S-values: Conventional context-minimal measures of the sturdiness of regression coefficients

Edward E. Leamer[1]

UCLA

August 2014

Abstract

This paper proposes a context-minimal range of alternative regression models that is used to generate a range of alternative estimates. If the range of estimates for a regression coefficient excludes zero, the sign of the coefficient is judged to be sturdy. The proposed measure of sturdiness is the s-value, which is the average of the minimum and maximum of the estimates divided by their difference.

The proposed range of alternative models is built on a Bayesian foundation in which doubt about the relevance of each variable is captured by a prior distribution for the regression coefficients located at zero with a particular variance. If this prior variance were known or estimable, we would have an estimation problem not a model ambiguity problem. The choice of the prior variance is facilitated by transformation to standardized variables which makes the prior expected R^2 equal to the sum of the prior variances. Three different ranges of the prior expected R^2 are used to define three different intervals of prior covariance matrices which are used to produce three different sets of s-values.

The approach is illustrated with a reexamination of the regression equation of Sala-i-Martin, Doppelhofer and Miller (2004) which has 67 variables that are intended to explain the growth in real per capita incomes of eighty-seven countries from 1960 to 1996. In contrast with the conclusion of Sala-i-Martin, Doppelhofer and Miller, I do not find many of these coefficient estimates to be sturdy, meaning their signs are ambiguous. Thus I conclude that context-minimal inference in this setting is not productive, and if something of value can come from these data it depends on context-dependent information that establishes a compelling preference for some variables or some combinations. In other words, it takes insight and wisdom, not just data and math and computer power.

[1] Prepared with the research assistance of Shekhar Mittal. Comments from my colleagues at UCLA including Sebastian Edwards and Romain Wacziarg are gratefully acknowledged. Special thanks to Paola Guiliano and Christian Dippel for suggestions that I have incorporated here. The discussion of g-priors included in this version is a response to a comment from Gary Chamberlain and to references brought to my attention by Mark Steel. I have also benefited much from discussions with Carlos Cinelli.

1 Introduction

Inferences from limited data sets are afflicted by both sampling uncertainty and model ambiguity. T-statistics and p-values are ubiquitous context-free conventional measures of sampling uncertainty of estimated coefficients in linear regression. These measures are designed to reveal if the signs of the estimated coefficients are statistically reliable. When drawing inferences from nonexperimental data it is usually acknowledged that these measures depend on untested assumptions concerning which variables should be included in the model and which should be excluded. This kind of model ambiguity is normally handled in an ad hoc manner by reporting the results from a variety of alternative models.[2]

I propose here context-minimal conventional measures of the extent to which the signs of regression coefficients are free of model ambiguity concerns. This requires a context-minimal convention regarding the range of allowable alternative models among which there is indifference, not meaning equal prior probabilities but instead hard-to-assess prior probabilities. The key innovation comes from the fact that an R^2 is a measure of the size of the coefficients and the prior expected R^2 is a function of the prior covariance matrix – larger prior expected R^2 corresponding with greater prior covariance matrix and more freedom for the coefficients to wander from zero. The minimal contextual input is taken to be an interval for the expected prior R^2, which corresponds with a set of prior covariance matrices, which in turn corresponds with sets of posterior means and posterior variances, from which extreme values can be selected.

The proposal here is to standardize the variables to have unit variance, to adopt a Bayesian approach that shrinks the coefficients to zero (prior mean equal to zero), and to allow an interval of prior covariance matrices with upper and lower bounds that depend on conventional context-free upper and lower values for the expected R^2 of the equation. The logic for this convention is explained herein.

If model ambiguity is extreme and the prior covariance matrix is free to vary over all positive definite matrices, then almost no coefficient has a sturdy sign, since there are prior covariance matrices that imply either positive or negative signs for all but one special linear combination of regression coefficients. It is surely sensible to have a lower bound on the prior covariance matrix, and thus to exclude dogmatic priors with zero variances since these allow the imposition of hard linear constraints on the coefficients which are completely impervious to the data evidence.

To let the data speak freely as the sample size grows, we need a lower bound for the prior covariance matrix, but we do not need an upper bound. On the contrary, an upper bound rules out the unconstrained ordinary least-squares estimate and thus prevents the data from speaking freely except asymptotically when the message is so loud it cannot be ignored. I had

[2] I have written extensively about this, e.g. Leamer(1978), as have others.

thus thought that an upper bound would not also be essential, but I changed my mind when the intervals of possible estimates became uselessly wide as the number of variables increased. An upper bound for the prior variance is thus also essential since it limits the reduction in sturdiness of the coefficients that comes from adding variables to the equation.

Both upper and lower bounds that are proposed here have the prior variance inversely proportional to the number of variables in the model, in other words greater doubt about the statistical importance of individual variables as the number of variables increases. This makes it less likely in a large model that any particular variable has a large beta-coefficient, while keeping constant across models of different size the total ignorance measured by the sum of the prior variances. [3]

An essential step toward defining context-minimal upper and lower bounds for the prior covariance matrix is the transformation to standardized variables with unit variance, thus estimating “beta-values” as the regression coefficients. With this parameterization and with all other information about the setting hidden from the analyst, the variables are exchangeable, and the only possible choice for a conventional prior distribution is i.i.d., since there is no information that would allow anything else.[4] Still, it needs to be emphasized that the proposed measures of model ambiguity are invariant to changes in scales of the variables (via the standardization), but not invariant to rotations, meaning if you use x_1 and x_2, while I use x_1-x_2, and x_1+x_2, we will get different answers regarding the sturdiness of the effect of x_1 holding fixed x_2. In other words, the use of an identity prior covariance matrix is context-dependent.[5] However, the proposed s-values use covariance matrices proportional to the identity matrix only

---

[3] Incidentally, in the model-selection framework discussed in a companion paper, Leamer(2014), it is the product of the variances (or the sum of the logarithms of the variances) that must be held constant in order for the larger models to compete on a level playing field with the smaller models.

[4] I understand that exchangeable does not rule out the equicorrelated case, but if the common correlation is treated as a known number, as done here, and nothing is known about the context, the only option is zero correlation.

[5] This brings to mind Hoerl and Kennard’s(1970) “ridge regression” which uses exactly the same formula for estimating the regression coefficients as a Bayesian analysis with a spherical prior, namely adding the same value to each and every diagonal element of the sample moment matrix of the explanatory variables. Hoerl and Kennard(1970) demonstrate that there exists a scalar to be added to all these diagonals which lowers the mean-squared-error of the estimator, reducing the variance by more than it increases the bias. Unfortunately, that secret scalar depends on the unknown size of the coefficient vector and this is not a way to do better than ordinary least squares regardless of the regression coefficients.

to establish upper and lower bounds for the prior covariance matrix, which relieves but does not eliminate the coordinate-dependence of the approach. Thus the adjective: context-minimal.[6] T

With the prior covariance matrix equal to ν^2times the identity matrix, it is shown below that ν^2 is equal to the prior expected R^2 divided k, the number of coefficients, $\nu^2 = E(R^2)/k$. For defining a range of alternative prior distributions all that is needed is a conventional range for the prior expected R^2. Three ranges for the prior expected R^2 are proposed. The broadest context-free range allows the expected R^2 to be any value between 0.1 and 1.0. Two other ranges are formed by splitting this interval in two, from 0.1 to 0.5, and from 0.5 to 1.0. The context would have to be referenced implicitly or explicitly to choose one of these narrow ranges of models. The context question is: Are you studying a setting in which the R^2 is likely to be in the higher range?

Although it is of interest to map out the estimates as the scalar ν^2 is varied, something that Hoerl and Kennard (1970) called the "ridge trace," this produces a one-dimensional set of estimates that is context-dependent, namely a context in which i.i.d. actually applies. It is proposed here that the assumption of an identity prior covariance matrix be used only to assist in selecting upper and lower bounds for prior variances of linear combinations of parameters. Then the context-limited set of allowable estimates is based on any prior covariance matrix that produces prior variances of linear combinations of coefficients between these upper and lower bounds. In mathematical terms we use a positive-semi-definite ordering with the prior covariance matrix bounded from above or below. Prior covariance matrices compatible with these two inequalities will be close to proportional to identity matrices at the extremes but between the extremes can have substantial non-zero covariances, and less reliance on the coordinate system.

Corresponding to any interval of prior covariance matrices there is a maximum and minimum estimate for each coefficient estimate. These bounds are recorded in a sturdiness indicator which I will call the s-value to accompany the t-value and the p-value. The s-value is the average of the minimum and maximum estimates divided by half the difference between them. The s-value summarizes the robustness interval of estimates in exactly the same way a t-value summarizes the one-standard error confidence interval – an s-value or a t-value greater than one means the corresponding interval is bounded away from zero.

The lower bound for the prior covariance matrix assures that, as the sample size increases and the data become more informative, the range of allowable estimates will diminish and will collapse on the unconstrained estimate. Thus both sampling uncertainty and model ambiguity are reduced as the sample size increases. That raises the question: Are sampling uncertainty and model ambiguity essentially-the-same ways of looking at the same problem – not enough

[6] If you prefer, you may call it instead: context-forgotten, meaning that temporarily even the names of the variables are forgotten.

data? More specifically, are the s-values some simple function of the t-values? An affirmative answer to this question is suggested by the result in Leamer(1975) that the omission of a variable can change the sign of another coefficient only if the t-value on the retained variable is less than the t-value of the omitted variable. Beyond just the suggestion of that result, if the interval of prior covariance matrices is expanded to include any positive definite matrix, the s-values are proportional to the t-values, and the difference between sampling uncertainty and model ambiguity is only what the t-value is compared with, respectively, the number 2.0 or the chi-square for testing the joint significance of call coefficients. (explained below) In addition, if the upper and lower covariance bounds are proportional not to the identity matrix but proportional to the sample covariance matrix, then there is a one-to-one exact correspondence between s-values and t-values, also explained below.

Though the actual s-value of a coefficient depends on its t-value and also on the correlation with other coefficients, for the data set studied here there is a substantial correlation between t-values and s-values, which in a sense is a comforting result, suggesting that we haven't gone too far astray by focusing on t-values.

Section 2 illustrates the main results with a graph of estimates for a two-variable regression model. Section 3 lays out the notation of regression used here. Section 4 surveys the limited literature on sensitivity analysis for linear regression. Section 4 explains the way that standardized coefficients can be used to produce conventional prior distributions with upper and lower prior covariance matrices. Two treatments are suggested: one in which all variables are treated the same and a second in which the energy of the data is concentrated on a set of "favorite" variables. Section 7 reviews the theorem in Leamer(1978) which is the basis for the ambiguity measures proposed here. This result describes the set of Bayes estimates of regression coefficients when the prior covariance matrix is bounded from above and below. A special case from Chamberlain and Leamer (1976) describes the set of Bayes estimates corresponding to the set of all prior covariance matrices. Other relevant sensitivity results are reported in Leamer, Edward E. (1975) Chamberlain and Leamer(1976), Leamer and Chamberlain(1976), Leamer, Edward E. (1978) and Leamer, Edward E. (1982).

I offer here some new sensitivity results for linear regression that extend these results. Worthy of special note are results that highlight the critical role of the simple correlations in determining the range of allowable estimates. I show below that, for the intervals of prior covariance matrices considered here, when the data are weak relative to the prior, the estimates will conform in sign with the simple correlations. It is this result that has led to my suggestion to include the simple one-variable-at-a-time regressions alongside the multivariate regression. These one-at-a-time regressions are a feature of the data, while the "partial" regression coefficients are cooked up by the analyst when he or she selects the control variables. This cooking needs special scrutiny when a simple correlation and a partial correlation are opposite in sign, in which case we need some way of deciding which is the better estimate of the sign. I argue that there is no preference for either sign when the ambiguity in the prior is great relative to the strength of the data information, something which is captured precisely with the proposed s-values. Section 7.1 is an example based on data of the Sala-i-Martin et.al.(2004) study of determinants of the growth rate of per capita real incomes from 1960 to 1996 in a sample of 87 countries. For dispositive purposes, I first report s-values for a regression with only 14 "favorite" explanatory variables selected "wisely" from the full set of 67 used by Sala-i-Martin et.al.(2004). Next are s-values in a regression with the full set of 67 variables, and last a proposed conventional concentration of the data information on the 14 favorites among the 67 variables. I conclude that without some favoritism, there is little to be learned from these data. This conclusion contrasts with the much more optimistic conclusion of Sala-i-Martin et.al.(2004) which has a parallel study of these data also using a Bayesian approach, though a different one. Both methods are built on doubt about the importance of the explanatory variables and both assume ambiguity in defining that doubt. Some brief comments on the similarities and differences are provided in Section 4.2 and in the conclusions, but a full discussion is relegated to a companion paper, Leamer(2014), titled "S-values and All-Subsets Regressions."

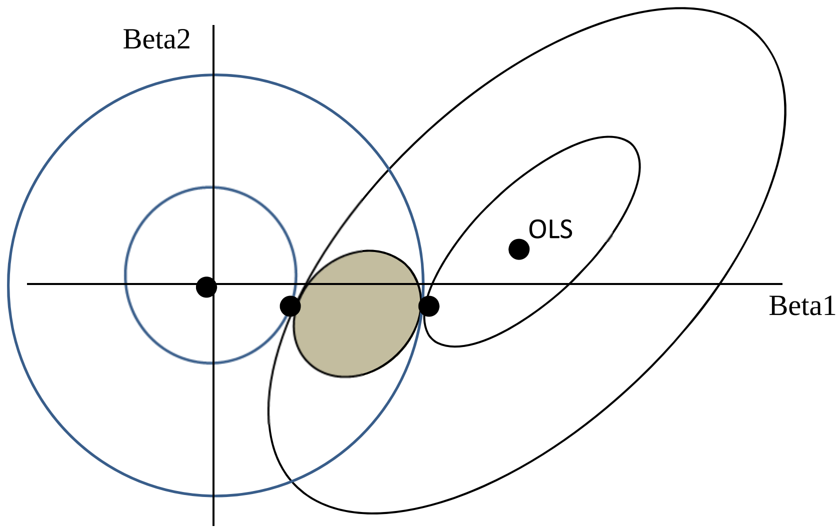
Finally, in Section 9, I offer concluding comments and some final promotional words in support of the convention proposed here.

2 Visual Illustration of the Proposed Reporting Style

Jumping ahead to the finish line, Figure 1 illustrates the features that form the proposed reporting style. This figure has the OLS estimate of the two parameters in the positive quadrant, surrounded by two concentric likelihood ellipses (or confidence sets). Two circles centered at the origin represent the prior opinion that these beta coefficients are probably small. The optimal compromises between the data and the prior are points of tangency between a likelihood ellipse and a prior circle. Two such points are illustrated, the one closer to the OLS estimate corresponding to the "strongest" sensible prior opinion (minimum expected R^2) and the other one corresponding to the "weakest" sensible prior opinion (maximum expected R^2).

Trace out with your mind's eye the locus of tangencies, and you will notice that it begins in the positive quadrant at the OLS point, but then experiences a sign change for Beta2. That suggests the sign of Beta2 is fragile. However, the estimates for the weakest and strongest prior both have Beta2 negative, so perhaps the sign of Beta2 is sturdy. Finally, the shaded ellipse is the set of estimates associated with the set of priors between the weakest and the strongest cases, meaning that the prior variances of any and all linear combinations are bounded from above and below by the prior variances for these linear combinations implied by the weakest and strongest priors. This set of prior opinions allows estimates off the locus of tangencies between the circles and the likelihood ellipses because noncircular priors are allowed. The shaded region is the full set of these estimates. This shaded "model-ambiguity" region includes both positive and negative estimates of Beta2, and we would thus conclude that the sign of Beta2 is fragile while the sign of Beta1 is sturdy.

Figure 1 OLS and Two Estimates Closer to Zero: Spherical Prior



The proposed reporting style describes the sampling uncertainty in terms of the OLS estimate, and also several alternatives on the locus of tangencies between the likelihood ellipsoids and the prior spheres. The model ambiguity is conveyed by the extreme estimates taken from the shaded model-ambiguity region.

The critical assumptions are (1) sphericity of the reference prior and (2) sensible choices of the two extreme priors. Sphericity seems a "natural" consequence ignorance about the setting combined with transformation to beta-coefficients – without knowing more about the setting it is impossible to select non-zero prior covariances or non-identical prior variances.[7] To find the extremes, I will show below that it is enough to have extreme values of the expected R^2. Specifically, I propose reporting s-values corresponding to three different intervals of the prior expected R^2: from 0.1 to 1.0, and the non-overlapping subintervals, 0.1 to 0.5, and 0.5 to 1.0.

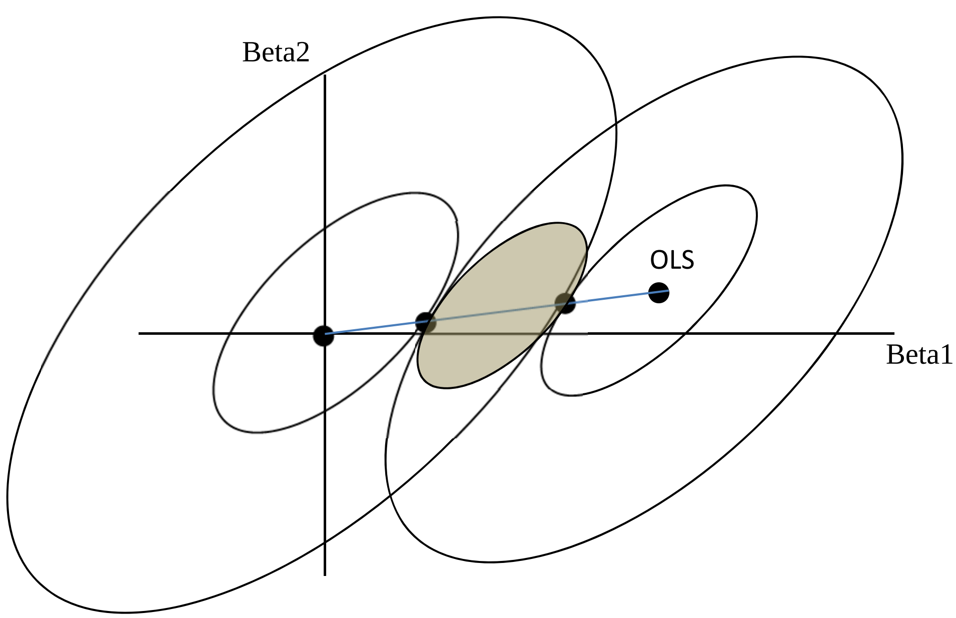
This proposal is not genuinely context-free since the shaded model-ambiguity region in Figure 1 is not invariant to all linear transformations of the data. Thus if you use β_1 and β_2 as your parameters, while I use $\beta_1+\beta_2$ and β_1-β_2 , we are going to draw different conclusions regarding the sturdiness of the estimates of β_1 and β_2. To get invariance, we need the upper and lower prior precision matrices proportional not to the identity matrix but to the sample precision matrix, **X'X**, per the g-prior proposed by Zellner(1986). This case is illustrated in Figure 2. Here

---

[7] Warning: this statement presumes normality. There are countless other distributions that could describe the prior state of mind and circles are not the only way to express the idea that the coefficients are probably small. For example, one might suppose that the product of the absolute beta values is small and instead of circles we would have hyperbolas, and thus shrink toward the sides of the quadrants.

the prior ellipses are not spherical but have the same shape as the likelihood ellipses around the OLS point. As a consequence the locus of tangencies that represent efficient compromises between the data and the prior lie on a straight line connected OLS with the origin. This means that there cannot be any sign changes in any of the estimated coefficients – just multiply the OLS vector by a suitably selected positive scalar less than one. In addition, the shaded sturdiness ellipse for measuring model ambiguity has exactly the same shape as a likelihood ellipse used for measuring statistical uncertainty. For these two reasons, as further explained below, s-values and t-values are proportional to each other. To determine if sampling uncertainty is small compare a t-value with a suitably selected number; to determine if model ambiguity is small, compare a t-value with a different number. Aside from this (disappointing) fact that these g-priors make statistical uncertainty and model ambiguity numerically the same, the use of g-priors requires some very strange behavior by the analyst and her audience, you are all forced to wait until you see the **X** matrix before you can express yourself regarding the sense in which the regression coefficient vector is close to the origin. Caveat: the same critique applies to my use of standardized variables, but, as argued already, in a nonexperimental setting, exchangeable beta-coefficients seems more natural than exchangeable coefficients or exchangeable t-values.

Figure 2 OLS and Two Estimates Closer to Zero: g-prior



3 Setting

This section describes the linear regression setting in which the data are analyzed: the sampling process and the prior state of mind, both assumed to be distributed normally.

3.1 Sampling Process

The data are assumed to consist of an ($n\times 1$) "dependent" variable vector **y** and an ($n\times k$) "explanatory" variable matrix **X.** It is assumed that the vector **y** conditional on **X** is normally distributed with mean $\mathbf{X}\beta$ and covariance matrix $\sigma^2\mathbf{I}$, were β is a ($k\times 1$) vector of unknown "regression coefficients" that link **y** with **X**, and where σ^2 is a scalar equal to the variance of each element of **y** given $\mathbf{X}\beta$. The ordinary least squares estimate of β is then

$\mathbf{b} = (\mathbf{X'X})^{-1}\,\mathbf{X'y} = \mathbf{N}^{-1}\,\mathbf{r}$, where

$\mathbf{N} = \mathbf{X'X}$, and

$\mathbf{r} = \mathbf{X'y}$

The corresponding sampling variance and precision matrices are

Variance(**b**) = $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} = \boldsymbol{H}^{-1}$

Precision(**b**) = $\mathbf{H} = (\mathbf{X'X})/\,\sigma^2$

3.2 Prior State of Mind and Approximate Posterior Mean

Prior to the observation of **y** and **X** it is assumed that the analyst and possibly the analyst's audience have the opinion that the vector of coefficients β is probably close to zero. This vague statement allows for doubtful variables that probably have small coefficients and also allows for similar variables with coefficients that are probably about the same.

This state of mind is approximated with a normal prior distribution for β with mean vector **0** and variance matrix ***V***. Then the posterior mean is a matrix weighted average of the OLS estimate ***b*** and prior mean ***0***, with weights proportional to the sample and prior precision matrices: $\widehat{\boldsymbol{\beta}}(\boldsymbol{V}) = \left(\boldsymbol{H} + \boldsymbol{V}^{-1}\right)^{-1}\boldsymbol{Hb}.$ This approximation of the prior state of mind is mostly for mathematical convenience, and it's a huge and incredulous step from the vague statement that the coefficients are probably small to the precise statement that the coefficients are drawn from a normal distribution with known variance matrix. Because of that, we need to know how sensitive the posterior mean is to changes in the prior variance matrix. That is the way we will be expressing model ambiguity problems.

The prior opinions, if not dogmatic, are irrelevant when the sample is sufficiently informative, but in most samples studied by economists the prior state of mind matters and can affect the inferences drawn. The two related questions addressed in this paper are thus: (1) *Is the data* *evidence strong enough that we can ignore the prior state of mind*? (2) *If the answer to (1) is negative, does ambiguity in the prior state of mind cause ambiguity in the signs of the regression coefficients?*

For studying nonexperimental data, specification ambiguity is especially important because the absence of experimental controls makes the list of potential regressors unlimited, and this makes the signs of the coefficients sensitive to the choice of explanatory variables. Consequently, it isn't easy to tease persuasive inferences from nonexperimental data. It is going to take some wisdom and some judgment and possibly some collective delusion to limit the list of explanatory variables in a way that allows useful and credible inferences.

3.3 Treatment of the residual variance

In what follows, the value of the sample precision ***H*** is taken as given and compared with hypothetical prior precision matrices, but in fact ***H*** depends on the uncertain residual variance σ^2 which has to be estimated. While the best estimate of σ^2 surely depends on the prior information, the most interesting and easily computable measures of specification ambiguity discussed below require a value of ***H*** that is fixed independent of the prior. Accordingly, I will use a traditional sample of estimate of σ^2 to determine the scale of the precision matrix ***H***, and I ignore any impact of the prior information about the coefficients β on the best estimate of σ^2.

4 Treatments of Estimation Uncertainty and Model Ambiguity

This section describes traditional and some nontraditional ways to uncover and to communicate estimation uncertainty and model ambiguity. These approaches compete with the proposals offered in this paper.

4.1 Tables of Alternative Constrained Regressions

It is common practice when estimating a regression to try different specifications with different subsets of variables or with other linear constraints imposed on the coefficients. This kind of specification search usually combines in an unclear way both estimation uncertainty and specification ambiguity.

If the model selection algorithm is predetermined, (e.g. omitting variables with t-values less than 2), then this procedure implicitly ignores specification ambiguity completely and is only a solution to the same estimation problem that gives rise to the unconstrained unbiased estimator. The design of an estimation algorithm with sample-dependent restrictions on the coefficients requires a study of the properties of this implicit and complex mapping of data sets into estimates, for example, studying the mean squared error of these biased estimators.

Usually, however, specification ambiguity is acknowledged by the way the results are reported – not one estimate, but a table of alternatives with different subsets of the restrictions imposed. If it is taken as given that each of the reported models is an equally reliable summary of the statistical uncertainty regardless of its statistical fit, then this table of alternative estimates is exclusively helping to define the amount of specification ambiguity. But if the evidence is said by the analyst to favor one model over another, because of a better fit or more plausible coefficients, then the table is serving partly to reveal the underlying details of a complex estimator that mixes the results from many models. The standard-operating-procedure of professional economists is presentation of a table of results interpreted informally in a way that mixes estimation uncertainty with specification ambiguity. One goal of this paper is to design a style of reporting that clearly distinguishes these two, and clearly reveals the prior state of mind that is the foundation for whatever are the reported inferences.

4.2 Averaging of Constrained Regressions

The coordinate system in which the omission of variables is appropriate is one in which the parameters are independent in the prior distribution, meaning "you" begin the study of the data with a state of mind such that if "you" learn something about β_1 alone that tells "you" nothing about β_2. The connection between the prior and the coordinate system comes from the result in Leamer and Chamberlain (1976) that when the prior covariance matrix $\boldsymbol{V}$ is diagonal, the Bayes estimator $\widehat{\boldsymbol{\beta}}(\boldsymbol{V}) = \left(\boldsymbol{H} + \boldsymbol{V}^{-1}\right)^{-1}\boldsymbol{Hb}$ can be written as a weighted average of the 2^k regressions formed by combinations of included variables selected from the full set of k explanatory variables.

If the diagonal elements of $\boldsymbol{V}$ are fixed, the weights on the 2^k regressions do not depend on how well a model that includes a subset of variables fits, which is a feature of a study with a normal prior distribution and a normal sampling distribution. But if the prior distribution has a nonnormal shape with a stronger central tendency and flatter tails, something that captures the idea that the restriction is probably true, but if it isn't then not much is really known, then the diagonal elements of the prior precision matrix $\boldsymbol{V}$ should be thought to be data-dependent.

This is an elliptical way of introducing Bayesian Model Averaging (BMA) used by Raftery(1995) and by Fernández, Ley and Steel(2001), and the same but misnamed BACE estimator used by Sala-i-Martin, Doppelhofer and Miller(2004). BACE stands for Bayesian Averaging of Classical Estimators, but in fact this estimator is an approximate Bayesian posterior mean for a special kind of prior distribution, which has an atom of mass (positive probability) assigned to each restriction but is otherwise diffuse (uninformative).

There is a technical problem with this kind a prior distribution that combines the dogmatic with the uninformed, since it wants to strongly favor the model with the fewest number of coefficients.[8] This technical problem is described in Leamer(1978) and repeated in Sala-i-Martin et.al(2004), but the solution to this technical problem chosen by Sala-i-Martin et. al.(2004) is

[8] The problem is that the diffuse prior puts "almost all" the weight at plus or minus infinity where the likelihood value is zero, and as the dimension of the model increases in a certain sense more and more of the weight is placed where the likelihood is zero, and thus the weighted likelihoods for larger models are infinitely smaller than the weighted likelihoods for small models.

eminently sensible, in other words it conforms with Leamer(1978) who provides warning labels and with Schwartz(1978) who does not provide warning labels.

In any case, BACE is solving an estimation problem in a more thoughtful way than stepwise regression, but it's not about model ambiguity. However, Sala-i-Martin, Doppelhofer and Miller(2004) do have a treatment of model ambiguity (a sensitivity analysis) carried out with respect the expected number of included variables, and Ley and Steel (2009) explore both estimation and sensitivity analysis with respect to the prior covariance multiplier "g".

4.3 Posterior Bounds Given A Range of Prior Distributions

The traditional sensitivity analysis involves a study of the way estimates change when some variables are omitted or some linear constraints are imposed. These estimates presume an unlikely combination of knowledge and ignorance – complete confidence that some constraints apply but total ignorance about the validity of other constraints. An appealing feature of a Bayesian approach is that it allows the intermediate imposition of "soft" constraints, some harder than others. The (Bayesian) estimates considered here are matrix weighted averages of the OLS vector **b** and the vector **0**:

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{V}) = \left(\boldsymbol{H} + \boldsymbol{V}^{-1}\right)^{-1}\boldsymbol{H}\boldsymbol{b} \qquad (1)$$

where **V** is the prior variance matrix. This is the Bayes estimator with a normal prior for the parameter vector **β** with mean **0** and precision matrix **V.** What lies behind this assumption is the same thing that lies behind the usual procedure of dropping variables with small t-values, namely, we begin with doubt about the importance of each variable and expect the data to overcome that doubt. That doubt is expressed precisely in Bayesian terms by the assumption that $\boldsymbol{\beta}'\boldsymbol{V}^{-1}\boldsymbol{\beta}$ is probably small. If the prior variance matrix is large enough, then the constraints are essentially ignored, and if the prior vaiance matrix is small enough, the constraints are imposed with little regard for the data evidence.

The leap from the feeling that β is probably small to the technical Bayesian assumption that β is drawn from a normal distribution with mean vector **0** and known covariance matrix **V** is preposterously large, so large that no one dares to make the jump except in the dreams of theorists. But that leap seems more comfortable if instead of a known value of the prior covariance matrix **V** we allow for a range of alternatives.

The title question "What are the signs of the coefficients?" is here translated into the specific mathematical question: "What alternative sign patterns emerge for the vector of estimates as the prior precision matrix **V** is varied?" This question emerges in a setting in which prior opinions have two special features: (1) There is wide agreement that the psychological starting point for the data analysis is one of doubt about the signs and sizes of the coefficients **β**.[9] (2)

[9] If there is wide agreement on some location for the prior other than zero, then by a change of parameters, the problem is mathematically identical to an omitted variables problem.

There is wide disagreement/confusion over how exactly to describe the nature of the doubt other than to say that, absent the data, the best estimate of $\boldsymbol{\beta}$ is the zero vector.[10]

5 Seeking context-minimal measures of model ambiguity

The central message of Bayesian statistics is also its greatest shortcoming: the context matters. To do a Bayesian analysis of a data set, an analyst is expected to think long and hard about what are the probable values of the parameters in the context under study, and then find just the right words to convince her clients and readers about the wisdom of her choice of informative prior distribution. The large literature on diffuse priors is an attempt to automate around these difficult tasks by recommending a conventional prior that can be used in a wide set of applications in which the prior information is weak compared with the data information. That allows automated conventional Bayesian measures of sampling uncertainty which typically are the same as sampling theory measures. Parenthetically, economists who routinely study what happens to estimates when variables are omitted from regressions are implicitly rejecting this advice, and resorting to ad hoc specification searches to input the contextual information.

While a diffuse prior distribution can be used under some circumstances to measure the amount of sampling uncertainty, a model with a diffuse prior cannot be a starting point for measuring model ambiguity since there is no (hyper) parameter that can be perturbed to perform the sensitivity analysis. The goal of this paper is find a measure of model ambiguity which has an ideally chosen minimal context dependence, enough so that measures of model ambiguity are useful, but not so much that the required inputs place an unbearably heavy credibility burden on the analyst and her readers.

Essentially what a prior distribution does is to establish a probable domain for the parameters. Instead of making a direct choice of probable domain, I allow a range of expected prior R^2 values to determine a range of prior distributions that are located at zero and vary in their degree of concentration, more concentrated for small expected R^2 and less concentrated (i.e. probably larger coefficients) for large expected R^2.

An analyst who can comfortably select a prior expected $R^{2.}$ has limited the set of potential prior covariance matrices, perhaps enough that further refinement of the prior wouldn't matter very much. The R^2 with standardized variables is equal to $\boldsymbol{\beta}'\boldsymbol{\Sigma}_{xx}\boldsymbol{\beta}$ where $\boldsymbol{\Sigma}_{xx}$ is the covariance matrix of the explanatory variables. This scalar has prior expected value equal to $trace(\boldsymbol{\Sigma}_{xx}\boldsymbol{V})$ where $\boldsymbol{V}$ is the prior covariance matrix. The set of prior covariance matrices compatible with a given prior expected R^2 is

[10] The proposed reporting separates sampling uncertainty and model ambiguity but a question that combines the two is: "How sure are you that the sign is positive?" That question could be translated into: "What range of t-values for this coefficient can be found from these data?"

$$\Omega_1 = \{\boldsymbol{V} | trace(\boldsymbol{\Sigma_{xx}V}) = E(R^2), 0 \leq \boldsymbol{V}\}$$

If an analyst could convince herself and her audience that some particular value of the prior expected R^2 is the unique value that applies, then a sensitivity analysis could perturb the prior covariance matrix only within the range Ω_1. This state of public opinion is surely rare and the sensitivity analysis would be more inclusive if it encompassed variability in the prior expected R^2. This implies a set of prior covariance matrices that depends *only* on the interval of prior expected R^2's:

$$\Omega_2 = \left\{\boldsymbol{V} | Min\big(E(R^2)\big) \leq trace(\boldsymbol{\Sigma_{xx}V}) \leq Max\big(E(R^2)\big), 0 \leq \boldsymbol{V}\right\}$$

Carrying out the sensitivity analysis with regard to this set of prior covariance matrices has appeal but is analytically difficult because of the complexity of the set of prior covariance matrices compatible with any given prior expected R^2. To make progress, I thus restrict the set of prior covariance matrices to lie within an interval $v_L^2\boldsymbol{\Phi} \leq \boldsymbol{V} \leq v_U^2\boldsymbol{\Phi}$ where the inequality $v_L^2\boldsymbol{\Phi} \leq \boldsymbol{V}$ means that $\boldsymbol{V} - v_L^2\boldsymbol{\Phi}$ is positive semi-definite, where $\boldsymbol{\Phi}$ is a selected matrix to be discussed next, and where the upper and lower scalars, v_L^2 and v_U^2, are selected to assure that that the prior expected R^2 lies between lower and upper value, $Min(E(R^2)) \leq v^2 trace(\boldsymbol{\Sigma_{xx}\Phi}) \leq Max(E(R^2))$.

For reasons already discussed, my preferred choice of $\boldsymbol{\Phi}$ is the identity matrix. This comes from the exchangeability of the beta-coefficients, absent additional information. Then the set of prior covariance matrices is

$$\Omega_3 = \{\boldsymbol{V} | Min(v^2)\boldsymbol{I} \leq \boldsymbol{V} \leq Max(v^2)\boldsymbol{I}, Min(E(R^2)) \leq v^2 trace(\boldsymbol{\Sigma_{xx}I}) \leq Max(E(R^2))\}$$

This is best thought as the first step in inputting context into the data analysis. If the model ambiguity is small when a minimal of contextual information is used, there is no need to be more careful, but if the model ambiguity is great, more work is needed to choose the contextual inputs.

The set of estimates of the regression coefficients implied by the set Ω_2 is invariant to linear transformations, while the set of estimates implied by Ω_3 is not. In other words, we have allowed context to matter beyond the prior expected R^2 – when the prior is extreme, the covariance is proportional to the identity matrix. This is a setting in which the coordinate system matters. Expressed simply, you have to think omitting variables makes sense, and imposing other constraints like equality of coefficients would require more thought and introspection.

A mathematically convenient way to obtain invariant results is to have the upper and lower bounds be proportional to the sample covariance matrix, $\boldsymbol{\Phi} = \boldsymbol{H^{-1}}$, the case illustrated in Figure 2. It will be shown below that this assumption yields measures of model ambiguity that are proportional to the usual measures of statistical uncertainty.

6 Conventional Upper and Lower Prior Variance for Beta coefficients

A conventional prior needs to compare the effects of different explanatory variables in a way that does not depend on arbitrary units of measurement. The usual t-statistics have this invariance property but these t-values depend in very complex ways on the matrix of covariances of the explanatory variables, which makes it unlikely that analysts could confidently claim prior knowledge of t-values. In settings in which the variables are strictly positive, a logarithmic model could be used and would have parameters that are unit-free.

When a log-linear model is not possible and even for a log-linear model, we can find conventional priors for "beta" coefficients, based on data normalized to have sample means equal to zero and sample variances all equal to one. The "beta" coefficients with these normalized data do not depend on the units of measurement of either the dependent variable or the explanatory variables. These coefficients measure the number of standard deviations of the dependent variable that is induced by a one-standard deviation change in an explanatory variable.[11]

Here is how to find a conventional prior when variables are normalized to have unit variance. A regression equation with k explanatory variables can be written as $y_t = \boldsymbol{x}_t{}'\boldsymbol{\beta} + \varepsilon_t$ where t indexes the observations, $\boldsymbol{x}_t$ is a vector of k observables, y_t is a scalar observable, $\boldsymbol{\beta}$ is a vector of k parameters to be estimated and ε_t is an unobservable assumed to be normally distributed with mean zero and unknown variance σ_ε^2. The corresponding variance of the dependent variable is $\sigma_y^2 = \boldsymbol{\beta}'\boldsymbol{\Sigma}_{xx}\boldsymbol{\beta} + \sigma_\varepsilon^2$ where $\boldsymbol{\Sigma}_{xx}$ refers to the covariance matrix of the explanatory variables. With standardized variables, the variance of the dependent variable is one and the variance of the residual is $1-R^2$ where R^2 is the squared multiple correlation coefficient. Substituting these into the variance equation $\sigma_y^2 = \boldsymbol{\beta}'\boldsymbol{\Sigma}_{xx}\boldsymbol{\beta} + \sigma_\varepsilon^2$ we obtain the result that the R^2 is the generalized beta-coefficient, $R^2 = \boldsymbol{\beta}'\boldsymbol{\Sigma}_{xx}\boldsymbol{\beta}$.

For a conventional prior, we cannot claim knowledge of the relative importance of the variables, nor can we claim any knowledge of the relationships between the coefficients. This, I propose, requires the conventional prior to have covariance matrix proportional to the identity matrix, $Var(\boldsymbol{\beta}) = v^2\boldsymbol{I}$, where v^2 is the variance that applies to all coefficients.

The quadratic form $\boldsymbol{\beta}'\boldsymbol{\Sigma}_{xx}\boldsymbol{\beta}$ can be written as $trace(\boldsymbol{\beta}'\boldsymbol{\Sigma}_{xx}\boldsymbol{\beta}) = trace(\boldsymbol{\Sigma}_{xx}\boldsymbol{\beta}\boldsymbol{\beta}')$, and we can replace matrix $\boldsymbol{\beta}\boldsymbol{\beta}'$ with it's prior expectation $E(\boldsymbol{\beta}\boldsymbol{\beta}') = Var(\boldsymbol{\beta})+ E(\boldsymbol{\beta})E(\boldsymbol{\beta}') = Var(\boldsymbol{\beta}) = v^2\boldsymbol{I}$,

[11] An economic theorist who chooses to explore the impact of x_1 on y but makes no mention in the theory of x_2, x_3, … must have chosen to focus on x_1 because of the probable importance of this variable compared with the others. That I conclude must be a reference to a beta coefficient, since I cannot think of any alternative that makes sense.

to obtain $E(R^2) = trace(\, v^2 \boldsymbol{\Sigma}_{xx}) = v^2 k$. Thus we have the prior variance of each beta-coefficient equal to the R^2 divided by the number of parameters, k, $v^2 = E(R^2)/k$.

Finally, to select a range of prior variances, we need to select a range of expected R^2.

$$v_L^2 = \frac{\min \mathrm{E}(R^2)}{k} \leq v^2 \leq \frac{\max \mathrm{E}(R^2)}{k} = v_U^2$$

Both ends of this interval of prior variances shift toward zero as the number of explanatory variables increases. The posterior ambiguity need not increase, however; that depends on the contribution of the additional variables to the model.

So here is my recommendation based on my own experience with regressions with economic data: The wideset context-minimal bound for the expected R^2 extends from 0.1 to 1.0. In anticipation that this interval implies fragile signs of the coefficients, the interval is split in half: a pessimistic view about the fit of the model with an expected R^2 extending from 0.1 to 0.5, and an optimistic view with an expected R^2 extending from 0.5 to 1.0. Thus there are three s-values reported: the context free s-values, and two context dependent s-values: pessimistic and optimistic regarding the quality of the fit of the model.

The prior standard errors associated with these three values for expected R^2 decline with the square root of the number of coefficients per the table below.

Conventional Prior Standard Errors

k	(0.1/k)^.5	(0.5/k)^.5	(1/k)^.5
1	0.31623	0.70711	1.00000
5	0.14142	0.31623	0.44721
10	0.10000	0.22361	0.31623
20	0.07071	0.15811	0.22361
50	0.04472	0.10000	0.14142
100	0.03162	0.07071	0.10000

This choice is analogous to the choice of conventional levels of statistical significance, 10%, 5%, or 1% and it plays the same role of determining the extent to which the hypothesis of no effect can be overcome by the data. If the expected R^2 is low, powerful data information is needed to overcome this class of priors, but if the expected R^2 is high, relatively weak data can be enough.

6.1 Favorite Variables Conventional Prior

The previous section offers a data analysis that is completely context-minimal. Press a button and you get the t-values and the s-values. No need to think. This section allows for "favorite" variables that are expected to be more important than the other variables. I have previously captured this idea with the words "free" and "doubtful" and used infinite prior variances for the free coefficients and any prior variance matrix on the doubtful variables. The free and doubtful variables bounds are usually too large to be useful, especially when the number of doubtful variables is large. The basic problems with this approach are two: the infinite prior variance on the free variables surely overstates their potential importance and the unbounded interval of prior covariance matrices is way too wide, allowing at the same time completely dogmatic opinions (zero prior variances) and completely uninformed opinions (infinite prior variances).

Here, to think about bounds for the prior variance we can use a block diagonal structure for the prior variance $E(\boldsymbol{\beta\beta}') = \begin{bmatrix} \nu_F^2\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \nu_D^2\boldsymbol{I} \end{bmatrix}$ where the subscripts F and D refer to the free and doubtful variables. Then the expected value of $(R^2) = trace\big(\boldsymbol{\Sigma_{xx}}E(\boldsymbol{\beta\beta}')\big) = k_F\nu_F^2 + k_D\nu_D^2$.

So here is my proposal. Take the overall lower bound of the expected R^2 to be 0.5. Let's suppose that the favorite variables are expected by themselves to make the R^2 at least 0.4 and the collection of other doubtful variables is expected to add another 0.1 to the R^2, thus adding up to 0.5 Then the lower bound prior variances depend on the number of free and doubtful variables (k_F and k_D) per the inequalities:

$$\nu_F^2 \geq \frac{0.4}{k_F}\ ,\ \ \nu_D^2 \geq \frac{0.1}{k_D}$$

Similarly, hold the upper value of the expected R^2 at one and distribute that between the free and doubtful variables, 0.8 to the free and 0.2 to the doubtful variables:

$$\nu_F^2 \leq \frac{0.8}{k_F}\ ,\ \ \nu_D^2 \leq \frac{0.2}{k_D}$$

7 Specification Ambiguity With The Prior Variance Bounded From Above and Below

The basis of the proposed conventional sturdiness measures is the following theorem which takes the prior covariance matrix to be bounded from both above and below. The lower bound for the prior covariance matrix excludes dogmatic priors that would impose hard linear restrictions which would be completely impervious to the data evidence, no matter how large the sample may be. The upper bound is necessary to limit the influence of unimportant variables in models with many explanatory variables.

Theorem 1 Leamer(1982,p729) Given that the prior variance matrix $\boldsymbol{V}$ is bounded from above and below, $\boldsymbol{V}^* \geq \boldsymbol{V} \geq \boldsymbol{V}_*$ **,** then the Bayes estimate $\widehat{\boldsymbol{\beta}}(\boldsymbol{V}) = \left(\boldsymbol{H} + \boldsymbol{V}^{-\mathbf{1}}\right)^{-\mathbf{1}}\boldsymbol{Hb}$ lies in the ellipsoid

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{f})\boldsymbol{G}(\widehat{\boldsymbol{\beta}} - \boldsymbol{f}) \leq c$$

where

$$\boldsymbol{G} = \left(\boldsymbol{H} + (\boldsymbol{V}^*)^{-1}\right)\left((\boldsymbol{V}_*)^{-1} - (\boldsymbol{V}^*)^{-1}\right)^{-1}\left(\boldsymbol{H} + (\boldsymbol{V}^*)^{-1}\right) + \ (\boldsymbol{H} + (\boldsymbol{V}^*)^{-1})$$

$$\boldsymbol{f} = \left[\boldsymbol{H} + (\boldsymbol{V}_*)^{-1}\right]^{-1}\left[\boldsymbol{H}\boldsymbol{b} + \left((\boldsymbol{V}_*)^{-1} - (\boldsymbol{V}^*)^{-1}\right)\left(\boldsymbol{H} + (\boldsymbol{V}^*)^{-1}\right)^{-1}\boldsymbol{H}\boldsymbol{b}/2\right]$$

$$c = \boldsymbol{b}'\boldsymbol{H}\left[\boldsymbol{H} + (\boldsymbol{V}^*)^{-1}\right]^{-1}\left[(\boldsymbol{V}_*)^{-1} - (\boldsymbol{V}^*)^{-1}\right]\left[\boldsymbol{H} + (\boldsymbol{V}_*)^{-1}\right]^{-1}\boldsymbol{H}\boldsymbol{b}/4$$

From this ellipsoid, the extreme estimates of the linear combination $\boldsymbol{\psi}'\widehat{\boldsymbol{\beta}}(\boldsymbol{V})$ are

$$\boldsymbol{\psi}'\boldsymbol{f} \pm \left(\boldsymbol{\psi}'\boldsymbol{G}^{-1}\boldsymbol{\psi}\right)^{1/2}(\boldsymbol{c})^{1/2}$$

The proposed measure of model sturdiness is the center of this interval divided by half the length

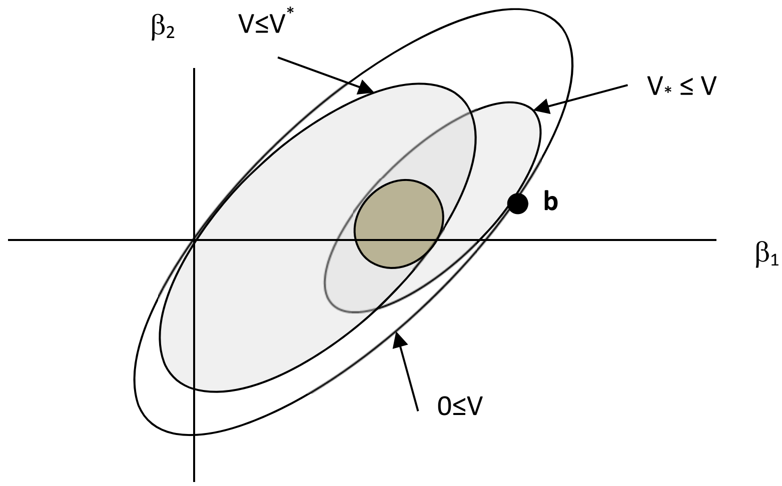
$$\boldsymbol{s} = \frac{\boldsymbol{\psi}'\boldsymbol{f}}{(\boldsymbol{\psi}'\boldsymbol{G}^{-1}\boldsymbol{\psi})^{1/2}(\boldsymbol{c})^{1/2}}.$$

Four ellipses of estimates implied by this theorem are illustrated in Figure 3. The largest ellipse has an unshaded interior. This “**feasible ellipsoid**” is the set of estimates corresponding to the full set of symmetric positive-semi-definite prior variance matrices, $\boldsymbol{0} \le \boldsymbol{V}$. The skin of this feasible ellipsoid is the set of estimates subject to homogenous linear restrictions. (e.g. lines through the origin) This skin includes the OLS estimate **b**, and the prior mean **0**. Properties of this feasible ellipsoid are discussed in an Appendix.

This figure has two lightly shaded ellipses, each of which includes one and only one point on the feasible ellipse. The set of prior covariance matrices only bounded from above, $\le \boldsymbol{V}^*$, produces the ellipse that includes the origin but excludes the OLS estimate **b**, which needs an infinite prior covariance matrix to be obtained. The set of prior covariance matrices only bounded from below,$\boldsymbol{V}_* \le \boldsymbol{V}$, produces the ellipse the excludes the origin, which requires a zero prior covariance matrix to be obtained, but includes the OLS estimate **b.**

The darkest ellipse in the interior of all the others is the ellipse corresponding to a set of prior covariance matrices bounded both from below and from above, $\boldsymbol{V}_* \le \boldsymbol{V} \le \boldsymbol{V}^*$**.** The extreme values of the coefficients on this ellipse determine the proposed measures of model ambiguity – when this ellipse is large alternative “reasonable” ways to describe the prior state of mind lead to substantially different estimates, but when this ellipse is small, details of the prior are less important. As the sample size grows and the data become more important, this ellipse will shrink and move toward the OLS point **b**. In other words, the treatment for model ambiguity is the same as the treatment for estimation uncertainty: more and better data.

Figure 3 Four Ellipses of Estimates



7.1 g-priors

I am advocating that the upper and lower prior covariance matrices be proportional to the identity matrix but an alternative to the identity matrix is the sample variance with the interval of prior covariance matrics bounded above and below by a "g-prior" covariance matrix, $v_L^2\mathbf{H^{-1}} \leq \boldsymbol{V} \leq v_U^2\mathbf{H^{-1}}$. Using Theorem 1, the s^2 value under these conditions becomes:

$$\boldsymbol{s^2} = \left(\frac{(\boldsymbol{\psi}'\boldsymbol{b})^2/\boldsymbol{\psi}'\boldsymbol{H^{-1}}\boldsymbol{\psi}}{\boldsymbol{b}'\boldsymbol{H}\boldsymbol{b}}\right)\left(\frac{2\dfrac{E_*(R^2)E^*(R^2)}{k(1-R^2)/n} + E_*(R^2) + E^*(R^2)}{\left(E^*(R^2) - E_*(R^2)\right)}\right)^{\mathbf{2}}$$

The first term in this expression is $(\boldsymbol{\psi}'\boldsymbol{b})^2/\boldsymbol{\psi}'\boldsymbol{H^{-1}}\boldsymbol{\psi}$, the square of the Z-statistic for testing $\boldsymbol{\psi}'\boldsymbol{\beta} = 0$, divided by $\boldsymbol{b}'\boldsymbol{H}\boldsymbol{b}$, the χ^2-statistic for test the multivariate hypothesis $\boldsymbol{\beta} = 0$. This ratio is the only data-dependent part of the expression and is invariant to linear transformations of the data. Thus there is a perfect one-to-one mapping connecting the traditional measure of sampling uncertainty, $(\boldsymbol{\psi}'\boldsymbol{b})^2/\boldsymbol{\psi}'\boldsymbol{H^{-1}}\boldsymbol{\psi}$, and the proposed measure of sturdiness. While the Z statistic is normally compared with 1.96 to determine statistical certainty, for sturdiness the comparison should be with the χ^2-statistic $\boldsymbol{b}'\boldsymbol{H}\boldsymbol{b}$**,** which like a Z statistic grows with sample size, which means that this part of the sturdiness statistic does not grow with sample size. If the lower expected R^2 is zero, the other part of the expression is equal to one. In other words, if you are willing to entertain the dogmatic zero prior as a possibility, increasing the sample size is not going to increase the sturdiness of the inference. As the sample size grows, the second term grows like

$$\left(\left(\frac{n}{k}\right)\frac{2E_*(R^2)E^*(R^2)}{(1-R^2)\left(E^*(R^2)-E_*(R^2)\right)}\right)^2$$

which increments with n/k, guaranteeing that as the sample size grows, the sturdiness grows without limit, and model ambiguity as well as statistical uncertainty dissipate. (Best to have a non-zero lower bound to eliminate dogmatic priors.)

8 Reporting Style for the Determinants of Economic Growth

This section offers a proposed reporting style that combines estimation information with model ambiguity information in a single table. The example is a country growth regression based on data from Sala-i-Martin, Doppelhofer and Miller (2004) which includes the sixty-seven explanatory variables listed in Table 1. The dependent variable is the average rate of growth of real (PPP adjusted) GDP per capita from 1960 to 1996 of eighty-seven countries.

Three different treatments are provided. For expository purposes, I begin in Section 8.1 with an equation that includes only 14 explanatory variables. These are my favorites, not necessarily yours. This is for discussion only, and is not an endorsement of a study with a subset. Better to cast the net widely from the start, which is what is done in Section 8.2 which has an analysis of the growth equation with the full set of 67 explanatory variables. Then in Section 8.3, I report a regression equation with all 67 explanatory variables but with the 14 treated as favorites per the discussion above. In other words, I first concentrate all the energy on the data on 14 favorite coefficients, then I spread the data over all 67, and last spread the data over all 67 but in a way that favors the 14.

8.1 Fourteen of Sixty-Seven Explanatory Variables

Table 2 has the proposed format for reporting regressions. To try to gain some intellectual control over this crowded shelf of alternative explanatory variables, I have assigned each to one of five categories: (1) catch-up based on initial per capita GDP, (2) culture, (3) geography, (4) government, and (5) resources. The full mapping of variables into categories can be found in the regression equation with all 67 variables reported in Table 5.

We have to include the catch-up effect of initial per capita GDP. Surely unpriced technological transfer is a big part of the growth story. My favorite fourteen variables includes only one from the culture category. I am an economist after all and I chose a variable that I think would facilitate contracting across national borders: the fraction who speak a foreign language. I really wanted to include the fraction English speaking population, but I held myself back on that one. From the geography variables I chose absolute latitude. A hot climate has traditionally been bad for the critical manufacturing sector because it is hard to operate (expensive) machinery for long hours at high pace in hot and humid climates. To spread the fixed capital costs over the largest number of units produced, manufacturing clustered together in the northern latitudes of both North America and Europe. In addition to cool climates, footloose manufacturing needs logistics infrastructure to get the goods to the global market, thus I have included the fraction of the land near navigable water. Also because I am an economist, I am inclined to think that the government's job is not to mettle too much with the private enterprise system, so the size of government might be a problem as would interference with cross-border trade which may or may not be captured by the outward orientation variable. Wars can't be good for growth, while the former British colonies might have a better government than most, something which represents unwarranted prejudice on my part. Among the resource variables, I chose higher education as one of my favorites. It's what I do and probably you too. It must be important. Based on my understanding of the lifetime profile of earnings which peaks in midlife, I include the youth variable which should be good for growth and the elderly variable which should not. Last are two variables that measure non-manufacturing exports, primary exports and oil producing, which will be good only in periods of elevating terms of trade.

I know. I am totally prejudiced for no good reason but these are my favorites, nonetheless. If you were doing the hard work, you could pick your own.

The fourteen explanatory variables in Table 2 are organized by category and are sorted within these categories by the OLS t-values which are reported in the eighth column. All variables are standardized to have unit standard errors, and the coefficients consequently are "beta-coefficients" which measure the number of standard deviations in growth of real GDP per capita that is associated with a one standard deviation change in an explanatory variable.

The table is divided between columns that describe the sampling uncertainty and columns that describe the model ambiguity. [12] The first five columns report five different sets of estimates of these beta-coefficients and the next five columns report the corresponding t-values. These are measuring sampling uncertainty. The last three columns have three different sets of s-values that characterize the model ambiguity

The fifth column labelled b-OLS reports the usual regression estimate with all 14 variables included. To the left of this column of OLS estimates are three "Bayes" estimates which shrink the OLS coefficients toward zero, greater shrinkage as we move toward the left in this table as the prior distribution gets more concentrated around zero. The first column labelled "b-SIMPLE" refers to 14 different regressions, each with one and only one explanatory variable. Although these one-at-a-time estimates are not obtainable with the kind of shrinkage estimator that is used here, a theorem presented in an appendix indicates that the most extreme shrinkage (all estimates close to zero) produces estimates with the same signs as b-SIMPLE, and the estimation problem roughly speaking is to seek compromises between the signs of the b-SIMPLE estimates and the signs of the b-OLS estimates. When these signs are in conflict the b-SIMPLE coefficient is printed in bold. The initial per capita GDP is an example, with a positive

[12] The Knightian words, "risk" and "uncertainty", in modern usage do not convey the Knightian distinction between known and unknown probabilities, and I prefer to use "uncertainty" and "ambiguity".

SIMPLE coefficient, suggesting that the rich get richer, while the b-OLS estimate is negative, suggesting that, after controlling for all the other "initial conditions" in the equation, there is a convergence effect picked up by this variable. To put the point a bit aggressively, the simple regression is a feature of the data, but the OLS multiple regression is something cooked up by the analyst when he or she chose the control variables. The s-values in the last three columns are intended to determine if that "cooking" is credible, answering whether reasonable changes in the choice of control variables can alter the sign of the estimates. More on the subtle meaning of the coefficient of initial GDP per capita below.

Outward orientation also has a positive simple correlation but a negative partial correlation in this 14 variable regression. In other words, outward orientation seems favorable to growth when it stands alone but unfavorable (though statistically insignificant) when one controls for the other 13 variables. Also of note, on a variable-by-variable basis, youth are bad for growth but elderly are good according to their simple correlations, but controlling for all the other variables, the opposite signs emerge.

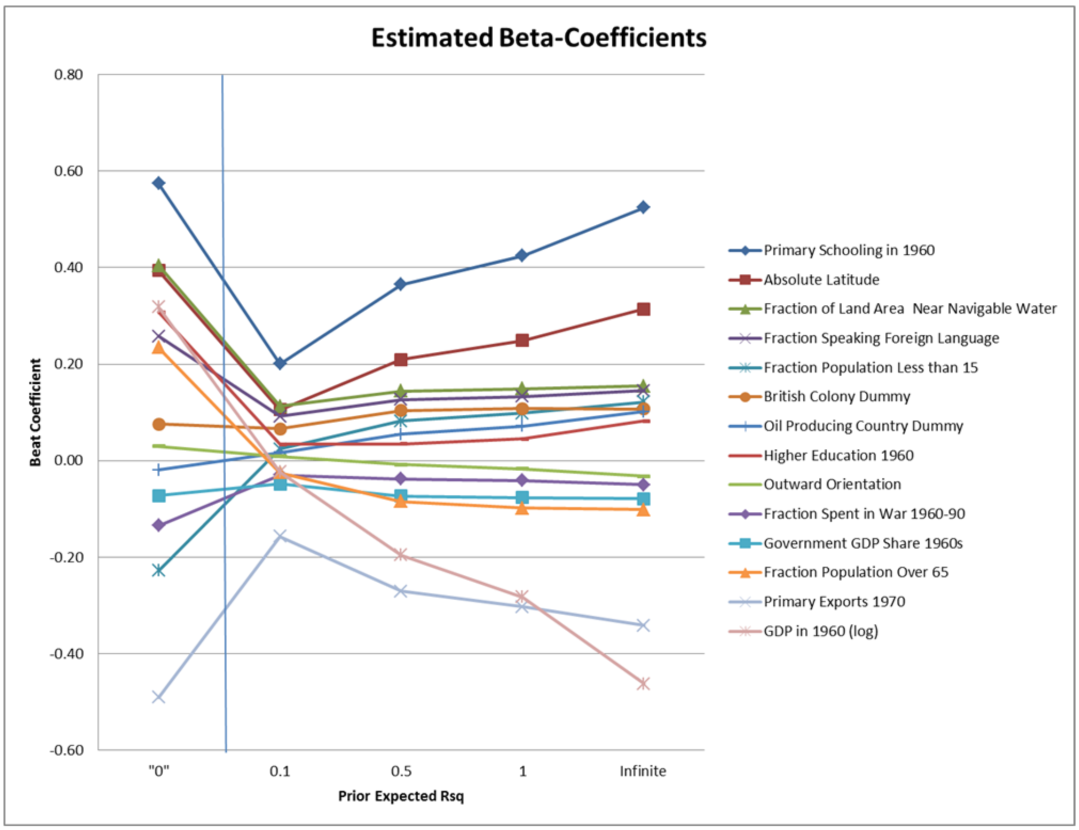
The first row labelled "Prior R-square" refers to the prior expected R-squared, with values equal to "0", 0.1, 0.5, 1.0 and infinite. The quotations around "0" alert you to the fact that what is placed in this column is not a shrunken estimate like the ones in the other columns but only the one-variable at a time regression. To interpret this as a shrunken OLS estimate, it is important that you ignore the magnitudes and look only at the signs, and think of "0" as "almost zero."

The estimates as a function of the prior expected R-square are illustrated in Figure 4 with the univariate regressions at the extreme left indicated by "0" and separated with a vertical line to make sure you remember that these are not shrinkage estimates in the same family as the other estimates. Use these to identify the signs when the shrinkage is extreme and to contrast with the signs of the OLS estimate corresponding to the "infinite" prior R-squared. The labelling in this figure is ordered by the size of the OLS beta-coefficient, from the largest positive (primary schooling) to the largest negative (initial per capita GDP).

As the "Prior R-square" decreases, the prior variances get small, which allows less play in the estimates and more shrinkage toward the origin, something which is very evident in this figure. The coefficient of initial per capita GDP stands out in this picture with a large negative OLS estimate that dives toward zero as the vector of estimates is shrunk toward zero. This might be taken as a reminder that this coefficient and in general the hypothesis of "convergence" is not well defined, since it depends on the other variables in the equation that identify the initial conditions, such as population fractions and education levels. If variables were included that perfectly explained the variability of per capita GDP in 1960, this variable would drop out from the equation. Referring to the usual growth accounting, the coefficient on initial per capita GDP could thus be interpreted as the effect of TFP (total factor productivity), meaning the component of variability that is not explained by all the other variables in the equation. To pursue this idea, I report in Table 3 a regression explaining the logarithm of initial per capita GDP with the other 13 variables (including the concurrent war variable), and the residuals from this equation in Table 4. If this is really TFP, it's a strange estimate with Trinidad and Tobago have the greatest TFP followed by Venezuela and South Africa, and with Taiwan and Korea among the worst.

This should really raise alarm bells about the difference between simple and partial correlations, and in particular the fact that the partial correlations depend on the control variables in the equation. What do those 14 coefficients really represent? Even more so: What could the coefficients in a 67 variable regression possibly refer to? Is there a sensible hypothetical that holds sixty six fixed and varies only one???

Figure 4 Estimates of Beta Coefficients: 14 Variable Model



The expected prior R-square in Figure 4 applies when the prior covariance matrix is proportional to the identity matrix, with the beta-coefficients independent and identically distributed. Larger variances (more diffuse priors) are needed to allow the estimates to conform more closely with the ordinary least squares estimates. Indeed it takes infinite variances to reproduce the OLS estimates exactly. But I am proposing that an expected R-squared equal to one with the i.i.d. assumption on the beta-coefficients is as diffuse a prior as we need to entertain. The estimates with expected R-squared equal to one are all the same signs as OLS estimates and close in magnitude. Thus this amount of shrinkage does not change individual estimates very much, though as reported in the last two rows of Table 2 the R-squared has been reduced from 0.537 to 0.436 and the averaged squared coefficient from 0.058 to 0.036.

The t-values in excess of two in absolute value are shaded in Table 2. In the 14-variable OLS equation initial per capita GDP, latitude, primary schooling and primary exports all have statistically significant coefficients. These t-values generally decline as the coefficients are shrunk toward the origin. Only primary schooling and primary exports maintain their statistical significance as the coefficients are shrunk closer to zero.

The last three columns report three different sets of s-values. The s-values in excess of one are shaded. These are the sturdy coefficients. The first column of s-values is built on the broad range of prior expected R^2 between 0.1 and 1.0. Only primary schooling has a sturdy coefficient with this broad set of prior distributions. From this we conclude: The presence of a population with a high rate of primary school education has a statistically significant and sturdy effect on predicted economic growth, holding fixed the other 13 variables in the equation. (Does that mean anything?)

The next two columns split this interval in two, from 0.1 to 0.5, and from 0.5 to 1.0. The sub-interval with a prior expected R^2 pessimistically between 0.1 and 0.5 produces s-values reported in the second column of s-values that are pretty close to the first column of s-values corresponding with the broad range of priors, while the third column has larger s-values. The interpretation of this result is that the ambiguity in signs of the coefficients occurs when priors are allowed that shrink the estimates close to the origin. This seems apparent in Figure 4 because the effect of the prior isn't that great for expected R^2 of 0.5 and above, while the shrinkage is substantial from 0.5 to 0.1.

If one were more optimistic about the collective explanatory power of these 14 variables, and adopted the lower bound of 0.5 for the expected R-squared, then 10 of the 14 coefficients are judged to be sturdy. The losers which are fragile even at the 0.5 level are: outward orientation, higher education, and the two demographic variables (fraction less than 15 and fraction over 65)

Both s-values and t-values grow with the sample size, since more data relieve both sampling uncertainty and model ambiguity, but in a finite sample, as we have seen, it is possible for these measures to diverge, but it is also possible that they offer essentially similar information. In this fourteen variable study, the s-values and t-values are highly correlated, as illustrated in Figure 5 which includes the s-values in the last two columns of Table 2 with the non-overlappping high and low R^2 intervals. [13] Moreover, the scatter of the two s-values reveals they are related almost perfectly 5:1.

---

[13] (Referring back to Figure 1 a sufficient condition for the s-values and t-values to be proportional to each other is if the shape of the shaded ellipse of estimates is the same as the shape of a likelihood ellipse and if the center of this shaded ellipse is on a line from OLS to the origin.)

Figure 5 s-values and t-values, 14 variable case

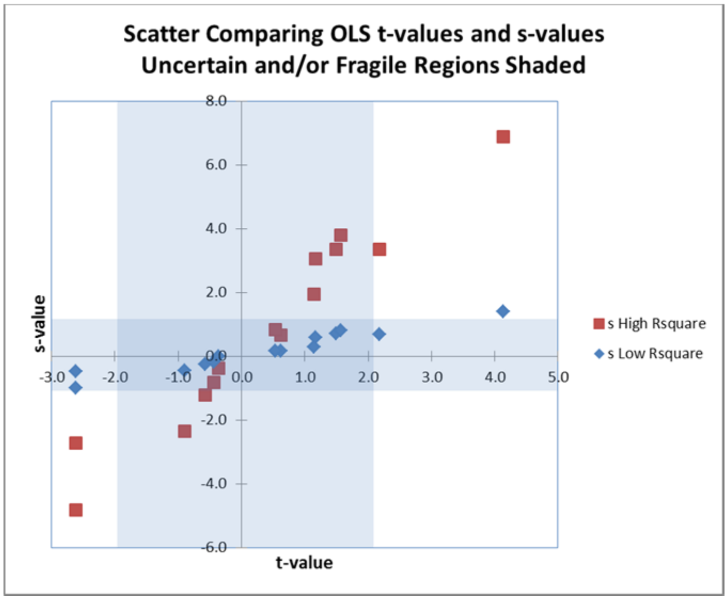
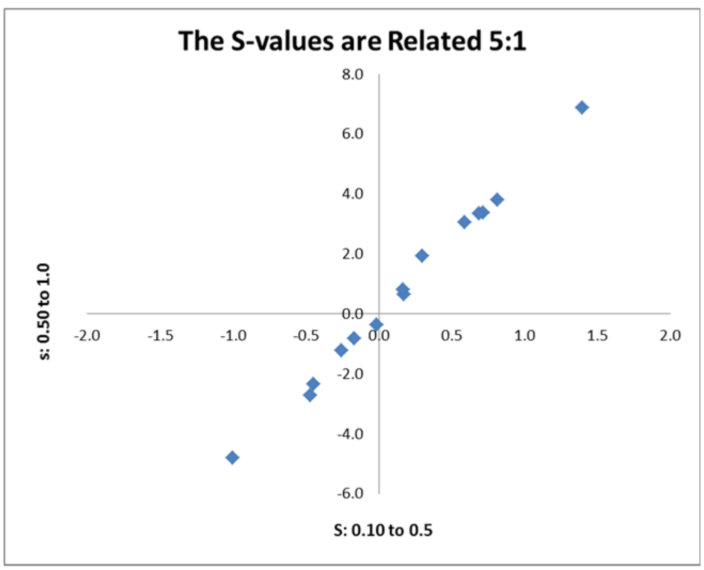


Figure 6 Two s-values, 14 variable case



8.2 Sixty-Seven Explanatory Variables

Table 5 contains the results for a regression with the full set of 67 variables used by Sala-i-Martin et. al. (2004). The t-values in excess of two in absolute value are shaded. Except for the simple univariate regressions there are not many of these. For the OLS estimate, only the price of investment goods is statistically significant.

The outliers among the beta-estimates are highlighted with a box. These are the coefficients for variables 34 and 35, two ways of measuring the size of government. One coefficient is the huge negative -3.401 and the other is the huge positive 3.379, almost equal but opposite in sign. Both are statistically insignificant and they have a correlation of -0.999. This is a symptom of two variables measuring the same thing, but notice how well the shrinkage has performed in erasing this abnormality. With the expected prior R^2 equal to one, the estimates of these two coefficients are -0.40 and -0.47, essentially the same effect for essentially the same variable. In this case, Bayesian shrinkage is working nicely. But Bayesian shrinkage doesn't create statistically significant coefficients except that it allows the East Asian Dummy to have a t-value in excess of 2. In other words, there remains a lot of sampling uncertainty here, even when the data are helped out with an informative prior distribution.

The third from last column in this table reports the s-values from the broadest range of prior covariances associated with the interval of prior expected R^2 from 0.1 to 1.0. The s-values in excess of one would be shaded but there are none; everything is fragile. As we discovered from the study of the 14 variable model, the s-values reported in the next to last column are about the same as the s-values reported in the third from last column, none greater than one in absolute value, while the it is the last column with a restricted set of prior covariance matrices with the interval of prior expected R^2 from 0.5 to 1.0 where there are some sturdy coefficients. These are the Fraction Confucian, the fraction Buddhist, the East Asian Dummy, the real exchange rate distortions, the price of investment goods and the primary school enrollment rate. In other words, there is a lot of model ambiguity here.

There are some important similarities in these results and the 14-variable results. As can be seen in Figure 7, here as before there is a high degree of correlation between the OLS t-values and the Bayesian s-values at the 0.1 level. The correlation is substantial but far from one, which means that s-values are offering information not captured by the t-values. As can be seen in Figure 8 the narrower interval of priors with a lower bound for the prior expected R^2 of 0.5 provides s-values that are about three times as large as the wider interval of priors with a lower bound for the prior expected R^2 of 0.1, compared with five times in the 14 variable case.

In summary, when we jump from a subset of 14 variables to a full complement of 67 variables, things have gone from not-so-good to really troubling since the fairly weak data resource has been spread very thinly over such a large set of parameters that only one regression coefficient is statistically certain and no coefficient is sturdy. Next we will see what happens if we focus the data resource on the same 14 variables we have studied separately.

Figure 7 t-values and s-values are correlated, 67 variable case.

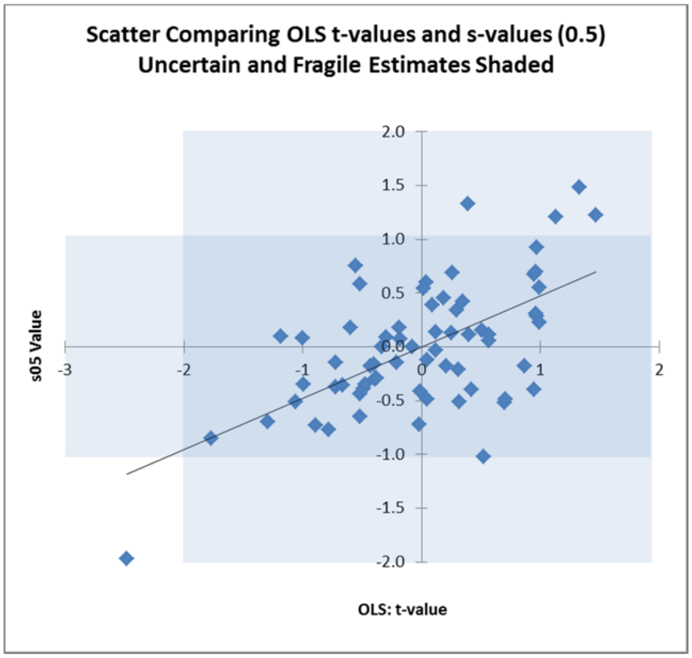
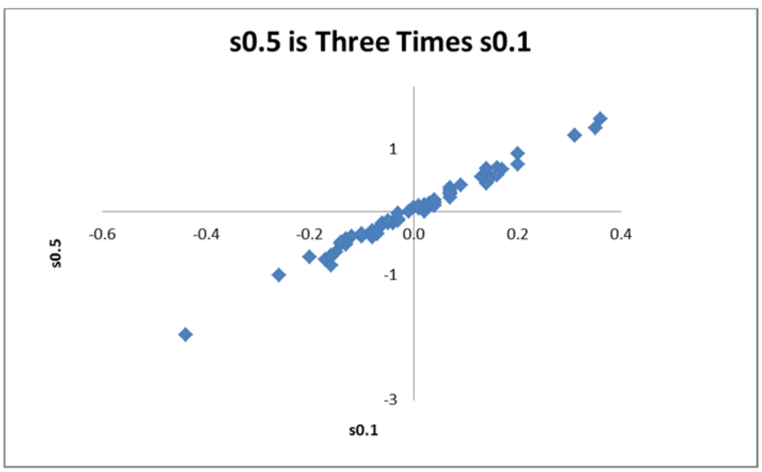


Figure 8 The Two Sturdiness Measures Are Highly Correlated



8.3 Fourteen Favorites among Sixty Seven Explanatory Variables

In an effort to make these data speak more clearly I will focus the sample energy on a subset of explanatory variables per the discussion above of favorite variables priors. For my favorites, I chose one from each category. Table 6, which has the new results, is headed by my 14 favorite variables.

The reason that the OLS t-values are so low is that the energy of the sample has been dissipated over 67 parameters. The Bayes t-values show what happens if the data wealth is concentrated on the fourteen favorite variables. Most of the Bayes t-values are larger in absolute value than the OLS t-values, but only primary schooling has emerged as statistically significant at traditional levels.

There are six s-values in excess of one in absolute value in the 67 variable, no-favorite case in Table 5 and also six in the 67 variable, 15 favorite results in Table 6. These have been summarized in Table 7 to make comparison easier. The effect of the favoritism has allowed the favorites, initial per capita GDP, fraction speaking a foreign language, and primary exports to emerge with sturdy signs, and has allowed a substantial increase in the s-value for primary schooling, which is sturdy in both cases. The East Asian Dummy and the Investment price are in both lists, and thus survive even though they are not favorites, while fraction Confucian, fraction Buddhist, and real exchange rate distortions are no longer sturdy when other variables are favored.

My bottom line: it's pretty hard to squeeze something of value from cross-country regressions with inclusive lists of explanatory variables.

9 Concluding Comments

Table 7 compares s-values of my fourteen favorite variables and other coefficients that are sometimes sturdy for different intervals of prior expected R2 and for three treatments of favoritism: 67 variables with no favoritism, 14 favorites among the 67, and only the 14 favorite variables included. The important but not altogether surprising message from this table is that there is not much to be learned from a study of these data unless you are willing to favor some of the 67 variables over others. To get ten sturdy coefficients out of fourteen, we need to include only these 14 variables and also adopt the high expected R^2 interval. With only 14 variables, only the primary school enrollment rate is sturdy if for the broad class of priors. And at the extreme left in this table we find no sturdy coefficients with 67 variables and no favorites. Of these 14 favorite variables, it is primary school enrollment that stands out, with the highest s-value in every instance, and with an s greater than one in absolute value except in the kitchen-sink regression reported at the far left.

Though thousands of regressions with growth as the dependent variable have probably been estimated by economists, the intensive mining of these data was much stimulated by Barro(1991). My negative conclusions[14] about the usefulness of data mining in this setting contrast with more optimistic conclusions of Sala-i-Martin et. al.(2004, p.833): "In fact, we find

[14] For other pessimistic voices see Brock, Durlauf, and West(2003), Brock and Durlauf (2001), Durlauf(2000), Levine and Renelt (1992), and Rodrik(2012) but a more optimistic view is offered by Fernández, Ley and Steel(2001), who like Sala-i-Martin also use Bayesian Model Averaging.

that about one-fifth of the 67 variables used in the analysis can be said to be significantly related to growth while several more are marginally related. The strongest evidence is found for primary schooling enrollment, the relative price of investment goods and the initial level of income where the latter reflects the concept of conditional convergence." I consider their approach quite interesting and a worthy competitor to the treatment described in this paper. Their method is a Bayesian mixture of the all-subsets estimation formed by including and excluding the 67 variables in all 2^{67} different ways, with a prior on the size of the model that favors models that are not "too big." Commenting on the similarities and differences takes too much space to include in this paper and instead forms a companion paper Leamer(2014): "S-values and All-Subsets Regressions." I like mine better, but I like theirs too.[15]

The goal of this paper is something that many Bayesians would object to: formulation of a conventional prior distribution that can be used in many if not all settings. One of the big messages of the Bayesian approach is that the context matters, and thus conventional priors need to be strongly discouraged. While I share that opinion, I note that it tells us *what not to do,* but *not what to do*. The Bayesian "what to do" is hopelessly naïve – simply describe your state of mind with a specific probability distribution. No analyst is going to feel honestly comfortable with that task, and all but the true believers are going to feel extreme discomfort when the results are described to an incredulous audience. In any case, the proof of the failure of Bayesian thinking about context-dependent prior distributions is the near complete absence of any such Bayesian studies.

The tension between knowing for sure what not to do, but having no real idea what to do, can be relieved with a conventional starting point that is not intended to be the final say, but only a wisely chosen point to begin the conversation about how the context matters.

As with any tool, there is room for mischief and abuse. I and others[16] are greatly concerned with the unwise use of hypothesis testing at conventional levels of statistical significance[17], and the

[15] Foreshawdowing: To study the model ambiguity, Sala-i-Martin et. al.(2004) produce a one-dimensional set of results as the prior expected model size is varied, which would be analogous to the one dimensional "ridge trace" when the prior expected R^2 is varied. Either one of these traces takes the basic structure of the prior as given, and makes nt allowance of potential context-dependent changes in variables, e.g. using sums and differences. In contrast, my model ambiguity ellipsoid, which is a consequence of upper and lower bounds for the prior covariance matrix, has full k-dimensional volume, a does allow for "local" reparameterizations.

[16] Leamer(1978, pp.93-98), Kruskal (1978), Ziliak and McCloskey (2008),

[17] A critical problem with any fixed level of statistical significance is that Type II error is altogether ignored, and as the sample size increases the added information in the data is fully deployed in reducing the probability of Type II error. Except for highly unlikely lexicographic preferences, it is necessary to use the additional information to reduce both error probabilities,

confusion of "significant" with "important" when "statistically significant" only means "measurable". Conventional measures of model ambiguity could also be misused, but I think that the conventional measures of model ambiguity proposed here are not so dangerous as the conventional choices of significance levels for hypothesis testing since at least the s-values are appropriately sensitive to both the sample size and the number of regression coefficients. Where nontransparent mischief can surely be done is in the choice of statistical "horizon" – the set of variables within which estimation and sensitivity analysis are performed.

But the reality is that until we have a cultural understanding about how the context should be allowed to matter when data are studied, data analysis will combine context-free measures of statistical uncertainty and (I hope) context-minimal measures of model ambiguity, supplemented with verbally expressed wisdom offered by analysts who understand the context. The s-values when there are favorite variables are an important first step in the direction of context-dependent data analysis.

The sturdiness statistics proposed here fit the circumstances in the sense of depending appropriately on the breadth of alternative models as measured by the number of variables considered and also in the sense of increasing properly as the sample becomes more informative and model ambiguity is relieved. These statistics are thus going to raise alarms in settings with an abundance of variables and weak data, such as a study of the determinants of long-run economic growth. That's a good thing. We shouldn't be pretending that there is a lot to learn about what makes countries grow from context-free regressions with countless variables.

10 References

Barro, Robert J. "Economic Growth in a Cross Section of Countries." *The Quarterly Journal of Economics*, 106 (No. 2 1991): 407-43.

Brock, W.A., Durlauf, S.N. and West, K.D. (2003). Policy evaluation in uncertain economic environments. *Brookings Papers on Economic Activity* 1:235-322.

Brock, William A. and Steven N. Durlauf (2001) "Growth Empirics and Reality," *The World Bank Economic Review* , Vol. 15, No. 2 (2001) , pp. 229-272

Chamberlain, Gary and Edward E. Leamer (1976) , "Matrix Weighted Averages and Posterior Bounds," *Journal of the Royal Statistical Society*, B, 38, 73-84.

---

and thus make the level of statistical significance decline with sample size. Thus fixed conventional significance levels are both arbitrary and mistaken, while a conventional expected R^2 is only arbitrary.

Durlauf, Steven N.(2000) Review of Growth, Inequality, and Globalization: Theory, History, and Policy. by Philippe Aghion; Jeffrey G. Williamson, *Journal of Economic Literature* , Vol. 38, No. 3 (Sep., 2000) , pp. 637-638

Fernández, Carmen , Eduardo Ley and Mark F. J. Steel(2001), "Model Uncertainty in Cross-Country Growth Regressions," *Journal of Applied Econometrics*, Vol. 16, No. 5 (Sep. - Oct., 2001), pp. 563-576

Hoerl, Arthur E. and Robert W. and Kennard (1970), "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics,* Vol. 12, No. 1 (Feb 1970), 55-67.

Kruskal, William (1978) "Tests of Significance" in William H Kruskal and Judith M. Tanur ed. (1978) *International Encyclopedia of Statistics*, The Free Press: New York, pp. 944-958

Leamer, Edward E. (1975) "A Result on the Sign of Restricted Least Squares Estimates," *Journal of Econometrics,* 3 (1975), 387-390.

Leamer, Edward E. (1978) *Specification Searches: Ad Hoc Inference With Nonexperimental Data*, New York: Wiley.

Leamer, Edward E. (1982), "Sets of Posterior Means with Bounded Variance Priors," *Econometrica,* Vol. 50, No. 3 (May 1982), pp. 725-736.

Leamer, Edward E (2014), "S-values and All-Subsets Regressions," working paper.

Leamer, Edward E. and Gary Chamberlain (1976) "A Bayesian Interpretation of Pretesting," with G. Chamberlain, *Journal of the Royal Statistical Society*, Series B, 38 (No. 1, 1976), 85-94.

Ley, E., Steel, M.F.J., (2009.) "On the effect of prior assumptions in Bayesian model averaging with applications to growth regression." *Journal of Applied Econometrics* 24, 651–674.

Levine, Ross and David Renelt (1992) "A Sensitivity Analysis of Cross-Country Growth Regressions," The American Economic Review, V 82, Issue 4 (Sept), 942-963.

Raftery, A.E. (1995). Bayesian model selection for social research. *Sociological Methodology*, 25:111{163.

Rodrik, Dani (2012)"Why We Learn Nothing from Regressing Economic Growth on Policies," Seould Journal of Economics, Vol 25, No. 2, 137-

Sala-I-Martin, Xavier X. (1997), "I Just Ran Two Million Regressions" , *The American Economic Review,* Vol. 87, No. 2, Papers and Proceedings . (May, 1997), pp. 178-183.

Sala-i-Martin, Xavier, Gernot Doppelhofer and Ronald I. Miller (2004) , "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," *The American Economic Review*, Vol. 94, No. 4 (Sep., 2004), pp. 813-835

Zellner, A., (1986) On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In: Goel, P.K., Zellner, A. (Eds.), Bayesian Inferenceand Decision Techniques: Essays in Honour of Bruno de Finetti. North-Holland, Amsterdam, pp. 233–243

Ziliak, Stephen, and McCloskey, Deirdre, (2008). *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*. Ann Arbor, University of Michigan Press, 2009.

11 Appendix: Properties of the Feasible Ellipsoid

The most extreme special case of Theorem 1, discussed in Leamer and Chamberlain(1976), has no constraint on the prior variance matrix except that it is positive semi-definiite. This produces an ellipsoid of feasible estimates which has the same shape as a likelihood ellipse, but located at the OLS estimate divided by two, passing through both the OLS estimate **b** and the prior mean **0**.

Theorem 2 (Feasible Ellipsoid) The matrix weighted average $\hat{\boldsymbol{\beta}}(\boldsymbol{V}) = \left(\boldsymbol{H} + \boldsymbol{V}^{-1}\right)^{-1}\boldsymbol{Hb}$ for any positive semidefinite matrix **V** lies in the ellipsoid

$$\left(\hat{\boldsymbol{\beta}} - \frac{\boldsymbol{b}}{2}\right)\boldsymbol{H}\left(\hat{\boldsymbol{\beta}} - \frac{\boldsymbol{b}}{2}\right) \leq \frac{\boldsymbol{b}'\boldsymbol{Hb}}{\boldsymbol{4}}$$

and conversely any point in or on this ellipsoid is equal to a Bayes estimate $\hat{\boldsymbol{\beta}}(\boldsymbol{V}) = \left(\boldsymbol{H} + \boldsymbol{V}^{-1}\right)^{-1}\boldsymbol{Hb}$ for some value of **V.**

Discussion: This set of equally valid estimates is so large that it is practically useless. The good news is very limited: the only orthant that is *not* attainable is the orthant opposite the orthant of the simple correlations, which is demonstrated below.

If this were an accurate description of your state of mind, you might as well not study any data since no data set is capable of materially relieving your extreme state of ambiguous doubt. To make some progress, you need to deploy a narrower set of prior covariance matrices.

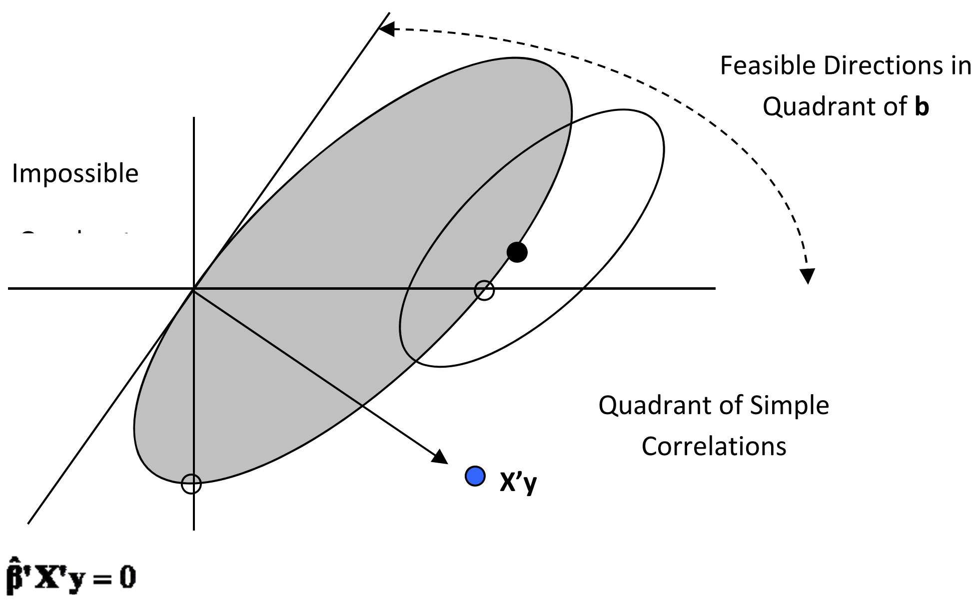
This extreme ambiguity is a surprising property of matrix weighted averages. If ***b*** were a single parameter, the weighted average would lie between zero and the OLS estimate. For matrix weighted averages it cannot be as simple as that, but still $\hat{\boldsymbol{\beta}}$, which is an average of the zero vector and the OLS vector **b,** must in some sense lie between the two. One might conjecture that the orthant of the OLS estimate ***b*** would at least be minimally preferred in the sense that $\hat{\boldsymbol{\beta}}$ and **b** would necessarily have at least one sign in common. Expressed differently, this conjecture is that it is impossible for $\hat{\boldsymbol{\beta}}$ and **b** to lie in opposite orthants. It is surprising that this is not true, and it is doubly surprising that it is the vector of unconditional covariances, **X'y** that actually plays this role. I show below that $\hat{\boldsymbol{\beta}}$ and **X'y** must have at least one sign in common. Thus the advice: For determining the signs of the coefficients, be sure to take a look at the simple correlations.

The results of this section are illustrated in Figure 9 which includes an OLS estimate **b** in the positive quadrant. Surrounding the OLS estimate **b** is a typical likelihood ellipse. The other shaded ellipse in this figure is the set of all possible compromise estimates mapped out as the prior covariance matrix **V** is varied over all symmetric positive semi-definite matrices. The skin of this ellipse is the set of constrained estimates when some linear combination is set exactly to zero.

This shaded "feasible ellipse" covers three quadrants, including the (-,-) quadrant with signs opposite the OLS **b**. The quadrant that this feasible ellipse does not intersect is the (-,+) quadrant. The two small circles on this ellipse are the two constrained estimates when one or the other coefficient is set to zero. Thus the vector of simple correlations is (+,-), which is the preferred orthant in the sense that the compromise estimates between **b** and zero take on all signs accept the opposite of (+,-).

The vector **X'y** is the inward normal of this feasible ellipse at the origin, and $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = 0$ is the tangent hyperplane to that inward normal. The picture has a feasible ellipse which includes values in all orthants but one: the orthant opposite the vector of simple correlations. Note also that the direction of feasibility (relative sizes of the coefficients) is unlimited in the orthant of the simple correlations, **X'y**, but is limited in the other orthants. That is another sense that the orthant of the simple correlations is special.

Figure 9 The Feasible Ellipse and the Impossible Orthant



Generally, there is only one excluded orthant – the orthant opposite the orthant of the simple correlations.

Theorem 3: By varying the prior covariance matrix, the vector of estimates $\hat{\boldsymbol{\beta}}(\boldsymbol{V}) = (\boldsymbol{H} + \boldsymbol{V}^{-1})^{-1}\boldsymbol{H}\boldsymbol{b}$ can lie in any orthant but one: the orthant opposite the vector of simple correlations, X'y.

Proof: The impossibility of the orthant opposite the orthant of the simple correlations is straightforwardly established by premultiplying the posterior mean $\widehat{\boldsymbol{\beta}}(\boldsymbol{V}) = \left(\boldsymbol{H} + \boldsymbol{V}^{-1}\right)^{-1}\boldsymbol{H}\boldsymbol{b} = \left(\boldsymbol{H} + \boldsymbol{V}^{-1}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}/\sigma^2$ by the vector of covariances **X'y** to obtain $\boldsymbol{y}'\boldsymbol{X}\widehat{\boldsymbol{\beta}}(\boldsymbol{V}) = \boldsymbol{y}'\boldsymbol{X}\left(\boldsymbol{H} + \boldsymbol{V}^{-1}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}/\sigma^2$ which must be positive because $(\mathbf{H}+\mathbf{V}^{-1})^{-1}$ is positive definite. At least one of the elements of this inner product $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$ must be positive for the sum to be positive which means at least one of the elements is the product of two numbers with the same sign.

We next show that the feasible ellipsoid covers every orthant except the orthant opposite the simple regressions **X'y**. Consider the estimate $\hat{\boldsymbol{\beta}} = \lambda\boldsymbol{\Lambda}\mathbf{s}$ where **s** is a vector of ones and minus ones selecting an orthant, and $\boldsymbol{\Lambda}$ is a positive diagonal matrix selecting the direction within the orthant and λ is a scalar that is used to shrink the size of vector. This vector is feasible if

$$\left(\lambda\boldsymbol{\Lambda}\mathbf{s} - \frac{\mathbf{b}}{2}\right)'\mathbf{H}\left(\lambda\boldsymbol{\Lambda}\mathbf{s} - \frac{\mathbf{b}}{2}\right) < \mathbf{b}'\mathbf{H}\mathbf{b}/4$$
$$\lambda^2\mathbf{s}'\boldsymbol{\Lambda}\mathbf{H}\boldsymbol{\Lambda}\mathbf{s} - \lambda\mathbf{s}'\boldsymbol{\Lambda}\mathbf{H}\mathbf{b} + \mathbf{b}'\mathbf{H}\mathbf{b}/4 < \mathbf{b}'\mathbf{H}\mathbf{b}/4$$
$$\lambda\mathbf{s}'\boldsymbol{\Lambda}\mathbf{H}\boldsymbol{\Lambda}\mathbf{s} < \mathbf{s}'\boldsymbol{\Lambda}\mathbf{H}\mathbf{b} = \mathbf{s}'\boldsymbol{\Lambda}\mathbf{X}'\mathbf{y}/\sigma^2$$

The expression on the left side of this inequality is a positive number that converges to zero as λ converges to zero. Thus this inequality is satisfied for some λ if the item on the right is positive. If **s** and **X'y** have at least one element with a common sign, then the positive diagonal matrix $\boldsymbol{\Lambda}$ can have the corresponding element with a large enough value that offsets whatever negatives might come from the other elements. Thus by choice of $\boldsymbol{\Lambda}$ and λ, this inequality can be satisfied for all **s** but the one with signs all opposite **X'y**.

The choice of the diagonal positive matrix $\boldsymbol{\Lambda}$ doesn't matter if **s** and **X'y** conform in sign, but otherwise has to be chosen to emphasize the common signs of **s** and **X'y.** Thus the corollary:

Corollary 3.1: The relative magnitudes of the elements of $\widehat{\boldsymbol{\beta}}(\boldsymbol{V}) = \left(\boldsymbol{H} + \boldsymbol{V}^{-1}\right)^{-1}\boldsymbol{H}\boldsymbol{b}$ are unlimited in the orthant of the simple correlations, X'y, but otherwise the relative magnitudes $\boldsymbol{\Lambda}$ are limited by the inequality $\mathbf{s}'\boldsymbol{\Lambda}\mathbf{X}'\mathbf{y} = \mathbf{s}'\boldsymbol{\Lambda}\mathbf{r} = \sum s_i\Lambda_i r_i > 0$.

Note that this condition depends on the simple correlations, **X'y**, not the partial correlations (regression coefficients), $(\mathbf{X}'\mathbf{X})^{-1}$ **X'y,** which is yet another reason for interest in these simple correlations:

Corollary 3.2: The range of the relative magnitudes of the elements of $\widehat{\boldsymbol{\beta}}(\boldsymbol{V}) = \left(\boldsymbol{H} + \boldsymbol{V}^{-1}\right)^{-1}\boldsymbol{H}\boldsymbol{b}$ depends on the simple correlations r=X'y = Nb, not on the regression estimates b.

Corollary 3.3: The OLS multiple regression and the one-at-a-time simple regressions must have at least one sign in common.

This follows from the fact that the inner product of the OLS estimate and the covariance vector must be positive: **b'X'y** = **y'X**(**X'X**) $^{-1}$ **X'y** >0

11.1 Extreme Bounds Analysis

That concludes the discussion of the orthants (sign patterns). Next we can find the intervals of estimates of linear combinations of parameters $\boldsymbol{\psi}'\widehat{\boldsymbol{\beta}}$ over the feasible ellipsoid. Since we are trying to determine if a coefficient or linear combination can be bounded away from zero, it is discouraging that all but one special linear combination has an interval of estimates that includes both positive and negative values.

First we report a result on the extreme values of linear combinations constrained to the feasible ellipsoid

Theorem: The linear combination $\psi'\hat{\beta}$ with $\hat{\beta}$ constrained to the feasible ellipsoid $\left(\widehat{\boldsymbol{\beta}}-\frac{\boldsymbol{b}}{2}\right)\boldsymbol{H}\left(\widehat{\boldsymbol{\beta}}-\frac{\boldsymbol{b}}{2}\right)\leq\frac{\boldsymbol{b}'\boldsymbol{H}\boldsymbol{b}}{\boldsymbol{4}}$ must lie in the interval

$$\frac{\boldsymbol{\psi}'\boldsymbol{b}}{2}-\left(\frac{(\boldsymbol{\psi}'\boldsymbol{H}^{-1}\boldsymbol{\psi})(\boldsymbol{b}'\boldsymbol{H}\boldsymbol{b})}{4}\right)^{\frac{1}{2}}\leq\boldsymbol{\psi}'\widehat{\boldsymbol{\beta}}\leq\frac{\boldsymbol{\psi}'\boldsymbol{b}}{2}+\left(\frac{(\boldsymbol{\psi}'\boldsymbol{H}^{-1}\boldsymbol{\psi})(\boldsymbol{b}'\boldsymbol{H}\boldsymbol{b})}{4}\right)^{\frac{1}{2}}$$

Proof: The problem of finding the extremes of ψ'x subject to ellipsoidal constraint: $\left(x-\frac{c}{2}\right)'A\left(x-\frac{c}{2}\right)\leq\frac{c'Ac}{4}$ leads to the Lagrangian derivative: $\psi-\lambda A\left(x-\frac{c}{2}\right)=0$. From this vector of equations we can solve for $\left(x-\frac{c}{2}\right)=A^{-1}\psi/\lambda$, and can insert this into the ellipsoidal constraint $\left(x-\frac{c}{2}\right)'H\left(x-\frac{c}{2}\right)=\frac{\psi'A^{-1}\psi}{\lambda^2}=\frac{c'Ac}{4}$ which allows us to solve for the Lagrange multiplier, $\lambda^2=\frac{\psi'A^{-1}\psi}{\frac{c'Ac}{4}}$. Thus the solution is

$$x=\frac{c}{2}\pm A^{-1}\psi\left(\frac{c'Ac}{4\psi'A^{-1}\psi}\right)^{1/2}$$

$$\psi'x=\frac{\psi'c}{2}\pm\psi'A^{-1}\psi\left(\frac{c'Ac}{4\psi'^{H^{-1}}\psi}\right)^{\frac{1}{2}}=\frac{\psi'c}{2}\pm\left(\frac{\psi'A^{-1}\psi c'Ac}{4}\right)^{\frac{1}{2}}$$

An s-value (sturdiness measure) is equal to the center of the ambiguity interval divided by half the length. The s-value corresponding to the feasible ellipsoid per the theorem above is equal to the z-statistic $\boldsymbol{\psi}'\mathbf{b}/(\boldsymbol{\psi}'\boldsymbol{H}^{-1}\boldsymbol{\psi})^{\mathbf{1/2}}$ for testing $\psi'\beta=0$, divided by the square root of the chi-squared statistic $(\boldsymbol{b}'\boldsymbol{H}\boldsymbol{b})$ for testing the multivariate hypothesis that $\boldsymbol{\beta}$=**0**.

s-value = $\frac{\boldsymbol{\psi}'\boldsymbol{b}}{(\boldsymbol{\psi}'\boldsymbol{H}^{-1}\boldsymbol{\psi})^{1/2}(\boldsymbol{b}'\boldsymbol{H}\boldsymbol{b})^{1/2}}=\frac{z_\psi}{\left(\chi^2_\beta\right)^{1/2}}$

This is confirmation that sampling uncertainty and model ambiguity can be two sides of the same coin. To determine if an estimate is adequately certain, compare the z-value with 1.96. To determine if an estimate is sturdy, compare the z-value with the square root of the chi-squared for testing $\boldsymbol{\beta}$=**0**.

This has a geometric interpretation. A confidence interval for β_1 in Figure 9 can be found by projecting a suitably chosen likelihood ellipse onto the x-axis. The ambiguity interval can be found by projecting the feasible ellipse onto the x-axis. Because the feasible ellipse and the likelihood ellipse have exactly the same shape, projections in all directions behave similarly. The shaded ellipse in Figure 3 describes the set of estimates that are attainable when the prior variance matrix is bounded from above and from below. This ellipse of estimates does not have same shape as the likelihood ellipse which raises the possibility that at least sometimes sampling uncertainty and model ambiguity; sometimes t-values and s-values are not identically ordered.

Next we can confirm that the s-value is less than one for all but one linear combination of coefficients, in other words, almost all model ambiguity intervals overlap zero.

Corollary: If the linear combination is proportional to the vector of covariances, $\boldsymbol{\psi} = \boldsymbol{Hb} \propto \boldsymbol{X'y}$ then the interval of feasible estimates extends from zero to the OLS estimate $\boldsymbol{\psi'b}$.

Proof: For $\psi = Hb$ the interval of estimates $\frac{\boldsymbol{\psi'b}}{2} \pm \left(\frac{(\boldsymbol{\psi'H^{-1}\psi})(\boldsymbol{b'Hb})}{4}\right)^{\frac{1}{2}}$ is $\frac{\boldsymbol{b'Hb}}{2} \pm \left(\frac{(\boldsymbol{b'Hb})(\boldsymbol{b'Hb})}{4}\right)^{\frac{1}{2}}$ $= \frac{\boldsymbol{b'Hb}}{2} \pm \frac{\boldsymbol{b'Hb}}{2} = \frac{\boldsymbol{\psi'b}}{2} \pm \frac{\boldsymbol{\psi b}}{2}$.

Corollary: If the linear combination $\boldsymbol{\psi}$ is anything other than proportional to the vector of covariances, then the interval of feasible estimates overlaps zero.

Proof: We need to show that center of the interval of estimates is less that the amount that is added and subtracted, in other words, $(\boldsymbol{\psi'b})^2 < (\boldsymbol{\psi'H^{-1}\psi})(\boldsymbol{b'Hb})$**,** or equivalently, $(\boldsymbol{\psi'b})^2/(\boldsymbol{\psi'H^{-1}\psi}) < (\boldsymbol{b'Hb})$**.** It is enough to find the maximum of the expression on the left (the χ^2-statistic for testing $\boldsymbol{\psi'\beta} = \boldsymbol{0}$) and show that it is less than the expression on the right (the χ^2-statistic for testing $\boldsymbol{\beta} = \boldsymbol{0}$). The maximum is found by setting the vector of derivatives to zero:

$$\boldsymbol{0} = \partial ln\left((\boldsymbol{\psi'b})^2/(\boldsymbol{\psi'H^{-1}\psi})\right)/\partial\boldsymbol{\psi} = \frac{2\boldsymbol{bb'\psi}}{(\boldsymbol{\psi'b})^2} - \frac{\boldsymbol{2H^{-1}\psi}}{\boldsymbol{\psi'H^{-1}\psi}}$$

Confirm that this vector of derivatives is zero when $\boldsymbol{\psi} = \boldsymbol{Hb}$, and the χ^2-statistic in that case becomes $(\boldsymbol{\psi'b})^2/(\boldsymbol{\psi'H^{-1}\psi}) = (\boldsymbol{b'Hb})^2/(\boldsymbol{b'Hb}) = (\boldsymbol{b'Hb})$.

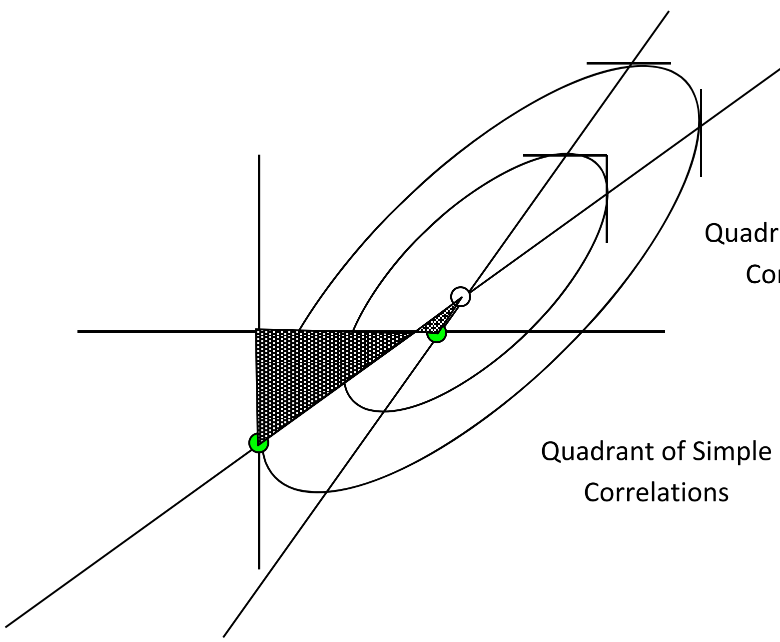
12 Appendix: Some Implications and non-implications of the diagonal case

As just discussed, the vector of signs of simple correlations take on special meaning since it is impossible to devise a shrinkage estimator with all signs opposite these, but any other sign pattern is attainable with a suitably chosen shrinkage "metric." If the prior precision matrix is restricted to be diagonal the vector of simple correlations is even more important. Here is one reason:

Theorem: If the prior covariance matrix is diagonal, then the compromise estimates closest to zero are in the orthant of the simple correlations.

This is illustrated in Figure 10. Rather than proving this one, I prove a more general result next.

Figure 10 Compromise Estimates with Diagonal Prior Precision



Theorem: If the prior precision matrix is diagonal, then estimates may not lie in opposite orthants.

Proof: Suppose, to the contrary, that $\hat{\boldsymbol{\beta}}_1$ and the orthant opposite $\hat{\boldsymbol{\beta}}_1$ are both feasible. Then there exists positive diagonal matrices such that:

$$\mathbf{D}_1\hat{\boldsymbol{\beta}}_1 = \mathbf{N}\left(\mathbf{b}-\hat{\boldsymbol{\beta}}_1\right), \mathbf{D}_1 = diag\{d_{11}, d_{12},..., d_{1k}\} > 0$$
$$\mathbf{D}_2\hat{\boldsymbol{\beta}}_2 = \mathbf{N}\left(\mathbf{b}-\hat{\boldsymbol{\beta}}_2\right), \mathbf{D}_2 = diag\{d_{21}, d_{22},..., d_{2k}\} > 0$$
$$\hat{\boldsymbol{\beta}}_2 = -\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}_1, \boldsymbol{\Lambda} = diag\{\lambda_1, \lambda_2,...\lambda_k\} > 0$$

Substitute the third line into the second to obtain the two equations:

$$\mathbf{D}_1\hat{\boldsymbol{\beta}}_1 = \mathbf{Nb} - \mathbf{N}\hat{\boldsymbol{\beta}}_1$$
$$-\mathbf{D}_2\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}_1 = \mathbf{Nb} + \mathbf{N}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}_1$$

Subtract the second from the first:

$$(\mathbf{D}_1 + \mathbf{D}_2\boldsymbol{\Lambda})\hat{\boldsymbol{\beta}}_1 = -\mathbf{N}(\mathbf{I} + \boldsymbol{\Lambda})\hat{\boldsymbol{\beta}}_1$$

Then premultiply by $(\mathbf{I} + \boldsymbol{\Lambda})\hat{\boldsymbol{\beta}}_1$ to obtain:

$$\hat{\boldsymbol{\beta}}_1'(\mathbf{I} + \boldsymbol{\Lambda})(\mathbf{D}_1 + \mathbf{D}_2\boldsymbol{\Lambda})\hat{\boldsymbol{\beta}}_1 = -\hat{\boldsymbol{\beta}}_1'(\mathbf{I} + \boldsymbol{\Lambda})\mathbf{N}(\mathbf{I} + \boldsymbol{\Lambda})\hat{\boldsymbol{\beta}}_1$$

which is a contradiction because the expression to the left of the equal sign is positive because it is the weighted sum of squares of the elements of $\hat{\boldsymbol{\beta}}_1$ while the expression to the right is negative because the quadratic form involves a positive definite matrix.

Corollary: If the prior precision matrix is diagonal, then the orthant opposite the OLS vector b is not possible.

Corollary: There can be at most $2^k/2$ orthants of estimates.

Corollary: In the two dimensional case there can be at most two orthants of estimates.

False k-orthant generalization: There can be at most k different orthants of estimates.

Counterexample available on request. (I had high hopes this would be true.)

13 Tables

Table 1 Sala-i-Matin et. al. (2004) variables

Table 1 - Data Description and Sources: SDM(2005)

Rank	Variable	Description and Source
	Average growth rate of GDP per capita 1960-1996	Growth of GDP per capita at purchasing power parities between 1960 and 1996. From Alan Heston et al. (2001).
1	East Asian dummy	Dummy for East Asian countries.
2	Primary schooling in 1960	Enrollment rate in primary education in 1960. Barro and Jong-Wha Lee (1993).
3	Investment price	Average investment price level between 1960 and 1964 on purchasing power parity basis. From Heston et al. (2001).
4	GDP in 1960 (log)	Logarithm of GDP per capita in 1960. From Heston et al. (2001).
5	Fraction of tropical area	Proportion of country's land area within geographical tropics. From John L. Gallup et al. (2001).
6	Population density coastal in 1960's	Coastal (within 100 km of coastline) population per coastal area in 1965. From Gallup et al. (2001).

7	Malaria prevalence in 1960's	Index of malaria prevalence in 1966. From Gallup et al. (2001).
8	Life expectancy in 1960	Life expectancy in 1960. Barro and Lee (1993).
9	Fraction Confucian	Fraction of population Confucian. Barro (1999).
10	African dummy	Dummy for Sub-Saharan African countries.
11	Latin American dummy	Dummy for Latin American countries.
12	Fraction GDP in mining	Fraction of GDP in mining. From Robert E. Hall and Charles I. Jones (1999).
13	Spanish colony	Dummy variable for former Spanish colonies. Barro (1999).
14	Years open 1950-1994	Number of years economy has been open between 1950 and 1994. From Jeffrey D. Sachs and Andrew M. Warner (1995).
15	Fraction Muslim	Fraction of population Muslim in 1960. Barro (1999).
16	Fraction Buddhist	Fraction of population Buddhist in 1960. Barro (1999).
17	Ethnolinguistic fractionalization	Average of five different indices of ethnolinguistic fractionalization which is the probability of two random people in a country not speaking the same language. From William Easterly and Ross Levine (1997).
18	Government consumption share 1960's	Share of expenditures on government consumption to GDP in1961. Barro and Lee (1993).
19	Population density 1960	Population per area in 1960. Barro and Lee (1993).
20	Real exchange rate distortions	Real exchange rate distortions. Levine and Renelt (1992).
21	Fraction speaking foreign language	Fraction of population speaking foreign language. Hall and Jones(1999).
22	Openness measure 1965-1974	Ratio of exports plus imports to GDP, averaged over 1965 to 1974. This variable was provided by Robert Barro.
23	Political rights	Political rights index. From Barro (1999).
24	Government share of GDP in 1960's	Average share government spending to GDP between 1960-1964. From Heston et al. (2001).
25	Higher education in 1960	Enrollment rates in higher education. Barro and Lee (1993).
26	Fraction population in tropics	Proportion of country's population living in geographical tropics. From Gallup et al. (2001).
27	Primary exports 1970	Fraction of primary exports in total exports in 1970. From Sachs and Warner (1997).
28	Public investment share	Average share of expenditures on public investment as fraction of GDP between 1960 and 1965. Barro and Lee (1993).
29	Fraction Protestant	Fraction of population Protestant in 1960. Barro (1999).
30	Fraction Hindu	Fraction of the population Hindu in 1960. Barro (1999).
31	Fraction population less than 15	Fraction of population younger than 15 years in 1960. Barro and Lee (1993).
32	Air distance to big cities	Logarithm of minimal distance (in km) from New York, Rotterdam, or Tokyo. From Gallup et al. (2001).
33	Nominal government GDP share 1960's	Average share of nominal government spending to nominal GDP between 1960 and 1964. Calculated from Heston et al. (2001).
34	Absolute latitude	Absolute latitude. Barro (1999).
35	Fraction Catholic	Fraction of population Catholic in 1960. Barro (1999).
36	Fertility in 1960's	Fertility in 1960's. Barro and Sala-i-Martin (1995).
37	European dummy	Dummy for European economies.
38	Outward orientation	Measure of outward orientation. Levine and Renelt (1992).
39	Colony dummy	Dummy for former colony. Barro (1999).
40	Civil liberties	Index of civil liberties index in 1972. Barro (1999).
41	Revolutions and coups	Number of revolutions and military coups. Barro and Lee (1993).
42	British colony	Dummy for former British colony after 1776. Barro (1999).
43	Hydrocarbond deposits in 1993	Log of hydrocarbond eposits in 1993. From Gallup et al. (2001).
44	Fraction population over 65	Fraction of population older than 65 years in 1960. Barro and Lee(1993)

45	Defense spending share	Average share public expenditures on defense as fraction of GDP between 1960 and 1965. Barro and Lee (1993).
46	Population in 1960	Population in 1960. Barro (1999).
47	Terms of trade growth in 1960's	Growth of terms of trade in the 1960's. Barro and Lee (1993).
48	Public education spending share in GDP in 1960's	Average share public expenditures on education as fraction of GDP between 1960 and 1965. Barro and Lee (1993).
49	Landlocked country dummy	Dummy for landlocked countries.
50	Religious intensity	Religion measure. Barro (1999).
51	Size of economy	Logarithm of aggregate GDP in 1960.
52	Socialist dummy	Dummy for countries under Socialist rule for considerable time during 1950 to 1995. From Gallup et al. (2001).
53	English-speaking population	Fraction of population speaking English. From Hall and Jones (1999).
54	Average inflation 1960-1990	Average inflation rate between 1960 and 1990. Levine and Renelt(1992).
55	Oil-producing country dummy	Dummy for oil-producing country. Barro (1999).
56	Population growth rate 1960-1990	Average growth rate of population between 1960 and 1990. Barro and Lee (1993).
57	Timing of independence	Timing of national independence measure: 0 if before 1914; 1 if between 1914 and 1945; 2 if between 1946 and 1989; and 3 if after 1989. From Gallup et al. (2001).
58	Fraction of land area near navigable water	Proportion of country's land area within 100 km of ocean or ocean-navigable river. From Gallup et al. (2001).
59	Square of inflation 1960-1990	Square of average inflation rate between 1960 and 1990.
60	Fraction spent in war 1960-1990	Fraction of time spent in war between 1960 and 1990. Barro and Lee (1993).
61	Land area	Area in km2. Barro and Lee (1993).
62	Tropical climate zone	Fraction tropical climate zone. From Gallup et al. (2001).
63	Terms of trade ranking	Barro (1999).
64	Capitalism	Degree Capitalism index. From Hall and Jones (1999).
65	Fraction Orthodox	Fraction of population Orthodox in 1960. Barro (1999).
66	War participation 1960-1990	Indicator for countries that participated in external war between 1960 and 1990. Barro and Lee (1993).
67	Interior density	Interior (more than 100 km from coastline) population per interior area in 1965. From Gallup et al. (2001).

Table 2 Proposed Reporting Style for a 14 variable Regression

Models Explaining Growth Rates of Real per capita Income from 1960 to 1966: 87 Countries

Data Source: Sala-i-Martin, Doppelhofer and Miller (2004)

Standardized Variables (Unit variance and zero mean)

All Variables Treated the Same

Sorted First by Category and then by OLS t-value

		Prior Expected R-sq	"0"	0.1	0.5	1	Infinite	"0"	0.1	0.5	1	Infinite	(0.1, 1.0)	(0.1, 0.5)	(0.5, 1.0)
	Category	Description	b-SIMPLE	b-BAYES	b-BAYES	b-BAYES	b_OLS	t-SIMPLE	t-BAYES	t-BAYES	t-BAYES	t-OLS	s-val	s-val	s-val
1	catchup	GDP in 1960 (log)	**0.318**	-0.023	-0.195	-0.283	-0.462	3.09	-0.32	-1.60	-1.99	-2.61	-0.48	-0.47	-2.72
2	culture	Fraction Speaking Foreign Language	0.258	0.092	0.127	0.133	0.145	2.46	1.52	1.54	1.51	1.50	0.67	0.72	3.36
3	geography	Absolute Latitude	0.394	0.105	0.209	0.249	0.314	3.95	1.53	1.90	2.02	2.18	0.67	0.69	3.36
4	geography	Fraction of Land Area Near Navigable Water	0.404	0.113	0.145	0.149	0.155	4.07	1.83	1.70	1.64	1.57	0.76	0.82	3.79
5	Government	British Colony Dummy	0.076	0.066	0.104	0.108	0.107	0.70	1.11	1.30	1.27	1.17	-0.42	-0.45	-2.34
6	Government	Government GDP Share 1960s	-0.073	-0.048	-0.073	-0.076	-0.079	-0.67	-0.81	-0.94	-0.93	-0.90	-0.24	-0.25	-1.22
7	Government	Fraction Spent in War 1960-90	-0.135	-0.029	-0.038	-0.041	-0.050	-1.25	-0.50	-0.48	-0.50	-0.57	0.56	0.59	3.06
8	Government	Outward Orientation	**0.030**	0.009	-0.008	-0.017	-0.032	0.28	0.16	-0.10	-0.20	-0.36	-0.03	-0.02	-0.38
9	resources	Primary Schooling in 1960	0.574	0.201	0.365	0.424	0.525	6.47	3.03	3.61	3.81	4.14	1.34	1.40	6.89
10	resources	Primary Exports 1970	-0.491	-0.157	-0.271	-0.302	-0.342	-5.19	-2.34	-2.63	-2.65	-2.61	0.17	0.17	0.82
11	resources	Oil Producing Country Dummy	**-0.019**	0.017	0.055	0.072	0.102	-0.18	0.28	0.70	0.86	1.14	0.16	0.18	0.66
12	resources	Higher Education 1960	0.308	0.034	0.035	0.045	0.082	2.98	0.51	0.34	0.40	0.62	-0.17	-0.17	-0.83
13	resources	Fraction Population Less than 15	**-0.228**	0.024	0.082	0.099	0.122	-2.16	0.34	0.63	0.62	0.53	-0.96	-1.00	-4.81
14	resources	Fraction Population Over 65	**0.234**	-0.025	-0.085	-0.098	-0.102	2.22	-0.35	-0.65	-0.62	-0.44	0.29	0.30	1.94
		R-squared		0.210	0.378	0.436	0.537								
		Mean Squared Coefficient		0.008	0.026	0.036	0.058								

Table 3 Regression of Initial Per Capita GDP on 13 Explanatory Variables

Dependent Variable: Log of Real GDP per Capita, 1960
Standardized Variables
Included observations: 87

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Fraction Speaking Foreign Language	0.151	0.062	2.45	0.02
Absolute Latitude	0.068	0.095	0.71	0.48
Fraction of Land Area Near Navigable Water	0.081	0.064	1.25	0.21
British Colony Dummy	-0.008	0.061	-0.14	0.89
Government GDP Share 1960s	-0.028	0.058	-0.48	0.63
Fraction Spent in War 1960-90	-0.075	0.058	-1.29	0.20
Outward Orientation	-0.021	0.059	-0.36	0.72
Primary Schooling in 1960	0.234	0.079	2.95	0.00
Primary Exports 1970	0.042	0.087	0.48	0.63
Oil Producing Country Dummy	0.071	0.058	1.21	0.23
Higher Education 1960	0.345	0.078	4.44	0.00
Fraction Population Less than 15	-0.126	0.151	-0.84	0.41
Fraction Population Over 65	0.241	0.150	1.61	0.11
Constant	0.000	0.053	0.00	1

R-squared	0.795	Mean dependent var	2.04E-17
Adjusted R-squared	0.759	S.D. dependent var	1.01
S.E. of regression	0.494	Akaike info criterion	1.57
Sum squared resid	17.8	Schwarz criterion	1.97
Log likelihood	-54.5	Hannan-Quinn criter.	1.73
F-statistic	21.8	Durbin-Watson stat	2.03
Prob(F-statistic)	0		

Table 4 Extreme TFP estimates

Top Ten and Bottom Ten of Residuals from Regression in Table 3

	Actual	Actual Rank	Residual	Residual Rank
Trinidad & Tobago	1.47	11	1.65	1
Venezuela	1.60	8	1.10	2
South Africa	0.40	26	0.81	3
Papua New Guinea	-0.25	46	0.71	4
Mexico	0.69	24	0.69	5
Finland	1.40	14	0.64	6
Guatemala	0.09	36	0.64	7
Australia	1.84	2	0.60	8
Zambia	-0.53	59	0.55	9
Senegal	-0.43	56	0.50	10
Taiwan	-0.23	45	-0.46	78
Tanzania	-1.78	85	-0.60	79
Korea	-0.60	62	-0.66	80
Togo	-1.62	84	-0.69	81
Malawi	-1.58	83	-0.73	82
Egypt	-0.72	65	-0.75	83
Philippines	-0.34	52	-0.78	84
Portugal	0.22	30	-0.81	85
Indonesia	-0.99	73	-0.88	86
Lesotho	-1.80	86	-1.41	87

Table 5 Regression on 67 Variables, No Favorites

Models Explaining Growth Rates of Real per capita Income from 1960 to 1966: 87 Countries

Data Source: Sala-i-Martin, Doppelhofer and Miller (2004)

Standardized Variables (Unit variance and zero mean)

All Variables Treated the Same

Sorted First by Category and then by OLS t-value

		Prior Expected R-sq	"0"	0.1	0.5	1	Infinite	"0"	0.1	0.5	1	Infinite	(0.1, 1.0)	(0.1, 0.5)	(0.5, 1.0)
	Category	Description	b-SIMPLE	b-BAYES	b-BAYES	b-BAYES	b_OLS	t-SIMPLE	t-BAYES	t-BAYES	t-BAYES	t-OLS	s-val	s-val	s-val
1	catchup	GDP in 1960 (log)	**0.318**	-0.01	-0.01	-0.121	-0.329	3.09	-0.28	-0.28	-1.18	-0.78	-0.17	-0.17	-0.77
2	culture	Fraction Confucian	0.474	0.068	0.068	0.131	0.276	4.96	1.99	1.99	1.71	1.33	0.37	0.38	1.48
3	culture	Fraction Buddhist	0.445	0.063	0.063	0.111	0.337	4.58	1.82	1.82	1.41	1.13	0.31	0.32	1.21
4	culture	Fraction Muslim	0.012	0.012	0.012	0.032	0.253	0.11	0.35	0.35	0.38	0.97	0.07	0.07	0.28
5	culture	Fraction Speaking Foreign Language	0.258	0.029	0.029	0.074	0.2	2.46	0.84	0.84	0.88	0.95	0.17	0.17	0.67
6	culture	English Speaking Population	0.065	-0.014	-0.014	-0.011	0.146	0.60	-0.40	-0.40	-0.14	0.87	-0.04	-0.05	-0.18
7	culture	Fraction Protestants	0.086	-0.024	-0.024	-0.056	0.348	0.80	-0.67	-0.67	-0.64	0.70	-0.13	-0.13	-0.53
8	culture	European Dummy	0.256	0	0	0.03	0.412	2.44	0.00	0.00	0.28	0.51	0.04	0.04	0.15
9	culture	Fertility in 1960s	-0.435	-0.018	-0.018	-0.028	-0.259	-4.46	-0.49	-0.49	-0.26	-0.43	-0.06	-0.06	-0.18
10	culture	Spanish Colony	-0.109	-0.02	-0.02	-0.048	0.159	-1.01	-0.56	-0.56	-0.52	0.42	-0.10	-0.10	-0.40
11	culture	Ethnolinguistic Fractionalization	**-0.431**	-0.033	-0.033	-0.054	0.073	-4.40	-0.91	-0.91	-0.61	0.32	-0.14	-0.14	-0.51
12	culture	Fraction Catholic	0.051	-0.009	-0.009	-0.031	0.195	0.47	-0.26	-0.26	-0.31	0.31	-0.06	-0.06	-0.21
13	culture	Fraction Othodox	0.069	0.011	0.011	0.024	0.036	0.64	0.34	0.34	0.35	0.30	0.08	0.08	0.34
14	culture	Religion Measure	**-0.202**	-0.019	-0.019	0.003	0.018	-1.90	-0.56	-0.56	0.04	0.12	-0.03	-0.04	-0.03
15	culture	Fraction Hindus	0.061	0.013	0.013	0.01	0.026	0.57	0.37	0.37	0.13	0.12	0.04	0.04	0.14

16	geography	Absolute Latitude	**0.394**	0.018	0.018	0.009	-0.629	3.95	0.50	0.50	0.09	-1.18	0.04	0.04	0.10
17	geography	Fraction of Tropical Area	-0.429	-0.034	-0.034	-0.071	-0.517	-4.38	-0.94	-0.94	-0.70	-1.06	-0.14	-0.14	-0.51
18	geography	Fraction GDP in Mining	**-0.051**	0.023	0.023	0.092	0.191	-0.47	0.65	0.65	1.16	0.97	0.20	0.20	0.92
19	geography	African Dummy	**-0.571**	-0.05	-0.05	-0.101	-0.362	-6.42	-1.38	-1.38	-1.00	-0.89	-0.20	-0.21	-0.73
20	geography	Latin American Dummy	-0.115	-0.022	-0.022	-0.051	-0.541	-1.07	-0.62	-0.62	-0.51	-0.72	-0.10	-0.10	-0.37
21	geography	Fraction Population In Tropics	-0.575	-0.044	-0.044	-0.051	-0.191	-6.47	-1.22	-1.22	-0.52	-0.52	-0.13	-0.14	-0.44
22	geography	Tropical Climate Zone	-0.242	-0.019	-0.019	-0.037	-0.085	-2.30	-0.55	-0.55	-0.46	-0.49	-0.10	-0.10	-0.39
23	geography	Land Area	-0.011	-0.008	-0.008	-0.018	-0.158	-0.10	-0.24	-0.24	-0.21	-0.40	-0.04	-0.04	-0.16
24	geography	East Asian Dummy	0.53	0.079	0.079	0.149	0.128	5.77	2.26	2.26	1.68	0.39	0.35	0.36	1.33
25	geography	Fraction of Land Area Near Navigable Water	**0.404**	0.019	0.019	-0.005	-0.11	4.07	0.53	0.53	-0.05	-0.33	0.02	0.03	0.00
26	geography	Landlocked Country Dummy	-0.234	-0.009	-0.009	-0.019	0.039	-2.22	-0.26	-0.26	-0.25	0.21	-0.05	-0.05	-0.18
27	geography	Population Density Coastal in 1960s	0.427	0.048	0.048	0.04	0.067	4.36	1.37	1.37	0.47	0.19	0.14	0.15	0.45
28	geography	Air Distance to Big Cities	**-0.345**	-0.008	-0.008	0.01	-0.051	-3.38	-0.22	-0.22	0.11	-0.18	0.00	0.00	0.07
29	Government	Public Investment Share	0.05	-0.001	-0.001	-0.082	-0.254	0.47	-0.02	-0.02	-1.13	-1.77	-0.16	-0.16	-0.85
30	Government	Revolutions and Coups	-0.215	-0.019	-0.019	-0.068	-0.248	-2.03	-0.54	-0.54	-0.88	-1.29	-0.16	-0.16	-0.70
31	Government	Capitalism	**0.255**	0.016	0.016	0.002	-0.171	2.43	0.47	0.47	0.02	-1.00	0.03	0.04	0.08
32	Government	Defense Spending Share	**0.277**	0.024	0.024	0.019	1.152	2.66	0.68	0.68	0.23	0.99	0.07	0.08	0.24
33	Government	Public Education Spending Share in GDP in 1960s	0.242	0.018	0.018	0.053	0.47	2.30	0.53	0.53	0.68	0.99	0.13	0.13	0.55
34	Government	Government Share of GDP in 1960s	-0.34	-0.033	-0.033	-0.040	-3.401	-3.33	-0.93	-0.93	-0.43	-0.99	-0.10	-0.11	-0.35
35	Government	Gov. Consumption Share 1960s	-0.446	-0.041	-0.041	-0.047	3.379	-4.60	-1.15	-1.15	-0.48	0.95	-0.12	-0.13	-0.40
36	Government	Timing of Independence	**-0.166**	0.008	0.008	0.017	-0.194	-1.55	0.24	0.24	0.19	-0.60	0.04	0.04	0.18
37	Government	Square of Inflation 1960-90	**-0.06**	-0.002	-0.002	0.017	0.232	-0.56	-0.07	-0.07	0.18	0.57	0.02	0.02	0.11
38	Government	Years Open 1950-94	**0.619**	0.055	0.055	0.084	-0.188	7.26	1.53	1.53	0.91	-0.55	0.21	0.21	0.75
39	Government	Nominal Government GDP Share 1960s	**-0.073**	-0.018	-0.018	-0.072	-0.118	-0.67	-0.51	-0.51	-0.87	-0.52	-0.15	-0.15	-0.65
40	Government	Real Exchange Rate Distortions	**-0.515**	-0.055	-0.055	-0.093	0.099	-5.54	-1.56	-1.56	-1.16	0.52	-0.26	-0.27	-1.02
41	Government	Average Inflation 1960-90	**-0.096**	-0.005	-0.005	0.018	0.156	-0.89	-0.14	-0.14	0.19	0.40	0.02	0.02	0.11
42	Government	Political Rights	-0.408	-0.019	-0.019	-0.039	-0.122	-4.12	-0.53	-0.53	-0.40	-0.39	-0.08	-0.08	-0.30
43	Government	War Particpation 1960-90	-0.131	-0.011	-0.011	-0.027	-0.06	-1.22	-0.31	-0.31	-0.35	-0.38	-0.07	-0.07	-0.29

44	Government	Socialist Dummy	**-0.097**	0.005	0.005	0.032	0.013	-0.90	0.15	0.15	0.45	0.09	0.07	0.07	0.39
45	Government	Fraction Spent in War 1960-90	-0.135	-0.01	-0.01	0.004	-0.015	-1.25	-0.29	-0.29	0.05	-0.08	-0.01	-0.02	-0.01
46	Government	Colony Dummy	**-0.357**	-0.026	-0.026	-0.061	0.016	-3.52	-0.74	-0.74	-0.66	0.05	-0.13	-0.13	-0.49
47	Government	British Colony Dummy	0.076	0.02	0.02	0.061	0.009	0.70	0.59	0.59	0.76	0.04	0.14	0.14	0.60
48	Government	Civil Liberties	0.221	-0.015	-0.015	-0.085	-0.006	2.09	-0.43	-0.43	-0.98	-0.02	-0.16	-0.16	-0.73
49	Government	Openness measure 1965-74	0.265	0.036	0.036	0.062	0.007	2.53	1.03	1.03	0.69	0.02	0.15	0.15	0.54
50	Government	Outward Orientation	**0.03**	-0.003	-0.003	-0.035	-0.003	0.28	-0.09	-0.09	-0.49	-0.01	-0.08	-0.08	-0.41
51	resources	Investment Price	-0.444	-0.063	-0.063	-0.16	-0.383	-4.57	-1.83	-1.83	-2.23	-2.48	-0.44	-0.45	-1.97
52	resources	Primary Schooling in 1960	0.574	0.057	0.057	0.149	0.457	6.47	1.59	1.59	1.62	1.47	0.31	0.31	1.22
53	resources	Fraction Population Less than 15	**-0.228**	0.01	0.01	0.047	0.584	-2.16	0.27	0.27	0.45	0.96	0.07	0.07	0.31
54	resources	Population in 1960	0.14	0.021	0.021	0.07	0.214	1.30	0.60	0.60	0.89	0.96	0.16	0.16	0.70
55	resources	Size of Economy	**0.33**	0.009	0.009	-0.029	-0.393	3.22	0.25	0.25	-0.28	-0.72	-0.03	-0.02	-0.15
56	resources	Malaria Prevalence in 1960s	**-0.557**	-0.047	-0.047	-0.052	0.222	-6.18	-1.29	-1.29	-0.55	0.71	-0.14	-0.15	-0.49
57	resources	Higher Education 1960	**0.308**	0	0	-0.047	-0.156	2.98	0.01	0.01	-0.52	-0.66	-0.07	-0.07	-0.36
58	resources	Fraction Population Over 65	0.234	-0.004	-0.004	0.015	0.253	2.22	-0.12	-0.12	0.15	0.57	0.01	0.01	0.06
59	resources	Life Expectancy in 1960	**0.548**	0.038	0.038	0.086	-0.236	6.03	1.02	1.02	0.82	-0.52	0.16	0.16	0.58
60	resources	Primary Exports 1970	-0.491	-0.033	-0.033	-0.039	-0.125	-5.19	-0.91	-0.91	-0.41	-0.47	-0.10	-0.11	-0.35
61	resources	Hydrocarbon Deposits in 1993	0.07	0.006	0.006	0.04	0.063	0.65	0.18	0.18	0.53	0.35	0.09	0.09	0.42
62	resources	Population Growth Rate 1960-90	-0.371	-0.008	-0.008	0.017	-0.143	-3.68	-0.21	-0.21	0.17	-0.30	0.01	0.01	0.09
63	resources	Population Density 1960	**-0.115**	0.01	0.01	0.068	0.051	-1.07	0.29	0.29	0.88	0.26	0.14	0.14	0.69
64	resources	Oil Producing Country Dummy	**-0.019**	0.006	0.006	0.012	0.055	-0.18	0.17	0.17	0.15	0.25	0.03	0.03	0.13
65	resources	Terms of Trade Ranking	**0.066**	-0.015	-0.015	-0.015	-0.056	0.61	-0.43	-0.43	-0.17	-0.21	-0.05	-0.05	-0.15
66	resources	Terms of Trade Growth in 1960s	**0.013**	0.004	0.004	0.021	-0.044	0.12	0.11	0.11	0.26	-0.19	0.04	0.04	0.18
67	resources	Interior Density	0.121	-0.006	-0.006	-0.007	0.007	1.12	-0.17	-0.17	-0.10	0.05	-0.03	-0.03	-0.12
		R-squared		0.3531	0.3531	0.707	0.92								
		Mean Squared Coefficient		0.0008	0.0008	0.004	0.4194								

Table 6 Regression on 67 Variables, 14 Favorites

Models Explaining Growth Rates of Real per capita Income from 1960 to 1966: 87 Countries
Data Source: Sala-i-Martin, Doppelhofer and Miller (2004)
Standardized Variables (Unit variance and zero mean)
First 14 Favorite Variables
Sorted First by Category and then by OLS t-value

		Prior R-sq	0	(0.4, 0.1)	(0.8, 0.2)	Infinite	0	(0.4, 0.1)	(0.8, 0.2)	Infinite	
	Category	Description	b-SIMPLE	b-BAYES	b-BAYES	b_OLS	t-SIMPLE	t-BAYES	t-BAYES	t-OLS	s-value
1	catchup	GDP in 1960 (log)	0.318	-0.201	-0.274	-0.329	3.091	-1.773	-1.989	-0.78	-1.416
2	culture	Fraction Speaking Foreign Language	0.258	0.123	0.133	0.2	2.457	1.575	1.478	0.946	1.338
3	geography	Absolute Latitude	0.394	0.128	0.12	-0.629	3.948	1.213	0.933	-1.181	0.794
4	geography	Fraction of Land Area Near Navigable Water	0.404	0.048	0.021	-0.11	4.075	0.57	0.216	-0.334	0.304
5	Government	Nominal Govertnment GDP Share 1960s	-0.073	-0.083	-0.098	-0.118	-0.672	-1.107	-1.128	-0.524	-0.976
6	Government	Fraction Spent in War 1960-90	-0.135	-0.022	-0.014	-0.015	-1.254	-0.309	-0.182	-0.076	-0.249
7	Government	British Colony Dummy	0.076	0.073	0.071	0.009	0.702	0.983	0.838	0.043	0.839
8	Government	Outward Orientation	0.03	-0.041	-0.055	-0.003	0.278	-0.587	-0.732	-0.014	-0.74
9	resources	Primary Schooling in 1960	0.574	0.31	0.338	0.457	6.466	3.352	3.165	1.468	2.822
10	resources	Fraction Population Less than 15	-0.228	0.084	0.105	0.584	-2.156	0.704	0.697	0.963	0.478
11	resources	Higher Education 1960	0.308	-0.014	-0.028	-0.156	2.981	-0.15	-0.259	-0.658	-0.178
12	resources	Fraction Population Over 65	0.234	-0.015	0.028	0.253	2.223	-0.124	0.189	0.572	0.035
13	resources	Primary Exports 1970	-0.491	-0.155	-0.137	-0.125	-5.193	-1.633	-1.236	-0.472	-1.194
14	resources	Oil Producing Country Dummy	-0.019	0.04	0.044	0.055	-0.179	0.539	0.518	0.255	0.481
15	culture	Fraction Confucian	0.474	0.056	0.074	0.276	4.957	1.437	1.447	1.332	0.911
16	culture	Fraction Buddhist	0.445	0.059	0.075	0.337	4.578	1.552	1.506	1.126	0.972
17	culture	Fraction Muslim	0.012	0.021	0.031	0.253	0.108	0.53	0.581	0.974	0.34
18	culture	English Speaking Population	0.065	-0.008	-0.005	0.146	0.598	-0.199	-0.101	0.875	-0.09
19	culture	Fraction Protestants	0.086	-0.019	-0.025	0.348	0.799	-0.473	-0.467	0.704	-0.287
20	culture	European Dummy	0.256	0.004	0.009	0.412	2.442	0.087	0.154	0.509	0.073
21	culture	Fertility in 1960s	-0.435	-0.015	-0.023	-0.259	-4.457	-0.362	-0.382	-0.435	-0.214
22	culture	Spanish Colony	-0.109	-0.025	-0.032	0.159	-1.009	-0.629	-0.589	0.422	-0.364
23	culture	Ethnolinguistic Fractionalization	-0.431	-0.019	-0.026	0.073	-4.398	-0.482	-0.487	0.321	-0.294
24	culture	Fraction Catholic	0.051	-0.024	-0.034	0.195	0.467	-0.593	-0.611	0.312	-0.357
25	culture	Fraction Othodox	0.069	0.016	0.02	0.036	0.642	0.437	0.424	0.298	0.284
26	culture	Religion Measure	-0.202	-0.023	-0.025	0.018	-1.899	-0.61	-0.496	0.118	-0.346
27	culture	Fraction Hindus	0.061	0.012	0.015	0.026	0.565	0.324	0.287	0.117	0.191
28	geography	Fraction of Tropical Area	-0.429	-0.025	-0.033	-0.517	-4.378	-0.605	-0.573	-1.058	-0.339
29	geography	Fraction GDP in Mining	-0.051	0.03	0.048	0.191	-0.473	0.779	0.939	0.971	0.549
30	geography	African Dummy	-0.571	-0.038	-0.051	-0.362	-6.419	-0.922	-0.91	-0.885	-0.538
31	geography	Latin American Dummy	-0.115	-0.033	-0.043	-0.541	-1.072	-0.82	-0.775	-0.718	-0.466

32	geography	Fraction Population In Tropics	-0.575	-0.028	-0.033	-0.191	-6.473	-0.689	-0.587	-0.519	-0.371
33	geography	Tropical Climate Zone	-0.242	-0.002	-0.001	-0.085	-2.301	-0.04	-0.025	-0.495	-0.02
34	geography	Land Area	-0.011	-0.002	-0.002	-0.158	-0.099	-0.062	-0.04	-0.398	-0.03
35	geography	East Asian Dummy	0.53	0.079	0.104	0.128	5.768	1.991	1.952	0.387	1.194
36	geography	Landlocked Country Dummy	-0.234	-0.008	-0.015	0.039	-2.218	-0.209	-0.302	0.21	-0.165
37	geography	Population Density Coastal in 1960s	0.427	0.037	0.044	0.067	4.359	0.949	0.846	0.192	0.547
38	geography	Air Distance to Big Cities	-0.345	-0.007	-0.009	-0.051	-3.385	-0.163	-0.171	-0.181	-0.1
39	Government	Public Investment Share	0.05	-0.005	-0.018	-0.254	0.466	-0.125	-0.366	-1.77	-0.169
40	Government	Revolutions and Coups	-0.215	-0.015	-0.023	-0.248	-2.032	-0.379	-0.465	-1.289	-0.273
41	Government	Capitalism	0.255	0.013	0.014	-0.171	2.43	0.341	0.278	-1.001	0.196
42	Government	Defense Spending Share	0.277	0.017	0.018	1.152	2.661	0.428	0.348	0.992	0.234
43	Government	Public Education Spending Share in GDP in 1960s	0.242	0.013	0.02	0.47	2.302	0.341	0.392	0.985	0.234
44	Government	Government Share of GDP in 1960s	-0.34	-0.019	-0.021	-3.401	-3.335	-0.477	-0.383	-0.993	-0.254
45	Government	Gov. Consumption Share 1960s	-0.446	-0.025	-0.027	3.379	-4.599	-0.617	-0.496	0.946	-0.327
46	Government	Timing of Independence	-0.166	0.011	0.013	-0.194	-1.549	0.271	0.246	-0.6	0.155
47	Government	Square of Inflation 1960-90	-0.06	-0.002	0.004	0.232	-0.558	-0.042	0.074	0.568	0.015
48	Government	Years Open 1950-94	0.619	0.049	0.061	-0.188	7.257	1.222	1.121	-0.554	0.7
49	Government	Real Exchange Rate Distortions	-0.515	-0.051	-0.067	0.099	-5.541	-1.306	-1.285	0.516	-0.806
50	Government	Average Inflation 1960-90	-0.096	-0.008	-0.005	0.156	-0.892	-0.194	-0.09	0.401	-0.082
51	Government	Political Rights	-0.408	-0.013	-0.021	-0.122	-4.123	-0.31	-0.372	-0.391	-0.205
52	Government	War Particpation 1960-90	-0.131	-0.015	-0.023	-0.06	-1.221	-0.392	-0.455	-0.376	-0.269
53	Government	Socialist Dummy	-0.097	0.009	0.014	0.013	-0.903	0.234	0.28	0.089	0.169
54	Government	Colony Dummy	-0.357	-0.03	-0.041	0.016	-3.521	-0.75	-0.737	0.05	-0.445
55	Government	Civil Liberties	0.221	-0.018	-0.027	-0.006	2.09	-0.46	-0.509	-0.021	-0.298
56	Government	Openess measure 1965-74	0.265	0.032	0.042	0.007	2.533	0.822	0.799	0.018	0.492
57	resources	Investment Price	-0.444	-0.067	-0.097	-0.383	-4.572	-1.776	-1.991	-2.483	-1.241
58	resources	Population in 1960	0.14	0.013	0.017	0.214	1.303	0.328	0.343	0.959	0.212
59	resources	Size of Economy	0.33	0	-0.003	-0.393	3.222	0.011	-0.059	-0.716	-0.017
60	resources	Malaria Prevalence in 1960s	-0.557	-0.03	-0.035	0.222	-6.181	-0.733	-0.633	0.71	-0.4
61	resources	Life Expectancy in 1960	0.548	0.022	0.031	-0.236	6.033	0.51	0.524	-0.521	0.299
62	resources	Hydrocarbon Deposits in 1993	0.07	0.017	0.029	0.063	0.648	0.431	0.576	0.354	0.327
63	resources	Population Growth Rate 1960-90	-0.371	-0.007	-0.008	-0.143	-3.679	-0.173	-0.139	-0.297	-0.089
64	resources	Population Density 1960	-0.115	0.03	0.05	0.051	-1.07	0.78	1	0.259	0.575
65	resources	Terms of Trade Ranking	0.066	-0.016	-0.016	-0.056	0.611	-0.389	-0.296	-0.21	-0.204
66	resources	Terms of Trade Growth in 1960s	0.013	-0.008	-0.014	-0.044	0.123	-0.203	-0.257	-0.194	-0.143
67	resources	Interior Density	0.121	-0.003	-0.003	0.007	1.121	-0.075	-0.054	0.045	-0.041
		R-squared		0.512	0.611	0.920					
		Mean Squared Coefficient, favorites		0.0156	0.0195	0.0877					
		Mean Squared Coefficient, others		0.0008	0.0014	0.5071					

Table 7 Summary table: s-values of the Favorites

s-values of the favorites at the 0.50 level: Three Approaches

	# of Explanatory Variables		67	67	67	14	14
	Favoritism		No Fav	No Fav	14 Fav		
	Lower R-square		0.1	0.5	0.5	0.1	0.5
	Upper R-square		1	1	1	1	1
	FAVORITE VARIABLES						
1	catchup	GDP in 1960 (log)	-0.17	-0.77	-1.42	-0.48	-2.72
2	culture	Fraction Speaking Foreign Language	0.17	0.67	1.34	0.67	3.36
3	geography	Absolute Latitude	0.04	0.10	0.79	0.67	3.36
4	geography	Fraction of Land Area Near Navigable Water	0.02	0.00	0.30	0.76	3.79
5	Government	Nominal Government GDP Share 1960s	-0.15	-0.65	-0.98	-0.42	-2.34
6	Government	Fraction Spent in War 1960-90	-0.01	0.00	-0.25	-0.24	-1.22
7	Government	British Colony Dummy	0.14	0.60	0.84	0.56	3.06
8	Government	Outward Orientation	-0.08	-0.41	-0.74	-0.03	-0.38
9	resources	Primary Schooling in 1960	0.31	1.22	2.82	1.34	6.89
10	resources	Fraction Population Less than 15	0.07	0.31	0.48	0.17	0.82
11	resources	Higher Education 1960	-0.07	-0.36	-0.18	0.16	0.66
12	resources	Fraction Population Over 65	0.01	0.06	0.04	-0.17	-0.83
13	resources	Primary Exports 1970	-0.10	-0.35	-1.19	-0.96	-4.81
14	resources	Oil Producing Country Dummy	0.03	0.13	0.48	0.29	1.94

OTHER STURDY COEFFICIENTS

Culture	Fraction Confucian	0.37	1.48	0.91
Culture	Fraction Buddhist	0.31	1.21	0.97
geography	East Asian Dummy	0.35	1.33	1.19
Government	Real Exchange Rate Distortions	-0.26	-1.02	-0.81
Resources	Investment Price	-0.44	-1.97	-1.24