \( z_2 \)) is an empirical question. For instance, variance in ROI for a firm over time might be due to variance in either S/I or ROS, or some combination, each possibility having its own set of implications for the management skills at work in the firm. Similarly, variation in sales over time may be due mainly to variation in the number of customers, their average purchase amount or their average purchase frequency. Again, each of these
possibilities implies a very different mechanism by which sales change. The extent to which two components covary is also insightful. Indeed, a composite variable, $Z$, might have very little variance in a given sample, even with substantial variation in both $z_1$ and $z_2$, if the covariance between the two components is negative. Which components contribute the most to variation in the composite variable, and which ones covary positively or negatively with one another? These are empirical questions, answers to which can provide important insights into the composite variable being studied. Thus, even though the definitional identity is known a priori, there is useful empirical information to be gleaned from the sources of variance in the composite. Such analyses are underutilized in the marketing literature, however, one exception that we know of being in the work of Farris, Parry and Webster (1989). These researchers use PIMS data to examine the extent to which each cost component of profitability contributes to its variance across businesses. Their primary findings are that most of the cross-sectional variance in ROI is due to variance in ROS, not the S/I ratio. Further, a substantial portion of the variance in ROS is attributable to the Purchases/Sales ratio, with other cost components contributing much less.

2.2 Models For a Composite Dependent Variable

Recognition of the components is also useful in a structural model that specifies a composite variable as function of only non-definitional explanatory variables:

$$Z = \gamma_0 + \gamma_1 X_1 + \varepsilon$$

(2.3)

This is a purely empirical model with no definitional relationships between variables on the two sides of the equation. Regressions of profit on market share and product quality, or sales on advertising are some examples of models with composite dependent variables. Although the components are not explicitly specified, the effect of the explanatory variables on the composite
dependent variable must occur through its logically prior components. Therefore, decomposing the composite dependent variable into its various components and determining the impact of the explanatory variables on each component can provide useful information to the researcher. This strategy is used quite frequently in the sales promotion literature, where the effect of promotion on sales or share may be modeled through number of purchasers, purchase frequency and purchase amount; brand switching, purchase acceleration and primary demand; or penetration, share of requirements, and category usage (e.g., Neslin and Shoemaker 1983; Gupta 1988; Bell Chiang and Padmanabhan 1999; Ailawadi, Lehmann, and Neslin 2001). Doing so provides useful information about the mechanism by which sale promotion operates, and also helps to evaluate its profitability. Boulding and Staelin (1993) and Ailawadi, Farris, and Parry (1999) have also used such component level analyses with PIMS data to better understand the effect of market share and unobserved management skill, respectively, on profitability.

2.3 Models With a Composite Independent Variable

A composite variable often appears as an independent variable in marketing models:

\[ Y = \gamma_0 + \gamma_1 X_1 + \gamma_2 Z + \epsilon \]  

(2.4)

Difference scores and measures of congruence, widely used as explanatory variables in behavioral models of person-job fit, stress, satisfaction etc. provide some different, non-financial examples of such specifications. In such models, recognition of the definitional components is not just desirable but necessary. To see why, consider the following form of equation (2.4):

\[ Y = \gamma_0 + \gamma_1 X_1 + \gamma_2 (a_1 z_1 + a_2 z_2) + \epsilon \]  

(2.5)

It shows that using the composite variable Z as an independent variable is mathematically equivalent to using both components, \( z_1 \) and \( z_2 \), as independent variables, but constraining their
coefficients, such that the ratio of the coefficients is equal to $a_1/a_2$. If the constraint is incorrect, it will result in biased estimates of all the coefficients in the model (Johnston 1984). Such constraints can and should be empirically tested by estimating the following unconstrained model:

$$Y = \gamma_0 + \gamma_1 X_1 + \gamma_2 Z_1 + \gamma_3 Z_2 + \epsilon$$

(2.6)

Edwards (1994) and Peter, Churchill and Brown (1993) provide a good review of these and other methodological issues in the use of congruence measures and difference scores.

2.4 Mixed Models of a Composite Dependent Variable

So far, we have discussed the definitional identity itself and the first two types of models listed in Table 1, where the components themselves are not explicitly included in the model. In such cases, it is always useful, though not always econometrically necessary, to decompose a composite variable into components. The remaining model specifications in Table 1 are more problematic – they include a composite variable and one of its components in a single model, one of them being an independent variable and the other the dependent variable. Thus, they mix \textit{a priori} definitional relationships with empirical ones that need to be estimated with observational data, resulting in serious econometric problems. There are a host of mixed models in the literature. They have been used, quite often, in PIMS based studies of the profit-market share relationship (e.g. Prescott, Kohli and Venkatraman 1986, Buzzell and Gale 1987, Jacobson and Aaker 1985) and in industrial organization studies of the Structure-Conduct-Performance paradigm (e.g. Comanor and Wilson 1967, Porter 1976, Ravenscraft 1983). For instance, Comanor and Wilson specify a model where profitability is regressed on one of its components (advertising/sales) while Prescott et al. regress ROI on ratios of investment, manufacturing, marketing, and R&D to sales.
These studies illustrate the following mixed specification:  

\[ Z = \gamma_0^* + \gamma_1^* X_1 + \gamma_2^* z_1 + \varepsilon^* \]  

(2.7)

Such mixing of known definitional relationships with structural relationships that must be empirically estimated has been a subject of concern amongst researchers for quite some time. This concern stems from two issues. The first questions the explanatory power of a model when a variable appears on both sides of the equation. For instance, Gale and Buzzell (1990), in their review of studies of profitability, note that the inclusion of several accounting variables in a regression model of ROS or ROI results in a "virtual algebraic identity which does not explain anything", despite its high \( R^2 \). Zenor and Leone (1991) state that, in a series of empirical analyses, they obtain a much higher \( R^2 \) when dollar revenue is predicted by price and other variables rather than when unit sales is the dependent variable. Similarly, Hanssens and Parsons (1992) question the explanatory power of a demand model that uses price to explain sales revenue, since price appears on both sides of the equation. Ailawadi and Farris (1993) quantify the inflation in \( R^2 \) that occurs when an additive component of a composite variable is included as an explanatory variable in a regression model of the composite. The inflation occurs due to the artificial "explanation" of the composite by the error term of the included component.

The second issue deals with the correlation between a composite and its component. For instance, Mahajan, Varadarajan and Kerin (1987) note that the negative association of ROI with the investment/sales ratio may be partly or wholly an "artifact" of their definitional relationship. Similarly, the substantial literature on the analysis of ratio variables discusses the spurious correlation between two ratio variables with a common component (see, for example, Schuessler

\[ \text{1 Throughout this paper, we use an asterisk to depict parameters in mixed models.} \]
However, the regression of a composite variable on its own component also affects the estimated coefficients of all the other explanatory variables in the model. Farris, Parry, and Ailawadi (1992) quantify this econometric bias, showing that the estimated coefficients of the non-definitional antecedents in the model reflect their effect, not on the dependent variable but only on those of its definitional components that are excluded from the right hand side of the model. They illustrate this bias with two PIMS based models of the market share – ROI relationship, Jacobson and Aaker (1985, hereafter denoted JA) and Buzzell and Gale (1987). They replicate these mixed models by decomposing the dependent variable, ROI, into its definitional components, and estimating the effect of the variables on the right hand side on each of the definitional components. By doing so, they illustrate the econometric result that the estimated coefficients of the variables in the mixed model reflect their effect not on ROI but only on its excluded components.

Jacobson and Aaker (1993) subsequently published a comment on this research. It is worth addressing the issue raised in that comment and clarifying its relevance to the primary message of the Farris, Parry, and Ailawadi (1992) work. We summarize the issue in the following statements of fact while making one simplification for expositional purposes – we use ROS instead of ROI as the dependent measure of profitability.  

1. JA regress profitability on its own lagged values, current values of two of its definitional components, Value Added/Sales and Marketing/Sales, and other explanatory variables. As they recognize in their 1993 comment, including the two definitional components on the right

---

2 The definitional relationship of ROS with its cost components is additive, which makes the exposition simpler than the multiplicative relationship with ROI. The analysis can be extended to ROI, as shown in Farris, Parry, and Ailawadi (1992).
hand side of their model biases all the estimated coefficients.

2. The estimated coefficients of their model with ROS as the dependent variable are exactly equal to the algebraic combination of the estimated coefficients when each of the additive components of ROS, (Value Added/Sales, Marketing/Sales, and Other Costs/Sales) is regressed on the same set of right hand side variables. To see this, consider the following:

\[ \text{ROS} \equiv \frac{VA}{S} - \frac{Mktg}{S} - \frac{Other\ Costs}{S} \]

\[ \text{ROS} = X\hat{\beta}_{\text{ROS}} + \varepsilon \]

\[ \hat{\beta}_{\text{ROS}} = (X'X)^{-1}X'ROS \]

\[ \equiv (X'X)^{-1}X'(VA/S - Mktg/S - Other\ Costs/S) \]  \hspace{1cm} (2.8)

\[ \equiv (X'X)^{-1}X'(VA/S) - (X'X)^{-1}X'(Mktg/S) - (X'X)^{-1}X'(Other\ Costs/S) \]

\[ \equiv \hat{\beta}_{VA/S} - \hat{\beta}_{Mktg/S} - \hat{\beta}_{Other\ Costs/S} \]

3. Thus, in order to replicate JA’s model at the component level, one must regress each component on the exact same explanatory variables as they had in their composite model, i.e., lagged profitability and the other variables in their model.

4. Jacobson and Aaker (1993) argue that the JA model was not correctly replicated and that each component should have been regressed on the lagged value of that component to control for unobserved variables. They further note that a structural model that would give rise to regressing each component on their set of explanatory variables is “not readily apparent”.

5. We make three points in this regard. First, the component level models used by Farris, Parry, and Ailawadi (1992) exactly and correctly replicate JA’s composite level model, as the equations above show quite simply. Using lagged components instead of lagged profitability on the right hand side would not replicate the composite model. Second, if the structural model underlying the component level replication of JA’s composite model is “not readily apparent”,
then one must question the composite model because algebraically combining the estimates from the component models will provide exactly the estimates of JA’s composite model.

Third, Farris, Parry, and Ailawadi’s (1992) purpose was not to find the best way to control for unobserved variables but to quantify and illustrate a fundamental econometric problem, i.e., mixing definitional components with other explanatory variables. Of course, we fully agree with them that unobserved variables should be controlled for. We might add though, that the question of how best to control for them, be it in composite or component level models, remains an open one. It is not clear that first differencing to control for fixed effects and/or rho-differencing to control for autocorrelated unobserved variables is always ideal, as is evidenced by the work of Gatignon and Christen in this book.

2.5 Mixed Simultaneous Systems of a Composite and its Component

Several researchers have noted that including a component as an independent variable in the model for a composite variable violates the assumption of independence between the error term and the independent variables and biases the OLS estimates (e.g., Comanor and Wilson 1974; Boulding 1990; Farris, Parry, and Ailawadi 1992). Therefore, based on standard econometric treatments of simultaneous systems (e.g., Intriligator 1978; Johnston 1984), an instrumental variable based estimation procedure is expected to solve the problem. In the next section, we show that this is, unfortunately, not true. Such an estimation procedure does not solve the problems associated with the mixing of definitional and structural relationships -- the solution to the problem lies in appropriate model specification, not estimation. In doing so, we reveal some important issues relating to the identification of simultaneous systems that are unique to mixed composite-component systems, and we also examine another type of mixed model, i.e., the regression of a
component on a composite variable.

3. COMPOSITE-COMPONENT RELATIONSHIPS IN SIMULTANEOUS SYSTEMS

One often encounters simultaneous equation systems in the marketing literature where a composite variable is specified as a function of one of its components and vice versa. In fact, as noted earlier, it is believed that (i) the definitional relationship between a composite variable and its component causes the component to be endogenous in the system that determines the composite, and (ii) the use of an instrumental variable based estimation procedure in the simultaneous framework solves the problem. One important example lies in the advertising-profit relationship studied under the well-known Structure-Conduct-Performance (SCP) paradigm. Since advertising is one of the cost components that is deducted from gross profit to compute net profit, models using a net rate of return mix definitional relationships with structural parameters to be estimated from observational data. Comanor and Wilson (1974) were probably the first to suggest that the advertising-profit relationship should be estimated in the simultaneous equation framework, at least partly to address the definitional problem:

"A simultaneous-equation model is appropriate not only because of the possible presence of reverse causality in the profit rate equation but also because the error terms in the two structural equations are likely to be negatively correlated. Profits and advertising are linked by the identity that profits net of all deductions are derived from gross cash flow by deducting depreciation, income taxes and advertising."

Several other researchers use similar models in their studies of the SCP paradigm (e.g., Porter 1976, Intriligator 1978, Ravenscraft 1983, Chang and Choi 1988, Kumar 1990), showing that Comanor and Wilson’s model was not an isolated case and making it especially important to determine whether such procedures are appropriate. Our analysis shows that simultaneous equation specification and estimation procedures do not solve the problem of separating the
definitional relationship between the two endogenous variables from structural effects. We use the following mixed simultaneous system, to illustrate our analysis:

\[ z_1 = \gamma_{10}^* + \gamma_{11}^* X_1 + \gamma_{12}^* X_2 + \beta_{12}^* Z + \varepsilon_1^* \]  \hspace{1cm} (3.1)

\[ Z = \gamma_{20}^* + \gamma_{21}^* X_1 + \gamma_{23}^* X_3 + \beta_{21}^* z_1 + \varepsilon_2^* \]  \hspace{1cm} (3.2)

where \( Z \equiv z_1 + z_2 \)  \hspace{1cm} (3.3)

Thus, the two endogenous variables in the system are \( Z \) and \( z_1 \), and, as is conventionally done, the system is identified by including at least one explanatory variable in each equation that does not appear in the other equation. We now evaluate the coefficients that are obtained when the above system is estimated using an instrumental variable technique, such as 2SLS. We begin with equation (3.2) for the composite, \( Z \), where an instrument is used for \( z_1 \).

### 3.1 Interpreting Coefficients of the Composite Equation

In the first step of the 2SLS procedure, \( z_1 \) is regressed on all the exogenous variables in the system:

\[ z_1 = \alpha_{10} + \alpha_{11} X_1 + \alpha_{12} X_2 + \alpha_{13} X_3 + \mu_1 \]  \hspace{1cm} (3.4)

Then, the predicted value of \( z_1 \) from the above equation is used as an instrument for \( z_1 \) in the \( Z \) equation:

\[ Z = \gamma_{20}^* + \gamma_{21}^* X_1 + \gamma_{23}^* X_3 + \beta_{21}^* \hat{z}_1 + \varepsilon_2^* \]  \hspace{1cm} (3.5)

To interpret the second-stage regression estimates in (3.5), we substitute for \( Z \) using:

\[ Z \equiv z_1 + z_2 \equiv \hat{z}_1 + \hat{\mu}_1 + z_2 \]  \hspace{1cm} (3.6)

Equation (3.5) can then be written as:
\[ \hat{x}_1 + \hat{\mu}_1 + z_2 = \gamma_{20}^* + \gamma_{21}^* X_1 + \gamma_{23}^* X_3 + \beta_{21}^* \hat{x}_1 + \epsilon_2^* \]  

Equation (3.7) combines three regressions. Its coefficient estimates are given by the algebraic sum of the corresponding coefficient estimates when \( \hat{x}_1, z_2 \) and \( \hat{\mu}_1 \) are regressed on all the regressors in the equation:

\[ \hat{x}_1 = \gamma_{40}^* + \gamma_{41}^* X_1 + \gamma_{43}^* X_3 + \beta_{41}^* \hat{x}_1 + \epsilon_{41}^* \]  

(3.8)

\[ \hat{\mu}_1 = \gamma_{\mu_0}^* + \gamma_{\mu_1}^* X_1 + \gamma_{\mu_3}^* X_3 + \beta_{\mu_1}^* \hat{\mu}_1 + \epsilon_{\mu_1}^* \]  

(3.9)

\[ \hat{z}_2 = \gamma_{2z_0}^* + \gamma_{2z_1}^* X_1 + \gamma_{2z_2}^* X_3 + \beta_{2z_1}^* \hat{z}_1 + \epsilon_{2z_2}^* \]  

(3.10)

and

\[ \gamma_{2i}^* = \gamma_{i0}^* + \gamma_{i1}^* + \gamma_{i2}^* \quad \text{for all } i \]

and

\[ \beta_{2i}^* = \beta_{i0}^* + \beta_{i1}^* + \beta_{i2}^* \]  

(3.11)

Note that a variable is a perfect predictor of itself. Therefore, when \( \hat{x}_1 \) is regressed on a set of variables including itself, the coefficient estimate of \( \hat{x}_1 \) is exactly one, with a standard error of zero, the coefficient estimates of all the other variables in the equation are zero, and the \( R^2 \) is exactly 1.0. Therefore, the estimates of equation (3.8) are:

\[ \gamma_{\hat{z}_0}^* = \gamma_{\hat{z}_1}^* = \gamma_{\hat{z}_3}^* = 0 \]

and

\[ \beta_{\hat{z}_1}^* = 1 \]

(3.12)

Also note that the residual \( \hat{\mu}_1 \) is, by definition, uncorrelated with all of the exogenous variables in the system and with \( \hat{z}_1 \). Therefore, \( R^2 \) is zero for equation (3.10) and the coefficient estimates are:
\begin{equation}
\hat{\gamma}_{\hat{\mu},0} = \hat{\gamma}_{\hat{\mu},1} = \hat{\gamma}_{\hat{\mu},3} = \hat{\beta}_{\hat{\mu},1} = 0 \tag{3.13}
\end{equation}

Equations (3.11) through (3.13) show that the estimates of $\gamma_{20}^*$ through $\gamma_{23}^*$ represent the effects of the corresponding variables on $z_2$ alone. The estimate of $\beta_{21}^*$ equals the structural effect of $z_1$ on $z_2$, plus the known definitional parameter 1. Thus, we find that, despite the instrumental variable estimation, the problem with coefficient estimates of the mixed composite model still remains. We now turn our attention to the equation for the component.

3.2 Interpreting Coefficients of the Component Equation

In equation (3.1), a logically prior component ($z_1$) is regressed on its composite ($Z$). The same two-stage procedure is followed for the estimation of the $z_1$ equation. $Z$ is regressed on all the exogenous variables in the system and its predicted value is used as an instrument for $Z$ in the $z_1$ equation. To analyze the coefficients obtained by this process, consider the following identity:

\[ Z \equiv z_1 + z_2 \equiv \hat{Z} + \hat{\mu} \equiv (\hat{z}_1 + \hat{z}_2) + (\hat{\mu}_1 + \hat{\mu}_2) \]

where $\hat{\mu}_1$ and $\hat{\mu}_2$ are residuals from the regression of $z_1$ and $z_2$ respectively, on all the exogenous variables in the system. Substituting in (3.1), i.e., the $z_1$ equation, we have:

\[ (\hat{z}_1 + \hat{\mu}_i) = \gamma_{10}^* + \gamma_{11}^* X_1 + \gamma_{12}^* X_2 + \beta_{12}^* (\hat{z}_1 + \hat{z}_2) + \epsilon_i^* \tag{3.15} \]

Estimates of the $\gamma_{1}^*$ vector and $\beta_{1}^*$ in (3.15) are treated as consistent estimates of the coefficients in the $z_1$ equation. But, this equation is, in effect, a constrained regression where the coefficients of $\hat{z}_1$ and $\hat{z}_2$ are forced to be equal. Further, the coefficient estimates of the variables in equation (3.15) are simply the sum of the corresponding coefficient estimates when they are used as regressors for $\hat{z}_1$ and $\hat{\mu}_i$. If the constraint on the coefficients were removed, as in:

\[ (\hat{z}_1 + \hat{\mu}_i) = \gamma_{10}^* + \gamma_{11}^* X_1 + \gamma_{12}^* X_2 + \beta_{12}^* \hat{z}_1 + \beta_{12}^* \hat{z}_2 + \epsilon_i^* \tag{3.16} \]
we would have:

- A coefficient of 1 for $\hat{z}_1$, since all the other variables would drop out of the $\hat{z}_1$ equation and, by definition, the residual and predicted values are uncorrelated.
- Coefficient estimates with expected values of zero, for all the exogenous variables in the system, since they drop out of the $\hat{z}_1$ regression and are uncorrelated with the residual.
- A non-zero coefficient estimate for $\hat{z}_2$, which reflects its association with the residual, i.e. its effect on $z_1$ once all the other variables have been accounted for.

Thus, the constraint imposed in the $z_1$ equation only serves to disguise these \textit{a priori} relationships. Since the only part of Z that remains to be defined once $z_1$ has been defined, is the other component, $z_2$, it is this component that should be used in the $z_1$ equation. Such a specification is intuitively appealing and, at the same time, it separates the definitional relationship of $z_1$ with Z from any structural relationship between the two components, $z_1$ and $z_2$:

$$
\begin{align*}
  z_1 &= \gamma_{10} + \gamma_{11}X_1 + \gamma_{12}X_2 + \beta_{12}z_2 + \varepsilon_1 \\
  & \quad \text{(3.17)}
\end{align*}
$$

Once the equations for both components have been correctly specified, an appropriate estimation procedure can be used to estimate their structural parameters. An instrumental variable based procedure like 2SLS is needed because the system is non-recursive, not because of any definitional endogeneity. When $z_1$ is included in the equation for Z, the only error remaining in the Z equation is that due to $z_2$, i.e., $\varepsilon_2$. This error term in the mixed Z equation is not \textit{definitionally} correlated with $z_1$. The OLS assumption of $\text{E}(X'\varepsilon) = 0$, which researchers like Comanor and Wilson (1974), Boulding (1990), and Farris, Parry and Ailawadi (1992) are concerned about, is defied if the two component errors, $\varepsilon_1$ and $\varepsilon_2$, are correlated, and not “by definition”.
3.3 Implications of Identifying Restrictions in the Equation For Z

From the above discussion, it would appear that 2SLS estimates of the mixed Z equation provide consistent estimates of the effect of non-definitional explanatory variables on $z_2$. Even this may not be true, however. It depends on whether the following equation for $z_2$ implied by (3.12) is correctly specified:

$$z_2 = \gamma_{20} + \gamma_{21} X_1 + \gamma_{23} X_3 + \beta_{21} \hat{z}_1 + \epsilon_2 \quad (3.18)$$

To see if the specification in (3.18) is correct, we determine the implications of the mixed system for the $z_2$ equation, focusing specifically on the exclusion restrictions incorporated in the Z equation.

Recall that the effect of an antecedent on Z is the sum of its effects on $z_1$ and $z_2$. Therefore, the effect of a variable on Z can be zero (as is implied by an exclusion restriction) only if (i) it does not affect either component or (ii) the sum of its effects on $z_1$ and $z_2$ is zero. Keeping this fact in mind, consider the role of $X_2$ in the mixed simultaneous system. Since it is excluded from the researcher's equation for Z, (s)he must believe that $X_2$ has no effect on Z. Given that $X_2$ does affect $z_1$, its effect on $z_1$ must be exactly offset by an effect on $z_2$, if the exclusion restriction is valid. The appropriate specification for $z_2$ should then be:

$$z_2 = \gamma_{20} + \gamma_{21} X_1 + \gamma_{22} X_2 + \gamma_{23} X_3 + \beta_{21} \hat{z}_1 + \epsilon_2 \quad \text{where} \quad \gamma_{22} = -\gamma_{12} \quad (3.19)$$

This discussion reveals an important consideration. If the exclusion restriction is appropriate, a relevant variable, $X_2$, has been excluded from the $z_2$ equation in (3.17), rendering it misspecified. But, how reasonable is it to expect the exclusion restriction to be valid? The
researcher must have very strong theoretical priors about the specific variable, $X_3$, to assume that its effect on $z_2$ is exactly equal in magnitude and opposite in sign, to its effect on $z_1$. Although we have illustrated this problem with only exclusion restrictions for a single explanatory variable, it applies to other types of identifying restrictions as well (e.g., equality restrictions). Recognition of such implications of their identifying restrictions not only encourages researchers to ensure that their model specification accurately reflects the underlying theory, but, we hope, it may cause them to rethink their theory in a more complete and logically consistent manner. The work by Moore and Morgan in this manuscript provides important methodological guidelines for testing the validity of identifying restrictions in simultaneous systems in general, and would be particularly helpful in the context of such mixed models.

4. SIMULTANEOUS MODEL OF ADVERTISING-PROFIT RELATIONSHIP

In this section, we use Comanor and Wilson's (1974) simultaneous model of the advertising-profit relationship as an illustration to summarize the key issues examined above. Their model is as follows:\(^3\)

\[
\frac{A}{S} = \gamma_{10}^* + \gamma_{11}^* LGR + \gamma_{12}^* DUR + \gamma_{13}^* LCONC + \gamma_{14}^* TECH + \beta_{12}^* ROS + \varepsilon_1^* \tag{4.1}
\]

\[
ROS = \gamma_{20}^* + \gamma_{21}^* LGR + \gamma_{25}^* LCAP + \gamma_{26}^* LOCAL + \beta_{21}^* \frac{A}{S} + \varepsilon_2^* \tag{4.2}
\]

where LGR is the natural log of demand growth rate, DUR is a dummy variable for durable goods industries, LCONC is the natural log of industry concentration ratio, TECH is a dummy variable for high technical barriers to entry, LCAP is the natural log of volume of absolute capital

\(^3\) They use Return on Equity instead of ROS but note that different rates of return are highly correlated and lead to similar empirical results. We illustrate their model using ROS to avoid the additional complexity of a multiplicative identity.
requirement, LOCAL is a dummy variable for local industry, ROS is the net return on sales, and A/S is the advertising/sales ratio.

4.1 Problems with the Mixed Model

Our analysis in the previous section leads us to note at least four important issues about this model. First, the coefficient of the included component, A/S, is equal to 1 plus the structural effect of A/S on the gross rate of return before advertising costs have been deducted (denoted as Π/S), and the corresponding t-statistic must be appropriately interpreted. Second, irrespective of the estimation procedure used, the coefficients of non-definitional explanatory variables in the ROS equation are effects only on Π/S, irrespective of whether or not advertising costs are deducted to compute the dependent profit rate measure. This finding renders almost irrelevant the discussion in the industrial organization literature about whether the profit measure should be gross or net of advertising (see, for example, Sawyer 1982, Schumacher 1991). If the researcher is really interested in the effects of non-definitional variables (e.g. concentration) on profit gross of advertising, there is no reason for using net profit as the dependent variable. If, on the other hand, the researcher wants to estimate the effects of these variables on profit net of advertising, then too, net ROS should not be used. Equations for both components, gross ROS and A/S, must be estimated, and total effects on net ROS computed from them.

Third, the specification of the A/S equation is incorrect, and its coefficients are biased and inconsistent. This is true for OLS as well as the 2SLS estimates analyzed above. Since ROS cannot be defined until A/S has been defined, at least contemporaneously, it is counter-intuitive to use it to predict A/S. Just as ROS cannot be and is not used to contemporaneously predict one component, gross ROS, it cannot predict the other component, A/S. The A/S equation should be:
Fourth, the exclusion restrictions that are incorporated in mixed simultaneous equation models of ROS and A/S, in order to identify the equations, have very stringent underlying implications about the magnitude and direction of the effects of the antecedent variables on the components. Three exogenous variables, TECH, DUR and LCONC, are excluded from the ROS equation although they are included in the A/S equation, while two other variables, LCAP and LOCAL, are excluded from the A/S equation. Consider the implications. First, if an exogenous variable in the ROS equation is excluded from the A/S equation (e.g., LCAP), it must have an effect on the other component, Π/S. Furthermore, the effect of that variable on ROS is equal to its effect on Π/S, since Π/S is the only component through which it affects ROS. Although not particularly stringent, it is useful for the researcher to recognize this implication so as to ensure its consistency with theory. Second, if an exogenous variable affects A/S but is excluded from the ROS equation (e.g. LCONC), its effect on Π/S must be exactly equal to its effect on A/S. Thus, if the exclusion restrictions in the model are indeed valid, each of the three exogenous variables, DUR, LCONC, and TECH, must have an exactly equal effect on Π/S as on A/S, leading to the following specification for Π/S:

\[
\frac{\Pi}{S} = \gamma_{20} + \gamma_{21}LGR + \gamma_{22}DUR + \gamma_{23}LCONC + \gamma_{24}TECH
\]

\[
+ \gamma_{25}LCAP + \gamma_{26}LOCAL + \beta_{21} \frac{A}{S} + \varepsilon_{2}
\]

where

\[
\gamma_{22} = \gamma_{12}; \gamma_{23} = \gamma_{13}; \gamma_{24} = \gamma_{14}
\]

The theoretical justification required to support this specification is certainly not obvious, if at all possible. The researcher specifying these systems of equations should explicitly recognize these
assumptions and evaluate their feasibility. Once suitable identifying constraints have been
determined, the model can be re-specified so as to separate any definitional parameters from
structural effects that must be empirically estimated. In the advertising-profit framework, this
implies that the effect of A/S on ROS basically occurs through $\Pi/S$, and vice versa. Both
components are consequently endogenous, and the first part of the system should model the non-
recursive relationship between them. Therefore, we must write down structural equations for the
two components, each containing only exogenous variables and a component as explanatory
variables. Finally, the composite variable, ROS, must only appear in the second part of the model,
comprising the definitional identity:

4.2 Empirical Illustration

We replicated Comanor and Wilson’s original data set as closely as possible for 39 of the 41
industries analyzed by them. Data on several of the variables in their model are included in their
book, and William Comanor kindly provided us with data on some of the other variables. Table 1
presents the OLS and 2SLS estimates of their A/S and ROS equations (4.1) and (4.2), obtained from
this data. It also provides OLS and 2SLS estimates when ROS is replaced by $\Pi/S$ as a dependent
variable:

$$\Pi / S = \gamma_{30}^* + \gamma_{31}^* \text{LGR} + \gamma_{35}^* \text{LCAP} + \gamma_{36}^* \text{LOCAL} + \beta_{31}^* \frac{A}{S} + \varepsilon_3^*$$  \hspace{1cm} (4.5)

Note that the coefficient estimates of all the exogenous variables in the mixed ROS equation
are exactly equal to the corresponding estimates for the $\Pi/S$ equation. This is true of both OLS and
2SLS estimates. Thus, although 2SLS does remove any inconsistency due to reverse causality in the Advertising-Profit relationship, it does not remove the definitional "inconsistency".

Next, we estimate the A/S and Π/S equations (4.3) and (4.4) implied by their model, and compute corresponding estimates for ROS. Note that 3SLS would provide more efficient estimates than 2SLS, since these restrictions apply to structural coefficients across equations, and the error terms for the two components are probably correlated. To maintain comparability with C&W's 2SLS estimation procedure, however, we make use of the 2SLS procedure in two steps. First, we estimate the coefficients of equation (4.3) for A/S by 2SLS. These estimates are then used to constrain the required coefficients in equation (4.4) for Π/S and 2SLS is used to estimate the constrained version.

Insert Table 2 About Here

Table 2 depicts the impact of this respecification on the coefficient estimates. It also depicts the impact of the exclusion constraints imposed on the original model, by computing effects on ROS from both Π/S specifications, (4.5) and (4.4). These computed coefficients for ROS are labeled ROS_{compute1} and ROS_{compute2} respectively. Note the difference between the A/S coefficients for the ROS_{compute1} and ROS_{compute2} equations -- 1.24 versus 0.66. Thus, the constraints implied by Comanor and Wilson’s exclusion restrictions substantially lower the A/S coefficient. The coefficients of all the other variables in the equation are also affected by these constraints although the differences are not statistically significant.\(^4\) The same is true for estimates of the re-specified A/S equation.

5. CONCLUSION

\(^4\) Since the C&W model does not fit the data well, standard errors of the estimated coefficients are high.
5.1 Summary of Findings

As we noted in the beginning of this paper, the definitional identity relating a composite variable to its components appears quite often in the literature, and in various model specifications. Table 4 summarizes our discussion of each, pointing out which models are misspecified and what the correct specification and analysis should be. We have seen that, as long as it is appropriately incorporated in the model, recognition of definitional relationships between marketing constructs can be very useful, assuming, of course, that data is available on the components. In summary, we have found that:

- Valuable insights into the mechanism by which the composite variable changes can be obtained by determining the extent to which variance in each component and covariance between two components contributes to the total variance of the composite variable.

- Similarly, estimating the effect of non-definitional determinants on each component of the composite variable provides an empirical understanding of the mechanism by which these determinants affect the composite variable that is not available from analyses of the composite variable alone.

- Using a composite as an explanatory variable can lead to biased and inconsistent results if the constraint that is implicitly placed on the coefficients of the components is incorrect. The components should be used as separate explanatory variables so that the constraint can be empirically tested.

- Composite models that include a component as an explanatory variable not only have biased and inconsistent coefficients, their explanatory power and tests of significance are also affected. Including a component always inflates the explanatory power of the model relative to the
corresponding model without the component.

- Models of a component where the composite variable is used as an explanatory variable also produce biased and inconsistent results due to the imposition of an incorrect constraint.
- Finally, contrary to commonly held beliefs, these problems in a mixed composite or component model are not solved by the use of an instrumental variable estimator in the simultaneous equation framework. Further, identifying conditions such as exclusion and equality constraints, that are commonly imposed upon the mixed equation for a composite variable, in order to identify the system, have very stringent theoretical implications for the effect of the excluded variables on the components.

<Insert Table 4 About Here>

The solution to these problems lies in appropriate model specification - a model that contains both a composite variable and one or more of its definitional components must be specified in two distinct parts. The first part models relationships between components and non-definitional antecedents, that must be empirically estimated. The second comprises the known definitional identity that has no impact on the estimation or interpretation of the first part. This two-part specification procedure makes the identity truly independent of the structural model.

The reasoning process by which the researcher makes the choice of composite and component variables is helpful in at least two ways. First, she is forced to think through the underlying theory in a more complete and logically consistent manner. And, second, she avoids both the misapplication of simultaneous equation procedures and misinterpretation of model estimates.

Of course, multiple schemes for defining and organizing these components may exist. For
instance, Return on Investment (ROI) might be specified as the product of Sales/Investment (S/I) and Return on Sales (ROS) or as the difference between S/I and Costs/Investment. Mathematically, neither is superior - both are identities. However, one set of components might be more useful than another for a particular theory or from a managerial viewpoint.

5.2 Identifying Definitional Components

Some definitional relationships such as the one between profit and its components, or between incremental sales and its components (e.g., Neslin and Shoemaker 1983) are quite easy to identify. Others, such as the relationship between total passengers flown during a given period and the number of flights operating in that period in Gatignon's (1984) model of advertising reactivity in the airline industry, may not be as obvious. Even for a given identity, definitional relationships may be more obvious in the use of some components than in others. For instance, the use of gross profit, along with other strategic variables like market share and product quality, to predict net profit would likely be immediately identified as tautological, yet, the definitional relationship of R&D and Value Added (which is nothing but 1 minus purchases) with net profit is not as readily questioned (e.g. Prescott, Kohli and Venkatraman 1986, Phillips, Chang and Buzzell 1983). However, the consequences of mixing remain the same, irrespective of which specific components are included in the mixed model.

Of course all of this is under the assumption that it is possible to distinguish a composite variable from its components in an identity. In the case of definitional identities such as those relating advertising and profit, the distinction between components and the composite is quite unambiguous since net profit cannot be determined until all the costs have been determined. For other identities, the context in which the model is being developed may determine which is the
composite and which are the components. Consider the identity that is implicit in Gatignon’s (1984) model of airline passengers flown:

\[
\text{Total Passengers Flown} \equiv \text{Number of Flights X Average Passengers per Flight}
\]

The model uses number of flights as an explanatory variable. One might argue that passengers per flight and number of flights are logically prior to the total passengers flown and, since the model does not attempt to model variations in number of passengers from flight to flight, number of flights and average passengers per flight should be considered components of total passengers flown. On the other hand, it may also be argued that average passengers per flight is computed only after we have data on total passengers flown and therefore should not be considered a component of the latter.\(^5\) Our view is that the context and the theory used by the researcher helps to distinguish the components from the composite variable. However, the fact remains that using number of flights to explain variations in total passengers flown automatically implies that the coefficients of the non-definitional variables in the model reflect their effect only on average passengers per flight, not total passengers flown. It is important for the researcher to recognize this fact and ensure that his/her theory is consistent with this implicit assumption.

The aim of research is to increase understanding (and to reduce error in our equations). Identities and tautologies may be useful ways to conceptualize some parts of the system that we seek to understand. The development of better and more complete theories can be encouraged by requiring them to be logically consistent with the existence of these unambiguous identities. A good description and understanding of a composite variable might use multiple decomposition schemes on a single dependent variable. Further, it may be that more complete and precise theories

\(^5\) We thank Hubert Gatignon for making this point.
will enable us to express variables that were previously considered empirical causes, as components. It would be fruitful to attempt to extend the domain of "components" to include variables that were previously considered empirical causes.

As models become more complicated and statistical techniques less transparent, the potential for simple truths to get obscured will accelerate. Recognition of the relationship between composite variables and their definitional components may help us ensure that identities are separated from empirical investigations, yet contribute significantly to an overall understanding of the phenomenon that is being studied.

For some types of identities, it may not be possible or even meaningful to make such a differentiation between components and the composite, even in the specific context of the theoretical framework being modeled. Equilibrium identities, which only hold at optimal points (e.g., Supply \(\equiv\) Demand), or synthetic relationships between ratios (e.g., US/UK exchange rate \(\equiv\) US/French rate \(\times\) French/UK rate) are examples of such identities. Our analysis in this paper is clearly not relevant to these non-definitional identities.

Also, our analysis in this paper has been limited to additive identities. Purely multiplicative identities can be analyzed similarly if the models are specified in log-linear form. It is equally incorrect to specify a mixed model for a composite variable that includes multiplicative components from one level of decomposition and additive components from another level as explanatory variables (e.g., models of ROI that contain both investment turnover and marketing/sales ratios) although interpreting the estimates of such a model is more difficult. We hope that our work serves to caution modelers against mixing any kind of definitional and structural relationships.
We conclude with a word of gratitude to Bob Buzzell and the other researchers who pioneered the PIMS program and have provided so much food for thought and material for research to empirical and methodological researchers alike, that issues raised by the early studies of PIMS data continue to be actively researched three decades after the program was launched.
REFERENCES


Variables,” …. 


Moore, Michael J., and Ruskin Morgan (2003), Specification Tests in Simultaneous Equation Marketing Models,” …. 


TABLE 1
MODEL SPECIFICATIONS WITH COMPOSITE VARIABLES IN THE LITERATURE

<table>
<thead>
<tr>
<th>SPECIFICATION</th>
<th>EXAMPLES OF</th>
<th>REPRESENTATIVE STUDIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z = \gamma_0 + \gamma_1 X_1 + \varepsilon$</td>
<td>ROS</td>
<td>Market Share</td>
</tr>
<tr>
<td>$Y = \gamma_0 * + \gamma_1 * X_1 + \gamma_2 * Z + \varepsilon *$</td>
<td>Job Satisfaction</td>
<td>Difference between (algebraic, absolute, squared) actual &amp; desired job attributes</td>
</tr>
<tr>
<td>$Z = \gamma_0 * + \gamma_1 * X_1 + \beta_1 * z_1 + \varepsilon *$</td>
<td>ROS</td>
<td>Marketing/Sales</td>
</tr>
<tr>
<td>$z_1 = \gamma_0 * + \gamma_1 * X_1 + \beta_1 * Z + \varepsilon$</td>
<td>Advertising/Sales Ratio</td>
<td>ROS</td>
</tr>
<tr>
<td>$Z = \gamma_{10} * + \gamma_{11} * X_1 + \beta_{11} * z_1 + \varepsilon_1$</td>
<td>ROS, A/S Ratio</td>
<td>ROS, A/S Ratio</td>
</tr>
</tbody>
</table>

$Z$=Composite variable; $z_i$=Component; $Y$=Non-definitional dependent variable; $X$=Non-definitional explanatory variable
<table>
<thead>
<tr>
<th>DEP.VAR.</th>
<th>CONST.</th>
<th>LGR</th>
<th>DUR</th>
<th>LCONC</th>
<th>TECH</th>
<th>LCAP</th>
<th>LOCAL</th>
<th>A/S</th>
<th>ROS</th>
<th>Π/S</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C&amp;W (Replication): OLS</td>
<td>39 Industries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A/S</td>
<td>1.32</td>
<td>-0.21</td>
<td>1.90*</td>
<td>0.34</td>
<td>-1.21</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>0.74*</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>(2.71)</td>
<td>(0.90)</td>
<td>(0.94)</td>
<td>(0.77)</td>
<td>(1.26)</td>
<td></td>
<td></td>
<td></td>
<td>(0.25)</td>
<td></td>
</tr>
<tr>
<td>ROS</td>
<td>1.15*</td>
<td>0.43</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>0.49*</td>
<td>0.84</td>
<td>0.24*</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>(0.45)</td>
<td>(0.52)</td>
<td></td>
<td></td>
<td></td>
<td>(0.17)</td>
<td>(0.97)</td>
<td>(0.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Π/S</td>
<td>1.15*</td>
<td>0.43</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>0.49*</td>
<td>0.84</td>
<td>1.24*</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>(0.45)</td>
<td>(0.52)</td>
<td></td>
<td></td>
<td></td>
<td>(0.17)</td>
<td>(0.96)</td>
<td>(0.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C&amp;W (Replication): 2SLS</td>
<td>39 Industries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A/S</td>
<td>2.18</td>
<td>0.20</td>
<td>-2.23*</td>
<td>0.38</td>
<td>-0.89</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>0.40</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>(3.19)</td>
<td>(1.11)</td>
<td>(1.12)</td>
<td>(0.86)</td>
<td>(1.45)</td>
<td></td>
<td></td>
<td></td>
<td>(0.49)</td>
<td></td>
</tr>
<tr>
<td>ROS</td>
<td>1.15</td>
<td>0.43</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>0.49*</td>
<td>0.84</td>
<td>0.24</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>(0.66)</td>
<td>(0.57)</td>
<td></td>
<td></td>
<td></td>
<td>(0.21)</td>
<td>(1.08)</td>
<td>(0.20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Π/S</td>
<td>1.15</td>
<td>0.43</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>0.49*</td>
<td>0.84</td>
<td>1.24*</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>(0.66)</td>
<td>(0.57)</td>
<td></td>
<td></td>
<td></td>
<td>(0.21)</td>
<td>(1.08)</td>
<td>(0.20)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Standard errors are in parentheses

* p < 0.05
TABLE 3
ESTIMATES OF THE RE-SPECIFIED C&W MODEL

<table>
<thead>
<tr>
<th>DEP.VAR.</th>
<th>CONSTT.</th>
<th>LGR</th>
<th>DUR</th>
<th>LCONC</th>
<th>TECH</th>
<th>LCAP</th>
<th>LOCAL</th>
<th>A/S</th>
<th>ROS</th>
<th>Π/S</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.92 (1.54)</td>
<td>0.69 (1.35)</td>
<td>-1.51</td>
<td>0.26</td>
<td>-0.65</td>
<td>0.60 (0.49)</td>
<td>0.51 (2.55)</td>
<td>0.66 (0.48)</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Π/Sconstrain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A/S</td>
<td>1.43 (3.53)</td>
<td>0.10 (1.14)</td>
<td>-1.51 (1.59)</td>
<td>0.26 (0.87)</td>
<td>-0.65 (1.38)</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>0.31 (0.34)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

COMPUTED EFFECT ON ROS:

| ROScompute1 | -0.28 | 0.33 | 1.51 | -0.26 | 0.65 | 0.49 | 0.84 | 1.24 | --- | -0.31 |
| ROScompute2 | 1.49 | 0.59 | 0.00 | 0.00 | 0.00 | 0.60 | 0.51 | 0.66 | --- | -0.31 |

Standard errors are in parentheses
* p < 0.05
TABLE 4
MODELS WITH COMPOSITE VARIABLES: SUMMARY OF APPROPRIATE ANALYSES

<table>
<thead>
<tr>
<th>SPECIFICATION IN LITERATURE</th>
<th>APPROPRIATE SPECIFICATION AND ANALYSES</th>
</tr>
</thead>
</table>
| \( Z = \gamma_0 + \gamma_1 X_1 + \varepsilon \) | • Analyze variances and covariances of components:  
  • \( \sigma_z^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2 + 2\sigma_{x_1x_2} \)  
  • Regression of composite on \( X \):  
  • \( Z = \gamma_0 + \gamma_1 X_1 + \varepsilon \)  
  • Regression of components on \( X \):  
  • \( z_1 = \gamma_{10} + \gamma_{11} X_1 + \varepsilon_1 \)  
  • \( z_2 = \gamma_{20} + \gamma_{21} X_1 + \varepsilon_2 \) |
| \( Y = \gamma_0^* + \gamma_1^* X_1 + \gamma_2^* Z + \varepsilon^* \) | • Use components (interactions and transformations, if necessary) as separate explanatory variables:  
  • \( Y = \gamma_0 + \gamma_1 X_1 + \gamma_2 z_1 + \gamma_3 z_2 + \varepsilon \) |
| \( Z = \gamma_0^* + \gamma_1^* X_1 + \beta_1^* z_1 + \varepsilon^* \) | • Specify & estimate equations for each component, with included component affecting the excluded one(s):  
  • \( z_1 = \gamma_{10} + \gamma_{11} X_1 + \varepsilon_1 \)  
  • \( z_2 = \gamma_{20} + \gamma_{21} X_1 + \beta_{21} z_1 + \varepsilon_2 \) |
| \( z_1 = \gamma_0^* + \gamma_1^* X_1 + \beta_1^* Z + \varepsilon^* \) | • Specify and estimate equation with excluded component(s) as explanatory variable(s), not \( Z \):  
  • \( z_1 = \gamma_{10} + \gamma_{11} X_1 + \varepsilon_1 \)  
  • \( z_2 = \gamma_{20} + \gamma_{21} X_1 + \gamma_{22} X_2 + \beta_{21} z_1 + \varepsilon_2 \) |
| \( Z = \gamma_{10}^* + \gamma_{11}^* X_1 + \beta_{11}^* z_2 + \varepsilon_{11}^* \) | • Specify equation for each component, with other component(s) as explanatory variables; impose suitable identifying constraints with caution.  
  • \( z_1 = \gamma_{10} + \gamma_{11} X_1 + \beta_{11} z_2 + \varepsilon_1 \)  
  • \( z_2 = \gamma_{20} + \gamma_{21} X_1 + \gamma_{22} X_2 + \beta_{21} z_1 + \varepsilon_2 \) |

Notes: \( Z \equiv z_1 + z_2 \); * Signifies parameters of a misspecified model.