Accepted Manuscript

Bivariate jointness measures in Bayesian Model Averaging: Solving the conundrum

Paul Hofmarcher, Jesus Crespo Cuaresma, Bettina Grün, Stefan Humer, Mathias Moser

PII: S0164-0704(18)30031-4
DOI: 10.1016/j.jmacro.2018.05.005
Reference: JMACRO 3027

To appear in: *Journal of Macroeconomics*

Received date: 16 January 2018
Revised date: 8 May 2018
Accepted date: 15 May 2018

Please cite this article as: Paul Hofmarcher, Jesus Crespo Cuaresma, Bettina Grün, Stefan Humer, Mathias Moser, Bivariate jointness measures in Bayesian Model Averaging: Solving the conundrum, *Journal of Macroeconomics* (2018), doi: 10.1016/j.jmacro.2018.05.005

Bivariate jointness measures in Bayesian Model Averaging: Solving the conundrum*

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

[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)
[f]Research Institute for Economics of Inequality (INEQ)
[g]Department of Socioeconomics, Vienna University of Economics and Business (WU)

May 16, 2018

Abstract

We introduce a new measure of bivariate jointness to assess the degree of inclusion dependency between pairs of explanatory variables in Bayesian Model Averaging analysis. Building on the discussion concerning appropriate statistics to assess covariate inclusion dependency in this context, a set of desirable properties for bivariate jointness measures is proposed. We show that none of the proposed measures so far meets all these criteria and an alternative measure is presented which fulfils all of them. Our measure corresponds to a regularised version of the Yule's Q association coefficient, obtained by combining the original measure with a Jeffreys prior to avoid problems in the case of zero counts. We provide an empirical illustration using cross-country data on economic growth and its determinants.

JEL Classification: C11, C55, O40.
Keywords: Bayesian Model Averaging, Jointness, Robust Growth Determinants, Association Rules.

*The authors would like to thank an anonymous referee, Gernot Doppelhofer, Sylvia Frühwirth-Schnatter and the participants at the *WU Workshop in Applied Econometrics* and the *Applied Econometrics Meeting* in Sevilla for helpful comments. Corresponding Author: Paul Hofmarcher. Address: Welthandelsplatz 1, 1020 Vienna, Austria, Tel: +43(0)131336-4514, Email: paul.hofmarcher@wu.ac.at.

1. Introduction

Bayesian Model Averaging (henceforth BMA, see Hoeting et al., 1999) has become a standard econometric tool to carry out inference in the presence of model uncertainty. In economics, a large body of literature has used BMA to assess the robustness of empirical determinants of economic growth differences across countries (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, 2016). In such applications, the importance of a specific variable is routinely measured by its posterior inclusion probability (PIP). The PIP is defined as the sum of the posterior probabilities of the model specifications which contain that particular variable. Following an early discussion of Bayesian measures of variable importance (Leamer, 1978; Mitchell and Beauchamp, 1988), PIPs have become a standard tool for interpreting results in econometric applications of BMA. While they provide valuable insight into the overall importance of single variables, they rely only on marginal posterior distributions and neglect the interdependence of inclusion and exclusion of variables. Using PIPs it is for example not possible to conclude if the importance of the variable is evenly spread across all potential model specifications or it is specific to a certain combination of explanatory variables (see Crespo Cuaresma et al., 2016).

To gain insights into the interdependence of the inclusion of sets of different variables, several existing studies focus on investigating the joint posterior inclusion of pairs of variables. This joint bivariate inclusion probabilities allow to infer whether two variables are complements, i.e., tend to be included together in models with high posterior probability, or substitutes, i.e., tend to exclude the inclusion of the other. To analyse the joint posterior inclusion probabilities, bivariate jointness measures within the BMA framework were proposed first in the working paper by Doppelhofer and Weeks (2005), which was published in a slightly different version as Doppelhofer and Weeks (2009a). In addition Ley and Steel (2007), Strachan (2009) as well as Ley and Steel (2009a) propose alternative measures. Ley and Steel (2007) list a set of properties which should be fulfilled by jointness measures and show that these properties are not fulfilled by the statistics put forward in Doppelhofer and Weeks (2009a). Strachan (2009) shows that the interpretability of the jointness measure proposed by Doppelhofer and Weeks (2009a) may be limited in contexts where one or both of the analysed variables have a negligible PIP and offers yet another measure in order to tackle this shortcoming. In their reply to these criticisms, Doppelhofer and Weeks (2009b) propose a further desirable property which requires that any jointness measure should test the dependence over the joint posterior distribution of the variables considered and show that this property is not fulfilled by the measures proposed by Ley and Steel (2007) and Strachan (2009).

In addition to these contributions on bivariate jointness indicators for analysing BMA output, different association measures between sets of variables have also been proposed in the machine learning literature on association rules (see, e.g., Tan et al., 2004; Hahsler et al., 2005; Glass, 2013, 2014). Association rules aim at finding "interesting" patterns in large binary databases which allow to infer the presence of a certain set of variables given that a different set of variables is also present. Due to the large number of possible rules, interestingness measures have been developed for rules which allow to restrict attention to rules with high values for these interestingness measures and certain properties have been proposed in this stream of literature in order to select adequate interestingness measures.

In this paper we present a rigorous analysis of the properties of bivariate jointness measures in BMA, combining the insights from these two strands of literature: the literature on jointness indicators based on post-processing of BMA output and the machine learning literature on interestingness measures for association rules. This leads us to propose a set of properties a bivariate jointness measure for BMA analysis should fulfil. Since none of the previously proposed statistics meets all these criteria, a suitable new bivariate jointness measure is presented. This new measure is a regularised version of the well known Yule's Q association coefficient and is derived based on an augmented contingency table of variable inclusion which allows us to avoid the problems that arise due to zero counts. More specifically, we combine the multinomial distribution underlying the bivariate contingency table with a Dirichlet prior. As the Dirichlet distribution is conjugate to the multinomial distribution, the posterior distribution of the parameters is again a Dirichlet distribution. The parameters correspond to the augmented (posterior) contingency table. Our proposed *modified Yule's Q* measure is obtained from the posterior inclusion contingency table for pairs of variables and corresponds to the Yule's Q measure for this table. It thus includes a small correction factor such that the measure fulfils all desired properties of a bivariate jointness measure in BMA post-processing analysis.

The remainder of this paper is organised as follows. Section 2 describes the modelling setting in which BMA applications tend to be carried out in econometric research, as well as the bivariate measures for post-processing BMA output that have been proposed hitherto in the econometric literature. Section 3 discusses a set of desirable properties for bivariate jointness measures in BMA analysis and assesses the extent to which the existing measures fulfil them. These results lead us to propose a new bivariate jointness measure which meets all the proposed criteria. Section 4 presents an empirical application of our measure for the cross-country dataset of economic growth and its determinants from Fernández et al. (2001a). Section 5 concludes.

2. Bayesian Model Averaging and Jointness Analysis

BMA methods are routinely used to obtain posterior distributions for the quantities of interest in a linear regression setting while accounting for specification uncertainty with respect to covariate inclusion. The linear regression models entertained in BMA are usually assumed to be of the form

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

where y is an $n \times 1$ vector containing the observations of the dependent variable, $\mathbb{1}$ a vector of ones of the same length and the $n \times k_j$ matrix X_j is composed of the observations of k_j variables out of a total set of K potential covariates. Considering the $K \times 1$ vector β containing the parameters associated to each of the K potential covariates, model uncertainty can be explicitly addressed by basing inference of the parameters of interest on the posterior distribution given by

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

where M_j denotes a specific model out of the possible 2^K model specifications where a subset of the K variables is included. Each specification-specific posterior distribution $p(\alpha, \beta, \sigma|y, M_j)$ is weighted by its corresponding posterior model probability, $p(M_j|y)$. Posterior model probabilities are proportional to the marginal likelihood of the specification 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 use improper non-informative priors for α and σ, i.e., $p(\alpha) \propto 1$ and $p(\sigma) \propto \sigma^{-1}$. A common choice for the prior of the slope coefficients β_j is Zellner's g–prior (Zellner, 1986),

$$p(\beta_j|M_j, \sigma) \sim N\left(0, \sigma^2 \left(\frac{1}{g_0} X_j' X_j\right)^{-1}\right). \tag{4}$$

The prior variance-covariance matrix of the vector of parameters associated to the covariates considered has thus the structure of the variance-covariance matrix of the ordinary least squares estimator and is scaled by the parameter g_0. Several approaches to select a specific value of g_0 have been proposed (see, e.g., Foster and George, 1994; Fernández et al., 2001b). To allow for more flexibility, hyper-priors on g_0 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 use a flat prior over all possible specifications, so that $p(M_j) = 2^{-K}$ for all j. Given that this prior implies a preference for models of size around $K/2$, Ley and Steel (2009b) argue for a beta-binomial prior on covariate inclusion. This setting is able to achieve a very flexible prior structure over the model size and induces an uninformative distribution on the number of included covariates.

Since analysing all 2^K models that can be obtained by combining covariates 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 approximate the relevant posterior distributions.

The output of the MC^3 procedure allows to investigate the posterior distributions for inclusion as well as other quantities of interest. In particular, the joint posterior distribution of covariate inclusion constitutes the basis to create measures of jointness. Following Doppelhofer and Weeks (2009a), let model specifications be represented by binary vectors of covariate inclusion profiles (as defined by the inclusion variables γ_k, $k = 1, \dots, K$), so that

$$p(M_j|y) = p(\gamma_1 = c_1, \gamma_2 = c_2, \dots, \gamma_K = c_K|y) \ , \tag{5}$$

where $c_k \in \{0, 1\}$ is the binary variable representing the inclusion or exclusion of covariate k in the model M_j. A posterior estimate for the jointness of two variables, $p(\gamma_i, \gamma_j|y), i \neq j$ can be derived from the contingency table containing the MCMC frequency counts of joint inclusion, joint exclusion and the inclusion of one variable while excluding the other variable. In particular, for two arbitrary variables A, B of the K potential variables in the regression model, the inclusion profiles Γ_A and Γ_B (which are vectors of length equal to the number of MCMC iterations, N) allow to construct a contingency table of the form given in Table 1.[1]

[Table 1 about here.]

Given these inclusion profiles, the posterior probability of variables A and B appearing jointly across models $(p(A \cap B|y) \equiv p(AB|y))$ can be estimated from the MCMC draws by a/N. The posterior probabilities of each variable appearing alone in a model $(p(A \cap \overline{B}|y) \equiv p(A\overline{B}|y)$ and $p(\overline{A} \cap B|y) \equiv p(\overline{A}B|y))$ can be estimated by b/N and c/N, respectively. Finally, the posterior probability that the true model does not contain any of the two variables $(p(\overline{AB}|y))$ can be obtained by d/N.

The bivariate jointness measures of interest are functions $J(a, b, c, d)$ that return a single value which allows to characterise the association between the pair of variables in terms of joint posterior inclusion. The values of the function should allow to distinguish between pairs of variables which are complements, occur independently or are substitutes in terms of their joint inclusion. Variables are considered complements if the inclusion of one variable implies that the other variable also tends to be included, while for substitutes the inclusion of one variable tends to imply the exclusion of the other covariate. For each jointness measure, the existence of a clearly defined threshold which allows to distinguish between complements and substitutes and which indicates when variables occur independently would be a desirable property.

So far, five different measures of jointness have been proposed in the econometric literature to post-process BMA output. These measures differ in the way they incorporate the different marginal and joint inclusion probabilities. The earliest jointness measure in the BMA context can be attributed to Doppelhofer and Weeks (2005), who propose to use

$$\tilde{J}_{\text{DW1}}(A, B) = \ln\left(\frac{p(AB|y)}{p(A|y)p(B|y)}\right), \tag{6}$$

[1] We assume that the output used to determine the jointness measures is obtained using MCMC methods. If the model space is tractable and all 2^K models can be estimated, the contingency table would be created using the actual posterior probabilities covered by models which include or exclude the respective variables and the jointness measures are not estimated, but are determined exactly.

which is estimated from the contingency table by

$$J_{\mathrm{DW1}} = \ln\left(\frac{aN}{(a+b)(a+c)}\right). \tag{7}$$

This measure corresponds to the log of the joint inclusion in relation to the product of the individual inclusion probabilities of the variables. Ley and Steel (2007) criticise this jointness indicator based on the fact that the measure may be misleading for variables with extremely high or low PIPs and that values of the measure can hardly be compared across different pairs of variables.

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

$$\tilde{J}_{\mathrm{DW2}}(A,B) = \ln\left(\frac{p(AB|y)p(\overline{A}\overline{B}|y)}{p(A\overline{B}|y)p(\overline{A}B|y)}\right), \tag{8}$$

which is estimated by

$$J_{\mathrm{DW2}} = \ln\left(\frac{ad}{bc}\right). \tag{9}$$

In their comment to this contribution, Ley and Steel (2009a) also question this approach, since J_{DW2} is not defined in cases where a variable has a PIP of zero or one. Instead, they propose two alternative measures (Ley and Steel, 2007),

$$\tilde{J}_{\mathrm{LS1}}(A,B) = \frac{p(AB|y)}{p(A|y)+p(B|y)-p(AB|y)}, \quad \tilde{J}_{\mathrm{LS2}}(A,B) = \frac{p(AB|y)}{p(A\overline{B}|y)+p(\overline{A}B|y)}, \tag{10}$$

which are computed using the contingency table as

$$J_{\mathrm{LS1}} = \frac{a}{a+b+c}, \qquad J_{\mathrm{LS2}} = \frac{a}{c+b}. \tag{11}$$

While J_{LS1} relates the joint inclusion to the probability of including either one of the two variables, J_{LS2} uses the probability of including either one but not both variables as a normalising factor.

Strachan (2009) proposes to ignore the joint exclusion and to concentrate on relevant variables in terms of PIP. This is accomplished by adapting J_{DW2} in such a way that the joint exclusion is omitted and that the marginal probabilities of both variables are included,

$$\tilde{J}_{\mathrm{St}}(A,B) = p(A|y)p(B|y)\,\ln\left(\frac{p(AB|y)}{p(A\overline{B}|y)p(\overline{A}B|y)}\right), \tag{12}$$

that is,

$$J_{\mathrm{St}} = \frac{a+b}{N}\frac{a+c}{N}\ln\left(\frac{aN}{bc}\right). \tag{13}$$

The treatment of $p(\overline{AB}|y)$, or equivalently its estimate d/N, has been controversially discussed and no unanimous preference for either taking it into account or ignoring it was established. This *exclusion margin* indicates the extent to which both variables tend to not be included in specifications. While Doppelhofer and Weeks (2009b) stress the importance of this aspect in their discussion, the jointness measures proposed by Ley and Steel (2009a) and Strachan (2009) do not take the exclusion margin into account to determine the inclusion dependency of a pair of variables.

In addition to the aforementioned measures proposed in the BMA literature, a parallel literature on data mining proposes alternative measures of association that are similar in nature. Recent surveys in this field collect as many as 40 different measures and try to provide a structural overview of the alternative indicators available (see for example Geng and Hamilton, 2006; Tan et al., 2004; Glass, 2013; Crespo Cuaresma et al., 2015).[2] Some of these measures resemble the ones proposed in the BMA jointness literature. The jointness measure in Doppelhofer and Weeks (2005) is equivalent to the so-called Log-Ratio or equivalently, the log of the Interest (Lift) measure (Geng and Hamilton, 2006). Ley and Steel (2007)'s J_{LS1} is identical to the Jaccard index and their J_{LS2} measure is an adjustment thereof. A similar measure to Strachan (2009)'s J_{St} has been introduced in this strand of

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

literature as Two-Way Support (Geng and Hamilton, 2006). Finally, a modification of the statistic proposed by Doppelhofer and Weeks (2009a) is known as the Odds-Ratio in the field of data mining (Tew et al., 2014).

Other prominent measures in the association rule literature include *support*, which is simply defined as the joint inclusion probability of A and B, i.e., $p(AB)$, *confidence*, defined as $conf(A,B) = \frac{p(AB)}{p(B)}$ or *Yule's Q*, defined as

$$\tilde{J}_{\mathrm{YQ}}(A,B) = \frac{p(AB)p(\overline{AB}) - p(A\overline{B})p(\overline{A}B)}{p(AB)p(\overline{AB}) + p(A\overline{B})p(\overline{A}B)}, \tag{14}$$

which is estimated from the contingency table by

$$J_{\mathrm{YQ}} = \frac{ad - bc}{ad + bc}. \tag{15}$$

In fact, J_{YQ} is a mapping of Doppelhofer and Weeks (2009a)'s J_{DW2} to the $[-1,1]$ interval.

3. On the Properties of Jointness Measures for BMA Analysis

3.1. Desirable Properties for Jointness Measures

For BMA applications, Ley and Steel (2007) propose four properties that BMA jointness measures should fulfil. A jointness measure should be *interpretable*, in the sense that it has a "clear intuitive meaning" and is well *calibrated* against a clearly defined scale. Furthermore, the property of *extreme jointness* implies that a measure should reach its maximum when both variables always appear together in the specifications considered. A measure should always be defined (*definition*) whenever at least one of the variables is included with positive probability. In the data mining literature on association rules (see Wu et al., 2010; Glass, 2014), other characteristics that are expected to be fulfilled by interestingness measures in association rules analysis are put forward. We merge and extend the properties proposed in both strands of literature in order to assess the suitability of existing jointness measures and derive an appropriate measure for jointness analysis in BMA.

The properties that the existing literature discusses in the BMA context partly correspond to those which are used to measure interestingness of association rules. The adequacy of a measure clearly depends on the properties that are required for a particular application. For example, while machine learning problems are often concerned with positive association, BMA results may additionally need to reflect negative association in order to be able to identify variables which are substitutes. Furthermore, some consideration needs to be given to the question whether two variables in a model are considered exchangeable, so that measures associated with *if A then B* ($A \rightarrow B$) lead to the same result as measures associated with $B \rightarrow A$. Based on the insights from the BMA and machine learning discussions, we select several properties which can be considered crucially relevant for BMA jointness analysis.

P1. Definition: This property follows from Ley and Steel (2007). A jointness measure should be defined whenever one of the variables is included with positive probability. Obviously this property is fulfilled by the measures proposed by Ley and Steel (2007) but not by the measures proposed by Strachan (2009) or Doppelhofer and Weeks (2009a). Yule's Q does not fulfil the definition property either.

P2. Monotonicity, boundedness and maximality: For interpretability, the range of an association measure is required to be *bounded* on $[-1, 1]$. In addition, the measure should be *monotonically increasing* in a and d between the two extreme cases (see Glass, 2013). The jointness measure should be maximal if and only if $b = c = 0$ and minimal if and only if $a = d = 0$ (Glass, 2014). This property is related to *extreme jointness*, the property introduced by Ley and Steel (2007) in the BMA literature. However, in contrast to their definition, the property proposed here ensures that a measure reaches an extreme value when both variables always appear together or are jointly excluded. The explicit consideration of joint exclusion d is the decisive difference between Ley and Steel (2007) and Glass (2014).

P3. Limiting behaviour: This property has not been discussed in the BMA literature hitherto. For any finite values $b, c \geq 0$ and $a \vee d \rightarrow \infty$ the jointness measure should converge to its maximum. Analogously, for given finite values $a, d \geq 0$ and $b \vee c \rightarrow \infty$ the jointness measure should converge to its minimum.

P4. Bayesian confirmation: This property is proposed in Glass (2013). A jointness measure J fulfils the property of a Bayesian confirmation measure if it holds that

$$J > 0 \iff p(A|B) > p(A),$$
$$J = 0 \iff p(A|B) = p(A),$$
$$J < 0 \iff p(A|B) < p(A).$$

Bayesian confirmation requires that the jointness measure equals zero if and only if the inclusion of variable A and the inclusion of variable B are statistically independent events. Complements are characterised by a positive jointness measure and a negative value of the jointness indicator means that A is less likely to be included as a covariate in a regression model where B is already included than expected from its marginal distribution (and vice-versa). The measure is thus anchored at zero, which corresponds to statistical independence of the inclusion of the two covariates. For the case of jointness measures discussed in the BMA literature, this property is fulfilled only for the measures J_{DW1} and J_{DW2}. This property is closely related to the property of monotonicity (P2) and to non null-invariance (P5).

P5. Non null-invariance: The non null-invariance property is extensively discussed in Glass (2013). Measures that are null-invariant ignore "null transactions", i.e., specifications where A and B are jointly excluded. When estimated from the contingency table, this implies that they do not depend on d. 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) criticise 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 models with low posterior probability, which might be spuriously assessed as "interesting" by non null-invariant measures where the common exclusion probability is respected.

However, non null-invariance seems to be a crucial assumption for bivariate jointness measures. Tables 3a and 3e present two exemplary contingency tables where the inclusion of covariates A and B are statistically independent events. For Table 3a J_{LS2} returns a different value than for Table 3e. As a consequence, for null-invariant measures, as those proposed by Ley and Steel (2007), there exists no uniquely defined threshold which allows for a categorisation into substitutes and complements in terms of inclusion. Moreover in the case of null-invariance this threshold depends on the marginal inclusion probabilities of the individual covariates, $p(A)$ and $p(B)$.

P6. Commutative symmetry: Measures that have a different value associated with $A \to B$ than with $B \to A$ are asymmetric. Since for BMA output analysis, jointness measures aim at measuring the degree of joint appearance of two explanatory variables in a model (or lack thereof) a posteriori, a suitable measure should be symmetric with regard to the ordering of variables. The symmetric assertion that certain pairs of covariates are "substitutes" or "complements" thus requires commutativity. All jointness measures proposed in the BMA literature fulfil this requirement. However, this property is discussed heavily in the data mining literature. While Tan et al. (2004)[3] and Wu et al. (2007) argue in favour of such an adjustment, Glass (2013) strictly opposes this property. In the context of BMA, where only the joint inclusion or exclusion of variables is of interest, commutativity appears to be a desirable property.[4]

P7. Hypothesis symmetry: Glass (2013) proposes hypothesis symmetry as a desirable property for association measures. Hypothesis symmetry states that $\tilde{J}(A, B) = -\tilde{J}(\overline{A}, B)$. In the context of BMA if A and B are complements to some degree, then $\overline{A}$ and B should be substitutes to the same degree, a property which facilitates the interpretation of a jointness measure. Together with the desired property of commutative symmetry, hypothesis symmetry also implies that $J(A, B) = -J(A, \overline{B})$, a property which is called evidence symmetry in the association literature (see Glass, 2013).

[Table 2 about here.]

3.2. An Illustration of the Properties of Jointness Measures

[Table 3 about here.]

The measures proposed in the BMA literature as well as Yule's Q measure do not fulfil all criteria suggested to be desirable for jointness measures. This leads to unfavourable results under certain specifications which are illustrated by making use of exemplary two-way contingency tables. We analyse the contingency tables given in Tables 3a–3f, which are assumed to contain the joint covariate inclusion/exclusion events recorded during the MCMC analysis in BMA, using the measures J_{DW1}, J_{DW2}, J_{LS1}, J_{LS2}, J_{St} and J_{YQ}.

[3]Tan et al. (2004) suggest to symmetrise any measure by using $\max(p(A|B, y), p(B|A, y))$.

[4]Note that in the context of market-basket analysis, which is a typical area of application for association rules, whether A implies B or vice versa might be of interest and therefore commutativity might not be desirable.

Table 3a displays an independent relationship between two variables A and B, while Table 3b is an example for strong positive jointness. As expected, for Table 3a, J_{DW2} as well as J_{YQ} equal zero, indicating independence of the considered variables, while the measures proposed by Ley and Steel (2007) and Strachan (2009) return values above zero. For Table 3b, J_{DW1} delivers a value practically equal to zero, while J_{DW2} is $\ln(0) = -\infty$, indicating a strong negative jointness between the inclusion of A and B. The results of both measures are thus misleading, as they indicate either no jointness or a counterintuitive value of the measure. The measures proposed by Ley and Steel (2007) capture the jointness correctly and both measures tend to reach their maxima for Table 3b. The same is true for the measure proposed by Strachan (2009). J_{YQ}, however, delivers a misleading negative value in this example.

Table 3c exchanges the inclusion and exclusion of variables compared to Table 3b. While J_{YQ} and J_{DW2} provide the same jointness measure values for Tables 3b and 3c, the measures proposed by Ley and Steel (2007) equal zero. Note that, replacing the zero ($a = 0$) entry in Table 3c with, for example, 10, as illustrated in Table 3d changes the results dramatically for J_{DW1}, J_{DW2} and J_{YQ}. Now, all three measures return a strong degree of positive jointness between A and B. On the other hand, according to J_{LS1} and J_{LS2} the same degree of jointness can be observed in Tables 3a and 3d. The different treatment of *null-invariance* is the essential property leading to these differences in results between Tables 3a and 3d. Furthermore for Table 3d J_{St} takes a value close to zero because the marginal probabilities are small for this table.

Further examples are given in Tables 3e and 3f. Compared to Table 3a, Table 3e displays a two-way table for independent events of inclusion of covariates A and B, where the inclusion probability of B is larger than that of its counterpart. As desired, the measures proposed by Doppelhofer and Weeks (2009a) as well as the Yule's Q measure are equal to zero for both Tables 3a and 3e, but the measures proposed by Ley and Steel (2007) differ for Tables 3a and 3e. This makes comparison of J_{LS1} and J_{LS2} across different marginal distributions quite difficult. For the former table J_{LS2} for example, returns a value of $\frac{1}{2}$ while for the latter example it returns $\frac{2}{3}$. Finally, Table 3f illustrates a contingency table where one variable is always included. Along the lines of Ley and Steel (2007) a jointness measure should be defined in such a case. The measures J_{DW2} and J_{YQ} are not defined, but the measures proposed by Ley and Steel (2007), J_{DW1} and J_{St} are.

3.3. The Modified Yule's Q, J_{YQM}

In order to overcome the problems illustrated by the existing jointness measures in the BMA literature, we propose a new measure which is a simple modification of Yule's Q, the modified Yule's Q (henceforth J_{YQM}). Essentially, J_{YQM} is a regularised version of J_{YQ}, where zero entries in the contingency table are avoided by augmenting the table with virtual counts as illustrated in Table 4.

The observed contingency table can be interpreted as representing the counts (a, b, c, d) for realisations from a multinomial distribution with parameter vector $\pi = (p(AB), p(\overline{A}B), p(A\overline{B}), p(\overline{AB}))$. Table 4 can then be interpreted as the posterior mode estimates for data from the multinomial distribution combined with a Dirichlet prior on π. The Dirichlet distribution with parameter α is the conjugate prior for the multinomial and the resulting posterior distribution follows a Dirichlet distribution $D(\alpha_1 + a, \alpha_2 + b, \alpha_3 + c, \alpha_4 + d)$. Put differently, if we base the bivariate jointness analysis on the result of a multinomial model in combination with a conjugate Dirichlet prior distribution we end up with a "posterior" contingency table such as the one illustrated in Table 4.

[Table 4 about here.]

Several choices of α_k for $k \in \{1, \ldots, 4\}$ might be appealing. These can be interpreted as virtual counts for the table entries before observing the actual realisations of a, b, c, d. Setting $\alpha_k = 1$ for all k corresponds to using a uniform prior with prior sample size equal to 4. For our modified Yule's Q, J_{YQM}, we propose to use the Jeffreys prior (Jeffreys, 1946) which equals $\alpha_k = \frac{1}{2}$ for all k. Gill (2014) argues that the Jeffreys prior has the main advantage that it is obtained from a mechanical procedure which results almost always in an uninformative prior. In fact in case of 2×2 tables the Jeffreys prior corresponds to adding two observations with all cells of the table being equally likely to occur. In addition Firth (1993) shows that the Jeffreys prior can be used to reduce the bias of maximum likelihood estimates in exponential family models. In the context of latent class analysis Galindo Garre and Vermunt (2006) also empirically indicate the general good performance of Bayesian posterior mode estimates resulting from the Jeffreys prior compared to maximum likelihood estimates. Except for a correction factor in the denominator to guarantee maximality (see Section 3.1), J_{YQM} corresponds to the Yule's Q measure estimated on the augmented contingency table and has the form

$$J_{\text{YQM}} = \frac{(a+\frac{1}{2})(d+\frac{1}{2}) - (b+\frac{1}{2})(c+\frac{1}{2})}{(a+\frac{1}{2})(d+\frac{1}{2}) + (b+\frac{1}{2})(c+\frac{1}{2}) - \frac{1}{2}}, \tag{16}$$

where the last term in the denominator denotes the correction factor.

Usually, when $N = a + b + c + d$ is large, as is the case in BMA applications, and there are no zeros in the contingency tables, the effect of Jeffreys prior vanishes. The prior effect only becomes relevant when (a) N is very small, which is usually not the case for BMA applications or (b) if zeros are present in the contingency table. The influence of N on the measure values using either J_{YQ} or the modified versions with Jeffreys prior (YQM_J with $\alpha_k = \frac{1}{2}$) or a uniform prior (YQM_U with $\alpha_k = 1$) is shown in Figure 1 for five different tables. These tables correspond to the cases of complete positive or complete negative association, no association and half-positive or half-negative association with balanced success probabilities for the two events A and $\overline{A}$ or B and $\overline{B}$. J_{YQ} gives the same values for each case, which center around ± 1, 0 and ± 0.5, regardless of N. The modified versions return the same measure values as J_{YQ} in case of complete or no association, and converge to the values returned by J_{YQ} rather quickly if the association is only partly positive or negative. The stronger regularisation induced by the uniform compared to the Jeffreys prior is visible by these measure values being relatively shrunk towards zero.

If zeros are present in the contingency table, the virtual counts added by the modified versions correspond to a correction for continuity to avoid pathological results as illustrated in Tables 3b, 3c and 3f. The results obtained for J_{YQ}, as well as the modified versions using either Jeffreys or a uniform prior for the six tables in Table 3, are given in Figure 1 on the right. The three measures give the same results for the non-pathological cases in Tables 3a, 3d and 3e. For the pathological cases, the unmodified version gives completely different and misleading results or is not defined, while the same values are obtained for the modified versions regardless of the correction factor added.

[Figure 1 about here.]

4. Jointness of Economic Growth Determinants Revisited

We illustrate the use of the modified Yule's Q measure as a jointness indicator and assess the differences observed when compared to the ordinary Yule's Q measure on the BMA results when analysing the dataset on the determinants of differences in economic growth across countries used in Fernández et al. (2001a).

The dataset provided by Fernández et al. (2001a), which is partly based on the data collection by Sala-i Martin (1997), contains information of GDP per capita growth and 41 potential explanatory variables for 72 countries. The variables included in the dataset are summarised in Table 5. With GDP per capita growth as a dependent variable, we apply BMA to the dataset employing a hyper-g prior over the parameters associated to the covariates, as proposed by Liang et al. (2008) and a beta-binomial prior for the inclusion of covariates in order to create the prior over the individual specifications, in line with Ley and Steel (2009b). The MCMC procedure is carried out with the R package **BMS** (Zeugner and Feldkircher, 2015) using 5 million iterations with one million burn-in runs.[5] The jointness measures given by J_{YQ} and J_{YQM} are computed based on the inclusion/exclusion patterns for all 820 pairs of variables in the models sampled using MCMC.

[Table 5 about here.]

Figure 2 presents histograms for both the values obtained for J_{YQ} and J_{YQM} over all variable pairs. The colour of the bars indicates the degree to which the pairs are evaluated as substitutes or complements in terms of joint inclusion, with the threshold between substitutes and complements indicated in white. A higher number of extreme jointness values are observed in J_{YQM} as compared to the standard Yule's Q measure. This is mainly driven by the fact that Yule's Q is not defined if one of the variables in the pair is always or never included in the models sampled by the MCMC procedure. The modified measure, J_{YQM}, however, allows to obtain an estimate of the the degree of jointness for these 68 pairs of variables. Out of those 68 pairs of variables for which Yule's Q is not defined, 20% indicate a strong negative jointness with J_{YQM} values between -1 and -0.8. For 23% of those pairs we observe a J_{YQM} between -0.8 and -0.3 and only 9% are near to statistically independence with jointness degrees between -0.3 and 0.3. Inspecting the complements gives similar results: 13% of those 68 pairs have a J_{YQM} between 0.3 and 0.8 and 35% result in jointness values above 0.8. The fact that J_{YQM} fulfils the definition property appears particularly important in this application, since the variables which are included in all sampled models (for which J_{YQ} is not defined) correspond to the most important variables in terms of PIP. For the dataset in Fernández et al. (2001a) these are the initial GDP per capita (*GDP60*), the fraction of the population of Confucian religion (*Confucian*), life expectancy (*LifeExp*), and the dummy variable for Sub-Saharan countries (*SubSahara*).

[5] The PIP of the individual variables, as well as the mean and standard deviation of the posterior distribution over the parameters associated to the covariates, are in line with those reported in previous BMA applications using this dataset. They are available from the authors upon request.

Figures 3a and 3b present graphically the degree of jointness implied by Yule's Q and J_{YQM} using mosaic plots. The pairs of variables are ordered in such a way to minimize the difference between jointness values of neighbouring elements (Hahsler et al., 2008). To determine the ordering the results for the Yule's Q measure are used and the same ordering is applied to the results for the modified Yule's Q measure results. This facilitates the comparison of results between the two mosaic plots. Pairs of variables which correspond to complements are represented by blue shaded tiles, while substitutes are indicated by red tiles.

As expected, these graphs depict a strong correspondence between the results of Yule's Q and its modified version although some important differences can be discerned. While the correlation in jointness measured by J_{YQ} and J_{YQM} is over 0.99 for the pairs of variables for which both indicators deliver measurements, the jointness level measured by J_{YQM} for the pairs involving the four covariates with undefined J_{YQ} measures tend to be at the extreme of the distribution (see the difference in histograms in Figure 2). The *GDP60* variable tends to act as a correlate of initial conditions and as such pairs of variables including this covariate usually present high negative values of J_{YQM}. Relevant exceptions are the pairs (*GDP60*, *LifeExp*), (*GDP60*, *Confucian*) or (*GDP60*, *SubSahara*), which include variables with extremely high PIPs and high positive jointness. The strong relationship of complementary inclusion existing among *GDP60*, *Confucian* and *LifeExp* is in line with the results presented in Ley and Steel (2007).

[Figure 2 about here.]

[Figure 3 about here.]

5. Conclusion

BMA has often been used in cross-country and panel regressions to account for model uncertainty and identify robust determinants of differences in economic growth across countries. Usually the importance of single covariates in the regression is measured by the PIP of the corresponding variables. PIPs do not allow to determine whether two covariates tend to appear together or tend to exclude each other within the set of considered regression specifications.

To gain insights into interdependencies in covariate inclusion patterns, bivariate jointness measures have been proposed in the economic literature as suitable tools for post-processing BMA output. A range of different measures has been put forward, with each of them fulfilling different properties and criteria suggested to be important. In this paper these properties are investigated and combined to arrive at an ultimate list of desirable characteristics of bivariate jointness measures. Since none of the existing measures meet all requirements, a new measure resulting from a modification of the Yule's Q measure is proposed. This modified Yule's Q can be interpreted as a regularised version given by a posterior estimate of the observed two-way inclusion contingency table combined with a non-informative Jeffreys prior.

In addition, we apply the proposed measure to a dataset previously used in the empirical economic growth literature on jointness measures and compare the results to those obtained for the original version of Yule's Q. This empirical illustration allows to assess which insights can be gained when employing a measure which fulfils all desirable properties of a jointness measure in a BMA post-processing analysis.

The analysis of bivariate jointness is only the starting point for investigating the interdependencies across covariate inclusion patterns in BMA analysis. Jointness measures for larger sets of variables could be of interest and Ley and Steel (2007) point out that generalisations of bivariate joint measures could be considered in this case, despite the drawback that results might often be difficult to communicate. An alternative approach to investigate the interdependence structure of covariate inclusion in BMA has been proposed in Crespo Cuaresma et al. (2016). Crespo Cuaresma et al. (2016) aim at succinctly and comprehensibly describing the dependence structure of inclusion across all variables in the model space using Dirichlet process clustering.

References

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). A comprehensive approach to posterior jointness analysis in bayesian model averaging applications. *WU Vienna - Department of Economics Working Paper Series*, (193).

Crespo Cuaresma, J., Grün, B., Hofmarcher, P., Humer, S., and Moser, M. (2016). Unveiling covariate inclusion structures in economic growth regressions using latent class analysis. *European Economic Review*, 81:189–202.

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

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.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1):27–38.

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

Galindo Garre, F. and Vermunt, J. K. (2006). Avoiding boundary estimates in latent class analysis by Bayesian posterior mode estimation. *Behaviormetrika*, 33(1):43–59.

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

Gill, J. (2014). *Bayesian Methods: A Social and Behavioral Sciences Approach*. Chapman & Hall/CRC, New York, 3rd edition.

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

Glass, D. H. (2014). Entailment and symmetry in confirmation measures of interestingness. *Information Sciences*, 279:552 – 559.

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.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 14(4):382–401.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 186(1007):453–461.

Leamer, E. (1978). *Specification Searches. Ad Hoc Inference with Nonexperimental Data*. Wiley.

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.

Mitchell, T. and Beauchamp, J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023–32.

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. (2016). Growth empirics in panel data under model uncertainty and weak exogeneity. *Journal of Applied Econometrics*, 31(3):584–602.

Sala-i Martin, X. (1997). I just ran two million regressions. *The American Economic Review*, 87(2):178–183.

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.

Wu, T., Chen, Y., and Han, J. (2007). Association mining in large databases: A re-examination of its measures. In *Knowledge Discovery in Databases: PKDD 2007*, pages 621–628. 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.

Zeugner, S. and Feldkircher, M. (2015). Bayesian model averaging employing fixed and flexible priors: The BMS package for R. *Journal of Statistical Software*, 68(4):1–37.

	B	$\overline{B}$	$\sum$
A	a	b	$(a+b)$
$\overline{A}$	c	d	$(c+d)$
$\sum$	$(a+c)$	$(b+d)$	$a+b+c+d=N$

Note: $\overline{A}$ ($\overline{B}$) indicates that event A (B) did not occur.

Table 1: General contingency table for two binary variables.

Property		**Jointness Measure**						
		J_{DW1}	J_{LS1}	J_{LS2}	J_{St}	J_{DW2}	J_{YQ}	J_{YQM}
Definition	P1		✓	✓				✓
Monotonicity & Maximality	P2							✓
Limiting Behaviour	P3							✓
Bayesian Confirmation	P4	✓				✓	✓	✓
Non Null-Invariance	P5	✓			✓	✓	✓	✓
Commutative Symmetry	P6	✓	✓	✓	✓	✓	✓	✓
Hypothesis Symmetry	P7					✓	✓	✓

Table 2: Properties of bivariate jointness measures.

(a)

	B	$\overline{B}$
A	10	10
$\overline{A}$	10	10

(b)

	B	$\overline{B}$
A	10^6	10
$\overline{A}$	10	0

(c)

	B	$\overline{B}$
A	0	10
$\overline{A}$	10	10^6

(d)

	B	$\overline{B}$
A	10	10
$\overline{A}$	10	10^6

(e)

	B	$\overline{B}$
A	10	5
$\overline{A}$	10	5

(f)

	B	$\overline{B}$
A	10	10
$\overline{A}$	0	0

Table 3: Exemplary contingency tables to illustrate properties of jointness measures.

	B	$\overline{B}$	$\sum$
A	$a+\alpha_1$	$b+\alpha_2$	$(a+b+\alpha_1+\alpha_2)$
$\overline{A}$	$c+\alpha_3$	$d+\alpha_4$	$(c+d+\alpha_3+\alpha_4)$
$\sum$	$(a+c+\alpha_1+\alpha_3)$	$(b+d+\alpha_2+\alpha_4)$	$(a+b+c+d+\sum_{k=1}^{4}\alpha_k)$

Table 4: Augmented contingency table for the modified Yule's Q measure.

	Abbreviation	Variable	Mean	Std. Dev.
1	Abslat	Absolute latitude	25.73	17.250
2	Age	Age	23.71	37.307
3	Area	Area (Scale Effect)	972.92	2051.976
4	BlMktPm	Black Market Premium	0.16	0.291
5	Brit	British Colony dummy	0.32	0.470
6	Buddha	Fraction Buddhist	0.06	0.184
7	Catholic	Fraction Catholic	0.42	0.397
8	CivlLib	Civil Liberties	3.47	1.712
9	Confucian	Fraction Confucian	0.02	0.087
10	EcoOrg	Degree of Capitalism	3.54	1.266
11	English	Fraction of Pop. Speaking English	0.08	0.239
12	EquipInv	Equipment investment	0.04	0.035
13	EthnoL	Ethnolinguistic fractionalization	0.37	0.296
14	Foreign	Fraction speaking foreign language	0.37	0.422
15	French	French Colony dummy	0.12	0.333
16	GDP60	GDP level in 1960	7.49	0.885
17	HighEnroll	Higher education enrollment	0.04	0.052
18	Hindu	Fraction Hindu	0.02	0.101
19	Jewish	Fraction Jewish	0.01	0.097
20	LabForce	Size labor force	9305.38	24906.056
21	LatAmerica	Latin American dummy	0.28	0.451
22	LifeExp	Life expectancy	56.58	11.448
23	Mining	Fraction GDP in mining	0.04	0.077
24	Muslim	Fraction Muslim	0.15	0.295
25	NequipInv	Non-Equipment Investment	0.15	0.055
26	OutwarOr	Outward Orientation	0.39	0.491
27	PolRights	Political Rights	3.45	1.896
28	Popg	Population Growth	0.02	0.010
29	PrExports	Primary exports, 1970	0.67	0.299
30	Protestants	Fraction Protestant	0.17	0.252
31	PrScEnroll	Primary School Enrollment, 1960	0.80	0.246
32	PublEdupct	Public Education Share	0.02	0.009
33	RevnCoup	Revolutions and coups	0.18	0.238
34	RFEXDist	Exchange rate distortions	121.71	41.001
35	RuleofLaw	Rule of law	0.55	0.335
36	Spanish	Spanish Colony dummy	0.22	0.419
37	stdBMP	SD of black-market premium	45.60	95.802
38	SubSahara	Sub-Saharan dummy	0.21	0.409
39	WarDummy	War dummy	0.40	0.494
40	WorkPop	Ratio workers to population	−0.95	0.189
41	y = Economic growth	GDP per capita growth	0.02	0.018
42	YrsOpen	Number of Years open economy	0.44	0.355

Table 5: Variable names and descriptive statistics – Fernández et al. (2001a) dataset.

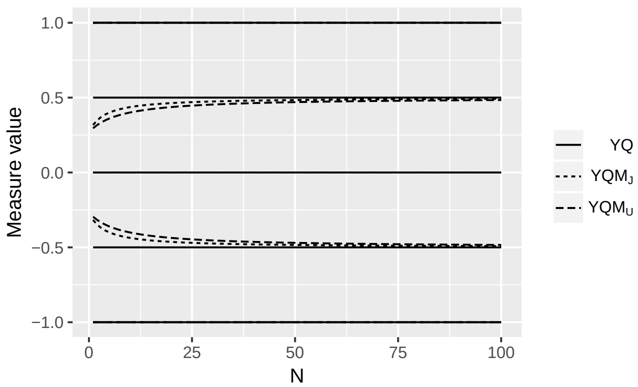


Table	YQ	YQM_J	YQM_U
(a)	0.00	0.00	0.00
(b)	−1.00	1.00	1.00
(c)	−1.00	1.00	1.00
(d)	1.00	1.00	1.00
(e)	0.00	0.00	0.00
(f)	−	0.00	0.00

Figure 1: Left: Influence of N on J_{YQ} and the modified versions thereof, using either Jeffreys prior (YQM_J with $\alpha_k = \frac{1}{2}$) or a uniform prior (YQM_U with $\alpha_k = 1$). Right: Measure values for J_{YQ} and the modified versions using either Jeffreys (YQM_J) or a uniform (YQM_U) prior, based on the exemplary contingency tables given in Table 3.

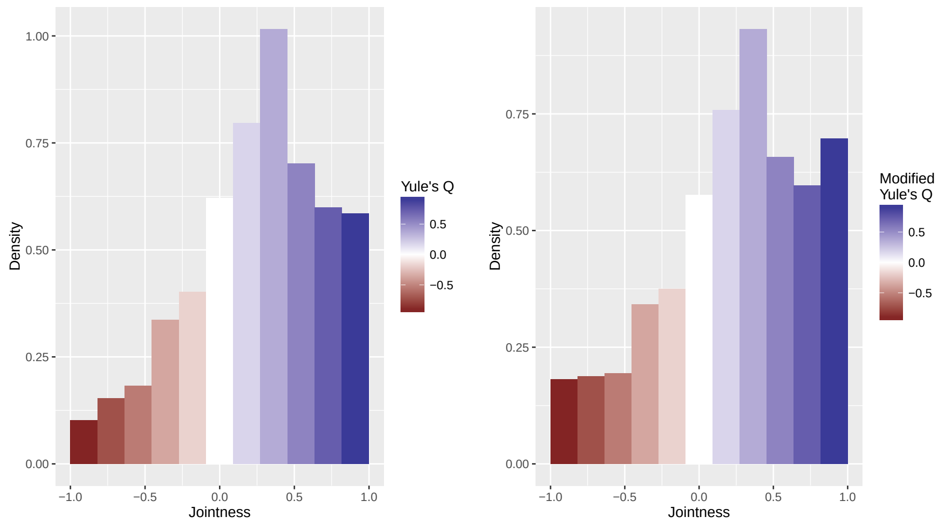
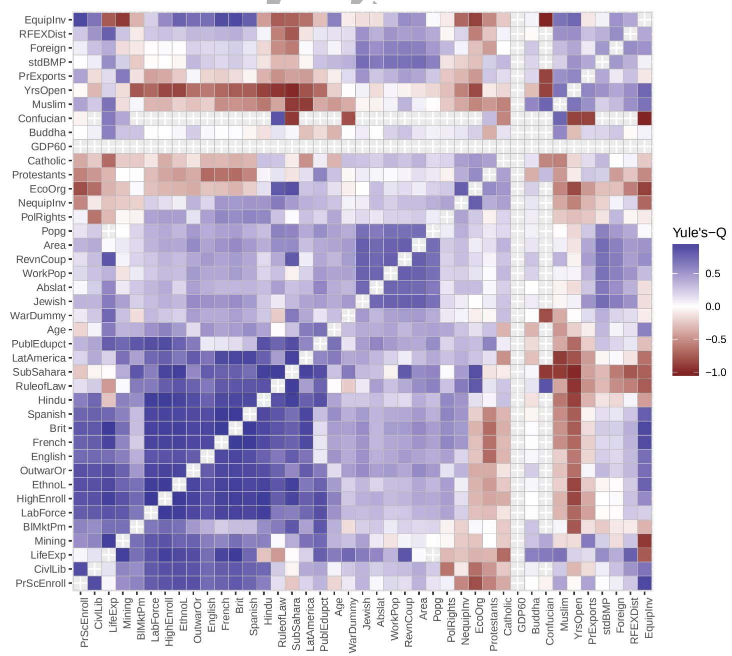
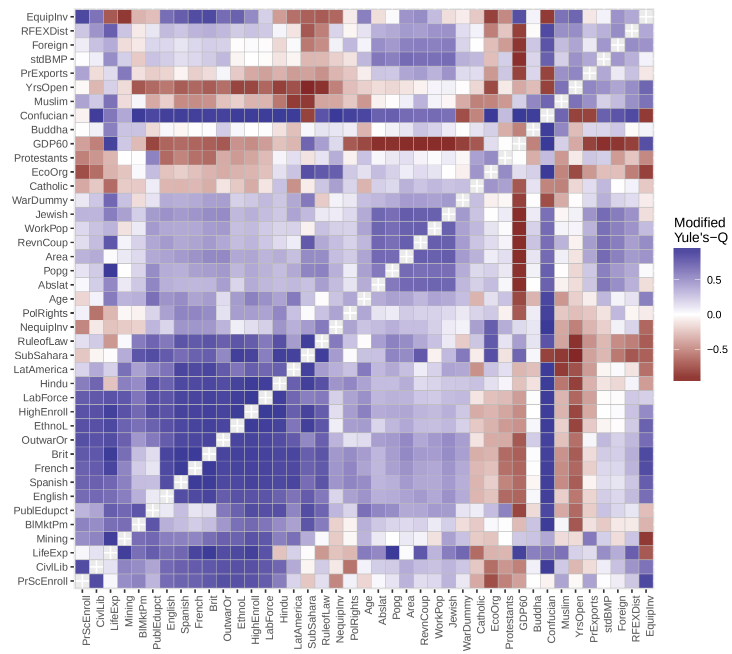


Figure 2: Histograms of obtained (finite) Yule's Q and modified Yule's Q measure values for all pairs of variables.



(a) Yule's Q.



(b) Modified Yule's Q.

Figure 3: Mosaic plots for Yule's Q (left) and modified Yule's Q (right).

A. Appendix: Properties of Jointness Measures

A.1. Jointness Measure DW1, J_{DW1}

$$\tilde{J}_{\mathrm{DW1}} = \ln\left(\frac{p(A \cap B)}{p(A)p(B)}\right)$$
$$J_{\mathrm{DW1}} = \ln\left(\frac{a}{(a+b)(a+c)/N}\right)$$

P1. Definition: J_{DW1} is not defined when $a = b = 0$, i.e., A is never included, or $a = c = 0$, i.e., B is never included.

P2. Monotonicity, boundedness and maximality: J_{DW1} fails to meet the criterion of boundedness as it converges to $-\infty$ when $b = c$ and both $\to \infty$ and to $+\infty$ when $d \to \infty$.

P3. Limiting behaviour: J_{DW1} converges to its maximum for $d \to \infty$. However it converges to 0 for $a \to \infty$. For $b \to \infty$ the measure converges to $\ln(a/(a+c))$ and analogously to $\ln(a/(a+b))$ for $c \to \infty$.

P4. Bayesian confirmation: It holds that

$$\frac{a(a+b+c+d)}{(a+b)(a+c)} = 1 \quad \Longleftrightarrow \quad ad - bc = 0$$

and $p(A|B) = p(A)$ iff $ad - bc = 0$.

It is straightforward to show that the corresponding inequalities of the Bayesian confirmation property are similarly fulfilled.

P5. Non null-invariance: Since $J_{\mathrm{DW1}} = \ln\left(\frac{aN}{(a+c)(a+b)}\right)$ with $N = a + b + c + d$, this measure obviously depends on d, the number of counts of joint exclusion.

P6. Commutative symmetry: This follows directly from the definition of J_{DW1}.

P7. Hypothesis symmetry: Consider the case where $a = 0$. Then $J(A, B) \to -\infty$. However, $J(\bar{A}, B) = \ln((b + c + d)/(c + d))$ which only equals ∞ if $c = d = 0$ with $b > 0$.

A.2. Jointness Measure DW2, J_{DW2}

This jointness measure is defined as

$$\tilde{J}_{\mathrm{DW2}} = \ln\left(\frac{p(A\cap B)p(\bar{A}\cap\bar{B})}{p(A\cap\bar{B})p(\bar{A}\cap B)}\right)$$
$$J_{\mathrm{DW2}} = \ln\left(\frac{ad}{bc}\right)$$

P1. Definition: J_{DW2} is not defined if one of the variables is always or never include, e.g., $p(A) = 1$ and $c = d = 0$.

P2. Monotonicity, boundedness and maximality: Obviously J_{DW2} is not bounded, monotonically increasing in a and d and decreasing in b and c. The maximality criterion is not met as it is sufficient that only one of the entries equals zero for the measure to reach an extreme value.

P3. Limiting behaviour: For $a \to \infty$ and $d = 0$, J_{DW2} is not defined.

P4. Bayesian confirmation: This is obvious because if $ad = bc$, the measure equals 0. It is positive if $ad > bc$ and negative if $ad < bc$.

P5. Non null-invariance: The measure depends on d.

P6. Commutative symmetry: Trivial.

P7. Hypothesis symmetry: It is trivial to show that $J(A, B) = -J(\bar{A}, B) = \ln\left(\frac{ad}{bc}\right)$.

A.3. Jointness Measure LS1, J_{LS1}

$$\tilde{J}_{\text{LS1}} = \frac{p(A \cap B)}{p(A \cup B)}$$
$$J_{\text{LS1}} = \frac{a}{a+b+c}$$

P1. Definition: J_{LS1} is defined whenever one variable is included with positive probability.

P2. Monotonicity, boundedness and maximality: This measure is bounded on $[0, 1]$ but fails to meet the minimality condition; i.e., to reach its minimum only when a *and* d are zero. This measure is independent of d and reaches its minimum whenever $a = 0$. It is constant in d, but meets the monotonicity condition.

P3. Limiting behaviour: For $d \to \infty$, J_{LS1} does not converge to its maximum.

P4. Bayesian confirmation: Given that the range is $[0, 1]$ the Bayesian confirmation criterion cannot be fulfilled.

P5. Non null-invariance: The measure is independent of d and therefore null-invariant.

P6. Commutative symmetry: By definition.

P7. Hypothesis symmetry: It follows directly from the fact that $J_{\text{LS1}} \geq 0$, that $J(A, B) = -J(\bar{A}, B)$ cannot hold.

A.4. Jointness Measure LS2, J_{LS2}

$$\tilde{J}_{\text{LS2}} = \frac{p(A \cap B)}{p(A \cap \bar{B}) + p(\bar{A} \cap B)}$$
$$J_{\text{LS2}} = \frac{a}{b + c}$$

P1. Definition: J_{LS2} is defined whenever one variable is included with positive probability.

P2. Monotonicity, boundedness and maximality: $J_{\text{LS2}} \in [0, \infty)$ and is therefore not bounded. It is constant in d and thus does not meet the maximality criterion as the minimum value of 0 is reached if $a = 0$ regardless of the value of d. The monotonicity criterion is fulfilled.

P3. Limiting behaviour: For $d \to \infty$, J_{LS2} does not converge to its maximum.

P4. Bayesian confirmation: $J_{\text{LS2}} \geq 0$ and thus fails to fulfil the Bayesian confirmation condition.

P5. Non null-invariance: The measure is independent of d and therefore null-invariant.

P6. Commutative symmetry: Trivial.

P7. Hypothesis symmetry: Since $J_{\text{LS2}} \geq 0$, $J(A, B) = -J(\bar{A}, B)$ cannot hold.

A.5. Jointness Measure St, J_{St}

$$\tilde{J}_{\text{St}} = p(A)p(B)\ln\left(\frac{p(A\cap B)}{p(A\cap\bar{B})p(\bar{A}\cap B)}\right)$$
$$J_{\text{St}} = \frac{a+b}{N}\frac{a+c}{N}\ln\left(\frac{aN}{bc}\right).$$

P1. Definition: J_{St} is not defined if $p(A)=0$, i.e., $a=b=0$, or $p(B)=0$, i.e., $a=c=0$.

P2. Monotonicity, boundedness and maximality: This jointness measure is not bounded as for $a=0$ and $b,c>0$ it takes the value $J_{\text{St}}=-\infty$, while for $a>0$ and $b=0$ and/or $c=0$, $J_{\text{St}}=\infty$. The measure is monotonically increasing in a, but not necessarily in d and not necessarily decreasing in b or d.

P3. Limiting behaviour: For $d\to\infty$, it is straightforward that $J_{\text{St}}\to 0$.

P4. Bayesian confirmation: In the case of independence, i.e., $ad=bc$, $(aN)/(bc)=N/d$ and thus the measure is positive.

P5. Non null-invariance: The measure depends on $N=a+b+c+d$ and thus on d.

P6. Commutative symmetry: Trivial.

P7. Hypothesis symmetry: A simple counterexample can be constructed using a contingency matrix. Consider the following case, where already relative frequencies are given in the table below as the measure does not depend on the absolute sum of the entries:

	B	$\bar{B}$	$\sum$
A	0.3	0.4	0.7
$\bar{A}$	0.1	0.2	0.3
$\sum$	0.4	0.6	1

Then $J(A,B) = 0.7 \cdot 0.4 \ln(\frac{0.3}{0.4 \cdot 0.1}) = 0.562$, but $-J(\bar{A},B) = 0.3 \cdot 0.4 \ln(\frac{0.1}{0.2 \cdot 0.3}) = 0.061$.

A.6. Jointness Measure Yule's Q, J_{YQ}

$$\tilde{J}_{\text{YQ}} = \frac{p(A \cap B)p(\bar{A} \cap \bar{B}) - p(A \cap \bar{B})p(\bar{A} \cap B)}{p(A \cap B)p(\bar{A} \cap \bar{B}) + p(A \cap \bar{B})p(\bar{A} \cap B)}$$
$$J_{\text{YQ}} = \frac{ad - bc}{ad + bc}$$

P1. Definition: J_{YQ} is not defined if either of the variables is always or never included, e.g., for $p(A) = 1$ and $c = d = 0$.

P2. Monotonicity, boundedness and maximality: J_{YQ} reaches its extreme value whenever *one* of the diagonal/off-diagonal entries is zero. The boundedness criterion is met by definition. The monotonicity condition is also fulfilled.

P3. Limiting behaviour: For $a \to \infty$ and $d = 0$, $J_{\text{YQ}} \to -1$, instead of converging to 1. The measure fulfils the limiting behaviour criterion if $d > 0$.

P4. Bayesian confirmation: Follows directly from the definition of a Bayesian confirmation measure because $p(A|B) = p(A)$ iff $ad - bc = 0$.

P5. Non null-invariance: J_{YQ} is a function of d and thus not null-invariant.

P6. Commutative symmetry: Trivial.

P7. Hypothesis symmetry: $J_{\text{YQ}}(\bar{A}, B) = \frac{bc-ad}{bc+ad} = -J_{\text{YQ}}(A, B)$.

A.7. Modified Yule's Q, J_{YQM}

$$J_{\text{YQM}} = \frac{(a+\frac{1}{2})(d+\frac{1}{2}) - (b+\frac{1}{2})(c+\frac{1}{2})}{(a+\frac{1}{2})(d+\frac{1}{2}) + (b+\frac{1}{2})(c+\frac{1}{2}) - \frac{1}{2}},$$

P1. Definition: The original Yule's Q is not defined if either a or d and either b or c are equal to zero. For the modified Yule's Q the adjustment ensures definition of the jointness measure.

P2. Monotonicity, boundedness and maximality: First, we discuss monotonicity: We prove that for given values b, c, J_{YQM} is monotonically increasing in a and d. Note that J_{YQM} can be rewritten as

$$J_{\text{YQM}} = \frac{x-y}{x+y-\frac{1}{2}}, \tag{17}$$

with $x = (a+\frac{1}{2})(d+\frac{1}{2})$ and $y = (b+\frac{1}{2})(c+\frac{1}{2})$.

This gives

$$J_{\text{YQM}} = \frac{x-y}{x+y-\frac{1}{2}} = \frac{x(1-\frac{y}{x})}{x(1+\frac{y}{x}-\frac{1}{2x})} = \frac{1-\frac{y}{x}}{1+\frac{y}{x}-\frac{1}{2x}}. \tag{18}$$

The denominator of equation (18) can be rewritten as $1+\frac{y}{x}-\frac{1}{2x} = 1+\frac{1}{x}(y-\frac{1}{2})$. Note that $y = bc + \frac{1}{2}c + \frac{1}{2}b + \frac{1}{4}$ and that $(y-\frac{1}{2}) > 0$ whenever b or c is greater than zero, while for $b = c = 0$, $J_{\text{YQM}} = 1$. This implies that J_{YQM} is monotonically increasing in x for b or c greater than zero as the numerator in equation (18) is increasing in x and the denominator is decreasing in x.

The boundedness as well as maximality conditions are trivially fulfilled. The modified Yule's Q takes values in $[-1, 1]$, the minimum value of -1 is attained if $a = d = 0$ and the maximum value if $b = c = 0$.

P3. Limiting behaviour: This follows directly from the fact that J_{YQM} can be rewritten as

$$\frac{x}{x+y-\frac{1}{2}} - \frac{y}{x+y-\frac{1}{2}}. \tag{19}$$

If $x \to \infty$, this expression converges to 1. Analogously J_{YQM} converges to -1 for $y \to \infty$.

P4. Bayesian confirmation: First, we prove the independency property $J(A, B) = 0 \iff p(A \cap B) = p(A)p(B)$, i.e.,

$$\frac{(a+\frac{1}{2})}{N+2} = \frac{(a+\frac{1}{2})+(b+\frac{1}{2})}{N+2}\,\frac{(a+\frac{1}{2})+(c+\frac{1}{2})}{N+2}. \tag{20}$$

Multiplying equation (20) with $(N+2)^2$ and using that $N = a+b+c+d$ results in

$$ad + \frac{1}{2}a + \frac{1}{2}d = bc + \frac{1}{2}b + \frac{1}{2}c, \tag{21}$$

which corresponds to $J_{\mathrm{YQM}} = 0$.

Analogously, if $p(A \cap B) > p(A)p(B)$ the left hand side of equation (21) dominates its right hand side, which implies a strictly positive J_{YQM}.

P5. Non null-invariance: J_{YQM} depends on d.

P6. Commutative symmetry: $J(A, B) = J(B, A)$ follows directly from the definition of J_{YQM}.

P7. Hypothesis symmetry: Hypothesis symmetry states that $J(A, B) = -J(\overline{A}, B)$.

Note that

$$\begin{aligned} J(A,B) &= \frac{(a+\frac{1}{2})(d+\frac{1}{2}) - (c+\frac{1}{2})(b+\frac{1}{2})}{(a+\frac{1}{2})(d+\frac{1}{2}) + (c+\frac{1}{2})(b+\frac{1}{2}) - \frac{1}{2}} \\ &= -\frac{(b+\frac{1}{2})(c+\frac{1}{2}) - (a+\frac{1}{2})(d+\frac{1}{2})}{(a+\frac{1}{2})(d+\frac{1}{2}) + (c+\frac{1}{2})(b+\frac{1}{2}) - \frac{1}{2}} = -J(\bar{A}, B) \end{aligned}$$

B. Appendix: Review of Jointness measures

Table 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	$N \times (p(AB) - p(A)p(B))$

Table 1: *(continued)*

	#	Measure	Value
31	rr	Relative Risk	$\ln\left[\frac{p(B\|A)}{p(B\|\bar{A})}\right]$
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	$\frac{\sqrt{p(AB)p(\bar{A}\bar{B})}-\sqrt{p(A\bar{B})p(\bar{A}B)}}{\sqrt{p(AB)p(\bar{A}\bar{B})}+\sqrt{p(A\bar{B})p(\bar{A}B)}}$