

Easy R scripts for Two-Stage Least Squares, Instruments, Inferential Statistics and Latent Variables

Douglas R. White, Ren Feng, Giorgio Gosti, and B. Tolga Oztan

Prepared for submission to *Sociological Methodology* annual volume of the ASA* Not for general distribution
2SLS-IS/causality project Shortened 2SLS-isReOrg10a.doc (MPI Leipzig) draft changed Tuesday Jan 10 2012

Abstract. The two-stage least-squares inferential statistics (2SLS-IS) scripts in R provided here help to discriminate the quality of results in regression models, improving on prototypes by Dow (2007) and Eff and Dow (2009) that implemented a new two-stage least squares (2SLS) standard for regression models with missing data imputation and controls for autocorrelation. Major 2SLS-IS modeling improvements with these inferential statistics (-IS) scripts are threefold. **First**, they provide a relative effect (*reff*) measure, analogous to use of percentaged variables in regression modeling (Fox 2002:27) and linear transformation of the regression coefficient that scales uniformly for comparing strengths of each variable alongside *iid* significance tests. **Second**, they evaluate competing variables within regression models by inferential statistics derived from random subsamples of observations to estimate model coefficients that are then used in remaining independent subsamples to test resilience of significance levels. **Third**, they optimize imputation of missing data by an option to impute missing observations on all variables and all cases (Rubin 1996) rather than just those for which the dependent variable is coded (Eff and Dow 2009, Dow 2007). **Fourth**, they link to latent variables and structural equation models (SEM) and will include the hierarchical partitioning of variance widely used for 2SLS (Brown and Eff 2010) and White and Lu's (2010) Hausman tests for model robustness. Each of these features helps to provide inferential statistics to evaluate regression and causal modeling, and to output results useful for analyzing networks of causal, bias, and control. The 2SLS-IS scripts and planned add-ons are modular, currently in three source files, only one of which requires editing by the user to define a new model or to change the database in *.Rdata format.

The advantage of working in the R computing environment, a widely used, cooperatively developed collection of free open-source software, is the flexibility not only of executing programs but writing or incorporating functions that operate on data or output of other scripts or databases. The 2SLS-IS scripts adapted to the Standard Cross-Cultural Sample (SCCS) build on James Dow's (2004) SCCS.Rdata and his editing of missing value codes to conform to R, and prototypes (by Dow 2007, Eff and Malcolm Dow 2009, and Brown and Eff 2010) for missing data imputation and efficient estimates of significance using 2SLS for regression models with autocorrelation controls. Potential collaborators or users and developers can communicate in further open-access program developments. Importing documented packages from the extensive R package archives make it easy to extend these benefits and for new authors to successively improve software.

1 – <i>Introducing regression models with controls for autocorrelation: 2SLS and 2SLS-IS</i>	Outline
1 – 1 The Basis of the 2SLS-IS Scripts	
1 – 2 Why model? Causal Modeling Advances and Structural Models	
2 – <i>The 2SLS-IS scripts</i>	
2 – 1 Causal Modeling Advances with Relative $\delta y/\delta x$ Effects and Variable Inflation Factors (vifs)	
2 – 2 Replication: Comparing Output of Eff and Dow's 2SLS with 2SLS-IS	
2 – 3 Resilience: Using the Random Inferential Subsample Training Ratio	
3 – <i>Robustness tests of prototype models and identification of improved models</i>	
3 – 1 Latent Variables: Moral Gods ("Hidden variables")	
3 – 2 Competition among Models using Significance Tests and Relative $\delta y/\delta x$ Effects	
4 – <i>Networks of Variables, 2SLS-IS Regression, and SEM</i>	
4 – 1 Direct and Indirect Variables, correcting for biased subsamples, and networks of variables	
4 – 2 Structural arrows and SEM: What can we learn about theory from Graphical language?	
4 – 3 Exogeneity and Multiple Imputation	
4 – 4 No Free Lunch: A coupled 2SLS-IS and SEM approach to networks of variables	<i>Appendix 1: Mathematical intuitions</i>
5 – Conclusions	<i>Appendix 2: Inferential Statistics Tables</i>

To do: *relaimpo*, Hal White/Lu Hausman test, option for *indep_vars* not subject to robustness tests, dot graphs that run *sem* models from *EQS*-type commands, *Kyono Commander* in R for *EQS* output, autocorrelation coeffs per *WX* autocorrelation variable), effective sample sizes each independent variables, and use of *scale()* to normalize variables for *probit*. Option to restrict X in *WX* to the *restrict_vars* to default to the full set of independent variables. Use separate *WX* *W*'s from *WYWX.Rdata* to estimate *sem* models (Fox 2006). Do that run with *Fa* *Cmty* fully imputed \hat{E} *WYWX*.

1 - Introducing regression models with controls for autocorrelation: 2SLS and 2SLS-IS

Nonindependence of cases is a foundational problem that applies to all nonexperimental studies and to those with experimental design. Overcoming problems due to autocorrelation was long ago designated as “Galton’s problem” in survey and comparative research, i.e., how to deal with nonindependence of cases. Incorporation of models of network effects can be considered “Galton’s Asset” (Korotayev and de Munck 2003) in dealing with many different problems that confound empirical research and require understanding of the effects of network interactions. One of the problems is the most misused of descriptive statistics: tests of significance (Henkel and Morrison 1970). Significance levels for statistical decisions are invalid without correction for endogeneity, i.e., correlations between the error term and the independent variables that arise from networks of interaction within the sample of observations. Two-stage least squares (2SLS) regression is a method of choice for attainment of exogeneity, with a first stage OLS regression that provides Instruments to measure endogeneity and a second stage that tests whether the Instruments designed to control for autocorrelation, and measured in the first stage, provide exogeneity when incorporated as Instrumental Variables in the second. This provides results on network influences among the cases in the sample, and measures effects of variables on one another if tests for controls for external network influences give evidence of exogeneity. Some researchers consider studies that include network effects to improve the quality of understanding of interdependence in network-linked social processes to be “Galton’s Opportunity” (Witkowski 1974) to understand how network effects of nonindependence combine with exogenous independent variables. Still, even with 2SLS models that attain exogeneity of independent variables, overconfident rejection of null hypothesis of significance tests that are efficiently estimated only on the condition of exogeneity may still not be adequate for deciding which variables to include in a regression model. We incorporate in our scripts alternative criteria (relative $\delta y/\delta x$ effects and Hausman tests of robustness, defined as persistence of a model’s characteristics under perturbation or random error) to use alongside significance tests in models that predict the dependent variables. Examining networks of regression-related variables leads to the possibility of exploring direct and indirect or mediated effects and measurement biases that link back, through SEM

1 – 1 The Basis of the 2SLS-IS Scripts

We provide here R scripts for two-stage least-squares (2SLS) inferential statistics (2SLS-IS) that add, to a basic 2SLS software prototype designed to control for autocorrelation in regression models (Dow 2007, Eff and Dow 2009), tests of robustness and resilience –inferential statistics derived from random subsampling of data to identify sets of variables with persistent characteristics in a regression model. Prototype scripts include multiple (independent) imputation of missing data (Rubin 1987, Little and Rubin 2010); “multiple” refers here to computing eight or more copies of the variables to be used in a regression, each with independent probabilistic estimates of what the missing values are likely to be given reliable and completely coded auxiliary variables, the “vaux.Rdata” described by Eff and Dow (2009), using the R ‘mice’ package for Multivariate Imputation by Chained Equations (van Buuren and Groothuis-Oudshoorn 2011) that computes regression statistics independently for each imputed dataset and then uses Rubin’s (1987) formulas to give combined estimates for coefficients, significance, and other statistics. The 2SLS prototype scripts of Eff and Dow also include tests for omitted variables among the potential candidates for inclusion in a model (Wald 1943, Wooldridge 2006:587) and tests for endogeneity (correlations of independent variables with regression errors) that invalidate accurate estimation of effects and their statistical significance (Hausman 1978, Wooldridge 2006:532–533, White and Lu 2010). We also go beyond Eff and Dow to allow a choice of whether to impute missing data for both dependent and independent variables for all cases in the full sample, or just for the independent variables. The former is normally done with survey data (Rubin 1996) and has a key advantage: It allows us to consider joint effects in networks of variables, variables that measure bias, and to test for effects of missing data. The ability to model networks of comparable variables also allows us to use transformations of unstandardized regression coefficients that are normalized for comparable measurement of effects. Relative $\delta y/\delta x$ effects (= regression coefficient/dep_var \times indep_var ranges) for independent and dependent pairs in a network of variables, derived from models of comparable effects in networks of variables developed by Pearl (2001, 2011), complement statistical significance and robustness measures from inferential statistics.

Aspects of our 2SLS-IS scripts are described in five parts. Part 1 (What is the basis of the software?) describes the prototype for the new scripts and ways they can be used. Part 2 (What are its practical features?) presents each script in turn. Part 3 - 1 (Does the model hold for random samples of the data? Do prior 2SLS models pass these tests?) validates the new scripts by rough equivalence of results with Eff and Dow (2009) and Brown and Eff (2010) and shows how the new software generates statistical distributions to supplement significance tests in evaluating models. These tests save results of fitting large random training subsamples, and using the holdout data to test for robustness of significance tests.

Part 3 - 2 (How do 2SLS-IS robustness tests help to improve on 2SLS models?) exemplifies questions raised by robustness tests. Are the model coefficients from the training sample predictive of similar results such as R^2 , significance, and other “holdout test” subsample results for some variables but not others? Are there weaknesses that might be improved? Part 3 - 3 (How might alternatives be investigated?) exemplifies how weaker variables, e.g., in the Brown and Eff (2010) model might be improved and tested in an alternative and the use of relative $\delta y/\delta x$ effects. Part 4 (Can we model results as causal graphs including multiple variables including measures of potential biases?) goes beyond these extensions of Eff and Dow to address logical problems and solution concepts for networks-of-variables models including bias controls, adjustment for common causes, and problems involved in combining regressions for single dependent variables into networks of variables suitable for consideration of causal graphs. Parts 5 -1 and 5 -2 extend the 2SLS-IS approach to networks of variables and conclude with summaries of the advantages of 2SLS-IS and further developments.

Difficulties in solving for network autocorrelation effects given nonindependence of cases –a problem in regression analysis that creates spurious correlations and tests of their significance– were studied by Dow, White and Burton (1982), Dow, Burton, White, and Reitz (1984), and Dow, Burton and White (1982). Dow (2007) found the solutions for these problems in cross-cultural research. Problems that remain with 2SLS include how to select, evaluate and justify variables chosen (Greenland 2010) and avoid overfitting a model with variables that are spurious but may happen to enhance significance or raise R^2 without passing tests of robustness. To avoid these problems it is important to have not only tests of robustness but prior theory, prior evidence, alternative samples, a deductive argument to connect theory to choices of variables, an acute sense of sources of bias (Dow 2007), and inferential statistics that avoid reliance on significance tests. This makes any study that relies on correlations, significance, OLS, or 2SLS regression alone problematic. 2SLS-IS models (with inferential statistics) help to deal with these problems (Greenland 2010) by providing a level of overall model evaluation that is not commonly used in regression and 2SLS regression modeling. Pg1Bib

The features of Eff and Dow (2009) and the earlier advances on which the new 2SLS-IS R scripts is built are encompassed within the new software, as summarized by Eff and Rionero (2011):

“We follow the methodology developed for cross-cultural data sets by Dow (2007), Dow and Eff (2009a, 2009b), Eff and Dow (2008, 2009), and Eff (2008). Multiple imputation addresses the problem of missing data. Weight matrices for geographical and linguistic proximity are used to model the influence of cultural transmission via borrowing and inheritance, respectively. The geographical distance weight matrix is created by calculating the great circle distance between centroids for each pair of countries, and the language phylogeny weight matrix is taken from Eff (2008). Since the two weight matrices are highly correlated, it is difficult to identify separate borrowing and ancestry effects. We accomplish this by employing the composite weight matrix method presented in Dow and Eff (2009a:142), and also used in Brown and Eff (2010): combining the two weight matrices, we select the combined matrix that results in highest model R^2 . Our model takes the form:

$$y = \rho Wy + X\beta + \varepsilon \tag{1}$$

In equation 1, y is the dependent variable, and Wy is the composite weight matrix times the dependent variable, giving us a measure of cultural transmission. Wy is endogenous, which requires that the model be estimated using two-stage least squares (Dow 2007).”

If this model is well specified with appropriate variables X and network measures W for linkages among the cases in the sample, it should pass statistical tests for the possibility that each independent variable in X and dependent variable Wy (estimated from X) and in the model will be independent of the error term ε in the regression results, i.e., *exogenous* (uncorrelated) with respect to ε . Only then are significance tests of the null hypothesis for independent variables correctly specified because if the prediction errors in ε can be assumed to be independently and identically distributed (*iid*) then they do not bias the measures of effect on the dependent variable or their significance. Conversely, if a variable in regression is *endogenous*, i.e., correlated with the error term ε , then its effect on the dependent variable and its significance cannot be correctly estimated. Solution of the $y = Wy + X + \varepsilon$ model in (1) cannot be solved directly using OLS because autocorrelation can be present either in the Wy or X terms, or both, and there is nothing to align their effect coefficients so that each set of Wy or X terms are exogenous. A first stage OLS regression, however, does suffice to solve $y = Wy + \varepsilon_w$, and allows a test of whether ε_w is uncorrelated with Wy (ε_w uncorrelated with the matrix product $y \% \% W$), giving a fixed estimate \hat{w}_y , after dropping the error term. Then a second stage of OLS regression suffices to solve $y = \hat{w}_y + X + \varepsilon$ and to allow an estimate of whether ε is uncorrelated with \hat{y} and each of the X variables. Whether the variables in these equations are exogenous must be tested *post hoc* using a LaGrange Multiplier test for network dependence in the residuals (H_0 : no autocorrelation of ε with any of the W matrix products in the first-stage regression) and Hausman tests, in Dawid (1979) notation, of $X_i \perp \varepsilon$ and $\hat{y} \perp \varepsilon$, meaning a null hypothesis H_0 : for no autocorrelation of ε with any of the independent variables (Wooldridge 2006:528-532) in the second stage regression. Whether these results are achieved depends on choice of an appropriate model of $y = Wy + \varepsilon_w$ in first stage regression and choice of W matrices that measure and control for the

network effects of interdependencies among the observations. Thus, to separate the effects of independent variables from the dependent variable, Dow (2007) and Eff and Dow (2009) determined that the best and most viable model for 2SLS models in cross-cultural studies uses a first-stage regression of $y = Wy + \varepsilon_w = WX + \varepsilon'_w$. The result is a “neighborhood weighted average” of each first-stage independent variable for each observation where the weights are the strength of direct links between neighbors. A variant of the 2SLS model takes weighted indirect effects into account, and is easily testable by including compound matrix products of the W matrixes such as $W \%*\% W$, i.e., second or third order autocorrelation (user options for $W \%*\% W \%*\% W$ matrix products are single operations in R). Appendix 1 gives further mathematical detail on 2SLS and the 2SLS-IS model given in equation (1) and describes weaknesses of several alternatives.

Still, reliance on significance tests to choose variables in a final regression model is not sufficient for a well-specified model and is a defect of the Eff and Dow (2009) 2SLS methodology. OLS and 2SLS regressions give a single result and not a family of likelihoods for gauging robustness as to which sets of variables replicate better than others. We will show by example that 2SLS-IS tests of robustness can offer improvements in model specification. To do so we estimate a model for a given set of independent variables on large random sample of the full dataset, and then analyze replicability of each variable in the model by comparing estimates based on fixing the effect coefficients found for each larger random sample and comparing the performance of independent variables for the unused portions of the remaining independent samples according to their effects and significance relative to other variables and as measured by a comparable measure of relative $\delta y/\delta x$ effects discussed above and defined in section 2.1. The rank order strengths of these relative $\delta y/\delta x$ effects differ from those of statistical significance and provide an alternative criterion for evaluating which variables should be included in a model relative to excluded variables.

1 – 2 Why model? Causal Modeling Advances and Structural Models

Economists use more theoretically motivated forms of regression to evaluate the effect of “interventions” as deliberate policy actions and incentives where the modeling problem is one of establishing the causes or drivers of change and effects on change. Hurwitz (1962) explained the concept of “structural models” in econometrics that identify how interventions create changes in parameters, equations, and observable or unobservable variables that are elements of the model. These models aim at accurate characterizations of the effect of an intervention on the dependent variable after the intervention. Some variables in the model are causal and others are not, holding fixed other variables that serve as controls. Wooldridge (2006:521; Chapter 16) concurs that a *structural equation* is one supposed to contain a measure of one or more causal relationships and may contain both exogenous and endogenous variables as well as controls. The Lucas (1976) critique is that, holding other equations fixed, changing only a policy equation to predict the effects of a policy change will fail because the other equations will change when the policy changes, i.e., due to response behavior to the policy and its effects. Structural equation models (SEM), with all of their problems and possibilities, date back to path analysis (Wright 1921, 1923, 1934), which uses the solution of simultaneous equations or complex computations of partial correlations. R packages for SEM and latent variables are *sem* (Fox 2006, 2009, Fox and Byrnes 2011, Grace 2009) and *lavaan* (Rosseel 2011a, 2011b), the latter aiming at providing all the state-of-the-art capabilities that are currently available in commercial packages like LISREL, EQS, Amos, and Mplus as well as graphic languages to not only to visualize but to create the commands for a SEM model (as was once the case with the dot language). In time, R packages for SEM will link to programs like *Commentator* (Kyono 2010), currently designed to work only with EQS, that “detect and list: (i) identifiable parameters, (ii) identifiable total effects, (iii) instrumental variables, (iv) minimal sets of covariates necessary for estimating causal effects, and (v) statistical tests to ensure the compatibility of the model with the data. These lists assist SEM practitioners in deciding what test to run, what variables to measure, what claims can be derived from the data, and how to modify models that fail the tests.”

Pg3bib2

Econometricians use regression methods in ways that have benefited from the lessons of SEM, causal and structural modeling (Pearl 2009:27) to distinguish and test claims (White and Lu 2010:4) of the form:

“an outcome or response of interest, Y , is structurally generated as

$$Y = r(D, Z, U) \tag{3}$$

where r is an unknown structural function, D represents observed causes of interest, and Z and U are other drivers of Y , where Z is observed and U is not. U represents not just ‘shocks,’ but all factors driving Y that are too costly or too difficult to observe precisely. So...”

If the effects of the drivers are linear, a different form of regression (White and Lu 2010:4) will apply, then

$$Y = D^t\beta + Z^t\alpha + U \tag{4}$$

“where β and α represent the effects of D and Z on Y , respectively” (and symbol t is the transpose).

Here, D includes the primary causes of interest, and equation (4) makes possible efficient estimation (“identification”) of the β parameters of interest with only a conditional form of exogeneity where, given $X = (Z^t, W^t)$, D is uncorrelated with U ,

i.e., ; and exogeneity need not apply to Z (Z often consists of control variables). Here (White and Lu 2010:31) the W^t can include proxies for U (which include the W matrix controls for autocorrelation of such as those equation 1'), proxies for unobserved drivers of D , or unobserved drivers of D . What this means in practice for 2SLS-IS and its extensions to SEM is that the Hausman (1978) test that diagnoses misspecification, where an independent variable is endogenously correlated with the error term, such as ϵ'' in equation 1' of Appendix 1, need only be applied to the core D of observed causes of interest, and not to variables that are included in the model (e.g., as controls) but have no causal content. In statistical theory this means wider latitude to omit tests that could disqualify some causal models because non-core variables are not exogenous. In testing substantive theory by model estimation this entails much more careful consideration of what variables are justified as causal core variables versus controls (see Greenland 2010), and what variables can be included in a structural model identified by regression methods with the Hausman test of misspecification limited to the core variables. Following after Brown and Eff (2010), where a standard Hausman test (Wooldridge 2006:532-533) is used uniformly for every independent variable in a regression model, White and Lu's (2010) new Hausman test of robustness for critical core regression model coefficients that apply to potentially causal variables. If we successfully implement this test –which merited consideration for a 2011 Nobel prize– it would mark another advance in development in R of 2SLS and our 2SLS-IS scripts, as with others that adopt it.

Pg4Bib

2 - The 2SLS-IS scripts

The 2SLS-IS scripts retain the objectives of Eff and Dow (2009) as a prototype for controlling for autocorrelation (see White 1993, 2007 for spatial and linguistic autocorrelation in cross-cultural data), for imputing missing data, and further generalization for use with survey data in which observations are almost always nonindependent and require instrumental variables or controls to measure and remove endogeneity. The 2SLS-IS scripts are also aimed at (1) new inferential “train-test” statistics fixing coefficients with a large random subsample to test, in the smaller remainder of the sample, resilience of significance tests, (2) inclusion of regression models with temporal data, and (3) increasing the potential for generalization of structural equation models (SEM), including dynamics and simulations. To that end, the 2SLS-IS scripts can be integrated with functions from SEM freeware in R (Fox 2006, Fox and Byrnes 2011, Rosseel 2011a), and models of networks of variables discussed in section 4.

The 2SLS-IS scripts are sourced with the following R commands that re-factor the Eff and Dow R scripts into larger linked modules to make the new code easier to use and modify. The user is expected to change only the 2nd source code, set the dependent variable and run the model for all their independent variables, and then narrow down to a smaller set of restricted variables that are effective predictors and may also be found to be resilient to perturbation. That may involve experimenting with the 1st, and variants of the 4th source code with the full sample or a “random subsample” test/retest percentage (50, 70, 79) for testing resilience of results. The iterated loop for the 4th and 5th source code is optional, and saves inferential statistics that are not part of the Eff and Dow (2009) program. The use of remotely sourced scripts means that the students can load the R libraries and the data used for collaborative teaching using remote access as described at http://intersci.ss.uci.edu/wiki/index.php/Scs_2SLS-IS_downloads.

Table 1: The R scripts for 2SLS-IS

```

setwd('/Users/drwhite/Documents/3spls/scs/')           #setup working directory
source('Libraries2SLS.R')                             #load R libraries
load('scs.Rdata')                                     #read scs or other data
load(url('http://intersci.ss.uci.edu/wiki/R/source/scs.Rdata')) #remote source
source("examples/create/autocorr/depvar_wghts.R")     #1st source code (optional re: stage-one regression)
source("examples/create/create_EduR_1/create_EduR_1DistMGdHV7.R$5") #2nd source code (changed in user's model)
source('http://intersci.ss.uci.edu/wiki/R/create_EduR_1DistMGdHV7$5.R') #2 remote source
source("R_3_s_ols/two_stage_ols.R")                  #3rd source code main functions; saves WYhat.Rdata
source("R_3_s_ols/two_stage_ols_full.R")             #3rd source code main functions; saves WYhat.Rdata
source('http://intersci.ss.uci.edu/wiki/R/two_stage_ols.R') #3 remote source
source('http://intersci.ss.uci.edu/wiki/R/two_stage_ols_full.R') #3 remote source (option full imputation)
for (i=1:1) {                                         # for (i=1:10) full run
source("examples/create/R_run/run_model_70.R")       #4th source code, "random subsample" = 70
source('http://intersci.ss.uci.edu/wiki/R/run_model_79.R') #4 remote source (test-retest subsamples)
source('http://intersci.ss.uci.edu/wiki/R/run_model.R') #4 remote source (full sample)
#source("examples/stats/inferential_stats.R")        } #5th source code, computes and saves results
source("examples/create/R_run/averageAll.R")         #6th source code, save imputed data for OLS

```

In the sequence of scripts of Table 1, the commands: set the working directory (`setwd`), which is illustrated for a Macbook but could be a shared directory in a classroom computer lab; and install the R scripts and data as described in White (2012). To change the two-stage model, including controls for autocorrelation, the user creates or changes only the contents of the 2nd source file (e.g., `create_EduR_1DistMGdHV7$5.R`). These entries define: 1) the dependent variable of a model and which of the autocorrelation variables are to be included in the model (these square W matrices

with zeros in the diagonals are row-normalized so that row entries sum to one¹, 2) a name and variable for the dependent variable, 3) a frame, labeled as in `my_sccs` after the database name, that contains user-labeled variables from the database that may be retrieved from an `.Rdata` file, 4) a frame of `indep_vars` as potential and some excess variables for the model; and 5) a frame of `restrict_vars` that will be successively adjusted by the user for goodness-of-fit. A useful feature is that after the first (#1st) source code is run, the variables selected for analysis are accessible from `my_sccs`, and can be further analyzed (White 2012), e.g., by OLS (`results=lm(dep_var ~ {list of variable names, in quotes, comma separated})`). Within any of the `create_...R` (#2nd) source codes, the `scale()` function is available in R to convert ordinal to normalized distributions, creating the possibility for a Probit analysis of normalized data through the same 2SLS-IS scripts. Missing data are imputed and the variables with imputed values for missing data are available after the 6th source code (`average.R/averageAll.R`) is run, as explained in White (2012).

Analytical steps are programmed in the 3rd-6th source programs. The 2nd source code has specified which `W` matrices to use to measure nonindependence among cases in the sample, defining one or more Instrumental Variables as controls for autocorrelation. The third source code, `two_stage_ols.R` (2SLS) is prepared to fit but not necessarily to optimize the model given inputs for how the model is created and how it is run. It is this program that creates matrix products of the independent variable row vectors (size n) by the chosen (or tested, each in turn) row-normalized n by n `W` matrices to define new row vectors (also size n) whose products with the `X` variables (“`WX`”) are regressed against the “`Wy`” autocorrelation neighborhoods (Appendix 1 equation 2). As shown in Appendix 1, the coefficients of nonindependence, e.g., for language and/or spatial proximity, estimate the “network effects” by which cases affect one another; once these are fixed the correlation with the dependent variable y is maximized by estimating the α_{ji} parameters, one for each `WX` matrix product in equation 1’ of the appendix. The `two_stage_ols_full.R` script is a variant of the third script. Instead of imputing only the independent variables, for those cases where the dependent variable is defined, the `..._full.R` script also imputes missing data for the dependent variable and for all cases of the independent variables. It is essential to use `..._full.R` when doing networks of variables. The fourth source code, `run_model.R`, does missing data imputation, and produces the results. These exceed the results of ordinary regression and 2SLS, and allow choice of a parameter for a 100% sample or 79% of a randomized subsample percentage for a model “train test,” e.g., `run_model_79.R` in place of `run_model_100.R`. The FOR loop for multiple runs of each `run_model...R` (e.g., 10 iterations), is a computational loop that would save results in a new file (currently done by hand by saving results of ten separate runs). The fifth source code, `inferential_stats.R`, is intended for summarization the repeated results of randomized train-test subsamples (if not 100%), summing the distributions of inferential statistics for the independent variables in the train and holdover test samples; and may be further scripted for analysis of how variables differ in the resilience of their significance or effect ratios relative to expected fluctuations due to subsample sizes. The fourth source code, that is, has the potential to provide inferential statistics on individual variables, while the fifth provides inferential statistics overall, e.g., from 10 iterations of the program. Practical advice for runs of the program is provided in (White 2012). The sixth source code, `averageAll.R`, averages the multiple independent columns of imputed data, both numeric and ordinal for the full sample ($N=186$ for the SCCS). Following execution of the 2SLS-IS scripts these averages of imputed variables (White 2012) also become available for OLS regression (using the R `Lm` package for linear models), without controls for autocorrelation, or can be reanalyzed by the 2SLS-IS scripts.

The purpose of the FOR loop for source codes four and five of the 2SLS-IS scripts (`run_model...R`, `inferential_stats.R`) is to save several kinds of inferential statistics for model evaluation, including the “train-test” inferential statistics commonly used by computer scientists, for example, in split-halves model testing, given parameters computed from one sample, as to how well R^2 performs in an independent sample. The key parameter of `run_model...R`, as in `run_model_50.R` or `run_model_79.R`, sets a ratio for choice of a subsample (e.g., from 50%-79%) that is randomly selected for estimating coefficients in both the first and second stage of the 2SLS regressions and observing robustness of significance tests in the retest samples. A minimum of six iterations of the `run_model` analysis (each taking about 40 seconds, depending on non-zero density of the `W` matrices) might be used, with each run (e.g., seven minutes run-time for ten iterations) selecting different random samples. These iterations generate distributions of four key parameters used for inferential statistics of the total model. These are: 1) relative $\delta y/\delta x$ effects and vif-variable inflation factors, 2) raw and finished (2nd-stage) correlations between independent variables and the dependant variable and R^2 for the first and second stage regressions, and 3) p-values. Of these, correlations and relative $\delta y/\delta x$ effects are dimensionless

¹ The user may edit an auxiliary (optional) source code (`depvar_wghts.R`), following a script of Brown and Eff (2010) that optimizes weightings of multiple `W` matrices used in the first stage of regression (see Appendix 1). This script calculates a fixed Instrument that measures autocorrelations from the independent variable neighborhoods of each observation to be added to the set of independent variables regressed on the dependent variable y in the second stage of regression. The first stage regresses “`Wy`” = $\sum_i W_i y = \sum_i \sum_j p_{ij} W_{ix_j} + \epsilon$ for multiple autocorrelation `W` matrices to take the estimated value “`Wy`” as the Instrument measuring total autocorrelation.

and potentially comparable across different studies (for effect sizes see Nakagawa and Cuthill 2007) while significance tests are not, as they depend on sample size and univariate distributions. Multiple measures of relative $\delta y/\delta x$ effects and vifs avoid overreliance on significance tests. The following section describes our use of relative $\delta y/\delta x$ effects and how they differ from conventional measures of effect sizes, which are also dimensionless. Pg5Bib

2 - 1 Causal Modeling Advances with $\delta y/\delta x$ Regression and Variable Inflation Factors (vifs)

Fox (2002:27) makes good use of percentages in his illustration of Duncan’s (1961) Occupational-Prestige regression, noting for example that “holding education constant, a 1 percent increase in high-income earners is associated on average with an increase of about 0.6 percent in high prestige ratings” (0.6 being the unstandardized regression coefficient for income). This is a “unit change” or *relative effect (reff)* obtained by standardizing the ranges of the independent and dependent variables. The 2SLS-IS software gives similar relative $\delta y/\delta x$ effects (*reffs*) for each independent variable. They are computed by dividing each independent variable’s unstandardized regression coefficient (Coef) by its numeric range (IVrange) and multiplying in each case by the dependent variable numeric range (DVrange), i.e., $reff = \text{Coef} * \text{IVrange} / \text{DVrange}$. These estimate *elasticities*, i.e., $\delta y/\delta x$, the ratio of a unit change in an independent variable x to its effect on the dependent variable y . They can be used for suitable ordinal, normalized, logged or exponentially transformed variables, and set to 1 for dichotomies. The idea is to unitize the independent and dependent variables, as with percentages. An option is to eliminate outliers in each IVrange and DVrange prior to calculating $\delta y/\delta x$.

Independent variables in a regression model need not be statistically independent of one another: the extent of multicollinearity is measured in regression by the variable inflation factor (vif) for each variable. The square root of vif measures how much larger is the standard error of the variable compared with what it would be if that variable were uncorrelated with other independent variables in the regression. Hence division of the relative $\delta y/\delta x$ effect by the vif gives an *adjusted relative effect* ($reff/vif$) measure for which the sum of absolute values may be less than 1, within error bounds. In regression, the absolute value of the $\delta y/\delta x$ ratios, for a robust and highly predictive model, will often sum to one (and tend to overshoot with high vifs). The *reff* thus serves as an alternative measure of whether an independent variable should be included in a regression model. A relative effect may be high relative to other independent variables that are more significant, or vice versa. In a structural model of a network of variables, relative $\delta y/\delta x$ effects are easier to interpret than unstandardized regression coefficients (i.e., they represent what economists call elasticities). The same is true of Pearl’s (2009:150,367-368) purely conceptual *do* operator. The two concepts are conceptually orthogonal, however, the latter applicable at the level of analyzing a causal graph as to whether its structure is consistent with causal inference.

Table 2 shows for one out of 10 runs of the Brown-Eff (2010) moral gods model (left panel) in 2SLS-IS that the correlation between *reff* or $\delta y/\delta x$ and significance (p-value) for the independent variables is $R^2=0.42$, compared to $R^2=0.65$ for an alternate 2SLS-IS model ($R^2=.71$ for $reff/vif$; and $R^2=0.65$ if the control variable, logdate, for date of observation, is included). Higher correlation for the alternate model might indicate higher model reliability. Whenever $\delta y/\delta x$ has a causal interpretation, so does the elasticity. If the $reff = \delta y/\delta x = 0.189$ then a 100% change in this independent variable predicts a 19% change in the dependent variable.² In Table 2 external war is the most significant variable in the Brown-Eff model while fifth in $reff/vif$ while the relative ranks of $reff/vif$ and p-value for the alternate model is nearly identical.

Table 2: Relative effect (reff) measure compared to significance tests for the Brown-Eff moral gods regression

create_EduR_1.5DistBrownEffSign.R*				HiddenVars create_EduR_1DistMGdHV7\$5.R (White et al. 2011)*			
Variable name	reff	reff/vif	p-value	Variable name	reff	reff/vif	p-value
+PCAP	0.293	0.257	0.089	+AnimXbwealth	0.237	0.177	0.034
+Anim	0.253	0.160	0.048	+FxCmtyWages	0.104	0.079	0.057
(-)PCsize	0.280	0.182	0.020	+SuperjrhWriting	0.172	0.123	0.047
-PCsize.2	-0.354	0.281	0.026	+No_rain_Dry	0.215	0.164	0.006
+Caststrat	0.215	0.191	0.040	+Missions	0.101	0.089	0.040
-Eeextwar	-0.179	0.164	0.003	Control variable:	(0.829)	(0.633)	= Partial sums
+Foodscarc	0.124	0.078	0.140	+logdate	(0.784)	0.700	0.010
Model $R^2=.355$	1.405	1.313	= Sums	Model $R^2=.411$	1.613	1.333	= Sums
	$\delta y/\delta x$	r/v	$p\text{-val } R^2=.42$		$\delta y/\delta x$	r/v	$p\text{-val } R^2=.71; .65$

*Full models in Table 5 lack the reff/vif ratios.

2 - 2 Replication: Comparing Output of Eff and Dow’s 2SLS with 2SLS-IS

² For example, $reff = \delta y/\delta x = 0.189$ if the range of the independent variable is $IVrange=9$, that of the dependent variable is $DVrange=4$, the regression coefficient Coef is .084 and $reff = \text{Coef} * \text{IVrange} / \text{DVrange} = (0.084 * 9.000) / 4 = 0.189$.

Reproducibility is the ability of a study or experiment to be independently replicated, a basic principle of the scientific method. In statistics this means repeating with the same or a different method the estimated variability of a phenomenon. *Replication* is the repetition of an experimental condition so that the variability associated with the phenomenon can be estimated, including drawing of samples and controls for endogeneity such experimental or observational context or control variables that may correlate with the error term in regression because the observations are not independent. Random sampling has no effect on the problems of endogeneity. Given that our 2SLS-IS software uses a prototype for these problems programmed in R by Eff and Dow (2009), the user can test replication of the Eff-Dow results with the same variables and controls for endogeneity. Stricter tests are done in each case by the same dependent, independent variables and restricted variables with no missing data (to avoid the small fluctuations in results due to estimation of missing data) and the same Instrumental Variables (IVs). Because our WY, control variables for endogeneity, are estimated from WX autocorrelations in the independent variables (the Ws being network autocorrelation matrices), we are currently studying whether to provide a choice of estimation only from the smaller set restricted variables for a final regression model while retaining a larger set of independent variables. Eff and Dow (2009) compute WX for the larger set of independent variables without allowing the user the more restrictive option that targets controls for endogeneity only for the precise variables in the model. Careful attention to alternatives such as these can help to achieve replication. Eff and Dow's method for reliability in probabilistic estimating missing data (R package *mice*) depends on a set of fully coded auxiliary variables (vaux.Rdata) that remain in the background but must be replaced by the user when constructing a new database for 2SLS regression. We also test replication with model results of Brown and Eff (2010) for a different set of variables.

In the achievement of replication, reproducibility and validation of results, we consider next the issues of perturbation, such as the stability or resilience of regression results varying the sample or random subsamples, and those of validation, such as using networks of variables or multiple regression models for linked variables, which may require construction of latent variables and that consider direct and indirect effects. Here, the problems include whether we can find better measures of our variables. Table 3 offers a useful comparison to results of Table 2, both of which consider two different models for the same dependent variable, first separately (Table 2), but now in competition (Table 3). Both models have approximately the same R². A very important point, however, is that the quality of a model is not measured by R², nor the significance of the individual variables. To illustrate, at the left and right of Table 3 are columns with *reff* effect ratios (*dy/dx* elasticities) and significance tests, along with R² for two distinct models predicting the occurrence of high gods (Swanson 1960) in the SCCS from various other variables in that database. The middle two sets of columns show results that enter variables from two models: those at the right a "Hidden variables" model (White et al. 2012) and at left Brown and Eff (2010). Test results show that in competition for variance and significance in a single model, none of the Brown and Eff independent variables are significant, while all of the "Hidden variables" but one (AnimXbwealth) are significant (*p* < .10).

Table 3: Evaluation of two competing models predicting High Gods (N=186) as coded by Snarey (1996)

```
source("examples/create/create_EduR_1/create_EduR_1.5DistBrownEffSign.R") #Brown-Eff Solo Model
source("examples/create/create_EduR_1/create_EduR_1.5DistB_EffReduce_2Models.R") #Mixed Model
source("examples/create/create_EduR_1DistMGdHV7$52.R") HiddenVars
```

Solo Effect Ratio	R ² =.46 R ² =.48* Signif.	Mixed Effect Ratio	R ² =NA Signif.	Brown & Eff Variables	White-Snarey "Hidden Variables"	R ² =.43 Signif.	Mixed Effect Ratio	Solo Effect Ratio	R ² .41 Signif.
0.635	.00000	0.576	.00000	Distance	Distance	.00000***	0.576	0.534	.0000002
.239	.089(.012)	0.156	.303	(-)PCAP					Omitted
.280	.020(.038)	0.005	.975	(-)PCsize	SuperjhWriting	.092 *	0.279	0.172	.047 **
-.354	.026(.011)	-.216	.261	PCsize.2					Omitted
-.179	.003(.025)	-.072	.244	Eextwar					Omitted
0.253	.048(.113)	-.005	.975	Anim	AnimXbwealth	.250	0.186	0.237	.034 **
0.215	.040(.034)	0.099	.384	Caststrat	FxCmtyWages	.080 *	0.106	0.104	.057 *
0.124	.139(.016)	0.036	.664	Foodscarc	No_rain_Dry	.053 *	0.174	0.215	.006 ***
	Omitted				Missions	.097 *	0.085	0.101	.040 **
1.634		0.589	RATIO SUMS	RATIO SUMS	RATIO SUMS		0.830	0.829	
(overfitted?)					Logdate	.081 *	0.593	0.784	.010 ***

*Brown and Eff (2010) R² components are adjusted with hierarchical partitioning but total R² is constant.

** Nonsignificance due to a missing data bias: N=132 using v238, N=144 using Snarey's HiGod4 (N=186). The significance tests in parentheses are those from Brown and Eff. *** Relative effect. ***** Nonsignificance due to additional Hidden Variables.

A key difference between these two models is that the Brown-Eff relies on significance tests to choose variables, ignoring the possibility for direct calculation of the dy/dx effects (as used by Eff and Routon 2012), and then testing the model overall for robustness in controlling for endogeneity. The “ratio sums” of these effects for the leftmost column, which we have calculated, are well in excess of unity (1.634), a possible indicator of overfitting, which is common when variables are selected by significance tests. On the right side these sums are $.828+.725=1.553$, but Logdate is considered a control variable. In the middle columns the leftmost sum is low, $.589$, when the Brown-Eff model competes with “Hidden variables” and $.830+.593$ for “Hidden variables” in competition with Brown-Eff. Discussion of the “Hidden variables” model can be found in White, Oztan, Gosti, Wagner and Snarey (2012). An advantage is that Snarey provided a number of fully coded variables, including a new High Gods variable, previously unpublished. Both models have very high significance for distance autocorrelation (in all runs, $p < .0000001$). Brown and Eff (2010:12) concluded that

“cultural transmission, Wy, turns out to be overwhelmingly the most important force conditioning the presence of moralizing gods, and that transmission is geographic, based on diffusion across space, rather than linguistic, based on transmission from a common ancestor.”

Table 4: 2SLS-IS regression results for the three models, separate and combined, in Table 3: column output is explained in Appendix 1. These 2SLS-IS results are similar to those of Brown and Eff except for the Principal component PCsize and PCsize.2 variables (adjusted by changing their signs). The significance values indicated by asterisks (***) < 0.01 , ** < 0.01 , * < 0.1) are the same for all but the PCA variables and food scarcity, which have slightly lower significance (noted by *, **). Model results are basically the same within error bars.

High God Model: Brown and Eff (2010)

```
V238 Source("examples/create/create_EduR_1/create_EduR_1.5DistBrownEffSign.R")
Source('http://intersci.ss.uci.edu/wiki/R/create_EduR_1.5DistBrownEffSign.R') #remote source
DEPV MoralGod coef xrange dx dy/dx Fstat df p-value VIF |dy/dx| %coded tot.cor part.cor
(Intercept) -0.178 NA NA NA 0.345 3598.955 0.55711584 NA NA 0.000 0.000 0.000
distance 0.953 2.000 1.906 0.635 48.212 4278.706 0.00000000 1.572 0.635 0.000 0.610 0.121 ***
PCAP (-)0.036 19.920 -0.718 -0.239 2.891 29611.062 0.08909325 1.172 0.239 1.000 -0.090 -0.120 */
foodscarc 0.093 4.000 0.372 0.124 2.229 94.609 0.13876607 1.032 0.124 0.774 0.143 0.153 //
anim 0.084 9.000 0.758 0.253 3.920 11854.137 0.04774623 1.580 0.253 1.000 0.503 0.276 **
PCsize (-)0.104 8.058 -0.841 -0.280 5.393 15088.622 0.02022705 1.514 0.280 0.988 -0.221 -0.102 **/
PCsize.2 -0.037 28.955 -1.062 -0.354 4.979 59811.483 0.02566309 1.258 0.354 0.988 0.033 -0.103 **/
caststrat 0.215 3.000 0.646 0.215 4.214 22270.164 0.04010317 1.128 0.215 0.970 0.281 0.213 **
eeextwar -0.034 16.000 -0.538 -0.179 8.795 980.250 0.00309332 1.089 0.179 0.827 -0.079 -0.190 ***
Train R2:final model Train R2:IV_distance (PCAP variable fixed)
0.4637752 0.9805554
```

High God Model: Brown and Eff (2010) and White, Oztan, Gosti, Wagner and Snarey (2012), combined

```
Source("examples/create/create_EduR_1/create_EduR_1.5DistB_EffReduce_2Models.R")
Source('http://intersci.ss.uci.edu/wiki/R/create_EduR_1.5DistB_EffReduce_2Models.R') #remote source
DEPV HiGod4 coef xrange dx dy/dx Fstat df p-value VIF |dy/dx| %coded tot.cor part.cor
(Intercept) -1.751 NA NA NA 6.499 56445.134 0.01079363 NA NA 0.000 0.000 0.000
distance 0.863 2.000 1.727 0.576 27.280 15802.823 0.00000018 1.740 0.576 0.000 0.550 0.173
PCAP (-)0.023 19.920 0.467 0.156 1.058 12828.717 0.30373279 1.423 0.156 1.000 0.090 0.093
PCsize (-)0.002 8.058 0.014 0.005 0.001 32016.100 0.97536684 2.320 0.005 0.984 -0.207 -0.114
SuperjhWriting 0.072 9.000 0.652 0.217 2.839 4192.523 0.09209900 2.907 0.217 0.989 0.324 0.226
PCsize.2 -0.022 28.955 -0.649 -0.216 1.264 6219.227 0.26097158 1.758 0.216 0.984 0.041 -0.064
eeextwar -0.014 16.000 -0.220 -0.073 1.361 385.979 0.24409391 1.195 0.073 0.828 0.016 -0.058
anim -0.002 9.000 -0.016 -0.005 0.001 461981.916 0.97537156 3.081 0.005 1.000 0.449 0.284
AnimXbwealth 0.056 10.000 0.557 0.186 1.325 58191.820 0.24968020 2.712 0.186 1.000 0.391 0.276
caststrat 0.099 3.000 0.297 0.099 0.756 115241.385 0.38459768 1.253 0.099 0.973 0.262 0.217
FxCmtyWages 0.317 1.000 0.317 0.106 3.102 217.861 0.07961140 1.425 0.106 0.602 0.245 0.233
foodscarc 0.027 4.000 0.109 0.036 0.189 146.749 0.66428446 1.172 0.036 0.774 0.126 0.147
No_rain_Dry 0.174 3.000 0.523 0.174 3.741 15668.047 0.05309694 1.785 0.174 1.000 0.407 0.277
Missions 0.256 1.000 0.256 0.085 2.757 5665.285 0.09687603 1.194 0.085 1.000 0.241 0.210
logdate 0.617 2.883 1.778 0.593 3.045 28482.940 0.08102152 1.394 0.593 1.000 0.051 0.161
Train R2:final model Train R2:IV_distance
0.4253451 0.9763177
Diagnostics Fstat df p-value
RESET 4.440 1831.042 0.035
Wald.on.restrs 3.465 3238.932 0.063
NCV 0.045 48199.869 0.832
SW.normal 1.547 1147.897 0.214
lag..distance 2.480 222530.561 0.115
```

High God Model: White, Oztan, Gosti, Wagner and Snarey (2012)

```
Source("examples/create/create_EduR_1/create_EduR_1DistMGdHV7$52.R") HiddenVars
Source('http://intersci.ss.uci.edu/wiki/R/create_EduR_1.5DistB_EffHiddenVariables2.R') #remote source
DEPVAR HiGod4 coef xrange dx dy/dx Fstat df p-value VIF |dy/dx| %coded tot.cor part.cor
(Intercept) -2.039 NA NA NA 10.591 29267.507 0.00113790 NA NA 0.000 0.000 0.000
distance 0.798 2.000 1.597 0.532 27.091 42390.073 0.00000020 1.532 0.532 0.000 0.550 0.208
AnimXbwealth 0.071 10.000 0.712 0.237 4.488 492405.829 0.03412642 1.336 0.237 1.000 0.391 0.285
FxCmtyWages 0.312 1.000 0.312 0.104 3.633 1707.006 0.05680562 1.315 0.104 0.602 0.226 0.223
SuperjhWriting 0.057 9.000 0.515 0.172 3.952 25658.157 0.04682820 1.395 0.172 1.000 0.340 0.261
No_rain_Dry 0.215 3.000 0.644 0.215 7.428 204011.553 0.00642152 1.398 0.215 1.000 0.407 0.289
Missions 0.302 1.000 0.302 0.101 4.209 64145.581 0.04020857 1.129 0.101 1.000 0.241 0.213
logdate 0.816 2.883 2.353 0.784 6.716 17880.892 0.00956263 1.120 0.784 1.000 0.051 0.152
```

```

100% sample R2:final model Train R2:IV_distance
0.4107085 0.9770213
Diagnostics Fstat df p-value
RESET 2.515 9887.941 0.113
Wald.on.restrs 1.441 770.873 0.230
NCV 0.217 18656.793 0.641
SW.normal 0.133 18127.015 0.715
lag..distance 2.849 178946.793 0.091

```

2 - 3 Resilience: Using the Random Inferential Subsample Training-Test Ratio

Resilience is the ability of a cohesive entity to recover from a shock or disturbance. Disturbing the sample used in testing a statistical model, e.g., by taking or comparing results from random subsamples, is a basic concept used in inferential statistics and computation. Computer scientists typically use independent (train-test) samples first to entrain model parameters for a sizeable random subample of observations (e.g., regression coefficients) and then to test whether, holding these constant in the complementary “holdover” subsample, other key features of the model are resilient. This does not require identical but rather comparable results. Measures of significance, for example, will be affected by variations in sample size, and will be most comparable in split-halves perturbations of the independent test-retest samples. If the training sample is larger the test sample will be expected to have lower significance results. What will be observed as lower resilience are certain variables that have much lower significance than due only to smaller sample size and the effect of missing data. Use of the FOR loop in our source code #4 and varying the subsample ratio from 50% to 79% to 100% (50% to 21%, or none, in the holdover sample) allows: 1) observation, given parameter estimates from in the larger train samples, of the resilience of effects in the smaller (random) test subsample to see if significance tests are comparable given the difference in sample size, 2) use of multiple runs to show how variables perform differently in their significance tests, relative $\delta y/\delta x$ effects and other measures that show differences in resilience of the statistical model, given that the regression coefficients differ in each training subsample. The percentage parameter is chosen by name in the fourth source code (as in `run_model_70.R` versus `run_model_79.R`) or the user can copy and change the name of a script and then change the train parameter inside the script.

One of the motivations for our reanalysis of the Brown-Eff model was that while the R^2 , significance tests and robustness for exogeneity were excellent overall, some of the R^2 hierarchically partitioned (Chevan and Sutherland 1991) into R^{2p} for individual variables (agricultural potential-PCAP, foodscarcity, PCsize and PCsize.2 —a principal component measure composed of aspects of political hierarchy and community size and its square, designed to provide measures for one of Swanson’s hypotheses about High Gods) were particularly low, and slightly suspect because of the use of composite principal component (PC) measures. When we ran the test-retest perturbations (option_79% for our #4th source code), these were the same variables that lacked resilience. The results in Table 5 show a model (see **INSET** for variable definitions) that is robustly estimated overall, with $R^2 = .467$, while of four runs of script #4 set for a 79% train sample, the 28 significance tests for the seven variables in the model, 21 were significant at $p < .10$. The PCAP variable (scs\$V921, V928 1st principal component of PCA, measuring agricultural potential) mostly lacked significance ($p = .37, .22, .21$ and $< .10$) in these runs. (In six extra runs $p = .40, .30, .17, .16, .11, .02$). Three other variables were nonsignificant at $p > .10$ in one out of four runs (foodscarc $p = .52$, anim $p = .44$, and negative PCsize.2 (squared) $p = .29$). (For 6 extra tests negative PCsize.2 $p = .28, .26$, and four significant values). Test results from the 21% holdout sample ($n = 39$ cases) tend to replicate at a somewhat lower level in significance tests, with 68% of the p-values $< .10$, 25% $< .15$, and 7% $> .15$. Brown and Eff (2010) is thus a robust model even for 79%/21% random independent subsamples. And that in spite of our small failures to replicate the PCA calculations used by Brown and Eff. Appendix 2 breaks out all these results for further statistical inferences about model resilience.

Table 5: Screen shot of 2SLS-IS results for the Brown-Eff (2010) model variables (defined in the inset)

```

depvar=scs$V238 source("examples/create/create_EduR_1/create_EduR_1.5DistBrownEffSign.R")
source('http://intersci.ss.uci.edu/wiki/R/create_EduR_1.5DistBrownEffSign.R')#remote source
Variables beginning with PC in these and all subsequent tables are effectively positive
DEPV MoralGod coef xrange dx dy/dx Fstat df p-value VIF |dy/dx| %coded tot.cor part.cor
(Intercept) -0.178 NA NA NA 0.345 3598.955 0.55711584 NA NA 0.000 0.000 0.000
distance 0.953 2.000 1.906 0.635 48.212 4278.706 0.00000000 1.572 0.635 0.000 0.610 0.121 ***
PCAP (-)0.036 19.920 -0.718 -0.239 2.891 29611.062 0.08909325 1.172 0.239 1.000 -0.090 -0.120 */
foodscarc 0.093 4.000 0.372 0.124 2.229 94.609 0.13876607 1.032 0.124 0.774 0.143 0.153 //
anim 0.084 9.000 0.758 0.253 3.920 11854.137 0.04774623 1.580 0.253 1.000 0.503 0.276 **
PCsize (-)0.104 8.058 -0.841 -0.280 5.393 15088.622 0.02022705 1.514 0.280 0.988 -0.221 -0.102 **/
PCsize.2 -0.037 28.955 -1.062 -0.354 4.979 59811.483 0.02566309 1.258 0.354 0.988 0.033 -0.103 **/
caststrat 0.215 3.000 0.646 0.215 4.214 22270.164 0.04010317 1.128 0.215 0.970 0.281 0.213 **
eeextwar -0.034 16.000 -0.538 -0.179 8.795 980.250 0.00309332 1.089 0.179 0.827 -0.079 -0.190 ***
Train R2:final model Train R2:IV_distance PCAP variable fixed (rerun 100% and 79%)
0.4637752 0.9805554
Diagnostics Fstat df p-value
RESET 0.398 930.596 0.528
Wald.on.restrs 1.995 37.009 0.166
NCV 0.157 29815.515 0.692
SW.normal 5.969 5664.569 0.015

```

```

lag..distance 0.966 18571.059 0.326
Sample 79% coef xrange dx dy/dx Fstat df p-value VIF |dy/dx| %coded tot.cor 2sls.cor
(Intercept) -0.133 NA NA NA 0.140 18855.943 0.70794958 NA NA 0.000 0.000 0.000
distance 0.899 2.000 1.798 0.599 30.673 2603.808 0.00000003 1.533 0.599 0.000 0.572 0.124 ***
PCAP (-)0.031 20.244 0.627 0.209 1.511 8186.546 0.21908178 1.245 0.209 1.000 0.101 0.110 //
foodscarc 0.132 4.000 0.529 0.176 3.908 501.539 0.04860373 1.043 0.176 0.759 0.199 0.206 **
anim 0.093 9.000 0.834 0.278 3.804 20917.047 0.05114962 1.510 0.278 1.000 0.480 0.287 **
PCsize (-)0.118 8.061 -0.953 -0.318 5.110 24415.196 0.02379171 1.537 0.318 1.000 -0.217 -0.126 **/
PCsize.2 -0.040 28.847 -1.145 -0.382 4.093 38226.480 0.04307062 1.306 0.382 1.000 0.031 -0.109 **/
caststrat 0.248 3.000 0.744 0.248 4.284 38188.803 0.03847757 1.126 0.248 0.970 0.304 0.238 **
eeextwar -0.032 16.000 -0.512 -0.171 4.641 127.366 0.03310760 1.111 0.171 0.812 -0.043 -0.155 **/
Train R2:final model Train R2:IV_distance
0.4430773 0.9788768
21% coef xrange dx reff Fstat df p-value VIF |dy/dx| %coded tot.cor part.cor
(Intercept) -0.133 NA NA NA 0.202 9151.776 0.65326794 NA NA 0.000 0.000 0.000
distance 0.899 2.000 1.798 0.599 43.572 1302.867 0.00000000 1.533 0.599 0.000 0.774 0.321
PCAP (-)0.031 14.249 0.441 0.147 2.164 4011.866 0.14135937 1.245 0.147 1.000 0.033 0.157
foodscarc 0.132 4.000 0.529 0.176 5.425 265.571 0.02059757 1.043 0.176 0.829 -0.067 -0.059
anim 0.093 6.000 0.556 0.185 5.467 10162.876 0.01940065 1.510 0.185 1.000 0.634 0.352
PCsize (-)0.118 7.583 -0.897 -0.299 7.348 11847.182 0.00672198 1.537 0.299 1.000 -0.311 -0.161
PCsize.2 -0.040 23.941 -0.950 -0.317 5.894 18484.006 0.01520444 1.306 0.317 1.000 0.097 0.081
caststrat 0.248 3.000 0.744 0.248 6.173 18444.499 0.01298320 1.126 0.248 0.971 0.084 0.032
eeextwar -0.032 16.000 -0.512 -0.171 6.182 74.281 0.01515514 1.111 0.171 0.886 -0.123 -0.186
Test R2:final model Test R2:IV_distance
0.57159750 0.9879918
Diagnostics Fstat df p-value INSET: Definitions of variables
RESET 1.006 9778.729 0.316
Wald.on.restrs 0.904 173.403 0.343 distance Autocorrelation effect for cultural transmission
NCV 0.051 487.747 0.821 PCAP Agricultural potential (low) *1st PCA cbind(v921,v928)
SW.normal 4.048 3242.058 0.044 foodscarc Chronic food scarcity sccs$v1685
lag..distance 0.582 28045.649 0.446 anim % subsistence: Animal husbandry sccs$v206
Diagnostics Fstat df p-value PCsize Community size (SMALL) *1st PCA cbind(v63,237)
RESET 1.006 9778.729 0.316 PCsize.2 PCsize squared (squared)
Wald.on.restrs 0.904 173.403 0.343 caststrat Degree of caste stratification sccs$v272
NCV 0.051 487.747 0.821 eeextwar Frequency of external war sccs$v1650
SW.normal 1.762 499.715 0.185
lag..distance 0.582 28045.649 0.446 *First principal component of two variables

```

Diagnostic statistics for regression results were part of the original Eff and Dow (2009) program and are included in the new 2SLS-IS software; nonsignificant p-values satisfy the tests. They include whether for the various tests: 1) RESET - the relationship of the independent and dependent variables are linear, 2) Wald - the appropriate variables are dropped, i.e., whether those of the larger list of indep_vars that were not included in the restrict_vars list were significant predictors of the dependent variables, 3) Breusch-Pagan homoskedasticity, NCV for non-constant variance - the error terms for the second-stage regression were not bunched, 4) Shapiro-Wilk - the residuals are normally distributed, and 5) lag - the residuals had no spatial lag for language autocorrelation (LM test). Test 4), however, fails for 2SLS-IS but passes for Brown and Eff, where p=0.370 (n.s.) but three out of four of the 2SLS-IS tests have p<.05. Given that there are a total of 25 significance tests, however, it is within the bounds of the null hypothesis that one or two of the p-values in a given row will depart from significance at p>.10.

Diagnostic tests for the 100% samples from Table 4 are summarized in Table 6, comparing Brown and Eff's results with those of runs of 2SLS-IS (White et al. 2012). These runs shown in all cases that the model has correct functional form, appropriate variables were dropped, residuals homoskedastic, and spatial lag of distance autocorrelation is exogenous with the independent variables. One of the 2SLS-IS runs shows normally distributed residuals but three do not.

Table 6: Diagnostic tests for the Brown and Eff (2010) and White et al. (2012) regression models for moral gods/HiGod4
create_EduR_1.5DistBrownEff.R

Significance levels for Diagnostic tests	Brown and Eff (2010)	Four runs of 2SLS-IS (2011, run 79%)			
		a	b	c	d
RESET test. H0: model has correct functional form	0.336	0.316	0.279	0.504	0.737 (all p-values, p>.10)
Wald test. H0: appropriate variables dropped	0.402	0.343	0.141	0.286	0.152 (all values, p>.10)
Breusch-Pagan test. H0: residuals homoskedastic	0.646	0.821	0.796	0.403	0.629 (all values, p>.10)
Shapiro-Walk test. H0: residuals normally distributed	0.370	0.044	0.307	0.011	0.006 (some values, <.10)
LM test. H0: Spatial lag (distance) not endogenous	0.298	0.446	0.845	0.393	0.524 (all values, p>.10)

With multiple significance tests we expect to get some number of type I errors. One way to deal with this is the Bonferroni (Holm 1979, Rice 1989) simple sequentially rejective multiple test procedure at some level of significance, e.g., $\alpha=.10$ for an ordered set of k tests, where the p-value of the i'th test in order of significance is rejected if its p-value $\leq \alpha / (1+k-i)$. In the Brown and Eff and Eff and Dow models, each with six p-values < .10, only the top three values remain significant in the sequential (Bonferroni) correction test, and only two if we set $\alpha=.10$. The lower part of Table 5 shows the output coefficients and p-values (deleting the foodscarc variable that was significant at p>.50), for which the top three variables in significance (-eeextwar, anim, and -PCsize.2) retain significance in the Bonferroni correction test at p<.10 and

two at $p < .05$. The vifs for three variables indicate covariation between anim, PCsize, and possible residual neighborhood clusters not removed by the Instrumental Variable (IV) for distance autocorrelation. This would account for the high relative effect sums for the independent variables. Dividing each by its vif, the sum is 1.14, reasonably close to one, and indicative of a good model. Three variables (distance, anim, and eextwar) show a noticeably lower correlation due to WY as compared to the raw correlations without the instrument (IV) created as a control for autocorrelation,

3 - Robustness tests of prototype models and identification of improved models

Robustness in regression means tests that exogeneity is not violated so that independence in the error terms can be assumed to give efficient estimates of significance and other measures (e.g., Hausman 1978, White and Lu. 2010). In OLS regression, without control for autocorrelation, methods of obtaining robustness include generalizations of maximum likelihood for locating the center of a distribution (*M-estimation*; Huber 1964, Fox 2002a,b,c) or bootstrapping (Davison and Hinkley 1997, Fox 2002c) from random sampling of the dataset to estimate confidence intervals on the parameters (e.g., coefficients) of the regression. In statistics generally a robust technique is one that performs well even if its assumptions are somewhat violated by the true model from which the data were generated. In evolution it is the persistence of a system's characteristic behavior under perturbations or uncertainty. A statistical model is also robust if it has the ability to remain valid under different assumptions, parameters and initial conditions, as is illustrated by our test-retest results (Appendix 2 and Table 5). These fix parameter estimates for a model given a random sample of the data and then test robustness of significance and R^2 for each independent variable using the observations not sampled. Robustness is also helped in regression by having more accurate variables, with greater high face validity and less random error. The dependent variable (in the models illustrated here, sccs\$V238, Guy Swanson's (1960) Moral gods variable or Snarey's HiGod fully-coded version) both coded by an expert in comparative religion. Murdock (1967) coded this variable for the Ethnographic Atlas of 1270 societies without any reported difficulty. The SCCS database includes Murdock's coding from the Atlas for 168 of the 186 cases in the sample, with 18 cases having missing data. A priori, this is a very likely to be a high-concordance variable that allows reliable judgments by coders for a given society, given information from the ethnographic sources.

Directness is an aspect of robustness in regression that is often forgotten because it requires networks of related variables dependent variable, e.g., $x_2 \rightarrow x_1 \rightarrow y$, where y is the initial dependent variables, x_1 an independent variable, and x_2 an independent variable with an effect on x_1 in a second regression. In this case x_1 is a direct effect but x_2 has an indirect effect on y . Robust inferences about indirect effects require, among other things, that the samples for each variable are either fully coded or fully imputed for missing data, which is an option for our (#3rd) 2SLS-IS script.

We concluded that the Brown-Eff model was overfitted and too dependent on background variables – probable indirect effects on Moral Gods. The various clues were these. **First**, consistent with their model, a variety of measures of scarcity were predictive, which increases the need for cooperation and the value of moral codes and makes the adoption of moralizing gods more valuable and more likely. **Second**, the dependence on animal variable did not specify the *type* of animal, which on examination for the high-dependence societies tend to be horses, camels, and donkeys with a scattering of sheep and goats, that is societies with a high potential for acquiring wealth in herds, wealth through trade, and increasing population size by greater accumulation of wives when animals are given as a purchase on a wife's commitment to the husband's group. The wealth-augmenting potential of these sorts of animals and their effect of differential rights in animals, access to brides, and subgroup sizes makes for the possibility of wealth inequalities that augment when total population growth begins to outstrip resources, making for increasing scarcities and lower remuneration to lower strata of labor tending animals contrasted with increasing value of productive resources in herds accruing to the upper strata. **Third**, the fixed stratification variable (castes) hinted at stratification as a factor associated with scarcity and lack of mobility in lower social strata. **Fourth**, the times at which lower social strata are known to suffer from inequality that is viewed as inequitable or unethical, requiring cooperative restitution between strata, are also those known to be crisis periods in empires or the state polities of which many of the societies in the sample are part, and thus impacted by fluctuations of scarcity at these higher levels. **Fifth**, when stratified political units also collect taxes, their periods of stress are likely to induce additional inequality that is viewed as inequitable or unethical, requiring cooperative restitution between strata, and again increasing the value of moral codes and making the adoption of moralizing gods more valuable and more likely. Variables such as levels of political jurisdiction that might make it seem that religious strata emulate political strata seem less likely, and it may be more likely that there are other mediating variables: that politically stratified societies undergoing the stress of rising population with diminishing and more valued productive resources held increasingly richer hands might be the cause of conflict of a societal realignment of more ethical concepts and practices and make the the value of moral codes and making the adoption of moralizing gods more valuable and more likely. **Sixth**, it would seem likely that this same kind of recurrent situation would occur as between landowners in societies with fixed communities and fixed property and wage laborers without landed property. We thought that these six hypotheses were

more likely as more direct explanations of effects on Moral gods and of Brown and Eff's results as identifying indirect effects that operated through our "hidden variables." The construction of a full network model (White, Oztan, Gosti, Wagner and Snarey 2012), however, is another story.

3 – 1 Latent Variables: Moral Gods ("Hidden variables")

Competition among sets of variables for a given model may involve whether some variables are direct and others indirect, and whether mediators are missing. *Latent variables* are model-driven measures that are not directly observed (measured) but inferred from other directly measured variables or generated by a mathematical model. "Hidden variables" are ones that could in principle be measured but for practical reasons have not, or difficulty to measure. They may be constructed from other indicators variables using latent structure analysis. Latent variables based on common correlations may be constructed by common factor models, or more abstract hypothetical concepts may be constructed from uncorrelated variables. Because they reduce the dimensionality of data while also linking observable data to underlying concept, they are often seen as serving a function like that of scientific theories.

Latent variable models are evaluated not only with a concern for rejecting the null hypothesis but the *sensitivity* or *power* of statistical tests: the probability of detecting true positives out of the total of true positives, and of rejecting the null when it is false and should be rejected, or $1 - \beta$, where β is the test probability of false negatives, rejecting the null hypothesis when it is true, i.e., type II error. Without exogeneity in a regression equation, error terms are usually underestimated (e.g., given many similar cases whose existences are near-duplicate or copies of one another because the cases are not independent), and type I errors are common, rejecting the null hypothesis when it is true. The sequential Bonferroni test tries to guard against type I errors, while the more power the test, the better type II errors are guarded against. At a given probability level the β error rate goes down the larger the sample and the higher the R^2 , and up the more predictors (<http://www.danielsoper.com/statcalc3/calc.aspx?id=3>). Power is diminished the lower the setting of the null hypothesis "significance" threshold.

Using the script for latent variable analysis in the R package of *lavaan* (Rossee 2011.a, 2011b), Table 7 specifies a latent variable ("Malth") that combines two agrarian variables (wage labor in fixed communities and jurisdictional hierarchy set as zero with no writing), and, from our 2SLS-IS regressions, three independent variables and the WY instrument for autocorrelation. The script is not equipped with missing data imputation, and thus uses only the 112 cases completely coded. The results show that WY which reduces the variance on the dependent variable (*HiGod4*) and that the wage labor in fixed communities independent variables works against the latent variable. This indicates that although *FxCmtyWages* is correlated with the other agrarian society variable, it has less effect on its own than the latent variable (episodes of Malthusian scarcity hypothesized to alter the dependent variable) would predict. This "corrects" for a nonlinearity that would have been more difficult to specify in the linear regression model.

Table 7: A *lavaan* script for a latent variable (episodes of Malthusian scarcity hypothesized to alter the the *HiGod4* dependent variable) and its results

```

setwd('/Users/drwhite/Documents/3sls/scs/')
load(url('http://intersci.ss.uci.edu/wiki/R/source/HiG.Rdata'))
library(lavaan)
model <- '
# latent variable definitions
Malth =~ FxCmtyWages + SuperjhWriting # + AnimXbwealth
# regressions
HiGod4 ~ No_rain_Dry+FxCmtyWages+AnimXbwealth + WY + Malth '
fit <- sem(model, data=HiG)
summary(fit, fit.measures=TRUE)

Results = Lavaan (0.4-9) converged normally after 22 iterations
Malth =~ FxCmtyWages + SuperjhWriting # + AnimXbwealth           HiGod4= dependent variable
.389      .206      .616                                           .288  variances
HiGod4 ~ No_rain_Dry+FxCmtyWages+AnimXbwealth + WY + Malth
      .209      -.352      .234  -.518      1.427  regression coefficients

```

P-values for the null hypothesis (type I error) are uniformly significant for total Chi-squared ($p=.002$), Chi-squared baseline model ($p=.000$), Root Mean Square Error of Approximation (RMSEA, $p=.011$). For fit (type II error), the Root Mean Square Residual is 0.087 and a value less than .08 is generally considered a good fit (Hu & Bentler, 1999). Comparing full model versus baseline indices for CFI and TLI (the former expected to be larger than the latter) are 0.828

and 0.587 (which are not bad for this number of variables, and which improve to .864 and .756 when a White-Murdock alignment IV is used complement Wy), and might be expected to improve with missing data imputation).

R offers a great advantage for use of the *lavaan* program in conjunction with our 2SLS-IS scripts because script #3 saves the WY Instrumental Variable to the file WYhat.Rdata, which is loaded along with the main *.Rdata file in the *lavaan* R session. Source #3_full and source#6 can be mobilized to save all other variables, after they are fully imputed. In this case those variables include *FxCmtyWages*, so only that one multiply imputed variable need be averaged across the imputations and copied from the output. Whether the *lavaan* model improves depends on the imputation. Bias in the imputation can be checked by including a dummy variable in the 2SLS-IS script #2 for this variable that is scored for cases coded and those imputed.

Use of 2SLS-IS script outputs also simplifies use of the R *sem* package (Fox 2006) because our script #3 also saves a write.csv(file="WYWX.csv") for read.csv("WYWX.csv",header=TRUE) or WYWX.Rdata that has all of WX variables used to construct the WY Instrumental Variable (IV). Each of the individual WX variables can be taken as separate IVs to provide at least as many instrumental variables as coefficients, as required to estimate sem equations.

3 – 2 Competition among Models using Significance Tests and Relative $\delta y/\delta x$ effects

Competition among statistical models is not determined by R^2 , significance tests or by $\delta y/\delta x$ relative effects. It is more concisely conceived as the *power* of tests as a probability of not committing type II error, β , where β is the probability that we have mistakenly failed to reject a false null hypothesis, that is, to fail to find an effect when one exists. Random or measurement error and bias, sample size, biased sampling, finding variables that are predictive but not explanatory all contribute to type II error, β . The power of a test is the probability $1 - \beta$, and the more powerful a test, the more likely it is to show, statistically, an effect that exists.

Another example of type II or β error is a model in which the predictive arise from indirect effects, that operate through mediators, rather than direct effects that have not yet been discovered. Such is the case with the “hidden variables” discussed in 3-1 arise from the hypothesis that crises of resource scarcities might create conditions where cooperation is needed, and thus enhance the adoption of moral gods, we began to look for other variables. Encouraged by prior research on what kinds of animals, under what conditions, might enhance the adoption of moral gods, we replaced the “anim” variable with a dichotomous intersection of two variables (“animXbwealth”: anim sccs\$v206 and bridewealth sccs\$v208=1) that defined pastoral economies with potential multiplier effects in crises of relative scarcity that augment inequality to the point that remediation of moral balance is needed. We hypothesized that pastoralists that use large animals both in trade and in marital bridewealth payments might build up inequalities in herd distribution that have multiplier effects. When some kin groups dominate wealth in herds and through trade, their benefits as owners of wealth in times of scarcity are multiplied relative to others. The increase in significance with this variable, animXbwealth, compared to the variable for animals (“anim”) alone, led us to add precision to other variables that would measure comparable conjunctive variables in other areas. *FxCmtyWages* combined the variables for fixed communities (sccs\$v61=6) with wage labor (sccs\$v1732 and 1009) where inequalities due to multiplier effects of inequality between landowners and agricultural workers. *SuperjhWriting*, combining superjurisdictional hierarchy (Superjh, sccs\$v236) with writing (sccs\$v249), was defined as a variable with possible multiplier effects between political elites with a capability to record taxes that may become excessive in times of overpopulation relative to resources. Taxation *per se*, however, even when considered burdensome, has no effect when added to this model. Although excluding external war, this model had some of the conceptual dimensions of the original Brown-Eff model for moral gods, now specified by pairs of variables that defined “hidden variables” that lent precision to conditions generating scarcities that would require enhanced means of cooperation and enhanced concepts of dealing with inequality with new principles of morality and justice. White, Oztan, Gosti, Wagner and Snarey (2012) give further explanation for why these variables were chosen. Table 6 shows results of three contrastive regressions, each with 2SLS distance autocorrelation. In the center of the table are six columns of a single model that combines all the variables from both the *hidden variables* and the Brown-Eff models for predictors of moral gods. When combined in direct competition, none for center Brown-Eff variables are significant, and all “hidden variables” are significant. On the far left is the model with only the variables of the Brown-Eff moral gods model ($R^2=.47$ or .48). On the far right is the “hidden variables” model ($R^2=.41$).

Our conjecture about relative $\delta y/\delta x$ effects, ignoring control variables such as Distance or Date of observation (logdate) shown in bold font in Table 6, is that the sum of the absolute values of relative $\delta y/\delta x$ effects is near unity when the model is well fitted, well over one when overfitted, and under one when underfitted. By this criterion the “hidden variables” model is underfitted and the Brown-Eff overfitted. The underfitted “hidden variables” model has lower R^2 but in mixed-

model its variables out-compete those of the Brown-Eff variables in terms of significance (4 of 5 $P < .10$ vs. all n.s.) and relative $\delta y / \delta x$ effects (averaging .166 for “hidden variables” vs. .084). Absolute values of relative $\delta y / \delta x$ effects act like percentaged variables that take values from 0 to 100 (here 0 to 1 over the minimum to maximum values in the IVrange of the regression coefficient relative to the IVrange: $reff = \text{Coef} * \text{IVrange} / \text{DVrange}$) in ordinary regression. If relative $\delta y / \delta x$ effects acted as a “do” operator (Pearl 2009:368) and all of the variables with positive relative $\delta y / \delta x$ effects are set at their maximum then those with negative relative $\delta y / \delta x$ effects are set at their minimum then a regression “do” operator would predict a maximum value of the dependent variable. And if all of the variables with positive relative $\delta y / \delta x$ effects are set at their minimum and those with negative relative $\delta y / \delta x$ effects are set at their maximum then the regression should predict a minimum value of the dependent variable.

The strategy of the “hidden variables” model is the clear winner in direct competition with the Eff and Brown model (center six columns in Table 6). In this model, precisely specified variables outperform broader and less well defined variables (such as 1st principal components of pairs variables). The “hidden variables” model also has an average $R^2 = .41$, 17% higher than the Brown-Eff $R^2 = .35$ as measured in 2SLS-IS without factoring in the distance effect by use of hierarchical partitioning, which gives an $R^2 = .48$ for the Brown-Eff variables. Only in its next iteration will 2SLS-IS have the capability doing this calculation. Rounton and Eff (2011) also use precisely specified interactive variables and hierarchically partitioned R^2 in their external war model with outstanding results.

4 – Networks of Variables, 2SLS-IS Regression, and SEM

The two models discussed in this introduction to new scripts for regression and latent variable analysis illustrate how scripts from different packages in R can yield complementary results, in this case helping to differentiate a first stage of modeling on a dependent variable, in this case by Brown and Eff (2010), followed by a second stage that builds on insights from the first to try to identify more direct causes of a focal variable. This requires not just a new model, however, but a series of linked models that follow from potential direct effects to the indirect effects that affect them, in turn, as in $x_2 \rightarrow x_1 - y$, where y is the dependent variable, x_1 the direct effects (in this case the “hidden variables” shown in Table 3 to drive out some of the variables in the Brown and Eff regression), and x_2 the effects found to act on an x_1 and indirectly on y .

4 – 1 Direct and Indirect Variables, correcting for biased subsamples, and networks of variables

To link regression results into a network of variables with both direct and indirect effects poses some new problems. The `two_stage_ols_full.R` script allows the study of networks of effects among directed chains of variables linked as independent and dependent variables with full imputation of missing data. Directionality of effects in these models must be carefully evaluated with reliance on domain expertise and Bayesian inference from new data in the face of prior substantive theory. In survey research with imputed data (Rubin 1996) missing data are normally imputed for all cases in the sample for both dependent and independent variables. Only one author (von Hippel 2007) argues for the special case of the more conservative approach taken by Eff and Dow (2009). Others (Rubin 1987, Rubin 1996, Roderick & Rubin 1987, 2010, Allison 2001, Wayman 2003) argue for the standard approach, with cross-checks where there is doubt for convergence of results with and without imputation of the dependent variable. With this caveat, the standard approach allows us to consider networks of variables with comparable relative $\delta y / \delta x$ effects between independent and dependent variables, and dummy variables showing missing data for each variable. This is especially effective for SCCS variables when the author coded every second, third or fourth case in the list of societies or where missing data appear at random. More problematic are variables from studies like that of Paige and Paige (1981) that contributed codes only for prestate societies. To use these variables correctly for the total sample, the researcher might have to select and extend the codes on these variables on a good-sized random subsample of the state-level societies. A dummy variable for state versus prestate societies is not useful when lacking relevant comparisons with state-level societies. When the dependent variable is one from the Paige and Paige study conditional imputation is needed.

The 2SLS regression model with imputation based on Eff and Dow’s (2009) script, emulated by `two_stage_ols.R`, can be used to treat single dependent variables but is not recommended for networks of variables because only those missing data cases are imputed that are coded for the dependent variable. For a network of variables, i.e., containing multiple dependent variables, using Eff and Dow’s method has the defect for networks of variables of assigning each pair of dependent and independent variables a sample that differs according to missing data for the dependent variable. This poses problems of nonuniformity and potential bias from missing data (Robins, Scharfstein and Rotnitzky 1999; Greenland 2010). When subsamples for dependent variables differ substantially, identification and correction for spurious correlations in the larger network of variables becomes a concern. Results of our `two_stage_ols.R` and

`two_stage_ols_full.R` scripts can be compared to see if they differ, but for networks of variables bias in subsamples would still remain a problem.

Bias in subsamples for networks of variables are easier to check, along with correcting for common causes and many other problems, with full imputation. They open the possibility for structural equation modeling (SEM), which includes potential bias and relevant control variables. The long history of SEM and path analysis (Wright 1921, 1923, 1934) has generated many software packages, the foremost among which for 2SLS-IS has been the *sem* freeware in R (Fox 2006, 2009, Grace 2009). SEM uses the solution of simultaneous equations or complex computations of partial correlations. In purely structural graphical form, SEM model estimation is equivalent to causal graph analysis, as shown by Pearl (2009, 2011). Most of the SEM software (AMOS, EQS, and LISREL for personal computers) allows the model to be drawn on the screen (e.g., in the dot language), which is translated in lines of code (Kline 1998:6), as SEM in R (Fox 1999, Fox and Byrnes 2011, Rosseel 2011a) might do.

Pg8bib Pg9bib

It is virtually impossible to create a valid SEM or causal model beyond a limited scale in numbers of variables. That is, while it is possible to construct a network of variables linked by independent/dependent variables in regression analysis with 2SLS adjustments for nonindependence of cases, it is impossible to focus on SEM or causal subgraphs for more than limited clusters of variables. There is an advantage in focusing on smaller clusters of variables where more intensive research can stimulate a gradual accretion of results in fields that possess adequate datasets. Results of regression models of relatively proximal common-causes among relatively small sets of dependent variables, along with variables needed to adjust for various types of bias, can contribute results on potentially causal relationships that bear further investigation.

4 – 2 Structural arrows and SEM: What can we learn about theory from Graphical language?

Regression analysis, even in the best case of complete coding of variables, does not yield causal effects for a network of variables. Even the solutions in SEM to multiple equations starting with correlations free of endogeneity do not offer a sure-fire method of testing causal models (Pearl 2009:150). For regression-based modeling the “unit change” interpretation of relative effect (*reff*) coefficients included in our 2SLS-IS results (Tables 1, 3) is consistent with a structural interpretation of SEM effects “operationalized through a precise mathematical definition” in which Pearl (2009: 367) makes explicit the equivalence between SEM and causal graphs.

Pearl’s massive contribution to causality was to formally define concepts and measures as the basis for consistent mathematical definitions and proofs for a language of graphs that expresses precisely what can and cannot be derived from assumptions about causal relationships among variables. His definition of *causal* is neither deterministic nor probabilistic but refers to a set of structural equations that are functions of the form $x_i = f(pa_i, u_i)$, $i = 1, \dots, n$ (Pearl 2009: 27) where pa_i denotes the (causal) “parents” of x_i , and u_i denotes “disturbances” due to omitted factors (including noise, one would infer). Linear parametric models are the special case, “which have become a standard tool in economics and social science” (Pearl 2009:27; see chapter 5), with parameters α_{ik} that differ from those of ordinary regression coefficients, and where the linear model is

$$x_i = \sum \alpha_{ik} x_k + u_i, i = 1, \dots, n_i \quad (3)$$

“In linear structural equation models, the hypothesized causal relationships between variables can be expressed in the form of a directed graph annotated with coefficients, some fixed *a priori* (usually to zero) and some free to vary” (2009:144). Pearl showed that the same elements of a graphical language for models of effects apply to experiments, observational data, and hypothetical operations on changing a variable in a system of relationships while fixing the value of other variables (his *do* operator). He demonstrates why probability distributions alone are the wrong language for causality, and that a better approach is only partly Bayesian – what assumptions have what consequences? – but also partly a consistent mathematical extension of common sense, in the context of particular sets of assumptions about what might be inferred to affect what, what can be taken as valid *a priori* Bayesian beliefs, what evidence is already given, and how to change beliefs given new evidence. Precursors for a language of graphs “originate mostly from Physics (Gibbs 1902), Genetics (Wright 1921, 1934) and Economics (Wold 1954)” (Lauritzen 2011).

Pg10bib

One assumption of a language of graphs for a network of effects between independent and multiple dependent variables is the possibility of delimiting a system of interaction where there are no variables that have serious common outside causes, or where the chains of common outside causes are so long that their indirect effects can be small enough to be ignored, thus included in the u_i terms of equation (3), given that indirect of causal chains involve multiplication of effects. That is, chained positive or negative effects on a scale from -1 to 1, when multiplied, are decreasing in absolute value the longer the chain of multiplications for magnitude of indirect effects.

Delimitation of the boundaries of a network of effects is a key issue that allows for definition of *structural parameters*. While definitions differ, one variant is that, if all degrees of freedom of a delimited system are observable, a parameter in a theoretical model of the system is structural when it has a distinct structure for its effect. Informally, changing or removing a structural parameter alters system behavior. The delimitation of such systems (and external perturbations) borders on problems of complex systems and complexity sciences that recognize complex interactions among effects over time in bounded networks (which may also have external perturbations or changing boundary conditions). Often, understanding such interactions may involve multiple set of empirical data, simulations, and studies of longitudinal interactions.

For regression analysis, an important question is: Can a structural parameter α between $x \rightarrow y$ ever be equated with a regression coefficient? As Pearl (2009:150) shows: Yes, when all paths between X and Y (or x and y) in their graph are blocked by converging arrows. This is again an assumption about relative isolation of a system of variables in which $x \rightarrow y$ is embedded. A path p between x and y in a causal graph is *d-separated* or blocked (Pearl 2009:16-17) by a set of nodes Z when “ p contains a chain $x \rightarrow m \rightarrow y$ or a fork $x \leftarrow m \rightarrow y$ where m is in Z; or, p contains an inverted fork (collider) $x \rightarrow m \leftarrow y$ such that m is not in Z and such that no descendant of m is in Z.” (A set Z is said to *d-separate* X from Y “if and only if Z blocks every path from a node X to a node Y”; the definition does not apply reversibly to Y and Z). The overwhelming bulk of justification for conclusions about (*d-*) separability and connectivity in causal graphs must come from theory about a particular system of variables and not from an unlimited and thus unknowable set of interactions. Graphical language for structural arrows within such a delimited system can show us where data are relevant, and arrows or their absence in a causal graph show only the possibility of an effect that may not be demonstrable in a given dataset. Only occasionally, where assumptions about structural arrows greatly limit the possibilities, i.e., for a well-bounded problem, will data be relevant to causal estimation: normally a majority of graphical structures of such arrows are irresolvably confounded, in which case data do not resolve questions of causality. Nevertheless a structural model is sometimes identifiable (i.e., values of structural parameters on the arrows could be inferred if valid treatments of empirical data were available), even if it cannot be fully solved (i.e., by unequivocally testing a model of the system with valid data).

Given sufficient advances in understanding the graphical language for identifying potential effects and how they may encapsulate in limited systems, some advances may be made by evaluating regression analysis in which *iid* assumptions are justified in conjunction with models that are strongly justified by past research and theoretical arguments. We can only begin to estimate structural parameters, even of limited theoretical models, if we bypass premature conclusions based on spurious correlations and significance tests and do estimations with 2SLS and 2SLS-IS, experimental data, or opportunistic and carefully qualified quasi-experiments (like the event-based or event-sequence analyses discussed by contributors to Vayda and Walters (2011)). Given the *iid* property in 2SLS regression, it is useful to raise the bar further to include the problem of whether the potential results would meet the higher standards of causal separability or interaction rather than fitting a statistical model for the sake of R^2 and statistical significance. (Unfortunately, that bar has rarely been raised in cross-cultural studies. In general, as is frequently the case in cross-cultural studies, the learning gradient about sociocultural systems is not helped by using inadequate means of fitting models.) Still, even with regression that produces *iid* residuals, “there can be no conclusion made regarding the existence or the direction of a cause and effect relationship only from the fact that A and B are correlated”; even “when the relationship between A and B is statistically significant, a large effect size is observed, or a large part of the variance is explained.” In such cases determining “whether there is an actual cause and effect relationship requires further investigation” (Wikipedia: Correlation does not imply causation). But even in the weak form of incomplete models, identifying effects in models that are only partially tested can make use of the benefits of 2SLS and 2SLS-IS. Pg11bib

4 – 3 Exogeneity and Multiple Imputation?

Statistical 2SLS models, now common in econometrics, genetics, epidemiology, physics and social science, create a context for eliminating poor estimation (e.g., misestimation in significance testing) and biases that come from data samples in which the cases are nonindependent or variables are endogenous (e.g., unequal clusters of values in the sample of observed cases). Endogeneity can arise as a result of autoregression with autocorrelated errors, from measurement error, simultaneity, omitted variables, and sample selection errors. 2SLS and 2SLS-IS offer ways to correct for endogeneities such as arise from endogenous “social effects” (e.g., spatial and language clustering of cases in samples of societies). In anthropology, for example, some societies may have features that are independently invented, but diffusion and common origin, alone or in combination, are recognized as sources of nonindependence of cases. Many sources of nonindependence need to be recognized and tested or corrected, based on thinking through the ways in which different kinds of variables are endogenous or autocorrelated, such as measured in our sample regressions here using regression terms like Wy

(multiple network effects on autocorrelation in dependent variable) or WX (similarly, for regression of Wy on columns of neighborhood effects of independent variables).

With the combination of exogeneity and (full or partial) multiple imputation of missing data, significance tests are more efficient and comparable with respect to the sample size for a dependent variable in the context of regression analysis. Significance test biases can be tested with respect to the proportion of imputed missing data cases in the sample for a given variable (the more imputation, the more randomness in the imputations, with measurable bias toward less significance expected in the significance test). More importantly, relative $\delta y/\delta x$ effects and percentage coded-versus-imputed dummy variables (Part 3, e.g., Tables 2 and 4) can be evaluated independently as against statistical significance, so that significance tests should not be the sole basis for rejecting a variable in a statistical model for which the relative effect is larger than those of other variables in the model. Choosing variables in a regression model on the basis of the relative significance of independence variables in the second stage of 2SLS regression is a dubious practice that may lead to failure of model robustness when using multiple train-test statistics, for example, that test independent random subsample replications. Use of other inferential statistical distributions in model testing may offer similar advantages. In addition to 2SLS and 2SLS-IS models for single dependent variables an extension of findings about how different dependent variables are linked into networks of variables presents potential (but not always) soluble approaches to problems such as subsample sizes that differ (e.g., when the subsamples for which missing data are imputed vary with the cases coded for the dependent variable) or delimitation of systems that contain structural parameters that can be partially or completely estimated.

4 – 4 No Free Lunch: A coupled 2SLS-IS, Latent Variable and SEM approach to networks of variables

Software may provide help in statistical modeling but cannot provide solutions for causal or uniquely-most-robust models. Maximizing R^2 cannot be the only criterion for choosing among models because of the problem of overfitting. Nor can significance tests, which will vary in regression models according to the other variables in the model, size of the sample, reliability or bias in measurement or missing data imputation, and other factors. Relative $\delta y/\delta x$ effects suffer from all of these factors except for size of sample, which make them useful alongside significance tests.

More basic problems are those of ignored or unmeasured variables and *a priori* assumptions as to what can and cannot cause what, or direction of causality or effect. Full, expert, and/or intuitive understanding backed by correct logical deduction, knowledge of other relevant findings and cases, working closely with primary and coded data and problems of measurement, coding, and potential sources of bias cannot be replaced by software.

Good models, new data, and experimentation with alternatives often lead to new conundrums. In the discussions here of models of moral gods used as examples for evaluating 2SLS-IS models of effects, discovery that new variables such as AnimXbwealth and FxCmtyWages seemed to confirm the hypothesis that social dilemmas of excessive inequality could anticipate adoption or spread of prosocial moralizing gods or religious beliefs. The availability of a new control for autocorrelation (Routon and Eff 2001), a W matrix for religious similarity, offered an additional test of this hypothesis. Controlling for religiously similar neighborhoods through W matrix weightings in 2SLS provided an improvement in R^2 that wiped out the significance of the AnimXbwealth and FxCmtyWages variables in the model shown in Tables 2 and 5, but not other variables. This could reflect a classical problem of the confound between homophily (similar cases seek out ties of influence, regardless of distance) and network effects (cases with ties of influence become more similar). Most AnimXbwealth societies, for example, are Islamic, and inhabit similar environments in Eurasia. AnimXbwealth societies may have a greater propensity toward Islam but the effect is wiped out controlling for co-religious ties among them. Adopting Islam because of its commonality in the social environment might lead to adoption of features such as co-wives that make it easier to increase pastoralism and engage in exchanges of animals as bridewealth. Time-series data are required to discriminate among these kinds of possibilities.

5 – Conclusions

What is envisaged in further development of 2SLS-IS scripts or an R package that is useful in the social sciences and observational survey data analysis generally and that can connect to R SEM software (Fox 2006, 2009, Fox and Byrnes 2011, Rosseel 2011a) – and the mathematics that limit but in some cases enables causal inferences – is a gradient of research possibilities that utilize data, existing models and new simulations, and that build on substantive theory and plausible Bayesian priors. Model building is not a sole or definitive end, but can open paths to new insights and research. The one example explored here simply aimed to show that the 2SLS-IS R scripts replicate their prototypes in previous 2SLS scripts (Eff and Dow 2009) but go beyond them in important ways. We showed in an example that after adding a new train-test component that provided new distributions for inferential statistics about reliability and robustness, we

were better able to (1) show sources of weakness in the model tested by Eff and Dow and (2) use knowledge of the strengths and weaknesses to point towards ways to improvement. 2SLS-IS adds a relative effect (*reff*) measure, with support from inferential statistics, as auxiliaries to significance tests in regression models. Thus, even once endogeneities were largely eliminated by the use of Instrumental Variables in 2SLS, the 2SLS-IS scripts go beyond exclusive reliance on significance testing and other tests of the null hypothesis.

We found the Eff-Dow 2SLS model (“value of children”) to lack robustness (i.e., in replicating their results with 2SLS-IS inferential statistics), while the 2SLS-IS scripts applied to the Brown and Eff (2010) model (for “moral gods”) found weaknesses that could be resolved by a new and more robust model of our making. Replacing one of the independent variables – dependence on animal husbandry (*anim*) – with a modified variable (*animXbridewealth*) that captured the larger context of a subtype of socioeconomic exchange found in pastoral societies provided ways of introducing questions about dynamics into the model that could be investigated further. In this example, the opening of new questions illustrates the estimation of models not as an ending in themselves, aiming at conclusive results, but as an opening to new questions. It is in these kinds of developments that the “higher bar” of asking questions about causality, and the mathematical language of causal separation and interdependencies in networks of variables, may help to push us out of an empirical wasteland of reliance on model-building software to the reopening of more fundamental concepts in social science research whose efficacy can be tested with appropriate logic, empirical data, and methods suitable to specific projects.

Acknowledgements

The authors owe a great debt to Anthon Eff, Malcolm Dow, and Christian Brown for baseline software and model data (and White thanks Eff for generous help in learning advanced features of R). We thank Halbert White, Nihat Ay, Sander Greenland, Judea Pearl and John Fox for inspiration in the area of causal graphs and SEM; Scott White for graceful programming the new 2SLS-IS R code that re-factors the Eff-Dow (2009) R software and for the train-test algorithm and other suggestions of the principal author; and John Snarey for use of his SCCS HiGod4 variable, coded for the full 186 society SCCS in place of more Murdock’s comparable codings on a subset of 168. White is responsible for the methodology, with assist from Giorgio Gosti, while Tolga Oztan and Feng Ren were TAs and Instructors, testing the software with students over three quarters at UC Irvine and one short course at Xi’an Jiaotong University hosted by Prof. Haifeng Du. We thank the 100+ students who participated so effectively in these courses. We hold none of these contributors responsible for any misstatements made in this draft. We thank Jürgen Jost, Director of the Max Planck Institute (MPI) for Mathematics in the Sciences, for a generous invitation to our UC Irvine-based research group, including Feng Ren, Giorgio Gosti, and Tolga Oztan, to work at the MPI during the latter two weeks of June, 2011 and to make presentations to MPI students, faculty and researchers. We thank the Santa Fe Institute, and faculty director David Krakauer, for hosting us in similar working group seminars at SFI in August-September 2010 and 2011, in which Scott White and Feng Ren were also able to attend in 2010 and Gosti in 2011, along with Oztan and Elliott Wagner. We thank the Institute of Mathematical Behavioral Sciences at UCI and its faculty research group in Social Dynamics and Complexity for supporting the Human Sciences and Complexity videoconferences in which Hal White, Judea Pearl, Sander Greenland and members of Causality project at UCI gave presentations. Those talks can be seen in streaming video at <http://itunes.apple.com/us/itunes-u/human-sciences-complexity/id429669291>.

References - R Software for Regression with Inferential Statistics (2SLS-IS)

- Pg1Bib (the order of references will be sequential or alphabetical depending on the journal)
- Dow, James W. 2004. Using R for Cross-Cultural Research. *World Cultures* 14(2):144-154.
- Dow, Malcolm M. 2007. Galton's Problem as Multiple Network Autocorrelation Effects. *Cross-Cultural Research* 41:336-363. <http://ccr.sagepub.com/content/41/4/336.short>
- Dow, Malcolm M., Douglas R. White, and Michael L. Burton. 1982. Multivariate Modeling with Interdependent Network Data. *Cross-Cultural Research* 17(3-4): 216-245. <http://ccr.sagepub.com/cgi/content/abstract/17/3-4/216>
- Dow, Malcolm M., Michael L. Burton, Douglas R. White, Karl P. Reitz. 1984. Galton's Problem as Network Autocorrelation. *American Ethnologist* 11(4):754-770. <http://www.jstor.org/pss/644404>
- Dow, Malcolm M., Michael L. Burton, & Douglas R. White. 1982. Network Autocorrelation: A Simulation Study of a Foundational Problem in the Social Sciences. *Social Networks* 4(2): 169-200. http://eclectic.ss.uci.edu/~drwhite/pw/NetworkAutocorrelation_A_SimulationStudy.pdf
- Greenland, Sander. 2010. Overthrowing the tyranny of null hypotheses hidden in causal diagrams. Ch. 22 in: Dechter, R., Geffner, H., and Halpern, J.Y. (eds.). *Heuristics, Probabilities, and Causality: A Tribute to Judea Pearl*. London: College Press, 365-382. http://intersci.ss.uci.edu/wiki/pdf/Pearl/22_Greenland.pdf
- Henkel, Ramon, and Denton Morrison (Eds.). 1970. *The Significance Test Controversy: A Reader (Methodological perspectives)*. Chicago: Aldine.
- Pearl, Judea. 2009 (2nd edition). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.
- Rubin, Donald B. 1996. Multiple Imputation after 18+ Years (with discussion). *Journal of the American Statistical Association* 91: 473-489.
- Wikipedia: Galton's problem. 2011.
- Appendix1Bib
- Anselin, Luc. 1988. Lagrange Multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis* 20:1-17.
- Bivand, R., with contributions by L. Anselin, R. Assunção, O. Berke, A. Bernat, M. Carvalho, Y. Chun, C. Dormann, S. Dray, R. Halbersma, E. Krainski, N. Lewin-Koh, H. Li, J. Ma, G. Millo, W. Mueller, H. Ono, P. Peres-Neto, M. Reeder, M. Tiefelsdorf and D. Yu. (2009). *spdep: Spatial dependence: weighting schemes, statistics and models*. R package version 0.4-34.
- van Buuren, Stef, & Karin Groothuis-Oudshoorn. 2011. R Package 'mice' version 2.10.
- Pg2Bib
- Brown, Christian and Anthon Eff, 2010. The State and the Supernatural: Support for Prosocial Behavior, Structure and Dynamics: *eJournal of Anthropological and Related Sciences*, 4(1) art 1. <http://escholarship.org/uc/item/7cm1f10b>
<http://intersci.ss.uci.edu/wiki/pdf/eScholarshipBrown&Eff.pdf>
- Dow, Malcolm M, and E. Anthon Eff. 2009a. Cultural Trait Transmission and Missing Data as Sources of Bias in Cross-Cultural Survey Research: Explanations of Polygyny Re-examined. *Cross-Cultural Research*. 43(2):134-151.
- Dow, Malcolm M., and E. Anthon Eff. 2009b. Multiple Imputation of Missing Data in Cross-Cultural Samples. *Cross-Cultural Research*. 43(3):206 - 229.
- Eff, E. Anthon. 2008. Weight Matrices for Cultural Proximity: Deriving Weights from a Language Phylogeny. *Structure and Dynamics: eJournal of Anthropological and Related Sciences*. 3(2): Article 9. <http://repositories.cdlib.org/imbs/socdyn/sdeas/vol3/iss2/art9>
- Eff, E. Anthon, and Malcolm M. Dow. 2008. Do Markets Promote Prosocial Behavior? Evidence from the Standard Cross-Cultural Sample. MTSU Working Paper. <http://econpapers.repec.org/paper/mtswpaper/200803.htm>
- Eff, E. Anthon, and Malcolm M. Dow. 2009. How to deal with Missing Data and Galton's Problem in Cross-Cultural Survey Research: A Primer for R. *Structure and Dynamics: eJournal of Anthropological and Related Sciences*. 3(3): Article 1. <http://repositories.cdlib.org/imbs/socdyn/sdeas/vol3/iss3/art1>
<http://intersci.ss.uci.edu/wiki/pdf/eScholarshipEff&Dow2009.pdf>
- Eff, E. Anthon and Giuseppe Rionero. 2011. The Motor of Growth? Parental Investment and per capita GDP. *World Cultures eJournal*, 18(1). <http://escholarship.org/uc/item/5zh0t0q4>
- Little, Roderick, J. A., and Donald B. Rubin. (1987, 2010: 2nd Ed.). *Statistical analysis with missing data*. Wiley Series in Probability and Mathematical Statistics.
- Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.
- von Hippel, Paul T. 2007. Regression with Missing Y's: An Improved Strategy for Analyzing Multiply Imputed Data *Sociological Methodology* 37:83-117.
- Routon, P. Wesley, and E. Anthon Eff. 2011 forthcoming. Farming and Fighting: An Empirical Analysis of the Ecological-Evolutionary Theory of the Incidence of Warfare. *Structure and Dynamics: eJournal of Anthropological and Related Sciences*, 5(2) art 2.
- Rubin, Donald B. 1996. Multiple Imputation after 18+ Years (with discussion). *Journal of the American Statistical Association* 91: 473-489.
- Allison, Paul D. 2001. *Missing Data (Quantitative Applications in the Social Sciences)*. Sage.
- Wald, Abraham. 1943. Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large. *Transactions of the American Mathematical Society* 54: 426-482.
- Wayman, Jeffrey. 2003. Multiple Imputation With Missing Data: What is it and how can I use it? Paper presented at the 2003 Annual Meeting of the American Educational Research Association, Chicago, IL.
- Wooldridge, J.M. 2006. *Introductory Econometrics: A Modern Approach*. NY, NY: Thomson South-Western.
- Pg3Bib1

- Kelejian, H., & Prucha, I. 1998. A generalized two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics*, 17, 99-121.
<http://econweb.umd.edu/~kelejian/Research/P071897.PDF>
- Ord, Keith. 1975. Estimation Methods for Models of Spatial Interaction. *Journal of the American Statistical Association* 70:120-126.
 Pg3Bib2
- Dawid, A. P. 1979. Conditional Independence in Statistical Theory. *Journal of the Royal Statistical Society. Series B (Methodological)* 41:1-31.
- Fox, John. 2006. Structural Equation Modeling with the sem Package in R. *Structural Equation Modeling* 13(3): 465-486.
<http://socserv.socsci.mcmaster.ca/jfox/Misc/sem/SEM-paper.pdf>
- Fox, John. 2002a. An R and S-Plus Companion to Applied Regression. Thousand Oaks, CA: Sage Publications.
- Fox, John. 2002b. Robust Regression: Appendix to An R and S-PLUS Companion to Applied Regression.
- Fox, John. 2002c. Bootstrapping Regression Models: Appendix to An R and S-PLUS Companion to Applied Regression.
- Fox, John. 2009. The sem Package in R. <http://www.statpower.net/Content/312/R%20Stuff/sem.pdf>
- Fox, John, Jarrett Byrnes. 2011. CRAN - Package *sem*: Structural Equation Models. <http://cran.r-project.org/web/packages/sem/>
- Grace, Jim. 2009. Modeling with Structural Equations/About SEM Software. <http://www.structuralequations.com/software.html>
- Rosseel, Yves. 2011a. Package 'lavaan', lavaan - latent variable analysis. <http://lavaan.ugent.be/>
- Rosseel, Yves. 2011b. Package 'lavaan'. <http://cran.r-project.org/web/packages/lavaan/lavaan.pdf>
- Pg4Bib
- Hausman, J. A. 1978. Specification Tests in Econometrics. *Econometrica* 46:1251-1271.
- White, Halbert, and Xun Lu. 2010. Robustness Checks and Robustness Tests in Applied Economics. <http://bit.ly/kK2M8g>
- Hurwicz, L. 1962. On the Structural Form of Interdependent Systems. In, Patrick Suppes, and Amos Tverski, eds., *Logic, Methodology and Philosophy of Science*, pp. 232-239. Stanford, CA: Stanford University Press.
- Lukas, Robert. 1978. Economic Policy Evaluation: A Critique. In, K. Brunner, A. Meltzer, eds., *The Phillips Curve and Labor Markets*. Carnegie-Rochester Conference Series on Public Policy. New York: Elsevier, pp. 19-46.
- Pg5Bib
- Nakagawa, Shinichi, and Innes C. Cuthill. 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews: Cambridge Philosophical Society* 82 (4): 591-605. doi:10.1111/j.1469-185X.2007.00027.x. PMID 17944619.
- Duncan, Otis. D. 1961. A Sociometric Index for All Occupations. Pp. 109-138 in, A. J. Reiss, Jr., ed.), *Occupations and Social Status*. New York:Free Press
- Huber, Paul J. 1964. Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics* 35:73-101.
- Davison, A. C. & D. V. Hinkley. 1997. *Bootstrap Methods and their Application*. Cambridge: Cambridge University Press.
- White, Douglas R. 1993. Spatial Levels in Cultural Organization. *Handbuch der Ethnologie*, pp. 459-88, T.Schweizer, Margarete Schweizer, and Waltraud Kokot, eds. Berlin: Reimer Verlag.<http://eclectic.ss.uci.edu/~drwhite/pub/DRW1993Spatial.pdf>
- White, Douglas R. 2007. Standard Cross-Cultural Sample. *International Encyclopedia of the Social Sciences*, 2nd edition. Vol "S": 88-95. New York: Macmillan Reference USA. <http://intersci.ss.uci.edu/wiki/pub/IntlEncyStdCross-CulturalSample.pdf>
- White, Douglas R. 2012. Sccs_2SLS-IS downloads. http://intersci.ss.uci.edu/wiki/index.php/Sccs_2SLS-IS_downloads
- Pg6Bib
- Christakis, Nicholas A. and James H. Fowler. 2009. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*. New York, NY: Little, Brown and Company.
- Granovetter, Marc S. 1973. The Strength of Weak Ties. *American Journal of Sociology* 78(6): 1360-1380.
- Korotayev, Andrey, and Victor de Munck. 2003. Galton's Asset and Flower's Problem: Cultural Networks and Cultural Units in Cross-Cultural Research. *American Anthropologist* 105 (2): 353-358.
- Witkowski, Stanley. 1974. Galton's Opportunity-Hologeistic Study of Historical Processes. *Behavior Science Research* 9 (1): 11-15.
- Pg7Bib
- Barry, Herbert, III, Lili Josephson, Edith Lauer, and Catherine Marshall. 1977. Agents and Techniques for Child Training. *Ethnology* 16:191-230. Reprinted in Herbert Barry, III, and Alice Schlegel. 1980. *Cross-Cultural Samples and Codes*. Pittsburgh: University of Pittsburgh Press. <http://www.jstor.org/stable/3773387>
- Dow, Malcolm M.. 1984. Categorical Analysis of Cross-Cultural Survey Data: Effects of Clustering on Chi-Square Tests. *Journal of Quantitative Anthropology* 1(4):335-352.
- White, Douglas R., Michael Burton, William Divale, Patrick Gray, Andrey Korotayev, Daria Khalturina. 2009. Standard Cross-Cultural Codes. <http://eclectic.ss.uci.edu/~drwhite/courses/SCCCodes.htm> (Last updated May 2011. See variable 238.
- Murdock, George Peter. 1967. *Ethnographic Atlas*, Pittsburgh: University of Pittsburgh Press.
- Swanson, Guy E. 1960. *Birth of the Gods*. Ann Arbor, MI: University of Michigan Press.
- Pg8bib
- Chevan, Albert, and Michael Sutherland. 1991. Hierarchical Partitioning. *The American Statistician* 45(2): 90-96.
<http://www.jstor.org/stable/2684366>
- Grömping, Ulrike. 2006. Relative Importance for Linear Regression in R: The Package relaimpo. *Journal of Statistical Software* 17, Issue 1. Relative importance of regressors in linear models: <http://cran.r-project.org/web/packages/relaimpo/index.html>
<http://prof.beuth-hochschule.de/groemping/relaimpo/>
- Grömping, Ulrike. 2007a. Estimators of Relative Importance in Linear Regression Based on Variance Decomposition. *The American Statistician* 61, 139-147. <http://prof.beuth-hochschule.de/fileadmin/user/groemping/downloads/amstat07mayp139.pdf>
- Grömping, Ulrike. 2007b. Response to comment by Scott Menard, re: Estimators of Relative Importance in Linear Regression Based on Variance Decomposition. In: *Letters to the Editor, The American Statistician* 61, 280-284.

- Das Gupta, P. 1978. A General Method of Decomposing a Difference Between Two Rates into Several Component. *Demography* 15: 99-112.
- Holm, S. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6:65-70.
- Rice, William R. 1989. Analyzing Tables of Statistical Tests. *Evolution* 43(1): 223-225.
- Paige, Karen and Jeffrey Paige. 1981. *The Politics of Reproductive Rituals*. University of California Press.
- Robins, James M., Daniel O. Scharfstein, and Andrea Rotnitzky. 1999. Sensitivity Analysis for Selection Bias and Unmeasured Confounding in Missing Data and Causal Inference Models. *Statistical Models in Epidemiology: The Environment and Clinical Trials*. Halloran, M.E. and Berry, D., eds. NY: Springer-Verlag, pp. 1-94. <http://biosun1.harvard.edu/~robins/sasbci-fnl.pdf>
- Snarey, John R. 1996. The natural environment's impact upon religious ethics: a cross-cultural study. *Journal for the Scientific Study of Religion* 35(2):85-96.
- Pg9bib
- Hu, Li-tze, and Peter M. Bentler. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal* 6(1): 1-55.
- Kyono, Trent Mamoru. 2010. Commentator: A Front-End User-Interface Module for Graphical and Structural Equation Modeling. Technical Report R-364. Cognitive Systems Laboratory. Department of Computer Science. UCLA. http://ftp.cs.ucla.edu/pub/stat_ser/r364.pdf
- Pearl, Judea. 2011. The Causal Foundations of Structural Equation Modeling. Chapter for R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling*. New York: Guilford Press. Dept. of Statistics Papers <http://www.escholarship.org/uc/item/490131xj>
- Revelle, William. 2007. Using R for psychological research: A simple guide to an elegant package Structural Equation Modeling in R. [http://www.personality-project.org/r/\(r.sem.html\)](http://www.personality-project.org/r/(r.sem.html))
- Wright, Sewall. 1921. Correlation and causation. *Journal of Agricultural Research* 20: 557-585.
- Wright, Sewall. 1923. The theory of path coefficients: A reply to Niles' criticism. *Genetics* 8: 239-255.
- Wright, Sewall. 1934. The method of path coefficients. *Annals of Mathematical Statistics* 5: 161-215.
- Pg10bib
- Gibbs, J. Willard. 1902. *Elementary Principles of Statistical Mechanics*. New Haven, Connecticut: Yale University Press.
- Lauritzen, Steffen. 2011 (January). <http://www.stats.ox.ac.uk/~steffen/teaching/grad/graphicalmodels.pdf> Graphical Models. PPT in PDF. Graduate Lectures, Oxford.
- Wold, Herman O. A. 1954. Causality and Econometrics. *Econometrica* 22: 162-177.
- Pg11bib
- Vayda, Andrew P. and Bradley B. Walters. 2011. *Causal Explanation for Social Scientists*. Lanham, MD: Rowman and Little.
- White, Halbert, and Xun Lu. 2010. "Robustness Checks and Robustness Tests in Applied Economics." *Journal of Financial Econometrics*, 8: 193-243.
- White, Douglas R., B. Tolga Oztan, Giorgio Gosti, Elliott Wagner and John Snarey. 2012. Proximal Causality in the Adoption of Ethical God Beliefs. Web version 'Draft'
- White, Douglas R., B. Tolga Oztan, Giorgio Gosti, Elliott Wagner and John Snarey. 2012. Discovery of Hidden Variables for the Evolution of Ethical Religions. Draft paper. MoralDraft9: Nov 23 SFI:RENAMED
- White, Douglas R., B. Tolga Oztan, Giorgio Gosti, Elliott Wagner, and John Snarey. 2012. Multiple Hypotheses and Divergent Explanations: The Evolution of Moralizing High Gods. PREVIOUS DRAFT

Appendices

Appendix 1: Mathematical intuitions behind the 2SLS model

The $y = Wy + X$ model as developed in Dow (2007:347) and used by Eff and Dow (2009:13), in its full expression, is:

$$y = \rho_0 + \rho_1 W_1 y + \rho_2 W_2 y + \dots + \rho_j W_j y + X\beta + \varepsilon \quad (1)$$

This measures autocorrelation in y by inclusion of the Wy dependent variables (estimated from WX) where the W matrices are row normalized so that a matrix product $y \% \% W$ is a vector of the same dimension as y and WX a potential predictor of Wy . The result for a given observation is a “neighborhood weighted average” of each independent variable regressed on the dependent variable “neighborhood weighted average,” i.e., also from the values of X proximity-weighted by W for networks of societies. The weights reflect the proximities to neighbors.

A *simpler* model than equation (1), $y = X\beta + \varepsilon$, is not adequate because, if the observations are not independent, autocorrelation due to a missing Wy term, as in (1), may be displaced into loops of causality through network effects for some independent variables (X_1, \dots, X_m) that correlate with the error term ε . Here, through the linkages among the cases as measured by Wy , “a loop of causality such as network effects that link the independent and dependent variables of a model leads to endogeneity” (Wikipedia 2011: Endogeneity). Thus direct regression results for equation 1 will simply split endogeneity with ε between X and Wy , so the valid solution for equation 1 cannot be achieved. Thus solution of equation (1) requires two stages of OLS regression and two tests of independence with the error terms. The first stage is to solve for $Wy = WX + \varepsilon_W$ and a LaGrange Multiplier test of whether the ε_W residuals are uncorrelated with any of the WX . This gives a fixed estimate \hat{W}_y , dropping the error term. The second stage of OLS regression solves $y = \hat{W}_y + X + \varepsilon$ and allows a Hausman test of whether ε is uncorrelated with each X independent variable. Pg2Bib

For a *stronger* model than equation (1) it is possible to include the values of Wy in the proximity weighting, but this does maintain the distinction between dependent and independent variables and requires a different “endogenous” estimation. A *weaker* model for inclusion of autocorrelation measure Wy for the regression model in equation (1) is to require only that X and ε are *conditionally* independent with respect to Wy (in Dawid notation: $X \perp \varepsilon | Wy$) so that although endogeneity in equation (1) may remain for correlations between y and Wy or Wy and ε , so that *controlling* for Wy eliminates the $X \sim \varepsilon$ correlations, and satisfies *conditional* exogeneity given Wy .

A variable in regression is *exogenous* if it is uncorrelated or conditionally uncorrelated with the error term ε ; and only the effects of exogenous variables can be efficiently estimated in regression analysis. The solution for equation (1), then, in the *weaker* model, is to test whether conditional control given Wy creates *conditional* exogeneity ($X \perp \varepsilon | Wy$) for X . To achieve this goal for X and the error term requires a first-stage regression that predicts not the dependent variable y but makes an independent estimate of the control variable Wy in (1) from WX , i.e., from appropriate transformations of the independent variables to their neighborhood effects, given in equation (2).

$$\begin{aligned} Wy &= \alpha_0 + \alpha WX + \varepsilon' && \text{(estimate } \alpha_0, \alpha, \varepsilon') \\ \hat{W}_y &= \alpha_0 + \alpha_{ij} \sum_{j=1}^m W_j X_j + \varepsilon' && (2) \end{aligned}$$

Dropping the error term in equation (2) gives $\hat{W}_y = \alpha_0 + \alpha WX$ as in (2') below that represents the local W -neighborhood (proximity-weighted) averages of each $W_j X_j$ regressed on the $W_j y$. Thus, equation 2 represents a neighborhood-level regression of WX -neighborhoods on Wy -neighborhoods. The same W matrices are used for both X and y , so that if the network effects that produce autocorrelation in equation (1) are captured by the W networks, then WX and the error term ε' are more likely to be independent as measured by W in regression equation (2), an assumption that can be tested, e.g., as to whether $W_y \perp \varepsilon'$, which depends on whether the W_i matrices are well constructed to capture the network effects that produce nonindependence among cases (higher order effects like transitivity or clustering within W can be measured and added to the model, but efficient estimation of X may be achieved with an initial set.) Dropping the error term in equation 2 gives as a result an estimate of Wy that can be substituted into equation 1 (see equation 1' below).

$$\hat{W}_y = \alpha_0 + \alpha WX \quad \text{(drop } \varepsilon' \text{ from equation 2)} \quad (2')$$

Stage-two regression inserts the \hat{W}_y estimate into equation 1 to create 1'. (Because \hat{W}_y is an estimate, double primes (“”) are used for new coefficients in model 1', as they vary from those in equation 1.) Here, whether X is exogenous (uncorrelated with ε') is directly measured by a Hausman (1978) test of whether each $X_i \perp \varepsilon'$ in equation (1').

$$(1, \text{ substituting } WX \text{ for } \hat{W}_y \text{ from } 2') \quad y = \hat{W}_y + X\beta + \varepsilon''$$

$$y = \alpha_0 + \alpha WX + \beta''_0 + X\beta'' + \varepsilon''$$

$$y = \alpha_0 + \beta''_0 + \sum_{j=1}^w \alpha_j W_j X_i + \sum_{i=1}^v X_i \beta''_i + \varepsilon'' \quad (1')$$

The ε'' in equation 1', however, like that in equation (1), is simply a prediction error with no causal content for prediction from X to y. Thus the question of which X, controlling for autocorrelation, have *causal* effects on y, as opposed to statistical predictions, has not been solved. This reflects the weakness of choosing independent variables for a regression model based on higher significance tests and R² results alone.

Pg3Bib1

A Lagrange Multiplier (LM) spatial lag test, when nonsignificant, indicates no further improvement is possible in autocorrelation controls, given the W matrices used (Brown and Eff 2009:17, Anselm 1988, Bivand et al. 2009), as illustrated in Table 4. If this and other diagnostic tests are satisfied, then the β'' coefficients and other measures for the effects of variables X on y can be efficiently estimated. 2SLS estimates of this sort have been shown to converge to maximal likelihood estimates (so this result can also be checked), including accurate results for statistical significance. Ord (1975), for example, shows convergences between maximum likelihood estimation and various alternative models for autocorrelation. A two-stage model that gives additional autocorrelation controls for the error terms is given by Kelejian and Prucha (1998) and programmed using the IMSL program library. Maximum likelihood estimation is computationally challenging when the sample size is large.

Among the features to be added to the 2SLS-IS program are the stage-one (equation 2) regression results for each independent variable and the Hausman test for whether $X_i \perp \varepsilon$ and $\hat{y} \perp \varepsilon$. The idea here is that if the average neighborhood properties of y (Wy) and X (WX) have no further autocorrelation, then ε may be random and X in equation (2) exogenous according to the Hausman test.

Appendix 2: Inferential Statistics Tables

Table 5A shows Brown and Eff results using the 168 societies coded by Murdock for moral gods in the SCCS. The top nine rows of Table 5A show for distance and each independent variable the percentage of missing data, significance of each variable for the full sample (100%) and the 79%/21% random subsamples in ten runs of the 2SLS-IS train/test results. Model R^2 are given for each of the 79%/21% samples, with the larger samples averaging 65% higher ($R^2=.493$) than the latter ($R^2=.299$). Still, four of the variables are still significant in 40% or more of the reduced 21% samples, with only the PCsize, and Foodscarc variables dropping lower and thus considered to be the weakest variables. The last four rows of Table 5A show the best (Eextwar) and worst (Foodscarc) variables. The Agricultural potential variable (PCAP AgriPot.) had to be corrected (PCAP AgriPot.2 as the square of AgriPot), which increased its replicability but also slightly diminished the significance of other variables.

Table 5A: 2SLS-IS Inferential statistics showing 10 runs of 79% train and 21% test samples
source("examples/create/create_EduR_1/create_EduR_1.5DistBrownEff100v238.R")

Brown and Eff (2010) Variables	Missing data	100% Model p-value	79% train % p < 0.10	21% test % p < 0.12	# 1 p-values	# 2 p-values	# 3 p-values	# 4 p-values	# 5 p-values	# 6 p-values	# 7 p-values	# 8 p-values	# 9 p-values	# 10 p-values
Distance	0%	<0.00001	100%	100%	1x10 ⁻⁴	1x10 ⁻⁴	1x10 ⁻⁴	1x10 ⁻⁴	1x10 ⁻⁴	1x10 ⁻⁴	1x10 ⁻⁴	1x10 ⁻⁴	1x10 ⁻⁴	1x10 ⁻⁴
Low Eextwar	17.2%	0.025	100%	70%	0.14	0.21	0.03	0.17	0.02	0.07	0.07	0.04	0.06	0.05
Caststrat	.7%	0.034	100%	60%	0.03	0.09	0.11	0.17	0.09	0.22	0.07	0.08	0.24	0.22
PCAP AgriPot2	0%	0.012	100%	40%	0.61	0.10	0.33	0.17	0.30	0.09	0.16	0.08	0.12	0.37
PCsize	0%	0.038	80%	40%	0.19	0.19	0.08	0.41	0.03	0.22	0.18	0.03	0.21	0.12
Foodscarc	22.6%	0.016	90%	30%	0.09	0.26	0.16	0.11	0.10	0.35	0.14	0.25	0.45	0.21
Low PCsize.2	0%	0.011	90%	20%	0.18	0.22	0.21	0.28	0.02	0.42	0.17	0.11	0.22	0.05
PCAP AgriPot.	0%	0.012	90%	10%	0.04	0.61	0.16	0.16	0.16	0.35	0.54	0.37	0.67	0.13
79% R^2 is 65% higher than 21% $R^2=.36\%$				$R^2=.493$	$R^2=.49$	$R^2=.55$	$R^2=.49$	$R^2=.45$	$R^2=.45$	$R^2=.48$	$R^2=.54$	$R^2=.50$	$R^2=.50$	$R^2=.48$
Brown & Eff 2010 Dep_var=v238 N=168				$R^2=.299$	$R^2=.36$	$R^2=.05$	$R^2=.32$	$R^2=.45$	$R^2=.57$	$R^2=.29$	$R^2=.15$	$R^2=.12$	$R^2=.28$	$R^2=.39$
Foodscarc	5 p > 0.20			3 p < 0.13			Worst variable							
Eextwar	1 p > 0.20			6 p < 0.13			Best variable							
PCAP Agri.Pot.	5 p > 0.20			4 p < 0.13			First approximation							
PCAP Agri.Pot2.	4 p > 0.20 redone			4 p < 0.13 redone			Corrected PCA calculation following Brown and Eff (2010)							

Snarey (1996) independently coded all 186 SCCS societies for the moral gods variable, with perfect agreement for the 40 of Murdock's codes for presence of moralizing gods. He lent his codes to our project, so Table 5B shows the SLS-IS results for all 186 societies as coded for moral gods by Snarey, although results for the corrected by AgriPot.2 variable is not shown nor are the 79% train results. Foodscarc remains a weak variable, but now Eextwar is also a weak variable. This suggests a sampling bias for missing data, i.e., in cases not coded by Murdock for moral gods. Snarey's codings for Murdock's missing-data cases may have included more warlike societies not associated with moral gods.

Table 5B: 2SLS-IS Inferential statistics showing 10 runs of 79% train and 21% test samples

Brown and Eff (2010) Variables	Missing data	100% Model p-value	79% train % p < 0.10	21% test % p < 0.11	# 1 p-values	# 2 p-values	# 3 p-values	# 4 p-values	# 5 p-values	# 6 p-values	# 7 p-values	# 8 p-values	# 9 p-values	# 10 p-values	
Distance	0%	<0.00001	100%	100%	1x10 ⁻⁴	1x10 ⁻⁴	1x10 ⁻⁴	1x10 ⁻⁴	1x10 ⁻⁴	1x10 ⁻⁴	1x10 ⁻⁴	1x10 ⁻⁴	1x10 ⁻⁴	1x10 ⁻⁴	
Anim	0%	0.113		60%	0.35	0.04	0.03	0.09	0.15	0.06	0.02	0.59	0.03	0.49	
Caststrat	.7%	0.034		60%	0.13	0.09	0.32	0.06	0.14	0.07	0.13	0.24	0.12	0.14	
PCsize	0%	0.038		50%	0.03	0.05	0.18	0.08	0.41	0.28	0.04	0.24	0.13	0.62	
PCsize.2	0%	0.011		50%	0.04	0.12	0.34	0.12	0.40	0.06	0.07	0.36	0.36	0.21	
Foodscarc	22.6%	0.016	5 > 0.20	40%	0.09	0.27	0.21	0.03	0.07	0.21	0.20	0.08	0.22	0.50	
Eextwar	17.2%	0.025	3	40%	0.35	0.18	0.18	0.06	0.13	0.24	0.13	0.30	0.10	0.19	
PCAP AgriPot.	0%	0.012	5 > 0.20	20%	0.13	0.09	0.66	0.26	0.17	0.28	0.20	0.67	0.41	0.45	
27% higher				Aver. R^2	$R^2=.368$	$R^2=.31$	$R^2=.32$	$R^2=.40$	$R^2=.34$	$R^2=.30$	$R^2=.41$	$R^2=.37$	$R^2=.36$	$R^2=.37$	$R^2=.32$
Snarey'96 Dep_var=HiGod4 N=186				$R^2=.290$	$R^2=.44$	$R^2=.33$	$R^2=.05$	$R^2=.44$	$R^2=.45$	$R^2=.12$	$R^2=.23$	$R^2=.20$	$R^2=.25$	$R^2=.30$	
Average 79% $R^2=0.33$ for 10 runs					0.49	0.32	0.40	0.34	0.30	0.41	0.37	0.36	0.37	0.32	
Average 21% $R^2=0.29$ for 10 runs					0.44	0.33	0.05	0.44	0.45	0.12	0.23	0.20	0.25	0.39	

source("examples/create/create_EduR_1/create_EduR_1.5DistBrownEff100HiGod4.R")

source("R_3_s_ols/two_stage_ols.R") #source("R_3_s_ols/two_stage_ols_full.R") results are similar.

source("examples/create/R_run/run_model_79.R").

Former title:

Two-Stage Least Squares and Inferential Statistics (2SLS-IS) for Fast and Robust OLS in R

Douglas R. White, Ren Feng, Giorgio Gosti, and B. Tolga Oztan