Department of Economics
Working Paper No. 193

A Comprehensive Approach to Posterior Jointness Analysis in Bayesian Model Averaging Applications

Jesus Crespo Cuaresma
Bettina Grün
Paul Hofmarcher
Stefan Humer
Mathias Moser

March 2015

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

A Comprehensive Approach to Posterior Jointness Analysis in Bayesian Model Averaging Applications*

Jesus Crespo Cuaresma[a,b,c,d], Bettina Grün[e], Paul Hofmarcher[a], Stefan Humer[a], and Mathias Moser[a]

[a]Department of Economics, Vienna University of Economics and Business (WU)

[b]Austrian Institute of Economic Research (WIFO)

[c]International Institute of Applied System Analysis (IIASA)

[d]Wittgenstein Centre for Demography and Global Human Capital (WIC)

[e]Department of Applied Statistics, Johannes Kepler University Linz (JKU)

This version: March 2015

Abstract

Posterior analysis in Bayesian model averaging (BMA) applications often includes the assessment of measures of *jointness* (joint inclusion) across covariates. We link the discussion of jointness measures in the econometric literature to the literature on association rules in data mining exercises. We analyze a group of alternative jointness measures that include those proposed in the BMA literature and several others put forward in the field of data mining. The way these measures address the joint exclusion of covariates appears particularly important in terms of the conclusions that can be drawn from them. Using a dataset of economic growth determinants, we assess how the measurement of jointness in BMA can affect inference about the structure of bivariate inclusion patterns across covariates.

JEL Classification: C11, C55, O40.

Keywords: Bayesian Model Averaging, Jointness, Robust Growth Determinants, Machine Learning, Association Rules.

*Corresponding Author: Mathias Moser. Address: Welthandelsplatz 1, 1020 Vienna, Austria, Tel: +43(0)131336-5481, Email: matmoser@wu.ac.at.

1. Introduction

Addressing model uncertainty concerns in econometric applications through the use of Bayesian model averaging (BMA) is becoming a standard practice in empirical studies where no unique theoretical guidelines exist. One of such areas in economics where BMA has established itself as a useful tool of analysis is economic growth. A growing number of studies aims at identifying robust determinants of income per capita growth differences across countries without having to rely on specific theoretical frameworks (see for example Fernández et al., 2001a; Brock and Durlauf, 2001; Sala-i Martin et al., 2004; Moral-Benito, 2012; Eicher et al., 2012; Moral-Benito, 2014). In these studies, the robustness of individual covariates as determinants of income growth differences is routinely measured through *posterior inclusion probabilities* (PIP), i.e., the posterior probability covered by all models that contain that particular variable. This represents an average over a (possibly) large number of very different models.

Moving beyond the development of robustness measures based on individual covariates, some contributions in the literature aim at identifying particular structures in the posterior distribution of joint covariate inclusion. The literature tends to concentrate on the assessment of measures based on bivariate inclusion structures and uses the term *jointness* to refer to the dependence in the inclusion of groups (most often, pairs) of variables. Doppelhofer and Weeks (2005), Ley and Steel (2007, henceforth LS), Doppelhofer and Weeks (2009a, henceforth DW) and Strachan (2009) are the most relevant references dealing with measuring posterior inclusion dependence of regressors in economic growth applications. Using a different approach from these studies, Crespo Cuaresma et al. (2015) employ clustering methods to identify covariate inclusion patterns over the structure revealed by the posterior model probabilities of BMA exercises.

To quantify the association of covariate inclusion, the BMA literature has proposed several measures of *jointness*. These measures and the properties that define them have been studied in a strand of independent literature in the field of data mining, which aims at evaluating the quality of so-called association rules. A common example for such a problem in data mining is finding sets of products that tend to be purchased together in a shopping basket. The development of rules that define the inclusion patterns existing between two or more items is conceptually very similar to finding jointness structures for a given set of covariates in the model space after the posterior model probabilities have been computed. However, the choice of measures to quantify these associations has generated a vivid discussion in the machine learning literature. Several studies provide comparisons of a large number of concepts and try to identify suitable measures through the kind of properties they fulfill (Geng and Hamilton, 2006; Glass, 2013). Besides these attempts to select measures based on objective criteria, some authors also adopt a subjective approach, in which the researcher tries to quantify a priori expectations (Tan et al., 2004). Some studies also show that many of the proposed measures produce similar rankings and therefore can be used exchangeably in many applications (Vaillant et al., 2004; Tew et al., 2014).

The controversy around measuring jointness in BMA applications was born from the contributions by Ley and Steel (2009a), Strachan (2009) and Doppelhofer and Weeks (2009b). In their exchange of ideas the different authors raised concerns about how the different measures in the BMA context were defined. These discussions especially revolved around cases were several measures are undefined, or give contradictory results. Especially the question of whether the probability that two variables are not included in a model should influence the value of a jointness measure or not was debated vividly. We bring insights from the literature on association measures used in data mining and provide a thorough analysis of the differential characteristics of a larger set of jointness measures which nests those proposed hitherto in BMA applications. More specifically, we review properties of jointness measures, which have been proposed in the machine learning literature and focus on the property of null-invariance. We show that, while most measures in the BMA literature have this property, it is not favorable in BMA applications. Based on this discussion, we select a subset of measures that fulfill the afore discussed properties and use them to investigate jointness in the data set of Fernández et al. (2001a).

The paper is structured as follows. In section 2 we briefly review the standard implementation of jointness measures in the context of BMA. We present a short summary of relevant concepts from the literature on association rule analysis and how these are related to jointness in section 3. The empirical application based on the cross-country growth regression dataset in Fernández et al. (2001b), is carried out in section 4. Section 5 concludes and puts forward avenues of further research.

2. BMA and Jointness Measures: A Review

BMA methods aim at obtaining posterior distributions of the quantities of interest in a regression model which incorporate the uncertainty concerning model specification. Let our quantity of interest be related to the parameters of a linear regression model of the form

$$y|\alpha, \beta_j, \sigma \sim N(\alpha\iota + X_j\beta_j, \sigma^2 I), \tag{1}$$

where y is an $n \times 1$ vector whose elements are the observations of the dependent variable of interest, ι a vector of ones of the same length and the $n \times k$ matrix X_j is composed by the observations of k variables out of a total set of K covariates. Model uncertainty can be explicitly addressed by basing our inference on the parameters of interest on the posterior distribution

$$p(\alpha, \beta, \sigma|y) = \sum_{j=1}^{2^K} p(\alpha, \beta, \sigma|y, M_j)p(M_j|y), \tag{2}$$

where each specification-specific posterior distribution $p(\alpha, \beta, \sigma|y, M_j)$ is weighted by the corresponding posterior model probability $p(M_j|y)$. The posterior model probability is in turn proportional to the marginal likelihood of the model multiplied with the prior model probability,

$$p(M_j|y) \propto p(y|M_j)p(M_j). \tag{3}$$

It is standard in BMA applications to elicit improper non-informative priors on α and σ, so that $p(\alpha) \propto 1$ and $p(\sigma) \propto \sigma^{-1}$. A common choice for the prior of the slope coefficients β is Zellner's g–prior (Zellner, 1986),

$$p(\beta|M_j, \sigma) \sim N(0, \sigma^2(\frac{1}{g_0}X_j'X_j)^{-1}) , \quad (4)$$

so that the prior variance matrix is scaled by the parameter g_0 and has the structure of the covariance matrix of the OLS estimator. Several fixed values for the g parameter have been proposed (see e.g. Foster and George, 1994; Fernández et al., 2001b). To allow for more flexibility, hyperpriors on g have also been put forward in the literature by Liang et al. (2008); Feldkircher and Zeugner (2009); Ley and Steel (2012).

For the prior model probabilities, a straightforward approach is to elicit a flat prior over all specifications entertained, so that $p(M_j) = 2^{-K}$ for all j. Given that this prior embodies a preference for models of size around $K/2$, Ley and Steel (2009b) argue for a binomial-beta prior on covariate inclusion, a setting which is able to achieve a very flexible prior structure over model size and includes a purely uninformative distribution over the number of included covariates.

Since analyzing the whole model space of 2^K models is often computationally infeasible, the relevant parts of the model space can be explored via Markov Chain Monte Carlo Model Composition (MC^3) methods (Madigan and York, 1995) in order to compute the relevant posterior distributions.

Among the many interesting features of the posterior over model specifications, the joint distribution of covariate inclusion constitutes the basis to create measures of jointness. Following Doppelhofer and Weeks (2009a), let model specifications be represented by a 0-1 vector of covariate inclusion profiles (as defined by the inclusion variables γ_k, $k = 1, \ldots, K$), so that

$$p(M_j|y) = p(\gamma_1 = c_1, \gamma_2 = c_2, ..., \gamma_K = c_K|y) , \quad (5)$$

where c_k is the binary variable representing the inclusion of covariate k in the model. Given these inclusion profiles, jointness quantifies to which degree two variables A and B tend to appear jointly across models $\big(p(A \cap B|y) \equiv p(AB|y)\big)$ as opposed to the posterior probability to appear without the respective other variable $\big(p(A \cap \bar{B}|y) \equiv p(A\bar{B}|y)$ and $p(\bar{A} \cap B|y) \equiv p(\bar{A}B|y)\big)$.

The comparison of these probabilities allows to consider two covariates as complements, substitutes or independent a posteriori, given their relative (common) appearance. The group of jointness measures that have been proposed in the BMA context uses these probabilities to generate a single statistic which allows a categorization of such pairs (or eventually, triplets) of variables. Positive values for these indicators typically refer to joint appearance (and therefore a certain degree of complementarity between them), while negative values are related to the fact that the two covariates act as substitutes in specifications. So far, five different measures of jointness have been proposed in the econometric literature dealing with BMA, which differ in the way they incorporate the different marginal and joint inclusion probabilities.

The earliest jointness measure in the BMA context is attributed to Doppelhofer and Weeks

(2005), who propose to use

$$J = \ln\left(\frac{p(AB)}{p(A) \times p(B)}\right), \tag{6}$$

which resembles the logarithm of the posterior odds ratio. The use of posterior odds ratios as jointness indicator was criticized by Ley and Steel (2007), who note that the measure may be misleading for variables with high PIP and that the measure hardly allows for comparisons across different pairs of variables.

In a later study Doppelhofer and Weeks (2009a) propose a cross-product ratio of inclusion probabilities as another measure,

$$\mathcal{J} = \ln\left(\frac{p(AB) \times p(\bar{A}\bar{B})}{p(A\bar{B}) \times p(\bar{A}B)}\right). \tag{7}$$

In a reply Ley and Steel (2009a) are again not in favor of this approach, since the DW measure is not defined in cases where a variable has a PIP of 1 or 0. Instead LS highlight two alternative measures (Ley and Steel, 2007):

$$\mathcal{J}^* = \frac{p(AB)}{p(A) + p(B) - p(AB)} \tag{8}$$

$$\mathcal{J}' = \frac{p(AB)}{p(A\bar{B}) + p(\bar{A}B)} \quad . \tag{9}$$

While $\mathcal{J}'$ relates the joint inclusion to the probability of including either one of the two variables, $\mathcal{J}^*$ uses the probability of including either one but not both variables in the denominator.

Another measure was introduced by Strachan (2009), who proposes to only look at relevant variables in terms of PIP. This is accomplished by adapting DW's cross-product ratio in such a way, that it includes the marginal probabilities of both variables,

$$\tilde{\mathcal{J}} = p(A)p(B) \ln\left(\frac{p(AB)}{p(A\bar{B}) \times p(\bar{A}B)}\right). \tag{10}$$

A major discussion in the jointness literature also involves the treatment of $p(\bar{A} \cap \bar{B}|y) \equiv p(\bar{A}\bar{B}|y)$. This *exclusion margin* indicates to which extent both variables do not tend to appear together in specifications and therefore may be considered as representing a measure of (un)importance of bivariate jointness. While DW stress the importance of this aspect in the discussion (Doppelhofer and Weeks, 2009b), this property is not included in the jointness measures proposed by Ley and Steel (2009a). The treatment of the information concerning joint exclusion of covariates constitutes a differential characteristic across association measures known as *null-invariance* in the data mining literature (Glass, 2013).

3. From Association Rules to Jointness Measures

The measures used in the literature on jointness of covariates in BMA analysis are often applied in data mining when describing association rules, although the linkages between the two strands of literature has not been explicitly acknowledged hitherto. Data mining is often concerned with the exploration of huge datasets of so-called transactions, which may for example each represent shopping baskets with different sets of items (products). Association analysis aims at finding patterns in these data structures to learn about consumer behavior and the interrelation across purchased items. The major tool used are association rules of the form $A \longrightarrow B$ (if A is included in the basket, B tends to be included), where A and B can include either individual items or disjoint itemsets.

For a large number of items, the count of rules can potentially grow very large. The number of itemsets is $2^K - 1$ for K items (variables) which implies $3^K - 2^{K+1} + 1$ possible association rules (excluding empty sets) between itemsets of all sizes. Therefore association rules are routinely *mined* to only include such rules which are "interesting" for the application. This refers on the one hand to associations which are frequent, as measured by the *support*. On the other hand, rules should be strong as measured by the *confidence*, which relates the occurrence of a pattern to the number of counterexamples in the data.

The most common strategy to extract such rules is the *apriori algorithm* (Aggarwal and Yu, 1998; Hahsler et al., 2005), which reduces the complexity of the problem by reasoning that all item subsets of a frequent itemset must also be frequent and vice versa. This approach is also related to *support-based pruning* and has been applied by a large number of studies in the data mining literature (Tan et al., 2004).

In addition to support and confidence — which are relevant to achieve computational feasibility — the interestingness of these rules can be quantified using several measures. Similar to the jointness literature, a number of such indicators has been proposed in the data mining context. Recent surveys in this field collect as many as 40 different measures and try to provide a structural overview of the alternative measures available (Glass, 2013; Geng and Hamilton, 2006; Tan et al., 2004).[1] Some of these measures resemble the ones proposed in the BMA jointness literature. The first jointness measure of Doppelhofer and Weeks (2005) is equivalent to the Log-Ratio or equivalently, the log of the Interest (Lift) measure (Geng and Hamilton, 2006). Ley and Steel (2007)'s $\mathcal{J}^*$ is identical to the long-used Jaccard index and their $\mathcal{J}'$ measure is a derivation thereof. As another alternative, Strachan (2009) introduces a measure ($\tilde{\mathcal{J}}$) that has been known as the Two-Way support (Geng and Hamilton, 2006). Finally, the statistic proposed by Doppelhofer and Weeks (2009a) has also been known as the Odds-Ratio in the field of data mining (Tew et al., 2014).

Another similarity between the two strands of literature is the debate on which measure is the most appropriate for a given application. Tan et al. (2004) propose the use of subjective measures, which depend on the user to rank a small predefined set of associations for a specific

[1] A detailed overview of interestingness measures can be found in Appendix A.1.

application. Using this approach, an appropriate measure can be selected, which reproduces the user's ranking. More generally, objective measures have been analyzed based on certain properties they are expected to fulfill. Ley and Steel (2007) propose four properties, that BMA jointness measures should fulfill: An indicator should be *interpretable* in such a sense, that it has a "clear intuitive meaning" and is well *calibrated* against a clearly defined scale. Furthermore, the property of *extreme jointness* states that a measure should reach its maximum when both variables always appear together. Also, a measure should always be defined (*definition*) when either variable is included with positive probability. In contrast, the association analysis literature tends to impose a larger number of characteristics that are expected to be fulfilled. In the following section we shortly review the most important properties proposed in the literature, discuss their implications for jointness and relate them to the measures in the BMA literature where applicable.

3.1. Desirable Properties of Interestingness Measures for BMA

The properties that have been independently discussed in the BMA context, partly reflect those which are used in the machine learning (ML) literature. Finding a suitable measure clearly depends on the properties that are required for a certain application. For example, while machine learning problems are often concerned with positive association, BMA results additionally need to reflect negative association in the form of variable substitutes. Furthermore the type of assertion that is being made, needs to be considered and especially the question whether two variables are considered exchangeable, so that $A \to B \equiv B \to A$. In the following we select four properties, which can be considered crucially relevant for jointness based on the insights from the BMA and ML discussions: Confirmation, symmetry, monotonicity and null-invariance. A symmetric, non null-invariant confirmation measure fulfills Ley and Steel (2007)'s property of *interpretability*, while the monotonicity requirement ensures that the *extreme jointness* condition is met. Additionally *calibration* is implicitly given for all measures. Additionally Ley and Steel (2007) argue that a measure should always be *defined*, regardless of the observed cases, this is not always the case for the measures considered here.

Interestingness vs. Confirmation A confirmation measure is an interestingness measure m that, for a given threshold τ, satisfies that

$$\begin{aligned}
m(A,B) > \tau &\iff Pr(A|B) > Pr(A),\\
m(A,B) = \tau &\iff Pr(A|B) = Pr(A),\\
m(A,B) < \tau &\iff Pr(A|B) < Pr(A).
\end{aligned}$$

The indicator is thus anchored at some threshold value τ that defines statistical independence (e.g. 0 for DW's Odds-Ratio). For the case of jointness indicators discussed in the BMA literature, this property is implicitly given for all proposed measures and seems to be a reasonable characteristic to be fulfilled. We therefore limit our empirical analysis to the set of confirmation measures that have been proposed in the data mining context (Glass, 2013).

Symmetry Implication rules that imply that the proposition $A \to B$ differs from $B \to A$ are asymmetric. Since jointness measures are interested in measuring the common appearance (or lack thereof) of two explanatory variables, a suitable measure should therefore be symmetric with regard to the ordering of variables. The assertion that certain covariates are "substitutes" or "complements" implies thus commutativity.[2] All jointness measures proposed in the BMA literature fulfill this requirement. A number of measures from the data mining literature are however asymmetric and thus excluded from the empirical analysis carried out in the following sections.[3]

Monotonicity and Maximality The range of interestingness measures should be bounded and monotonically increasing between the two extreme cases. This property is partly reflected in the more restrictive Piatesky-Shapiro conditions: $m = 0$ if $p(AB) = 0$, m monotonically increases with $p(AB)$ and m monotonically decreases with $p(A)$ or $p(B)$ (Piatetsky-Shapiro, 1991; Tan et al., 2004). Maximality corresponds to *extreme jointness*, the property introduced by Ley and Steel (2007) in the jointness literature. This property defines that a measure should reach its maximum when both variables always appear together.

Table 1: Interestingness Measures for Jointness

	Value	Range	k
Non null-invariant			
Collective Strength	$\ln\left[\frac{p(AB)+p(\bar{A}\bar{B})}{p(A)p(B)+p(\bar{A})p(\bar{B})} \times \frac{1-p(A)p(B)-p(\bar{A})p(\bar{B})}{1-p(AB)-p(\bar{A}\bar{B})}\right]$	$]-\infty,\infty[$	
Relative Risk	$\ln\left[\frac{p(B\|A)}{p(B\|\bar{A})}\right]$	$]-\infty,\infty[$	
Yule's Q	$\frac{p(AB)p(\bar{A}\bar{B})-p(A\bar{B})p(\bar{A}B)}{p(AB)p(\bar{A}\bar{B})+p(A\bar{B})p(\bar{A}B)}$	$[-1,1]$	
Normalized Difference	$p(B\|A)-p(B\|\bar{A})$	$[-1,1]$	
ϕ-Coefficient	$\frac{p(AB)-p(A)p(B)}{\sqrt{p(A)p(B)p(\bar{A})p(\bar{B})}}$	$[-1,1]$	
Null-invariant			
AllConf	$\min(p(B\|A), p(A\|B))$	$[0,1]$	$-\infty$
Coherence	$(p(A\|B)^{-1}+p(B\|A)^{-1}-1)^{-1}$	$[0,1]$	-1
Cosine	$\frac{p(AB)}{\sqrt{p(A)p(B)}}$	$[0,1]$	0
Kulczynski	$(p(A\|B)+p(B\|A))/2$	$[0,1]$	1
MaxConf	$\max(p(B\|A), p(A\|B))$	$[0,1]$	$+\infty$

[2]This property is often called commutative symmetry (Glass, 2013).
[3]Tan et al. (2004) suggest to symmetrize measures by using $\max(p(A|B), p(B|A))$.

Null-invariance Measures that are null-invariant ignore so-called null transactions, in which neither A nor B occur. Whether null-invariance is a desirable property for an association measure depends on the nature of the empirical application under scrutiny. For the case of jointness measures in BMA analysis, different views concerning the desirability of null-invariance have been voiced in the literature. Doppelhofer and Weeks (2009b) criticize null-invariance, since "*[...] jointness can manifest itself in both the inclusion and exclusion margin of the joint posterior distribution*". In contrast, Strachan (2009) and Ley and Steel (2009a) stress the effect of low-probability models, which are represented only sparsely in the model matrix and which would be "uninteresting" for most non null-invariant measures where the common exclusion probability is respected.

3.2. Confirmation Measures for Jointness Analysis

Based on the extensive surveys of interestingness measures in the data mining literature (Tan et al., 2004; Geng and Hamilton, 2006; Glass, 2013; Tew et al., 2014), we select a subset of indicators which fulfill the properties put forward above and that are therefore potentially suitable to analyze jointness in BMA applications. More specifically, all interestingness measures analyzed here are (a) confirmation measures, (b) symmetric around a threshold that implies inclusion independence and (c) reach their maxima when both variables are highly complementary. We group these measures by whether they fulfill null-invariance or not. Table 1 provides an overview of these indicators.[4]

Table 2: Comparison of Interestingness Measures: Independency

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Probabilities								
$p(A)$	0.10	0.50	0.90	0.70	0.50	0.60	0.50	0.90
$p(B)$	0.10	0.10	0.10	0.20	0.50	0.40	0.90	0.90
$p(A\|B)$	0.10	0.50	0.90	0.70	0.50	0.60	0.50	0.90
$p(B\|A)$	0.10	0.10	0.10	0.20	0.50	0.40	0.90	0.90
$p(AB)$	0.01	0.05	0.09	0.14	0.25	0.24	0.45	0.81
$p(A\bar{B})$	0.09	0.05	0.01	0.06	0.25	0.16	0.45	0.09
$p(\bar{A}B)$	0.09	0.45	0.81	0.56	0.25	0.36	0.05	0.09
$p(\bar{A}\bar{B})$	0.81	0.45	0.09	0.24	0.25	0.24	0.05	0.01
Non null-invariant								
Collective Strength	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Relative Risk	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Yule's Q	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Normalized Difference	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ϕ-Coefficient	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Null-invariant								
AllConf	0.10	0.10	0.10	0.20	0.50	0.40	0.50	0.90
Coherence	0.05	0.09	0.10	0.18	0.33	0.32	0.47	0.82
Cosine	0.10	0.22	0.30	0.37	0.50	0.49	0.67	0.90
Kulczynski	0.10	0.30	0.50	0.45	0.50	0.50	0.70	0.90
MaxConf	0.10	0.50	0.90	0.70	0.50	0.60	0.90	0.90

Note: Independency defined as $p(AB) = p(A)p(B)$

[4] A full list of the interestingness measures used in the literature and that have been considered to select the particular indicators considered here is presented in Appendix A.1.

This choice of measures subsumes all the indicators used in the BMA jointness literature, while we adhere to the naming conventions used in data mining. We replace the *Odds Ratio* with its projection on the $[-1, 1]$ interval, which is known as *Yule's Q*.[5] The *Collective Strength* measure was introduced by Aggarwal and Yu (1998) and compares the violation rate of an itemset to its expected value under statistical independence. It is defined between zero and ∞, where a value of unity signals statistical independence, a lower value indicates substitutability and a larger value complementarity. We use the log transformed measure which is defined around 0 as the independence threshold. *Relative Risk* is a measure widely used in case studies, where an exposed group (numerator) is compared to a non-exposed group (denominator). Log-transforming this measure, we define independence at a value of zero and substitutes (complements) below (above) this value. *Normalized Difference* is simply the difference between two probabilities and hence defined in $[-1, 1]$. The ϕ*-Coefficient* is basically a correlation measure and closely related to the χ^2 statistic, bounded in the interval $[-1, 1]$. We do not include the non null-invariant measure of Strachan (2009), known as *Two-Way-Support*, since it does not fulfil the monotonicity requirement (Glass, 2013).

Table 3: Comparison of Interestingness Measures: Complementarity

	Substitutes					**Complements**				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Probabilities										
$p(A)$	0.10	0.90	0.70	0.50	0.70	0.10	0.50	0.90	0.30	0.40
$p(B)$	0.10	0.10	0.20	0.50	0.30	0.10	0.40	0.10	0.30	0.20
$\boldsymbol{p(A\|B)}$	**0.01**	**0.09**	**0.07**	**0.05**	**0.07**	**0.90**	**1.00**	**1.00**	**1.00**	**1.00**
$p(B\|A)$	0.01	0.01	0.02	0.05	0.03	0.90	0.80	0.11	1.00	0.50
$p(AB)$	0.00	0.01	0.01	0.02	0.02	0.09	0.40	0.10	0.30	0.20
$p(A\bar{B})$	0.10	0.09	0.19	0.48	0.28	0.01	0.00	0.00	0.00	0.00
$p(\bar{A}B)$	0.10	0.89	0.69	0.48	0.68	0.01	0.10	0.80	0.00	0.20
$p(\bar{A}\bar{B})$	0.80	0.01	0.11	0.03	0.02	0.89	0.50	0.10	0.70	0.60
Non null-invariant										
Collective Strength	-0.12	-2.48	-1.43	-2.94	-2.80	2.38	2.20	0.13	Inf	1.15
Relative Risk	-2.40	-4.51	-3.43	-2.94	-3.43	4.39	Inf	Inf	Inf	Inf
Yule's Q	-0.85	-1.00	-0.98	-0.99	-1.00	1.00	1.00	1.00	1.00	1.00
Normalized Difference	-0.10	-0.90	-0.60	-0.90	-0.90	0.89	0.80	0.11	1.00	0.50
ϕ-Coefficient	-0.10	-0.90	-0.69	-0.90	-0.90	0.89	0.82	0.11	1.00	0.61
Null-invariant										
AllConf	0.01	0.01	0.02	0.05	0.03	0.90	0.80	0.11	1.00	0.50
Coherence	0.01	0.01	0.02	0.03	0.02	0.82	0.80	0.11	1.00	0.50
Cosine	0.01	0.03	0.04	0.05	0.05	0.90	0.89	0.33	1.00	0.71
Kulczynski	0.01	0.05	0.04	0.05	0.05	0.90	0.90	0.56	1.00	0.75
MaxConf	0.01	0.09	0.07	0.05	0.07	0.90	1.00	1.00	1.00	1.00

Notes: Substitutes defined as $p(AB) = 0.1 \times p(A)p(B)$
Complements defined as $p(AB) = \min(1, 9 \times p(A)p(B))$

As described by Wu et al. (2010), five common null-invariant measures can be represented by the generalized mean of the two conditional probabilities $p(A|B)$ an $p(B|A)$ with parameter k. This representation nests the *AllConf* measure (*Confidence*), *Coherence* (*Jaccard*, Ley and Steel 2007), *Cosine* (similar to Doppelhofer and Weeks, 2005), *Kulczynski* and *MaxConf*, which we

[5]The *Odds Ratio*, *Yule's Q*, and the log transformation of *Yule's Q*, *Yule's Y*, produce the same rankings of association rules and are therefore considered equivalent (Tew et al., 2014).

employ as examples of alternative null-invariant measures. These measures present themselves as differently weighted means, so that *Coherence* describes the harmonic mean, *Cosine* the geometric and *Kulczynski* the arithmetic mean of the two probabilities (Wu et al., 2010).

Based on the reasoning by Doppelhofer and Weeks (2009b) concerning the fact that a sensible jointness measure should equal zero for independence, we provide a synthetic example for different measures in Table 2. The eight columns provide scenarios where A and B are statistically independent, so that $p(AB) = p(A)p(B)$, but differ in the values for $P(A)$ and $P(B)$. Based on this assumption, we calculate the different jointness measures for each scenario. Column 5 depicts the scenario described in Doppelhofer and Weeks (2009b), which is the special case of $p(A) = p(B) = 0.5$. While Doppelhofer and Weeks (2009b) only argued based on an example with equal posterior inclusion probability across covariates ($p(A) = p(B)$), we also consider differing individual posterior probabilities of inclusion in Table 2.

As expected, the non null-invariant measures regard all eight scenarios presented in Table 2 as independent, since they explicitly take care of the exclusion margin $p(\bar{A}\bar{B})$. In contrast, the null-invariant measures only agree in terms of the absolute size of the indicator for cases where the posterior inclusion of both variables is equally likely (see columns 1, 5 and 8). Even in these scenarios, the measures do not provide a clear independence threshold. The value defining independence varies with $p(A)$ and $p(B)$, the posterior inclusion probabilities of both variables. *AllConf* and *MaxConf*, which only consider the minimum or maximum of the two conditional probabilities, $P(A|B)$ and $P(B|A)$, are exceptions. It has been argued that null-invariant measures are hardly able to correctly quantify positive and negative association, since they do not account for varying sizes of the exclusion margin (Glass, 2013). This can be considered less of a problem for certain applications in data mining, where only a small set of positive relationships out of a large set of transactions containing many zeros is of interest. However for the application to jointness the researcher mostly faces "balanced" datasets, where variable inclusion and exclusion are both similarly frequent.

Table 3 provides insights to the reaction of these different measures to substitutes and complements. We choose substitution relations in joint inclusion (columns 1 to 5) in such a way that the probability of common occurrence is one tenth of the independence threshold, or $P(AB) = 0.1 \times p(A)p(B)$. We find that non null-invariant measures regard all these scenarios as substitutes, leading to jointness values below zero. *Yule's Q* is in this regard very consistent, as it finds values close to its absolute minimum of -1 for all five cases. As a counterexample, *Normalized Difference* and the *ϕ-Coefficient* agree in the scenarios entertained where the exclusion margin is low (columns 2, 4 and 5) by regarding the pair as highly substitutes, but gain in value (towards independence) when this margin increases (columns 1 and 3). For the extreme case of $p(A) = p(B) = 0.1$ this results in a large exclusion margin of 0.8, while at the same time both indicators approach zero (-0.1).

While the independence threshold is not uniquely defined for null-invariant measures, most of these present very low values, close to their common lower bound of zero.

Columns 6–10 in Table 3 present five examples of bivariate complements, where p(AB) is

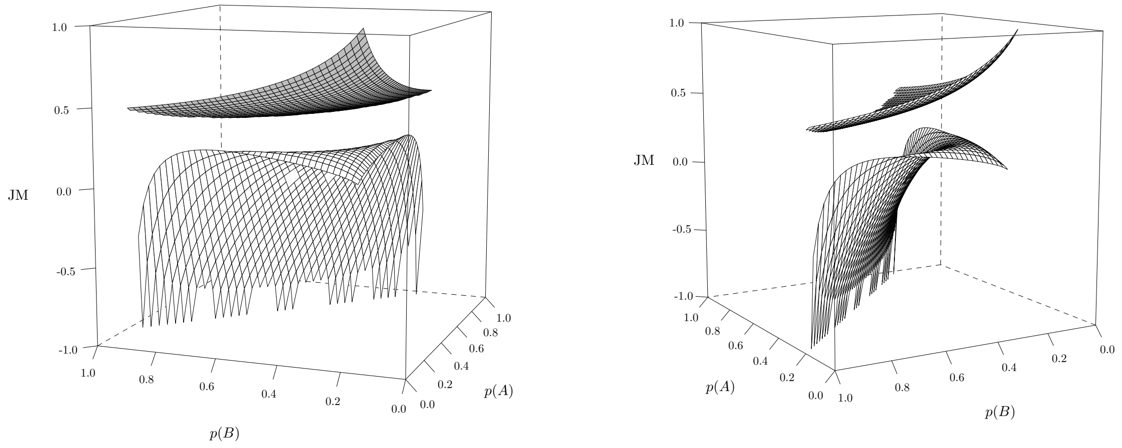


Figure 1: Jointness (JM) for Cosine (transparent) and Yule's Q (grey), $p(AB) = 0.2$

set to be a multiple of the independence threshold, $p(AB) = 9 \times p(A)p(B)$. Our findings for non null-invariant measures suggest that we correctly identify complements in all of these five scenarios. For $p(A|B) = 1$ the *Relative Risk* measure is infinite by construction. As before, the *Normalized Difference* and the *ϕ-Coefficient* are more ambiguous in their assertion of the complementarity relationships between pairs. Especially for cases which have a high level of mutual exclusion (column 8) — $p(\bar{A}B) = 0.8$ in this case — both measures shift in value towards independence. A similar case can be made for *Collective Strength* and for the scenario depicted in column 10.

Identifying complements via null-invariant measures seems to be a harder task, since we need to interpret these values relative to the (non unique) independence point. *MaxConf* always represents $p(A|B)$, which was chosen to be large, and therefore also ranges at its upper border of unity. *AllConf, Coherence* and *Kulczynski* all represent similar patterns to *Normalized Difference* and the *ϕ-Coefficient*, that is, high values when mutual exclusion is low and a drop in the indicator level as soon as either of these probabilities rise.

The effect of extreme values for the exclusion margin can also be grasped by assessing the jointness measures graphically. Figure 1 depicts the sensitivity of two measures, Cosine (null-invariant) and Yule's Q (Odds Ratio, non null-invariant) for a given level of joint occurrence $p(AB) = 0.2$ and varying values of $p(A)$ and $p(B)$ (X-, Y-axes). The Cosine measure is represented by a slightly convex plane, varying between 0.33 and 1, whereas mean and median lie close to 0.45. The maximum of the measure is found at the minimum values of $p(A)$ and $p(B)$, which correspond to the joint probability of 0.2. The measure then decreases towards the extreme values $\{1, 0.2\}$ and $\{0.2, 1\}$. In both cases the exclusion margin $p(\bar{A}\bar{B})$ is zero, however the probabilities of $p(A\bar{B})$ and $p(\bar{A}B)$ vary and cause the measure to react. The non null-invariant measure, Yule's Q, varies in a stronger fashion in an interval between 0.2 and its absolute minimum of -1. We find that this indicator is rather stable for individual inclusion values up to 0.4, which is twice the value of $p(AB)$, and for cases where $p(A) \gg p(B)$ or vice versa. For cases where the inclusion probabilities of both variables become large, the measure drops sharply indicating substitutability instead of complementarity. In our opinion this is a desirable indication. If the joint probability of occurrence if far below the marginal inclusion probabilities of the two variable, a measure should not classify them as complements.

4. Jointness of Economic Growth Determinants Revisited

Table 4: Results of the BMA routine for the FLS data set

	PIP	Post Mean	Post SD
GDP60	1.000	-0.016	0.003
Confucian	0.993	0.060	0.016
LifeExp	0.971	0.001	0.000
EquipInv	0.907	0.124	0.062
SubSahara	0.885	-0.016	0.008
Mining	0.815	0.031	0.020
Hindu	0.717	-0.050	0.042
NequipInv	0.696	0.034	0.029
RuleofLaw	0.666	0.008	0.007
LabForce	0.655	0.000	0.000
EcoOrg	0.614	0.001	0.001
Muslim	0.598	0.007	0.008
BlMktPm	0.566	-0.004	0.004
LatAmerica	0.563	-0.006	0.007
EthnoL	0.561	0.006	0.007
Protestants	0.559	-0.005	0.006
HighEnroll	0.554	-0.049	0.055
PrScEnroll	0.495	0.008	0.011
CivlLib	0.430	-0.001	0.002
Spanish	0.427	0.004	0.006

In our empirical application we apply alternative jointness measures to the dataset used in Fernández et al. (2001a, henceforth FLS), which includes information on income per capita growth and 41 potential determinants of economic growth differences for 72 countries.[6] In a first step, we apply BMA methods to obtain the posterior inclusion probabilities for all variables, as well as the mean and standard deviation of the posterior distribution of the parameters associated with each covariate. For this application we employ a hyper-g prior over the parameters (Liang et al., 2008) and a Binomial-Beta model prior following Ley and Steel (2009b). The BMA results are obtained using five million Markov Chain Monte Carlo iterations over the model space, where the first two million are disregarded as burn-in. Out of the three million visited models, approximately two thirds are unique, with a mean number of 19.8 included explanatory variables. Table 4 presents the posterior inclusion probabilities for the top 20 variables, together with the mean and standard deviation of the posterior distribution of their respective parameters. The BMA results confirm

[6]See Appendix B.1 for a description of the variables, as well as some descriptive statistics.

the robustness of several economic growth determinants such as *GDP60, Confucian, LifeExp* or *EquipInv*, which have a PIP above 0.9.

Using the top 10,000 unique models weighted by posterior model probabilities, we construct the binary matrix of model profiles, defined by the inclusion binary variables, γ_j. Since the top 10,000 models have been included approximately 130,000 times in the three million MCMC draws, this matrix has dimensions $130,000 \times 41$, where each cell describes whether covariate 1–41 is included (1) or not (0) in a given model. From this model profile matrix we can construct rules based on joint inclusion of variables.[7] A common further step in association analysis involves support-based pruning, where the rules are reduced given a minimum and/or maximum value for support, i.e., the frequency of a rule, and confidence, which measures the occurrences of a rule relative to the number of counterexamples. Pruning with respect to support eliminates infrequent rules, which only appear very rarely in the data. Table 5 shows the number of bivariate rules found for the FLS data set given different thresholds for support and confidence. We find a total of 1,640 bivariate rules if we impose no restrictions, which is no lower bounds for support or confidence. Twelve rules satisfy the most rigorous pruning, implied by only keeping highly frequent pairs which have a support value larger than 0.9. Following the association rules analysis literature, we use a low level of support pruning (0.1), so that we end up with a set of 582 distinct rules to analyze. In addition, we prune rules with extremely high support, namely the twelve resulting from a support level of 0.9. These may be of interest in the data mining context, but do not provide enough variation to analyze whether the covariates involved are substitutes or complements in the jointness context. The rules selected involve 29 of the 41 covariates.

Table 5: Number of rules by minimum confidence and support

Support/Confidence	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	1640	1002	750	611	530	467	403	347	291	217
0.1	304	304	264	227	215	201	168	145	128	101
0.2	154	154	154	140	133	124	99	83	75	62
0.3	110	110	110	110	106	102	81	66	59	48
0.4	88	88	88	88	88	86	70	60	55	44
0.5	74	74	74	74	74	74	61	53	48	41
0.6	38	38	38	38	38	38	38	34	30	25
0.7	28	28	28	28	28	28	28	28	25	20
0.8	20	20	20	20	20	20	20	20	20	16
0.9	12	12	12	12	12	12	12	12	12	12

For the overall set of identified joint variable inclusions and the pruned subset we obtain the interestingness measures described in Table 1 and calculate Spearman rank correlations, to quantify the concordance of the orderings implied by the different measures. Table 6 presents the results from this exercise. In the lower triangle, the results for the total of 1,640 rules are presented, while in the upper-right triangle we show the correlations for the pruned subset. The rank correlations within the group of non null-invariant measures imply highly congruent rankings

[7] We concentrate on bivariate jointness. A straightforward extension would be to analyze jointness based on triplets, for which tools such as the *apriori algorithm* can be used.

by these indicators. These measures provide rankings that are only loosely correlated with those delivered by their null-invariant counterparts. Comparing rank correlations for the full and pruned sets of associations, we find that the agreement increases above the diagonal in Table 6, which indicates that the exclusion of extreme cases causes the rankings implied by the measures to converge.

Within the set of null-invariant measures we find significantly less within-group correlation. While the measures *Coherence*, *Cosine* and *Kulczynski* tend to agree in terms of ranking bivariate inclusion relationships, this is not the case for *AllConf* and *MaxConf*. Since these two measures actually represent minima and maxima functions over the conditional inclusion probabilities $p(A|B)$ and $p(B|A)$, they frequently take extreme values at 0 or 1 and therefore produce rankings with a large number of ties around these values. Similarly, the rank correlations for the pruned set of bivariate inclusions are higher than for the full set.

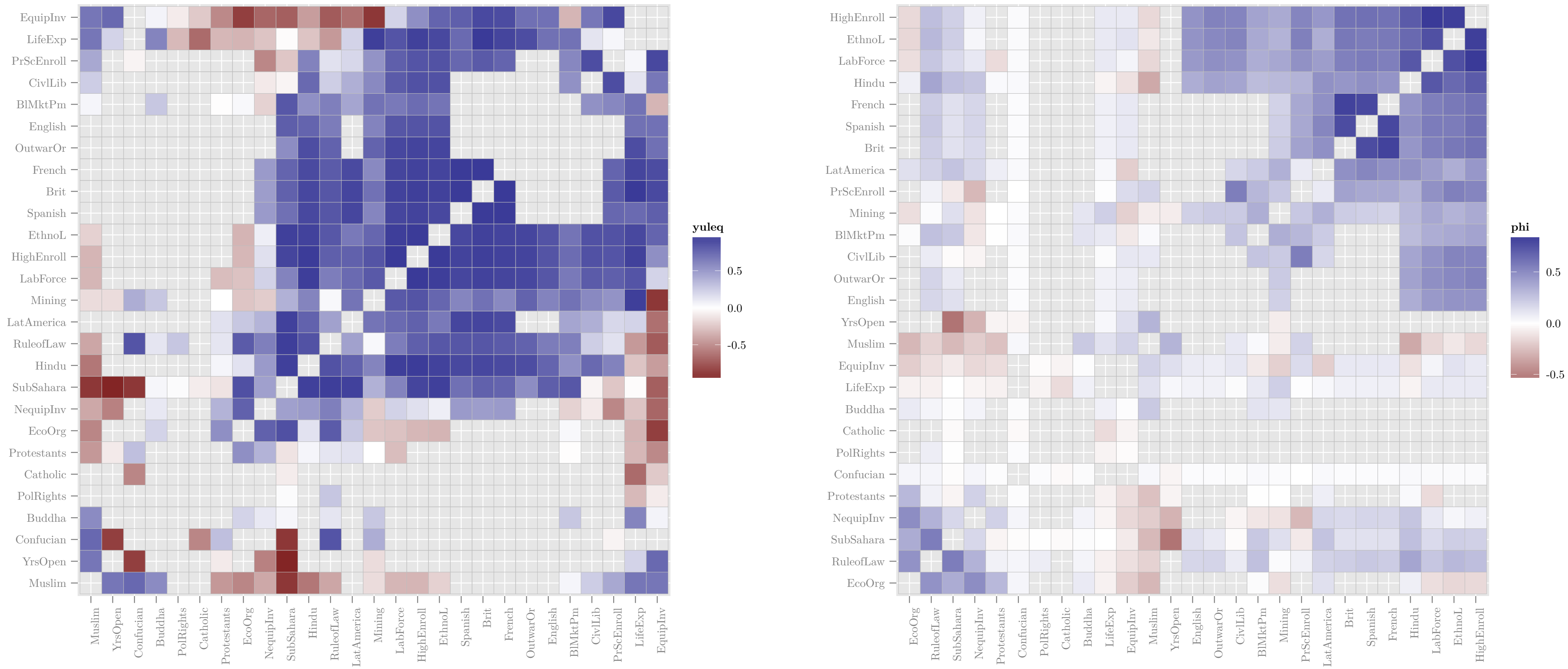
Given these results, we restrict our subsequent analysis to four distinct measures. On the one hand, we select the *Yule's Q* (an Odds-Ratio transformation) and the ϕ-*Coefficient*, which have been shown to react differently to the exclusion margin in the simulations. On the other hand, we concentrate on the null-invariant measures *Cosine* (Jaccard) and *Kulczynski*.

Table 6: Spearman Correlation of Ranked Measures

	Non null-invariant					**Null-invariant**				
	1	2	3	4	5	6	7	8	9	10
Non null-invariant										
1 Collective Strength		0.89	0.91	0.93	0.99	0.38	0.39	0.36	0.28	0.14
2 Relative Risk	0.83		0.91	0.86	0.90	0.24	0.24	0.22	0.19	0.10
3 Yule's Q	0.81	0.94		0.87	0.93	0.34	0.36	0.38	0.38	0.34
4 Normalized Difference	0.95	0.86	0.87		0.94	0.33	0.35	0.34	0.30	0.19
5 ϕ-Coefficient	0.99	0.82	0.86	0.94		0.43	0.44	0.43	0.37	0.26
Null-invariant										
6 AllConf	0.04	-0.15	-0.08	0.13	0.13		0.99	0.86	0.64	0.38
7 Coherence	0.03	-0.12	-0.03	0.13	0.14	0.98		0.91	0.71	0.47
8 Cosine	0.05	-0.06	0.04	0.16	0.17	0.94	0.99		0.93	0.76
9 Kulczynski	-0.02	-0.01	0.11	0.13	0.13	0.81	0.91	0.95		0.93
10 MaxConf	-0.22	0.12	0.25	-0.03	-0.06	0.17	0.32	0.43	0.64	

Notes: Lower-left triangle: Rank correlations for all 1640 rules
Upper-right triangle: Rank correlations for 568 pruned rules (support min 0.1/max 0.9)

Figures 2 and 3 represent graphically the degree of jointness implied by these four measures. The pairs of variables in these figures are ordered in such a way that high jointness patterns can be found along the diagonal of the matrix depicted in them (Hahsler et al., 2008; Tan et al., 2004). For Yule's Q in Figure 2a, we find a number of strong complementary relationships, represented by the blue shaded tiles. These clusters are primarily composed of the colonial dummies (*Brit, English, Spanish* and *French*) as well as geographical factors (*Latin America, SubSahara, EthnoL*). We also find a number of complements in the set of economic system-related variables, *OutwardOrientation, RuleOfLaw, LabForce* and *BlackMarketPremium.* In contrast, Yule's Q unveil very few substitutability relationships between pairs of variables. These are mainly related to religious variables (*Muslim, Confucian*) and their relation to the Sub-Saharan African dummy.



(a) Yule's Q (Odds Ratio)

(b) ϕ-Coefficient

Figure 2: non null-invariant Jointness Measures, FLS Dataset

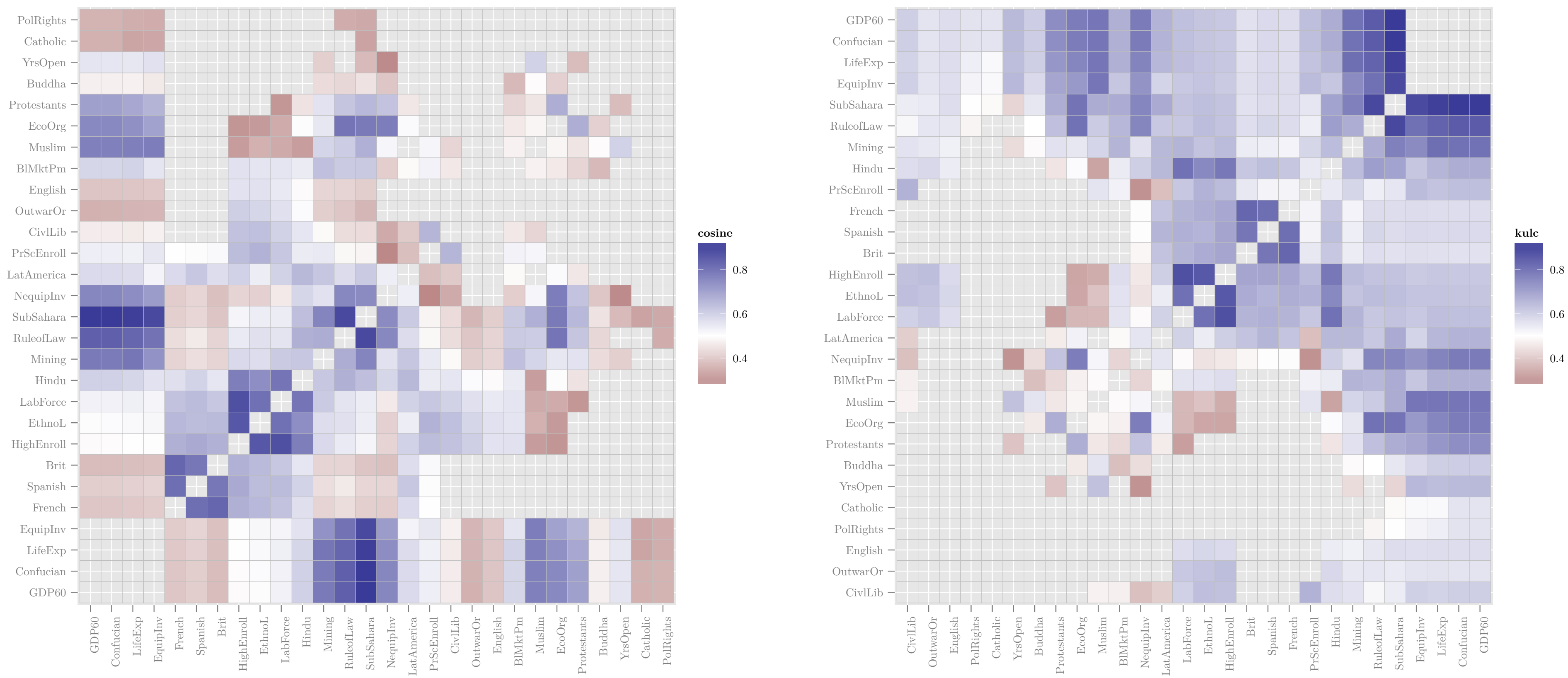
The ϕ-Coefficient (see Figure 2b) presents a similar picture with respect to colonial variables and *RuleOfLaw* or *OutwardOrientation*. However, it highlights even less substitutability relationships than Yule's Q besides the connection between *SubSahara* and *YrsOpen*.

The two null-invariant measures in Figure 3 show very similar patterns for complementarity of colonial and geographical variables. However, they tend to emphasize bivariate relationships of variables that have very high PIPs in the BMA exercise. For these covariates there are hardly any models where they do not appear together, so that these types of measures consider them to be strongly related in a complementarity sense. This applies to all the variables that present very high PIPs: *GDP60, Confucian, EquipInv* or *LifeExp*. The Cosine and the Kulczynski measures also find a number of substitutes, with *YearsOpenEconomy* and *NequipInv* being an example of these.

To sum up, both types of measure provide similar insights into the bivariate covariate inclusion structure in the model space. On the one hand, the additional weighting for the probability of joint exclusion in the non null-invariant measures causes relationships with high individual PIPs to lose importance as compared to the bivariate jointness of variables with average PIP. On the other hand, null-invariant measures ignore this exclusion margin and stress the importance of variable relationships where both variables have a high individual PIP.

In contrast to the results of LS for this data set, the jointness results found here are not exclusively related to variables with high PIP. For the measures introduced by LS, high jointness is concentrated among the top 5 regressors (*GDP60, Confucian, LifeExp, EquipInvest* and *Sub-Sahara*). This can be reproduced by restricting the analysis to the two null-invariant measures considered here. If however, the exclusion margin is included into the analysis, other jointness relationships are discovered. One example for these are colonial variables, which are less frequent, but still exhibit complementary behavior.

In their analysis, DW employ the dataset of Sala-i Martin et al. (2004, SDM data set), for which PIPs tend to be more concentrated on a few variables. Accordingly LS also find less jointness in this data set, using their null-invariant measures.



(a) Cosine

(b) Kulczynski

Figure 3: Null-invariant Jointness Measures, FLS Dataset

5. Conclusion

In this paper we investigate the issue of measuring jointness of robust growth determinants as raised by Ley and Steel (2007), Doppelhofer and Weeks (2009a) and others in the BMA literature. We link the measurement of joint inclusion of covariates to the field of assessing association in data mining, where similar problems are studied. We argue that the search for substitutes and complements in model profiles is similar to the data mining issue of finding “interesting” combinations of e.g. products in a shopping basket.

We link the properties that have been introduced for jointness to the concepts that are used for categorizing interestingness measures for association rules analysis. In particular, the jointness literature in BMA is concerned with a subset of these interestingness measures, referred to as confirmation measures. Furthermore, we highlight the role of null-invariance, that is, the effect of cases were both variables in a bivariate inclusion relationship are excluded. Based on these properties we select a set of interestingness measures and show how they relate to the jointness indicators proposed in the literature.

We show that null-invariant measures fail to give a comprehensive view on jointness since they cannot gauge the effect of statistical independence consistently across different dependence structures. We examine further how sensitive different measures are with regard to varying dependence structures across included covariates. Finally, we provide an empirical application of these measures to the well known dataset of economic growth determinants used by Fernández et al. (2001a) and discuss the complementarity and substitutability inclusion structures found.

Using non null-invariant measures, such as Yule’s Q, we find a large number of complementary relationships but only few substitutes among bivariate pairs of variables. The latter are primarily related to the combination of socioeconomic specifics (*Confucian, Muslim*) and geographical variables (*SubSahara*). Complementary relationships are manifold and can be found for example between different colonial variables, such as *Brit, English, Spanish* or *French.* Furthermore the quality of institutions (*RuleOfLaw*) and economic variables (*OutwardOrientation, BlackMarketPremium*) seem to exhibit such relationships.

As highlighted by Doppelhofer and Weeks (2009b), the treatment of the *exclusion margin* is highly relevant for an analysis of jointness. Null-invariance may lead to ambiguous results since these measures cannot quantify substitutes and complements in an appropriate fashion (Glass, 2013). Given this theoretical justification, we do find differences in the rank correlations between the two types of measures, but these only partly influence the general picture of complementary and substitute covariates found in the FLS dataset.

References

Aggarwal, C. C. and Yu, P. S. (1998). A new framework for itemset generation. In *PODS 98, Symposium on Principles of Database Systems*, pages 18–24, Seattle, WA, USA.

Brock, W. A. and Durlauf, S. N. (2001). What have we learned from a decade of empirical research on growth? Growth empirics and reality. *The World Bank Economic Review*, 15(2):229–272.

Crespo Cuaresma, J., Grün, B., Hofmarcher, P., Humer, S., and Moser, M. (2015). Unveiling covariate inclusion structures in economic growth regressions using latent class analysis.

Doppelhofer, G. and Weeks, M. (2005). Jointness of growth determinants. Cambridge working papers in economics.

Doppelhofer, G. and Weeks, M. (2009a). Jointness of growth determinants. *Journal of Applied Econometrics*, 24(2):209–244.

Doppelhofer, G. and Weeks, M. (2009b). Jointness of growth determinants: Reply to comments by Rodney Strachan, Eduardo Ley and Mark FJ Steel. *Journal of Applied Econometrics*, 24(2):252–256.

Eicher, T. S., Helfman, L., and Lenkoski, A. (2012). Robust fdi determinants: Bayesian model averaging in the presence of selection bias. *Journal of Macroeconomics*, 34(3):637–651.

Feldkircher, M. and Zeugner, S. (2009). Benchmark priors revisited: On adaptive shrinkage and the supermodel effect in bayesian model averaging. *IMF Working Papers*, pages 1–39.

Fernández, C., Ley, E., and Steel, M. F. J. (2001a). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics*, 16(5):563–576.

Fernández, C., Ley, E., and Steel, M. J. F. (2001b). Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, 100(2):381–427.

Foster, D. P. and George, E. I. (1994). The risk inflation criterion for multiple regression. *The Annals of Statistics*, 22(4):1947–1975.

Geng, L. and Hamilton, H. J. (2006). Interestingness measures for data mining: A survey. *ACM Computing Surveys (CSUR)*, 38(3):9.

Glass, D. H. (2013). Confirmation measures of association rule interestingness. *Knowledge-Based Systems*, 44:65–77.

Hahsler, M., Grün, B., and Hornik, K. (2005). arules – A computational environment for mining association rules and frequent item sets. *Journal of Statistical Software*, 14(15):1–25.

Hahsler, M., Hornik, K., and Buchta, C. (2008). Getting things in order: An introduction to the r package seriation. *Journal of Statistical Software*, 25(3):1–34.

Ley, E. and Steel, M. F. J. (2007). Jointness in Bayesian variable selection with applications to growth regression. *Journal of Macroeconomics*, 29(3):476–493.

Ley, E. and Steel, M. F. J. (2009a). Comments on 'Jointness of growth determinants'. *Journal of Applied Econometrics*, 24(2):248–251.

Ley, E. and Steel, M. F. J. (2009b). On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. *Journal of Applied Econometrics*, 24(4):651–674.

Ley, E. and Steel, M. F. J. (2012). Mixtures of g-priors for Bayesian model averaging with economic applications. *Journal of Econometrics*, 171(2):251–266.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, 103(481):410–423.

Madigan, D. and York, J. (1995). Bayesian graphical models for discrete data. *International Statistical Review*, 63(2):215–232.

Moral-Benito, E. (2012). Determinants of economic growth: a bayesian panel data approach. *Review of Economics and Statistics*, 94(2):566–579.

Moral-Benito, E. (2014). Growth empirics in panel data under model uncertainty and weak exogeneity. *Journal of Applied Econometrics*.

Piatetsky-Shapiro, G. (1991). Discovery, analysis and presentation of strong rules. In Piatetsky-Shapiro, G. and Frawley, W., editors, *Knowledge Discovery in Databases*, pages 229–248. MIT Press, Cambridge, MA.

Sala-i Martin, X., Doppelhofer, G., and Miller, R. I. (2004). Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach. *American Economic Review*, 94(4):813–835.

Strachan, R. W. (2009). Comment on 'Jointness of growth determinants' by Gernot Doppelhofer and Melvyn Weeks. *Journal of Applied Econometrics*, 24(2):245–247.

Tan, P.-N., Kumar, V., and Srivastava, J. (2004). Selecting the right objective measure for association analysis. *Information Systems*, 29(4):293–313.

Tew, C., Giraud-Carrier, C., Tanner, K., and Burton, S. (2014). Behavior-based clustering and analysis of interestingness measures for association rule mining. *Data Mining and Knowledge Discovery*, 28(4):1004–1045.

Vaillant, B., Lenca, P., and Lallich, S. (2004). A clustering of interestingness measures. In *Discovery Science*, pages 290–297. Springer.

Wu, T., Chen, Y., and Han, J. (2010). Re-examination of interestingness measures in pattern mining: A unified framework. *Data Mining and Knowledge Discovery*, 21(3):371–397.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g–prior distributions. *Bayesian Inference and Decision Techniques Essays in Honor of Bruno de Finetti*, 6:233–243.

A. Review of measures of interestingness and confirmation

Table A.1: Definition of Jointness measures

	#	Measure	Value
1	ϕ	ϕ-Coefficient	$\frac{p(AB)-p(A)p(B)}{\sqrt{p(A)p(B)p(\bar{A})p(\bar{B})}}$
2	AV	Added Value	$p(B\mid A) - p(B)$
3	AC	AllConf	$\min(p(B\mid A), p(A\mid B))$
4	b	Carnap	$p(AB) - p(A)p(B)$
5	cf	Certainty Factor	$\frac{p(B\mid A)-p(B)}{1-p(B)}$ if $p(B\mid A) > p(B)$
6	χ^2	Chi-square (χ^2)	$\frac{(p(AB)-p(A)p(B))^2 N}{p(A)p(\bar{A})p(B)p(\bar{B})}$
7	κ	Coehen's Kappa (κ)	$\frac{p(B\mid A)p(A)+p(\bar{B}\mid\bar{A})p(\bar{A})-p(A)p(B)-p(\bar{A})p(\bar{B})}{1-p(A)p(B)-p(\bar{A})p(\bar{B})}$
8	coh	Coherence	$(p(A\mid B)^{-1} + p(B\mid A)^{-1} - 1)^{-1}$
9	cs	Collective Strength	$\ln\left[\frac{p(AB)+p(\bar{A}\bar{B})}{p(A)p(B)+p(\bar{A})p(\bar{B})} \times \frac{1-p(A)p(B)-p(\bar{A})p(\bar{B})}{1-p(AB)-p(\bar{A}\bar{B})}\right]$
10	conf	Confidence	$p(B\mid A)$
11	conv	Conviction	$\ln\left[\frac{p(A)p(B)}{p(A,\bar{B})}\right]$
12	IS	Cosine	$\frac{p(AB)}{\sqrt{p(A)p(B)}}$
13	G	Gini index	$p(A)(p(B\mid A)^2 + p(\bar{B}\mid A)^2) + p(\bar{A})(p(B\mid\bar{A})^2 + p(\bar{B}\mid\bar{A})) - p(B)^2 - p(\bar{B})^2$
14	IR	Imbalance Ratio	$\frac{\lvert p(A\mid B-p(B\mid A)\rvert}{Pr(A\mid B)+p(B\mid A)-p(A\mid B)p(B\mid A)}$
15	I	Interest	$\frac{p(AB)}{p(A)p(B)}$
16	J	J-Measure	$p(AB)\,\log\frac{p(B\mid A)}{p(B)} + p(A\bar{B})\,\log\frac{p(\bar{B}\mid A)}{p(\bar{B})}$
17	ζ	Jaccard (ζ)	$\frac{p(AB)}{p(A)+p(B)-p(AB)}$
18	k	Kemeny-Oppenheim	$\frac{p(A\mid B)-p(A\mid\bar{B})}{p(A\mid B)+p(A\mid\bar{B})}$
19	kl	Klosgen	$\sqrt{p(AB)} \times \max(p(B\mid A) - p(B), p(A\mid B) - p(A))$
20	kulc	Kulczynski	$(p(A\mid B) + p(B\mid A))/2$
21	L	Laplace	$\frac{N\times p(AB)+1}{N\times p(A)+2}$
22	l	Lift	$\frac{p(B\mid A)}{p(B)}$
23	ll	Log-Likelihood	$\ln\left[\frac{p(A\mid B)}{p(A\mid\bar{B})}\right]$
24	r	Log-Ratio	$\ln\left[\frac{p(B\mid A)}{p(B)}\right]$
25	MC	MaxConf	$\max(p(B\mid A), p(A\mid B))$
26	M	Mutual Information	$p(AB)\,\log\frac{p(AB)}{p(A)p(B)} + p(A\bar{B})\,\log\frac{A\bar{B}}{p(A)p(\bar{B})} + p(\bar{A}B)\,\log\frac{p(\bar{A}B)}{p(\bar{A})p(B)} + p(\bar{A}\bar{B})\,\log\frac{p(\bar{A}\bar{B})}{p(\bar{A})p(\bar{B})}$
27	s	Normalized Difference	$p(B\mid A) - p(B\mid\bar{A})$
28	α	Odds Ratio	$\ln\left[\frac{p(AB)p(\bar{A}\bar{B})}{p(A,\bar{B})p(\bar{A}B)}\right]$
29	ows	One-Way Support	$p(B\mid A)\ln\left[\frac{p(AB}{p(A)p(B}\right]$
30	PS	Piatetsky-Shapiro's	$N \times (p(AB) - p(A)p(B))$
31	rr	Relative Risk	$\ln\left[\frac{p(B\mid A)}{p(B\mid\bar{A})}\right]$
32	sup	Support	$p(AB)$
33	tws	Two-Way Support	$p(AB)\ln\left[\frac{p(AB)}{p(A)p(B)}\right]$
34	yq	Yule's Q	$\frac{p(AB)p(\bar{A}\bar{B})-p(A\bar{B})p(\bar{A}B)}{p(AB)p(\bar{A}\bar{B})+p(A\bar{B})p(\bar{A}B)}$
35	yy	Yule's Y	$\frac{\sqrt{p(AB)p(\bar{A}\bar{B})}-\sqrt{p(A\bar{B})p(\bar{A}B)}}{\sqrt{p(AB)p(\bar{A}\bar{B})}+\sqrt{p(A\bar{B})p(\bar{A}B)}}$

B. Data description

Table B.1: Variable Names and Descriptive Statistics — FLS

	Abbreviation	Variable	μ	σ
1	Age	Age	23.71	37.307
2	Area	Area (Scale Effect)	972.92	2051.976
3	BlMktPm	Black Market Premium	0.16	0.291
4	Brit	British Colony dummy	0.32	0.470
5	Buddha	Fraction Buddhist	0.06	0.184
6	Catholic	Fraction Catholic	0.42	0.397
7	CivlLib	Civil Liberties	3.47	1.712
8	Confucian	Fraction Confucian	0.02	0.087
9	EcoOrg	Degree of Capitalism	3.54	1.266
10	English	Fraction of Pop. Speaking English	0.08	0.239
11	EquipInv	Equipment investment	0.04	0.035
12	EthnoL	Ethnolinguistic fractionalization	0.37	0.296
13	Foreign	Fraction speaking foreign language	0.37	0.422
14	French	French Colony dummy	0.12	0.333
15	GDP60	GDP level in 1960	7.49	0.885
16	HighEnroll	Higher education enrollment	0.04	0.052
17	Hindu	Fraction Hindu	0.02	0.101
18	Jewish	Fraction Jewish	0.01	0.097
19	LabForce	Size labor force	9305.38	24906.056
20	LatAmerica	Latin American dummy	0.28	0.451
21	LifeExp	Life expectancy	56.58	11.448
22	Mining	Fraction GDP in mining	0.04	0.077
23	Muslim	Fraction Muslim	0.15	0.295
24	NequipInv	Non-Equipment Investment	0.15	0.055
25	OutwarOr	Outward Orientation	0.39	0.491
26	PolRights	Political Rights	3.45	1.896
27	Popg	Population Growth	0.02	0.010
28	PrExports	Primary exports, 1970	0.67	0.299
29	Protestants	Fraction Protestant	0.17	0.252
30	PrScEnroll	Primary School Enrollment, 1960	0.80	0.246
31	PublEdupct	Public Education Share	0.02	0.009
32	RevnCoup	Revolutions and coups	0.18	0.238
33	RFEXDist	Exchange rate distortions	121.71	41.001
34	RuleofLaw	Rule of law	0.55	0.335
35	stdBMP	SD of black-market premium	45.60	95.802
36	SubSahara	Sub-Saharan dummy	0.21	0.409
37	WarDummy	War dummy	0.40	0.494
38	WorkPop	Ratio workers to population	−0.95	0.189
39	YrsOpen	Number of Years open economy	0.44	0.355