# Inbound Open Innovation and Innovation Performance

## A Robustness Study

Ebersberger, Bernd; Galia, Fabrice; Laursen, Keld; Salter, Ammon

*Document Version*
Final published version

*Published in:*
Research Policy

*DOI:*
10.1016/j.respol.2021.104271

*Publication date:*
2021

Research Policy 50 (2021) 104271

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

# Inbound Open Innovation and Innovation Performance: A Robustness Study

Bernd Ebersberger[a,*], Fabrice Galia[b], Keld Laursen[c], Ammon Salter[d]

[a] University of Hohenheim, Schwerzstraße 40, 70599 Stuttgart, Germany
[b] Researcher in Innovation and Organization, France
[c] DRUID, Department of Strategy and Innovation Copenhagen Business School, Kilevej 14a, 2000 Frederiksberg, Denmark
[d] School of Management, University of Bath, Claverton Down, Bath, United Kingdom, BA2 7AY

ARTICLE INFO

KEYWORDS:
Model Uncertainty
Variable Robustness
Innovation
Open Innovation
Innovation Surveys
Innovation Studies

ABSTRACT

In studies of firm's innovation performance, regression analysis can involve a significant level of model uncertainty because the 'true' model, and therefore the appropriate set of explanatory variables are unknown. Drawing on innovation survey data for France, Germany, and the United Kingdom, we assess the robustness of the literature on inbound open innovation to variable selection choices, using Bayesian model averaging (BMA). We investigate a wide range of innovation determinants proposed in the literature and establish a robust set of findings for the variables related to the introduction of new-to-the-firm and new-to-the-world innovation with the aim of gauging the overall healthiness of the literature. Overall, we find greater robustness for explanations for new-to-the-firm rather than new-to-the-world innovation. We explore how this approach might help to improve our understanding of innovation.

## 1. Introduction

In innovation studies, researchers have long sought to understand the firm level determinants of innovation and why some firms are better able to introduce new products, processes, and services to the market. The early innovation literature tends to focus on firm-internal factors, including research and development (R&D) spending, firm size, age, and managerial structure, and how these aspects affect the pattern of innovation at both the firm and industry levels (Cohen, 1995). Over time, attention has shifted toward the impact on innovation performance of factors external to the firm such as collaboration and search. This more recent body of work builds on the concept of absorptive capacity which highlights the importance of R&D as enabling the development of new products and processes and as a means that allows the firm to learn from external sources of information (Cohen & Levinthal, 1990).

In the stream of research on firm-external factors related to inbound open innovation, the two contributions by Laursen and Salter (2006) and Cassiman and Veugelers (2006) are important. These studies use innovation survey data for the United Kingdom (UK) and Belgium, respectively. Laursen and Salter (2006) show that firm innovation performance benefits from openness to external sources of knowledge, which they describe as external search breadth and external search depth – but that after a certain point, decreasing and even negative returns set in. Cassiman and Veugelers (2006) demonstrate that firms benefit from making and buying innovations, suggesting complementarity between internal and external innovation sources. These two studies triggered a large body of research exploring whether and how external search and collaboration shape innovation performance including attempts to extend, reconceptualize, replicate, and revise those original works while also making a contribution in their own right (e.g., Garriga et al., 2013; Leiponen & Helfat, 2010; Love et al., 2014; Tether & Tajar, 2008). These studies adopt Laursen and Salter's (2006) and Cassiman and Veugelers's (2006) modeling choices but pay little attention to their appropriateness. The present paper focuses on the choice of variables, that is, the choice of which independent variables to include in the model. Given how this literature has evolved, there is a risk that this research stream does not build on sufficiently empirically robust results. By robust, in this context we mean results that hold for a substantial number of variable selection choices. The present study investigates the robustness of the central analyses in this literature based on a single empirical model, the same dependent variable, and a large number of different independent variables to try to derive a set of empirical results that are robust to different variables selection choices. We illustrate that Bayesian model averaging (BMA) which is the method

* Corresponding author.
E-mail addresses: ebersberger@uni-hohenheim.de (B. Ebersberger), fabrice.galia.25.26@gmail.com (F. Galia), kl.si@cbs.dk (K. Laursen), a.j.salter@bath.ac.uk (A. Salter).

https://doi.org/10.1016/j.respol.2021.104271
Received 8 May 2020; Received in revised form 19 April 2021; Accepted 19 April 2021
Available online 8 May 2021

we apply, is useful to investigate a wide range of topics in innovation research.[1]

This endeavor reflects the idea that "the goal of science is empirical generalization or knowledge development leading to some degree of understanding" (Hubbard et al., 1998, p. 251). However, in innovation studies, as in many scientific fields, the integrity of the relevant empirical literature has generally been questioned on two distinct counts. The first is that rather frequent errors and biases in the reporting of findings imperil the reliability of empirical research (Bruns et al., 2019). Bruns and colleagues (2019) report that about 45% of the papers they analyzed contained at least one reporting error. The second is that published empirical research may be burdened by what Rosenthal (1979) called the "file drawer problem" and by what is commonly referred to as "p-hacking", respectively. That is, papers that do not contain significant results do not get published, and that only significant results are presented (see Bruns et al., 2019) or only data leading to significant results will be retrieved (see Bruns & Kalthaus, 2020). This implies that the papers published in journals may be dominated by those reporting results with Type I errors (erroneous rejection of the null hypothesis) while papers with null outcomes are filed away by the researchers. In other words, when researchers "search for asterisks" (Bettis, 2012), in published papers, Type I error rates will be well in excess of the prescribed, conventional 5% level of significance (Denton, 1985). For instance, Goldfarb and King (2016) report that between 24% and 40% of the findings published in the top five strategic management research journals are likely the result of chance rather than reflecting true relationships. Bettis (2012) and Hubbard et al. (1998) suggest that robustness studies could help to mitigate this problem by establishing a set of reliable empirical results. Even if researchers do not "search for asterisks", statistical tests by their nature produce Type I errors.

Overall and especially from the point of view of variables selection our knowledge is limited about which results are empirically generalizable and whether they potentially add to our understanding. In principle, when researchers work on similar datasets and use similar or identical dependent variables, the robust (and less robust) results can be extracted while controlling for a host of other factors. Many currently available standard datasets provide information on firm-level variables (e.g., Compustat), strategic alliances (Schilling, 2009), patents (Hall et al., 2001), and innovation processes and outcomes based on various national Community Innovation Surveys (CIS), all of which are conducted according to a standardized procedure (Smith, 2005). The existence of these datasets for innovation research enables scholars to conduct a range of studies. However, in these studies as in most empirical applications, the "true" model, and therefore appropriate selection of the explanatory variables are essentially unknown. Here, "true model" refers to "the model that generated the data" (Ley & Steel, 2012, p. 256). The fact that the true model is unknown to the researcher leads to the phenomenon of "model uncertainty" (Chatfield, 1995; George & Clyde, 2004; Hoeting et al., 1999). Disregarding model uncertainty results in too small standard errors and over-confidence in the statistical findings (Raftery, 1995a,b; Hoeting et al., 1999). Model uncertainty can also facilitate the search for asterisks. Because the specification of the true model in terms of which variables to include is unknown, researchers may be tempted to opt for a specification that results in estimator significances that fit the researcher's line of reasoning. This can lead to a plethora of different models which all try to explain the same phenomenon.

In this paper, we address two claims made by Bruns et al. (2019). First, they consider robustness studies a tool to help generate more reliable empirical findings. We respond to this claim and contribute by presenting a cross-country robustness study of the association between inbound open innovation and innovation performance to examine whether model uncertainty is a concern in this influential literature. We consider this effort as providing a "health check" on this domain with respect to model uncertainty.[2] Second, Bruns et al. (2019) claim that in a field where the theory does not lead immediately to a uniquely accepted model specification, there are many possibilities for its empirical specification. Searching among potential models threatens the reliability of the research if the researcher confuses explorative research with hypothesis testing. This suggests that ensuring robust knowledge in innovation studies requires methods that constrain the researcher's options to search for individual models that fit the research expectations. One strategy might be to base inferences not on a single model but on a collection of models (Bruns & Ioannidis, 2020). In this study, we suggest a model averaging approach to provide a tractable way to avoid searching for a single, well-fitting model specification. We employ BMA and apply it to a key issue in the innovation literature. The BMA approach was originally applied to analyses of the macroeconomic determinants of growth (see for instance Bruns & Ioannidis, 2020 and Fernandez et al., 2001) and has not previously been employed in innovation studies. We believe that robustness studies of this type can make an important contribution to help alleviating the problem of our understanding of innovation being dominated by papers reporting results with Type I errors. In particular, the publication of robustness studies should discourage undesirable behaviors such as p-hacking.

Overall, our BMA analysis of pooled data from France, Germany, and the UK, shows that prior research in this area has a high degree of robustness which suggests that these early and influential attempts to link innovative performance to inbound open innovation variables (external search, collaboration, and make or buy decisions) are largely reliable. However, we find that the results related to new-to-the-world innovation are less robust than those related to new-to-the-firm innovation. We highlight the implications of these findings for this literature and discuss the potential application of a BMA approach to check the "health" of the field to other innovation research domains.

## 2. Innovation performance and innovation surveys

The literature on the drivers of firm-level innovative performance is vast. The traditional variables of interest are firm size and R&D activity or R&D spending. Papers testing the so-called Schumpeterian Mark II hypothesis which suggests that large firms are more innovative than small firms, have generated a large stream of generally inconclusive and even contradictory evidence (Cohen, 1995). Over time, attention has shifted to firm age based on the assumption that frequently firm size and firm age are related (Acs & Audretsch, 1988). Although research on the (dis)advantages of young and old firms in different industries is inconclusive, it suggests some advantages for young firms in new industries and for mature firms in older industries (Sørensen & Stuart, 2000). It has long been assumed that R&D investment helps firms to innovate and that the benefits of R&D reside not only in the direct effects of new goods, services, and processes but also in the building of firm absorptive capacity (Cohen & Levinthal, 1990). Yet, to a degree R&D investment may be determined endogenously by the intent to innovate or may fail to generate significant value for the organization (Hall et al., 2010).

In recent years, interest in the benefits that firms derive from

[1] We follow Clemens (2017, p. 327) and define a robustness analysis as an analysis in which the parameter estimates are drawn from a sampling distribution that is different from the sampling distribution in the original study, using the same or a different type of statistical analysis. This contrasts with replication analysis where the parameter estimates are drawn from the same sampling distribution as in the original study and the the statistical analysis is also the same as in the original study. Note that the type of robustness analysis we consider in this paper concerns the robustness of the model selection, not robustness with respect to other econometric problems such as endogeneity.

[2] Obviously, in our empirical set up there are many other concerns than model uncertainty that could cause non-robust results; we therefore do not claim that our results are robust in general.

external collaboration or search for innovative performance has increased significantly. Early work considered the effect of formal external collaboration on innovation performance (Ahuja, 2000; Powell et al., 1996) but the focus shifted later to the impact of internal and external search (Katila & Ahuja, 2002; Laursen & Salter, 2006) and external knowledge acquisition (Cassiman & Veugelers, 2006). Several scholars have broadened the range of potential variables of interest to include among others appropriability strategy (e.g., Ballot et al., 2015; Cassiman & Veugelers, 2002; Cohen et al., 2000; Laursen & Salter, 2014), managerial practices (e.g., Foss et al., 2011), internationalization (e.g., Basile, 2001; Cassiman & Golovko, 2011), and mergers and acquisitions (e.g., Paruchuri et al., 2006; Puranam et al., 2006). An important part of this research stream focuses on the exploration of the effects of search efforts on different types of innovative outcomes, and contrasting new-to-the-firm, associated with more "incremental" forms of innovation, with new-to-the-world, associated with more "radical" types of innovation.

A problem common to innovation studies is lack of data on innovation outputs across economic systems. In part, this reflects the lack of agreement on the definition and operationalization of the concept of innovation. The existence of many different types of innovation (e.g. product, process, organization, marketing, business model, service, modular, architectural, and incremental versus radical) has been acknowledged. The availability of NBER patent data and other online resources (Hall et al., 2001) has resulted in much research focused on patents as the primary measure of innovation. This has sparked numerous studies of patent-intensive industries such as semiconductors, robotics, biotechnology, and pharmaceuticals. At the same time, innovation survey data have become a major resource for scholars seeking to understand the determinants of innovation.

The primary focus of early work using innovation survey data was on the effects of collaboration and external knowledge sourcing on innovation performance. Laursen and Salter (2006) and Cassiman and Veugelers (2006) are two particularly influential papers which capture inbound open innovation practices such as collaboration, knowledge search, and knowledge sourcing. Laursen and Salter (2006) use data from an innovation survey of manufacturing firms in the UK to suggest that the breadth and depth of firms' external search are curvilinearly (inverted U-shaped) related to their innovation performance. They also compare the results for new-to-the-firm and new-to-the-world innovation. Cassiman and Veugelers (2006) use survey data for Belgium and focus on the complementarities in the make or buy decision with respect to innovation. They show that there are strong complementarities between use of in-house and external knowledge with respect to product innovation. These two papers—and others published before and after them—have stimulated a stream of work on the costs and benefits of external search and collaboration for innovation performance, focusing on a range of firm-level contingencies that might moderate or mediate this relationship (e.g., Garriga et al., 2013; Leiponen & Helfat, 2010; Love et al., 2014; Tether & Tajar, 2008). The research shows that absorptive capacity may mediate the effect of external search and collaboration on performance (Escribano et al., 2009), that the breadth of external search and its objectives have mutual effects on innovative performance (Leiponen & Helfat, 2010), that the obstacles to innovation can mediate the effect of external search on innovative outcomes (Garriga et al., 2013), and that prior search and collaboration can shape future search and innovative outcomes (Love et al., 2014). Cassiman and Valentini (2016) suggest that the complementary benefits of buying in and selling knowledge from outside the firm are small due to the associated external engagement and internal coordination costs.

Given the influence of the abovementioned studies, it is important to establish which elements of the empirical parts of these studies have a degree of robustness. The objective is not to suggest that any of the authors of these studies engaged in dubious research practices or made unsustainable model choices but to provide insights into whether the different findings across these studies may be considered robust with respect to model uncertainty. By doing so, we hope to inspire future work to examine the robustness of research in other domains of innovation studies.

**Table 1**
Sectoral distribution (in percent).

| Sector | Total (N=7,841) | France (N=3,681) | Germany (N=1,762) | UK (N=2,398) |
|---|---|---|---|---|
| Food | 11.49 | 15.65 | 4.94 | 9.92 |
| Textile | 6.19 | 8.23 | 4.48 | 4.30 |
| Pulp and Paper | 11.07 | 9.51 | 12.03 | 12.76 |
| Chemicals | 8.79 | 10.38 | 9.42 | 5.88 |
| Rubber and Plastics | 6.39 | 5.73 | 6.70 | 7.17 |
| Metal | 3.99 | 4.56 | 3.23 | 3.67 |
| Fabricated Metal | 12.09 | 10.65 | 14.76 | 12.34 |
| Equipment & Electronics | 26.36 | 22.44 | 34.40 | 26.48 |
| Other Manufacturing | 13.63 | 12.85 | 10.04 | 17.48 |
| Total | 100 | 100 | 100 | 100 |

Of course, these studies do not provide evidence of causal relationships. Although cross-sectional estimations can help to identify the direction of the relationship, they do not provide causal evidence of the effects of openness and/or make-buy decisions on firm-level performance per se. However, these studies can provide a stimulus for other researchers to experiment with research designs that could get closer to causal identification of the proposed relationship. They may also inspire more research into the mechanisms that underpin the correlations. For example, von Hippel and colleagues' early work on the role of lead users in the innovation process (e.g., Urban & von Hippel, 1988) spurred a lot of research into how users shape and enable innovative efforts and led to more insights into the ways engagement with users shapes innovative outcomes. These subsequent studies are careful attempts at causal identification, and specification of the more precise mechanisms underlying the phenomenon (see for instance Lilien et al., 2002; Dahl, Fuchs, & Schreier, 2015). Similarly, in studies of innovation performance at firm level where causal identification is notoriously difficult, correlations can provide a focusing device for subsequent research. However, given the usefulness of correlations in the research process and the aim of ensuring that future research is based on productive and reliable foundations, it is important that these correlations are robust, in the sense of being robust to variable selection choices. Our aim in this paper is to establish robust correlations.

## 3. Data and measures

Our empirical analysis uses firm-level innovation survey data from the 4th CIS for France, Germany, and the UK (Mairesse & Mohnen, 2010; Smith, 2005). The innovation outcome variables refer to year 2004, and the independent variables refer to year 2004 or the period 2002–2004. We restrict our sample to 7,841 manufacturing firms; Table 1 presents their sectoral breakdown.

Our analytical variables are constructed by focusing on the most influential papers in the relevant literature, such as Laursen and Salter (2006) and Cassiman and Veugelers (2006). We assume that these studies have influenced the modeling strategies adopted in subsequent work. We also include measures taken from the studies by Leiponen (2005), Schmiedeberg (2008), Roper et al. (2008, 2017), Grimpe and Kaiser (2010), Love et al. (2014), and Ballot et al. (2015). The list of the variables examined in these studies is extensive, and the contexts of the studies vary. To render our analysis tractable, we include only the key variables from each of the studies and the core control variables included in the innovation research literature. Table 2 lists the variables and the related innovation research.

**Table 2**
Description of the variables used.

| Variables | Description | Used in |
|---|---|---|
| INNOWORLD, INNOFIRM | Sales share of new-to-the-world or new-to-the-firm products, respectively. [log] as logarithmic transformation, and [r] as raw percentage. | Laursen and Salter (2006)[log]; Cassiman and Veugelers (2006)[r]; Schmiedeberg (2008)[r]; Love et al. (2014)[r]; Leiponen and Helfat (2010)[r]; Roper et al. (2008)[r]; Leiponen (2005a)[r]; Grimpe and Kaiser (2010)[r]; Garriga et al. (2013)[r]; Ballot et al. (2015)[log]; Roper et al. (2017)[r] |
| BREADTH | Number of sources of information used by the firm. The knowledge sources include internal sources, suppliers, customers, competitors, consulting firms, universities, governmental research organizations, conferences, publications, and business associations (0–10). | Laursen and Salter (2006); Love et al. (2014); Leiponen and Helfat (2010); Garriga et al. (2013) |
| DEPTH | Number of sources of information used to a high degree by the firm. The knowledge sources include internal sources, suppliers, customers, competitors, consulting firms, universities, governmental research organizations, conferences, publications, and business associations (0–10). | Laursen and Salter (2006); Garriga et al. (2013); Ballot et al. (2015) |
| USER | Firm uses customer information to a high degree (0/1) | Laursen and Salter (2006); Garriga et al. (2013) |
| COMPINFO | Importance of competitors as an information source on a scale of one (low) to three (high) (0–3). | Cassiman and Veugelers (2006); Roper et al. (2008) |
| BASICINFO | Importance of universities and research institutes relative to suppliers and customers as information sources (0.00–1.00). | Cassiman and Veugelers (2006) |
| PUBINFO | Importance of publications and conferences relative to suppliers and customers as information sources (0.00–1.00). | Cassiman and Veugelers (2006) |
| COLLAB | Firm is involved in innovation collaboration (0/1). | Laursen and Salter (2006); Schmiedeberg (2008); Garriga et al. (2013); Ballot et al. (2015) |
| COLDEPTH | Number of partner types used for innovation collaboration. Partner types include suppliers, customers, competitors, consulting firms, universities, and governmental research organizations (0–6). | Laursen and Salter (2006); Grimpe and Kaiser (2010); Schmiedeberg (2008) |
| RD | R&D expenditure as a share of sales (0.00–1.00). [log] as logarithmic transformation, and [r] as raw percentage. | Laursen and Salter (2006)[log]; Cassiman and Veugelers (2006)[r]; Schmiedeberg (2008)[r]; Love et al. (2014)[r]; Leiponen and Helfat (2010)[r]; Roper et al. (2008)[r]; Leiponen (2005a)[r]; Grimpe and Kaiser (2010)[r]; Garriga et al. (2013)[r]; Ballot et al. (2015)[log] |
| INTMKT | Main markets of the firm are international (0/1). | Laursen and Salter (2006); Ballot et al. (2015) |
| NATMKT | Main markets of the firm are national (0/1). | Laursen and Salter (2006) |
| IPF | Number of formal methods of protection for innovation, including the registration of designs, trademarks, or patents and the use of copyrights (0–4). | Cassiman and Veugelers (2006); Ballot et al. (2015) |
| IPNF | Number of informal methods of protection for innovation, including secrecy, complexity of design, or lead time advantage (0–3). | Cassiman and Veugelers (2006); Ballot et al. (2015) |
| OBSFIN | Lack of finance inside or outside the firm is 'very important' or 'important' (0/1). | Roper et al. (2008); Ballot et al. (2015) |
| OBSKNOW | Lack of qualified personnel, lack of information on technology, or lack of information on markets is 'very important' or 'important' (0/1). | Cassiman and Veugelers (2006); Ballot et al. (2015) |
| OBSMKT | Market domination by established enterprises or uncertain demand for innovative goods and services is 'very important' or 'important' (0/1). | Cassiman and Veugelers (2006); Ballot et al. (2015) |
| MAKEONLY | Internal R&D activities (0/1). | Cassiman and Veugelers (2006) |
| BUYONLY | Acquisition of technology through external R&D contracts, purchase of machinery for innovation, purchase of knowledge through patents, licensing, etc. (0/1). | Cassiman and Veugelers (2006) |
| MAKEBUY | Internal R&D activities and acquisition of technology through external R&D contracts, purchase of machinery for innovation, purchase of knowledge through patents, licensing, etc. (0/1). | Cassiman and Veugelers (2006) |
| LOGEMP | Logarithm of the number of employees. | Laursen and Salter (2006); Love et al. (2014); Leiponen and Helfat (2010); Roper et al. (2008); Leiponen (2005); Grimpe and Kaiser (2010); Garriga et al. (2013); Ballot et al. (2015) |
| STARTUP | Firm was a startup in the three years of the observation period (0/1). | Laursen and Salter (2006); Garriga et al. (2013) |

### 3.1. Dependent variables

In line with other research using CIS data, we measure innovation performance as share of sales based on innovation defined as sales of new products in 2004 in the total sales in 2004. We use two variants of the innovation performance measure, based on innovation novelty: INNOWORLD is the share of sales of new-to-the-world products, and INNOFIRM captures the share of sales of new-to-the-firm products.

### 3.2. Potential predictors

Unless otherwise stated, all the independent variables refer to the period 2002–2004.

***Search*:** Firms' innovation search activities are captured by BREADTH and DEPTH. We integrate particular information flows from users or from competitors through the variables USER and COMPINFO. We capture the value of the information from the science system by BASICINFO and from publicly available sources by PUBINFO.

***Collaboration*:** Patterns of innovation collaboration are captured by a collaboration dummy variable COLLAB and by a variable indicating collaboration depth COLDEPTH.

***R&D*:** R&D activities are reflected by total R&D expenditure as a share of sales in year 2004.

***Markets*:** The firm's main market is indicated by INTMKT for international and NATMKT for national market. The reference category is regional market.

***Appropriability strategy*:** IPF and IPNF denote formal and informal appropriability strategy, respectively.

***Obstacles to innovation*:** We include financial obstacles to innovation (OBSFIN), knowledge obstacles to innovation (OBSKNOW), and market obstacles to innovation (OBSMKT).

***Make or buy*:** Three dummy variables capture the make or buy decisions—MAKEONLY, BUYONLY, and MAKEBUY—with the neither make nor buy decision as the reference category.

***Firm demographics*:** Firm size is measured as the log of the number of employees (LOGEMP) in year 2004. STARTUP indicates firm foundation between 2002 and 2004. We also include eight industry dummies to capture the firms' main area of activity. The firms from France, Germany, and the UK are differentiated by two country dummies.

Table B1 in the Supplementary Material presents the correlations among the potential predictors, and Table B2 in the Supplementary Material provides a more detailed descriptive account of the data by country.

## 4. Methodology

### 4.1. Model uncertainty and model averaging

If in an empirical endeavor the true model specification in terms of the selection of independent variables is essentially unknown, model uncertainty (Chatfield, 1995) occurs in the form of variable selection uncertainty (George & Clyde, 2004; Hoeting et al., 1999). This applies to most of the analyses in innovation research. That is, at the outset, the model is unknown and the predictors that should be included in the regression model are unclear. In general, different sets of predictors and therefore different models can lead to dramatically different conclusions. Furthermore, neglecting model uncertainty and not addressing the dependence of the results on the chosen model leads to over-confident inferences based on the statistical estimates (Hoeting et al., 1999; Raftery, 1995b).

Model averaging approaches address this model uncertainty and recognize that in addition to the model parameters, the structure of the model needs to be estimated. Model averaging approaches achieve this without incurring the common problem of data mining: searching for and selecting a single best model but not presenting the process leading to its selection (Brock et al., 2007; Chatfield, 1995).

Model averaging includes both Bayesian and non-Bayesian approaches. Non-Bayesian approaches are documented in Hansen (2007) and Hjort and Claeskens (2003). Recently, BMA has attracted increased attention in diverse areas of economic research such as macroeconomic growth models (e.g., Bruns & Ioannidis, 2020; Crespo Cuaresma et al., 2011; Durlauf, 2001; Fernandez et al., 2001; Magnus et al., 2010), forecasting (Liu & Maheu, 2009; Wright, 2009), and agricultural economics (Balcombe & Rapsomanikis, 2010; Tiffin & Balcombe, 2011). Steel (2020) provides an extensive overview of the use of model averaging in economics. In management research there are only two studies studies so far that use an averaging approach (Arin et al., 2015; Melián-González et al., 2011). To the best of our knowledge, model uncertainty has not received attention in research on innovation.

### 4.2. Bayesian model averaging in a nutshell

Model averaging proposes a solution to model uncertainty by estimating several plausible models, averaging over the models, and drawing inferences based on weighted averages.

Relative to single variable selection approaches, BMA has certain advantages (Hinne et al., 2020): it reduces over-confidence in the estimation (e.g., Hoeting et al., 1999), improves prediction (e.g., Jacobson & Karlsson, 2004), avoids the "winner-takes-it-all" nature of single variable selection approaches (e.g., Wang et al., 2004), and is relatively robust with respect to model misspecification. We focus on the first of these advantages in our brief exposition of the model averaging process. Assume that, for inference, we are interested in an unknown statistic $\Delta$. We can think of $\Delta$ being the regression coefficients . If we estimate this statistic with a single regression model, then the estimate is conditional on this model. As already indicated, model uncertainty arises when the researcher is not certain that the model is the correct one; then, subsequently, not accounting for this uncertainty leads to over-confident conclusions from the analysis. Although it is common practice in empirical analyses to report a number of different models to illustrate robustness, this does not eliminate model uncertainty because the inference typically is still based on a single model.

To address this uncertainty, we are interested in an estimate of $\Delta$ that is unconditional on any specific model. BMA offers a viable approach to accomplish this since it does not search for a single good or best fit model and does not base all future conclusions on this single model. Rather, BMA uses all possible models to estimate $\Delta$. In short, for each of the possible models, BMA computes a posterior model probability (PMP) by updating our prior belief in the model with a measure of how well the model captures the given data. This PMP is used as a weight for averaging over all models to obtain an estimate of $\Delta$ which then is unconditional on a single individual model.

### 4.3. Bayesian model averaging in more detail

Bayes's rule tells us how our prior belief about $\Delta$ changes after we have seen the available data $\mathbf{D}$. This gives us the posterior belief $p(\Delta|\mathbf{D})$. Assume now that there are several models $M_j$ that can be used to describe the relationship between $\Delta$ and data $\mathbf{D}$. Again, we can use Bayes's rule and compute the posterior model probability (PMP), $p(M_j|\mathbf{D})$, for each of the several models. $p(M_j|\mathbf{D})$ then serves as a measure of plausibility for model $M_j$ after we have seen the data. Were we to perform variable selection, we would be searching for the single most plausible model and would identify the model with the highest PMP and base our inference on this single model. In the case of uncertainty about the underlying true model, Bayesian model averaging (BMA) allows us to base our inference not on a single model but on an ensemble of estimated models. It does this by averaging over all the estimated distributions of $\Delta$ in all the estimated models using the PMP as a weight.

The general idea in model averaging is to estimate a large number of models for the data $\mathbf{D}$ with $N$ observations which consist of the $N \times 1$ vector of the dependent variable $\mathbf{y}$ and the $N \times K$ matrix $\mathbf{X}$ of predictors. The models $M_j \in \{M_1 \ldots M_{2^K}\}$ all differ in terms of the predictors $\mathbf{X}_j$ included. We assume a linear regression model, where $\alpha$ is the constant, $\boldsymbol{\beta}_j$ are the coefficients, and $\boldsymbol{\varepsilon} \sim N(0, \sigma^2\mathbf{I})$ are the error terms.

$$\mathbf{y} = \alpha + \mathbf{X}_j\boldsymbol{\beta}_j^{'} + \varepsilon$$

The distribution of the parameter of interest $\Delta$ is computed as the weighted sum of the parameter distributions derived from the estimations of the models $M_j$:

$$p(\Delta|\mathbf{D}) = \sum_{j=1}^{2^K} p(\Delta|\mathbf{D}, M_j)p(M_j|\mathbf{D}).$$

$p(\Delta|\mathbf{D}, M_j)$ is the posterior distribution of $\Delta$ based on the model $M_j$.

**Table 3**
Results of BMA analyses of the pooled data – innovations new-to-the-firm.

| | Including squared terms and interactions$^{+}$ | | | | No squared terms and interactions | | | |
|---|---|---|---|---|---|---|---|---|
| Var | (a) PIP | (b) Post Mean | (c) Post SD | (d) Cond. Pos.Sign | (a) PIP | (b) Post Mean | (c) Post SD | (d) Cond. Pos.Sign |
| BREADTH | **0.776** | **0.416** | **0.387** | **0.941** | 0.423 | -0.004 | 0.066 | 0.524 |
| BREADTH2 | **0.677** | **-0.035** | **0.031** | **0.000** | | | | |
| DEPTH | **0.977** | **0.427** | **0.252** | **1.000** | **0.963** | **0.423** | **0.177** | **1.000** |
| DEPTH2 | 0.414 | 0.002 | 0.034 | 0.811 | | | | |
| COLDEPTH | **0.998** | **0.488** | **0.353** | **1.000** | **0.997** | **0.524** | **0.158** | **1.000** |
| COLDEPTH2 | 0.439 | 0.011 | 0.057 | 0.925 | | | | |
| INTMKT | **0.644** | **0.463** | **0.538** | **0.998** | **0.643** | **0.470** | **0.543** | **0.999** |
| RD | **0.722** | **-0.096** | **0.156** | **0.000** | 0.475 | -0.064 | 0.135 | 0.000 |
| COLLAB | 0.483 | -0.231 | 0.524 | 0.003 | 0.477 | -0.208 | 0.444 | 0.006 |
| USER | 0.465 | -0.140 | 0.345 | 0.016 | 0.451 | -0.122 | 0.333 | 0.030 |
| LOGEMP | **1.000** | **-0.604** | **0.134** | **0.000** | **1.000** | **-0.634** | **0.133** | **0.000** |
| STARTUP | **0.997** | **3.989** | **1.682** | **1.000** | **0.991** | **4.701** | **1.505** | **1.000** |
| STARTUP*RD | **0.558** | **8.519** | **10.002** | **1.000** | | | | |
| MAKEONLY | **0.508** | **0.235** | **0.679** | **0.905** | **0.508** | **0.308** | **0.700** | **0.937** |
| BUYONLY | 0.481 | -0.155 | 0.649 | 0.277 | 0.457 | -0.061 | 0.634 | 0.511 |
| MAKEBUY | **0.929** | **1.333** | **0.708** | **1.000** | **0.952** | **1.459** | **0.700** | **1.000** |
| OBSFIN | 0.433 | -0.065 | 0.252 | 0.000 | 0.423 | -0.061 | 0.250 | 0.000 |
| OBSKNOW | 0.481 | -0.186 | 0.388 | 0.000 | 0.473 | -0.182 | 0.386 | 0.000 |
| OBSMKT | 0.468 | 0.138 | 0.322 | 1.000 | 0.457 | 0.136 | 0.320 | 1.000 |
| BASICINFO | **0.564** | **-0.257** | **0.388** | **0.000** | **0.638** | **-0.366** | **0.417** | **0.000** |
| PUBINFO | 0.447 | -0.075 | 0.249 | 0.117 | 0.415 | -0.010 | 0.215 | 0.593 |
| COMPINFO | **0.971** | **0.640** | **0.257** | **1.000** | **0.987** | **0.689** | **0.239** | **1.000** |
| NATMKT | **0.629** | **0.487** | **0.593** | **0.996** | **0.627** | **0.491** | **0.597** | **0.997** |
| IPF | **0.792** | **0.251** | **0.195** | **1.000** | **0.797** | **0.258** | **0.196** | **1.000** |
| IPNF | **0.482** | **0.066** | **0.141** | **1.000** | **0.471** | **0.062** | **0.139** | **1.000** |
| N | 7,384 | | | | 7,384 | | | |
| N. of burn-in steps | $1\times10^{6}$ | | | | $1\times10^{6}$ | | | |
| N. of iteration steps | $5\times10^{6}$ | | | | $5\times10^{6}$ | | | |
| Corr. PMP | 0.997 | | | | 0.996 | | | |
| Threshold | 0.500 | | | | 0.500 | | | |

Note: Expected model size = 0.5 $K$. Results for the industry and country dummies are not reported. Variables with PIP>threshold are in bold. + Strong heredity enforced (Crespo Cuaresma 2011). Based on the 1,000 best models.

Note that $\Delta$ can be any statistic. As noted above in our case it is the regression coefficients $\beta$ that we are interested in.

$p(M_j|\mathbf{D})$ is the posterior model probability (PMP). It is used as the weight in the averaging process and indicates the probability that the model $M_j$ is the correct model. This of course, is conditional on one of the models $M_1 \ldots M_{2^K}$ being the correct model. $p(M_j|\mathbf{D})$ is given by:

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

Here, $p(\mathbf{D}|M_j)$ is the marginal likelihood of model $M_j$:

$p(\mathbf{D}|M_j) = \int p(\mathbf{D}|\alpha,\boldsymbol{\beta}_j,\sigma,M_j)\cdot p(\alpha,\boldsymbol{\beta}_j,\sigma|M_j)d\alpha\, d\boldsymbol{\beta}_j\, d\sigma$ .

$p(\mathbf{D}|\alpha,\boldsymbol{\beta}_j,\sigma,M_j)$ is the likelihood of model $M_j$, and $p(\alpha,\boldsymbol{\beta}_j,\sigma|M_j)$ is the prior density of the coefficients of model $M_j$.

### 4.4. Specification of the BMA

As suggested by Fernandez et al. (2001) and used very recently by Bruns and Ioannidis (2020), we assume for $\alpha$ and $\sigma^2$ an improper and non-informative prior $p(\alpha)\propto 1$ and $p(\sigma)\propto\sigma^{-1}$, respectively. For $\boldsymbol{\beta}_j$, we assume a $g$-prior (Zellner, 1986):

$\boldsymbol{\beta}_j|\sigma, M_j \sim N(0, \sigma^2 g(\mathbf{X}_j^{'}\mathbf{X}_j)^{-1})$.

As Bruns and Ioannidis (2020) point out, the analysis can be sensitive to the selection of the hyperprior $g$. Selecting a small $g$ would indicate a strong belief in a distribution tightly wrapped around zero. In contrast, a large $g$ leads to results that focus on a few models, giving rise to the so-called supermodel effect. This can make the findings rather fragile (Feldkircher & Zeugner, 2009, 2012), since any selection of the fixed $g$ entails the risk that the value is either too small or too large given the noise in the data. To prevent this, Feldkircher and Zeugner (2009) suggest avoiding fixed $g$-priors completely and substituting the fixed $g$-prior by a prior distribution on the $g$-parameter to "let … the data choose" (Feldkircher & Zeugner, 2009, p. 4). Liang et al. (2008) suggest a hyper-$g$-prior approach which essentially puts a Beta prior on the shrinkage factor $g/(1+g)$:

$\frac{g}{1+g} \sim \text{Beta}\left(1, \frac{a}{2}-1\right)$

We use $a = 2 + 2/N$ as $N > K^2$ for all our samples and subsamples. This ensures that the expected shrinkage is $\max\{N, K^2\}$ which is the benchmark proposed by Fernandez et al. (2001). The model priors $p(M_j)$ are:

$p(M_j) \ = \pi^{k_j}(1-\pi)^{K-k_j}$.

There, $\pi$ is the probability of inclusion of each regressor, and $k_j$ is the number of regressors in model $M_j$. Analogously, $1-\pi$ is the probability for each regressor of not being included in the model, and $K-k_j$ is the number of regressors that are not in model $M_j$. In general, it is preferable to prevent the prior from affecting the posterior distribution by assuming an inclusion probability of $\pi = 0.5$ (Brock et al., 2007) which is intended to be non-informative. However, Bruns and Ioannidis (2020) remind us that Ley and Steel (2009, p. 672) "strongly discourage the use of the fixed [$\pi$] prior as a 'non-informative' prior, as it has clearly been shown to be quite informative" because a fixed inclusion probability leads to a prior model size distribution that is quite concentrated, and hence rather informative. As suggested by Ley and Steel (2009) and implemented recently by Bruns and Ioannidis (2020), we impose a hyperprior on the inclusion probability, where $\pi$ is drawn from a Beta distribution. In contrast to macroeconomic growth regressions, where Sala-i-Martin et al. (2004) and Sala-i-Martin (1997) suggest an expected size of the growth regression models, we have no information about the expected model size. Hence, we assume an expected model size of 0.5 $K$. In the Appendix, we report analyses with expected model sizes of 0.4 $K$ and 0.6 $K$ (see Table A2 and A3).

 

**Table 4**
Results of BMA analyses of the pooled data – innovations new-to-the-world.

| | Including squared terms and interactions+ | | | | No squared terms and interactions | | | |
|---|---|---|---|---|---|---|---|---|
| Var | (a) PIP | (b) Post Mean | (c) Post SD | (d) Cond. Pos.Sign | (a) PIP | (b) Post Mean | (c) Post SD | (d) Cond. Pos.Sign |
| BREADTH | 0.319 | 0.006 | 0.109 | 0.415 | 0.331 | -0.008 | 0.052 | 0.271 |
| BREADTH2 | 0.099 | -0.001 | 0.008 | 0.002 | | | | |
| DEPTH | **0.945** | **0.493** | **0.287** | **1.000** | **0.936** | **0.365** | **0.167** | **1.000** |
| DEPTH2 | 0.374 | -0.024 | 0.043 | 0.000 | | | | |
| COLDEPTH | **0.942** | **0.408** | **0.279** | **1.000** | **0.931** | **0.406** | **0.180** | **1.000** |
| COLDEPTH2 | 0.223 | 0.002 | 0.038 | 0.402 | | | | |
| INTMKT | **0.995** | **1.451** | **0.418** | **1.000** | **0.996** | **1.474** | **0.426** | **1.000** |
| RD | **1.000** | **1.468** | **0.163** | **1.000** | **1.000** | **1.467** | **0.163** | **1.000** |
| COLLAB | 0.357 | 0.262 | 0.518 | 1.000 | 0.456 | 0.305 | 0.514 | 1.000 |
| USER | 0.262 | 0.065 | 0.254 | 0.998 | 0.359 | 0.104 | 0.295 | 1.000 |
| LOGEMP | **1.000** | **-0.776** | **0.120** | **0.000** | **1.000** | **-0.776** | **0.120** | **0.000** |
| STARTUP | **0.555** | **1.168** | **1.472** | **1.000** | **0.592** | **1.297** | **1.491** | **1.000** |
| STARTUP*RD | 0.148 | 0.606 | 3.501 | 1.000 | | | | |
| MAKEONLY | **0.947** | **2.260** | **0.850** | **1.000** | **0.964** | **2.318** | **0.815** | **1.000** |
| BUYONLY | 0.269 | -0.136 | 0.632 | 0.176 | 0.334 | -0.086 | 0.615 | 0.469 |
| MAKEBUY | **0.953** | **2.198** | **0.751** | **1.000** | **0.970** | **2.262** | **0.716** | **1.000** |
| OBSFIN | 0.220 | 0.000 | 0.160 | 0.431 | 0.297 | 0.001 | 0.186 | 0.585 |
| OBSKNOW | 0.218 | 0.000 | 0.213 | 0.474 | 0.301 | 0.001 | 0.250 | 0.562 |
| OBSMKT | **0.998** | **-1.504** | **0.393** | **0.000** | **0.998** | **-1.501** | **0.392** | **0.000** |
| BASICINFO | 0.239 | 0.034 | 0.199 | 0.758 | 0.322 | 0.046 | 0.227 | 0.807 |
| PUBINFO | 0.476 | -0.219 | 0.306 | 0.000 | **0.570** | **-0.262** | **0.317** | **0.000** |
| COMPINFO | 0.341 | -0.078 | 0.161 | 0.008 | 0.435 | -0.099 | 0.175 | 0.008 |
| NATMKT | 0.237 | -0.045 | 0.251 | 0.013 | 0.317 | -0.057 | 0.289 | 0.009 |
| IPF | **1.000** | **0.775** | **0.155** | **1.000** | **1.000** | **0.776** | **0.155** | **1.000** |
| IPNF | **1.000** | **1.382** | **0.164** | **1.000** | **1.000** | **1.382** | **0.164** | **1.000** |
| N | 7,841 | | | | 7,841 | | | |
| N. of burn-in steps | $1\times10^6$ | | | | $1\times10^6$ | | | |
| N. of iteration steps | $5\times10^6$ | | | | $5\times10^6$ | | | |
| Corr. PMP | 0.997 | | | | 0.997 | | | |
| Threshold | 0.500 | | | | 0.500 | | | |

Note: Expected model size = 0.5 *K*. Results for the industry and country dummies are not reported. Variables with PIP>threshold are in bold. + Strong heredity implemented (Cuaresma 2011). Based on the 1,000 best models.

Our set of regressors comprises the 21 variables summarized in Table 2, eight sector dummies, and two country dummies. Additionally, we include the second-order terms of BREADTH, DEPTH, and COLDEPTH and an interaction term (STARTPU*RD) of STARTUP and RD to capture all the variables in Laursen and Salter's (2006) main regressions. To ease "interpretability" (Moser & Hofmarcher, 2014, p. 346) of the second order or interaction term estimates the literature (Crespo Cuaresma, 2011; Hasan et al., 2018) suggests implementing the strong heredity principle (Chipman, 1996). It implies that every second order or interaction effect also has a main effect in the model. As in Hasan et al. (2018) we implement two different types of BMA analyses: one with second order and interaction variables and employing the strong heredity principle with 35 predictors, and one with only first order terms and 31 predictors.

The model space consists of all $2^K$ models that can be constructed from the respective $K$=35 and K=31 potential predictors. Searching this extensive model space requires use of a Markov Chain Monte Carlo (MCMC) algorithm. The MCMC algorithm runs through the model space by generating draws from a Markov chain on the model space and approximates the posterior model distribution (Fernandez et al., 2001). We perform 1,000,000 burn-in steps and 5,000,000 iterations to compute the results. Burn-in steps are required for the MCMC algorithm to zoom in on the important parts of the model space which are those with high PMPs. In the Supplementary Material, we include the details of our implementation using the R-package BMS (Zeugner & Feldkircher, 2015).

## 5. Results

Tables 3 and 4 report the results of the BMA analysis of the pooled data for the sales shares of the new-to-the-firm and new-to-the-world innovations, respectively, as dependent variables. To document Markov chain's convergence, we report the correlation coefficient *Corr. PMP* (Fernandez et al., 2001). Correlation coefficient values of 0.999 and 0.997 indicate that the algorithm converges to important areas of the model distribution. The left sides of Tables 3 and 4 report the analysis with the strong heredity principle imposed for the squared terms and for the interaction terms. The right sides of the tables report the analysis without the squared terms and without the interaction terms.

The posterior inclusion probability (*PIP*) reported in column (a) in Table 3 and Table 4 is the sum of the probabilities of the models that include this predictor. The PIP is the sum of the posterior model probabilities which are used as the weights in the averaging process. The PIP directly addresses model uncertainty. It represents the probability of each potential predictor to be part of the true model (see e.g., Crespo Cuaresma, 2010). The *Post Mean* in column (b) presents the direction and magnitude of the parameter estimates; it reports the mean parameter estimate for all the models that include the respective variable. In column (c), *Post SD* is the posterior standard deviation of the parameter estimates conditional on the variable being part of the model. Finally, *Cond. Pos. Sign* in column (d) is the fraction of the models with a positive parameter estimate when the variable is included in the model.

When interpreting the findings from the BMA analysis, it is important to consider the PIP relative to the prior inclusion probability which acts as the threshold for interpreting the variables as "significant" (Sala-i-Martin et al., 2004). We assume a prior inclusion probability that yields an expected model size of 0.5 *K*. If the results show that the PIP of the variables exceeds this threshold this supports our belief that the variables are part of the true model; in other words, that the variables "belong in the regression" (Sala-i-Martin, 2004, p. 823). Below, we label these variables "robust correlates" of innovation performance. In the analyses in the Appendix, we use the expected model size 0.4 *K* and 0.6 *K* to illustrate how much our findings depend on the prior inclusion

**Table 5**
Determinants' effects (open innovation variables only).

| | All (expected model size = 0.5 K) | | All (expected model size = 0.4 K) | | All (expected model size = 0.6 K) | |
|---|---|---|---|---|---|---|
| Squared & interaction terms | Yes | No | Yes | No | Yes | No |
| Threshold | 0.5 | 0.5 | 0.4 | 0.4 | 0.6 | 0.6 |
| **Sales share of new-to-the-firm innovations** | | | | | | |
| **BREADTH** | **Inverse U (6)** | - | **Inverse U (6)** | - | **Inverse U (6)** | - |
| **DEPTH** | **Positive** | **Positive** | **Positive** | **Positive** | **Positive** | **Positive** |
| **COLDEPTH** | **Positive** | **Positive** | **Positive** | **Positive** | **Positive** | **Positive** |
| COLLAB | - | - | Negative | Negative | - | - |
| USER | - | - | Negative | - | - | - |
| **MAKEONLY** | **Positive** | **Positive** | **Positive** | **Positive** | **Positive**[+] | **Positive**[+] |
| BUYONLY | - | - | Negative | Negative | - | - |
| **MAKEBUY** | **Positive** | **Positive** | **Positive** | **Positive** | **Positive** | **Positive** |
| **BASICINFO** | **Negative** | **Negative** | **Negative** | **Negative** | **Negative** | **Negative** |
| PUBINFO | - | - | - | - | - | - |
| **COMPINFO** | **Positive** | **Positive** | **Positive** | **Positive** | **Positive** | **Positive** |
| **IPF** | **Positive** | **Positive** | **Positive** | **Positive** | **Positive** | **Positive** |
| IPNF | Positive | Positive | Positive | Positive | - | - |
| **Sales share of new-to-the-world innovations** | | | | | | |
| BREADTH | - | - | - | - | - | - |
| **DEPTH** | **Positive** | **Positive** | **Positive** | **Positive** | **Positive** | **Positive** |
| **COLDEPTH** | **Positive** | **Positive** | **Positive** | **Positive** | **Positive** | **Positive** |
| COLLAB | - | - | - | Positive | - | - |
| USER | - | - | - | - | - | - |
| **MAKEONLY** | **Positive** | **Positive** | **Positive** | **Positive** | **Positive**[+] | **Positive**[+] |
| BUYONLY | - | - | - | - | - | - |
| **MAKEBUY** | **Positive** | **Positive** | **Positive** | **Positive** | **Positive** | **Positive** |
| BASICINFO | - | - | - | - | - | - |
| PUBINFO | - | Negative | Negative | Negative | - | - |
| COMPINFO | - | - | - | Negative | - | - |
| **IPF** | **Positive** | **Positive** | **Positive** | **Positive** | **Positive** | **Positive** |
| IPNF | Positive | Positive | Positive | Positive | - | Positive |

**Note:** This table condenses the information contained Table 3, Table 4 and Tables A1-A4 in the Appendix. It reports the direction and shape of robust (PIP > threshold) correlations that open innovation variables have with the sales share of innovation. – indicates regressors that we have not identified as robust correlates (PIP ≤ threshold) [+]indicates 0.550 < PIP ≤ 0.600. Consistent findings across these six (or three) models are in bold.

probability choice.[3] Table 5 presents a summary. It reports the shape and direction of the correlation of the inbound open innovation variables with innovation performance. The prior literature finds that with some exceptions most of these variables affect innovation performance.[4]

We focus on those inbound open innovation variables where all of the six models or the three models for the respective squared terms, summarized in Table 5, show agreement on the robustness and direction of the correlation.

Applying BMA to the pooled data shows that the search effort intensity matters for new-to-the-firm innovations, measured as both DEPTH of search and collaboration depth (COLDEPTH).[5] Only make (MAKEONLY) and simultaneous make and buy (MAKEBUY) are above the threshold with positive means. Note that the estimates of the posterior means of the three different make and buy decisions fulfill the condition for complementarity (Cassiman & Veugelers, 2006, p. 70) between make and buy.

Information from competitors (COMPINFO) with a positive posterior mean is a robust correlate given a PIP well above the threshold, whereas information from basic R&D such as universities and research organizations (BASICINFO) is a robust but negative correlate with innovation performance measured by the share of sales of new-to-the-firm innovations.

Formal IP protection measures (IPF) has a PIP larger than the threshold. The posterior mean of IPF points to a positive association with the share of sales of new-to-the-firm innovations.

It is important to note that search breadth (BREADTH) has a robust correlation with the share of sales of new-to-the-firm innovations and takes an inverted U-shape when second order and interaction terms are included in the analysis. Across the three models with the expected model sizes of 0.5 $K$, 0.4 $K$, and 0.6 $K$, the peak of the inverted-U is consistently at six out of ten information sources which is close to the median for the search breadth distribution. The BMA analyses which do not include second order and interaction terms, are not able to pick up this correlation. Rather, these analyses suggest no correlation which is inconsistent with the observation that more than 50% of the observations have a search breadth that suggests over-searching (see Laursen & Salter 2006).

The determinants of the share of sales of new-to-the-world innovations (bottom panel of the summary Table 5) overlap with the determinants of the share of sales of new-to-the-firm innovations. Both DEPTH and COLDEPTH are robust correlates with PIPs above the threshold. The overlap also contains MAKEONLY and MAKEBUY. Here again we find that the size of the posterior means of the three make and buy variables fulfil Cassiman and Veugelers's (2006, p. 70) condition for complementarity. Analysis of the pooled data shows that COMPINFO is not correlated with the share of sales of new-to-the-world innovations. Formal IP protection measures are correlated with a positive posterior

[3] In the Supplementary Material, in Tables B6-B8 we also report the BMA analyses for France, Germany, and the UK. We do not interpret these findings in detail but refer the reader to the summary table (Table B9) in the Supplementary Material.

[4] For instance, in Cassiman and Veugelers (2006), among the three critical make and buy variables, only the variable reflecting MAKEBUY is consistently strongly significant in the regressions. However, these results—including the weak or insignificant results for MAKEONLY and BUYONLY—play an important part in establishing that there is complementarity between make and buy, which is in line with the aim of the analysis.

[5] For comparison, traditional ordinary least square regressions of the pooled dataset are reported in Table B10 and Table B11 in the Supplementary Material.

mean and a PIP consistently at the value of 1. Search breadth is not identified as a robust correlate with innovation performance when measured as the share of sales of the more radical new-to-the-world innovations.

If we compare the results in Tables 3 and 4 for new-to-the-firm innovations and new-to-the-world-innovations, we can identify more robust correlations in the case of new-to-the-firm innovations (16 variables), compared to new-to-the-world-innovations (11 variables) based on the full models including the squared and the interaction terms.

## 6. Conclusions

### 6.1. Summary and contributions

In this study, we set out to assess the robustness of the literature on inbound open innovation and its impact on firm-level innovation performance by accounting for model uncertainty employing BMA. Similar to the original literature, our results do not identify causal relationships. Rather, we establish a set of "robust correlations" in the sense that they hold for a very large number of variable selection choices. We would argue that robust correlations could be a fruitful starting point for future research addressing causal interpretations and explicit modeling of the underlying mechanisms.

Overall, we found a degree of conformity between the BMA analysis and the original empirical literature, suggesting that the prior studies generate some robust findings. More specifically, we found robust correlations for critical inbound open innovation variables (external search, collaboration, and innovation make or buy decisions) considered in the literature and the related variables for appropriability strategies. However, we found some important differences in terms of some of the variables capturing some aspects of inbound open innovation. For example, while we found consistently robust correlations suggesting inverted U-shaped relationships such as those hypothesized by Laursen and Salter (2006) with respect to new-to-the-firm innovations, we did not find consistently robust correlations suggesting such relationships with respect to new-to-the-world innovations. More generally, the correlations suggest that the results for new-to-firm innovations are consistently more robust than the results for new-to-the-world innovations. In part, this might be because radical innovation is more difficult and therefore more uncertain than incremental innovation in terms of the factors shaping their outcomes. However, it would be unwise to see these results as definitive statements of a 'true' model for assessing external sources and collaboration and their effect on innovation performance. Our analysis has some important limitations with respect to the data used, the contexts covered, and variables available for inclusion all of which we discuss in greater detail below.

This paper makes two principal contributions. First, our analytical exercise shows that implementation of a model averaging approach to analyze innovation survey data allows for simultaneous estimation of the parameters and structure of the innovation performance model which contributes to reducing model uncertainty and our aim to examine the robustness of the literature. As stated in the introduction to this study, we believe that the publication of robustness studies would discourage undesirable research behaviors particularly "p-hacking". They would also encourage researchers to consider their modeling choices more carefully. Second, we contribute by presenting a cross-country robustness study of the association between inbound open innovation and innovation performance to examine whether model uncertainty, and therefore the robustness of the variable choices, is an issue in this literature. In other words, we conducted a "health check" on the literature and can cautiously conclude that it is in a quite healthy state. However, we found that the robustness of the variables for new-to-the-world innovation was markedly lower than in the case of the variables for new-to-the-firm innovation which highlights the need to consider modeling choices with respect to these two different types of innovation outcomes with great care.

Based on our study, we conjecture that averaging approaches that address model uncertainty could be a valuable tool in a range of innovation research contexts. We highlighted the particular value of this technique for a robustness study in a context where a relatively large number of studies (with many—potentially competing—explanatory variables) use the same or similar datasets while relying on the same dependent variable. Although many areas within innovation studies meet this condition, we would point to three potentially fruitful applications. First, many studies of firm-level innovation rely on patent counts or citation weighted patents as the dependent variable and use a wide range of explanatory variables from patent data and other sources as the independent variables. To improve the consistency and transparency of patent measures (Bruns & Kalthaus, 2020), BMA might allow greater transparency about the effect of model uncertainty in patent studies, and help to reduce the potential for Type 1 errors in the literature. Second, science, technology, and innovation studies have for long focused on scientific performance and used various measures of academic productivity such as citations, and a range of different explanatory variables to explain productivity differences among academics. It would be useful to know which among this wide gamut of potential variables are robust to model uncertainty. Finally, diffusion studies explore the uptake of different technologies over time, highlighting a wide array of factors that might explain technology adoption. Tackling model uncertainty in these diffusion studies could allow a richer appreciation of the factors driving technology adoption and a better understanding of the innovation process.

### 6.2. Limitations

Although BMA approaches can help to assess the health of a domain with respect to robustness to variable selection and model uncertainty in that particular context, these approaches evidently do not address problems related to unobserved heterogeneity or other endogeneity problems. Other methods and research tools such as experiments, regression discontinuity, or instrumental variables would be required to generate stronger evidence on the causal link between search or collaboration and innovation performance, and to understand the specific mechanisms that link these practices to the expected outcomes. Moreover, although BMA approaches can help to assess the robustness of operationalizations and the implications of the theoretical arguments in prior work, they should not be seen as alternatives to theory development. Indeed, our robustness analysis does not confirm or reject the theoretical predictions in the research considered. However, BMA analysis could act as an antidote to strong assumptions about the validity of theory based on current operationalizations in the empirical literature, and thus, could encourage scholars to build their theories on more reliable empirical foundations.

Although our study suggests new ways to assess the robustness of the literature in an important research area in innovation studies, it is subject to some important limitations. First, we assume that linear regression as the chosen econometric model is the correct specification, since unlike other model specifications linear regression models are well-established in the BMA approach. This means we cannot use the methodology to cross-check the appropriateness of the functional form of the model. Second, our approach assumes that the true model can be built from a subset of the potential predictors included in the analysis. In reality, predictors not yet identified in the literature could play a central role in identifying the true model. So, our findings should be interpreted contingent on this. Third, our approach relies on information contained in the CIS dataset and might suffer from omissions such as exclusion of important additional variables, and/or commission such as design and measurement errors emerging from the questionnaire. The approach proposed in this paper is appropriate for environments with relatively stable research designs and limited use of supplementary data. Fourth, and as pointed out above, model averaging of cross-sectional data does not help to resolve the endogeneity issues related to innovation data. A

**Table A1**
Results of BMA analyses of the pooled data – innovations new-to-the-firm (Robustness, expected model size = 0.4 K).

| | Including squared terms and interactions+ | | | | No squared terms and interactions | | | |
|---|---|---|---|---|---|---|---|---|
| | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) |
| Var | PIP | Post Mean | Post SD | Cond. Pos.Sign | PIP | Post Mean | Post SD | Cond. Pos.Sign |
| BREADTH | **0.723** | **0.377** | **0.388** | **0.927** | 0.374 | -0.004 | 0.062 | 0.527 |
| BREADTH2 | **0.614** | **-0.032** | **0.032** | **0.000** | | | | |
| DEPTH | **0.972** | **0.426** | **0.242** | **1.000** | **0.957** | **0.419** | **0.179** | **1.000** |
| DEPTH2 | 0.354 | 0.002 | 0.031 | 0.766 | | | | |
| COLDEPTH | **0.998** | **0.485** | **0.328** | **1.000** | **0.997** | **0.521** | **0.156** | **1.000** |
| COLDEPTH2 | 0.384 | 0.010 | 0.053 | 0.916 | | | | |
| INTMKT | **0.601** | **0.443** | **0.539** | **0.998** | **0.607** | **0.453** | **0.544** | **0.998** |
| RD | **0.658** | **-0.088** | **0.152** | **0.000** | **0.422** | **-0.057** | **0.129** | **0.000** |
| COLLAB | **0.424** | **-0.203** | **0.490** | **0.004** | **0.425** | **-0.185** | **0.425** | **0.006** |
| USER | **0.402** | **-0.117** | **0.323** | **0.021** | 0.399 | -0.104 | 0.316 | 0.037 |
| LOGEMP | **1.000** | **-0.606** | **0.134** | **0.000** | **1.000** | **-0.633** | **0.133** | **0.000** |
| STARTUP | **0.996** | **4.105** | **1.679** | **1.000** | **0.989** | **4.714** | **1.521** | **1.000** |
| STARTUP*RD | **0.483** | **7.402** | **9.782** | **1.000** | | | | |
| MAKEONLY | **0.454** | **0.211** | **0.638** | **0.900** | **0.460** | **0.271** | **0.662** | **0.930** |
| BUYONLY | **0.423** | **-0.153** | **0.608** | **0.272** | **0.409** | **-0.078** | **0.597** | **0.446** |
| MAKEBUY | **0.929** | **1.339** | **0.683** | **1.000** | **0.950** | **1.444** | **0.677** | **1.000** |
| OBSFIN | 0.372 | -0.055 | 0.235 | 0.000 | 0.370 | -0.052 | 0.234 | 0.000 |
| OBSKNOW | **0.419** | **-0.161** | **0.367** | **0.000** | **0.420** | **-0.160** | **0.368** | **0.000** |
| OBSMKT | **0.408** | **0.119** | **0.303** | **1.000** | **0.407** | **0.120** | **0.305** | **1.000** |
| BASICINFO | **0.511** | **-0.240** | **0.379** | **0.000** | **0.594** | **-0.340** | **0.411** | **0.000** |
| PUBINFO | 0.388 | -0.064 | 0.233 | 0.139 | 0.366 | -0.011 | 0.201 | 0.584 |
| COMPINFO | **0.968** | **0.648** | **0.260** | **1.000** | **0.986** | **0.694** | **0.239** | **1.000** |
| NATMKT | **0.583** | **0.464** | **0.592** | **0.995** | **0.587** | **0.471** | **0.596** | **0.997** |
| IPF | **0.761** | **0.244** | **0.198** | **1.000** | **0.770** | **0.251** | **0.199** | **1.000** |
| IPNF | **0.427** | **0.059** | **0.135** | **1.000** | **0.419** | **0.056** | **0.134** | **1.000** |
| N | 7,384 | | | | 7,384 | | | |
| N. of burn-in steps | $1\times10^6$ | | | | 1×106 | | | |
| N. of iteration steps | $5\times10^6$ | | | | 5×106 | | | |
| Corr. PMP | 0.996 | | | | 0.997 | | | |
| Threshold | 0.400 | | | | 0.400 | | | |

Note: Expected model size = 0.4 *K*. Results for the industry and country dummies are not reported. Variables with PIP>threshold are in bold. + Strong heredity implemented (Cuaresma 2011). Based on the 1,000 best models.

**Table A2**
Results of BMA analyses of the pooled data – innovations new-to-the-world (Robustness, expected model size = 0.4 K).

| | Including squared terms and interactions+ | | | | No squared terms and interactions | | | |
|---|---|---|---|---|---|---|---|---|
| | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) |
| Var | PIP | Post Mean | Post SD | Cond. Pos.Sign | PIP | Post Mean | Post SD | Cond. Pos.Sign |
| BREADTH | 0.289 | 0.004 | 0.100 | 0.387 | 0.302 | -0.008 | 0.049 | 0.255 |
| BREADTH2 | 0.082 | -0.001 | 0.007 | 0.002 | | | | |
| DEPTH | **0.940** | **0.484** | **0.284** | **1.000** | **0.933** | **0.364** | **0.167** | **1.000** |
| DEPTH2 | 0.351 | -0.023 | 0.042 | 0.000 | | | | |
| COLDEPTH | **0.939** | **0.412** | **0.269** | **0.999** | **0.929** | **0.409** | **0.180** | **1.000** |
| COLDEPTH2 | 0.202 | 0.002 | 0.036 | 0.374 | | | | |
| INTMKT | **0.994** | **1.448** | **0.418** | **1.000** | **0.996** | **1.470** | **0.424** | **1.000** |
| RD | **1.000** | **1.470** | **0.163** | **1.000** | **1.000** | **1.468** | **0.163** | **1.000** |
| COLLAB | 0.342 | 0.254 | 0.511 | 1.000 | **0.427** | **0.291** | **0.510** | **1.000** |
| USER | 0.245 | 0.064 | 0.250 | 0.998 | 0.330 | 0.098 | 0.286 | 1.000 |
| LOGEMP | **1.000** | **-0.776** | **0.120** | **0.000** | **1.000** | **-0.777** | **0.120** | **0.000** |
| STARTUP | **0.527** | **1.112** | **1.458** | **1.000** | **0.561** | **1.230** | **1.480** | **1.000** |
| STARTUP*RD | 0.131 | 0.538 | 3.301 | 1.000 | | | | |
| MAKEONLY | **0.942** | **2.252** | **0.860** | **1.000** | **0.963** | **2.317** | **0.811** | **1.000** |
| BUYONLY | 0.258 | -0.145 | 0.639 | 0.178 | 0.305 | -0.089 | 0.602 | 0.431 |
| MAKEBUY | **0.949** | **2.192** | **0.761** | **1.000** | **0.969** | **2.262** | **0.710** | **1.000** |
| OBSFIN | 0.200 | 0.000 | 0.153 | 0.411 | 0.271 | 0.001 | 0.178 | 0.544 |
| OBSKNOW | 0.202 | 0.000 | 0.205 | 0.462 | 0.272 | 0.001 | 0.238 | 0.535 |
| OBSMKT | **0.998** | **-1.506** | **0.392** | **0.000** | **0.998** | **-1.503** | **0.392** | **0.000** |
| BASICINFO | 0.221 | 0.029 | 0.189 | 0.736 | 0.291 | 0.040 | 0.214 | 0.786 |
| PUBINFO | **0.457** | **-0.210** | **0.302** | **0.000** | **0.544** | **-0.249** | **0.314** | **0.000** |
| COMPINFO | 0.317 | -0.072 | 0.156 | 0.010 | **0.404** | **-0.091** | **0.170** | **0.009** |
| NATMKT | 0.221 | -0.042 | 0.242 | 0.014 | 0.291 | -0.053 | 0.277 | 0.011 |
| IPF | **1.000** | **0.775** | **0.155** | **1.000** | **1.000** | **0.776** | **0.155** | **1.000** |
| IPNF | **1.000** | **1.384** | **0.164** | **1.000** | **1.000** | **1.383** | **0.164** | **1.000** |
| N | 7,841 | | | | 7,841 | | | |
| N. of burn-in steps | $1\times10^6$ | | | | $1\times10^6$ | | | |
| N. of iteration steps | $5\times10^6$ | | | | $5\times10^6$ | | | |
| Corr. PMP | 0.998 | | | | 0.998 | | | |
| Threshold | 0.400 | | | | 0.400 | | | |

Note: Expected model size = 0.4 *K*. Results for the industry and country dummies are not reported. Variables with PIP>threshold are in bold. +Strong heredity implemented (Cuaresma 2011). Based on the 1,000 best models.

**Table A3**
Results of BMA analyses of the pooled data – innovations new-to-the-firm (Robustness, expected model size = 0.6 K).

| | Including squared terms and interactions+ | | | | No squared terms and interactions | | | |
|---|---|---|---|---|---|---|---|---|
| Var | (a) PIP | (b) Post Mean | (c) Post SD | (d) Cond. Pos.Sign | (a) PIP | (b) Post Mean | (c) Post SD | (d) Cond. Pos.Sign |
| BREADTH | **0.815** | **0.446** | **0.383** | **0.951** | 0.472 | -0.004 | 0.069 | 0.502 |
| BREADTH2 | **0.726** | **-0.037** | **0.031** | **0.000** | | | | |
| DEPTH | **0.981** | **0.427** | **0.260** | **1.000** | **0.967** | **0.425** | **0.176** | **1.000** |
| DEPTH2 | 0.473 | 0.003 | 0.036 | 0.851 | | | | |
| COLDEPTH | **0.999** | **0.490** | **0.378** | **1.000** | **0.997** | **0.528** | **0.160** | **1.000** |
| COLDEPTH2 | 0.494 | 0.011 | 0.061 | 0.935 | | | | |
| INTMKT | **0.684** | **0.479** | **0.537** | **0.999** | **0.680** | **0.487** | **0.543** | **0.999** |
| RD | **0.765** | **-0.102** | **0.158** | **0.000** | 0.518 | -0.070 | 0.139 | 0.000 |
| COLLAB | 0.537 | -0.256 | 0.554 | 0.002 | 0.523 | -0.228 | 0.458 | 0.005 |
| USER | 0.519 | -0.160 | 0.361 | 0.013 | 0.502 | -0.138 | 0.349 | 0.025 |
| LOGEMP | **1.000** | **-0.603** | **0.134** | **0.000** | **1.000** | **-0.634** | **0.133** | **0.000** |
| STARTUP | **0.997** | **3.898** | **1.681** | **1.000** | **0.993** | **4.686** | **1.494** | **1.000** |
| STARTUP*RD | **0.615** | **9.353** | **10.072** | **1.000** | | | | |
| MAKEONLY | 0.556 | 0.262 | 0.713 | 0.918 | 0.553 | 0.345 | 0.735 | 0.944 |
| BUYONLY | 0.530 | -0.146 | 0.679 | 0.274 | 0.503 | -0.042 | 0.670 | 0.573 |
| MAKEBUY | **0.936** | **1.339** | **0.725** | **1.000** | **0.955** | **1.477** | **0.721** | **1.000** |
| OBSFIN | 0.489 | -0.075 | 0.267 | 0.000 | 0.473 | -0.069 | 0.263 | 0.000 |
| OBSKNOW | 0.536 | -0.208 | 0.404 | 0.000 | 0.519 | -0.201 | 0.401 | 0.000 |
| OBSMKT | 0.523 | 0.156 | 0.337 | 1.000 | 0.507 | 0.153 | 0.334 | 1.000 |
| BASICINFO | **0.607** | **-0.269** | **0.393** | **0.000** | **0.675** | **-0.388** | **0.421** | **0.000** |
| PUBINFO | 0.507 | -0.088 | 0.264 | 0.096 | 0.464 | -0.009 | 0.227 | 0.577 |
| COMPINFO | **0.974** | **0.633** | **0.256** | **1.000** | **0.988** | **0.686** | **0.238** | **1.000** |
| NATMKT | **0.669** | **0.504** | **0.592** | **0.997** | **0.660** | **0.505** | **0.596** | **0.998** |
| IPF | **0.820** | **0.257** | **0.191** | **1.000** | **0.821** | **0.263** | **0.193** | **1.000** |
| IPNF | 0.539 | 0.073 | 0.146 | 1.000 | 0.518 | 0.067 | 0.144 | 1.000 |
| N | 7,384 | | | | 7,384 | | | |
| N. of burn-in steps | $1\times10^6$ | | | | $1\times10^6$ | | | |
| N. of iteration steps | $5\times10^6$ | | | | $5\times10^6$ | | | |
| Corr. PMP | 0.999 | | | | 0.999 | | | |
| Threshold | 0.600 | | | | 0.600 | | | |

Note: Expected model size = 0.6 *K*. Results for the industry and country dummies are not reported. Variables with PIP>threshold are in bold. + Strong heredity implemented (Cuaresma 2011). Based on the 1,000 best models.

**Table A4**
Results of BMA analyses of the pooled data – innovations new-to-the-world (Robustness, expected model size = 0.6 K).

| | Including squared terms and interactions+ | | | | No squared terms and interactions | | | |
|---|---|---|---|---|---|---|---|---|
| Var | (a) PIP | (b) Post Mean | (c) Post SD | (d) Cond. Pos.Sign | (a) PIP | (b) Post Mean | (c) Post SD | (d) Cond. Pos.Sign |
| BREADTH | 0.333 | 0.008 | 0.115 | 0.432 | 0.355 | -0.008 | 0.054 | 0.279 |
| BREADTH2 | 0.110 | -0.002 | 0.009 | 0.002 | | | | |
| DEPTH | **0.947** | **0.500** | **0.289** | **1.000** | **0.941** | **0.367** | **0.166** | **1.000** |
| DEPTH2 | 0.389 | -0.025 | 0.043 | 0.000 | | | | |
| COLDEPTH | **0.943** | **0.403** | **0.286** | **0.999** | **0.933** | **0.403** | **0.180** | **1.000** |
| COLDEPTH2 | 0.235 | 0.003 | 0.040 | 0.426 | | | | |
| INTMKT | **0.994** | **1.452** | **0.421** | **1.000** | **0.996** | **1.479** | **0.429** | **1.000** |
| RD | **1.000** | **1.468** | **0.163** | **1.000** | **1.000** | **1.466** | **0.163** | **1.000** |
| COLLAB | 0.373 | 0.273 | 0.526 | 1.000 | 0.479 | 0.316 | 0.517 | 1.000 |
| USER | 0.272 | 0.065 | 0.257 | 0.998 | 0.384 | 0.109 | 0.300 | 1.000 |
| LOGEMP | **1.000** | **-0.775** | **0.120** | **0.000** | **1.000** | **-0.776** | **0.120** | **0.000** |
| STARTUP | 0.571 | 1.196 | 1.478 | 1.000 | **0.613** | **1.342** | **1.495** | **1.000** |
| STARTUP*RD | 0.162 | 0.663 | 3.657 | 1.000 | | | | |
| MAKEONLY | **0.947** | **2.258** | **0.852** | **1.000** | **0.966** | **2.322** | **0.814** | **1.000** |
| BUYONLY | 0.281 | -0.136 | 0.638 | 0.174 | 0.355 | -0.081 | 0.620 | 0.502 |
| MAKEBUY | **0.953** | **2.196** | **0.755** | **1.000** | **0.972** | **2.265** | **0.715** | **1.000** |
| OBSFIN | 0.229 | 0.000 | 0.163 | 0.456 | 0.325 | 0.002 | 0.195 | 0.622 |
| OBSKNOW | 0.230 | 0.000 | 0.219 | 0.482 | 0.326 | 0.001 | 0.260 | 0.589 |
| OBSMKT | **0.998** | **-1.503** | **0.392** | **0.000** | **0.999** | **-1.499** | **0.392** | **0.000** |
| BASICINFO | 0.254 | 0.039 | 0.206 | 0.776 | 0.350 | 0.053 | 0.237 | 0.826 |
| PUBINFO | 0.492 | -0.227 | 0.310 | 0.000 | 0.593 | -0.273 | 0.320 | 0.000 |
| COMPINFO | 0.356 | -0.082 | 0.164 | 0.008 | 0.464 | -0.106 | 0.180 | 0.007 |
| NATMKT | 0.249 | -0.047 | 0.257 | 0.013 | 0.344 | -0.061 | 0.301 | 0.008 |
| IPF | **1.000** | **0.775** | **0.155** | **1.000** | **1.000** | **0.776** | **0.155** | **1.000** |
| IPNF | 0.333 | 0.008 | 0.115 | 0.432 | **1.000** | **1.381** | **0.164** | **1.000** |
| N | 7,841 | | | | 7,841 | | | |
| N. of burn-in steps | $1\times10^6$ | | | | $1\times10^6$ | | | |
| N. of iteration steps | $5\times10^6$ | | | | $5\times10^6$ | | | |
| Corr. PMP | 0.997 | | | | 0.998 | | | |
| Threshold | 0.600 | | | | 0.600 | | | |

Note: Expected model size = 0.6 *K*. Results for the industry and country dummies are not reported. Variables with PIP>threshold are in bold. +Strong heredity implemented (Cuaresma 2011). Based on the 1,000 best models.

larger number of innovative firms might have inherently better quality (in terms of management, routines, etc.); therefore, cross-sectional data in part may reflect an omitted variable that would help to explain both the firms' use of different innovation strategies and their different innovation performance. It might also be the case that innovation performance drives strategic choice which in turn drives innovation. In the absence of evidence based on panel data and strong instruments, it is difficult to make strong inferences about the causal effects of particular variables on innovation outcomes. Fifth, although our study involves pooling of data for three countries, we found some important differences in terms of the robustness of some variables across these settings (see the Supplementary Material). Since we lack information on the national factors that might give rise to these differences systematically, in the present study we put this issue to the side. Future research based on data for a larger number countries could explore how these national (or even regional) differences shape the robustness of the key variables. Finally, it should be pointed out that BMA and model averaging more generally, is not a universal cure for model uncertainty but is itself subject to more research. For example, the findings by Sala-i-Martin et al. (2004) obtained using Bayesian averaging of classical estimators have been shown to be rather fragile (Ciccone & Jarociński, 2010).

Despite these limitations, we believe that robustness studies would help to ensure that the results in the innovation literature are generalizable in terms of variable selection, and also might encourage further efforts to find the most appropriate analytical model. This could lead to new research directions, and trigger questions about common modeling choices and new findings that lack robustness. Note that model uncertainty can be addressed from other than empirical perspectives. Theoretical advances—particularly in relation to the development of exact analytical models grounded in current innovation and management theory—might be helpful and complement work to address model uncertainty. However, we hope that our study provides a tool to help increase the research community's confidence in the robustness of a class of studies using similar dependent variables. In doing so, such tools might help to anchor the literature around common understandings and to direct the attention of scholars to new areas that require further research.

## CRediT authorship contribution statement

**Bernd Ebersberger:** Conceptualization, Methodology, Data curation, Formal analysis, Writing - original draft. **Fabrice Galia:** Data curation, Writing - original draft. **Keld Laursen:** Conceptualization, Methodology, Writing - original draft. **Ammon Salter:** Conceptualization, Methodology, Writing - original draft.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

Bernd conducted this research as part of the Research Area "Innovation, Entrepreneurship, and Finance (INEF)" at the University of Hohenheim's Faculty of Business, Economics, and Social Sciences. Fabrice was with Burgundy School of Business when this project started. We thank the participants in the OUI 15th International Open and User Innovation Conference (2017), the SMS Annual Conference (2016), the AOM Annual Meeting (2016), the International Schumpeter Society (ISS) Conference (2016), and the DRUID16 20th Anniversary Conference (2016) for their comments and suggestions on a previous version of the paper. We especially thank the editor and four anonymous referees for providing comments that substantially improved the paper. The usual disclaimer applies.

## Appendix

Table A1-A4

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.respol.2021.104271.

## References

Acs, Z.J., Audretsch, D.D., 1988. Innovation in large and small firms: An empirical analysis. American Economic Review 78 (4), 678–690.
Ahuja, G., 2000. Collaboration networks, structural holes and innovation: A longitudinal study. Administrative Science Quarterly 45 (3), 425–455.
Arin, K.P., Huang, V.Z., Minniti, M., Nandialath, A.M., Reich, O.F.M, 2015. Revisiting the determinants of entrepreneurship: A Bayesian approach. Journal of Management 41 (2), 607–631.
Balcombe, K., Rapsomanikis, G., 2010. An analysis of the impact of research and development on productivity using Bayesian model averaging with a reversible jump algorithm. American Journal of Agricultural Economics 92 (4), 985–998. https://doi.org/10.1093/ajae/aaq050.
Ballot, G., Fakhfakh, F., Galia, F., Salter, A.J., 2015. The fateful triangle: Complementarities in performance between product, process and organizational innovation in France and the UK. Research Policy 44 (1), 217–232.
Basile, R., 2001. Export behaviour of Italian manufacturing firms over the nineties: The role of innovation. Research Policy 30 (8), 1185–1201.
Bettis, R.A., 2012. The search for asterisks: Compromised statistical tests and flawed theories. Strategic Management Journal 33 (1), 108–113.
Brock, W., Durlauf, S.N., West, K., 2007. Model uncertainty and policy evaluation: Some theory and empirics. Journal of Econometrics 136 (2), 629–664.
Bruns, S.B., Ioannidis, J.P., 2020. Determinants of economic growth: Different time different answer? Journal of Macroeconomics 63, 103185. https://doi.org/10.1016/j.jmacro.2019.103185. December 2019.
Bruns, S.B., Asanov, I., Bode, R., Dunger, M., Funk, C., Hassan, S.M., Hauschildt, J., Heinisch, D., Kempa, K., König, J., Lips, J., Verbeck, M., Wolfschütz, E., Buenstorf, G., 2019. Reporting errors and biases in published empirical findings: Evidence from innovation research. Research Policy 48 (9), 103796. https://doi.org/10.1016/j.respol.2019.05.005.
Bruns, S.B., Kalthaus, M., 2020. Flexibility in the selection of patent counts: Implications for p-hacking and evidence-based policymaking. Research Policy 49 (1), 103877. https://doi.org/10.1016/j.respol.2019.103877.
Cassiman, B., Golovko, E., 2011. Innovation and internationalization through exports. Journal of International Business Studies 42 (1), 56–75.
Cassiman, B., Valentini, G., 2016. Open innovation: Are inbound and outbound knowledge flows really complementary? Strategic Management Journal 37 (6), 1034–1046.
Cassiman, B., Veugelers, R., 2002. Spillovers and R&D cooperation: Some empirical evidence. American Economic Review 92 (4), 1169–1184.
Cassiman, B., Veugelers, R., 2006. In search of complementarity in innovation strategy: Internal R&D and external knowledge acquisition. Management Science 52 (1), 68–82.
Chatfield, C., 1995. Model uncertainty, data mining and statistical inference. Journal of the Royal Statistical Society Series A (Statistics in Society) 158 (3), 419–466.
Chipman, H., 1996. Variable selection with related predictors. Canadian Journal of Statistics 24 (1), 17–36.
Ciccone, A., Jarociński, M., 2010. Determinants of economic growth: Will data tell? American Economic Journal: Macroeconomics 2 (4), 222–246.
Clemens, M.A., 2017. The meaning of failed replications: A review and proposal. Journal of Economic Surveys 31 (1), 326–342.
Cohen, W. M. (1995). Empirical studies of innovative activity. In P. Stoneman (Ed.), *Handbook of the economics of innovation and technological change* (pp. 342–365). Blackwell.
Cohen, W.M., Levinthal, D.A., 1990. Absorptive capacity: A new perspective of learning and innovation. Administrative Science Quarterly 35 (1), 128–152.
Cohen, W.M., Nelson, R.R., Walsh, J., 2000. Protecting their intellectual assets: Appropriability conditions and why U.S. manufacturing firms patent (or not). National Bureau of Economic Research Working Paper 7552.
Crespo Cuaresma, J., 2010. Natural disasters and human capital accumulation. World Bank Economic Review 24 (2), 280–302.
Crespo Cuaresma, J., 2011. How different is Africa? A comment on Masanjala and Papageorgiou. Journal of Applied Econometrics 26 (6), 1041–1047. https://doi.org/10.1002/jae.1180.
Crespo Cuaresma, J., Foster, N., Stehrer, R., 2011. Determinants of regional economic growth by quantile. Regional Studies 45 (6), 809–826. https://doi.org/10.1080/00343401003713456.
Dahl, D.W., Fuchs, C., Schreier, M., 2015. Why and When Consumers Prefer Products of User-Driven Firms: A Social Identification Account. Management Science 61 (8), 1978–1988.
Denton, F.T., 1985. Data mining as an industry. The Review of Economics and Statistics 67 (1), 124–127.
Durlauf, S.N., 2001. Manifesto for a growth econometrics. Journal of Economic Growth 100 (1), 65–69.

Escribano, A., Fosfuri, A., Tribó, J.A., 2009. Managing external knowledge flows: The moderating role of absorptive capacity. Research Policy 38 (1), 96–105.

Feldkircher, M., Zeugner, S., 2009. Benchmark priors revisited: On adaptive shrinkage and the supermodel effect in Bayesian model averaging. IMF Working Papers.

Feldkircher, M., Zeugner, S., 2012. The impact of data revisions on the robustness of growth determinants-a note on 'determinants of economic growth: Will data tell? Journal of Applied Econometrics 27 (4), 686–694. https://doi.org/10.1002/jae.2265.

Fernandez, C., Ley, E., Steel, M.F.J, 2001. Model uncertainty in cross-country growth regressions. Journal of Applied Econometrics 16 (5), 563–576. https://doi.org/10.1002/jae.623.

Foss, N.J., Laursen, K., Pedersen, T., 2011. Linking customer interaction and innovation: The mediating role of new organizational practices. Organization Science 22 (4), 980–999.

Garriga, H., von Krogh, G., Spaeth, S., 2013. How constraints and knowledge impact open innovation. Strategic Management Journal 34 (9), 1134–1144.

George, E.I., Clyde, M.A., 2004. Model uncertainty. Statistical Science 19 (1), 81–94. https://doi.org/10.1214/088342304000000035.

Goldfarb, B., King, A.A., 2016. Scientific apophenia in strategic management research: Significance tests & mistaken inference. Strategic Management Journal 37 (1), 167–176.

Grimpe, C., Kaiser, U., 2010. Balancing internal and external knowledge acquisition: The gains and pains of R&D outsourcing. Journal of Management Studies 47 (8), 1483–1509.

Hall, B., Jaffe, A., Trajtenberg, M., 2001. The Nber patent citations data file: Lessons, insights, and methodological tools 8498. NBER Working Paper.

Hall, B. H., Mairesse, J., & Mohnen, P. (2010). Measuring the returns to R&D. In *Handbook of the economics of innovation* (Vol. 2, pp. 1033–1082). North-Holland.

Hansen, B.E., 2007. Least squares model averaging. Econometrica 75 (4), 1175–1189.

Hasan, I., Horvath, R., Mares, J., 2018. What type of finance matters for growth? Bayesian model averaging evidence World Bank Economic Review 32 (2), 383–409. https://doi.org/10.1093/wber/lhw029.

Hinne, M., Gronau, Q.F., van den Bergh, D., Wagenmakers, E.-J., 2020. A conceptual introduction to Bayesian model averaging. Advances in Methods and Practices in Psychological Science 3 (2), 200–215.

Hjort, N.L., Claeskens, G., 2003. Frequentist model average estimators. Journal of the American Statistical Association 98 (464), 879–899.

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

Hubbard, R., Vetter, D.E., Little, E.L., 1998. Replication in strategic management: Scientific testing for validity, generalizability, and usefulness. Strategic Management Journal 19 (3), 243–254.

Jacobson, T., Karlsson, S., 2004. Finding good predictors for inflation: A bayesian model averaging approach. Journal of Forecasting 23 (7), 479–496.

Katila, R., Ahuja, G., 2002. Something old, something new: A longitudinal study of search behaviour and new product introduction. Academy of Management Journal 45 (8), 1183–1194.

Laursen, K., Salter, A.J., 2006. Open for innovation: The role of openness in explaining innovative performance among UK manufacturing firms. Strategic Management Journal 27 (2), 131–150.

Laursen, K., Salter, A.J., 2014. The paradox of openness: Appropriability, external search and collaboration. Research Policy 43 (5), 867–878.

Leiponen, A., 2005. Organization of knowledge and innovation: The case of Finnish business services. Industry and Innovation 12 (2), 185–203.

Leiponen, A., Helfat, C.E., 2010. Innovation opportunities, knowledge sources, and the benefits of breadth. Strategic Management Journal 31 (2), 224–236.

Ley, E., Steel, M.F.J, 2009. On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. Journal of Applied Econometrics 24 (4), 651–674. https://doi.org/10.1002/jae.

Ley, E., 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., Berger, J.O., 2008. Mixtures of g priors for Bayesian variable selection. Journal of the American Statistical Association 103 (481), 410–423. https://doi.org/10.1198/016214507000001337.

Lilien, G., Morrison, PD, Searls, K, Sonnack, M, von Hippel, E, 2002. Performance Assessment of the Lead User Generation Process for New Product Development. Management Science 48 (8), 1042–1060.

Liu, C., Maheu, J.M., 2009. Forecasting realized volatility: A Bayesian model-averaging approach. Journal of Applied Econometrics 24 (5), 709–733.

Love, J.H., Roper, S., Vahter, P., 2014. Learning from openness: The dynamics of breadth in external innovation linkages. Strategic Management Journal 35 (11), 1703–1716.

Magnus, J.R., Powell, O., Prüfer, P., 2010. A comparison of two model averaging techniques with an application to growth empirics. Journal of Econometrics 154 (2), 139–153.

Mairesse, J., & Mohnen, P. (2010). Chapter 26 - Using innovation surveys for econometric analysis. In B.H. Hall & N. Rosenberg (Eds.), *Handbook of the economics of innovation* (Vol. 2, pp. 1129–1155). North-Holland.

Melián-González, A., Moreno-Gil, S., Araña, J.E., 2011. Gay tourism in a sun and beach destination. Tourism Management 32 (5), 1027–1037.

Moser, M., Hofmarcher, P., 2014. Model priors revisited: Interaction terms in BMA growth applications. Journal of Applied Econometrics 29 (2), 344–347. https://doi.org/10.1002/jae.2365.

Paruchuri, S., Nerkar, A., Hambrick, D.C., 2006. Acquisition integration and productivity losses in the technical core: Disruption of inventors in acquired companies. Organization Science 17 (5), 545–562.

Powell, W.W., Koput, K.W., Smith-Doerr, L., 1996. Interorganizational collaboration and the local of innovation: Networks of learning in biotechnology. Administrative Science Quarterly 41 (1), 116–145.

Puranam, P., Singh, H., Zollo, M., 2006. Organizing for innovation: Managing the coordination-autonomy dilemma in technology acquisitions. The Academy of Management Journal 49 (2), 263–280.

Raftery, A.E., 1995a. Bayesian model selection in social research. Sociological Methodology 25, 111–164.

Raftery, A.E., 1995b. Rejoinder: Model selection is unavoidable in social research. Sociological Methodology 25, 185–195.

Roper, S., Du, J., Love, J.H., 2008. Modelling the innovation value chain. Research Policy 37 (6/7), 961–977.

Roper, S., Love, J.H., Bonner, K., 2017. Firms' knowledge search and local knowledge externalities in innovation performance. Research Policy 46 (1), 43–56.

Rosenthal, R., 1979. The file drawer problem and tolerance for null results. Psychological Bulletin 86 (3), 638–641.

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

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

Schilling, M.A., 2009. Understanding the alliance data. Strategic Management Journal 30 (3), 233–260.

Schmiedeberg, C., 2008. Complementarities of innovation activities: An empirical analysis of the German manufacturing sector. Research Policy 37 (9), 1492–1503.

Smith, K., 2005. Measuring innovation. In J. Fagerberg, D. C. Mowery, & R. R. Nelson (Eds.). The Oxford handbook of innovation. Oxford University Press, pp. 148–179.

Steel, M.F.J, 2020. Model averaging and its use in economics. Journal of Economic Literature 58 (3), 644–719. https://doi.org/10.1257/JEL.20191385.

Sørensen, J.B., Stuart, T.E., 2000. Aging, obsolescence, and organizational innovation. Administrative Science Quarterly 45 (1), 81–112.

Tether, B., Tajar, A., 2008. Beyond industry-university links: Sourcing knowledge for innovation from consultants, private research organisations and the public science-base. Research Policy 37 (6/7), 1079–1095.

Tiffin, R., Balcombe, K., 2011. The determinants of technology adoption by UK farmers using Bayesian model averaging: The cases of organic production and computer usage. Australian Journal of Agricultural and Resource Economics 55 (4), 579–598.

Urban, G.L., von Hippel, E., 1988. Lead User Analysis for Development of New Industrial Products. Management Science 34 (5), 569–582.

Wang, D., Zhang, W., Bakhai, A., 2004. Comparison of Bayesian model averaging and stepwise methods for model selection in logistic regression. Statistics in Medicine 23 (22), 3451–3467.

Wright, J.H., 2009. Forecasting US inflation by Bayesian model averaging. Journal of Forecasting 28 (2), 131–144.

Zellner, A., 1986. On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In P. Goel & A. Zellner (Eds.). Bayesian inference and decision techniques: Essays in honour of Bruno de Finetti. Elsevier, pp. 233–243.

Zeugner, S., Feldkircher, M., 2015. Bayesian Model Averaging Employing Fixed and Flexible Priors: The BMS Package for R. Journal of Statistical Software 68 (4), 1–30. https://doi.org/10.18637/jss.v068.i04.