

Jointness of Growth Determinants*

Gernot Doppelhofer[†]
University of Cambridge

Melvyn Weeks[‡]
University of Cambridge

August 2005

Abstract

Model uncertainty arises from uncertainty about correct economic theories, data issues and empirical specification problems. This paper investigates mutual dependence or *jointness* among variables in explaining the dependent variable. Jointness departs from univariate measures of variable importance, while addressing model uncertainty and allowing for generally unknown forms of dependence. Positive jointness implies that regressors are complements, representing distinct, but interacting economic factors. Negative jointness implies that explanatory variables are substitutes and act as proxies for a similar underlying mechanism. In a cross-country dataset, we show that jointness among 67 determinants of growth is important, affecting inference and economic policy.

Keywords: Model Uncertainty, Dependencies among Regressors, Jointness, Determinants of Economic Growth

JEL Classifications: C11, C52, O20, O50

*Ron Miller contributed to this project at early stages. We thank Steven Durlauf, Bruce Hansen, Chris Papageorgiou, Hashem Pesaran, Adrian Raftery, Jon Temple and an anonymous referee for helpful comments. Also thanks to seminar participants at Alicante, Barcelona, Bristol, Cambridge, Nicosia, Seattle and Madison for helpful comments. Doppelhofer thanks the Economics departments at UW-Madison and UPF Barcelona for their hospitality during his visits there and the Faculty of Economics at Cambridge for financial support. All errors are our own.

[†]Faculty of Economics and Trinity College, University of Cambridge, Cambridge CB3 9DD, UK. Email: gd237@cam.ac.uk, Tel: +44 1223 335200, Fax: +44 1223 335475.

[‡]Faculty of Economics, University of Cambridge, Cambridge CB3 9DD, UK. Email: mw217@econ.cam.ac.uk

1 Introduction

Model uncertainty is encountered in many areas of empirical work in economics. Typically there are a number of dimensions to such uncertainty, including theory, data and specification uncertainty. A policymaker faces considerable uncertainty given a set of multiple, overlapping theories which emphasize different channels in making inference or forecasting. Brock and Durlauf (2001) refer to this as “open-endedness” of economic theories, in the sense that one theory being true does not imply that another one is wrong. Within each theoretical channel there may be multiple competing measures representing the same theory. Furthermore, data problems might also be present. In many non-experimental settings the number of observations is often limited, and this problem may be compounded by the presence of missing data. Since it is often not clear *a priori* which theory is correct and which variables should be included in the “true” regression model, a naive approach that ignores specification and data uncertainty generally results in biased parameter estimates, overconfident (too narrow) standard errors and misleading inference and predictions.

Model averaging addresses issues of model uncertainty explicitly, and there is a recent and growing literature on this topic¹. Sala-i-Martin, Doppelhofer and Miller (2004) [henceforth SDM 2004] developed a method called *Bayesian Averaging of Classical Estimates* or *BACE* to estimate the posterior distribution of parameters of interest. From this distribution, which is unconditional with respect to the set of regression models, one can calculate the posterior inclusion probability, which measures the degree of confidence of including a variable with non-zero coefficients in the regression model. Such an unconditional, scale-free probability measure of relative variable importance can be a useful tool for policy decisions, inference and prediction, over and above parameter estimates such as posterior mean and standard deviation.

Much of the model averaging literature has neglected dependence among explanatory variables, in particular in specifying prior probability distributions over the space of models and in posterior inference. To address dependence among explanatory variables, we introduce the concept of *jointness* which is positioned between measures of model uncertainty and univariate measures of variable uncertainty. In contrast to univariate measures of variable uncertainty, jointness and related statistics are calculated by conditioning on a specific set of one or more other variables. However, jointness, like univariate measures, is unconditional with respect to the space of models and therefore takes model uncertainty into account. Although conditioning on some variables moves away from purely unconditional in-

¹See, for example, the surveys by Raftery (1995) and Hoeting, Madigan, Raftery and Volinsky (1999) on Bayesian model averaging, by Hjort and Claeskens (2003) on frequentist model averaging, and by Brock, Durlauf and West (2003) on applying model averaging to economic policy decisions.

ference, jointness investigates the sensitivity of posterior distribution and inference to dependence across regressors that might be hidden in the unconditional posterior distribution.

The *jointness* statistic introduced in this paper is defined as the log ratio of joint posterior inclusion probabilities of variables over the product of individual posterior inclusion probabilities. In other words, jointness is the log posterior odds ratio of joint compared to independent inclusion of explanatory variables. We also calculate the standardized posterior mean and so-called sign certainty of explanatory variables conditional on joint inclusion with other variables in the regression. We conceive of jointness in two fundamental ways.

Complementary or *positive jointness* among regressors could be due to two or more variables representing distinct, but complementary economic theories. In the context of economic growth for example, an economy's openness to international trade may be an important factor related to long term growth. However, a policymaker may be interested in the extent to which this factor interacts with other variables such as geography measures or distortionary policies such as real exchange rate distortions. In models of financial crises, institutional structures such as the rule of law and property rights may be required to successfully reform corporate governance. We call variables \mathbf{x}_i and \mathbf{x}_l *complements* if they exhibit positive jointness and our assessment of posterior probability depends positively on the mutual inclusion in the regression.

Alternately, *substitutable* or *negative jointness* may be manifest when two or more variables are proxies for a similar underlying mechanism. For example, in explaining economic growth there are competing measures of geographic and cultural characteristics and "quality of life" measures such as life expectancy or the extent of primary school enrolment. In propensity score models of financial loans, banks face a large amount of information designed to represent the risk of default. In the well known dataset, originally described in Lee (1996), there are 23 regressors measuring stability, demographic and financial variables, and there is substantial correlation between these indicators of default. We call indicators \mathbf{x}_i and \mathbf{x}_l *substitutes* if they exhibit negative jointness and are measuring the same underlying mechanism.

There are two main differences between the jointness approach proposed in this paper and other approaches suggested in the literature. First, jointness is revealed in the posterior distribution and is not dependent upon prior information on the nature of interactions that are likely to be important. In this respect, our jointness approach does not require to impose an *a priori* structure on interactions among explanatory variables or different theories. Second, our approach searches over a space of models and then estimates the degree of jointness among regressors while averaging across many different models. Parameters are therefore not estimated conditional on one particular model, but we estimate the unconditional distribution

of parameters and jointness across models. The jointness approach is therefore more general than existing approaches because it takes model uncertainty explicitly into account without requiring the assumption of a particular, generally *unknown* form of dependence or interaction.

We apply jointness to the dataset of SDM (2004), containing observations for 88 countries and 67 candidate variables in explaining economic growth. We find an important role for jointness among those growth determinants. In particular, we find evidence for significant negative jointness only between variables with relatively high (univariate) posterior inclusion probability and other regressors. This finding implies that only variables found to be significant growth determinants are being flagged up as significant substitutes for other regressors. Evidence for significant positive jointness or complementarity is found among a broader set of explanatory variables, including some regressors labeled “insignificant” according to univariate measures of variable importance. In contrast to negative jointness, complementary variables showing positive jointness reinforce size and significance of their mutual effects on economic growth. Compared to the potentially very large number of dependencies among growth determinants, we find a moderate number of significant jointness, implying that inference and policy decisions are not too complex, even when taking those dependencies into account.

The remainder of the paper is organized as follows: section 2 describes the statistical method of Bayesian model averaging. Section 3 derives the jointness statistic and discusses its importance for posterior inference and prior specification. Section 4 presents the empirical results for jointness of growth, and section 5 concludes.

2 Bayesian Model Averaging

Consider the following general linear regression model

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

where \mathbf{y} is a $(T \times 1)$ vector of observations of the dependent variable, for example, cross-sectional or panel observations of growth of income per capita, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_k]$ is a $(T \times k)$ matrix of explanatory variables (including an intercept) with $(k \times 1)$ coefficient vector β , and ε is a $(T \times 1)$ vector of residuals which are assumed to be normally distributed, $\varepsilon \sim N(\mathbf{0}, \sigma^2\mathbf{I})$, and *conditionally* homoscedastic².

Consider the problem of making inference on the determinants of the dependent variable, given the data $\mathbf{D} \equiv [\mathbf{y}, \mathbf{X}]$. Suppose that there are K potential regressors, then the model space \mathcal{M} is the set of all 2^K linear models. Each model M_j is

²If the assumption of conditionally homoscedastic residuals is violated, the mean slope coefficients (and therefore also model weights) would be consistently estimated. However, the estimation would be inefficient and inference would be affected.

described by a $(k \times 1)$ binary vector $\gamma = (\gamma_1, \dots, \gamma_K)'$, where a one (zero) indicates the inclusion (exclusion) of a variable \mathbf{x}_i in regression (1). For a given model M_j , the unknown parameter vector β represents the effects of the variables included in the regression model. We can estimate its density $p(\beta|\mathbf{D}, M_j)$ conditional on data \mathbf{D} and model M_j .

Given the (potentially large) space of models \mathcal{M} , there is uncertainty about the correct model and it is appropriate to consider the *unconditional* effects of model parameters by averaging over models. The density $p(\beta|\mathbf{D})$ permits unconditional inference by integrating out all aspects of model uncertainty, including the space of models \mathcal{M} (cf. Leamer 1978). The posterior density of slope coefficient estimates β given the data \mathbf{D} equals

$$p(\beta|\mathbf{D}) = \sum_{j=1}^{2^K} p(\beta|\mathbf{D}, M_j) \cdot p(M_j|\mathbf{D}) \quad (2)$$

where $p(\beta|\mathbf{D}, M_j)$ is the conditional density of β given model M_j . The posterior model probability $p(M_j|\mathbf{D})$ propagates model uncertainty into the posterior distribution of model parameters. By Bayes' rule it can be written as

$$\begin{aligned} p(M_j|\mathbf{D}) &= \frac{l(\mathbf{D}|M_j) \cdot p(M_j)}{p(\mathbf{D})} \\ &\propto l(\mathbf{D}|M_j) \cdot p(M_j) \end{aligned} \quad (3)$$

The posterior model probability (weight) is proportional to the product of the model-specific marginal likelihood $l(\mathbf{D}|M_j)$ and the prior model probability $p(M_j)$. The model weights are converted into probabilities by normalizing relative to the set of all 2^K models

$$p(M_j|\mathbf{D}) = \frac{l(\mathbf{D}|M_j) \cdot p(M_j)}{\sum_{i=1}^{2^K} l(\mathbf{D}|M_i) \cdot p(M_i)} \quad (4)$$

We follow SDM (2004) by assuming diffuse priors with respect to the error standard deviation σ and proper Normal-Gamma prior density for the slope coefficients β , centered at zero. However, we assume that information contained in the sample (through the marginal likelihood) “dominates” the prior information (see Leamer 1978). With these assumptions, the posterior model probability of model M_j relative to all 2^K possible model is given by

$$p(M_j|\mathbf{D}) = \frac{p(M_j) \cdot T^{-k_j/2} \cdot SSE_j^{-T/2}}{\sum_{i=1}^{2^K} p(M_i) \cdot T^{-k_i/2} \cdot SSE_i^{-T/2}} \quad (5)$$

where k_j is the number of regressors and $SSE_j = (\mathbf{y} - \mathbf{X}_j\beta)'(\mathbf{y} - \mathbf{X}_j\beta)$ is the sum of squared errors in model M_j . The posterior model weights are similar to the Schwarz model selection criterion (exponentiated) which is a function of the likelihood, but

penalizes relatively large models through the penalty term $T^{-k_j/2}$ on including additional regressors. This term will to a certain extent address collinearity among regressors, in the sense that models that include collinear regressors will receive lower weight compared to models with no added collinear variables. The intuition is that the explained sum of squares will not increase by much (due to collinearity), but the model is larger and hence is weighted down³.

Leamer (1978, p. 118) shows that mean and variance of slope parameters β can be calculated in a straightforward manner from conditional (model specific) parameter estimates. The mean of the posterior distribution of model parameter β_i , unconditional with respect to space of models \mathcal{M} , but conditional on including variable \mathbf{x}_i , equals

$$E(\beta_i|\gamma_i = 1, \mathbf{D}) = \sum_{j=1}^{2^K} \mathbf{1}(\gamma_i = 1|M_j, \mathbf{D}) \cdot p(M_j|\mathbf{D}) \cdot \hat{\beta}_{ij} \quad (6)$$

where $\hat{\beta}_{ij} = E(\beta_i|\mathbf{D}, M_j)$ is the OLS estimate for slope β_i given model M_j . The posterior variance of slope β_i , conditional on including variable \mathbf{x}_i , is given by

$$\begin{aligned} V(\beta_i|\gamma_i = 1, \mathbf{D}) &= \sum_{j=1}^{2^K} \mathbf{1}(\gamma_i = 1|M_j, \mathbf{D}) \cdot p(M_j|\mathbf{D}) \cdot V(\beta_i|\mathbf{D}, M_j) + \\ &+ \sum_{j=1}^{2^K} \mathbf{1}(\gamma_i = 1|M_j, \mathbf{D}) \cdot p(M_j|\mathbf{D}) \cdot \left[\hat{\beta}_{ij} - E(\beta_i|\mathbf{D}) \right]^2 \end{aligned} \quad (7)$$

where the conditional variance is estimated by the maximum likelihood⁴ estimator $V(\beta_i|\mathbf{D}, M_j) = \hat{\sigma}_j^2[\mathbf{X}'\mathbf{X}]_{ii}$, with error variance estimate $\hat{\sigma}_j^2 \equiv SSE_j/(T - k_j)$. Notice that the posterior variance of coefficient β_i consists of two terms: the sum of conditional (model-specific) variances and an additional term, taking into account the difference between conditional and posterior estimates of mean coefficients.

To ease comparison across variables, we calculate the *posterior standardized coefficient* associated with variable \mathbf{x}_i

$$E(\beta_i/\sigma_i|\gamma_i = 1, \mathbf{D}) \equiv \frac{E(\beta_i|\gamma_i = 1, \mathbf{D})}{\sqrt{V(\beta_i|\gamma_i = 1, \mathbf{D})}} \quad (8)$$

by dividing the posterior mean coefficient (6) by its posterior standard deviation σ_i , which simply equals the square root of the posterior variance (7).

SDM (2004) also suggest to estimate the probability the coefficient has the same sign as the posterior mean. This *posterior sign certainty* probability is given by

³We discuss alternative prior specifications in section 3.2 and the sensitivity of results with respect to alternative *AIC* model weights in section 4.3.3.

⁴Conditional on each model M_j , the maximum likelihood estimator is optimal, given that we consider a prior structure that is dominated by sample information. See for example Leamer (1978) or Raftery (1995) for further discussion.

$$p(s_i|\gamma_i = 1, \mathbf{D}) = \sum_{j=1}^{2^K} \mathbf{1}(\gamma_i = 1|\mathbf{D}, M_j) \cdot p(M_j|\mathbf{D}) \cdot CDF(t_{ij}) \quad (9)$$

where for each model M_j a non-centered cumulative t -distribution function is evaluated at standardized parameter estimates $t_{ij} \equiv (\hat{\beta}_{ij}/\hat{\sigma}_{ij}|\mathbf{D}, M_j)$ and $\hat{\sigma}_{ij}$ is the square root of the conditional (model-specific) variance estimate $V(\beta_i|\mathbf{D}, M_j)$.

A policymaker might be interested to know how important variables are in explaining the dependent variable. The *posterior inclusion probability* of variable \mathbf{x}_i

$$p(\gamma_i|\mathbf{D}) = \sum_{j=1}^{2^K} \mathbf{1}(\gamma_i = 1|\mathbf{D}, M_j) \cdot p(M_j|\mathbf{D}) \quad (10)$$

represents the probability that, conditional on the data, but unconditional of any particular model M_j , variable \mathbf{x}_i is relevant in explaining the dependent variable (cf. Leamer 1978 and Mitchell and Beauchamp 1988). This measure is therefore a model-weighted measure of relative importance of including variable i in the regression.

The posterior statistics presented above – standardized coefficient estimate (8), sign certainty (9), and inclusion probability (10) – have in common that they are calculated conditional on a variable’s inclusion⁵, but unconditional of models M_j . Hence, they allow for unconditional inference on the relative importance of the variable \mathbf{x}_i . However, such *unconditional* objects have the drawback of not revealing relationships between explanatory variables, unless all regressors \mathbf{X} are independent. To address this issue, the next section introduces a posterior object, positioned between the posterior measures of variable importance and model uncertainty, which will allow us to capture dependencies or jointness among explanatory variables.

3 Jointness

We start by equating model uncertainty with the best subset variable selection problem and by rewriting the posterior model probability as

$$p(M_j|\mathbf{D}) = p(\gamma_1 = s, \gamma_2 = s, \dots, \gamma_K = s|\mathbf{D}, M_j) \quad (11)$$

where $s = 1$ or 0 , depending on whether variable \mathbf{x}_i is included or not. Although the posterior probability $p(M_j|\mathbf{D})$ characterizes the degree of model uncertainty,

⁵The posterior mean $E(\beta_i|\mathbf{D})$, unconditional of including variable \mathbf{x}_i , equals the posterior mean (6) times the posterior inclusion probability (10). Similarly, the unconditional variance can be calculated from posterior estimates of moments (6), (7), conditional on including variable \mathbf{x}_i , and the posterior inclusion probability (10):

$$V(\beta_i|\mathbf{D}) = \{V(\beta_i|\gamma_i = 1, \mathbf{D}) + [E(\beta_i|\gamma_i = 1, \mathbf{D})]^2\} \cdot p(\gamma_i|\mathbf{D}) - [E(\beta_i|\mathbf{D})]^2$$

it is difficult (if not impossible) to detect dependencies among regressors, which contribute to the emergence of any given model and also determine the form of the posterior distribution over the space of models \mathcal{M} .

For the bivariate case, the posterior *joint* probability of inclusion for indicator i and l , is given by

$$p(\gamma_i \cap \gamma_l | \mathbf{D}) = \sum_{j=1}^{2^K} \mathbf{1}(\gamma_i = 1 \cap \gamma_l = 1 | \mathbf{D}, M_j) \cdot p(M_j | \mathbf{D}) \quad (12)$$

which simply sums posterior model probabilities for each model in which the two indicator appears. Analogous expressions of joint probabilities can be written for groups of more than two variables.

To formalize the degree of dependence or *jointness* among explanatory variables, a natural object of interest is the logarithm of the ratio of *joint* inclusion probabilities of a group of variables divided by the product of individual inclusion probabilities. For example, in the bivariate case the *jointness* statistic between two variables \mathbf{x}_i and \mathbf{x}_l is defined as

$$J_{il} = \ln \left(\frac{p(\gamma_i \cap \gamma_l | \mathbf{D})}{p(\gamma_i | \mathbf{D}) \cdot p(\gamma_l | \mathbf{D})} \right) \quad (13)$$

where γ_h indicates the inclusion of a variable $h = i, l$ in the regression with non-zero slope $\beta_h \neq 0$. The jointness statistic takes on the following values:

$$\begin{array}{ll} J_{il} < 0 & \text{substitutes} \\ J_{il} = 0 & \text{if variables are independent in posterior} \\ & \text{(as in prior)} \\ J_{il} > 0 & \text{complementary} \end{array}$$

The jointness statistic can also be viewed as logarithm of a *posterior odds ratio* of models that include variables \mathbf{x}_i and \mathbf{x}_l jointly, relative to models where they enter individually. The posterior odds ratio in turn can be decomposed into the product of relative prior probabilities and the ratio of marginal likelihoods, the so-called *Bayes factor*. As we describe in section 3.2, we assign *a priori* equal probability to both sets of models since the prior model size is not affected by two variables being jointly or separately included in the regression. We can therefore use a classification of Bayes factors, similar to the one suggested by Jeffreys (1961), to assess the significance of jointness among variables. In particular, we label variables with jointness $J_{il} < -1$ *significant substitutes* and variables with jointness $J_{il} > 1$ *significant complements*⁶.

Given the definition of joint inclusion probabilities, we may now redefine the posterior, univariate objects associated with variable \mathbf{x}_i introduced in section 2

⁶This corresponds to the modified version of Jeffreys' classification in Kass and Raftery (1995, p. 777). In addition, they call evidence with logarithmic Bayes factors in excess of 3 (less than 1/3) "strong" which corresponds to a posterior odds ration in excess of 20 (less than 0.05). We do not encounter such decisive values in the empirical application to growth determinants in this paper.

above, taking into account the inclusion of another variable \mathbf{x}_l . The *conditional mean* of the posterior distribution of model parameter β_i unconditional with respect to space of models \mathcal{M} , but conditional on the inclusion of another variable \mathbf{x}_l is given by

$$E(\beta_i|\gamma_i = 1 \cap \gamma_l = 1, \mathbf{D}) = \sum_{j=1}^{2^K} \mathbf{1}(\gamma_i = 1 \cap \gamma_l = 1|\mathbf{D}, M_j) \cdot p(M_j|\mathbf{D}) \cdot \widehat{\beta}_{ij} \quad (14)$$

In comparing the mean of the conditional distribution with the unconditional mean (6), we note the following. Unless $p(\gamma_i \cap \gamma_l|\mathbf{D}) = p(\gamma_i|\mathbf{D})$, the conditional mean is defined on a subset of models, relative to the unconditional mean. The extent to which the two means differ then depends upon whether the selection rule $\mathbf{1}(\gamma_i = 1 \cap \gamma_l = 1|\mathbf{D}, M_j)$ generates a non-random sample from the posterior density of β_i . Specifically, if the magnitude of β_i is generally independent of the inclusion of \mathbf{x}_l , we will observe a shrinking of $|\beta_i|$ towards zero. To isolate the heterogeneity of parameters over the conditional space of models from shrinkage due to joint model probabilities, we *normalize* all conditional expressions such as (14) by dividing by the sum of joint model probabilities (12). The amount of shrinkage can be inferred from comparing the joint with the individual inclusion probabilities contained in the jointness statistic.

Analogously to the derivation of the conditional mean, we can calculate the posterior variance for the slope coefficient $V(\beta_i|\gamma_i = 1 \cap \gamma_l = 1, \mathbf{D})$, conditional on inclusion of variable \mathbf{x}_l . In the expression for the unconditional variance (7), we replace the selection vector $\mathbf{1}(\gamma_i = 1|M_j, \mathbf{D})$ by the corresponding one for jointly including variables \mathbf{x}_i and \mathbf{x}_l , $\mathbf{1}(\gamma_i = 1 \cap \gamma_l = 1|\mathbf{D}, M_j)$. The *conditional standardized coefficient* is calculate by dividing the conditional mean (14) by the corresponding conditional standard deviations

$$E(\beta_i/\sigma_i|\gamma_i = 1 \cap \gamma_l = 1, \mathbf{D}) \equiv \frac{E(\beta_i|\gamma_i = 1 \cap \gamma_l = 1, \mathbf{D})}{\sqrt{V(\beta_i|\gamma_i = 1 \cap \gamma_l = 1, \mathbf{D})}} \quad (15)$$

Exactly the same replacement of selection vectors is used to calculate the *conditional sign certainty* probability for coefficient β_i , conditional on inclusion of variable \mathbf{x}_l

$$p(s_i|\gamma_i = 1 \cap \gamma_l = 1, \mathbf{D}) = \sum_{j=1}^{2^K} \mathbf{1}(\gamma_i = 1 \cap \gamma_l = 1|\mathbf{D}, M_j) \cdot p(M_j|\mathbf{D}) \cdot CDF(t_{ij}) \quad (16)$$

from the unconditional sign certainty (9). Note that in contrast to jointness which is symmetric $J_{il} = J_{li}$, the conditional standardized coefficients and sign certainty statistics are not symmetric, since conditional moments are not, e.g. $E(\beta_i|\gamma_i = 1 \cap \gamma_l = 1, \mathbf{D}) \neq E(\beta_l|\gamma_i = 1 \cap \gamma_l = 1, \mathbf{D})$.

3.1 Jointness and Posterior Inference

As mentioned in the Introduction, a policymaker can be confronted with several dimensions of model uncertainty, including partially overlapping theories and a potentially very large set of competing explanatory variables. These types of uncertainty have been used to motivate a model averaging approach to conducting policy inference. For example, Brock et al. (2003, p.281f) argue that a policymaker, following a particular “*t*-statistic” decision rule based on a quadratic loss function, evaluates policies utilizing *unconditional* standardized coefficients (8). However, there are limits to the inference about the posterior distribution of a policy effect in an purely unconditional setting.

Although model uncertainty with respect to individual variables is taken into account, dependence among variables is not revealed. Such dependence may manifest itself in the form of parameter heterogeneity, which may be of importance to a policymaker. For example, when evaluating a given policy, it matters if the effect is constant over a population, or whether it varies according to one or more characteristics of the observations (e.g. countries or firms). Brock and Durlauf (2001) argue that this type of heterogeneity can be readily integrated within a model averaging context. They suggest to impose prior information on possible sources of heterogeneity, and treat it as a variable inclusion problem by using simple interactions with dummy variables. However, as the number of competing theories increases, dependencies may require increasingly specific and complex prior information to become apparent.

Dependence among variables could also occur in situations where policies are not considered independently, but administered as a package⁷. If two or more policies or economic theories are highly complementary, then the use of simple *univariate* posterior objects, such as variable inclusion probability or standardized coefficients, may obscure such relationships. Alternatively, different theories might operate independently of one another in many economic applications, and Brock and Durlauf (2001) call this “open-endedness” of theories. To consider in more detail the additional information revealed by jointness and related conditional statistics (standardized coefficients and sign certainty), we consider a number of hypothetical scenarios.

1. A situation where we find very little significant positive jointness implies, both in terms of variables acting as policy instruments and controls, there is little complementarity. One might then conclude that little positive jointness is good for policymakers, since it implies that the distribution of policy instruments and conditioning variables is less complex (and unknown).

⁷In laying out a framework for policy evaluation in the face of various dimensions of uncertainty, Brock et al. (2003) treat the probabilities of each theory as approximately independent. Section 3.2 discusses implications of making different *prior* assumptions about variable (in)dependence for posterior inference and policy decisions.

2. Jointness over a set of regressors appears as significantly negative. This finding implies the reverse of the above. Namely, policies act as substitutes in terms of instruments to affect growth. This then means that collinearity and issues of prior distributions are important.
3. For a particular regressor \mathbf{x}_i the univariate posterior inclusion probability $p(\gamma_i|\mathbf{D})$ is low, but one or more joint inclusion probabilities $p(\gamma_i \cap \gamma_h|\mathbf{D})$, $h = 1, 2, \dots$ are high, suggesting that \mathbf{x}_i need complementary variables \mathbf{x}_h to explain the dependent variable. In such a situation, it is also like that the conditional standardized coefficient will increase significantly once the complementary variable is conditioned upon. In this instance, if policies are implemented based upon a package of measures, an analysis of univariate measures, such as $p(\gamma_i|\mathbf{D})$, may be misleading. An example could be that a particular determinant of growth requires appropriate conditioning on a set of other explanatory variables (e.g. to control for steady state conditions) before an effect can be realised. We would also expect that the sign certainty of the coefficient β_i to be strengthened when complementary variables are included in the regression.
4. Suppose the univariate posterior inclusion probability $p(\gamma_i|\mathbf{D})$ is high, but joint inclusion probabilities $p(\gamma_i \cap \gamma_h|\mathbf{D})$, $h = 1, 2, \dots$ are low, implying that variable \mathbf{x}_i with high univariate inclusion probability is a close substitute for \mathbf{x}_h . An example, could be a “catch-all” variable that measures similar underlying characteristics of the dependent variable as other regressors. We would expect conditional normalized coefficients to be shrinking towards zero when variables measuring “similar” underlying concepts are included in the regression.

In the empirical application to economic growth in section 4, we will illustrate implications of jointness and related conditional statistics for choice of prior, inference and policy decisions.

3.2 Jointness and Prior Specifications

A number of alternative approaches have been suggested in the literature to deal with dependencies among explanatory variables. We will discuss them in turn and relate them to our jointness approach. Alternative approaches can be interpreted as introducing different types of prior information when dealing with such problems. The choice of prior distribution over the model space, specifically whether an informative prior with emphasis on parsimonious (saturated) models with a few (large) number of variables, or an uninformative one is chosen, is critical in this respect.

One prominent example, where inference is sensitive to changes in sample and non-sample (prior) information, is the problem of collinearity amongst a set of regressors, indicating a lack of independent variability or the presence of “weak data”

(see Leamer 1983 and Judge et al 1985). Solutions to the collinearity problem correspond to the introduction of additional information, either by having more data or alternatively through prior information.

3.2.1 A “Classical” Response: Specification Search

One possible strategy in dealing with weak and collinear data might be to test across models and identify a single model chosen by a set of selection criteria. Hendry (2000) proposes the general-to-specific (*Gets*) modeling strategy to deal with variable selection. There are some general problems related to specification tests and post-selection inference. For a critical view see for example Pötscher (1991) and Leeb and Pötscher (2004). Recently there have been a number of advances in both the theory and operational aspects of the *Gets* methodology (see for example, Hoover and Perez 1999, Krolzig and Hendry 2000, and Campos and Ericsson 2000). There are however still a number of problems which restrict the applicability of this approach for our particular study.

First, in many non-experimental data there is a trade-off between model size and data availability. The more general our initial model, the smaller the set of non-missing observations. Faced with this occurrence it becomes difficult to differentiate between true economic insignificance and statistical insignificance driven by a small sample. Second, there is the related problem of collinearity. This problem exists as a result of both the similarity of candidate indicators and the relatively small sample size. Krolzig and Hendry (2000) note that this problem significantly complicates a general-to-specific modeling strategy, resulting in a large number of candidate models which differ marginally. Granger, King and White (1995) extend this point, noting data problems such a sample size and collinearity confer a favorable advantage to a null model, which in our case is particularly pertinent. This is essentially a small sample or “weak data” problem that would tend to disappear in large enough samples as more data are becoming available.

Perez-Amaral, Gallo and White (2003) developed the so-called *RETINA* approach which tests interactions among main variables and incorporates interactions into the selected model using cross-validation and estimation steps. We differ from this approaches by allowing for generally unknown forms of dependencies among explanatory variables and by estimating dependencies across variables unconditionally over the space of models.

In the presence of model uncertainty, we do not wish to condition inference and policy on one particular model, since inference and prediction could be misleading. Furthermore, we cannot analyze the unconditional dependence among explanatory variables across models once we condition on a model.

3.2.2 A Bayesian Response: Independence Priors

Another possibility is to ignore the problem of potential dependence among variables and treat the regressors \mathbf{x}_i as if they were independent with diagonal design matrix $\mathbf{X}'\mathbf{X}$. In Bayesian analysis, this corresponds to imposing uniform priors on the model space and assume that variables are independently included in the model. The corresponding prior probability for model M_j can be written as

$$p(M_j) = \prod_{i=1}^K \pi_i^{\gamma_i} (1 - \pi_i)^{1-\gamma_i} \quad (17)$$

where π_i is the prior inclusion probability of variable \mathbf{x}_i in M_j , with corresponding indicator γ_i . *Uniform priors* corresponds to setting $\pi_i^U = 1/2$ for all i (cf. George and McCulloch 1993).

As a consequence of the Bernoulli structure, departures from independence in the inclusion probabilities for all variables within a given model are not accounted for. One could argue that uniform priors may be appropriate if the regressors are likely to capture substantively different theories, but even then theories could interact with each other which would show up in our jointness measure. For example, in instances where two indicators \mathbf{x}_i and \mathbf{x}_l are considered close proxies (complements), then for $\gamma_i = 1$ the likelihood that \mathbf{x}_k is included in a given model may be less (greater) than that implied by an independence assumption. Interestingly, George and McCulloch (1993, p. 884) mention the possibility of correlation (or dependence) between explanatory variables and suggest to look at “conditional frequencies” without elaborating this point further. Jointness formalizes the dependence among regressors and brings out their relation contained in the posterior distribution.

With a relatively large number of regressors K , the choice of a uniform prior implies that the great majority of prior probability is focussing on models with a large number of variables. As an alternative, SDM (2004) introduce a prior expected model size \bar{k} and corresponding prior inclusion probability $\pi_i^{BACE} = \bar{k}/K$. Figure 1 shows the prior distribution over model sizes for our benchmark case with $\bar{k} = 7$ (implied prior inclusion probability $\pi_i = 7/67 = 0.104$) and for the uniform prior case with $\pi_i = 1/2$ (and implied prior model size $\bar{k} = 33.5$). A disadvantage of such an approach is that the notion of what constitutes a reasonable prior model size may vary across analysts. In response to this criticism Godsill, Stone and Weeks (2004) introduce another layer of prior information by combining independent Bernoulli sampling for each variable with a conjugate Beta prior for the binomial proportion parameter π_i .

[Figure 1 about here]

3.2.3 Another Bayesian Response: Dilution Priors

George (1999) observes that when the number of candidate explanatory variables and corresponding model space is relatively large, posterior model probabilities can be “diluted” and spread over many similar models when averaging over models. He suggests to use *dilution priors* instead, to reduce the emphasis on “similar” areas of the model space where several variables are proxying for the same underlying mechanism and to increase the importance of unique and economically important explanations.

Brock and Durlauf (2001) criticize independent Bernoulli priors (17) on the ground that the inclusion of one variable in the regression should affect the probability that other variables are included⁸. The issue of dilution of model probabilities is closely related to the “blue bus/red bus” problem in discrete choice theory, captured by the assumption of *irrelevance of independent alternatives (IIA)*. This is illustrated by the following example (see George 1999).

Suppose variables $\mathbf{x}_1, \mathbf{x}_2$ are true determinants of \mathbf{y} in (1) with the following posterior probabilities.

M_j	\mathbf{x}_1	\mathbf{x}_2	$\mathbf{x}_1, \mathbf{x}_2$
$p(M_j \mathbf{D})$	0.3	0.4	0.2

\mathbf{x}_3 which is highly correlated with \mathbf{x}_2 is added to the set of regressors. With a uniform prior on the model space the posterior probabilities become something like:

M_j	\mathbf{x}_1	\mathbf{x}_2	\mathbf{x}_3	$\mathbf{x}_1, \mathbf{x}_2$...
$p(M_j \mathbf{D})$	0.15	0.2	0.2	0.1	...

We observe that in order to accommodate the inclusion of an additional variable the total probability (across models with just \mathbf{x}_1 , just \mathbf{x}_2 and \mathbf{x}_1 and \mathbf{x}_2) has been diluted in a proportionate fashion. This is exactly analogous to the effects of the *IIA* assumption on the substitution effects in the “vanilla logit” model. Namely, the expansion of the choice set to include an additional alternative which is very close to an existing alternative over a space of attributes, will draw probability away from existing alternatives in a proportionate fashion. Analogously, in the context of expanding the space of regressors to include \mathbf{x}_3 , the set of regressors has not added any new information in the sense that \mathbf{x}_3 is merely seen as a substitute \mathbf{x}_2 .

Dilution priors, first advocated by George (1999), represent one solution to this problem by avoiding to put excess probability on many similar models, and thereby allocate more probability to unique and informative models. In the above example the posterior could look as follows.

⁸The uniform prior suggested by Brock and Durlauf (2001) does not solve the dilution problem, since the inclusion of variables is still assumed to be independent. As pointed out earlier, the only difference is a larger expected model size compared to the BACE priors.

M_j	\mathbf{x}_1	\mathbf{x}_2	\mathbf{x}_3	$\mathbf{x}_1, \mathbf{x}_2$...
$p(M_j \mathbf{D})$	0.3	0.2	0.2	0.05	...

George (2001) suggests the following modification to the independence prior (17)

$$p(M_j) = |\mathbf{R}| \prod_{i=1}^K \pi_i^{\gamma_i} (1 - \pi_i)^{1-\gamma_i}$$

where \mathbf{R} is the correlation matrix of regressors and proportional to $\mathbf{X}'\mathbf{X}$. Note that $|\mathbf{R}| = 1$ when the regressors are orthogonal, and $|\mathbf{R}| \rightarrow 0$ with columns of \mathbf{X} becoming more collinear. This prior therefore penalizes models with “similar” \mathbf{x}_i ’s.

We do not use dilution priors in this paper because we want to emphasize possible dependencies among regressors. In contrast to George’s (1999) *prior* dilution of the model space, we obtain a *posterior* measure of jointness among regressors that is unconditional of other variables. However, the estimated jointness statistic could be used to “dilute” the weight on variables capturing similar economic phenomena. Alternatively, one can use prior information (based on economic theory, for example) to interpret the jointness results further⁹.

3.2.4 Hierarchical Priors and A Priori Interactions

Another possibility is to impose a hierarchical structure on the model space (see for example Chipman, George and McCulloch 2001). For example, Brock et al (2003) suggest partitioning the model space along the following lines: First, they allow for different (distinct) theories to explain economic growth. Second, the variables are sorted into categories proxying for different theories. Third, heterogeneity of parameters and specifications is allowed.

We do not wish to follow this approach for two reasons: First, as discussed in the Introduction, theories of economic growth could exhibit interesting interactions in determining economic performance. These would not be considered by imposing independence priors (homogeneity) over different theories in the first layer. Second, variables could be correlated with each other and not exclusively proxying for one underlying theories of growth. In the hierarchical setup, variables are restricted to fall into certain categories *a priori*, and one could miss important parts of the model space. Ideally, the sensitivity of results with respect to these prior restrictions should be tested.

Chipman (1996) provides a framework for conducting Bayesian variable selection when the standard independence prior is not appropriate. Chipman (1996) considers a number of examples where it is appropriate to build in dependence across regressors in formulating model priors. These include cases where an analyst considers the significance of polynomial terms, say \mathbf{x}^h , or more generally an interaction effect

⁹Leamer (1973) advises using non-data information to interpret weak evidence.

$\mathbf{x} \times \mathbf{z}$. Prior distributions are constructed which in considering notions of inheritance, impose constraints on whether \mathbf{x} and \mathbf{z} should be included if $\mathbf{x} \times \mathbf{z}$ is present, and to force the inclusion of all terms \mathbf{x}^a $a < h$. Our jointness approach differs in that in contrast to utilizing prior knowledge as to the exact form of interactions, jointness allows for general (unrestricted) forms of dependence among regressors which is revealed in the posterior distribution.

3.2.5 Orthogonalization of Regressors

An additional possibility is to transform the data and make the design matrix $\mathbf{X}'\mathbf{X}$ diagonal, for example by decomposing \mathbf{X} into principle components¹⁰ or other factors. This approach is frequently suggested in the model averaging literature (see for example Clyde, Desimone and Parmigiani 1996). The main advantages are computational savings and speed. The disadvantage is the loss of economic interpretation of parameter estimates. The latter might be less relevant if one is primarily interested in prediction. For example, Stock and Watson (1999) propose to forecast inflation by combining a large number of macroeconomic variables into factors through principle components analysis. However, we are particularly interested in the interpretation of determinants of growth and therefore do not follow this route in this paper.

A more subtle problem with orthogonalization procedures is that factors are calculated *conditionally* on the available data sample. If new data are introduced, factors and their respective loadings will change. Using the original (untransformed) variables has the advantage that their interpretation is not changing as more data become available. The distribution of parameters will of course change in general, as indicated by the dependence of posterior distributions on the sample data \mathbf{D} .

4 Jointness of Growth Determinants

This section presents results of applying jointness to the determinants of economic growth dataset of SDM (2004). The explanatory variables are chosen from regressors found to be related to economic growth in earlier studies (see for example the list in Durlauf and Quah, 1999). SDM (2004) select variables that represent “state variables” in economic growth models and measure them as close as possible to the start of the sample period in 1960. Furthermore, the dataset is restricted to be balanced, i.e. without missing observations. Under these criteria the total number of explanatory variables is $K = 67$ and with observations for $T = 88$ countries¹¹.

¹⁰One can detect the problem of collinear regressors that are associated with “small” eigenvalues of $\mathbf{X}'\mathbf{X}$. However, Leamer (1983) points out that there is always a reparametrization of \mathbf{X} that makes $\mathbf{X}'\mathbf{X} = \mathbf{I}$, but the data could still be non-informative about certain parameter regions.

¹¹For list of data sources and countries with complete observations see SDM (2004), Table 1 and Appendix A1. We are addressing data issues, such as missing observations, in ongoing research.

The Data Appendix shows the data, the dependent variable, *average growth rate of GDP per capita between 1960-96*, and the 67 explanatory variables. Also shown are short names, a brief variable description, and sample mean and standard deviations. In the Data Appendix and Tables, explanatory variables are ranked by posterior inclusion probability (10), which is shown in the fourth column of the Data Appendix table. The posterior can be compared to the prior probability of inclusion, which equals $\pi_i^{BACE} = \bar{k}/K = 7/67 = 0.104$ in the benchmark case with prior model size $\bar{k} = 7$. SDM (2004) call the 18 highest ranked explanatory variables with posterior inclusion probability greater than the prior, “significantly” related to economic growth.

Table 1 shows the results of bivariate jointness, defined in equation (13), between the explanatory variables. As described in section 3, a negative jointness value indicates that two explanatory variables are substitutes in explaining economic growth, whereas a positive value indicates that they are complements. We call absolute values of jointness in excess of unity “significant”, reflecting evidence from the posterior odds ratio¹². Figure 2 shows the kernel density of jointness for all 67 explanatory variables and selected subsets of regressors (top, 21 top and last ranked by posterior inclusion probability). This figure shows the relative frequency of significant negative or positive values of jointness among groups of regressors.

We also report the conditional standardized coefficients (15) and the conditional sign certainty statistics (16) in tables 2 and 3, respectively. Notice that these tables contain conditional statistics associated with variable \mathbf{x}_i (in rows numbered 1 to 67), *conditional* on inclusion of variable \mathbf{x}_l (in columns numbered in the first row of each table). The diagonal entries (with coinciding row i and column l index) contain the *unconditional* standardized coefficients (8) and sign certainty statistics (9).¹³

The main empirical findings for the growth determinants are as follows:

1. We find evidence for both negative, as well as positive jointness among growth determinants. Jointness detects dependencies among regressors in an unconditional sense (across many possible growth models) and differs therefore from simple correlation measures which are independent of the models under consideration. We find no simple relationship between jointness and bivariate correlations¹⁴.
2. *Significant negative* jointness (with $J_{il} < -1$) occurs only between “significant”

¹²Variables exhibiting *significant negative* jointness (substitutes with $J_{il} < -1$) are set in *italics*. Variables with *significant positive* jointness (complements with $J_{il} > 1$) are set in **boldface**.

¹³The diagonal entries are unconditional with respect to inclusion of other variables \mathbf{x}_l , since clearly $\mathbf{1}(\gamma_i = 1 \cap \gamma_l = 1) = \mathbf{1}(\gamma_i = 1)$. The ease comparison with the other (conditional) entries, the diagonal entries are also set in **boldface** in tables 2 and 3.

¹⁴Simple bivariate correlations are not reported due to space constraints. Full set of tables and results are made available on our website www.econ.cam.ac.uk/doppelhofer.

variables, i.e. those with posterior inclusion probability greater than the prior, and other explanatory variables (including significant variables themselves). This indicates that only variables with relatively high inclusion probability in explaining growth are flagged as *significant substitutes* for other variables. Variables with posterior inclusion probability smaller than the prior inclusion probability do not exhibit significant negative jointness with variables outside the top 18 regressors.

3. *Significant positive* jointness (with $J_{il} > 1$) is found among a wider set of growth determinants, that are labeled “significant” and “insignificant” in an unconditional sense according the posterior inclusion probability. Some growth determinants benefit from inclusion of complementary variables in explaining economic growth and become more important in explaining economic growth, conditional on including the complimentary variable in the regression.
4. The *conditional standardized coefficients* are behaving similar to the jointness results. Negative jointness is associated with smaller conditional standardized coefficients (in absolute value terms). The inclusion of substitutes in the regression shrinks standardized coefficients towards zero. Positive jointness gives the opposite results: it strengthens standardized coefficients and increases their importance in a conditional sense. These findings are potentially very important for a policymaker interested in the the size and significance of variables, conditional on a set of controls or other policies.
5. The conditional *sign certainty* statistic follows the same qualitative pattern as standardized coefficients: positive jointness is associated with higher conditional sign certainty of effects and negative jointness with lower values of sign certainty. This is not surprising, given that the sign certainty averages t -statistics across models, whereas standardized coefficients are calculated from averaged coefficient estimates and their standard deviations.

[Figure 2 about here]

The following two subsections discuss in greater detail the results for growth determinants exhibiting significant negative and positive jointness, respectively. The results are presented in approximate univariate order of variable importance. We conclude by showing sensitivity of jointness with respect to prior model size \bar{k} , inclusion of additional regressors and alternative AIC model weights.

4.1 Negative Jointness and Substitutes

Negative jointness indicates that after averaging over alternative models, two determinants of growth have lower probability of joint inclusion in those models than

entering individually. These variables therefore act as substitutes for each other in an unconditional sense (after averaging over models). Interestingly, significant negative jointness among growth determinants is limited to variables with high posterior inclusion probability. Only the posterior distributions of variables with high posterior weight (relative to the prior inclusion probability) are significantly affected by separate or joint inclusion with other variables capturing “similar” aspects of economic growth.

There are several groups of variables exhibiting negative jointness. First, some variables measuring geographic or cultural differences across countries are substitutes. The *East Asian* dummy has negative jointness with several other regional and cultural variables, significantly so with the *Sub Saharan Africa* dummy and marginally with the *Latin America* dummy, the *Fraction Confucian*, and the *Fraction Buddhist* of the population. The conditional standardized coefficient shown in Table 2 follows the same pattern: the size of the coefficient for the East Asian dummy falls from 3.55 unconditionally to 2.62 conditional on including the Sub Saharan Africa dummy. This pattern is perhaps not surprising since the East Asian dummy acts as a “catch-all” variable that loses significance once other regional and cultural variables are included as regressors¹⁵. The significant negative jointness between the *Latin America* and *Spanish Colony* dummy could have been expected from the relatively high bivariate correlation 0.84. Negative jointness is however more general than simple correlation since it captures dependence in an unconditional sense, i.e. after averaging across models.

A second related group exhibiting negative jointness are between geography measures and the *Malaria Prevalence* capturing the disease environment a country is facing¹⁶. The *Fraction of Tropical Area* and the *Sub Saharan Africa* dummy show significant negative jointness with *Malaria Prevalence*, consistent with unfavorable disease environments present in these regions. Malaria Prevalence has also significant negative jointness with *Coastal Population Density* and overall *Population Density*, showing interesting dependence between adverse disease environment captured by Malaria Prevalence and population density. Conditional on including both variables exhibiting negative jointness the conditional standardized coefficients are smaller in absolute value, less negative for Malaria Prevalence and less positive for both population density variables. *Fraction of Tropical Area* shows also a significant negative jointness with the *Latin America* and *Spanish Colony* dummy, and Spanish colonies in turn share a negative jointness with *Coastal Population Density*. This last jointness pattern reflects geographic characteristics of former Spanish colonies

¹⁵This jointness pattern is consistent with the finding that the unconditional inclusion probability for the East Asian dummy falls with larger prior model sizes as more explanatory variables are included. See table 3 of SDM (2004).

¹⁶This finding agrees with Gallup and Sachs (2001) on the interaction of geographic factors and the economic burden of Malaria.

in Latin America.

A third group of variables exhibiting negative jointness consists of sectoral and geographic variables. The *Fraction of GDP in Mining* exhibits strong negative jointness with *Population Density*. Including both variables simultaneously in the regression significantly shrinks the positive standardized coefficients, indicating that both capture similar aspects of cross-country variation in economic growth. *Coastal Population Density* and the share of *Exports and Imports in GDP* act as substitutes. Countries with high coastal density are likely to have more opportunities to trade leading to higher trade shares of GDP. This finding can be interpreted as a version of the “gravity equation” in the empirical trade literature.

A fourth group consists of variables measuring government spending¹⁷. *Government Consumption Share* and *Government Share of GDP* are significant substitutes for one another, reflected also in the very high correlation 0.93. An analyst might conclude that these two variables measure very similar aspects of government spending, adding little information when being jointly included in the regression. Table 2 shows that the standardized posterior coefficient for the Government Share of GDP switches sign (to positive) conditional on joint inclusion with Government Consumption Share. The sign certainty reported in Table 3 falls for both variables conditional on joint inclusion: from 0.97 to 0.87 for the Government Consumption Share, and from 0.93 to 0.69 for the Government Share of GDP.

From observing the posterior distribution of coefficients and the negative jointness described above, an analyst or policymaker can learn about groups of variables acting as substitutes for one another in explaining economic growth across countries. In response, they might want to put less weight on models that jointly include substitutable variables and emphasize models with more independent explanatory variables instead (see discussion on dilution priors in section 3.2.3). In other words, observing this part of the posterior distribution can affect (prior) opinions on dependencies among regressors. In some cases, negative jointness can be spotted upfront looking at simple correlations. However, jointness is more general and allows to uncover dependencies in an unconditional sense after averaging across models.

4.2 Positive Jointness and Complements

Positive jointness indicates that across different models, determinants of growth have a higher probability of joint inclusion than when they enter individually in the regression. We therefore label such variables complements since they reinforce each other in explaining economic growth. In contrast to negative jointness, we find

¹⁷There is one additional case of *marginal* negative jointness among the growth determinants outside the top 18 regressors: the real and nominal government shares of GDP (ranked 24th and 33rd, respectively) have jointness equal to -0.95 . The simple correlation between the two government shares equals 0.48.

evidence for positive jointness also among variables with unconditional posterior inclusion probability lower than the prior. Some variables need the right conditioning variables to achieve relatively higher (joint) posterior inclusion probability. Some variables classified “insignificant” determinants of growth in an unconditional sense, become “significant” once conditioning on a particular set of variables.

Strong evidence for complementary relationships implies more complex dependencies among determinants of growth. This would complicate policy analysis in the sense that conditioning on the right variables has to be taken into account. If, on the other hand, we find little evidence for positive jointness, both inference (both statistical and economic) and resulting policy actions are well guided by unconditional characteristics of the posterior distribution of parameters.

We find evidence of positive jointness among some groups of variables. First, some geographic or cultural variables act as complements for one another. For example, the *Fraction Confucian*, the *Sub Saharan Africa* as well as the *Latin America* dummies are significant complements for one another. The *Fraction Confucian* is also complementary to the *Fraction Buddhist*. This pattern of complementarity indicates that these regional and cultural variables should be jointly included in the model and form a “conditioning set” in explaining growth. As pointed out earlier, the *East Asian* dummy is a substitute for all variables in this conditioning set of variables. If not included in the regression model, a larger set of the conditioning dummy variables is suggested. The *Fraction Catholic* and *Fraction Protestant* are also significant complements. In all cases, positive (complementary) jointness strengthens the standardized coefficients in Table 2 and increases the sign certainty in Table 3, conditional on including the significant complementary variables.

Second, some variables measuring disease environment, geographic factors and colonial history form a complementary group. *Malaria Prevalence* has significant positive jointness with the *Spanish Colony* dummy, reinforcing the negative coefficient of both variables with economic growth when being jointly included in regression models. The *Fraction of Population in Tropics* and *Fraction of Land Area Near Navigable Water* are significant complements in explaining growth. Table 2 shows that the standardized *negative* size of the coefficient for proximity to Navigable Water is strengthened from -0.44 unconditionally to -1.64 conditional on controlling for the Tropical Population (the corresponding conditional sign certainty in Table 3 equals 0.94, compared to 0.65 unconditionally). The dummies for former *British Colony* and *Colony* are also complementary variables, allowing for and strengthening the positive and negative effects associated with Former British and Non-British colonies, respectively.

Third, variables related to geography and openness interact positively. The *Population Density* and *Air Distance to Big Cities* are significant complements. Conditional on high population density a small distance to big cities are important and

vice versa. *Terms of Trade Growth in the 1960s* and *Real Exchange Rate Distortions* are complements, as are terms of trade growth and the *Population* size. Conditional on the growth of terms of trade in the 1960s, the negative effect of real exchange rate distortions and the positive effects of population size are strengthened, the latter not being statistically significant.

Fourth, the *Mining Share of GDP* has significant positive jointness with the *Public Investment Share* in GDP. The signs of the effects on economic growth are again reinforced when conditioning on the complementary variable, implying a stronger positive effect of the Mining share and a more *negative* (though not statistically significant) effect of the Public Investment Share on economic growth. These conditional estimates could be of interest to policymakers, in particular if a country has a relatively important Mining industry.

Finally, the variables measuring *Political Rights* and *Civil Liberties* are significant complements and also have a high correlation coefficient -0.83 . Neither variable on its own is significantly related to economic growth, but conditional on joint inclusion the *positive* effect¹⁸ of Political Rights and *negative* effect of Civil Liberties are reinforced. It appears that conditional on joint inclusion, the two variables measure different aspects of how political institutions effect economic growth in a cross-section of countries.

So what can an analyst or policymaker learn from positive jointness? First, we find that some growth determinants require a set of conditioning variables to jointly explain growth across countries. This appears to be true in particular for some variables measuring geographic and cultural differences across countries, capturing heterogeneity of cross-country growth rates. Second, some positively related variables capture different aspects of the cross-country growth mechanism, and provide a differentiated explanation. Examples include the Mining share and Public Investment Spending share or real exchange rate distortions and terms of trade growth. When implementing a particular policy, a policymaker might want to take these dependencies into account. In contrast to negative jointness, positive jointness implies that growth determinants reinforce the size and significance of their effect. However, given the number of growth determinants under consideration and the large number of possible dependencies, the number of cases of significant positive jointness is relatively small, implying that inference and policy decisions is not rendered unmanageably complex.

¹⁸Note that an increase in the Political Rights index implies *less* political freedom. The negative standardized coefficient associated with Political Rights in Table 2 corresponds implies a positive relation with economic growth.

4.3 Sensitivity Analysis of Jointness

This section presents results¹⁹ of sensitivity analyzes of jointness compared to the benchmark case presented above. We analyze sensitivity with respect to changes in (i) prior model weight \bar{k} , (ii) posterior model weights $P(M_j|D)$ and (iii) allowing for additional interaction terms as regressors.

4.3.1 Sensitivity of Jointness to Prior Model Size

First, we investigate the sensitivity of Jointness with respect to changing the prior model size \bar{k} . We summarize jointness results for different prior model sizes $\bar{k} = 4, 14, 21, 28$ and contrast them with the benchmark case with $\bar{k} = 7$. Figures 3 (for the top 21 regressors) and 4 (for the remaining regressors) show that the degree of dependence among explanatory variables weakens with increasing prior model size. This is perhaps not surprising, because larger model sizes imply a smaller penalty of adding regressors in the posterior model weights (5).

[Figures 3 and 4 ABOUT HERE]

Larger models imply that there is “less competition” among regressors to enter the regression models, which explains reduced negative jointness (substitutability) in larger models. For example, only one combination of variables exhibits significant negative jointness with prior model size $\bar{k} = 28$, namely the *Government Consumption Share* and overall *Government Share of GDP*. As we pointed out in section 4.1, these two variables have very highly correlation 0.93, explaining why they are significant substitutes even in relatively generously specified models.

Perhaps more surprising is the observation that positive dependence (complementarity) is also falling in larger models. Larger models also imply that a richer set of other conditioning variables is allowed to enter in each model M_j , which explains the reduced role of complementarity among any particular pair of regressors. For example, for relatively large prior model size $\bar{k} = 28$, there is not a single pair of complementary variables. For slightly smaller models with prior size $\bar{k} = 21$, there is one pair of variables exhibiting significant complementarity, namely the population *Fraction Catholic* and *Protestant* that also showed the strongest positive jointness in the benchmark case in section 4.1.

Allowing for relatively larger models *a priori*, reduces the degree of jointness for any particular combination of variables. This comes at the price of allowing for a richer set of control variables and interdependence within each regression model.

¹⁹Detailed results of the sensitivity analysis in this section are not shown in the paper, but are available on request.

4.3.2 Sensitivity of Jointness to Added Interaction Terms

Next, we analyze the effect of adding additional interaction terms as regressors for variables that exhibit significant positive jointness (complementarity) in the benchmark case shown in Table 1. In particular, we add two interaction terms, one for (*Malaria Prevalence* \times *Spanish Colony*), and the other for (*Fraction Protestant* \times *Fraction Catholic*). The number of candidate regressors increases therefore to 69 and we leave the prior model size unchanged at the benchmark value $\bar{k} = 7$.

The two new interaction terms do not have significant jointness with others variable in the original set, except for the marginal negative jointness (-0.99) of the (*Malaria Prevalence* \times *Spanish Colony*) interaction term with the *Sub-Saharan Africa* dummy. As expected, the jointness between the originally significant complements *Malaria Prevalence* and *Spanish Colony* is weakened, from 1.15 in the benchmark case to 0.80 when the interaction term is added to the regressors. Surprisingly, the already significant positive jointness between the population *Fractions Protestant and Catholic* is strengthened from 2.05 for the benchmark case to 2.25 when the interaction terms are allowed for.

For the remaining variables, adding the two interaction terms as candidate regressors acts much like a reduction in prior model size, i.e. we observe a slightly higher incidence of negative jointness. For some variables, e.g. some geographic variables, positive dependence is also strengthened with the added possibility of interaction terms.

4.3.3 Sensitivity to AIC Posterior Model Weights

Finally, we consider the effect on jointness of allowing for alternative posterior model weights. In particular, the *AIC* model weights developed by Akaike (1973) have been suggested as alternatives in the model averaging literature. The *AIC* model selection criterion is given by $AIC \equiv l(\mathbf{D}|M_j) - k_j$, where $l(\mathbf{D}|M_j)$ is model-specific marginal likelihood and k_j is the number of regressors. Exponentiating, normalizing by the sum over all 2^K models and ignoring constants gives the *AIC* model weights

$$p^{AIC}(M_j|\mathbf{D}) = \frac{p(M_j) \cdot e^{-k_j} \cdot SSE_j^{-T/2}}{\sum_{i=1}^{2^K} p(M_i) \cdot e^{-k_j} \cdot SSE_i^{-T/2}} \quad (18)$$

The difference between the *AIC* model weights (18) and the posterior model weights (5) used is the benchmark case is the degrees of freedom penalty term, e^{-k_j} instead of $T^{-k_j/2}$. Given $T = 88$ observations in this application, the *AIC* weights penalize larger models less severely than the benchmark model weights.

The effect of using the *AIC* model weights as alternative is therefore similar to allowing for relatively larger prior model sizes than the benchmark case, i.e. $\bar{k} > 7$. Jointness among regressors is generally less pronounced, with the exception

of stronger negative jointness among the two government variables, the *Government Consumption Share* and overall *Government Share of GDP*.

In summary, we find that jointness is sensitive to changes in the prior model size and that dependence among explanatory variables is reduced in larger models. However, this result is not surprising given the richer set of conditioning set of variables permitted in larger models. We also confirm this intuition when using a version of *AIC* model weights that penalize inclusion of additional regressors less severely than the benchmark posterior model weights.

5 Conclusion

In this paper, we propose a new measure of mutual dependence or *jointness* among variables in explaining the dependent variable. Jointness differs from existing approaches in two respects: First, jointness among regressors is calculated unconditionally by averaging across models, thereby fully addressing model uncertainty. Second, jointness emphasizes dependence among explanatory variables in the posterior distribution and does not introduce any prior, possibly misleading dependence among regressors. Positive values of jointness imply that variables are complements, representing distinct, but interacting economic factors. Regressors exhibiting negative jointness are substitutes and measure similar underlying mechanisms.

We estimate jointness among determinants of economic growth, using data from SDM (2004). We find evidence of significant negative jointness among some relatively important growth determinants, indicating that these variables are substitutes in explaining cross-country economic growth. Our results suggest that economists interested in empirical growth research should reduce the weight on models that jointly include such substitutable variables.

We find some evidence of complementary relationships in the form of significant positive jointness among growth determinants. In contrast to negative jointness, significant positive jointness is also present among some variables which would be labeled “insignificant” by univariate measures of variable importance. These variables require a richer conditioning set of complementary regressors in explaining growth. In contrast to negative jointness, complementary variables showing positive jointness reinforce the size and significance of their mutual effect on economic growth.

Compared to the possibly very large number of dependencies among growth determinants, we find a relatively small number of significant positive jointness or complementarity. This implies, that policy decisions and inference are not too complex, even when taking jointness into account. Our jointness results can inform a policymaker about potentially important heterogeneity of effects and interdependence between policy instruments and control variables.

References

- [1] Akaike H. 1973. Information Theory and an Extension of the Maximum Likelihood Principle. In *Second International Symposium on Information Theory*, B. Petrov and F. Csake (eds). Akademiai Kiado: Budapest.
- [2] Brock W, Durlauf S. 2001. Growth Empirics and Reality. *World Bank Economic Review* 15(2): 229-72.
- [3] Brock W, Durlauf S, West K. 2003. Policy Evaluation in Uncertain Economic Environments (with Comments and Discussion.) *Brookings Papers of Economic Activity*: 235-322.
- [4] Campos J, Ericsson N. 2000. Constructive Data Mining: Modeling Consumers' Expenditure in Venezuela. *Board of Governors of the Federal Reserve System, International Finance Discussion Papers No. 663*.
- [5] Chipman H, George E, McCulloch R. 2001. The Practical Implementation of Bayesian Model Selection. In *Model Selection*, Lahiri P. (ed). Institute of Mathematical Statistics Lecture Notes-Monograph Series Vol. 38: Beachwood, Ohio.
- [6] Durlauf S, Quah D. 1999. The New Empirics of Economic Growth. In *Handbook of Macroeconomics*, Vol. 1a, Taylor J, Woodford M (eds). Elsevier: North Holland.
- [7] Clyde M, Desimone H, Parmigiani G. 1996. Prediction via Orthogonalized Model Mixing. *Journal of the American Statistical Association* 91(435): 1197-208.
- [8] Chipman H. 1996. Bayesian Variable Selection with Related Predictors. *The Canadian Journal of Statistics* 24: 17-36.
- [9] Fernandez C, Ley E, Steel M. 2001. Benchmark Priors for Bayesian Model Averaging. *Journal of Econometrics* 100(2): 381-427.
- [10] Gallup J, Sachs J. 2001. The Economic Burden of Malaria. *The American Journal of Tropical Medicine and Hygiene*, vol. 64 (1, 2) S: 85-96.
- [11] George E. 1999. Discussion of Bayesian Model Averaging and Model Search Strategies by M.A. Clyde. *Bayesian Statistics 6*: 175-77. Oxford.
- [12] George E. 2001. Dilution Priors for Model Uncertainty. *Notes for MSRI Workshop on Nonlinear Estimation and Classification*. Berkeley.
- [13] George E, McCulloch R. 1993. Variable Selection Via Gibbs Sampling. *Journal of the American Statistical Association* 88(423): 881-89.

- [14] Godsill S, Stone M, Weeks M. 2004. Assessing the Impact of Private sector Balance Sheets on Financial Crises: A Comparison of Bayesian and Information-Theoretic Measures of Model Uncertainty. *Working Paper*. University of Cambridge.
- [15] Granger C, King M, White H. 1995. Comments on Testing Economic Theories and the Use of Model Selection Criteria. *Journal of Econometrics* 67: 173-187.
- [16] Hendry D. 2000. *Econometrics: Alchemy or Science?* Oxford University Press: London.
- [17] Hjort N, and Claeskens G. 2003. Frequentist Model Average Estimators. *Journal of the American Statistical Association* 98(464): 879-99.
- [18] Hoeting J, Madigan D, Raftery A, Volinsky C. 1999. Bayesian Model Averaging: A Tutorial. *Statistical Science* 14: 382-417.
- [19] Hoover K and Perez S. 1999. Data Mining Reconsidered: Encompassing and the General-to-Specific Approach to Specification Search. *Econometrics Journal* 2, 1-25.
- [20] Jeffreys H. 1961. *Theory of Probability*. Third edition. Oxford University Press: London.
- [21] Judge G, Griffiths W, Hill R, Lutkepohl H, Lee TC. 1985. *The Theory and Practice of Econometrics*. Second edition. Wiley.
- [22] Kass R, Raftery A. 1995. Bayes Factors. *Journal of the American Statistical Association* 90(430): 773-95.
- [23] Krolzig H, Hendry D. 2000. Computer Automation of General-to-Specific Model Selection Procedures. *Working Paper, Institute of Economics and Statistics*. Oxford.
- [24] Leamer E. 1973. Multicollinearity: A Bayesian Interpretation. *Review of Economics and Statistics* 55(3): 371-80.
- [25] Leamer E. 1983. Model Choice and Specification Analysis. In *Handbook of Econometrics*, Vol. 1, Griliches Z, Intriligator M. (eds). Elsevier: North-Holland.
- [26] Leeb H, Pötscher B. 2005. Model Selection and Inference: Fact and Fiction. *Econometric Theory* 21: 21-59.
- [27] Mitchell T, Beauchamp J. 1988. Bayesian Variable Selection in Linear Models. *Journal of the American Statistical Association* 83(404): 1023-32.

- [28] Perez-Amaral T, Gallo G, White H. 2003. A Flexible Tool for Model Building: the Relevant Transformation of the Inputs Network Approach (RETINA). *Oxford Bulletin of Economics and Statistics* 65, 821-38.
- [29] Pötscher B. 1991. Effects of Model selection on Inference. *Econometric Theory* 7: 163-85.
- [30] Raftery A. 1995. Bayesian Model Selection in Social Research. *Sociological Methodology* 25, 111-63.
- [31] Sala-i-Martin X, Doppelhofer G, Miller RI. 2004. Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach. *American Economic Review* 94(4): 813-35.
- [32] Stock J, Watson M. 1999. Forecasting Inflation. *Journal of Monetary Economics* 44: 293-35.

A Data Appendix

Rank	Short Name	Variable Description	$p(\gamma_i \mathbf{D})$	Mean	S.D.
Dep. Var.	GROWTH	Growth of GDP per capita at PPP between 1960–1996.	–	0.0182	0.019
1	EAST	East Asian Dummy	0.82	0.11364	0.31919
2	P60	Primary Schooling Enrollment	0.80	0.72614	0.29321
3	IPRICE1	Investment Price	0.77	92.47	53.68
4	GDPCH60L	Log GDP in 1960	0.68	7.35494	0.90108
5	TROPICAR	Fraction of Tropical Area	0.56	0.57024	0.47160
6	DENS65C	Population Coastal Density	0.43	146.87	509.83
7	MALFAL66	Malaria Prevalence	0.25	0.33943	0.43089
8	LIFE060	Life Expectancy	0.21	53.72	12.06
9	CONFUC	Fraction Confucian	0.21	0.01557	0.07932
10	SAFRICA	Sub-Saharan Africa Dummy	0.15	0.30682	0.46382
11	LAAM	Latin American Dummy	0.15	0.22727	0.42147
12	MINING	Fraction GDP in Mining	0.12	0.05068	0.07694
13	SPAIN	Spanish Colony Dummy	0.12	0.17045	0.37819
14	YRSOPEN	Years Open 1950-94	0.12	0.35545	0.34445
15	MUSLIM00	Fraction Muslim	0.11	0.14935	0.29616
16	BUDDHA	Fraction Buddhist	0.11	0.04659	0.16760
17	AVELF	Ethnolinguistic Fractionalization	0.10	0.34761	0.30163
18	GVR61	Gov't Consumption Share	0.10	0.11610	0.07454
19	DENS60	Population Density	0.09	108.07	201.44
20	RERD	Real Exchange Rate Distortions	0.08	125.03	41.71
21	OTHFRAC	Fraction Speaking Foreign Language	0.08	0.32092	0.41363
22	OPENDEC1	Openness Measure 1965-74	0.08	0.52307	0.33591
23	PRIGHTS	Political Rights	0.07	3.82250	1.99661
24	GOVSH61	Government Share of GDP	0.06	0.16636	0.07115
25	H60	Higher Education Enrollment	0.06	0.03761	0.05006
26	TROPPOP	Fraction Population In Tropics	0.06	0.29998	0.37311
27	PRIEXP70	Primary Exports	0.05	0.71988	0.28270
28	GGCFD3	Public Investment Share	0.05	0.05216	0.03882
29	PROT00	Fraction Protestant	0.05	0.13540	0.28506
30	HINDU00	Fraction Hindu	0.04	0.02794	0.12465
31	POP1560	Fraction Population Less than 15	0.04	0.39251	0.07488
32	AIRDIST	Air Distance to Big Cities	0.04	4324	2614
33	GOVNOM1	Nominal Government Share	0.04	0.14898	0.05843
34	ABSLATIT	Absolute Latitude	0.03	23.21	16.84
35	CATH00	Fraction Catholic	0.03	0.32826	0.41459
36	FERTLDC1	Fertility	0.03	1.56202	0.41928
37	EUROPE	European Dummy	0.03	0.21591	0.41381
38	SCOUT	Outward Orientation	0.03	0.39773	0.49223
39	COLONY	Colony Dummy	0.03	0.75000	0.43549
40	CIV72	Civil Liberties	0.03	0.50947	0.32593
41	REVCoup	Revolutions and Coups	0.03	0.18489	0.23223
42	BRIT	British Colony Dummy	0.03	0.31818	0.46844
43	LHCPC	Hydrocarbon Deposits	0.02	0.42115	4.35121
44	POP6560	Fraction Population Over 65	0.02	0.04881	0.02898
45	GDE1	Defense Spending Share	0.02	0.02589	0.02463
46	POP60	Population in 1960	0.02	20308	52538
47	TOT1DEC1	Terms of Trade Growth in 1960s	0.02	-0.00208	0.03455
48	GEEREC1	Public Education Spending Share	0.02	0.02441	0.00964
49	LANDLOCK	Landlocked Country Dummy	0.02	0.17045	0.37819
50	HERF00	Religion Measure	0.02	0.78032	0.19321
51	SIZE60	Size of Economy	0.02	16.15	1.82
52	SOCIALIST	Socialist Dummy	0.02	0.06818	0.25350
53	ENGFRAC	English Speaking Population	0.02	0.08398	0.25224
54	PI6090	Average Inflation 1960-90	0.02	13.13	14.99
55	OIL	Oil Producing Country Dummy	0.02	0.05682	0.23282
56	DPOP6090	Population Growth Rate 1960-90	0.02	0.02153	0.00946
57	NEWSTATE	Timing of Independence	0.02	1.01136	0.97667
58	LT100CR	Land Area Near Navigable Water	0.02	0.47216	0.38021
59	SQPI6090	Square of Inflation 1960-90	0.02	394.54	1119.70
60	WARTIME	Fraction Spent in War 1960-90	0.02	0.06955	0.15241
61	LANDAREA	Land Area	0.02	867189	1814688
62	ZTROPICS	Tropical Climate Zone	0.02	0.19002	0.26869
63	TOTIND	Terms of Trade Ranking	0.02	0.28127	0.19038
64	ECORG	Capitalism	0.02	3.46591	1.38089
65	ORTH00	Fraction Orthodox	0.02	0.01867	0.09829
66	WARTORN	War Participation 1960-90	0.02	0.39773	0.49223
67	DENS65I	Interior Density	0.02	43.37	88.06

Variables ranked by Posterior Inclusion Probability $P(\gamma_j|\mathbf{D})$. Prior inclusion probability for benchmark case equals $\bar{k}/67 = 7/67 = 0.10$.

Figure 1: Prior Probabilities by Model Size: Benchmark Case with Prior Model Size $\bar{k} = 7$ and Uniform Prior with $\bar{k} = 33$.

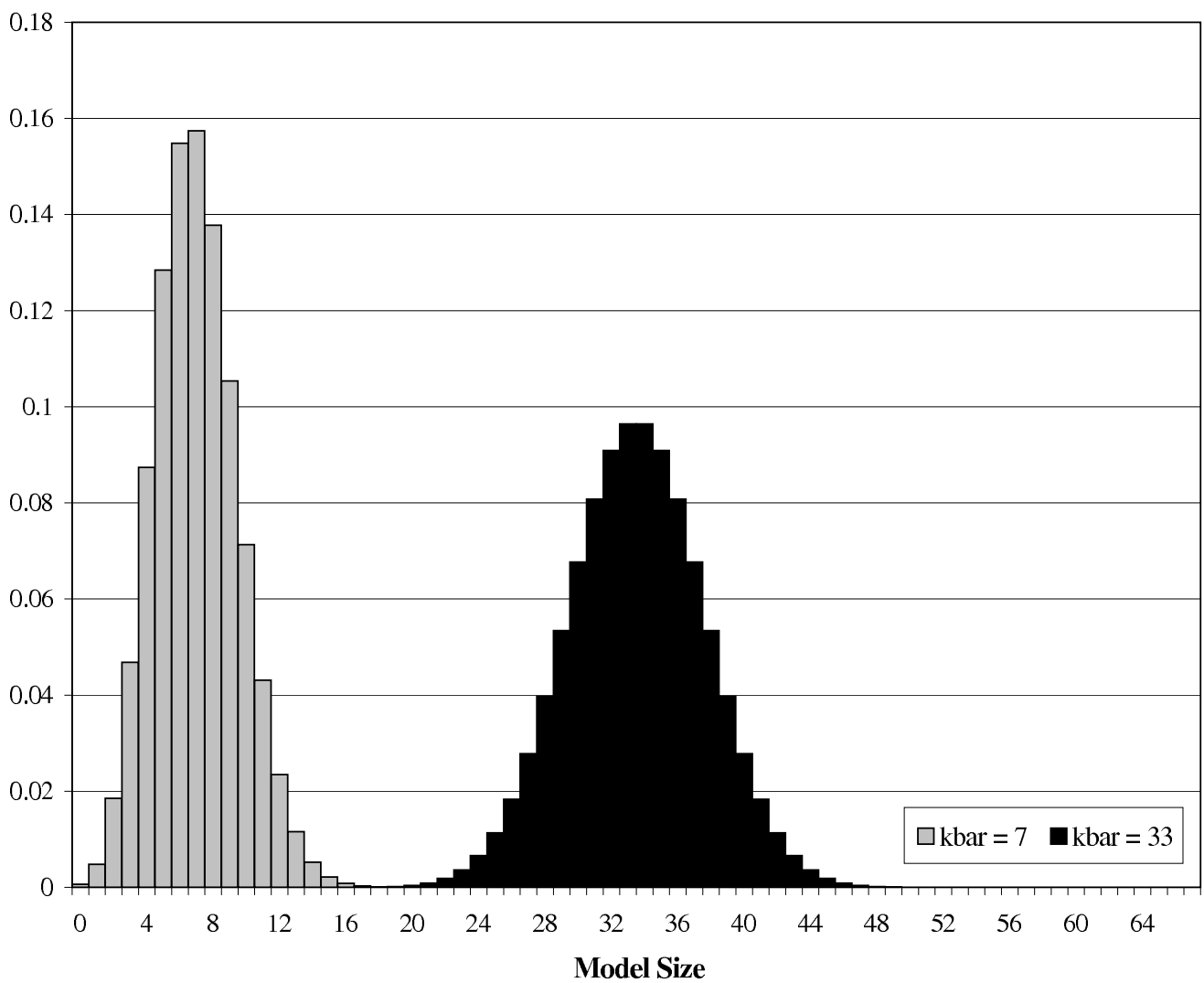


Figure 2: Jointness Among Selected Regressors

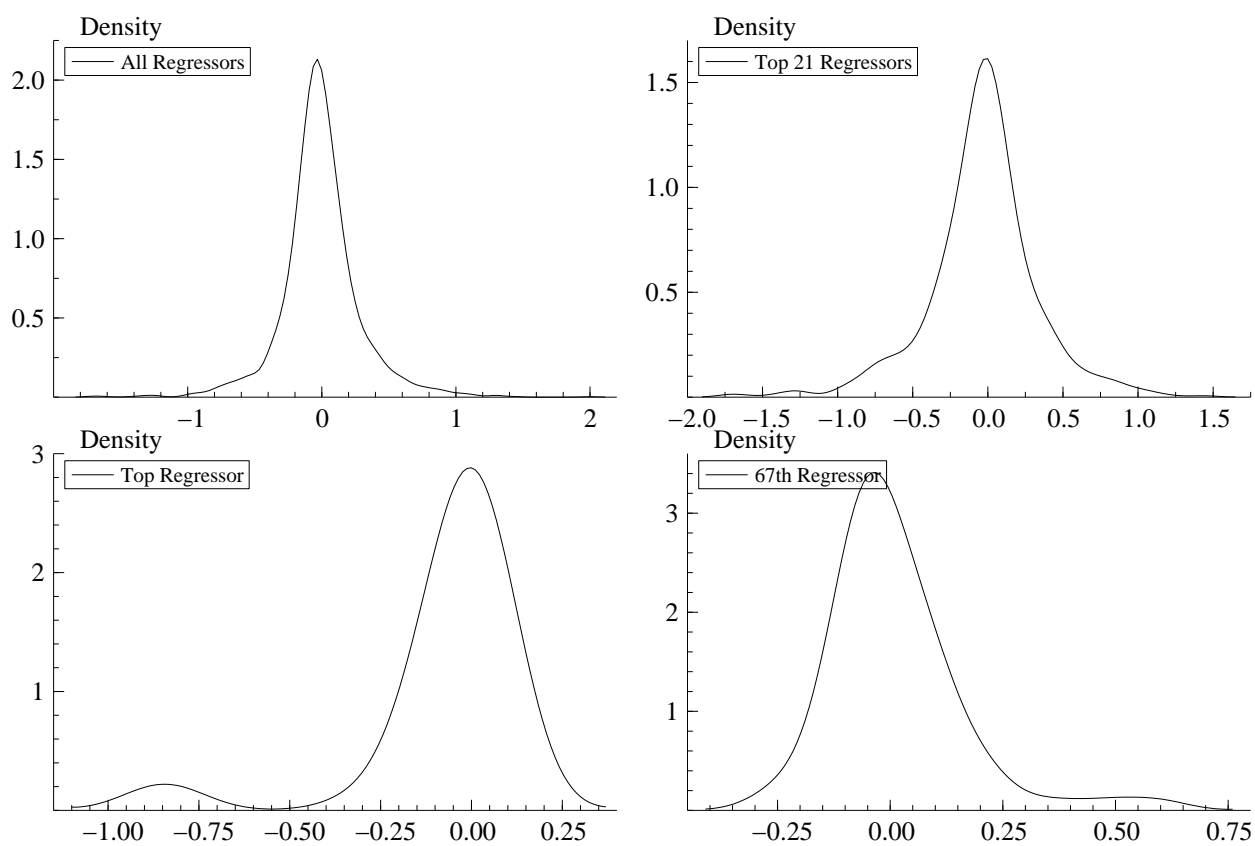


Figure 3: Robustness to Prior Model Size: Top 21 Regressors

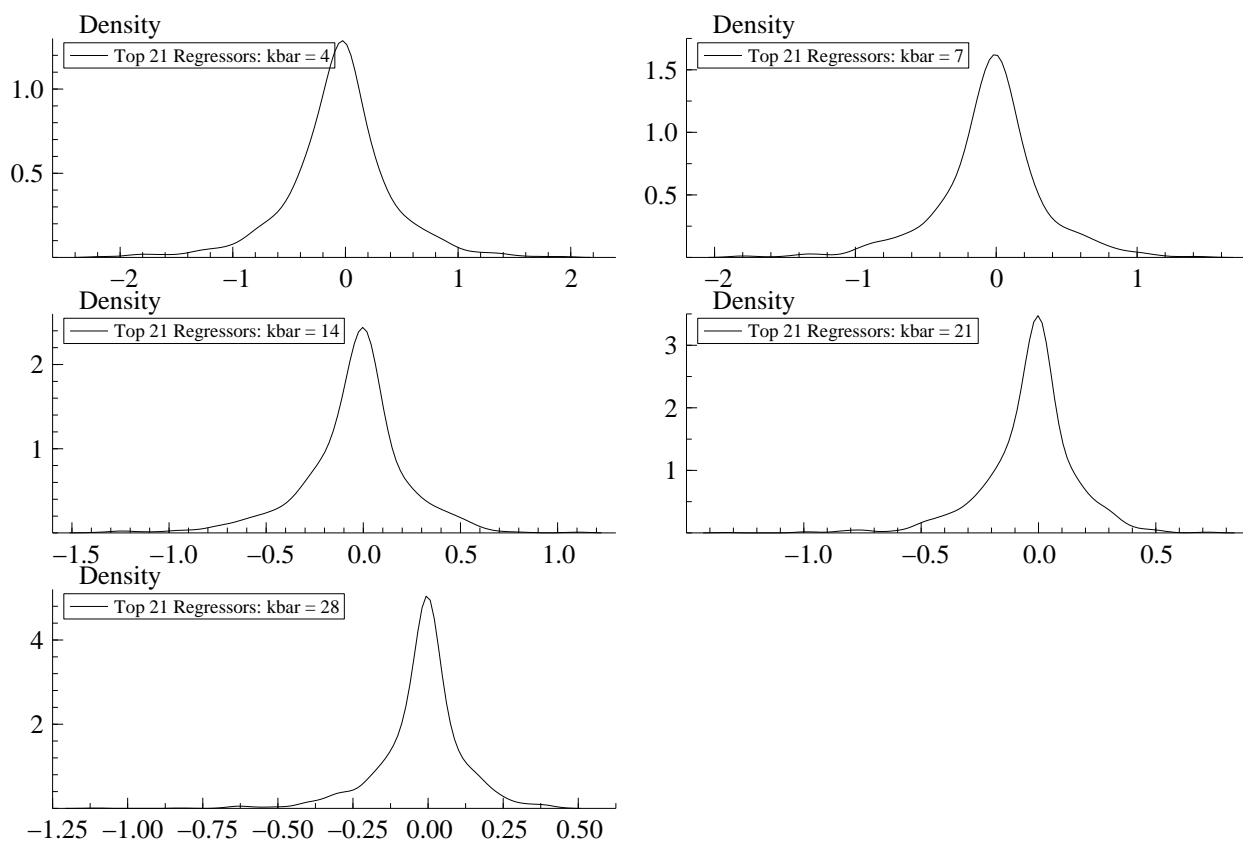


Figure 4: Robustness to Prior Model Size: Regressors Ranked 22-67

