Some New Regression Methods for Predictive and Construct Validation*

Stanley N. Rabinowitz
WestEd

David Rule
Marist College

Robert M. Pruzek
The University at Albany
State University of New York

September 1997

*Published in: *Social Indicators Research (Special Issue on Validity Studies*, Bruno Zumbo, Editor) 1998, 45, 201-231.

Some New Regression Methods for Predictive and Construct Validation

Both predictive and construct validation are essential to instrument development in all social and behavioral sciences. Ideally, both types of validation entail theoretical as well as empirical studies; moreover, the term validation implies a process that takes place over time, often in a sequentially articulated fashion. The choice of methods and methodology for empirical data analyses is of course central to the viability of validation studies.

In this article we shall describe and report on the empirical functionality of some modern methods for linear prediction, methods that appear to hold special promise for improving both the theoretical and empirical usefulness of validation studies in the social and behavioral sciences. Because *ordinary least squares* (OLS) methods are, far and away, the most popular forms of multiple regression, particular attention will be given to comparing the new methods with OLS regression.

As suggested, validation is never complete. The process is generally ongoing and sequential. It follows that each individual validation study should yield results that can be linked to those that came before and to those that will follow. The methods on which we focus accommodate both of these needs better than do conventional regression methods that provide no mechanisms for incorporating prior information. It will be shown below that our methods – which are adaptive – also tend to yield more stable results in small sample situations with many variables, compared to their OLS counterparts. This, in turn, suggests greater efficiencies may be possible in using what are often limited resources for empirical studies than has come to be expected based on conventional methods.

Cronbach and Messick (*cf*., Messick, 1989) distinguish between weak and strong construct validation. What mostly distinguishes these concepts is the degree to which the researcher can provide evidence of the form and type of expected interrelationships among variables defining the model in question. Although limited to linear systems, our methods help accommodate validation efforts across the weak-to-strong continuum. This is because the new methods can be used even if prior information is vague, as will often be the case at the early stages of research.

Guion (1980) discusses potential threats to the generalizability of validation studies. Primary among these is quality of the design and analysis of studies used to provide evidence of validity. The methods to be described and studied in the following pages tend to encourage coherence among goals of the validation process, beginning with the design, on through to analysis of data, and planning of subsequent efforts. In this sense, the new methods have potential to increase the "dependability and extendibility of research inferences" (Messick, 1989, p. 57) by reducing threats that diminish chances of this occurring (*i.e*., sampling and measurement error).

Finally, according to Messick (1989), validation should be viewed as a unitary process such that traditional content and criterion-related validity fall within the general construct validity umbrella. The logic of the newly developed methods in the context of design and analysis seems consistent with this view. We shall argue that results from studies that incorporate certain new methods can be expected to be interpretable in terms of the full range of relationships among the variables thought to operationalize the relevant constructs.

Limitations of OLS Regression

Numerous regression methods are of course available to researchers interested in studying relationships among several prospective predictors and one or more criterion variables (*c.f.* Darlington, 1990; Draper and Smith, 1981; and Rozeboom, 1966). Developed over many years, such methods can be used to answer questions concerning how to choose predictor variables that satisfy various standards of judgment. There have been a number of statistical and psychometric developments in recent years that bear on such questions. It follows that that researchers aiming to do validation studies must make a series of decisions, often among complicated alternatives, when selecting among available regression methods. It can be particularly difficult to do effective studies in typical social science situations where, there are – potentially – many variables, when their reliabilities are at least to some extent limited, and when sample sizes may be less than desired.

Although historically OLS regression methods have been commonly taught and continue to be heavily used, there are excellent reasons to be skeptical about these methods in validity studies. However, the problems brought about by the inadequacy of OLS methods have led many researchers to a variety of inadequate solutions. In a major paper that summarized many of the issues, Smith and Campbell (1980) – and their discussants – examined some of these attempts, which resulted in:

- concluding that the problems cannot be overcome; develop and analyze only simple models;
- ignoring the problems (either tacitly or explicitly) and crunch away at the data with such methods as stepwise regression that "disguise the imprecision of their estimates" (Smith and Campbell, 1980, p.75);
- selecting only a small number of nearly orthogonal predictor variables; or
- developing composite variables either through statistical approaches such as principal components analysis (Mosteller and Tukey, 1977) or more "logical" means such as combining test items into scales or total scores.

The first two of these are fatalistic and forms of denial. Each of the two latter approaches also is problematic. Because there are numerous prospective predictors in typical social science prediction studies, it is particularly important that investigators have available methods that can accommodate many variables. Reducing the number of predictors through the use of composites is widely taught, and frequently recommended. However, this step is not necessarily either wise or helpful. Composites can lead to confounding of interpretations of derived regression weights. In particular, composites can result in masking of relationships between individual predictors and criterion variable(s), and can lead to reductions in the predictive value of the system of constituent variables, in relation to what may be possible with more comprehensive approaches to using predictive information.

Another variation on OLS methodology features reducing the number of predictors by examining all subsets of predictors in relation to a criterion variable. But, while all subsets methods are generally acknowledged to constitute an improvement over hierarchical methods, to examine all subsets is so computationally intensive that the strategy is not feasible when there are very large numbers of predictors. Furthermore, all subsets methods generally make no allowances for measurement errors in predictor variables. Both predictors and criterion variables are nearly always non-ignorably fallible in social science applications.

Alternative methods proposed to ameliorate problems associated with OLS methods include ridge regression (Hoerl and Kennard, 1970) and reduced rank regression (Lawley and Maxwell, 1973; Pruzek and Frederick, 1978). Campbell and Smith discuss problems associated with use of ridge regression. Although reduced rank (RR) procedures do accommodate these needs, such methods are strongly dependent on the ability of a particular common factor model to soundly represent the joint predictor/criterion covariance matrix. In contrast to RR methods, OLS regression makes no provisions for an analyst to incorporate prior beliefs about variable interrelationships into data analysis, nor do OLS methods take measurement errors into account.

In the following section we briefly summarize WSR methods; next, we shall examine the applicability of these methods to validation processes, and go on to discuss the role of prior information in validation studies. Finally, we shall describe simulation studies that will help quantify the effects of using various WSR methods when they are applied to the same prediction problems.

Weighted Structural Regression

Recently, a new class of linear methods called Weighted Structural Regression (WSR) has been developed (Pruzek and Lepak (1992)) that subsumes OLS, ridge, and reduced rank regression, as well as many other cases, within its general framework. Most notably, certain forms of these new methods are adaptive, in that WSR methods allow one to incorporate prior information about variable interrelationships into analysis in a flexible fashion. When used in their adaptive forms WSR methods make it possible to capitalize on prior information (only) to the extent that it is supported by empirical data. When one's data strongly support the prior model, derived WSR estimators may be (far) more

generalizable than their OLS counterparts. When the prior model is not supported at all, the fallback is to OLS regression. A particular attraction of WSR methods for validation studies is that structural models can be used that account for measurement errors, as well as prior beliefs about structures. Accordingly, WSR methods may be expected to lead to enhanced interpretations of results.

Although WSR procedures can in principle accommodate a wide range of population covariance structures, the discussion here will be limited to the use of prior models of common factor form. WSR based on such models are both conceptually and computationally simple to implement, and may not be particularly limiting in the context of instrument validation. Other linear models that fall within the WSR framework include multiple dependent variables models, or those that specify relationships among error terms, as often used in structural equation modeling studies. Our discussion here is consistent with, but much more abbreviated than that of Pruzek and Lepak (1992).

For any set of k predictor variables and a criterion OLS regression equations can be simply derived from the joint covariance matrix of all observed variables. In particular, if y designates the criterion variable and x the set of k predictors, then the (1 | k) symmetrically partitioned covariance matrix

$$(1)\quad \mathbf{C} = \begin{bmatrix} \mathbf{c_{yy}} & \mathbf{C_{yx}} \\ \mathbf{C_{xy}} & \mathbf{C_{xx}} \end{bmatrix}$$

can provide the basis for generating $\boldsymbol{\beta}_{y.x}$, the k x 1 vector of ordinary least squares regression coefficients. Specifically,

$$(2)\quad \boldsymbol{\beta}_{y.x} = \mathbf{C_{xx}}^{-1}\,\mathbf{C_{xy}}$$

where $\mathbf{C_{xx}}$ represents the k x k covariance matrix for the independent variables, $\mathbf{C_{xy}}$ is the vector of k predictor-criterion covariances ($\mathbf{c_{yy}}$ represents the variance of the dependent or criterion variable).

Given the same joint predictor-criterion covariance matrix **C**, suppose a common factor analysis is completed. Now, let $\mathbf{A}_m$ denote the matrix of common factor coefficients, and $\mathbf{U}^2$ depicts the corresponding diagonal matrix of uniqueness variances (*cf.* Lawley and Maxwell, 1971). Then the *m*-factor approximation to the covariance matrix **C** can be written as

$$(3)\quad \mathbf{C}_m = \mathbf{A}_m\,\mathbf{A}_m' + \mathbf{U}^2 .$$

WSR entails "combining" the observed (sample) matrix **C** with its counterpart $\mathbf{C}_m$. That is, WSR procedures define a particular convex sum of the model free estimator of the joint predictor-criterion covariance matrix and a common factor estimate associated with this same covariance matrix, written as

$$(4)\quad \mathbf{C}_{mw}^{*} = w\,\mathbf{C} + (1-w)\,\mathbf{C}_m ,$$

where **C** is the conventional (maximum likelihood [model-free]) estimator of the *joint* predictor-criterion population covariance matrix; $\mathbf{C}_m$, defined in Expression (3), is the *m*-factor common-factor approximation for the same population covariance matrix; and w is a scalar in the interval (0,1), usually defined in terms of n and another scalar, γ_m; viz., $w = n/(n+\gamma_m)$. Pruzek and Lepak (1992) discuss the rationale for defining $\gamma_m = (p(1+ r_m)-2)/(p - r_m)$, where p is the number of variables and r_m is a function of the $p - m$ smallest ("rejected") eigenvalues for the *m*-factor model. The function r_m is defined as follows: let $a_m = \Sigma e_j$ for j=m+1 to p; and $b_m = \Sigma e_j^2$, also for j=m+1 to p. Define $r_m = a_m^2/b_m$. [*see footnote * below*]

Thus, WSR regression weights can be written simply as the vector

$$(5)\quad \boldsymbol{\beta}_{y.x(m,w)}^{*} = \mathbf{C_{xx}}^{*-1}\,\mathbf{C_{xy}}^{*},$$

where each term has the same form as Expression (2), except now the asterisks mean that the

---

* RP notes that William W. Rozeboom (Professor Emeritis, University of Alberta) wrote a substantial paper (unpublished) concerning properties of r_m. He found this function to be most interesting. His work makes it obvious that it is r_m, not γ_m, that does the work in WSR; we would like to reopen studies to extend his work in the WSR context. A copy of WWR's paper available from RMP on request.

constituent elements of Expression (5) derive from the partitioned form of the matrix in (4). WSR weights are explicitly dependent on the choice of m, the number of common factors; w, in turn, depends on the scalar γ_m, which further depends on r_m.

If the scalar w is defined a priori (as unity for OLS regression, as zero for RR regression) then the WSR method is said to be non-adaptive. When w is derived adaptively, as in $w = n/(n+\gamma_m)$, where γ_m is estimated (in what follows, non-iteratively), then WSR weights in (5) reflect how well the m-factor model fits extant data.

It should also be noted that if the number of common factors is set at zero, a priori, then the general from of $\mathbf{C}_m$ in Expression (3) is that of a diagonal matrix; this case corresponds to ridge regression. Of course several methods of common factor analysis could be used to form the m-factor approximation $\mathbf{C}_m$, and recent work has been particularly instructive about what seem to be effective prospects for these in different applications. Three alternative ways of generating w adaptively are noted later, in the context of simulation studies. Pruzek and Lepak (1992) provide relevant technical details. Moreover, software to implement such methods is also available from one of us (RP).

Some further technical information may help explain the underlying logic of the adaptive system. When a common factor model is used to generate the model-based estimate in Expression (3), then the computation of γ_m, and hence w, is based on the set of $p - m$ rejected eigenvalues of the (scaled) joint predictor/criterion covariance matrix. To the extent that these smallest eigenvalues are relatively homogeneous, the structural model is supported; this is because variance in these smallest eigenvalues, which is summarized by r_m, simply reflects lack of fit of the off-diagonals of $\mathbf{C}_m$ to the matrix $\mathbf{C}$. When an m-factor model fits well, w becomes small, so that WSR weights derive mostly from parameter estimates associated with the matrix $\mathbf{A}_m$. Particularly when m is relatively small, WSR weights will tend to be interpretable in terms of common factors; furthermore, the stability of WSR predictor weights will then tend to exceed that of their OLS counterparts.

Applicability of WSR Methods to Validation Processes

The use of adaptive WSR methods for validation studies, paired with an effective approach to planning and design of prediction batteries, seems to offer special potential for enhancing both interpretability and generalizability of derived results. As Messick (1989) notes, the validation process involves testing of a theoretical rationale with evidence obtained from applications of the measure(s) of interest. At any point in the validation process adaptive WSR methodology facilitates making connections between past theorizing and results of current empirical studies. Indeed, standard WSR methods generally require an investigator to articulate his or her theoretical expectations in the form of a prior structural model. It follows that WSR methods should help encourage investigators to inform themselves comprehensively about characteristics of their measures, as well as the respondents to whom they are to be administered.

Nevertheless, these particular structural regression methods enable the investigator to incorporate *vague* prior structural models, ones that can perhaps usually be justified in situations typical of those in which OLS regression is most commonly used. It is in just such situations that exploratory common methods seem most appropriate. (At relatively more mature stages of a validation process, more confirmatory structural models may be warranted; WSR methods can in principle be seen as relevant in such situations, but possibly with confirmatory models replacing exploratory common factor models.)

WSR methods can be described with reference to a general Bayesian perspective within which prior beliefs are incorporated in a straightforward way in most applications. Because adaptive WSR methods treat predictors as random variables, these methods seem especially appropriate in validation studies that seem amenable to use of multiple linear regression. In relation to conventional regression methods, WSR also makes more realistic assumptions in the way that they accommodate prior information, as well as variable unreliabilities. As noted above, adaptive WSR methods use prior information in such a way that the prior model is weighted according to how much support it

has from extant data.

Many decisions must be made in selecting predictor variables in a particular validation study. It is becoming increasingly clear, however, that one's prior beliefs play an especially important role in determining if a study is to be relatively successful. WSR methods draw attention to beliefs a researcher holds about various variable interrelationships, and how this information can be used to enhance usefulness of empirical data through comprehensive analysis. The key point, again, is that one must have available methods of analysis that can incorporate this information, even when it may be somewhat vague or diffuse.

Successful prediction studies combine the reliable information contained in the predictor set with a method of analysis capable of using this information. In many, perhaps most, behavioral applications, numerous predictors are either available or could, with some effort, be made available. WSR methods tend to encourage use of as many predictors as may be needed to "cover" the reliable criterion variance. Unlike conventional regression methods, those based on adaptive WSR algorithms tend not to break down as the number of predictors is increased, even if the predictors are mutually intercorrelated. For this reason, use of WSR methods tends to ameliorate the multicollinearity problem.

In our applications it has been interesting to learn that model-derived weights, that may often "look" quite different that OLS weights, often "work" nearly as well as "optimal" OLS weights, and are quite interpretable. The latter weights necessarily achieve the maximum possible multiple R^2 between the set of predictors and the criterion variable. But OLS weights routinely have poor generalizability properties, compared to weights derived by methods that incorporate prior information in analysis. In particular, WSR weights tend to generalize better to successive samples, meaning that they tend to exhibit less sampling variability than do their OLS counterparts.

In general, the suitability of a method of analysis depends on the match between the researcher's prior assumptions and the actual conditions represented by the obtained data (Laughlin, 1986). Also relevant is just how suitability is assessed. In regression studies, the focus of interest may be on a theoretical interpretation of the various weights assigned to a set of predictors relative to one or more criteria; or more practically, on the ability of the set of regressors to predict criterion scores accurately in some future application. We shall attend to both such issues in the simulations below.

Role of WSR in the Validation Process

Although most investigators appear to have been taught that factor analytic methods and regression methods are distinct and independent from one another, the introduction of WSR methodology suggests that these two methods are two sides of the same methodological coin. When using WSR methods for validation, it is feasible to treat the problem of choosing or selecting effective measures of constructs at the same time as such constructs are examined in relation to the criterion variable(s). In large, comprehensive, batteries communality estimates for individual predictor variables may be treated as reliability measures. Explicitly, derived factor coefficients for the joint predictor-criterion battery, typically following some form of transformation, can serve to facilitate naming of measures based on empirical relationships among all variables. Consequently, such factor coefficients can facilitate development of a theoretical network that is consistent with data in the context of construct validation.

On the other hand, criterion-related validity is generally viewed as falling within a regression framework. Researchers frequently appear uninterested in examining relationships among variables relative to the obtained regression weights, and thus ignore the question of why one predictor may receive a large (small) weight compared to another theoretically similar predictor. By their nature, OLS methods tend to focus on the magnitude of the squared multiple correlation between predictor sets and criteria, and examination of which variables contribute "significantly" to the criterion. But to focus narrowly only on such questions is to ignore the larger question of what variables to commence with in a given regression study, including individual test items, subtests or composites, or other (linear) combinations.

It seems important to consider the underlying structure of the most basic predictor variables available, and when possible, also to examine criterion validities in relation to these variables. In validation studies it is most helpful to think in terms of the prediction with reference to constructs associated with the domain of variables under study. What the conventional perspective often ignores is the notion that *both* the specific predictors and criterion variable employed represent just a sample of many possible measures that might have been used to represent the underlying construct(s) the researcher is attempting to use in prediction, or to explain. Use of WSR methods reinforces consideration of constructs in general prediction situations. In addition, WSR methods can help by facilitating thinking about latent (underlying) relationships, as well as reliabilities of predictor variables, in the context of one or more criterion variable.

Factor-based models tend to impose only loose or exploratory models about structural relationships among variables or constructs they may represent. But the adaptive feature of WSR methods, based on the concept of diffuse prior information, can reduce the importance of the specific models chosen, which should be especially beneficial in the common validation situation where prior information tends to be vague. It is anticipated that WSR methods may be quite useful even at the earliest stages of a series of validation studies. However, we do not argue that common factor methods cover all potential needs in the context of providing prior structural information or knowledge. Rather, we suggest that use of adaptive WSR forms, perhaps based on common factor starting points, can improve on work-a-day practice in the context of validation studies for the reasons that have been given. Such methods are regarded as a prospective beginning, not an end of methodological improvements in the context of validation studies.

Of special relevance to those using WSR in validation studies is that a particular advantage accrues to the theorist whose prior or advance knowledge is greatest at the time of composing the prediction/criterion battery. Such an individual entertains a good chance of being able to compose this battery so that relatively few factors, *m*, will be sufficient to explain or account for all off-diagonal elements of the joint covariance matrix. This, in turn, will help make the scalar γ_m relatively large, so that *w* will be relatively small. It is just this prospect, that of making the vector of WSR weights for the predictors derive mostly from the common factor coefficients in $\mathbf{A}_m$, that offers greatest hope for major improvements in interpretability and generalizability in contexts of empirical validation.

Simulation Methods

Two population systems are studied here; each was derived from real data. Although simulated data are used exclusively in the studies reported below, certain procedures were followed to help ensure realism, and to provide a distinctive challenge in the context of comparing the various regression methods. These steps included using real data sets of varying complexity as the basis for the simulation, as well as testing methods across a range of sample sizes. After the initial simulation results are presented, a further side study is reported to examine the robustness of these methods to a realistic level of outlier contamination. Methods for the outlier study will be described briefly after the first part of the simulation study has been presented.

Population Data Sets: The correlation matrices used to simulate data in this study were chosen to vary in the complexity of relationships among predictors. A second facet, sample size, was included to test its possible interaction with the other characteristic with respect to one or more of the evaluative criteria.

Two population data sets were selected: 1. Multi-trait-Multimethod Matrix (MTMM) (Campbell and Fiske, 1959) and, 2. Hauser's data (Hauser, 1966). These populations consist of ten, and 12 variables, respectively. They differ in their relative conformity to the common factor model, and in the relative predictability of the variables among themselves.

As noted above, all variables are assumed to be random, befitting the typical condition for observational research in the behavioral sciences. By considering all variables to be random, the

predictability of the overall system for each data set can be examined as a whole, in addition to the specific individual criterion variables selected for each data set.
Given each population, 100 samples of size n = 35, 70, and 140 were simulated using *normal* bootstrapping procedures (Efron and Tibshirani, 1986). This method entails generation of one hundred multivariate normal samples of size n, for each population; each sample is derived to be stochastically consistent with the respective population (correlation) matrix. However, for each sample, the regression analyses were begun from a sample variance-covariance matrix.

In more detail, the population sets used as the basis for simulation were:
Multitrait-Multimethod (MTMM): Campbell and Fiske (1959) discuss a 15 variable data set of MTMM form consisting of five trait and three method factors. As Campbell and Fiske indicate, the data fit the MTMM model loosely, particularly for the third set of method variables (peer ratings) for which correlations were relatively low. To make the problem more manageable for this work, without sacrifice, the latter five peer rating variables were dropped. As may be apparent from examination of Table 1a, this ten variable set creates a reasonable test for the factor based methods, particularly for low values of m.
Hauser: The Hauser population derives from relationships among twelve sociological variables and a sample size of 3427. The criterion variable for this system was chosen by the author (Hauser, 1966) and that is used here as well. The original predictor data was gathered in 1957 and consists of four measures of socioeconomic status, two measures of academic ability, three measures of student's perceptions, and two measures that addressed student's plans following high school. Criterion variable data (level of educational attainment) was collected in 1964. The Hauser population correlation matrix, SMCs, and eigenvalues are presented in Table 1b. These variables appear to constitute a fairly well defined two-factor system with only moderately high SMCs (and communalities).

Methods of Analysis

For each combination of data set, and sample size, the following regression methods were used to compute regression weights. In each case where a common factor method was used in these simulations, a variant of Joreskog's (1969) image factor analysis was employed.
(See Pruzek and Lepak (1992) for more details about these methods.):
1. Ordinary Least Squares (OLS): OLS is equivalent to a MR solution (see below) with $w = 1$.
2. MinRisk Ridge (Ridge), $m = 0$: the model-based information becomes diagonal in form, representing the ridge (biasing) diagonal.
3. MinRisk (MR), m = 1,2,3: Here models of common factor form were first used to obtain the so-called model-based estimator of the covariance matrix; see Expression (4) above. For MR, $w_m = n / (n + \hat{\gamma}_m)$; γ_m reflects how well the common factor structure fits extant data.
4. MinRisk*2 (MR2), m=1,2,3: This is a modified form of MR where $w = n / (n + 2\gamma_m)$, so that more weight is given to the common factor model in the convex sum, with respect to (3);
5. GFI: This method uses matrix traces, where $w = (d_1 / d_2)(\mathrm{tr}(\mathbf{C}_m{}^{*-1}\mathbf{C} - I)^2 / \mathrm{tr}(\mathbf{C}_m{}^{*-1}\mathbf{C})^2)$; (d_1 / d_2), based on degrees of freedom, equals $p(p+1) / ((p-m)^2 + p - m + 2)$ (*cf.* Jöreskog & Sörbom, 1986). GFI usually results in greater weight to the prior structural model than either MR or MR2;
6. Reduced Rank (RR): This method sets $w = 0$, a priori, in Expression (4). In this case RR regression entails use of a small sample version of image factor analysis suggested by Pruzek and Lepak (1992). The same common factor method was used for methods MR and MR2.

Criteria for Comparison of Methods:
This study examined both the reproducibility of regression weights and cross validation squared errors of prediction. Specifically, the methods were compared with the following evaluation criteria. (See the Pruzek and Lepak reference for more detailed descriptions.)

1. *Mean Squared Errors (MSEs):* these are the average squared differences between sample regression weights from the 100 normal bootstraps and population OLS counterparts;
2. *Predictive Mean Squared Errors (PMSEs): PMSEs,* or more accurately, the estimable parts of the predicted mean squared errors, are analogs of the *MSE*s, except *PMSE*s entail use of a weight matrix in computation. *PMSEs* are useful in that they indicate how well (actually, poorly) estimated sample-based regression estimates of the criterion variable (*y*-hats) should work when they are substituted for their population counterparts;
3. *Cross Validation Average Squared Errors (Cross Valid)*: These indices describe how much error is associated with use regression weights from each normal bootstrap sample, applied to a second (independent) normal bootstrap sample. *Cross Validation Avgerage Squared Errors* summarize how (poorly) weights holds up in an arbitrary random sample of the same size drawn from the same population.

Results

Simulation results for both population systems are presented. For each system, summaries are provided for the specific designated criterion variable, as well as the overall system (each variable predicted from p - 1 others). Simulation results are presented for n = 35, 70, and 140, and for values of m = l, 2, and 3. Use of all three values of m helps show how model-based methods work across a range of ms, to facilitate generalizations to other real data problems. No subject-specific interpretations are discussed for the data sets since the focus is on the fit of the different common factor models to the data under varying conditions of interest (sample size x population complexity).
Multitrait-Multimethod (MTMM): Table 2a summarizes *MSE*s, *PMSE*s, *Cross Validation Average Squared Errors* and *GOF* indices (1 - *w*) for the designated criterion variable for the MTMM data set. Table 2b presents this same information for the "overall" system. A few further comments will be made about the simulation results, beyond those that can be discerned from these tables, as seems needed for a comprehensive assessment of results.
MSEs: For this population, with n = 35, MR2 and GFI solutions perform the best, particularly for m = 2 with *MSE* values approximately one-third as large as those obtained from OLS. For the smallest values of n in particular, the improvements over OLS are dramatic! Ridge regression weights were nearly as poor as those from OLS, although the standard deviation of their obtained *MSE* values (not presented here) were substantially lower. RR *MSE*s improved greatly as m increased from one to three for all sample sizes, but only begin to be reasonably small for m = 3.

For n = 70, the MR and MR2 methods again outperformed OLS. GFI weights did less well with this larger data set. MR2 gave the best solution with *MSE*s roughly one-half as large as their OLS counterparts. Ridge weights worsened relative to all the other methods, with *MSE* values 40 percent larger than those for OLS. RR once again improved as m was raised from 1 to 3.

With n = 140, MR2 *MSE*s remained the best, approximately 50 percent better than those from OLS and 100 percent better than ridge values. The standard deviations of the MR2 *MSE*s remained less than 60 percent as large as those for OLS. Little difference was found across values of m from 1 to 3. GFI *MSE*s varied significantly across the values of m, as did those for RR; the MR and MR2 results did not change greatly over the three values of n. The relatively high *MSE*s for this specified criterion variable reflect the relative lack of fit of any low rank common factor model to this data set.

For the overall system, referring to Table 2b, *MSE*s remained high for this data set compared to both the circumplex and Hauser data. OLS weights were more than three times as large as those from MR2 and GFI (with standard deviations five times as large), with n = 35. MR2 was clearly the best method for both n = 70 and 140. GFI did less well at these larger sample sizes. RR remained mostly unaffected by increases in the sample size though raising m from 1 to 3 did improve RR significantly. OLS *MSE* standard deviations remained high relative to the other methods.
PMSEs: For n = 35, for the specified dependent variable, OLS *PMSE*s were nearly double those from MR and MR2, with standard deviations (not presented) 50 percent larger. For n = 70, OLS *PMSE*s remained larger by about 25 percent compared to both MinRisk methods. Even at n = 140, OLS

values were still 20 percent larger than their adaptive MR counterparts. Neither ridge nor GFI did as well as either MinRisk method, though both out-performed OLS. The *PMSE*s for RR were notably higher than for the other methods, especially for smaller value of *m*.

As seen from Table 2b, GFI did as well as MR and MR2 for *n* =35, for the overall system. Otherwise, a similar pattern was found (for *n* = 70 and 140) as described for the designated criterion variable.

Cross Valid: For this index, for both the specified dependent variable and the overall system, rather minor differences were found among methods. However, results again favored MR, MR2, and GFI for *n* = 35 compared to OLS (by roughly 10 percent for the average values, 20 percent for the associated standard deviations). For this criterion, and this population, most differences among methods were no longer discernable for *n* = 70 and 140. RR solutions were the worst for this criterion, particularly for smaller values of *m*. For *n* = 70 and 140, changes in *m* had little effect on the convex sum methods.

Goodness of Fit (1-w): Goodness of fit values, which merely reflect how much weight was given to the *m*-factor model in each adaptive WSR application, were relatively low for this population. For *n* = 140, the model was weighted from five to just below 20 percent for *m* = 3. Even for *n* = 35, the highest weighting (for *m* = 3) was only .385.

Hauser: Tables 3a provide *MSEs, PMSEs*, *Cross Validity Error* indices and GOF indices (1 - *w*) for the designated criterion variable for the Hauser data. Table 3b provides this same information for the overall system (excluding the goodness of fit index which is only relevant for a specific criterion variable).

MSEs: The pattern described above for the circumplex data set continues for the Hauser set, though with more dramatic differences between the conventional OLS methods and their adaptive WSR counterparts. For the specified criterion variable, with *m* = 2 (the best solution across sample sizes), MR2 and GFI methods, for *n* = 35, perform as well as OLS for *n* = 140! For *n* = 35, OLS *MSE* standard deviations are seven times as large as the MR and GFI counterparts. Even for *n* = 140, the OLS *MSE*s are twice as large as those obtained from MR an MR2 and the OLS standard deviations remain two to three times as large as the convex sum analogs.

Ridge regression outperformed OLS at all sample sizes for the Hauser system, but Ridge produced *MSE*s 36 percent larger than those from MR2 with *n* = 35, and 67 percent larger when *n* = 140. Reduced rank did particularly well for *n* = 35; however, with the extra data available at *n* = 140, RR no longer outperformed the convex sum methods, particularly MR2. GFI also slipped behind both MR and MR2 for *n* = 140, although, as expected, it outperformed RR for the Hauser system.

For the overall system, the same general pattern was found. GFI performed best with *n* = 35; MR2 passed GFI for *n* = 70 and outperforms it by nearly 50 percent for *n* = 140. The ratio of OLS *MSE* standard deviations to those from the convex sum methods varies from about 2.2 for *n* = 140, up through 6.3 with *n* = 35.

PMSEs: For the Hauser problem, with *n* = 35, OLS *PMSE*s are approximately four times as large as those from MR2 and GFI with a standard deviation also four times the size as their convex sum counterparts, for the specific criterion variable. With *n* = 70, the ratio drops to 3:1 with standard deviations more than twice as large. OLS *PMSE*s for *n* = 140 are still 20 percent larger than those from MR with a standard deviation also 20 percent larger.

As with *MSE*s, GFI and MR2 *PMSE*s are comparable for *n* = 35. Increasing the sample size to 140 decreased the overall value of the *PMSE*s for all the methods. For the overall system, MR2 and GFI *PMSE*s were similar for *n* = 35. Compared to *PMSE*s from OLS solutions, MR and MR2 values averaged less than one-third as large. With *n* = 140, MR and MR2 continue to notably outperform all other methods, including GFI.

Cross Valid: For the specific criterion variable, all non-OLS methods outperformed OLS for *n* = 35 by approximately 30 percent using this standard of comparison. With *n* = 70, the improvement dropped to ten percent for MR, MR2, GFI, and RR. For *n* = 140, average cross validation errors were

quite similar across all methods. A similar pattern was found for the overall system. It is notable that RR, which was found to reproduce population weights especially poorly, especially for $m = 1$, did better than OLS for the smallest n, even for $m = 1$ when $n = 35$. In general, it appears that *Cross Validation Error* does not provide a sensitive criterion for comparing regression methods.
Goodness of Fit (1-w): Hauser goodness of fit values were relatively high, ranging up to .452 for $n = 35$. This should not be surprising since the Hauser structure is fit relatively well by a low-rank common factor model. It is worth noting that the "best" adaptive methods, MR2 in particular, never yielded *GOF*s that even began to approach unity, the value that would be expected if the common factor model really was "completely satisfactory."

A final summary of all results is reported below, after the next study has been reported.

A Modest Simulation Study to Examine Outlier Robustness

This part of our study entailed similar methods to those used for the preceding simulations, except that the initial multivariate normal data matrix was deliberately contaminated by adding a small percentage of "outlier" values in order to compare the same methods with OLS in the presence of non-normal data. The particular methods for data contamination are briefly described below; the goal was to produce data that would simulate outlier systems of the kind that might commonly be encountered in work-a-day practice in validation studies. The particular reason for this extension to the preceding work is that some reviewers have questioned whether the advantages of WSR methods depend strongly on the use of multivariate normal sampling. The ensuing results speak to this question, if only in a limited way.

Following the construction of the initial simulation data matrix (X_{sim}), using multivariate normal sampling, a second compatible matrix was constructed, X_{cont}; this was done for each sample, for each value of n. Thus, each sample X_{cont} had n rows and p columns, but only k percent of those rows were non-null. The percentage of non-null rows was specified to be either two percent or five percent for each combination of population, common factor structure, and sample size. The non-null rows had exactly the same form as X_{sim} ; however it's columns were mutually independent. The matrix of these non-null rows was multiplied by the constant *three* so as to exaggerate the effect of these contaminates as outliers; this had the effect of producing outliers that were based on a simulation population a variance nine times larger than that of the original population.

Further, rows of the contamination matrix X_{cont} were bootstrapped so as to provide different numbers of contaminated rows in different samples, averaging k percent in number. This bootstrapped version of X_{cont} matrix was then added to the initial X_{sim} to produce the matrix X_{simcon}. All data matrices employed in this simulation had this form. This data generation procedure (i.e., with $k = 2\%$ or 5%) generally resulted in kurtosis statistics that were consistently larger than those of a normal distribution.

Because of the relatively more detailed results that this methodology involved, only two sample sizes were used here: $n = 60$ and $n = 120$. Three values were specified for *m:* $m = 2,3$ and 4. Again, one hundred bootstrap samples were used for each experimental combination in the simulation study. Cross validation errors are not provided in this phase of the simulation work.

Table 4 summarizes *MSE* and *PMSE* results for the Hauser data in a fashion that differs from the preceding Tables. In particular, to simplify how efficiencies of non-OLS methods compared with those for OLS, *efficiency ratios* were computed for each combination of method, sample size and value of m. Consequently, once the overall *MSE* and *PMSE* results are given (in the first row) for OLS, subsequent row numbers *exceeding* unity reflect an *MSE* or *PMSE* smaller, and efficiency that is greater (i.e. better), than that of OLS. Efficiencies less than unity reflect systematically worse performance, relative to OLS.

In nearly all cases, 116 of 120 (i.e., 60 for *MSE* & 60 for *PMSE*) results for the non-OLS regression methods showed greater efficiencies than OLS for both *MSE*s and *PMSE*s. In all instances

the non-OLS regression methods that used an *adaptive* weighting scheme yielded greater regression stability than least squares regression for both *MSE* and *PMSE* efficiencies. The only non-OLS method that did not consistently out-perform least squares regression was, the non-adaptive method Reduced Rank, which performed less well than least squares three times for the *PMSE* statistics at $m = 2$ and once at $m = 3$.

Examination of the *MSE*s and *PMSE*s for the specific method(s) that performed best (bold faced results in Table 4) shows that the GFI and MR2 methods were generally best: GFI performed best fourteen times, and MR2 performed best nine times. Interestingly, a pattern between GFI and MR2 appears to be delineated by the two efficiency statistics. With respect to the MSEs, the GFI method performed best with MR2 the next best eight times; the remaining four next best results are for Reduced Rank (including ties). In contrast, focusing on the *PMSEs* the MR2 performed best nine times of the possible 12, while GFI performed best three times, and Minimum Risk Ridge performed best once (including ties). *PMSE* results divided the next best performances among MR (six times), MR Ridge (three times), and GFI (one time).

The pattern of the GFI method performing best with respect to MSE criteria and the MR2 method performing best with *PMSE* criteria was also seen in the single overall best performance advantage over OLS for each statistic. Both instances occurred for $m = 2$, $k = 2\%$, and $n = 60$ with an *MSE* ratio of 2.51 for GFI and an *PMSE* ratio of 1.82 for MR2.

In general, the advantages of the adaptive WSR methods over OLS were greater for smaller sample sizes, and for the smaller level of contamination. But with *m* set at a reasonable value of two for the common factor model, the MR2 and GFI methods typically showed very substantial advantages over OLS, particularly with respect to the MSE criterion.

Conclusions

The viability of adaptive WSR methods for the analysis of data for a range of prediction problems, as developed by Pruzek and Lepak (1992), has been supported and extended by this study. Taken across all simulations, with two populations, varying sample sizes, and both normal and non-normal sampling processes, several results seem clear.

The MR2 method appeared to be the best method overall, even though MR and GFI showed the potential to improve on MR2 with respect to some criteria. Taken across both population systems, and all evaluative criteria, foregoing results suggest that for any given sample size – in the range of those examined – use of the best WSR method can yield substantial dividends. Specifically, the MR2 method for a given sample size *n* (and appropriate *m*) worked systematically at least as well as OLS for a doubled sample size, and sometimes as well as OLS for $4n$! Examination of Tables 2a and 2b, as well as Tables 3a and 3b, show consistently for most evaluative criteria that OLS methods were distinctly inferior to the best WSR counterparts.

The simulation study that examined effects of outliers further suggested that the special advantages of WSR methods are not confined to situations where multivariate normal sampling is used. The same WSR methods, especially MR2, showed themselves to be systematically better across both sample sizes and both levels of contamination than OLS methods. However, the WSR method closest to OLS, viz. MR, was not as effective as were MR2 and GFI in the presence of outliers. Also, for the larger *n*, the RR method did not perform well with respect to the *PMSE* criterion, especially at the higher level of contamination. These same findings were substantiated in the more comprehensive outlier simulation studies that were completed, but for reasons of space could not be included here.

The GFI method showed less robustness to variations in *m* than did MR2. Although methods such as the so-called Scree criterion seem often to work satisfactorily in selecting *m* for real data factor analyses, there is often some ambiguity associated with this choice. It is therefore pleasant to have seen that results from MR2 (especially) appear largely invariant to the choice of *m*, at least if one's choice for this scalar is in the "right ballpark."

That the better adaptive WSR methods often worked so much better than conventional OLS regression seems especially striking. Although some non-adaptive regression methods may have

certain features described below, adaptive WSR procedures are notable because they:

- make provision for incorporation of prior (possibly vague) structural models into analyses;
- weight the prior model adaptively, to the extent supported by observed data;
- can distinguish in principle between observed data and underlying latent traits;
- are invariant to shape preserving linear transformations of the data;
- are relatively robust to changes in the specification of *m* for the common factor model;
- can help enhance interpretability and generalizability of regression results even when the number of predictors approaches sample size;
- encourage the selection of predictors on the basis of theoretical considerations; and
- do not break down as the number of manifest variables is increased, provided that an appropriate method for common factor analysis is used.

In addition, it should be clear that adaptive WSR methods are both relatively easy to teach, and to program. Furthermore, these particular WSR methods are computationally highly efficient. They entail only about twice as much computation as a single OLS regression, and are far less computationally intensive than typical hierarchical or all-subsets methods.

By their nature, the adaptive WSR methods provide a coherent means of combining design considerations with analyses, in the context of what may be planned, sequential approaches to instrument validation. As many readers will know, results based on OLS methods tend to deteriorate rapidly as the number of correlated predictors grows, especially for anything but huge samples. However, as shown in preceding simulations, adaptive WSR methods may dramatically outperform OLS, the standard method for multiple regression, whether samples are multinormal, or contaminated. What could not be demonstrated in simulations, but what has become clear to the authors in many circumstances, is that the relative advantages of WSR methods over OLS methods can be far greater when there are many more variables than were included in the MTMM and Hauser populations.

That the new WSR methods can work so very much better than OLS methods may lead to different strategies for validation studies, compared to what has been most common in the past. In particular, small sample studies might be sequentially articulated to investigate particular construct domains, such that both structure and predictability of predictor-criterion systems become the focus of study. Based on newly available methods, one may expect to learn more, about more variables, than would have been possible had conventional validation strategies are followed.

Given sufficiently large sample sizes, results from adaptive WSR and OLS can be expected to converge (cf. Pruzek and Lepak, 1992). However, for smaller sample sizes such as those used in our simulations, OLS never performed as well as the best convex sum WSR alternative with respect to any of the evaluative criteria. Moreover, as the MTMM results suggest, even relatively conservative goodness of fit values can still bring about increased generalizability of regression results.

Particularly in cases where battery design has been such as to enhance the likelihood of a low-rank common factor model being supported by data, advantages deriving from application of WSR methods may be expected to be substantial.

Further research to develop effective structural models and to assess methods that accommodate multiple criterion variables, as well as multiple predictors, has been progressing well, increasing the promise of these methods. Extending WSR methods to structural models beyond those of factor analytic form is desirable but little work of this kind has been done.

Software Availability

Readers who wish to acquire (R) functions that can be used to perform WSR analyses are welcome to contact R. Pruzek at rpruzek@uamail.albany.edu .

References (selected, as I did not have time to enter all)

Laughlin, J.E 1986. Contrasting alternatives to least squares in regression using diagnostics for identifying influential data. *Multivariate Behavioral Research*, 21, 103-118.

Lawley, D.N. & Maxwell, A.E. 1973. Regression and factor analysis. *Biometrika*. 60, 331-338.

Pruzek, R.M. & Frederick, 1978. Weighting predictors in linear models: alternatives to least squares and limitations of equal weights. *Psych. Bull.* 85, 254-266.

Pruzek, R.M. & Lepak, 1992. Weighted structural regression: A broad class of adaptive methods for improving linear prediction. *Multivariate Behavioral Research,* 27, 95-129.

Smith, G. & Campbell, 1980. A critique of some ridge regression methods. *Jour. of the Amer. Statistical Association*. 75, 74-80.

Table 1a

Correlations, Squared Multiple Correlations and Eigenvalues for Multitrait-Multimethod Population System (Source: Campbell and Fiske, 1959)

CORRELATION MATRIX

1	1.00									
2	.27	1.00								
3	.62	.57	1.00							
4	.36	.47	.90	1.00						
5	.69	.32	.28	-.06	1.00					
6	.57	.11	.19	-.01	.53	1.00				
7	.28	.65	.42	.26	.37	.26	1.00			
8	.44	.25	.53	.45	.29	.31	.32	1.00		
9	.31	.20	.54	.52	.13	.11	.21	.47	1.00	
10	.15	.30	.12	.04	.34	.10	.12	.04	.06	1.00

SMCs

1	2	3	4	5	6	7	8	9	10
.77	.62	.94	.92	.75	.42	.52	.40	.37	.21

EIGENVALUES

1	2	3	4	5	6	7	8	9	10
4.067	1.772	1.247	.893	.662	.489	.445	.258	.134	.034

Table 1b

Correlations, Squared Multiple Correlations and Eigenvalues for Hauser Population System (Source: Hauser, 1963)

CORRELATION MATRIX

1	1.000										
2	.500	1.000									
3	.494	.318	1.000								
4	.389	.291	.523	1.000							
5	.244	.230	.212	.203	1.000						
6	.151	.149	.127	.116	.586	1.00					
7	.159	.141	.144	.146	.352	.439	1.00				
8	.299	.269	.290	.288	.369	.335	.424	1.00			
9	.278	.256	.284	.288	.318	.321	.327	.418	1.00		
10	.306	.269	.299	.304	.435	.469	.438	.542	.496	1.000	
11	.287	.246	.301	.288	.455	.466	.413	.507	.474	.766	1.000
12	.344	.292	.325	.319	.481	.545	.489	.491	.659	.659	.594 1.00

SMCs

1	2	3	4	5	6	7	8	9	10	11	12
.403	.287	.389	.329	.414	.480	.307	.396	.337	.672	.626	.565

EIGENVALUES

1	2	3	4	5	6	7	8	9	10	11	12
5.016	1.633	.864	.767	.683	.590	.530	.508	.431	.416	.342	.222

Table 2a
Summary Results for MTMM Data--Selected Criterion Variable

MSEs									
	n=35			*n=70*			*n=140*		
Method									
OLS	.421	.387	.426	.125	.142	.164	.064	.075	.079
RIDGE	.309	.309	.310	.184	.200	.170	.085	.104	.101
	m=1	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*
MR	.148	.163	.174	.075	.094	.086	.052	.056	.054
MR2	**.125**	.134	.141	**.066**	**.086**	**.071**	**.047**	**.049**	**.046**
GFI	.140	**.131**	**.134**	.115	.100	.071	.105	.091	.056
RR	.621	.244	.149	.590	.217	.095	.590	.203	.077
PMSEs									
OLS	.089	.091	.082	.034	.034	.036	.016	.017	.016
RIDGE	.072	.077	.069	.033	.035	.034	.016	.017	.016
	m=1	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*
MR	**.059**	.057	.051	**.027**	**.027**	.026	**.015**	**.014**	**.013**
MR2	.064	**.056**	**.049**	.030	.028	**.025**	.016	.014	.013
GFI	.085	.061	.050	.076	.045	.029	.073	.041	.020
RR	.459	.142	.067	.442	.113	.040	.440	.110	.028
CROSS-VALID									
OLS	.032	.031	.028	.025	.025	.024	.023	.022	.023
RIDGE	.030	.031	.026	.025	.025	.024	.023	.022	.023
	m=1	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*
MR	**.029**	**.029**	**.025**	**.025**	**.024**	**.023**	**.023**	**.022**	**.023**
MR2	.030	.029	.025	.026	.025	.023	.023	.022	.023
GFI	.032	.029	.025	.031	.026	.023	.027	.025	.024
RR	.070	.037	.026	.069	.033	.024	.061	.031	.024
GOF: (1-w)									
MR	.169	.284	.385	.100	.199	.291	.052	.114	.197
GFI	.432	.544	.657	.464	.630	.736	.479	.657	.780

Table 2b
Summary Results for MTMM Data--Overall System

*MSE*s *Method*	*n=35*			*n=70*			*n=140*		
OLS	.771	.770	.754	.293	.284	.317	.138	.154	.150
RIDGE	.294	.295	.304	.194	.190	.191	.104	.115	.110
	m=1	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*
MR	.273	.289	.279	.177	.166	.171	.101	.110	.100
MR2	**.231**	.230	.231	**.158**	**.147**	**.150**	**.092**	**.097**	**.092**
GFI	.238	**.222**	**.224**	.210	.184	.171	.191	.171	.154
RR	.543	.349	.278	.534	.335	.240	.590	.323	.222
*PMSE*s									
OLS	.148	.154	.153	.063	.061	.065	.028	.029	.029
RIDGE	.094	.099	.102	.051	.051	.053	.026	.026	.026
	m=1	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*
MR	.086	.088	.091	.047	.046	.047	.025	.025	.024
MR2	**.081**	**.082**	**.085**	**.045**	**.046**	**.046**	**.024**	**.024**	**.023**
GFI	.088	.084	.085	.068	.064	.054	.060	.055	.040
RR	.300	.152	.108	.285	.130	.075	.279	.119	.059
CROSS-VALID									
OLS	.052	.052	.054	.044	.043	.044	.042	.040	.042
RIDGE	.047	.047	.049	.043	.042	.043	.042	.040	.042
	m=1	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*
MR	.047	.046	.048	.043	.042	.043	.042	.040	.042
MR2	**.046**	**.046**	**.047**	**.043**	**.042**	**.042**	**.042**	**.040**	**.042**
GFI	.047	.046	.047	.045	.044	.043	.045	.043	.043
RR	.069	.053	.049	.067	.050	.046	.064	.050	.045

Table 3a
Summary Results for Hauser Data, For Selected Criterion Variable

	n=35			*n=70*			*n=140*		
Method									
OLS	.382	.376	.342	.155	.148	.156	.074	.064	.071
RIDGE	.144	.140	.141	.094	.090	.098	.060	.050	.055
	m=1	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*			
MR	.109	.101	.114	.067	.058	.068	.047	.034	.037
MR2	.076	.074	.089	.050	.047	.056	**.039**	**.030**	**.034**
GFI	**.066**	**.069**	.085	**.044**	**.045**	.052	.040	.040	.040
RR	.067	.069	**.082**	.059	.055	.057	.054	.051	.046
*PMSE*s									
OLS	.219	.209	.195	.084	.081	.086	.041	.036	.040
RIDGE	.099	.093	.094	.056	.055	.060	.034	.030	.033
	m=1	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*
MR	.076	.070	.078	.040	.038	.044	.027	.022	.023
MR2	.056	.054	.064	.031	.031	.038	.023	**.019**	**.021**
GFI	.049	.051	.062	**.025**	**.028**	**.035**	**.020**	.021	.022
RR	**.047**	**.049**	**.059**	.032	.032	.037	.027	.026	.024
CROSS VALID									
OLS	.061	.062	.058	.051	.048	.052	.046	.047	.045
RIDGE	.050	.051	.049	.048	.046	.049	.045	.046	.045
	m=1	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*
MR	.048	.049	.047	.047	.044	.047	.045	.046	.044
MR2	.046	.047	.046	**.046**	**.043**	**.047**	**.044**	**.045**	**.044**
GFI	**.045**	**.047**	**.045**	.046	.043	.046	.044	.046	.044
RR	.045	.047	.045	.047	.044	.046	.045	.046	.044
GOF: 1-w									
MR	.367	.452	.493	.291	.405	.459	.196	.324	.408
GFI	.621	.680	.713	.704	.782	.820	.743	.836	.887

Table 3b
Summary Results for Hauser Data--Overall System

*MSE*s									
	n=35			*n=70*			*n=140*		
Method									
OLS	.501	.526	.482	.212	.208	.198	.093	.087	.089
RIDGE	.189	.190	.188	.127	.124	.121	.073	.068	.070
	m=1	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*
MR	.152	.152	.170	.096	.085	.089	.059	.048	.049
MR2	.111	.115	.136	**.076**	**.071**	.075	**.051**	**.044**	**.044**
GFI	**.101**	**.108**	.130	.078	.073	**.072**	.073	.064	.052
RR	.140	.116	**.126**	.134	.098	.082	.130	.088	.060
*PMSE*s									
OLS	.280	.283	.266	.113	.110	.108	.050	.048	.048
RIDGE	.127	.126	.125	.076	.075	.074	.042	.040	.040
	m=1	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*
MR	.105	.105	.116	.061	.057	.059	.035	.031	.031
MR2	.086	.087	**.100**	**.053**	**.051**	**.053**	**.032**	**.030**	**.030**
GFI	**.083**	**.083**	**.098**	.063	.054	.053	.057	.042	.035
RR	.131	.091	.099	.117	.070	.059	.111	.056	.039
CROSS VALID									
OLS	.082	.082	.079	.065	.065	.064	.059	.058	.059
RIDGE	.068	.067	.066	.062	.062	.061	.058	.058	.058
	m=1	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*	*m=1*	*m=2*	*m=3*
MR	.066	.065	.065	.060	.060	.059	.058	.057	.057
MR2	.064	.063	.064	**.060**	**.059**	**.059**	**.058**	**.057**	**.057**
GFI	**.063**	**.063**	**.063**	.061	.059	.059	.060	.058	.058
RR	.068	.063	.063	.066	.061	.059	.065	.060	.058

Table 4

Estimated Relative Efficiencies for Mean Squared Errors and Predictive Mean Squared Errors in the presence of Contaminated Data: Hauser Problem

	Mean Squared Errors[a]				Pred. Mean Squared Errors[b]			
$m = 2$	$n = 60$		$n = 120$		$n = 60$		$n = 120$	
Method	k=2%	k=5%	k=2%	k=5%	k=2%	k=5%	k=2%	k=5%
OLS	.27	.44	.23	.29	.15	.23	.11	.14
Ridge	1.77	1.66	1.33	1.35	1.56	1.44	1.23	1.24
MR	2.07	1.65	1.51	1.38	1.72	1.40	1.33	1.24
MR2	2.46	1.93	1.75	1.60	**1.82**	**1.47**	**1.40**	**1.31**
GFI	**2.51**	**2.10**	**1.81**	**1.81**	1.63	1.39	1.08	1.08
RedR	1.78	1.76	1.35	1.40	1.02	.90	.70	.68
$m = 3$	$n = 60$		$n = 120$		$n = 60$		$n = 120$	
Method	k=2%	k=5%	k=2%	k=5%	k=2%	k=5%	k=2%	k=5%
OLS	.29	.51	.21	.33	.15	.26	.10	.16
Ridge	1.76	1.80	1.31	1.36	1.52	1.56	1.20	1.23
MR	2.00	1.68	1.49	1.39	1.64	1.47	1.29	1.24
MR2	2.29	1.98	1.67	1.59	**1.73**	**1.60**	**1.33**	**1.31**
GFI	**2.38**	**2.20**	**1.74**	**1.81**	1.67	**1.60**	1.21	1.22
RR	2.12	2.10	1.61	1.62	1.41	1.28	1.06	.95
$m = 4$	$n = 60$		$n = 120$		$n = 60$		$n = 120$	
Method	k=2%	k=5%	k=2%	k=5%	k=2%	k=5%	k=2%	k=5%
OLS	.35	.38	.20	.38	.18	.20	.10	.18
MR Ridge	1.75	1.78	1.35	1.36	1.51	**1.51**	1.24	1.21
MR	1.84	1.67	1.53	1.31	1.57	1.44	1.33	1.20
MR2	2.08	1.85	1.68	1.45	1.68	1.50	**1.38**	1.25
GFI	**2.21**	**1.93**	**1.74**	**1.61**	**1.71**	1.49	1.32	**1.26**
RR	2.17	1.85	1.67	1.60	1.63	1.36	1.23	1.18

[a] Table entries represent the average across 100 samples. For each value of m the first row is the actual OLS BMSE. Subsequent rows give ratios of overall OLS BMSE to the BMSE of the given method. Ratios greater than 1.00 indicate advantages of the alternative methods. The best value for each column is printed in bold type face.

[b] For each value of m the first row is the actual OLS Predicted Mean Squared Error (PMSE) followed by the PMSE ratios for the other methods. PMSE ratios are calculated in a manner identical to that of the BMSE ratios.