Intern. J. of Research in Marketing 31 (2014) 356–367

Contents lists available at ScienceDirect

Intern. J. of Research in Marketing

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

Full Length Article

Variable selection in international diffusion models

Sarah Gelper [a,*], Stefan Stremersch [b,c]

[a] Eindhoven University of Technology, Innovation, Technology Entrepreneurship & Marketing Group, Den Dolech 2, PO Box 513, 5600 MB Eindhoven, The Netherlands
[b] Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands
[c] IESE Business School Universidad de Navarra, Barcelona, Spain

ARTICLE INFO

Article history:
First received on November 3, 2011 and was under review for 6½ months
Available online 10 July 2014

Area Editor: Eitan Muller;
Guest Editor: Marnik G. Dekimpe

Keywords:
Variable selection
Shrinkage
Bayesian Lasso
Bayesian Elastic Net
International diffusion
Bass model

ABSTRACT

Prior research comes to different conclusions as to what country characteristics drive diffusion patterns. One prime difficulty that may partially explain this divergence between studies is the sparseness of the data, in terms of the periodicity as well as the number of products and countries, in combination with the large number of potentially influential country characteristics. In face of such sparse data, scholars have used nested models, bivariate models and factor models to explore the role of country covariates. This paper uses Bayesian Lasso and Bayesian Elastic Net variable selection procedures as powerful approaches to identify the most important drivers of differences in Bass diffusion parameters across countries. We find that socio-economic and demographic country covariates (most pronouncedly so, economic wealth and education) have the strongest effect on all diffusion metrics we study. Our findings are a call for marketing scientists to devote greater attention to country covariate selection in international diffusion models, as well as to variable selection in marketing models at large.

1. Introduction

Since the 80s (Heeler & Hustad, 1980), international diffusion of new products has strongly established itself as a research stream within the international marketing literature. International diffusion[1] studies predominantly seek to explain variation in new product growth patterns across countries using country characteristics, such as economics, culture or demographics (for recent contributions, see Chandrasekaran & Tellis, 2008; Talukdar, Sudhir, & Ainslie, 2002; Stremersch & Lemmens, 2009; Stremersch & Tellis, 2004; Tellis, Stremersch, & Yin, 2003; Van den Bulte & Stremersch, 2004; van Everdingen, Fok, & Stremersch, 2009).

An important difference among these studies – beyond the difference in the products or countries included – is the set of country-level covariates included in the model. Model specification in terms of covariates in international diffusion models is particularly challenging. There is no consensus in the literature about which country characteristics should or should not be included in an international diffusion model. Marketing scholars justify their choice for a certain set of explanatory variables by theoretical reasoning. Especially in international diffusion, the theory is very rich and thus the number of variables that one could consider including is very large. At the same time, the data is often sparse, in terms of periodicity, and number of countries and products. Standard statistical estimation techniques often have difficulties to fit such large models on such sparse data. Therefore, scholars may drop one or more of the available variables through subjective choice and iterative testing of smaller models, at the risk of omission.

Scholars who do not restrict their model ex ante, often face ill-conditioning of the design matrix – or harmful multicollinearity – as a significant problem (see Chandrasekaran & Tellis, 2008; Tellis et al., 2003). An ill-conditioned design matrix may pre-empt inference from the full model, by which people resort again to dimensionality reduction techniques, such as estimating nested models (Stremersch & Tellis, 2004), bivariate models (Chandrasekaran & Tellis, 2008), composite models (Gatignon, Eliashberg, & Robertson, 1989) or factor models (Helsen, Jedidi, & Desarbo, 1993; Tellis et al., 2003). Nested models and bivariate models, however, also face the risk of omitted variable bias. Composite and factor models are difficult to interpret and are unable to disentangle the effects of distinct country covariates.

This paper uses Bayesian Lasso (Hans, 2009; Park & Casella, 2008) and Bayesian Elastic Net (Hans, 2011; Li & Lin, 2010) to explore which country characteristics matter most in international diffusion. These procedures can cope with sparse data (i.e., many variables and few data points) by specifying an appropriate informative prior, which leads to a specific form of Bayesian regularization (Fahrmeir, Kneib, & Konrath, 2010). By construction of the Lasso and Elastic Net priors, some of the estimated regression coefficients will be exactly zero, identifying a subset of most important variables. The procedure simultaneously executes shrinkage and variable selection, while alternative

* Corresponding author at: Rotterdam School of Management, Erasmus University, PO Box 1738, 3000 DR Rotterdam, The Netherlands. Tel.: +31 10 4082732.
E-mail addresses: sgelper@rsm.nl (S. Gelper), stremersch@ese.eur.nl (S. Stremersch).

[1] We use the term international diffusion as a synonym to international new product growth.

http://dx.doi.org/10.1016/j.ijresmar.2014.04.001
0167-8116/© 2014 Published by Elsevier B.V.

shrinkage methods (e.g. Ridge regression) do not include variable selection and alternative variable selection methods (e.g. Bayesian model averaging) do not include shrinkage. The advantage of the Lasso and Elastic Net procedures over shrinkage methods without variable selection is that it leads to more stable estimation results and to the identification of a relatively small subset of variables that exhibit the strongest effects (Tibshirani, 1996). The advantage over variable selection methods without shrinkage is that the latter methods still lack power in a sparse data setting because the shrinkage is crucial for dealing with correlated covariates, as we show in a simulation study.

We estimate a Bayesian version of the Bass diffusion model (Bass, 1969) which was introduced by Lenk and Rao (1990) and subsequently extended by Talukdar et al. (2002). Bayesian analysis is particularly well suited for international diffusion models because of the multilevel structure of the data. The model decomposes the product- and country-variance, which is important, given that the sample of countries is typically not the same for all products and the product variance is typically larger than the country variance. Also, regularization to deal with sparse data comes natural in a Bayesian setting via the use of an informative prior. Scholars in both marketing (Lenk & Orme, 2009) and statistics (Fahrmeir et al., 2010) show an increasing attention for the usefulness of Bayesian regularization by informative priors.

We have data on the penetration levels of 6 high technology products (CD players, internet, ISDN, mobile phones, personal computers, and video cameras) in a total of 55 countries around the world. These data are also used in van Everdingen et al. (2009) and were graciously made available to us by Yvonne van Everdingen. We complement these data with an extensive set of country characteristics that encompasses the country characteristics used in previous studies on new product adoption, ranging from socio-economic over cultural to demographic and geographic characteristics.

The results indicate that even though many country characteristics have been related to new product growth in the past, in our particular set of countries and products, the following small set of variables explains most of the between-country variation. A first predominant variable is economic wealth. It has a strong positive effect on all three parameters of the Bass diffusion model. A second important variable is education which positively affects both the market potential (m) and the innovation coefficient (p). Beyond economic wealth and education, income inequality has a negative effect on the market potential (m), economic openness affects the innovation coefficient (p), while mobility affects the imitation coefficient (q) in the Bass diffusion model. Future application of variable selection techniques on other samples of international diffusion data, may yield a promising path towards generalizable findings.

2. Prior literature on international diffusion

Table 1 inventories the international diffusion literature using variations of the Bass diffusion model. For every study, we list which country characteristics are studied, whether a dimensionality reduction method is used, and which country characteristics the authors found to influence diffusion. A more general overview of diffusion and new product growth models can be found in Peres, Muller, and Mahajan (2010).

Gatignon et al. (1989) construct three country-level constructs (cosmopolitanism, mobility and sex roles), using 9 variables and find that the three constructs significantly relate to the parameters of the Bass diffusion model. This finding was confirmed in Kumar, Ganesh, and Echambadi (1998). Takada and Jain (1991) use two dummies to account for cultural and communication differences in four Pacific Rim countries and find them to affect the adoption rate. Helsen et al. (1993) cluster countries based on six factors extracted from a total of 23 country characteristics and conclude that life style and health status are related to the parameters of the Bass diffusion model. Dekimpe, Parker, and

Table 1
Overview of international diffusion literature using country characteristics in the Bass diffusion model.

Reference	Included country characteristics	Dimensionality reduction method	Important country characteristics
Gatignon et al. (1989)	Quantity of foreign mail sent and received, international telegrams received, foreign travel, foreign visitors received, number of telephones in use, percentage of population owning at least one car, number of cars per inhabitant, per capita mileage driven, women in labor force.	3 composites: cosmopolitanism, mobility and sex roles	Cosmopolitanism, mobility, sex roles
Takada and Jain (1991)	Culture dummy (high vs low context), communication dummy (homophilous vs heterophilous).	No reduction	Culture dummy, communication dummy
Helsen et al. (1993)	Number of air passengers/km, air cargo, number of newspapers, population, cars per capita, motor gasoline consumption, electricity production, life expectancy, physicians per capita, political stability, imports, exports, GDP per capita, phones per capita, electricity consumption per capita, foreign visitors per capita, tourist expenditures per capita, tourist receipts per capita, consumer price index, newspaper circulation, hospital beds, education expenditures/government budget, graduate education in population per capita.	6 factors: mobility, health status, trade, life style, cosmopolitanism, miscellaneous	Life style, health status
Kumar et al. (1998)	Quantity of foreign mail sent and received, international telegrams received, foreign travel, foreign visitors received, number of telephones in use, percentage of population owning at least one car, number of cars per inhabitant, per capita mileage driven, women in labor force.	3 composites: cosmopolitanism, mobility and sex roles	Cosmopolitanism, mobility, sex roles
Dekimpe et al. (1998)	Population growth, number of population centers, GNP per capita, crude death rate, communism, number of ethnic groups.	No reduction	Population growth, no. population centers, crude death rate, no. ethnic groups
Talukdar et al. (2002)	Income per capita, dependents–working ratio, Gini index, urbanization, international trade, TV penetration, newspapers per capita, illiteracy rate, number of ethnic groups, women in labor force, minutes of international telephone calls.	No reduction	Income per capita, urbanization, international trade, illiteracy
Van den Bulte and Stremersch (2004)	Individualism, uncertainty avoidance, power distance, masculinity, GDP per capita, Gini index.	No reduction	Individualism, uncertainty avoidance, power distance, masculinity, Gini index
Albuquerque et al. (2007)	Population size, GDP per capita, sustainability, literacy, urbanization.	No reduction	Population size

Note: Composites are constructed based on a fixed set of pre-selected country characteristics per construct; factors are obtained by principle component analysis on the complete set of country characteristics; "No reduction" means that all country characteristics are included in the model without transformation.

Sarvary (1998) find a significant effect on the diffusion process of four out of six covariates under consideration, mainly related to demographics. Talukdar et al. (2002) specify a hierarchical Bayesian Bass model, in which per capita income, urbanization and international trade affect a new product's market potential, a country's illiteracy rate affects the innovation coefficient, and no country covariate affects the imitation coefficient. Van den Bulte and Stremersch's (2004) meta-analysis shows that the q/p ratio reported in prior applications of the Bass diffusion model varies with national culture, income inequality and the presence of competing standards. Albuquerque, Bronnenberg, and Corbett (2007) study cross-country spillovers in the adoption of ISO certifications and find that only population size has an influence on market potential.

While we focus on the Bass diffusion model, there are a number of notable studies on international new product growth beyond applications of the Bass model. Dekimpe, Parker, and Sarvary (2000b) study the time between a product's first worldwide introduction and a country's adoption time and identify economic wealth (GNP per capita) and number of ethnic groups to be the main drivers. Chandrasekaran and Tellis (2008), Stremersch & Tellis (2004) and van Everdingen et al. (2009) study cross-country variation in time-to-takeoff and international spill-overs in takeoff. These studies include a large set of country-level predictors, such as economic wealth, income inequality and culture, but find mixed effects as to the influence they have on time-to-takeoff. Stremersch and Tellis (2004) study the growth phase of the product life cycle, after takeoff and identify economic wealth (GDP per capita) as the main growth driver. Stremersch and Lemmens (2009) and Putsis, Balasubramanian, Kaplan, and Sen (1997) develop flexible models to study international new product growth. Stremersch and Lemmens (2009) find that, in the context of pharmaceuticals, regulatory regimes are an important determinant of cross-country variation in new product sales growth. Lemmens, Croux, and Stremersch (2012) propose a method to dynamically segment countries based on the observed penetration pattern of new products. They exploit such dynamic segments to predict the national penetration patterns of new products prior to launch. Putsis et al. (1997) fit a flexible mixing model with cross-country influence and find significant effects of GDP per capita and number of televisions in use on differences in international diffusion patterns.

If scholars have used dimensionality reduction methods in this literature, they are mainly of two kinds, often executed in parallel. A first kind is to estimate a series of shorter models that are nested in the full model (see for instance Tellis et al., 2003). The estimation of such nested models comes at the risk of omitted variable bias or pretest error bias in the remaining regression coefficients. Such bias can even result in estimated parameters that switch signs as a consequence of omission. For instance, Chandrasekaran and Tellis (2008) report a significantly negative influence of uncertainty avoidance on time-to-takeoff when it is the only variable in the model, while the same coefficient is significantly positive in a model that also includes other country characteristics.

A second dimensionality reduction method is factor analyzing the explanatory variables and only retaining a set of factors that explain a large part of the variance (for instance Chandrasekaran & Tellis, 2008; Helsen et al., 1993; Tellis et al., 2003). The most important factors capture most of the variation in the complete set of variables and represent underlying unobserved constructs. In practice, however, it may be hard to give a meaningful interpretation to these unobserved constructs and this interpretation may not be universally accepted among scholars. Another drawback of factor analysis is that the commonly used estimation procedures (i.e. principal components or maximum likelihood) do not take into account the response variable in the model. This is a limitation, because in a regression context one wants to use different information in the explanatory variables depending on the response. Partial least squares or sliced inverse regression (Li, 1991; Naik, Hagerty, & Tsai, 2000) do take into account the response variable in the construction of the factors, but the interpretation of the resulting factor model becomes even more difficult. It is hard to argue that the factors represent an underlying construct if they by definition are different depending on the response variable in the model.

3. Method

In this section, we first review three penalized likelihood methods, Ridge regression, the Lasso and the Elastic Net. The latter two have a variable selection property which allows exploring which variables matter most. Next, we draw the analogy with Bayesian regularization through the choice of appropriate priors on the regression coefficients. We then describe the Bass diffusion model and illustrate the properties of the three regularization methods, as compared to the standard regression using diffuse normal priors, in the Bass diffusion model using a simulation study.

3.1. Penalized likelihood and Bayesian regularization

Consider the multiple linear regression model

$$y = Xb + e \tag{1}$$

where y is the response vector and X is the $(N \times k)$ matrix containing k regressors. Assume the response to be mean-centered and the regressors to be standardized such that no intercept is included. Furthermore, let $b = (b_1, \ldots, b_k)'$ denote the vector of regression coefficients. Assuming that the error term e follows a $N(0, \sigma_e^2)$ distribution, the penalized likelihood estimator maximizes the likelihood under a constraint on the coefficients. The constraints we consider here are designed to shrink the estimated parameters towards zero. In particular, the three penalized estimators we consider are all of the form

$$\hat{b} = \arg\min_b \sum_{i=1}^{N} (y_i - X_i b)^2 \text{ subject to } (1-\alpha)\sum_{j=1}^{k} \left|b_j\right| + \alpha \sum_{j=1}^{k} b_j^2 < t \tag{2}$$

for some positive value of t. For $\alpha = 1$, the estimator defined by Eq. (2) is the Ridge estimator, which puts a constraint on the sum of the squared coefficients. For $\alpha = 0$, the constraint is on the sum of the absolute values of the coefficients, which yields the Lasso estimator. Any value of α such that $0 < \alpha < 1$, results in the Elastic Net estimator. The Elastic Net constraint on the coefficients is a combination of the Ridge and the Lasso constraints.

To illustrate the difference between Ridge, Lasso and Elastic Net, the shrinkage obtained by each method is illustrated in Figs. 1, 2 and 3 for the case with only two regressors. The gray area in the figures specifies the region within which the coefficients on the axis are subject to the constraint in Eq. (2) for a certain value of t. A larger value of t would correspond to a less stringent constraint on the parameters, which would be represented by a larger gray constraint region. The value of t is typically chosen by cross-validation. The ellipses represent equi-mean-squared-error lines. The inner x-mark represents the maximum likelihood solution, which is the solution to problem (2) without a constraint on the coefficients. The inner ellipses are closer to the maximum likelihood solution, and thus have lower mean squared error values. For each shrinkage type (Ridge, Lasso or Elastic Net), we present a case with uncorrelated and correlated regressors in subfigures (a) and (b) respectively.

The solution to minimization problem (2) is given by the tangent point between the gray constraint region and the ellipsoid. Fig. 1 illustrates why Ridge regression does not result in variable selection. Because of the circular shape of the constraint region, the Ridge solution will only rarely result in zero coefficient estimates. A problem with Ridge regression is the sensitivity of the outcome to changes in the constraint region, especially when the regressors are correlated (Fig. 1b). If the amount of shrinkage is strong enough, the Ridge coefficients can change signs as compared to the least squares solution, as is the case in Fig. 1b.

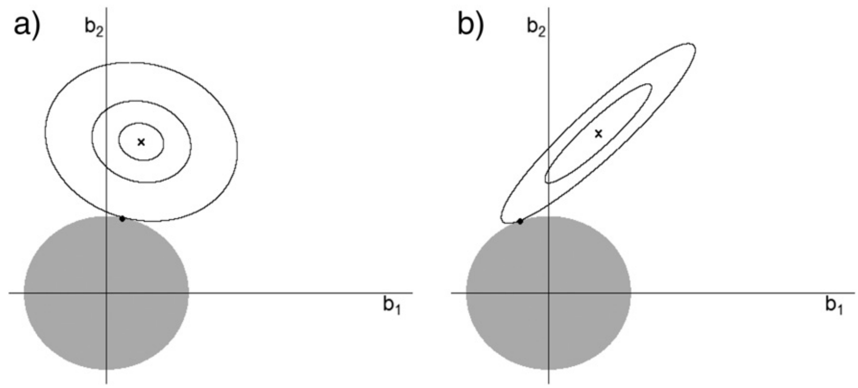

Fig. 1. Ridge solution under uncorrelated regressors (a) and correlated regressors (b). The gray region indicates the region satisfying the regression constraint, the ellipses represent equi-mean-squared-error lines, the least squares solution (x) and the regularized estimates (dot).

The variable selection property of the Lasso is illustrated in Fig. 2. Because of the squared shape of the gray constraint region, the Lasso solution can result in zero coefficients, ensuring variable selection. The tangency point between the gray constraint region and the ellipsoid is on the b_2 axis, resulting in a parameter estimate for b_1 which is exactly equal to zero, both in the uncorrelated case (Fig. 2a) and in the correlated case (Fig. 2b). The Lasso solution is in general more stable than the Ridge solution.

The Elastic Net constraint region presented in Fig. 3 for $\alpha = 0.5$ is an intermediate to the Ridge circular constraint region and the Lasso squared constraint region. The main difference with Ridge is that, similar to the Lasso, the corners of the Elastic Net constraint region facilitate variable selection. The difference with the Lasso is that due to the rounding of the constraint region in between the axes, the Elastic Net tends to select strongly correlated variables jointly in or out the model, which is often referred to as the *grouping effect* (Zou & Hastie, 2005). For instance, in Fig. 3b the correlated variables are selected together by the Elastic Net, while only one variable is selected by the Lasso in Fig. 2b.

The solution to Eq. (2) has a Bayesian interpretation as well. The link between regularization methods and hierarchical Bayes is well documented (e.g. Evgeniou, Pontil, & Toubia, 2007; Fahrmeir et al., 2010). In particular, the solution is equivalent to the posterior mode of the regression coefficients under a specific prior. Bayesian Ridge specifies a normal prior given by

$$b\Big|\sigma_e^2, \lambda_r^2 \sim N\left(\mathbf{0}, \frac{\sigma_e^2}{\lambda_r^2} I_k\right), \tag{3}$$

where the prior mean is zero for all regression parameters and the shrinkage parameter λ_r^2 controls the precision of the prior. A more precise posterior is obtained for larger values of the shrinkage parameter. Taking the prior mean equal to zero in combination with a tight prior is a conservative choice. If after combination with the data the posterior of a parameter is located away from zero, we safely conclude that the corresponding regressor is important in the model. A prior specification for the shrinkage parameter is defined as

$$\lambda_r \sim \text{Gamma}(r_r, s_r) \tag{4}$$

and for the error variance

$$\sigma_e^{-2} \sim \text{Gamma}(u_r, v_r)\ . \tag{5}$$

Posterior evaluation is obtained via the Gibbs sampler.

The disadvantage of Ridge regression is that it does not achieve variable selection. Moreover, the amount of effective shrinkage is hard to control. It not only depends on the shrinkage parameter but also on the amount of correlation in the data. The more correlation, the less stable Ridge regression becomes, which makes it a poor method for data with harmful multicollinearity like ours. This instability is shown by Tibshirani (1996) in a penalized likelihood setting, but also holds in the Bayesian setting as we illustrate in Appendix A.

Following the work of Hans (2009) and Park and Casella (2008), the Lasso point estimator for regression model (1), is defined as the mode of the posterior density of the regression parameters when imposing an independent Laplace prior with mean zero on the regression coefficients

$$b_j|\sigma_e^2, \lambda_l \sim \text{Laplace}\left(0, \frac{\sigma_e}{\lambda_l}\right) = \frac{\lambda_l}{2\sigma_e} \exp\left(\frac{-\lambda_l}{\sigma_e}\Big|b_j\Big|\right). \tag{6}$$

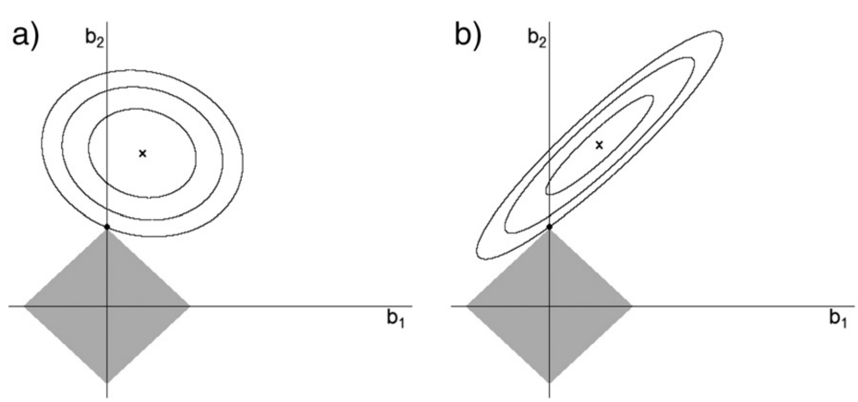

Fig. 2. Lasso solution under uncorrelated regressors (a) and correlated regressors (b). The gray region indicates the region satisfying the regression constraint, the ellipses represent equi-mean-squared-error lines, the least squares solution (x) and the regularized estimates (dot).

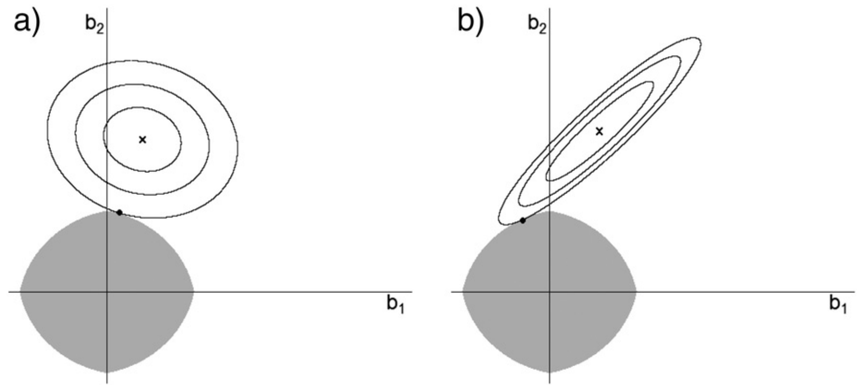

Fig. 3. Elastic Net solution for $\alpha = 0.5$ under uncorrelated regressors (a) and correlated regressors (b). The gray region indicates the region satisfying the regression constraint, the ellipses represent equi-mean-squared-error lines, the least squares solution (x) and the regularized estimates (dot).

As for the Ridge, a more precise posterior is obtained for larger values of the shrinkage parameter at the cost of more shrinkage. Similar to the term $|b_j|$ in the constraint in Eq. (2), the term $|b_j|$ in the prior in Eq. (6) facilitates variable selection. The key to variable selection using this procedure is that, depending on the value of the shrinkage parameter, the posterior mode of some regression coefficients can become exactly zero. Even though there is posterior mass located away from zero, whether the posterior mode of a regression coefficient is zero or not has important consequences for model interpretation. By construction, the mode will always be included in the highest posterior density region. Therefore, a regression parameter with zero posterior mode will never be "significant". Posterior evaluation is achieved via the Gibbs sampler described in Hans (2009). The latter requires a rejection sampling step to draw from the conditional distribution of the scale parameter, which we implemented using the R package ars by Perez Rodriguez (2009).

The Laplace prior puts more prior mass close to zero and in the tails as compared to a normal prior, as illustrated in Fig. 4, reflecting the idea that there are many small effects and a number of important effects. Other variable selection procedures build on the belief that some of the true regression coefficients are exactly zero, which is hard to defend (O'Hara & Sillanpaa, 2009). Especially in the international diffusion model, it is likely that all country characteristics influence the diffusion process, but some variables to a much lesser extent than others. In this context, variable selection should be considered as a tool to help the researcher distinguish between the small and the large important effects

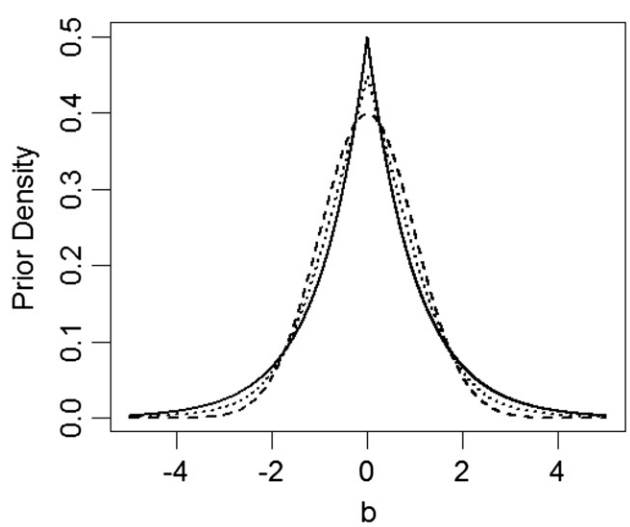

Fig. 4. Normal prior used for Ridge (dashed line), Laplace prior used for Lasso (full line) and mixed prior used for Elastic Net (dotted line).

rather than identifying zero-effects. The Elastic Net prior on the regression coefficients is a compromise between the Gaussian prior of Ridge regression and the Laplace prior of the Lasso (Li & Lin, 2010)

$$p\left(b_j|\sigma_e^2, \lambda_{1n}, \lambda_{2n}\right) \propto \exp\left(-\frac{1}{2\sigma_e^2}\left(\lambda_{1n}\left|b_j\right| + \lambda_{2n} b_j^2\right)\right). \tag{7}$$

A comparison between the priors is given in Fig. 4. The elastic net prior is an intermediate between the Normal and the Laplace prior. The spike at zero facilitates variable selection. The Bayesian Elastic Net has been used in marketing research before by Rutz, Trusov, and Bucklin (2011) in the context of paid search advertising.

3.2. Bayesian representation of the international Bass diffusion model

We use the Bayesian regularization methods as described in the previous section to identify which country characteristics best explain differences in diffusion patterns. To specify a Bayesian version of the Bass diffusion, denote by $S_{ij}(t)$ the penetration level of product j in country i at period t after commercialization. The diffusion process of product j in country i is given by

$$\Delta S_{ij}(t) = \left(p_{ij} + q_{ij}\frac{S_{ij}(t-1)}{m_{ij}}\right)\left(m_{ij} - S_{ij}(t-1)\right) + \varepsilon_{ij}(t), \tag{8}$$

where $\Delta S_{ij}(t) = S_{ij}(t) - S_{ij}(t-1)$, and $\varepsilon_{ij} \sim N(0, \sigma_j^2)$. The first parameter m_{ij} captures the market potential, p_{ij} is the coefficient of innovation, and q_{ij} is the coefficient of imitation for product j in country i. We include an additive error term in Eq. (8) following Albuquerque et al. (2007) to ensure that penetration levels are allowed to show small decreases over time, as is observed in our data.

To know which country characteristics influence the diffusion process, the diffusion parameters m_{ij}, p_{ij} and q_{ij} are first decomposed into a country- and product-specific component after controlling for the product-country specific introduction lag denoted by L_{ij}. Denote the vector of Bass model parameters for product j in country i by $\theta_{ij} = (m_{ij}, p_{ij}, q_{ij})$, then the variance decomposition is given by

$$\text{logit}\left(\theta_{ij}\right) = \alpha_i + \beta_j + \gamma L_{ij} + \xi_{ij} \quad \text{with} \quad \xi_{ij} \sim N\left(\mathbf{0}, \Sigma_\xi\right), \tag{9}$$

where we allow a full covariance matrix Σ_ξ. Since the values of $\theta_{ij} = (m_{ij}, p_{ij}, q_{ij})$ are between zero and one, we use a logit transformation to obtain values on the whole real line, which is similar to the approach in Lenk and Rao (1990). The first component of the γ vector is fixed at zero because the introduction lag only affects the growth rate towards the market potential (determined by p_{ij} and q_{ij}), and not the market potential m_{ij} itself.

Since our interest is in the country-specific parameters in vector α_i, it is further regressed on the country characteristics. These are represented in the matrix X of dimension $(C \times k)$, with C the number of countries and k the number of country characteristics. The third level of the Bass diffusion model then is of the form

$$\alpha_i = X_i\delta + \eta_i \qquad \text{with} \qquad \eta_i \sim N\left(\mathbf{0}, \Sigma_\eta\right), \tag{10}$$

where X_i is the row vector of length k with country characteristics for country i. The regression parameter matrix δ is of dimension $(k \times 3)$ and captures the effect of the country characteristics on the diffusion process. The matrix δ is our primary object of interest – it captures the influence of the country characteristics on the diffusion pattern – and is estimated using Bayesian regularization as described in Section 3.1.

The product-specific effects are captured in the parameter vector β_j which is modeled as a random effect with mean zero (for identification)

$$\beta_j \sim N\left(\mathbf{0}, \Sigma_\beta\right). \tag{11}$$

We assume Σ_η and Σ_β to be diagonal. All prior specifications are given in Appendix B1. The posterior and estimation details of the first level are given in Appendix B2.

Posterior evaluation of the parameters is achieved through MCMC draws. In the Lasso and Elastic Net case, apart from the posterior MCMC draws we are interested in the posterior mode of the regression coefficients in δ because the mode marks selection. The mode is obtained by maximum a posteriori (MAP) estimation. MAP estimation in the Bayesian Lasso setting is common, see e.g. Figueiredo (2003) and Genkin, Lewis, and Madigan (2007). The MAP estimator is obtained using Rao-Blackwellization as in Hans (2009) and Hans (2011). For each draw in the MCMC chain, we store the conditional distribution of δ on a fine grid. This conditional distribution is orthant normal for both Lasso and Elastic Net and sometimes has a zero-mode due to the shape of the prior. We then average the stored conditionals over the MCMC draws for each grid point to obtain an estimate of the marginal posterior from which we can easily obtain the mode as the Lasso or Elastic Net point estimate.

3.3. Simulation study

We run a simulation study to assess the performance of the Bayesian regularization methods described in Section 3.1 for estimating the country-level regression model parameters in the Bass diffusion model of Section 3.2. To assess in which conditions the Bayesian Lasso and Elastic Net perform better than Ridge or regression using diffuse normal priors, we run a 2×2 simulation design. As country covariates are typically highly correlated, the first dimension we vary is the amount of multicollinearity. We compare the accuracy of the regularization procedures across two settings, one in which covariates are correlated and one in which covariates are uncorrelated. The second dimension we take into consideration is the sparseness of the true model, i.e. whether some of the country covariates have an actual *zero* effect on the diffusion process. Due to their variable selection properties, these sparse models are favored by the Lasso and the Elastic Net. But since we do not know whether there truly are zero effects, we study the methods' performance in a situation where all country covariates have an effect but some have a stronger effect than others. This leads to four simulation settings where we have either correlated covariates or uncorrelated covariates, and either true model sparseness or not.

The specifics of the simulation setting are as follows. We simulate data according to the multi-product multi-country Bass diffusion model specified by Eqs. (8) to (11). The dimensions of the model are the same as in our data, i.e. we simulate 6 products, 55 countries and 17 country covariates ($k = 17$). We generate the country covariates X from a normal distribution with mean zero. In the correlated settings, the correlation between x_i and x_j equals $\rho^{|i-j|}$ with $\rho = 0.5$, following the simulation setup of Tibshirani (1996). In the uncorrelated settings, we set $\rho = 0$. In the sparse settings, we again follow Tibshirani (1996) and set $\delta_{\cdot j} = (3,1.5,0,0,2,0,\ldots,0)'$ for $j \in \{1,2,3\}$ corresponding to the diffusion metrics m, p and q respectively. In the non-sparse settings, we set $\delta_{\cdot j} = (3, 1.5, 1, 3/4, 3/5, \ldots, 3/17)'$ such that each covariate influences the diffusion process but the last covariates are gradually less important than the first. To make the simulation specification complete, we set $\sigma_j^2 = 0.01$, $L_{ij} = 0$ for all i and j, $\gamma = \mathbf{0}$, $\Sigma_\xi = \Sigma_\eta = \Sigma_\beta = I_3$ and we generate $N_s = 200$ data sets in each simulation setting.

As our main interest is in the performance to retrieve the parameters of the country-level regression models in δ, we compare the mean squared error

$$\text{MSE} = \frac{1}{3kN_s}\sum_{j=1}^{3}\sum_{i_s=1}^{N_s}\left(\hat{\delta}_{\cdot j} - \delta_{\cdot j}\right)'\left(\hat{\delta}_{\cdot j} - \delta_{\cdot j}\right), \tag{12}$$

where $\hat{\delta}$ is the vector of point estimates of the country-covariate effects. For the Lasso and Elastic Net, we use the posterior mode as described above. For Ridge regression and regression using diffuse normal priors, we use the posterior median as a point estimator. All MSE values are computed based on standardized variables.

For the sparse simulation settings, we also assess how well the Lasso and Elastic Net perform in terms of identifying those variables that have a non-zero coefficient. We compute the true positive rate (TPR) as the proportion of non-zero coefficients that are estimated to be non-zero, i.e. are correctly selected into the model. We also compare the true negative rate (TNR) as the proportion of zero coefficients that are estimated to be zero, i.e. correctly estimated as having a zero-effect:

$$\begin{aligned} \text{TPR} &= \frac{\#\left\{(ij) : \hat{\delta}_{ij} \neq 0 \text{ and } \delta_{ij} \neq 0\right\}}{\#\left\{\delta_{ij} \neq 0\right\}}, \\ \text{TNR} &= \frac{\#\left\{(ij) : \hat{\delta}_{ij} = 0 \text{ and } \delta_{ij} = 0\right\}}{\#\left\{\delta_{ij} = 0\right\}}. \end{aligned} \tag{13}$$

The mean squared error values are presented in Table 2. Overall, the Lasso achieves the best MSE values in all simulation settings. The benefit of the Lasso over the other methods, however, differs across the settings. The advantage of the Lasso is most pronounced when the covariates are correlated and the true model is sparse (Setting 4) and least pronounced when there is no multicollinearity and all predictors have an influence on the diffusion process (Setting 1). When the covariates are uncorrelated and the true model is sparse, both the Lasso and Elastic Net – which favor models with zero-coefficients – perform better than estimation based on diffuse normal priors and Ridge (Setting 2).

Table 2
Estimation accuracy (MSE ∗ 1000).

			Lasso	Elastic Net	Diffuse normal priors	Ridge
Setting 1	Uncorrelated covariates	Non-sparse model	8.96	9.10	10.96	9.10
Setting 2		Sparse model	6.12	8.04	11.81	11.52
Setting 3	Correlated covariates	Non-sparse model	9.28	10.98	15.47	11.89
Setting 4		Sparse model	7.00	9.13	16.88	12.18

Table 3
Variable selection accuracy.

			True positive rate (TPR)		True negative rate (TNR)	
			Lasso	Elastic Net	Lasso	Elastic Net
Setting 2	Uncorrelated covariates	Sparse model	0.93	0.99	0.81	0.34
Setting 4	Correlates covariates	Sparse model	0.90	0.98	0.60	0.23

Table 4
Products and countries in our data set.

Products	Mobile phone, CD player, video camera, PC, Internet, ISDN
Countries	Argentina, Australia, Austria, Belgium, Brazil, Bulgaria, Canada, Chile, China, Colombia, Croatia, Czech Republic, Denmark, Ecuador, Estonia, Finland, France, Germany, Greece, Hong Kong, Hungary, India, Indonesia, Ireland, Israel, Italy, Japan, Malaysia, Mexico, Morocco, Netherlands, New Zealand, Norway, Pakistan, Peru, Philippines, Poland, Portugal, Romania, Russia, Singapore, Slovakia, Slovenia, South Africa, South Korea, Spain, Sweden, Switzerland, Taiwan, Thailand, Turkey, UK, USA, Venezuela, Vietnam

When the covariates are correlated and all of them have an effect, the Lasso and Elastic Net clearly outperform estimation based on diffuse normal priors and Ridge (Setting 3). In sum, whether the true model is sparse or not, methods like the Lasso and Elastic Net should be considered as methods that lead to superior outcomes when multicollinearity is present in the data.

The variable selection accuracy of the Lasso and the Elastic Net are reported in Table 3. Elastic Net has a better true positive rate than the Lasso, at the cost of a lower true negative rate. This holds true in a setting where the covariates are uncorrelated as well as when they are correlated. The correlated setting (Setting 4) is especially of interest because the Elastic Net was introduced as a method that performs better when the covariates are correlated. The *grouping effect* states that the Elastic Net tends to select groups of correlated variables jointly. In our sparse settings, the first two variables both have an effect and are correlated. The fifth variable also has an effect and is correlated with variables that have a zero-effect. In this setting, the Lasso has a true positive rate of 90% while the Elastic Net achieves 98%. However, the Elastic Net tends to select too many variables into the model that are correlated with those variables that have an effect. As a result, the true negative rate of the Elastic Net is only 23%, while that of the Lasso is 60%, which illustrates the difference between both methods in terms of variable selection when the covariates are correlated.

4. Data

We use penetration data of six consumer durables in 55 countries listed in Table 4, gathered from publicly available sources, such as Euromonitor and the International Telecommunications Union. The country characteristics were gathered from publicly available sources such as the Statistical Yearbook of the United Nations, CIA World Factbook, World Development Indicators, U.S. Census Bureau, Euromonitor online, and Hofstede (2001). Country characteristics with multiple data points over the observation period were averaged.

We rely on the new product adoption and diffusion literature to specify our model in terms of country covariate inclusion. Table 5 gives an overview of the covariates we include, where the inclusion criterion is whether the variable has been used in previous diffusion literature. The country characteristics cover socio-economic, cultural, communication and demographic dimensions. The last column of Table 5 indicates to which growth metric (market potential, coefficient of innovation or coefficient of imitation) prior studies related each country covariate. To showcase the ability of variable selection methods to deal with long models, we link all available country characteristics to each diffusion metric. This procedure will allow us to explore whether or not there are important relationships that have not been identified or theorized on before.

To assess the degree of multicollinearity in our dataset, we compute the condition index of X as in Belsley, Kuh, and Welsh (1980). To obtain the condition index, we scale the variables in the X-matrix to have unit variance. According to Belsley et al. (1980), condition indices above 30 indicate moderate to strong multicollinearity. In our case, we obtain a condition index of 79.63, which is well beyond the threshold.

Table 5
List of included country characteristics.

Dimension	Variable	Operationalization	Related to diffusion metrics in previous literature
Socio-economic	Economic wealth	GDP per capita	m, p, q
	Inequality	GINI index on the household level based on net income	m, p, q
	Poverty	Under 5 year mortality rate	p, q
	Economic openness	(import + export)/GDP	m
	Education	Number of third-degree (university) students as a percentage of total population.	p
	Activity rate of women	% of women employed in nonagricultural sector	p, q
	Economic participation	Working to dependents ratio	m, p
Cultural	Individualism	Hofstede IND	p, q
	Uncertainty avoidance	Hofstede UAI	p, q
	Masculinity	Hofstede MAS	p, q
	Power distance	Hofstede PDI[a]	p, q
Communication	Media intensity	Number of newspapers per 1000 inhabitants	p
	Mobility	Number of cars per 1000 inhabitants	q
	Tourism	Number of incoming tourists per 1000 inhabitants	p, q
Demographic	Population growth	Annual population growth rate	p, q
	Population concentration	Population per square km	p, q
	Urbanization	% op people living in urban areas	m

[a] We do not include the fifth cultural dimension, Hofstede later added to his cultural framework, long-term orientation, because it is not available for a large number of countries in our dataset.

5. Results

5.1. Variable selection: Bayesian Lasso and Elastic Net

Table 6 presents the selected variables obtained by the Lasso and the Elastic Net and the posterior mode for a sequence of 10,000 draws after 2000 burn-in draws. The *prop*-values are the proportion of draws on the other side of zero than the mode. Because a variable is unselected from the model when the posterior mode is equal to zero, a *prop*-value cannot be calculated in such cases.

For all diffusion metrics, the predominant variable is economic wealth. Both the Lasso and the Elastic Net find that economic wealth has a positive effect on all three diffusion metrics. Talukdar et al. (2002) also found a strong effect of economic wealth on market potential but did not allow for an effect on the innovation and imitation coefficients, while according to our results this effect is strong. A second important variable is education, which influences both the market potential (m) and the innovation coefficient (p).

Apart from economic wealth and education, we find a distinct set of additional country covariates that affect the three diffusion metrics. We find a negative effect of income inequality on market potential. That is, all other things being equal, product adoption reaches a lower ceiling in such countries. We also find a positive effect of tourism on the market potential. Cosmopolitanism, which includes tourism, was one of the core constructs in the early studies of Gatignon et al. (1989) and Helsen et al. (1993). For the innovation coefficient, we find a positive effect of education and economic openness. We find a positive effect of mobility on the imitation coefficient, supporting the hypothesis that if people are more mobile they get in contact with more people and thus have a higher probability of influencing each other. All the remaining variables were not selected. Thus, after controlling for the included variables, they do not provide any additional information about the diffusion process, in our sample of products and countries. The latter subsentence is important and applies to all our findings reported in the present paper; to our experience, findings on international diffusion of new products are sensitive not only to the variable selection technique employed, but also to the sample composition in terms of which products and countries are covered as well as the extent to which such sample is balanced (i.e., the same products are covered across the same set of countries).

Table 5 summarizes which variables have been used as a driver of which metric in the previous literature. Including all variables as determinants of all diffusion metrics allowed us to extract three new findings on international diffusion. The first is the effect of education on market potential. All else equal, in a more educated population a higher proportion of the population will adopt new technologies. The second is the effect of tourism on market potential. The more touristic a country is, the more the population will get into contact with new technologies and thus the more people will eventually adopt. The third new effect is that of economic openness on the innovation coefficient.

The variable selection results obtained by the Lasso and the Elastic Net are very similar. Even though the Elastic Net is more sophisticated – as it chooses the intermediate between Ridge and Lasso in a data-adaptive way – this extra level of sophistication does not lead to substantially different insights in our setting. Fig. 5 compares the marginal densities of the effects of economic wealth and population growth on the market potential as estimated by the Lasso and the Elastic Net. Both methods identify economic wealth as an important variable. The Elastic Net posterior shrinks a bit more to zero, but there is no difference in substantive interpretation. As an illustration of an unselected variable, the right panel of Fig. 5 plots the posterior densities of the effect of population growth on market potential. Both posterior modes are zero, while the Lasso posterior is a bit more spiked. Similar comparisons between Lasso and Elastic Net posteriors are reported in Hans (2011) who finds small differences in the Lasso and Elastic Net posteriors using prostate cancer data (Stamey et al., 1989).

5.2. Diffuse normal priors and ridge

In Table 7, we report the results after estimating the Bass diffusion model using diffuse normal priors on all regression coefficients in Eq. (10) and Ridge regression. Diffuse normal priors are the most standard choice in Bayesian regression and are used in the Bass diffusion model by Talukdar et al. (2002), while Ridge regression is an alternative shrinkage method without the variable selection property as described above in Section 3. In the case of diffuse normal priors, none of the estimated effects is significant and so no conclusions can be drawn with respect to which variables influence which metric. When we do Bayesian regularization using Ridge, we only identify a positive effect of economic wealth on market potential. These poor conclusions with respect to

Table 6
Variable selection and significance for the Bayesian Lasso and Elastic Net procedure for each diffusion metric.

	Lasso						Elastic Net					
	Market potential		Innovation coefficient		Imitation coefficient		Market potential		Innovation coefficient		Imitation coefficient	
	Posterior mode	*Prop*-val	Posterior mode	*Prop*-val	Posterior mode	*Prop*-val	Posterior mode	*Prop*-val	Posterior mode	*Prop*-val	Posterior mode	*Prop*-val
Economic wealth	**.49**	**.02**	**.39**	**.02**	**.21**	**.03**	**.38**	**.02**	**.29**	**.05**	**.02**	**.04**
Inequality	**−.05**	**.04**	0		0		**−.04**	**.03**	0		0	
Poverty	0		0		0		0		0		.01	.42
Economic openness	.11	.16	**.10**	**.03**	0		.03	.32	.20	.23	0	
Education	**.20**	**.02**	**.15**	**.04**	0		**.01**	**.04**	.06	**.04**	.03	.11
Activity rate of women	0		0		0		0		0		0	
Economic participation	.05	.33	.04	.43	0		−.08	.20	0		0	
Individualism	0		0		0		0		0		−.25	.15
Uncertainty avoidance	0		0		0		−.03	.34	0		0	
Masculinity	0		0		.10	.45	0		.05	.35	.11	.24
Power distance	0		0		0		0		0		0	
Media intensity	0		0		0		0		0		0	
Mobility	.11	.22	0		**.46**	**.04**	.21	.09	0		**.30**	**.03**
Tourism	**.07**	**.03**	.10	.39	0		0		−.09	.31	0	
Population growth	0		0		0		0		.15	.25	0	
Population concentration	0		0		.05	.41	0		0		0	
Urbanization	0		.03	.25	0		.17	.07	0		.01	.40

Note: The posterior mode is the point estimate of the Lasso or Elastic Net. The *prop*-value is the proportion of draws on the other side of zero than the mode, indicating significance. The *prop*-value cannot be computed if the mode is zero, resulting in blank entries. Parameter estimates and *prop*-values are indicated in bold when the *prop*-value is less than .05.

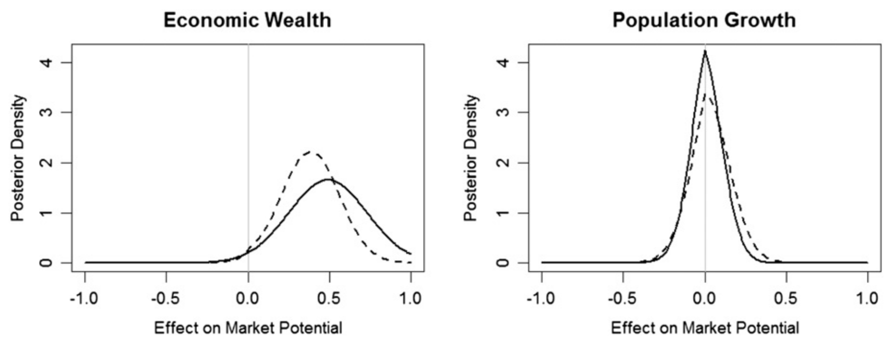

Fig. 5. Density estimates of the marginal posterior distributions of the effect of economic wealth (left) and population growth (right) on the market potential estimated using Lasso (full line) and Elastic Net (dashed line).

which country characteristic influences which diffusion metric are the result of the sparseness of the data. As we illustrated in the simulation section above, diffuse normal priors and ridge regression are poorly suited for a multicollinear setting like ours.

6. Discussion

Using the Bayesian Lasso and Elastic Net estimation procedures, we have shown that international variation in new product growth in our sample of products and countries is predominantly driven by economic wealth and education. In addition, economic inequality limits a new product's market potential. The innovation coefficient is also higher the higher the level of economic openness in a country. The imitation coefficient is higher, the higher the mobility of a country's citizens.

The application of Bayesian Lasso and Elastic Net on a larger sample of new products beyond high technology products (our present sample), such as laundry and appliances (e.g. Kumar & Krishnan, 2002), fast moving consumer goods, services, pharmaceuticals and entertainment products, may bring strong generalizable insights (on main effects or contingencies) to the international diffusion literature. The set of countries and products used in international diffusion studies will always have large effects on the findings given large product-country interactions (Talukdar et al., 2002). An update to the meta-analytic approach, such as in Van den Bulte and Stremersch (2004) could therefore prove to be a valuable contribution to the international diffusion literature.

Such applications could also easily further enlarge the set of country covariates to variables that so far received little attention, such as distribution infrastructure, competition, or regulation (see Stremersch & Lemmens, 2009, for an exception), to yield newly discovered strong determinants of international diffusion patterns. The methodology we propose is ideally suited to handle even larger covariate sets. One particular fruitful challenge lies in the study of interaction effects among country covariates. While, the Bayesian Lasso and Elastic Net cannot guarantee the inclusion of a main effect conditional on the inclusion of an interaction, Bien, Taylor, and Tibshirani (2013) propose a non-Bayesian variant of the Lasso which does exactly that. There is room for a methodological contribution to extend such an approach to the Bayesian world.

In addition to the above applications, the present paper shows several additional limitations the reader should be aware of. It is well known that model averaging approaches substantially improve the prediction accuracy as opposed to fitting one single model (Eklund & Karlsson, 2007; Raftery, Madigan, & Hoeting, 1997; Wright, 2008). A model

Table 7
Parameter estimates and significance for procedures using diffuse normal priors and Ridge for each diffusion metric.

	Diffuse normal priors						Ridge					
	Market potential		Innovation coefficient		Imitation coefficient		Market potential		Innovation coefficient		Imitation coefficient	
	Posterior median	*Prop*-val	Posterior median	*Prop*-val	Posterior median	*Prop*-val	Posterior median	*Prop*-val	Posterior median	*Prop*-val	Posterior median	*Prop*-val
Economic wealth	.42	.15	−.10	.42	−.22	.32	**.30**	**.01**	.04	.24	.10	.40
Inequality	−.07	.40	−.01	.49	−.03	.46	−.06	.26	.12	.12	−.10	.30
Poverty	−.08	.32	−.02	.46	.02	.45	−.06	.21	.01	.36	−.07	.42
Economic openness	.07	.39	.09	.39	.01	.48	.05	.32	.00	.39	−.10	.35
Education	.02	.46	−.05	.44	−.03	.46	.04	.33	−.01	.27	−.02	.32
Activity rate of women	−.02	.46	.02	.48	.00	.50	.01	.44	−.02	.26	−.10	.21
Economic participation	−.11	.33	.01	.48	−.05	.43	−.08	.19	.02	.26	.06	.37
Individualism	−.10	.37	−.26	.25	−.33	.19	.00	.50	−.41	.39	.38	.39
Uncertainty avoidance	−.06	.41	.11	.38	.06	.44	−.04	.32	.01	.36	.25	.35
Masculinity	−.01	.48	.09	.34	.14	.25	−.03	.35	.01	.46	−1.32	.44
Power distance	−.12	.31	−.09	.39	−.03	.47	−.11	.13	.02	.26	−.06	.22
Media intensity	.10	.37	.04	.45	.07	.42	.15	.08	−.03	.25	.10	.23
Mobility	.24	.25	.10	.41	.07	.43	.19	.05	−.05	.34	.09	.30
Tourism	−.03	.45	−.14	.31	−.06	.41	.01	.47	−.06	.43	−.19	.47
Population growth	.04	.45	.19	.30	.16	.32	.03	.38	.29	.40	−.08	.38
Population concentration	.00	.49	−.02	.48	−.12	.38	.06	.28	.00	.39	.34	.31
Urbanization	.20	.21	.05	.44	.08	.40	.16	.04	.00	.28	−.26	.18

Note: The posterior median is the point estimate of the estimation using diffuse normal priors and Ridge. The *prop*-value is the proportion of draws on the other side of zero than the median, indicating significance. Parameter estimates and *prop*-values are indicated in bold when the *prop*-value is less than .05.

averaging approach to the Bayesian Lasso and Elastic Net as in Hans (2010) and Hans (2011) respectively could be used for predicting the diffusion metrics of a product in a certain country before launch (van Everdingen, Aghina, & Fok, 2005).

Second, the Bayesian version of the international Bass diffusion model could be formulated more flexibly. Previous research suggested making the error term of the Bass diffusion model autocorrelated and heteroskedastic (Fok & Franses, 2007). The Bayesian Lasso and Elastic Net procedures we use in this paper can be easily implemented in such alternative diffusion models.

Third, the variable selection techniques introduced in this paper can be extended to other models that capture international new product growth patterns, such as duration models for time-to-adoption, time-to-takeoff (e.g. Tellis et al., 2003; van Everdingen et al., 2009) or duration of the growth stage (Stremersch & Tellis, 2004), as well as to semiparametric sales models (e.g. splines), etc.

Despite these limitations, this paper contributes to marketing scholars' knowledge on dealing with sparse data, and offers a solution that is relatively easy to implement. It is clear that the on-going modeling practice, as documented here for international diffusion, can be improved substantially by implementing regularization methods, such as the Lasso and Elastic Net used in this paper. Such improvement would not only benefit the reliability of scholarly evidence, but would also allow to simultaneously explore a larger set of covariates and derive new empirical evidence on factors that remained uninvestigated so far. Within the research area of international diffusion, duration models (Dekimpe, Parker, & Sarvary, 2000a; Dekimpe et al., 2000b) such as the time-to-takeoff model (Chandrasekaran & Tellis, 2008; Stremersch & Tellis, 2004; Tellis et al., 2003; van Everdingen et al., 2009) could lead to additional insights into which variables explain international differences in the timing pattern of diffusion. Areas that come to mind outside the research area of new product diffusion (Naik et al., 2008) include churn modeling, in which datasets with more than 100 explanatory variables are not an exception (Lemmens & Croux, 2006; Naik, Wedel, & Kamakura, 2010), or the vector autoregressive (VAR) modeling tradition of marketing effectiveness (Dekimpe & Hanssens, 1999; Srinivasan, Pauwels, & Nijs, 2010), in which the number of regression parameters dramatically explodes as the number of variables increases. We hope that the benefit of the procedure proposed in the present paper does not remain contained to international diffusion models, but rather diffuses to other research areas in marketing science as well.

Appendix A. The Lasso versus the Ridge prior

Tibshirani (1996) shows that Ridge regression is sensitive to the amount of correlation between regressors, while the Lasso is not. We show that the same conclusions hold in a Bayesian setting. We generate 1000 data points according to the following model

$$y = \beta_1 x_1 + \beta_2 x_2 \ ,$$

without error term. The regression parameters are fixed at $\beta_1 = 6$ and $\beta_2 = 3$. The regressors x_1 and x_2 are obtained from a standard normal distribution with correlation ρ. Since no error term is included in the data generating process, we use the original Lasso prior introduced in Tibshirani (1996), which is unconditional on the scale parameter (Hans, 2009).

The parameter estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ for different shrinkage parameters and different correlations ($\rho = 0, 0.23, 0.45, 0.68, 0.9$) are plotted in Fig. A.1. It clearly illustrates the sparseness of the Lasso. If much shrinkage is applied, the estimate of β_2 is zero. In contrast to the Lasso, the Ridge solution strongly depends on ρ. For high correlations ($\rho = 0.9$), the Ridge procedure sometimes even overestimates the true parameter instead of shrinking it to zero. The variation in the Lasso estimates for different values of ρ is not systematic and only due to the variation in the random generation of the regressors.

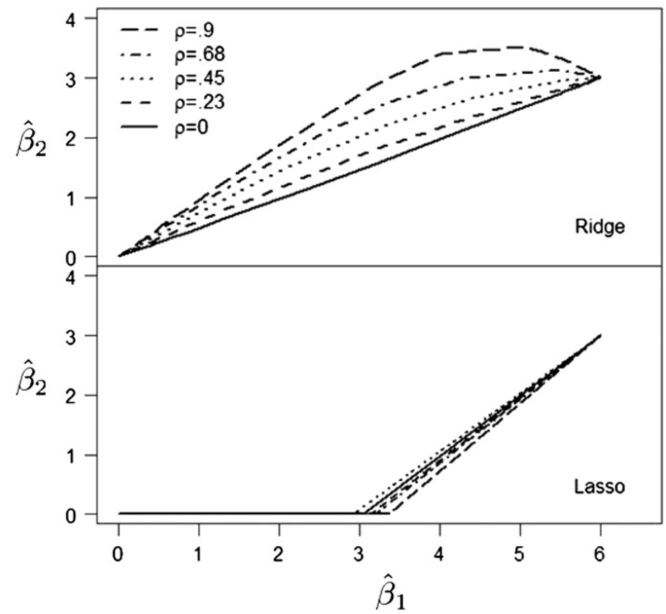

Fig. A.1. Ridge (upper panel) and Lasso (lower panel) regression estimates for different levels of shrinkage and for different values of the correlation coefficient ρ.

Appendix B. Model specification and estimation

Appendix B1. Prior specifications

Diffuse normal priors
$\sigma_j^{-2} \sim \text{Gamma}(1, 10)$
$\gamma \vert \Sigma_\xi \sim N(\mathbf{0}, 100\Sigma_\xi)$
$\Sigma_\xi \sim \text{Wishart}(5, 0.1I_3)$
$\delta_j^c \vert \sigma_{\eta,c}^2 \sim N(0, 100\sigma_{\eta,c}), \ c = 1, 2, 3 \text{ and } j = 1, \ldots, k$
$\sigma_{\eta,c}^{-2} \sim \text{Gamma}(1, 10)$
$\sigma_{\beta,c}^{-2} \sim \text{Gamma}(1, 10)$
Ridge priors
$\sigma_j^{-2} \sim \text{Gamma}(1, 10)$
$\gamma \vert \Sigma_\xi \sim N(\mathbf{0}, 100\Sigma_\xi)$
$\Sigma_\xi \sim \text{Wishart}(5, 0.1I_3)$
$\delta_j^c \vert \sigma_{\eta,c}^2 \sim N(0, \sigma_{\eta,c}/\lambda_{r,c}^2), \ c = 1, 2, 3 \text{ and } j = 1, \ldots, k$
$\sigma_{\eta,c}^{-2} \sim \text{Gamma}(1, 10)$
$\lambda_{r,c}^2 \sim \text{Gamma}(1, 1)$
$\sigma_{\beta,c}^{-2} \sim \text{Gamma}(1, 10)$
Lasso priors
$\sigma_j^{-2} \sim \text{Gamma}(1, 10)$
$\gamma \vert \Sigma_\xi \sim N(\mathbf{0}, 100\Sigma_\xi)$
$\Sigma_\xi \sim \text{Wishart}(5, 0.1I_3)$
$\delta_j^c \vert \sigma_{\eta,c}^2 \sim \text{Laplace}(0, \sigma_{\eta,c}/\lambda_{l,c}), \ c = 1, 2, 3 \text{ and } j = 1, \ldots, k$
$\sigma_{\eta,c}^{-2} \sim \text{Gamma}(1, 10)$
$\lambda_{l,c} \sim \text{Gamma}(1, 1)$
$\sigma_{\beta,c}^{-2} \sim \text{Gamma}(1, 10)$
Elastic Net priors
$\sigma_j^{-2} \sim \text{Gamma}(1, 10)$
$\gamma \vert \Sigma_\xi \sim N(\mathbf{0}, 100\Sigma_\xi)$
$\Sigma_\xi \sim \text{Wishart}(5, 0.1I_3)$
$\delta_j^c \vert \sigma_{\eta,c}^2 \propto exp((-1/(2\sigma_{\eta,c}))(\lambda_{1n,c}\vert\delta_j^c\vert + \lambda_{2n,c}(\delta_j^c)^2)), \ c = 1, 2, 3 \text{ and } j = 1, \ldots, k$
$\sigma_{\eta,c}^{-2} \sim \text{Gamma}(1, 10)$
$\lambda_{1n,c}^2 \sim \text{Gamma}(1, 1)$
$\lambda_{2n,c}^2 \sim \text{Gamma}(1, 1)$
$\sigma_{\beta,c}^{-2} \sim \text{Gamma}(1, 10)$

Appendix B2. MCMC draws in the first level of the Bass diffusion model

The parameters are estimated by drawing from their conditional posterior. In the first level of the Bass diffusion model, $\theta_{ij} = (m_{ij}, p_{ij}, q_{ij})'$ is obtained by a Metropolis–Hastings step. The posterior of θ conditional on σ^2 can be written as

$$p\left(\theta|\Delta S(t), \sigma^2\right) \propto \ell(\theta)\ p(\theta)\ ,$$

where we drop subscripts to avoid notational clutter. The first component on the right hand side is the likelihood function and the second component the prior. The likelihood is given by

$$\ell(\theta) \propto \prod_{t=2}^{T} \exp\left\{-\frac{1}{2\sigma^2}\left[\left(p+q\frac{S(t-1)}{m}(m-S(t-1))\right]^2\right\}.$$

The prior follows a logistic normal distribution given by

$$p(\theta) = \frac{\prod_{i=1}^{3}\theta_i(1-\theta_i)}{|2\pi\Sigma|^{\frac{1}{2}}} exp\left\{-\frac{1}{2}[\text{logit}(\theta)-\mu]'\Sigma^{-1}[\text{logit}(\theta)-\mu]\right\}.$$

More details on the logistic-normal distribution can be found in Atchison and Shen (1980). The parameter vector μ and matrix Σ are obtained from the second-level estimation. To obtain a candidate draw from $p(\theta|\Delta S(t), \sigma^2)$, we use a normal random-walk candidate generating function with variance such that the acceptance level is approximately 0.3. Denote the current value of θ by θ^c, then the candidate θ^* is accepted with probability $\min(1, p(\theta^*|\Delta S(t), \sigma^2)/p(\theta^c|\Delta S(t), \sigma^2))$.

Next, we draw σ_j from its conditional posterior distribution

$$\sigma_j^{-2}\Big|\Delta S(t), \theta\ \sim\ \text{Gamma}\left(1+n_j, \frac{10+n_j s_j^2}{1+n_j}\right),$$

where n_j is the total number of observations for product j and s_j^2 is given by

$$s_j^2\ = \sum_i \sum_t \left[\left(\Delta S_{ij}(t)\right) - \left(p_{ij} + q_{ij}\frac{S_{ij}(t-1)}{m_{ij}}\right)\left(m_{ij} - S_{ij}(t-1)\right)^2\right]$$

References

Atchison, J., & Shen, S. M. (1980). Logistic-normal distributions: Some properties and uses. *Biometrika*, *67*(2), 261–272.
Albuquerque, P., Bronnenberg, B. J., & Corbett, C. J. (2007). A spatiotemporal analysis of the global diffusion of ISO9000 and ISO14000 certification. *Management Science*, *53*(3), 451–468.
Bass, F. (1969). A new product growth model for consumer durables. *Management Science*, *15*(5), 215–227.
Belsley, D., Kuh, E., & Welsh, R. E. (1980). *Regression diagnostics.* New York: John Wiley & Sons, Inc.
Bien, J., Taylor, J., & Tibshirani, R. (2013). A Lasso for hierarchical interactions. *Annals of Statistics*, *41*(3), 1111–1141.
Chandrasekaran, D., & Tellis, G. J. (2008). Global takeoff of new products: Culture, wealth, or vanishing differences? *Marketing Science*, *27*(5), 844–860.
Dekimpe, M. G., & Hanssens, D.M. (1999). Sustained spending and persistent response: A new look at long-run marketing profitability. *Journal of Marketing Research*, *36*(4), 397–412.
Dekimpe, M. G., Parker, P.M., & Sarvary, M. (1998). Staged estimation of international diffusion models: An application to global cellular telephone adoption. *Technological Forecasting and Social Change*, *57*(1–2), 105–132.
Dekimpe, M. G., Parker, P.M., & Sarvary, M. (2000a). Global diffusion of technological innovations: A coupled-hazard approach. *Journal of Marketing Research*, *37*(1), 47–59.
Dekimpe, M. G., Parker, P.M., & Sarvary, M. (2000b). Globalization: Modeling technology adoption timing across countries. *Technological Forecasting and Social Change*, *63*(1), 25–42.
Eklund, J., & Karlsson, S. (2007). Forecast combination and model averaging using predictive measures. *Econometric Reviews*, *26*(2–4), 329–363.
Evgeniou, T., Pontil, M., & Toubia, O. (2007). A convex optimization approach to modeling consumer heterogeneity in conjoint estimation. *Marketing Science*, *26*(6), 805–818.
Fahrmeir, L., Kneib, T., & Konrath, S. (2010). Bayesian regularisation in structured additive regression: A unifying perspective on shrinkage, smoothing and predictor selection. *Statistics and Computing*, *20*(2), 203–219.
Figueiredo, M.A. T. (2003). Adaptive sparseness for supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*(9), 1150–1159.
Fok, D., & Franses, P. H. (2007). Modeling the diffusion of scientific publications. *Journal of Econometrics*, *139*(2), 376–390.
Gatignon, H., Eliashberg, J., & Robertson, T. S. (1989). Modeling multinational diffusion patterns: An efficient methodology. *Marketing Science*, *8*(3), 231–247.
Genkin, A., Lewis, D.D., & Madigan, D. (2007). Large-scale Bayesian logistic regression for text categorization. *Technometrics*, *49*(3), 291–304.
Hans, C. (2009). Bayesian lasso regression. *Biometrika*, *96*(4), 835–845.
Hans, C. (2010). Model uncertainty and variable selection in Bayesian lasso regression. *Statistics and Computing*, *20*(2), 221–229.
Hans, C. (2011). Elastic Net regression modeling with the orthant normal prior. *Journal of the American Statistical Association*, *106*(496), 1383–1393.
Heeler, R. M., & Hustad, T. P. (1980). Problems in predicting new product growth for consumer durables. *Management Science*, *26*(10), 1007–1020.
Helsen, K., Jedidi, K., & Desarbo, W. S. (1993). A new approach to country segmentation utilizing multinational diffusion patterns. *Journal of Marketing*, *57*(4), 60–71.
Hofstede, G. (2001). *Culture's consequences comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). London: Sage Publications.
Kumar, V., Ganesh, J., & Echambadi, R. (1998). Cross-national diffusion research: What do we know and how certain are we? *Journal of Product Innovation Management*, *15*(3), 255–268.
Kumar, V., & Krishnan, T. V. (2002). Multinational diffusion models: An alternative framework. *Marketing Science*, *21*(3), 318–330.
Lemmens, A., & Croux, C. (2006). Bagging and boosting classification trees to predict churn. *Journal of Marketing Research*, *43*(2), 276–286.
Lemmens, A., Croux, C., & Stremersch, S. (2012). Dynamics in the international market segmentation of new product growth. *International Journal of Research in Marketing*, *29*(1), 81–92.
Lenk, P. J., & Orme, B. (2009). The value of informative priors in Bayesian inference with sparse data. *Journal of Marketing Research*, *46*(6), 832–845.
Lenk, P. J., & Rao, A. G. (1990). New models from old: Forecasting product adoption by hierarchical Bayes procedures. *Marketing Science*, *9*(1), 42–53.
Li, K. C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, *86*(414), 316–327.
Li, Q., & Lin, N. (2010). The Bayesian Elastic Net. *Bayesian Analysis*, *5*(1), 151–170.
Naik, P. A., Hagerty, M. R., & Tsai, C. L. (2000). A new dimension reduction approach for data-rich marketing environments: Sliced inverse regression. *Journal of Marketing Research*, *37*(1), 88–101.
Naik, P. A., Wedel, M., Bacon, L., Bodapati, A., Bradlow, E., Kamakura, W., et al. (2008). Challenges and opportunities in high-dimensional choice data analyses. *Marketing Letters*, *19*(3–4), 201–213.
Naik, P. A., Wedel, M., & Kamakura, W. (2010). Multi-index binary response analysis of large data sets. *Journal of Business and Economics Statistics*, *28*(1), 67–81.
O'Hara, R. B., & Sillanpaa, M. J. (2009). A review of Bayesian variable selection models: What, how and which? *Bayesian Analysis*, *4*(1), 85–118.
Park, T., & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, *103*(482), 681–686.
Peres, R., Muller, E., & Mahajan, V. (2010). Innovation diffusion and new product growth models: A critical review and research directions. *International Journal of Research in Marketing*, *27*(2), 91–106.
Perez Rodriguez, P. (2009). *ars: Adaptive rejection sampling original C++ code from Arnost Komarek based on ars.f written by P. Wild and W. R. Gilks. R package version 0.4.* (http://CRAN.R-project.org/package=ars).
Putsis, W. P., Balasubramanian, S., Kaplan, E. H., & Sen, S. K. (1997). Mixing behavior in cross-country diffusion. *Marketing Science*, *16*(4), 354–369.
Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, *92*(437), 179–191.
Rutz, O. J., Trusov, M., & Bucklin, R. E. (2011). Modeling indirect effects of paid search advertising: Which keywords lead to more future visits? *Marketing Science*, *30*(4), 646–665.
Srinivasan, S., Pauwels, K., & Nijs, V. (2010). Demand-based pricing versus past-price dependence: A cost–benefit analysis. *Journal of Marketing*, *72*(2), 15–27.
Stamey, T. A., Kabalin, J. N., McNeal, J. E., Johnstone, I. M., Freiha, F., Redwine, E. A., et al. (1989). Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate: II. Radical prostatectomy treated patients. *Journal of Urology*, *141*(5), 1076–1083.
Stremersch, S., & Lemmens, A. (2009). Sales growth of new pharmaceuticals across the globe: The role of regulatory regimes. *Marketing Science*, *28*(4), 690–708.
Stremersch, S., & Tellis, G. J. (2004). Understanding and managing international growth of new products. *International Journal of Research in Marketing*, *21*(4), 421–438.
Takada, H., & Jain, D. (1991). Cross-national analysis of diffusion of consumer durable goods in Pacific Rim countries. *Journal of Marketing*, *55*(2), 48–54.
Talukdar, D., Sudhir, K., & Ainslie, A. (2002). Investigating new product diffusion across products and countries. *Marketing Science*, *21*(1), 97–114.
Tellis, G. J., Stremersch, S., & Yin, E. (2003). The international takeoff of new products: The role of economics, culture, and country innovativeness. *Marketing Science*, *22*(2), 188–208.
Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, *58*(1), 267–288.
Van den Bulte, C., & Stremersch, S. (2004). Social contagion and income heterogeneity in new product diffusion: A meta-analytic test. *Marketing Science*, *23*(4), 530–544.

van Everdingen, Y. M., Aghina, W. B., & Fok, D. (2005). Forecasting cross-population innovation diffusion: A Bayesian approach. *International Journal of Research in Marketing*, *22*(3), 293–308.
van Everdingen, Y. M., Fok, D., & Stremersch, S. (2009). Modeling global spillover of new product takeoff. *Journal of Marketing Research*, *46*(5), 637–652.
Wright, J. H. (2008). Bayesian model averaging and exchange rate forecasts. *Journal of Econometrics*, *146*(2), 329–341.
Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, *67*(2), 301–320.