

[NOTE: This layout and wording of this document may differ somewhat from that of the published version. Consult the published paper or its scanned image before quoting any passage.]

## **REASSESSING CHURCH GROWTH: Statistical Pitfalls and their Consequences\***

Laurence R. Iannaccone

\* Address all correspondence to Laurence R. Iannaccone, Department of Economics, Santa Clara University, Santa Clara, CA 95053. Thanks to Roger Finke, Daniel Olson, Darren Sherkat, Rodney Stark, and three *JSSR* referees for comments and suggestions.

## **ABSTRACT**

Studies of church growth have fallen prey to numerous statistical pitfalls. Flawed methods and inadequate data have reduced the predictive power of past research while systematically biasing its results. The biases work to overstate the importance of the demographic context in which a church exists and understate the importance of the church's own institutional characteristics, such as organizational strictness. This paper uses theory, simulations, and the data re-analysis to explore the empirical difficulties confronting church growth research and to reassess the role of strictness.

**Reassessing Church Growth:  
Statistical Pitfalls and their Consequences for Applied Research**

**I. INTRODUCTION**

Studies of church growth stretch back to the early part of this century, but the touchstone of all modern research appeared in 1972 when Dean Kelley (1986) first published his landmark book, *Why Conservative Churches are Growing*. Kelly called attention to the unprecedented decline in the membership of America's mainline Protestant denominations, which had begun around 1965, and he proposed a compelling but highly controversial explanation. Kelley argued that mainline denominations had become insufficiently *strict*, shedding their distinctive demands and thereby losing their capacity to create meaning and to generate commitment. The story seemed all the more persuasive because many strict denominations (Mormons, Adventists, Jehovah's Witnesses, Southern Baptists, and Assemblies of God) were *not* in decline -- indeed they were growing by leaps and bounds.

In Kelley's (1978: 165) own words, the book generated considerable "controversy, if not actual scandal." Mainline denominational leaders rightly read it as a repudiation of the direction in which they had been moving their churches -- the direction of greater individualism, ecumenical cooperation, social activism, and so forth. Critics accused Kelley of "deceptive statistics" (Bangs 1974: 852), "shallow" research (Bouma 1979: 136), and erroneous conclusions.<sup>1</sup> Yet none could deny the reality of the mainline's membership declines, nor the relative success of their more conservative counterparts. Spurred by these facts, the mainline denominations and a bevy of mainline-affiliated researchers began studying church growth. By the normal standards of

religious research, their efforts were well-funded and highly sophisticated, applying modern statistical techniques to mountains of census and survey data collected from thousands of congregations across America's largest Protestant denominations.

The fruits of this labor appeared in numerous books and articles, but was most fully summarized in two edited volumes published more than a decade apart: *Understanding Church Growth and Decline* (Hoge and Roozen 1979a) and *Church and Denominational Growth* (Roozen and Hadaway 1993).<sup>2</sup> Both volumes grew out of large research projects, and both made heavy use of surveys commissioned by major mainline denominations, including the United Church of Christ, the United Methodist Church, and the United Presbyterian Church. Both books garnered praise from noted religious scholars and church leaders.<sup>3</sup>

Two conclusions stand out in this work. First, "contextual factors" (the demographic and socioeconomic characteristics of the communities surrounding a church) are at least as important, and generally *more* important, in determining church growth and decline than are the "institutional factors" that a church might hope to control. Second, despite the importance of some institutional factors (such as evangelism) and despite the fact that conservative denominations continue to gain relative to their more liberal counterparts, Kelley was wrong. Strictness does not influence or explain the growth. In an overview chapter to the 1993 volume, editors Hadaway and Roozen (1993: 42) emphasize that "[r]esearch has provided little support for what has been called 'the Kelley thesis'... Strictness ... is unrelated to growth within liberal or conservative families. Kelley's 'theory' is best understood as one of sectarian survival -- not congregational or denominational growth." Authors repeatedly express this view throughout both volumes, and its status as demonstrated fact may explain why only one study in the first volume (Hoge 1979) and

none in the second were explicitly designed to test the strictness thesis.<sup>4</sup>

These two conclusions may have given some solace to the shrinking mainline. The first suggests that they really can not do much (nor could they have done much) to stave off decline. The second asserts that insofar as they can do something, it need not be anything so radical as a turn toward the strictness, separatism, and other-worldliness that characterizes the conservative wing of American Protestantism.

A review of the data and methods underpinning most church growth research casts serious doubt on the basis for both conclusions. Statistical problems abound in past studies, problems that have gone largely unexamined.<sup>5</sup> These not only reduce the predictive power of the research; they systematically bias its results. Nearly all the biases work to overstate the importance of demographic context and understate the importance of a church's own institutional characteristics, most notably its organizational strictness. This paper uses theory, simulations, and the re-analysis of old data to identify the statistical problems in past research. My basic message is summed up by a methodological warning found in Hadaway (1989: 160): "If the researcher knows where the mines lay and how to avoid them, then good research is possible, but for the unwary ... the chances for fatal errors are very high."

The paper has obvious relevance for churches that seek to grow, but it also relates to a developing body of theory and data that model religious organizations in terms of the costs they impose on their members (Stark and Bainbridge 1985, 1987, Finke and Stark 1992, Iannaccone 1992, 1994). According to this work, strictness promotes strength by discouraging free riding and screening out less committed members. The rehabilitation of Kelley's thesis gains wider significance when viewed against the backdrop of work that embeds strictness within a broad

theory of sectarianism, rational action, and the evolution of religious organizations.

## **II. UNDERSTANDING THE PITFALLS**

The pitfalls I address are all instances of standard statistical problems. One may therefore turn to any intermediate statistics text to see the general problems analyzed in abstract, mathematical terms. For present purposes, however, equations are less illuminating and less compelling than concrete examples. Moreover, since the relationships behind any real-world data set remain forever uncertain, I will illustrate each of the pitfalls with statistical simulations. The simulations can be fully specified and repeated at will, leaving no doubt as to their underlying structure and statistical properties. Though one may debate the extent to which any particular simulation approximates real-world situations, I seek merely to show that widely used methods lead to incorrect inferences when applied to reasonable data. The plan therefore is to take popular methods, apply them to simulated data structured to mimic the theories that researchers have proposed, and show how the methods break down.<sup>6</sup> Having identified problems in this fashion, I will then suggest remedial actions that can avoid or mitigate them when working with real data.

To keep the analysis as concrete and realistic as possible, I will adopt the conceptual framework that has dominated previous research. Membership growth will be attributed to "contextual" and "institutional" factors, and denominations will vary along a continuum of strictness/conservatism.

**Pitfall #1: Restricting the sample (to a single denomination) understates the impact of institutional characteristics**

When scholars speak of Kelley's thesis having been refuted, they usually are referring to large-scale, census and survey-based statistical studies of congregational growth. Nearly all these studies looked for the statistical correlates of church growth within a *single* denomination. So, for example, Hadaway (1980) and Roof, Hoge, Dyble, and Hadaway (1979) identified the correlates of church growth in a sample of 681 United Presbyterian congregations; McKinney (1979, c.f. McKinney and Hoge 1983) undertook a similar analysis of 263 United Church of Christ congregations; Perry and Hoge (1981) studied 205 Presbyterian congregations; and Thompson, Carroll and Hoge (1993) studied yet another 593 Presbyterian congregations. All these studies paid some attention to strictness or conservatism, and nearly all concluded that strictness had but "weak" effects on congregational growth. Moreover, "institutional" variables of *every* sort tended to look weak relative to "contextual" variables such as population growth, neighborhood characteristics, incomes, and the like.

There is, however, an overwhelming problem with these studies. A denomination is by its very nature committed to a particular set of institutional characteristics -- theological, behavioral, and organizational. Hence, the congregations *within* a given denomination inevitably manifest much less institutional variation than that which actually exists across the entire religious market. This narrowness is likely to be especially great with regard to the attributes that constitute strictness. No Mormon congregation, however liberal, is free to ignore LDS church rules regarding tobacco, alcohol, tithing, temple rites, organizational structure, excommunication, and the like. And conversely, no Presbyterian congregation, however conservative, is in a position to

impose these rules. On the other hand, most contextual/demographic attributes vary greatly across the congregations of a single denomination. The average Presbyterian may be relatively well off, but some congregations are poor and less educated, have many large families or high rates of unemployment, are located in rural areas or areas of increasing population, and so forth.

In short, looking across the "observations" of single denomination data set, the contextual variables always exhibit much more variation (relative to that of society as a whole) than do the institutional variables. The situation parallels that of an agricultural experiment in which rainfall levels are nearly identical on every test plot, but fertilizer levels vary substantially. Such an experiment will inevitably find that actual variation in crop yield depends more on the observed variation in fertilizer than the *observed* variation in water (whatever the actual importance of water). Single-denomination studies are likewise bound to find that institutional variables (particularly those most tightly associated with denominational differences) account for a relatively low fraction of the observed variation in growth.

A simulation highlights the problem. Assume that congregational growth is, in fact, the sum of three independent factors: a contextual variable (growth in the local population), an institutional variable (strictness), and unobservable effects. Hence, the true regression equation for growth is,

$$\text{GROWTH} = \text{POP\_GROWTH} + \text{STRICTNESS} + \epsilon \quad (1)$$

Let each observable variable account for 25% of the variation in membership growth so that the true  $R^2$  is .50. (To keep things simple, let all the independent variables be normally distributions across the full population of Protestant congregations. This assumption is not necessary, but it guarantees the applicability of standard regression statistics.) Assume, moreover, that each

congregation belongs to one of eight denominations arrayed along the strictness continuum. Within each denomination, congregations are drawn randomly from a strictness "band" that covers a limited (1.5 standard deviation) portion of the entire strictness distribution. The bands of adjacent denominations overlap, so that the congregations of the first denomination are not all less strict than the congregations of the second denomination, and so forth. On the other hand, assume that local population growth rates vary independent of denomination.

To simulate this situation using standard statistical software, one may randomly generate numerical values for each simulated congregation's POP\_GROWTH, STRICTNESS, and  $\epsilon$ ; and then calculate the congregation's GROWTH according to equation (1). Figure 1 shows the results from one such simulation, computed with the Stata statistics package. The figure displays the relationship between growth and strictness for a thousand randomly generated "congregations." The stipulated positive relationship between growth and strictness is quite visible. Applying standard regression techniques to a random sample of one hundred congregations, one obtains the coefficients, t-statistics, and  $R^2$  found in columns (1) through (3) of table 1.

Note that the regressions recover the "right" results from the simulated data, which is to say results consistent with the actual, underlying model specified by equation (1). Working from a random sample of all the congregations, we correctly infer that POP\_GROWTH and STRICTNESS are each significant, equally important, and each account for about a fourth of the variance in GROWTH.

But what if we restrict the analysis to a single denomination? The circles in figure 1 correspond to the congregations of just one, fairly liberal denomination. Following the

assumptions of the simulation, these congregations all fall within a fairly narrow segment of the strictness continuum. Within this *restricted* sample, the actual, underlying relationship between growth and strictness is far from clear. On the other hand, the effect of population growth remains strong since the denomination's congregations vary widely with respect to this "contextual" variable.

Regressions over one hundred simulated congregations drawn randomly from this single denomination yield the results printed in columns (4) through (6) of table 1. In these regressions, the estimates for POP\_GROWTH remain approximately correct, but the  $R^2$  and t-statistics for STRICTNESS are far too low.<sup>7</sup> Whereas population growth still explains around one fourth of the variance in growth, strictness now accounts for virtually nothing. Faced with these results, a researcher would certainly conclude that institutional effects are "weak" relative to contextual effects. But the weakness is an illusion, an artifact of the restricted, denominational sample. The *true* relationship remains that of equation (1).

Although the specific numbers in table 1 reflect the assumptions built into the simulation, the basic results remain valid for virtually any conceivable real-world situation. One thus finds that *any* study of congregations within a single denomination inevitably underestimates the relative importance of institutional variables such as strictness, doctrine, organizational structure, and the like.<sup>8</sup> Moreover, the problem persists even in multi-denominational studies as long as the denominations occupy nearby points on the denominational spectrum. A random sample of congregations from the Episcopal, UCC, United Presbyterian, and United Methodist churches, still represents a small fraction of the strictness spectrum that extends to Adventists, Mormons, Jehovah's Witnesses, and beyond.<sup>9</sup> This fact validates the concern voiced years ago by Dean

Kelley (1979:335):

"Our authors [of church growth research] ... eschew the effort to obtain data on the most intense forms of religious behavior -- 'new religious movements, sects, and cults' -- and confine themselves to the culture-affirming mainline churches ... [A] *whole wing of the spectrum of religious behavior in this country* -- by far the most interesting and informative -- *has been left out.*"

**Pitfall #2: Searching over numerous variables inflates  $R^2$  and invalidates significance tests**

Significance tests were designed for simpler times. Decades ago, an agricultural researcher might seed a hundred plots of land, varying the level of fertilizer from one to another. He would then hand calculate a single bivariate regression to test whether fertilizer had a "statistically significant" effect. If the estimated coefficient was significantly greater than zero at the 5% level, he might reasonably infer that fertilizer *does* affect crop yield, since the odds of getting such a strong result purely by chance was less than one in twenty.

The appropriate inferences change dramatically if the researcher had not only tested for the effects of fertilizer, but had also measured and calculated yield regressions for each of 99 other factors -- soil composition, drainage, rainfall, shade, wind, and so forth. With a total of 100 explanatory variables, standard t-tests can be expected to turn up five "statistically significant" even if *none* of the variables exerts any real influence on crop yield. This, after all, is what a "5 percent significance level" means -- a statistic so extreme that mere chance would cause it to arise only five times out of a hundred. With five or perhaps even ten "significant" correlations out of a hundred, the researcher has no business inferring *anything* from the data.

Sadly, the methods of most large-scale congregational studies exactly parallel those of the hapless farmer. Working with hundreds of hypothetical explanatory variables, they typically correlate *each* variable with growth and then retain those that prove "significant" at the 5, 10, or 15 percent level. But the significance test in question applies only to a *single* comparison involving a single, preselected variable! Is it any wonder that these studies routinely identify 20 to 30 "significant predictors" of growth? Mere chance would yield 15 spurious predictors in data set with 300 purely random variables.

This criticism might sound harsh, but consider several authors' own accounts of their procedures in recent studies:

"Because of the extensiveness of the data (more than 700 variables) a series of preliminary analyses was conducted to reduce the sets of predictor variables to a more manageable number that could be incorporated into the final analyses... All potential predictor variables that displayed statistically significant (at  $p < .05$  level) and theoretically meaningful correlations were retained and grouped into several subsets." (Welch 1993: 327, 370)

"The length and complexity of data displays discussed thus far preclude cogent summary. Such mind-numbing 'laundry lists' of variables ... becloud larger issues concerning which characteristics are most strongly related to congregational growth and decline. ... The table displays only those predictors that had partial R values of .10 or higher, after controlling for all of the other characteristics in that theme. ... Having thus sifted through the variables ... a 'grand model' was constructed." (Donahue and Benson 1993:235-237).

"Over 200 predictor variables were available for the analysis. We correlated all of them with the membership change measure and retained only those having correlations significant at or near the .10 significance level. .... A total of 32 variables were retained ..." (Thompson, Carroll, and Hoge 1993: 192)

Such procedures can not help but identify "significant" variables that are, in fact, totally irrelevant. A Nobel laureate, Ronald Coase, once put the matter bluntly: "if you torture the data long enough, Nature will confess" (Kennedy 1985:76). These false confessions grow in direct proportion to the number of variables and/or models over which one searches. Hence, searching for the significant predictors among a list of several regressors may not be too misleading (and is in fact common), but searching over several hundred predictors almost always yields nonsense. When the quoted authors identify 32 explanatory factors (significant at the .10 level) from an initial set of 200, one must recall that 20 such factors would arise in a purely random set of data.

The errors multiply when the researcher proceeds from searching to regressing, testing groups of "significant" variables for their contribution to  $R^2$ . The resulting regressions inevitably explain a substantial portion of their variance, *even if the underlying variables are pure noise*. The problem is multicollinearity and insufficient degrees of freedom masked by the initial search.

In fairness to the church growth literature, most researchers *do* make a point of adjusting their  $R^2$  calculations to account for degrees of freedom. For example, Thompson, Carroll, and Hoge (1993: 207) report that "All the  $R^2$ 's are adjusted downward using the standard adjustment formula ... based on the number of cases and the number of variables *in the regression*." (The

italics are mine.) McKinney and Hoge (1983: 60), Welch (1993:371), Roof, et. al. (1979:368), and most other previously cited studies likewise adjust their regressions for the number of explanatory variables *in that regression*.

But the authors overlook that meaningful adjustments must take account of the entire set of variables that were examined, not just the few that happen to have been retained. Regressing growth onto the ten "significant" variables in a fifty-variable data set is tantamount to regressing onto the entire set of 50 variables. After all, the forty "insignificant" variables add little or nothing to  $R^2$  once we have already included the ten "significant" variables. It follows that in such cases the appropriate adjustment to  $R^2$  comes not from the ten variables that were retained but rather from the entire fifty that were examined.

The numerical consequences are immense. The standard adjustment is  $\text{adj-}R^2 = 1 - (1 - R^2)(n-1)/(n-k-1)$ , where "n" denotes the number of observations and "k" the number of regressors. This inflates the unexplained portion of the variance by  $(n-1)/(n-k-1)$ . A one-hundred-observation, fifty-variable regression with a raw  $R^2$  of .50 thus converts to an adjusted  $R^2$  of zero.

But consider the actual studies. Roof, et. al. (1979: 199) derive their regressions from "681 congregations ... [and] nearly 500 variables." This implies that any raw  $R^2$  under .73 adjusts down to zero. McKinney (1979: 233) works with 263 congregations and 70 predictors, implying that any  $R^2$  below .27 may be entirely spurious. Thompson, Carroll, Hoge (1993:189,192) work with 593 congregations and "[o]ver 200 predictor variables," implying that we discount any  $R^2$  under .34. Similar considerations apply to all the previously cited studies. Sadly, the actual  $R^2$  obtained in most of these studies fall *below* these levels. Roof, et. al. obtain an  $R^2$  of .26; McKinney reports an  $R^2$  of .25, and Thompson, Carroll, and Hoge's most comprehensive

regression has an  $R^2$  of only .32.

The reader will note that the cited studies never actually ran their regressions over all available predictors, but instead built their regressors from the strongest zero-order correlates of growth. This makes my preceding re-adjustments too severe. But simulations detailed in my technical appendix show that the two-step, search and regress procedure remains deeply flawed in any case (yielding long lists of spurious predictors and  $R^2$  values that overstate equations' predictive power).<sup>10</sup>

The only solution is restraint. To produce valid test statistics, the researcher must not work with hundreds of explanatory variables, must swear off searching, and must limit regressions to a few well-defined, predetermined variables. Kitchen sink correlations may have use in purely exploratory work (though they inevitably turn up more junk than truth), and they may help show that a particular pre-selected theory is truly robust (by proving that its variables remain significant even when one tries to drowning them with controls), but they must not be confused with hypothesis testing.

**Pitfall #3: Measurement errors (that affect institutional variables more than contextual variables) understate the importance of institutions:**

Applied researchers have a hard life. Even when their models are correctly specified, and even when they have identified the true causes and have excluded the spurious correlates, they must struggle with measurement error. Josiah Stamp (1929: 258-9) summarized the dilemma long ago:

The Government are very keen on amassing statistics -- they collect them, add them, raise them to the  $n$ th power, take the cube root and prepare wonderful diagrams. But what you must never forget is that every one of those figures comes in the first instance from the village watchman, who just puts down what he damn pleases.

Substitute "denominational headquarters" for "government," and "congregational secretary" for "village watchman," and you have the problem confronting church researchers.

Statistical horror stories abound and need no repetition. It suffices simply to note that most congregations maintain sloppy records, have few incentives for getting their numbers right, operate with little or no oversight, and face no penalties for error. Church members are likewise prone to misstate facts regarding, for example, their rates of church attendance or contributions, through thoughtlessness, forgetfulness, miscommunication, or outright deceit. A still greater margin for error and miscommunication exist when polling people about their personal feelings, beliefs, and attitudes toward their congregations.

Church researchers must live with these problems, but they must also acknowledge the biases they introduce. When researchers test a theory about factor X (e.g., "strictness") but have at their disposal only some loose *proxy* for X, then the estimated impact of X will be biased toward zero (Kennedy 1985: 113, 123). Moreover, if the researchers are seeking to assess the effects of two different sets of factors X and Y, and random measurement errors plague X more than Y, then the regression results will be systematically biased in favor of Y. The estimated coefficients, betas, t-statistics, and  $R^2$  will all suggest that Y has more impact than X, even if both are equally important.

Now it would seem that the "contextual" variables in most church growth models are much less error-filled than the "institutional" variables, since the former are demographic quantities derived from systematic, large-sample government census statistics (of population growth, property values, incomes, family size, racial distributions, and so forth), whereas the latter are mostly attitude scales derived from relatively small, non-random samples of congregation's membership.<sup>11</sup> Measures of "strictness" may be particularly error-prone, because very few mainline, congregational surveys even mention the inventory of activities and attitudes that Kelley (1986:84) equated with strictness. This means that most tests of demographics versus strictness, or more generally, external context versus internal institutions, will overstate the relative impact of the former.

To see the bias at work, assume that membership growth is, in fact, equally determined by population growth and strictness (as was previously assumed in equation [1]). In table 1, columns (1) through (3), we saw that regression yields correct conclusions when working with the true variables and a random sample of simulated congregations. Assume now that the researcher can not observe true population growth and strictness, but must instead use proxy variables. Assume moreover that the population proxy is much less "noisy" than the strictness proxy. Specifically, let

$$\text{POP\_PROXY} = \text{POP\_GROWTH} + \epsilon_1$$

$$\text{STRICT\_PROXY} = \text{STRICTNESS} + \epsilon_2$$

where all the right hand side variables have independent normal distributions, with the variance in  $\epsilon_1$  one-tenth that of POP\_GROWTH and the variance of  $\epsilon_2$  equal to that of STRICTNESS.

Regressing GROWTH onto the true variables and their proxies (in a 1000-observation, simulated

data set) then yields the results printed in table 2.<sup>12</sup>

Columns (1) through (3) show the results obtained when GROWTH is regressed onto its true determinants, POP\_GROWTH and STRICTNESS, either singly or jointly. The results correctly imply that both factors are equally important, each accounting for about one fourth of the total variance in growth.

Columns (4) through (6) show the corresponding regression of growth onto the *proxy* variables for population growth and strictness. Note that these results *seem* to show that demographics are far more significant than strictness -- the population proxy explains 20% of the variance in growth, whereas the strictness proxy explains only 5%; the population proxy beta coefficients are nearly twice the strictness betas; and the population t-statistics are nearly twice the strictness t-statistics.<sup>13</sup> But the apparently obvious conclusion is wrong. We know that both factors are equally important, and both contribute 25%, because we simulated the data using equation (1). Strictness appears less important only because its measurement involved a relatively large error. In a case like this, the only practical solution is a major effort to obtain better measures of strictness, something largely absent from the growth literature.

**Pitfall #4: Stepwise-hierarchical regressions that force institutional effects to follow contextual effects are biased:**

Church growth studies typically assess the relative importance of contextual versus institutional factors by means of stepwise-hierarchical regressions that *first* calculate the explanatory power of contextual variables and *then* estimate the additional explanatory power derived from institutional variables. For example, Thompson, Carroll, and Hoge (1993:199)

divide their explanatory variables into three groups that they call "community context," "congregational demography," and "institutional factors" (which include recruitment efforts, theological orientation, church facilities, organizational structure, congregational attitudes, and the like). Regressing growth onto the first set of factors yields an adjusted  $R^2$  of .16; adding the second set of factors increases  $R^2$  to .20; and adding the final set increases  $R^2$  to .32. From these numbers the authors infer that "community context accounts for 16% of the variance, congregational demography accounts for an additional 4%, and institutional factors account for another 12%." Taking the same approach, Roof, et. al (1979:220) report that contextual variables account for 14.6% of the growth in church membership and that the addition of institutional variables accounts for another 11.4%. The overall conclusion? Thompson, Carroll, and Hoge (1993: 205) observe that in both studies "contextual factors were a bit stronger than institutional factors." (Similar procedures and results characterize Hoge 1979, McKinney 1979, McKinney and Hoge 1983, Olson 1993, and Welch 1993.)

There is good reason to question this conclusion, if only because it follows a table indicating that institutional factors are substantially *more* important than community context (Thompson, Carroll, and Hoge 1993:193-196). The table groups all 32 explanatory variables into clusters, seven of them contextual and seven institutional, and lists the percent of variance in growth independently explained by each cluster (i.e, the adjusted- $R^2$  when growth is regressed onto all the variables in that cluster). The highest contextual-cluster explains 11% (followed by other contextual clusters that explain 8, 7, 4, 4, 2, and 1 percent), whereas the highest institutional cluster explains 14% (followed other institutional clusters that explain 12, 11, 6, 6, 5, and 4 percent). Why conclude that contextual factors are more important than institutional factors,

when each institutional cluster explains more variance than the corresponding contextual cluster? And why say that all the institutional variables together account for only 12% of the variance in church growth, when one of seven institutional clusters predicts 14% of the variance all by itself?

As it turns out, the authors' conclusions derive entirely from the *predetermined* order in which variables entered the regressions: *first* context, *then* institutions. Very different results would have emerged had the order of entry been reversed. As things stand, anything that *can* be (statistically) attributed to context *is* attributed to context; institutional factors only get a shot at what is left over. To be sure, Roof, et. al. (1979:200) do acknowledge that "the size of the coefficients can be influenced by the order in which the clusters are introduced" -- see also McKinney and Hoge (1983:63) and Welch (1993:371). But this critical proviso is glossed over on the grounds that the order of inclusion parallels the "causal order" of the variables. The researchers assume that in any statistical model demographic variables *must* precede congregational attitudes because the latter are not likely to influence the former, whereas the former might very well shape the latter.

The assumption is wrong. *Even if contextual factors are causally prior to institutional factors, the stepwise-hierarchical approach yields biased results that systematically underestimate the importance of institutions.*<sup>14</sup>

To illustrate the fallacy, consider a study of how mold and excessive rain inhibit the growth of grapes. Rain can damage grapes directly (by causing them to split) or indirectly (by stimulating the growth of mold). Rain is thus "causally prior" to mold. Does it therefore make sense to regress crop yields onto rainfall (obtaining an  $R^2$  of, let us say, .80), then expand the model to include the level of mold (obtaining an  $R^2$  of .90), and then conclude that "mold explains

only 10% of the variance in growth" or that "rain is eight times more important than mold when it comes to determining grape yields"? By no means. The stepwise strategy fails on two counts: first, by describing the situation so as to understate the impact of mold (which might well account for most crop losses, while grape-splitting accounts for a little); and second, by directing attention away from mold-oriented solutions. Farmers are led to focus on rain, a factor beyond their control, rather than inexpensive fungicides like sulfur. In the same manner, a stepwise approach to church growth leads to misleading descriptions and biased prescriptions.

Simulations highlight the problem and suggest a solution. Consider the various path models of church growth depicted in figure 2. Arrows depict causation, dotted lines depict *non-causal* correlation. The figures seem to encompass all the major scenarios that scholars have suggested. In the first, income and strictness affect growth equally and independently. In the second, income and strictness are correlated, but only income influences growth. In the third, the situation is reversed and only strictness influences growth. In the fourth, income and strictness are correlated (due, perhaps, to some common prior factor) and both influence growth. In the fifth, income affects strictness, which in turn affects growth, but income itself has no direct impact upon growth. The sixth case parallels the grape example, since income now affects growth both directly and indirectly.<sup>15</sup>

With a few additional assumptions (such as the assumption that all observable and unobservable factors are normally distributed and the assumption that each true cause contribute 33% to the actual variance in growth) one may readily simulate all six models. For example, to simulate the first model, I generated three independent and identically distributed random variables; labeled these variables INCOME, STRICTNESS, and  $\epsilon$ ; and then calculated GROWTH

according to the equation  $GROWTH = (-1) \cdot INCOME + (+1) \cdot STRICTNESS + \epsilon$ . To simulate the second model, I generated the independent variables in such a way that  $INCOME$  correlates with  $STRICTNESS$ , then calculated  $GROWTH$  according to the equation  $GROWTH = (-1) \cdot INCOME + \epsilon$ . Proceeding in this manner, I produced six sets of simulated data corresponding to the six models in figure 2. Each data set was then analyzed using the same stepwise-hierarchical regression methods characteristic of past research.

Table 3 shows the results. In every case except the first two, the stepwise-hierarchical approach *overstates* the influence of income, the contextual factor, and understates the influence of strictness, the institutional factor. Consider, for example, the third column which summarizes results for model 3. The data for model 3 were simulated in such a way that strictness and income are correlated but only strictness affects growth. Hence, as is noted in the "actual  $R^2$ " rows, income contributes nothing to this model's *actual* variation in growth, whereas strictness contributes .33. But looking at the "estimated  $R^2$ " numbers, we see that regressing growth onto income alone yields an *estimated*  $R^2$  of .15, and regressing growth onto income and strictness yields an estimated  $R^2$  of .30. This leads to the erroneous conclusion, listed under "imputed  $R^2$ ," that income and strictness are equally important. The error has a simple cause: in this model, and every model except the first two, the underlying correlation between income and strictness guarantees that the initial (context-only) regression suffers from a classic "omitted variable" problem that inflates  $R^2$  and biases the estimated coefficient. To justify the stepwise approach, a researcher must thus posit an *absence* of correlation between contextual and institutional variables (which clearly is false) or the *impossibility* of a true causal effect running from the institutional factors to growth -- the very hypothesis that the approach was supposed to test!<sup>16</sup>

Reversing the order of stepwise entry will not solve the problem, though it may highlight the problem's existence. The only real alternative lies in reporting the complete results for a multiple regression in which both factors appear simultaneously. Correlation among the independent variables will still make it difficult to portion out  $R^2$ ; but comparing the magnitudes of t-statistics and/or betas in a single, multiple regression is much more defensible than comparing the increments to  $R^2$  in a series of hierarchical regressions. Table 4 reports the results of standard multiple regressions using the same data. In every case the results are consistent with the actual underlying model used to generate the data.

The moral is clear. Stepwise-hierarchical regressions that force institutional effects to follow contextual effects are biased and must be avoided. The conclusions drawn from such statistics can not be trusted.

### **III. ESCAPING PITFALLS: STRICTNESS REVISITED**

It is by now clear that potentially serious problems plague past studies of church growth. To demonstrate that the problems have, in fact, led to inappropriate conclusions, and to see how they can be overcome, I will re-analyze the data from a study entitled "A Test of Theories of Denominational Growth and Decline." I have chosen this study for several reasons. It is one of the few studies that work with denominational data, thereby avoiding the problems of a single-denomination sample. It is the only study designed from the ground up to test Kelley's theory of strictness and growth. As such, the study's author, Dean Hoge, went to great lengths to measure the social, organizational, and doctrinal attributes that Kelley identified as the determinants of church strength. Most importantly, the author documented his work with such care that one may

fully recover his data and fully replicate his analysis.<sup>17</sup> In the years since Kelley first wrote about strictness, no empirical study has addressed his thesis with anything approaching the precision, ingenuity, and thoroughness of this test.

Since Hoge described and documented his work so well, it suffices merely to summarize its major features. First, the study worked with denominations as the unit of analysis. This limited the data to 16 cases corresponding to the major Protestant denominations for which a wide range of data were available. Second, the study sought to explain denominational growth rates in terms of other denominational attributes, both "contextual" and "institutional." Hoge derived his contextual measures (including average income, education, occupational prestige, family size, and regional distribution) for each denomination from government data, the Glenmary census of religious bodies, and NORC's General Social Surveys. In order to obtain institutional measures corresponding to Kelley's theory, Hoge asked twenty-five experts (church historians, sociologists of religion, and denominational leaders) to complete a questionnaire in which they ranked each of the denominations along a series of seven-point scales. The scales concerned the denominations' strength of ethnic identity, theological conservatism, attitudes toward ecumenism, centralized or congregational polity, emphasis on local and community evangelism, involvement in social action, emphasis on distinctive life-style and morality, and attitude toward pluralism of beliefs.

Given the relatively large number of explanatory variables and the very small number of cases, searching for significant correlations is problematic. Fortunately, all the institutional ratings (except "ethnic identity" and "polity") were so strongly intercorrelated that one may view them as proxies for a single underlying trait. It therefore suffices to sum the ratings into a single scale or

select any one rating and ignore the rest.<sup>18</sup> Perhaps the most reasonable scale is the one derived from evangelism, distinctiveness, and pluralism, since Hoge (1979: 192) describes these ratings as "[t]he factors Kelley stressed most."

Hoge did not follow this approach, but instead ran separate zero-order correlations between denominational growth and all the variables. Working with the nine contextual variables two at a time, he found that average family size and regional distribution together explained 59% of the variance in growth among the 16 denominations. Despite the fact that most strictness measures predicted much better than this, Hoge (1979:190) took this result as proof that "contextual factors alone can explain over half the total variance in denominational growth rates in 1965-77."<sup>19</sup> His conclusion bears repeating:

"... the main causation should be seen as being from contextual factors first, then from institutional factors, not the opposite. The contextual factors taken alone can explain over half the variance in denominational growth in 1965-75... This conclusion is somewhat different from Kelley's view which puts most emphasis on institutional factors. Perhaps Kelley underemphasized contextual factors due to the hortatory purpose of his book. ... Our purpose is explanation, not advocacy, and we estimate that contextual factors comprise more than half the explanation for denominational growth or decline rates." (Hoge 1979: 194-195)

But we have already seen that in cases like this a stepwise hierarchical approach is biased *even if* contextual factors are "causally prior" to institutional factors. Multiple regression provides a more appropriate basis for inference, and at the very least one must check how sensitive the conclusions are to the order in which variables enter the regressions.

Table 5 shows the result of a more complete, multiple regression analysis. The first column replicates Hoge's finding that the best two-variable combination of contextual variables does indeed yield an  $R^2$  of .59.<sup>20</sup> But columns two, three, and four show that the strictness ratings do *far* better. The best rating, distinctiveness, explains 94% of the variance in growth. A comprehensive additive scale of all the ratings (except polity and ethnic identity) explains 83%. A three-item scale, based on the factors Kelley stressed most, explains 88% of the variance. Indeed, *every* institutional variable (except ethnic identity and polity) explains at least 70% of the variance in growth, more than the best *two* contextual variables combined.<sup>21</sup>

The most important results are found in columns five through seven. These show that contextual effects wash out of the equation once the strictness variables enter. In every case, the contextual variables' betas and t-statistics approach zero, while the statistics for the strictness variables remain extremely strong.

Figure 3 explains all. The relationship between growth and strictness is nearly perfect. It is no wonder, therefore, that the contextual variables drop out -- a single strictness scale derived from Kelley's theory, leaves nothing else to be explained!<sup>22</sup>

In short, given the inherent limitations of the data, a more powerful empirical vindication of Kelley's theory simply is not possible. Strictness measures trounce the competition in this head-to-head test. No other conclusion is possible.<sup>23</sup>

#### **IV. CONCLUSIONS**

Three major conclusions emerge from the preceding methodological review of the empirical research on church growth.

First, like all social scientists, church growth researchers must pay careful attention to the limitations of their data and methods. As computers have become vastly more powerful and statistical programs more convenient, the dangers of data mining have increased. These problems are in no way mitigated by ever-larger data sets and ever-more sophisticated computational techniques -- indeed, larger data sets are more prone to throw up spurious findings, and sophisticated statistics (including most non-linear procedures) tend to be more sensitive to specification errors than their simpler counterparts.

Second, statistical problems have almost certainly *biased* the results of past studies. Researchers have used data and methods that systematically privilege contextual factors relative to institutional factors. From an applied perspective, this means that church growth researcher has understated the extent to which churches and their leaders can alter their fate. From a theoretical perspective, it means that researchers have been too quick to dismiss particular theories of religious organization (particularly those that trace the vitality of religious organizations to their strictness, costliness, sectarianism, and so forth).

Third, future research must not repeat the mistakes of the past, and wherever possible old data must be re-analyzed using methods less prone to error and bias. This paper undertook one such re-analysis, thereby reversing the original conclusion. The data from most other studies still exist and should be subjected to similar re-analyses. It is difficult to predict which findings will stand and which will fall. Some results will certainly prove robust, but many others may not. Despite a raft a reasonable-sounding conclusions, it is not yet clear that large data sets and complex statistics have led to more precise and objective insights than those derived from the personal experiences of church leaders and church growth consultants.

Not all the news is bad. Each of the previously-described pitfalls carries with it a solution or an alternative, many of them straightforward. I have already demonstrated a solution to the bias inherent in hierarchical models: Do not make stepwise comparisons of  $R^2$  to assess the relative impact of different variables. Where multiple factors are believed to be at work, limit the analysis to multiple regressions and accept the fact that, absent better data or a controlled experiment, the individual impact of some factors can not be fully assessed. The solution to errors in the independent variables and restrictions on the independent variables is, wherever possible, better data. Rather than ransacking a 500-item survey distributed to thousands of members in a single denomination, direct a few carefully worded questions (based on theory and past research) to a reasonable number of congregations across a wide range of denominations. If at all possible, assess the level of response error in key items and take account of this error when drawing conclusions about the relative importance of different items. Where better data are unobtainable, identify the limits and potential biases inherent in the current sample. Faced with a large number of survey items, the researcher must resist the temptation to search for "significant predictors." The most defensible approach is to identify a small number of test items and controls (implied by a carefully-devised, tightly-stated theory) before doing any analysis. All other items should be ignored or introduced only to check that the pre-selected variables are truly robust. If this much self-restraint proves untenable, then try the following alternative: use a random sample of the data for ad-hoc exploration and then test the apparent predictors against the *remaining* data. Most of the spurious correlations will disappear in the second stage and most of the true correlations will remain.<sup>24</sup>

If we do all these things, we may finally learn why churches *really* grow. We may even discover that both the reasons and the evidence were in plain sight all along.

## Notes:

1. See also "Why Conservative Churches *Really* are Growing: Kelley Revisited" (Bibby 1978) and "'Strictness' and Church Membership," (McFaul 1974: 281), which went so far as to claim that "[c]ontrary to Dean Kelley's thesis ... a return to 'strictness' is the cause of, rather than the solution for, the mainline Protestant denomination' membership decline." For his own part, Kelley (1978: 171) remained "unrepentant and unreconstructed." Inskip (1993) offers an excellent summary and critique of Kelley's critics.
2. See Roozen and Hadaway's (1993) bibliography for dozens of other books and articles on church growth.
3. Church historian Martin Marty, admonished readers of the first volume to "not open your mouth about trends and patterns in church membership and participation unless you have read this book." The renowned church growth consultant, Lyle Shaller, declared of the second volume: "Never before has so much high quality research on church growth been available in one book." And Ken Bedell, editor of the *Yearbook of American and Canadian Churches* called the second volume the "best sociological analysis to date on church and denominational growth."
4. Several studies addressed strictness *en passant*. In the course of correlating numerous survey items against growth, they note which if any reflect some aspect of strictness. According to Kelley (1979: 340), however, the items tested in this fashion bear little relationship to what he called strictness.
5. In a special issue of the *Review of Religious Research* that includes five essays about methodological problems in congregational studies, Roozen and Carroll (1989: 116) note the continuing "lack of a cumulative, critical methodological literature on the study of congregations." The essays in the special issue examine a variety of sampling problems (such as insufficient response rates, non-representative samples, and difficulties identifying the relevant population), but they do not address any of the statistical pitfalls that I will examine. One reviewer has stated that these pitfalls were "fully debated" and well-known to "everyone" involved in the 1970s church growth research. Unfortunately, the debate never found its way into print.
6. Readers may write the author for disk and printed copies of all the data and simulation programs described in this paper. Though written for Stata, they can be translated to SPSS, Systat, or any other full-featured microcomputer statistics package.
7. The sample restriction does not bias the regression coefficient, but this is of little help since virtually all the institutional variables in previous studies are attitude scales that defy ratio level measurement. In practice, therefore, church growth researchers never concern themselves with the numerical value of these coefficients, only their beta values, and as table 1 shows these betas *are* biased downward. For a mathematical treatment of the range restriction problem, which verifies and extends the simulation results, readers may write for the technical appendix to this

paper.

8. Although the specific numbers obtained above depend on the assumptions built into the simulation, the basic results remain valid for virtually any conceivable real-world situation. The technical appendix, available upon request, analyzes the mathematics of range restrictions, and thereby verifies and generalizes the simulations.

9. The problem persists when comparing the results of different, single-denomination studies, *even if* some of the denominations in the studies are quite liberal and some quite conservative. As long as the congregations of each individual denomination occupy a relatively narrow segment of the strictness spectrum, the observed strictness effect within each denomination will be small. This fact may explain Roozen and Hadaway's (1993:42) claim that strictness "is unrelated to growth within liberal or conservative families."

10. The simulations generate GROWTH as a function of ten true causal variables (CAUSE1,...,CAUSE10) and a large random error term. In addition to these causal variables, I generate 290 "junk" variables (JUNK1,...,JUNK290) completely independent of GROWTH. Correlating GROWTH with all 300 variables and retaining those that are "significant" at the .10 level typically yields a list of twenty or thirty variables, *most of which are junk*. The calculated  $R^2$  exceeds the variables' true predictive power, many junk variables appear to be more significant than several of the true causes, and some causes never make the list.

11. The typical study combines census data on community demographics with church-record data on membership and giving and survey responses (regarding the activities and attitudes of the members) from a small sample of members and/or leaders. Sometimes a survey is given to all members attending church on a given Sunday, but even this approach greatly oversamples more committed members.

12. The algebraic derivations in the technical appendix confirm and extend the simulation results.

13. Both sets of t-statistics are, of course, extremely high because of the large number of observations and the perfect linearity of the simulation.

14. For the sake of argument, I have not challenged the notion that contextual variables are always "causally prior" to institutional variables. Kelley (1979:338), however, has correctly noted that this is tantamount to assuming that religion is "a dependent variable," incapable of affecting the environment in which it thrives. Certainly one can point to some instances where the order of causation runs from religion to demographics, as when a religion encourages large families, healthy life-styles, habits of thrift, or a commitment to high levels of education.

15. The last two cases probably come closest to capturing the relationships envisioned by most researchers. Although the strength of the causal links are a matter of dispute, everyone

acknowledges the observed (negative) correlation between income and strictness/sectarianism. Moreover, formal theories of strictness/sectarianism derive this correlation as a matter of logical necessity (Iannaccone 1992, Stark and Bainbridge 1987).

16. See Kennedy (1985:69) for a brief discussion of omitted variable bias. See my technical appendix for a mathematical analysis that generalizes the simulation and confirms its results.

17. Hoge printed some of the data in his chapter and listed the rest in a technical appendix (Hoge and Roozen 1979b:E1-E14). See Hoge and Roozen (1979a:19) for details about its availability. Readers can write me for the data in printed, spreadsheet, and Stata formats.

18. The fact that six out of eight institutional attributes collapse to but one dimension is itself strong support for Kelley's thesis and its theoretical cousin, church-sect theory as advanced by Johnson (1963) and Stark and Bainbridge (1985), and Iannaccone (1992, 1994). All these authors approach religious organizations from a *unidimensional* perspective. They order denominations along a single continuum of "strictness, tension, costliness" and explain the social attributes of the religion in terms of its location along the continuum. *A priori*, there is no obvious reason why the institutional attributes listed by Kelley and measured by Hoge should all hang together. That they do hang together is evidence for the logical coherence of Kelley's thesis and modern church-sect theory.

19. Hoge's states most of his results in terms of simple and partial correlations, but in order to maintain consistency with the rest of this paper, I have translated his correlations into their adjusted- $R^2$  equivalents.

20. Recall, however, that this  $R^2$  is the product of a *search* over nine variables. In repeated simulations regressing using the actual growth variable onto nine randomly generated variables, the best two variable model had an adjusted  $R^2$  over .22 and about one-fifth of the regressions yielding adjusted  $R^2$ 's over .40.

21. Hoge (1979:192) recognized this and observed that "the denominational characteristics attributed by Kelley to growing denominations are strongly upheld by independent ratings of experts. The factors Kelley stressed most -- emphasis on evangelism, emphasis on distinctive life-style and morality, and disallowing individualism in belief -- came out strongest." But since the context-only regressions had already explained 59% of the variance in growth, he inferred that strictness could not possibly count for more than the remaining 41 percent.

22. Kelley (1979:338) himself has pointed this out, but the significance of his critique seems to have been overlooked. He notes that the "three contextual factors the researchers found strongly related to denominational growth ... are said to 'explain over half the total variance in denominational growth rates ...' -- *if they are taken first*. If the institutional factors ... are taken first, they explain virtually *all* the total variance, and little is left for contextual factors to explain!"

23. The regressions in table 5 extend Hoge's analysis in a straightforward manner; they do not represent the approach to the data that I would recommend on theoretical and empirical grounds. Given the small number of available observations and the fact that the study was designed to test Kelley's theory, I would suggest approaching the data as follows: First, attempt to extract from the institutional data a single variable or scale that best captures Kelley's notion of "strictness." Fortunately, since all but two of the institutional measures scale along a single dimension, there is no problem deriving such a scale (by simple summation or choice of a single item). Second, regress growth onto this "strictness" scale to see whether it does, in fact, predict growth. (It does, yielding  $R^2$ 's around .90.) Finally, introduce the contextual variables -- one or two at a time -- to see if the initial result, predicted by Kelley's theory remains significant. (As it turns out, strictness not only remains "significant;" it totally overwhelms all other effects, even when those other effects are introduced through a somewhat suspect "search" procedure.)

24. This method can not be applied retroactively to demonstrate the validity of an effect identified in a prior correlations involving all the data.

## References:

- Bangs, Carl. 1972. Deceptive Statistics. *The Christian Century* 89 (August 30): 852-853.
- Bibby, Reginald D. 1978. Why Conservative Churches *Really* are Growing: Kelley Revisited. *Journal for the Scientific Study of Religion* 17(2): 129-137.
- Finke, Roger and Rodney Stark. 1992. *The Churching of America, 1776-1990: Winners and Losers in America's Religious Economy*. New Brunswick, NJ: Rutgers.
- Hadaway, C. Kirk. 1980. Conservatism and Social Strength in a Liberal Denomination. *Review of Religious Research* 21(3): 302-314.
- Hadaway, C. Kirk. 1989. Will the Real Southern Baptists Please Stand Up: Methodological Problems in Surveying Southern Baptist Congregations and Members. *Review of Religious Research* 31(2): 149-161.
- Hoge, Dean R. 1979. *Understanding Church Growth and Decline: 1950 - 1978*. New York: The Pilgrim Press.
- Hoge, Dean R. and David A. Roozen. 1979a. A Test of Denominational Growth and Decline. In *Understanding Church Growth and Decline: 1950-1978*, edited by Dean R. Hoge and David A. Roozen, 179-197. New York: The Pilgrim Press.
- Hoge, Dean R. and David A. Roozen. 1979b. *Technical Appendix to Understanding Church Growth and Decline: 1950-1978*. Hartford, CT: Hartford Seminary Foundation.
- Iannaccone, Laurence R. 1992. Sacrifice and Stigma: Reducing Free-Riding in Cults, Communes, and Other Collectives. *Journal of Political Economy* 100(2): 271-291.

- Iannaccone, Laurence R. 1994. Why Strict Churches are Strong. *American Journal of Sociology* 99(5): 1180-1211.
- Inskeep, Kenneth W. 1993. A Short History of Church Growth Research. In *Church and Denominational Growth*, edited by David Roozen and C. Kirk Hadaway, 135-148. Nashville: Abingdon Press.
- Johnson, Benton. 1963. On Church and Sect. *American Sociological Review* 28:539-49.
- Johnson, Benton. Church and Sect Revisited. 1971. *Journal for the Scientific Study of Religion* 10:124-137.
- Kelley, Dean M. [1972] 1986. *Why Conservative Churches Are Growing: A Study in the Sociology of Religion with a New Preface for the ROSE Edition*. Macon: Mercer University Press.
- Kelley, Dean M. 1979. Is Religion a Dependent Variable? In *Understanding Church Growth and Decline: 1950-1978*, edited by Dean R. Hoge and David A. Roozen, 334-343. New York: The Pilgrim Press.
- Kelley, Dean M. 1978. Why Conservative Churches Are Still Growing. *Journal for the Scientific Study of Religion* 17(2): 165-172.
- Kennedy, Peter. 1985. *A Guide to Econometrics*. Cambridge, MA: The MIT Press.
- McKinney, William J. Jr. 1979. Performance of United Church of Christ Congregations in Massachusetts and in Pennsylvania. In *Understanding Church Growth and Decline: 1950-1978*, edited by Dean R. Hoge and David A. Roozen, 224-247. New York: The Pilgrim Press.

- McKinney, William J. Jr., and Dean R Hoge. 1983. Community and Congregational Factors in the Growth and Decline of Protestant Churches. *Journal for the Scientific Study of Religion* 22:51-66.
- McFaul, Thomas R. 1974. "Strictness" and Church Membership. *The Christian Century* XX (March 13, 1974): 281-284.
- Olson, Daniel V. A. 1993. Congregational Growth and Decline in Indiana Among Five Mainline Denominations. In *Church and Denominational Growth*, edited by David Roozen and C. Kirk Hadaway, 208-224. Nashville: Abingdon Press.
- Perry, Everett L., and Dean R. Hoge. 1981. Faith Priorities of Pastor and Laity as a Factor in the Growth or Decline of Presbyterian Congregations. *Review of Religious Research* 22:21-32.
- Roof, Wade Clark, Dean R. Hoge, John E. Dyble, and C. Kirk Hadaway. 1979. Factors Producing Growth or Decline in United Presbyterian Congregations. In *Understanding Church Growth and Decline: 1950-1978*, edited by Dean R. Hoge and David A. Roozen, 198-223. New York: The Pilgrim Press.
- Roozen, David A., and Jackson W. Carroll. 1989. Methodological Issues in Denominational Surveys of Congregations. *Review of Religious Research* 31(2): 115-131.
- Roozen, David A., and C. Kirk Hadaway, eds. 1993. *Church and Denominational Growth*. Nashville: Abingdon Press.
- Stark, Rodney, and William Sims Bainbridge. 1985. *The Future of Religion*. Berkeley: University of California Press.

Stark, Rodney, and William Sims Bainbridge. 1987. *A Theory of Religion*. New York: Peter Lang.

Stamp, Josiah. 1929. *Some Economic Factors in Modern Life*. London: King and Son.

Thompson, Wayne L., Jackson W. Carroll, and Dean R. Hoge. 1993. Growth or Decline in Presbyterian Congregations. In *Church and Denominational Growth*, edited by David Roozen and C. Kirk Hadaway, 188-207. Nashville: Abingdon Press.

Welch, Michael R. 1993. Participation and Commitment Among American Catholic Parishioners. In *Church and Denominational Growth*, edited by David Roozen and C. Kirk Hadaway, 324-345. Nashville: Abingdon Press.

**TABLE 1:**  
**Growth Regressions in Cross-Denomination**  
**and Single-Denomination Data**

	(1)	(2)	(3)	(4)	(5)	(6)
pop_growth	1.05*** (6.8)		0.96*** (7.4)	0.57*** (5.0)		0.58*** (5.1)
strictness		1.03*** (5.8)	0.93*** (6.5)		0.03 (0.1)	0.22 (0.6)
constant	0.38	-2.58	-2.40	-0.96	-1.26	-1.46
R <sup>2</sup>	.32	.26	.52	.21	.00	.21
Obs	100	100	100	100	100	100

Notes: Asterisks \*, \*\*, and \*\*\* denote significance at the .05 and .01, and .001 levels, respectively. Dependent variable is congregation growth. Regression coefficients are in plain type, t-statistics in parentheses. Data are simulated according to equation (1) in the text. Columns (1) through (3) regress over a random sample of congregations from all denominations; columns (4) through (6) regress over congregations from one denomination.

**TABLE 2:**  
**Growth Regressions with Errors in the Variables**

	(1)	(2)	(3)	(4)	(5)	(6)
pop_growth	.48 (17.39)		.48 (20.65)			
strictness		.48 (17.25)	.48 (20.52)			
pop_growth proxy				.45 (15.98)		.46 (16.97)
strictness proxy					.26 (7.16)	.24 (8.86)
R <sup>2</sup>	.23	.23	.46	.20	.05	.26
Observations	1000	1000	1000	1000	1000	1000

Notes: Data are simulated via  $\text{growth} = \text{pop\_growth} + \text{strictness} + \text{error}$ . Proxy variable definitions are in the text.

**TABLE 3:**  
**Stepwise-hierarchical Regression Results**  
**(Actual and Imputed Contributions to Variance in Growth)**

	model 1	model 2	model 3	model 4	model 5	model 6
<b>Estimated R<sup>2</sup>:</b>						
income only	.32	.30	.15	.52	.24	.61
income + strictness	.65	.30	.30	.63	.32	.66
<b>Imputed R<sup>2</sup>:</b>						
to income	.32	.30	.15	.52	.24	.61
to strictness	.33	.00	.15	.11	.08	.05
<b>Actual R<sup>2</sup>:</b>						
from income	.33	.33	.00	.33+	.00	.33+
from strictness	.33	.00	.33	.33+	.33	.33+
<b>Observations:</b>	1000	1000	1000	1000	1000	1000

Notes: Each model is simulated according to the corresponding path diagram in figure 2. A ".33+" denotes models in which income and strictness are correlated, contribute equally to growth, and jointly account for 66% of the variance in growth. "Estimated R<sup>2</sup>" rows list R<sup>2</sup> for regressions of growth onto income alone and then onto income and strictness combined. "Imputed R<sup>2</sup>" indicates how the regression results are interpreted following a stepwise-hierarchical scheme. "Actual R<sup>2</sup>" gives the true structure of the simulated data.

**TABLE 4:**  
**Simple versus Multiple Regressions for Growth**

	model 1	model 2	model 3	model 4	model 5	model 6
simple regression:						
income only	.56 (21.43)	.55 (20.63)	.39 (13.22)	.72 (32.67)	.49 (17.93)	.78 (39.5)
R <sup>2</sup>	.32	.30	.15	.52	.24	.61
simple regression:						
strictness only	.57 (22.11)	.39 (13.22)	.55 (20.63)	.73 (34.00)	.32 (21.64)	.77 (38.32)
R <sup>2</sup>	.33	.15	.30	.54	.32	.59
multiple regression:						
income	.56 (29.80)	.57 (14.63)	-.03 (8.82)	.41 (15.65)	.07 (1.46)	.45 (13.36)
strictness	.57 (30.44)	-.03 (0.82)	.57 (14.63)	.45 (17.27)	.51 (10.64)	.39 (11.62)
R <sup>2</sup>	.65	.30	.30	.63	.32	.66
observations	1000	1000	1000	1000	1000	1000

Notes: Beta coefficients in plain text; t-statistics in parentheses. Same data as in table 2, simulated from the models depicted in figure 2.

**TABLE 5:  
Denominational Growth Regressions**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
children	.48* (2.73)				-.00 (0.03)	.11 (0.85)	.07 (0.60)
west	.50* (2.81)				-.05 (0.47)	.22 (1.82)	.14 (1.26)
distinctiveness		.97*** (15.10)			1.01*** (7.89)		
strictness1			.91*** (8.65)			.72*** (4.86)	
strictness2				.94*** (10.70)			.81*** (5.71)
adj-R <sup>2</sup>	.59	.94	.83	.88	.93	.85	.88
Obs	16	16	16	16	16	16	16

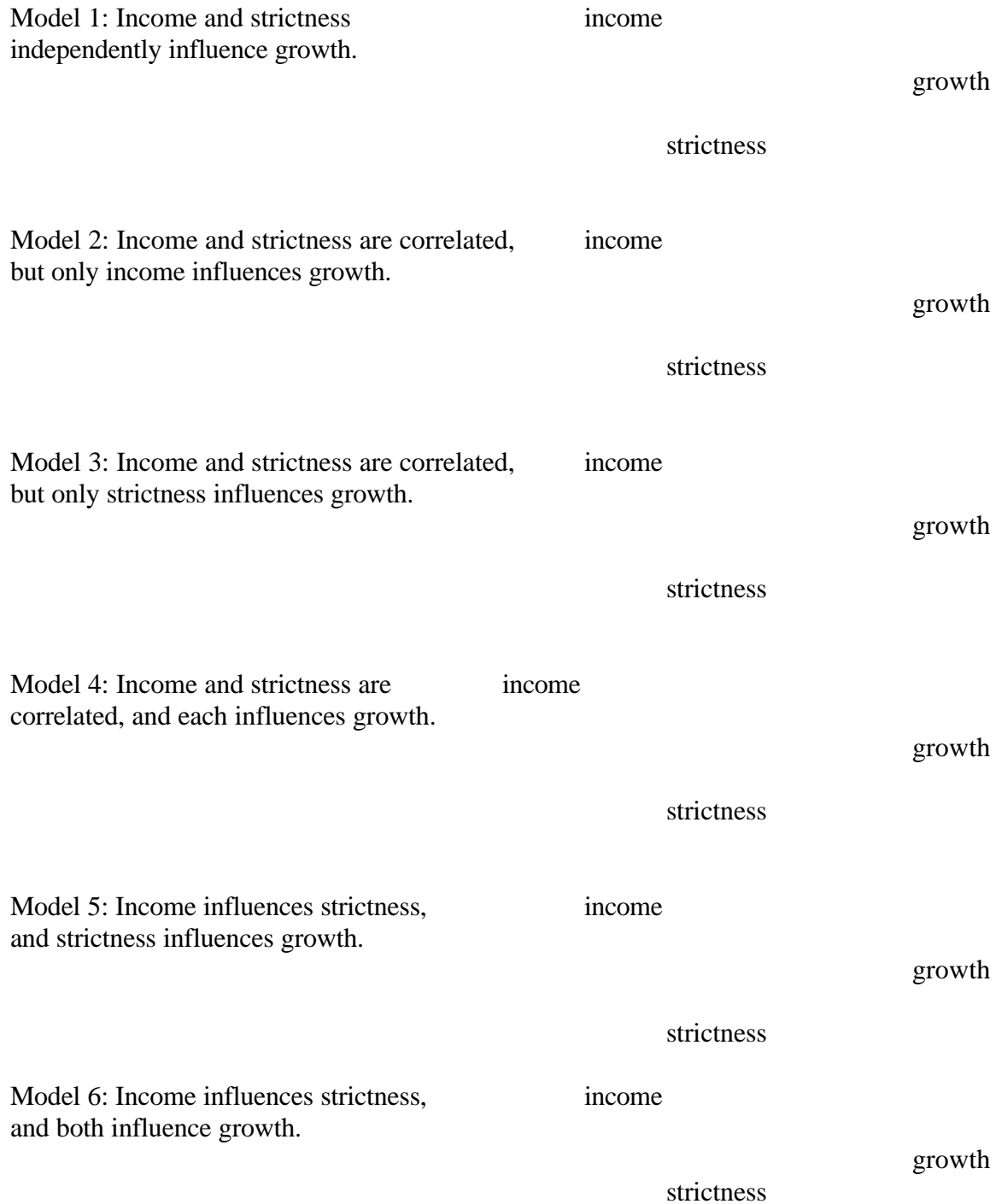
Notes: Asterisks \*, \*\*, and \*\*\* denote significance at the .05, .01, and .001 level, respectively. Beta coefficients in plain text; t-statistics in parentheses. Growth is the dependent variable. Variable definitions: growth = percentage change in denominational membership, 1965-75; west = percentage of the denomination's membership living in western states; children = average number of children in members' families; distinctiveness = expert scale; strictness1 = denomination's mean score for distinctiveness, theology, evangelism, pluralism, social activism, and ecumenism; strictness2 = mean score for distinctiveness, evangelism, and pluralism.

Source: Hoge (1979).

**FIGURE 1:**  
**Growth Versus Strictness**  
**Across a Random Sample of Simulated Congregations**

Note: Circles depict the congregations of a single denomination. Dots depict the congregations of all other denominations.

**FIGURE 2:  
Proposed Models of Growth  
as a Function of Income and Strictness**



**FIGURE 3:  
Growth Versus Strictness in Actual Data**

Source: Hoge (1979: 194)