Determinants of foreign direct investment

Bruce A. Blonigen *University of Oregon*
Jeremy Piger *University of Oregon*

Abstract. Empirical studies of bilateral foreign direct investment (FDI) activity show substantial differences in specifications with little agreement on the set of included covariates. We use Bayesian statistical techniques that allow one to select from a large set of candidates those variables most likely to be determinants of FDI activity. The variables with consistently high inclusion probabilities include traditional gravity variables, cultural distance factors, relative labour endowments and trade agreements. There is little support for multilateral trade openness, most host-country business costs, host-country infrastructure and host-country institutions. Our results suggest that many covariates found significant by previous studies are not robust.

Résumé. *Les déterminants de l'investissement direct à l'étranger*. Les études empiriques des déterminants des activités d'investissement direct bilatéral à l'étranger ont des spécifications substantiellement différentes et peu d'accord sur les variables co-reliées incluses. On utilise des techniques statistiques bayesiennes qui permettent de balayer un vaste ensemble de variables à la recherche de celles qui sont davantage susceptibles d'être des déterminants des activités d'investissement direct à l'étranger. Les variables qui se retrouvent de manière régulière dans la liste de haute probabilité d'impact sont les variables reliées à la gravité, les facteurs liés à la distance culturelle, les dotations relatives en facteur travail, et les accords commerciaux. Il y a peu de support pour des variables comme l'ouverture au commerce multilatéral, la plupart des coûts d'affaires, les infrastructures et les institutions dans les pays hôtes. Ces résultats suggèrent que plusieurs co-variations qu'on a jugées significatives dans les études antérieures ne sont pas robustes.

JEL classification: F21, F23, C52

1. Introduction

Empirical analyses of the factors determining foreign direct investment (FDI) across countries have employed a variety of econometric specifications. Many previous studies of cross-country FDI activity have used a gravity equation, which controls mainly for the economic size of the parent and host countries, the

We thank Theo Eicher, Gordon Hanson, Kathryn Russ, Deborah Swenson, anonymous referees and session participants at a Western Economic Association session for helpful comments. Blonigen is also a research associate of the National Bureau of Economic Research. Any errors or omissions are completely our own.
Corresponding author: J. Piger, jpiger@uoregon.edu

Canadian Journal of Economics/Revue canadienne d'économique, Vol. 47, No. 3
August 2014. Printed in Canada / Août 2014. Imprimé au Canada

0008-4085 / 14 / 775–812 /

geographic distance separating the countries and proxies for certain economic frictions. Like trade flows, this specification does a reasonably good job of fitting the observed data, but leaves one wondering if such a parsimonious specification captures all relevant factors.

Recent papers by Carr, Markusen and Maskus (2001; CMM) and Bergstrand and Egger (2007) have developed theoretical models of multinational enterprise's (MNE's) foreign investment decisions that suggest additional possible factors that determine FDI patterns. These studies point out a number of modifications to a standard gravity model that may be necessary to accurately explain FDI patterns. First, while gravity variables may adequately capture "horizontal" motivations for FDI, where firms look to replicate their operations in other countries to be more proximate to consumers in those markets, additional controls are necessary to allow for "vertical" motivations of FDI, where firms look for low-cost locations for labour-intensive production. For example, these studies introduce measures of relative labour endowments in the host country with the expectation that countries with relatively high shares of unskilled labour will be attractive locations for MNEs due to lower wages. In addition, these studies show that FDI decisions by MNEs are complex enough that interactions between key variables (e.g., GDP and skilled labour endowments) may be necessary to account for nonlinear effects of these variables on FDI patterns. Head and Ries (2008) differs from these previous studies by modelling FDI as arising from decisions by firms to acquire and control foreign assets (i.e., cross-border mergers and acquisitions), rather than development of new (or greenfield) plants. Their analysis of FDI patterns highlights the potential role of common culture and language between countries.

While these prior studies have been important in deepening our understanding of the factors that determine cross-country FDI patterns, they have generally focused on regression models involving specific sets of covariates determined by the researcher and the particular theoretical framework for FDI they chose to examine. By conditioning on a particular regression model specification, this practice ignores uncertainty regarding the model specification itself, which can have dramatic consequences on inference.[1] Most notably, inference regarding the effects of included covariates can depend critically on what other covariates are included versus excluded.

In this paper, we take a Bayesian approach to confront uncertainty regarding the appropriate set of covariates to include in a regression model explaining FDI activity. From a Bayesian perspective, incorporating such uncertainty is conceptually straightforward. The choice of covariates, or "model," is treated as an additional parameter that lies in the space of potential models, which allows us to compute the posterior probability that each potential model is the true model that generated the data. Posterior distributions for objects of interest, such as

1 For discussion and examples, see Leamer 1978, Hodges 1987, Moulton 1991, Draper 1995, Kass and Raftery 1995, Raftery 1996 and Fernandez, Ley and Steel 2001a.

the effect of a particular covariate, are then averaged across alternative models, using the posterior model probabilities as weights. This procedure, known as Bayesian Model Averaging (BMA), produces inferences that are not conditioned on a particular model.

To be clear, we are taking a purely empirical approach to determine the correlates with observed FDI patterns. As we discuss in the next section, there is very little consistency in the empirical FDI literature about the covariates one should use when empirically modeling cross-country FDI. We view this paper as a first step in pointing out these inconsistencies and providing evidence of the empirically robust determinants of FDI.

Although conceptually straightforward, BMA is practically difficult when the set of possible models is large, as direct calculation of posterior probabilities for all models becomes infeasible. In our application, we have a large set of potential covariates, which yields an extremely large set of possible models ($> 7 \times 10^{16}$). To sidestep this difficulty, we use techniques designed to obtain random draws of models from the probability distribution defined by the posterior model probabilities. Such draws are made possible even when the posterior model probabilities are unknown by using the MC^3 algorithm of Madigan and York (1995). These random model draws are then used to construct estimates of the posterior model probabilities.[2]

Our set of potential FDI determinants is meant to be comprehensive and includes a combination of covariates proposed by the previously mentioned studies, as well as other prior literature on FDI. We examine mainly cross-sectional patterns for the year 2000.[2] We examine both levels and log-linear regressions, placing more weight on our results for the log-linear regressions because most previous studies have used a logarithmic transformation to address skewness in the FDI variable. We also examine three measures of FDI—FDI stock, affiliate sales and cross-border mergers and acquisitions activity—in order to better compare with a broader set of prior studies. At the end, we also explore a specification that first differences observations across the years 1990 and 2000 to control for bilateral country-pair fixed effects as well as a negative binomial specification to better model the nature of our dependent variable.

Our analysis indicates that many of the covariates used in prior FDI studies (and often found statistically significant) do not have a high probability of inclusion in the true FDI determinants model once we consider a comprehensive set of potential determinants using BMA. A fairly parsimonious set of covariates is suggested by our analysis. The covariates with consistently high inclusion probabilities include traditional gravity variables, cultural distance factors, relative labour endowments and trade agreements. Variables with little support for inclusion are multilateral trade openness, most host-country business costs, host-country infrastructure (including credit markets) and host-country institutions.

2 Focusing on the year 2000 maximized our available sample size by allowing us to use datasets that have not been updated recently along with datasets that began being collected in 2000.

A few variables that have rarely been included in prior FDI studies, namely host-country remoteness, parent-country real GDP per capita and host is an oil-exporting country, have surprisingly high inclusion probabilities.

The remainder of the paper proceeds as follows. The next section reviews previous empirical literature on the determinants of FDI and makes the case that the appropriate model specification for explaining FDI patterns is far from settled. Section 3 then lays out the BMA methodology we use to assess model uncertainty. Section 4 describes the data and its sources, while section 5 reports the results and compares to the existing literature. Section 6 concludes.

2. Prior FDI literature

There is little consensus on how to empirically model bilateral FDI patterns, with many past empirical FDI papers using a base model consisting of gravity-type covariates (country-level GDP and distance) because of its popularity for explaining trade flows. As mentioned in the introduction, there have been a few recent efforts to develop specifications based on theoretical models—namely the knowledge-capital (K-K) model developed by James Markusen and co-authors, which was brought to data in CMM (2001), Bergstrand and Egger's (2007) model incorporating physical capital and Head and Ries' (2008) model of acquisition FDI.

There is little consistency in the covariates that are postulated to explain world-wide FDI patterns across these three papers. To see this, the first three columns of table 1 lists the covariates used in each of these papers. Distance between countries is the only covariate common to all three studies. There are 22 different covariates between the three studies, even though each study averages only about 10 covariates. While all three specifications postulate a role for economic size and trade frictions as driving forces of FDI, it is surprising how differently they construct and define variables meant to proxy for these common factors.

Of course, there have been many other papers that have empirically examined FDI patterns using specifications that differ from these three papers. Columns 4 through 8 of table 1 list the covariates used in a number of other highly regarded recent papers. Across these eight studies in columns 1 through 8, there are a combined 47 covariates. However, no covariate is shared by all eight studies and, on average, a covariate is used in only 1.7 of the eight studies. Interestingly, almost 85% of the covariates included in these 8 studies are found to be statistically significant. Given that the average study includes very few of the total set of possible covariates, the possibility of spurious correlations is quite real.

In addition to the substantial differences in covariates used across FDI studies, there are also differences across studies in whether variables are logged or whether panel data were used (these are noted in the first few rows of table 1). Given these wide differences in specifications, there clearly is no consensus on how to specify the determinants of bilateral FDI patterns.

TABLE 1
Specifications of prior studies of FDI determinants

	Carr, Markusen, Maskus (2001)	Bergstrand and Egger (2007)	Head and Ries (2008)	Eaton and Tamura (1994)	Wei (2000)	di Giovanni (2005)	Stein and Daude (2007)	Chakrabarti (2001)
Data and specifications								
Dependent variable	Sales	Sales	Stock and M&A	Stock	Stock	M&A	Stock	Flows
Variables logged	No	Yes	Yes	Yes	Yes	Yes	Some	No
Panel data	Yes	Yes	No	Yes	Yes	Yes	No	No
Two-way or one-way flows	Two-way	Two-way	Two-way	Two-way	One-way	Two-way	Two-way	One-way
Gravity measures								
PARENT GDP		x				x		
HOST GDP		z			x	x		x
Distance	z	x	x		x	x	x	
Other GDP-related terms								
PARENT per capita GDP			x					
HOST per capita GDP			x	x				x
PARENT population			x					
HOST population			x	x	x			
GDP similarity		x						
GDP sum	x	x					x	
GDP difference	x							
GDP per capita differences						x	x	
HOST GDP growth								z
Rest-of-the-world GDP		x						
Other geography measures								
Contiguous border							z	
Time zone differences							x	
Country-level endowments								
Relative skilled-unskilled labour endowments (skill difference)	x	x					x	
Interaction of skill differences and GDP differences	x							

(*Continued*)

TABLE 1
(*Continued*)

	Carr, Markusen, Maskus (2001)	Bergstrand and Egger (2007)	Head and Ries (2008)	Eaton and Tamura (1994)	Wei (2000)	di Giovanni (2005)	Stein and Daude (2007)	Chakrabarti (2001)
Relative capital-labour endowments		x						
HOST wages					x			z
HOST population density				x				
HOST education levels				x				
Bilateral cultural and colonial linkages								
Common language		x	x		x	x	z	
Colonial links			x				x	
Multilateral trade openness								
HOST trade costs	x							z
PARENT trade costs	x							
HOST trade openness (imports plus exports divided by GDP)								z
HOST trade costs times skill difference term squared	x							
Bilateral trade openness								
BILATERAL transport costs		z						
BILATERAL trade flows/deficit						x		z
Regional trade agreement		x				z	z	
Customs union						z		
Common service sector agreement						x		
Host country FDI/business costs								
HOST FDI costs	x	x						
HOST taxes					x			z

(*Continued*)

TABLE 1
(*Continued*)

	Carr, Markusen, Maskus (2001)	Bergstrand and Egger (2007)	Head and Ries (2008)	Eaton and Tamura (1994)	Wei (2000)	di Giovanni (2005)	Stein and Daude (2007)	Chakrabarti (2001)
PARENT taxes						x		
PARENT country has tax credit system					z			
Change in HOST consumer prices								z
Bilateral tax and investment agreements								
Tax treaty						x	x	
Investment treaty							z	
Host country communications infrastructure								
Telephone traffic						x	x	
Host country financial infrastructure								
HOST market capitalization						x		
HOST domestic credit						x		
Political environment and institutions								
HOST political stability					x			z
HOST legal institutions							x	
HOST corruption					x			
Exchange rates								
Exchange rates						z		z
Volatility of exchange rates						x		

NOTES: An “x” signifies that a variable is included and statistically significant in the majority of specifications reported in the paper. A “z” signifies that a variable is included, but it is not statistically significant in the majority of specifications reported in the paper. We exclude from this table variables that Chakrabarti (2001) posited as *ex ante* doubtful and that did not come in statistically significant in that analysis. The type of dependent variable in these studies varied in construction but can be characterized by data on affiliate sales, which we term *Sales* in the table; FDI stock, *Stock*; FDI flows, *Flows*; and/or counts (or value) of cross-border mergers and acquisitions activity, *M&A*.

The final paper documented in table 1 (last column) is Chakrabarti (2001). This paper is similar to ours in its motivation to understand which covariates are more likely to be robust determinants of bilateral FDI. However, the analysis considers a surprisingly small set of possible covariates, perhaps because it came before some of the recent advances in the literature. Also, it follows a different methodology (extreme bounds analysis) from ours, feasible implementation of which requires the model space be restricted *a priori*. The approach we take to implement BMA requires no such restriction and is designed to identify and explore relevant portions of the entire model space. That said, Chakrabarti (2001) serves as a potential warning signal for the literature and motivation for further study, as it finds that most of the covariates investigated are not statistically robust using typical extreme bounds criteria.

On a final note, Eicher, Helfman and Lenkoski (2010; EHL) and Jordan and Lenkoski (2012) are recent works that are similar to ours in their use of BMA to evaluate an extensive set of potential FDI determinants (including many of those included in table 1). However, there are a number of major differences. Both of these prior papers focus on determinants of FDI flows, whereas our focus is on the (static) cross-country distribution of FDI, typically measured by FDI stock or affiliate sales. This is an important distinction. Examination of FDI flows has been the purview of primarily the international finance literature, where the role of exchange rates, capital market shocks and short-run changes to other financial variables are the focus. In contrast, we wish to inform the empirical FDI literature that has focused on stock measures of FDI in order to directly assess the main general equilibrium theories of the long-run factors that explain the distribution FDI across countries. General equilibrium predictions are static in nature and therefore pertain to levels, not (short-run) changes, of the variables of interest. An additional focus of these papers is on modeling the selection issue of whether there is any FDI activity between bilateral country pairs in the first place. Since almost all prior empirical FDI studies do not address this issue, and our primary focus is on comparing our BMA results directly with these prior studies, we do not explore this issue either.

3. Methodology

3.1. FDI determinants model and Bayesian model averaging

To study the determinants of bilateral foreign direct investment (FDI), we focus on the linear regression model:

$$Y = \alpha \iota_N + X_j \beta_j + \varepsilon, \tag{1}$$

where Y is an N x 1 vector holding the measure of bilateral foreign direct investment, ι_N is an N x 1 vector of 1's, X_j is an N x k_j matrix of FDI determinants and ε is an N x 1 vector of independent, normally distributed disturbances, each with mean zero and variance σ^2. We are interested in the realistic case where there is uncertainty about the appropriate variables to include in X_j. In particular, suppose there are K potential determinants of FDI, collected in the N x K matrix X, and the variables in X_j are chosen as a subset of X, so that $k_j \leq K$. We assume that the only aspect of model uncertainty in (1) is the selection of X_j, so that a particular selection of X_j defines the j^{th} model, denoted M_j. If we place no restrictions on the combinations of the variables in X that can enter the regression model, there are $R = 2^K$ different models to consider.

The Bayesian approach to comparing alternative models is based on the posterior probability that M_j is the true model that generated the data:

$$Pr\left(M_j|\,Y\right) = \frac{f\left(Y|M_j\right)Pr\left(M_j\right)}{\sum_{i=1}^{R} f\left(Y|M_i\right)Pr\left(M_i\right)},\ j = 1, ..., R, \tag{2}$$

where (2) follows directly from application of Bayes' rule. In (2), $Pr\left(M_j\right)$ is the researcher's prior probability that M_j is the true model, while $f\left(Y|M_j\right)$ is the marginal likelihood:

$$f\left(Y|M_j\right) = \int f\left(Y|\alpha, \beta_j, \sigma, M_j\right) p\left(\alpha, \beta_j, \sigma|M_j\right) d\alpha\, d\beta_j\, d\sigma, \tag{3}$$

where $f\left(Y|\alpha, \beta_j, \sigma, M_j\right)$ is the likelihood function for model M_j and $p\left(\alpha, \beta_j, \sigma|M_j\right)$ is the researcher's prior density function for the parameters of M_j. In words, the marginal likelihood function is the likelihood function integrated with respect to the researcher's prior density function. It thus has the interpretation of the average value of the likelihood function, and therefore the average fit of the model, over different parameter values, where the averaging is done with respect to the prior density of model parameters.

The posterior model probabilities in (2) can be used to confront the model uncertainty present in the FDI determinants regression. One approach for using $\Pr\left(M_j|\,Y\right)$ is to select the model with highest posterior probability and then make inferences about the effects of alternative FDI determinants based on this "best" model alone. However, this focus on one chosen model (which mimics much of the model selection literature based on hypothesis tests and information criteria) ignores information in models other than the chosen model and thus does not yield inferences that fully incorporate model uncertainty. When the posterior model probability is dispersed widely across a large number of models, basing inferences on a single model can yield grossly distorted results.

Instead of basing inference on a single highest probability model, BMA proceeds by averaging posterior inference regarding objects of interest across alternative models, where averaging is with respect to posterior model probabilities.

Specifically, for a generic object of interest λ, the BMA posterior distribution is calculated as:

$$p(\lambda|Y) = \sum_{j=1}^{R} p(\lambda|Y, M_j)\, Pr(M_j|Y), \tag{4}$$

where $p(\lambda|Y, M_j)$ is the posterior distribution for λ conditional on model M_j. For common choices of λ, this conditional posterior distribution will often be available analytically. We discuss several such cases in section 3.4 below. The BMA posterior distribution in (4) follows from direct application of rules of probability and is thus the obvious solution to incorporate model uncertainty into inference from the Bayesian perspective.[3] It is worth emphasizing that $p(\lambda|Y)$ is not conditioned on a particular model being the true model, but is instead conditioned only on the data. That is, BMA has integrated out uncertainty regarding the identity of the true model.[4]

3.2. Priors

To implement BMA, we require posterior model probabilities. From (2) and (3), calculation of these probabilities requires a choice for both the prior density function for the parameters of M_j, $p(\alpha, \beta_j, \sigma|M_j)$ and the prior model probability, $Pr(M_j), j = 1, ..., R$. In this section, we describe how each of these priors are set in our study of FDI determinants.

In BMA applications, specification of the prior parameter densities poses a significant challenge. One approach is to elicit prior densities for the parameters of each model individually. However, this becomes intractable when the space of potential models is large, as will be true for the FDI determinants model. In such cases, it is useful to use prior parameter densities that are “automatic,” in that they are set in a formulaic way across alternative models. One simple, and seemingly attractive, way to do this is to use non-informative priors for the parameters of all models under consideration. Unfortunately, the use of non-informative priors for those parameters not common to all models will yield posterior model probabilities that mechanically favour models with fewer parameters over those with more. For our application, the slope parameters β_j are not common to all models, as they depend on the set of variables included in X_j. Thus, using non-informative priors for β_j is not an option, as it will paradoxically generate model comparison results that are solely a consequence of the prior. This is not the case for parameters that are common to all models, for which non-informative priors

3 For an introduction to BMA and a review of related literature, see Hoeting, Madigan, Raftery and Volinsky (1999).

4 In addition to BMA, we have produced results using the weighted average least squares (WALS) procedure of Magnus, Powell and Prüfer (2010), which is an alternative approach to model averaging from BMA. The results using WALS were quite similar to those from BMA and are available upon request.

yield posterior model probabilities that are not a function of the prior but only of sample information. For this reason, non-informative priors are a popular choice for parameters common to all models.

Here we use two different automatic procedures for setting priors. For our primary analysis, we use the priors suggested by Fernandez, Ley and Steel (2001a), hereafter FLS, who provide an automatic procedure for setting parameter prior densities for a group of linear regression models that differ only with respect to the choice of covariates. This procedure is designed for the case where the researcher wishes to use as little subjective information in setting prior densities as possible and was shown by FLS to both have good theoretical properties and perform well in simulations for the calculation of posterior model probabilities. As a robustness check, we also present results for a prior advocated by Eicher, Papageorgiou and Raftery (2011; EPR). We will describe the FLS prior in detail here, while the alternative prior is discussed in section 5.5.

The FLS procedure begins by factoring the prior parameter density function as follows:

$$p\left(\alpha, \beta_j, \sigma | M_j\right) = p\left(\beta_j | \alpha, \sigma, M_j\right) p\left(\alpha, \sigma | M_j\right). \tag{5}$$

For parameters common to all models, namely α and σ, FLS use the standard, improper non-informative prior density for location and scale parameters:[5]

$$p\left(\alpha, \sigma | M_j\right) \propto \sigma^{-1}. \tag{6}$$

To set $p\left(\beta_j | \alpha, \sigma, M_j\right)$, FLS use the natural conjugate Normal-Gamma prior density:

$$\beta_j | \sigma, M_j \sim N\left(\beta_j^0, \sigma V_j^0\right). \tag{7}$$

This natural conjugate form is advantageous as it allows for analytical calculation of the integrals in (3), which greatly speeds computing time. We set the prior mean, β_j^0, to a k_j x 1 vector of zeros. This centres the prior distribution for all model slope parameters on values consistent with the FDI determinants in X_j having no effect on FDI. To set the prior variance-covariance matrix, FLS suggest the g-prior specification of Zellner (1986):

$$V_j^0 = \left(g X_j' X_j\right)^{-1}. \tag{8}$$

5 This prior specification is independent of the model and thus assigns a common prior density for the intercept and conditional variance parameters across models. To ensure that the model intercept has the same interpretation across all models, we demean the FDI determinant variables before inclusion in the regressions. This gives the intercept parameter the role of the unconditional mean of the bilateral FDI measure for all models.

This prior specification is useful as it reduces the input from the researcher to a single hyperparameter, g, rather than needing to specify the entire k_j x k_j matrix V_j^0. FLS discuss theoretical motivations for alternative choices of g and, based on this theory and extensive Monte Carlo experiments, suggest the following rule:

$$g = \left\{ \begin{array}{ll} \frac{1}{K^2} & \text{if} \quad N \leq K^2 \\ \frac{1}{N} & \text{if} \quad N > K^2 \end{array} \right\}. \tag{9}$$

In our study of FDI determinants, we consider several possible measures of Y with corresponding varying values for N. For all of these variations on the dependent variable, we have either $N < K^2$ or $N \approx K^2$, and thus $g \approx 1/K^2$ in our analysis.

To specify the prior model probability, we begin by defining an indicator variable, τ_i, which is one if the i^{th} variable is included in the true model and is zero otherwise. Our prior assumption is that each potential regressor enters the true model independently of all others with prior probability θ, so that $\Pr(\tau_i = 1) = \theta, \forall\, i$. This implies prior model probabilities of the form:

$$\Pr\left(M_j\right) = \theta^{k_j}(1-\theta)^{K-k_j}.$$

A popular choice in the BMA literature is to set $\theta = 0.5$, which implies equal prior probability across all possible models:[6]

$$\Pr\left(M_j\right) = \frac{1}{2^K} = \frac{1}{R}. \tag{10}$$

Because it is uniform across individual models, the model prior in (10) implies a lack of prior information about which specific model is the true model. However, this prior does not imply a uniform prior for the model size, defined as the number of covariates included in the true model.[7] Indeed, as shown in Ley and Steel (2009), the prior probability distribution over model size implied by (10) will be binomial:

$$\Pr\left(\sum_{i=1}^{K} \tau_i\right) = Bin(K, \theta).$$

This binomial distribution will peak near $K/2$ and, for moderate to large K, place very low probability on models with either only a few or a very large number

6 See, for example, Raftery, Madigan and Hoeting (1997) and Fernandez, Ley and Steel (2001a, 2001b).

7 This is because the number of models for alternative model sizes can be different. For example, there is a single model with no covariates, but K models with one covariate.

of potential covariates. For example, in our study of FDI determinants, this prior would peak near 26 and place cumulative prior probability of less than 0.001 on all model sizes below 16 or above 40.

Here, we instead use a prior suggested in Ley and Steel (2009). Rather than fix θ as a prior hyperparameter, we treat this prior inclusion probability as a random variable that follows a $Beta(a, b)$ distribution, where a and b are hyperparameters of the prior. This is an example of a hierarchical prior, which Ley and Steel (2009) argue increases the flexibility of the prior and reduces the dependence of posterior model probabilities on prior assumptions. In this particular case, the hierarchical prior implies a beta-binomial prior distribution for model size, where a and b can be set to accommodate a wide variety of prior beliefs regarding model size. Ley and Steel (2009) recommend setting $a = 1$ and setting b to match a prior mean for model size, denoted m. Here we set b so that $m = K/2$, which generates a uniform prior for model size:

$$\Pr\left(\sum_{i=1}^{K} \tau_i\right) = \frac{1}{K+1}.$$

Thus, our prior over models will be agnostic regarding the number of covariates that are in the true model.[8]

3.3. Calculating posterior model probabilities

Given these specifications for the prior densities, posterior model probabilities are conceptually straightforward to calculate. In particular, model probabilities can be computed directly by calculating the marginal likelihood for all possible models, each is available analytically for the linear regression model in (1) and the parameter prior densities in (6) to (9). However, when K is large, the size of the model space makes direct calculation of $Pr\left(M_j|Y\right)$ based on (2) practically infeasible. For example, we will consider $K = 56$ potential FDI determinants, meaning there are greater than $R = 7 \times 10^{16}$ possible models to consider. Even if each model could be considered in $1/100{,}000^{\text{th}}$ of a second, an ambitious estimate at current computing speeds, it would still take over 22,000 years to evaluate all possible models.

When the model space becomes too large for direct calculation of posterior model probabilities, a popular alternative approach is to estimate these probabilities by sampling the model space. In particular, define a model indicator that takes on values from $1, \ldots, R,$ with a value of j indicating that model M_j is the true model and assume that this model indicator follows a multinomial probability distribution with probabilities given by $\Pr\left(M_j|Y\right)$. Further, suppose that

8 We also considered a prior in which $m = K / 10$; this places substantially more prior weight on smaller models than the prior with $m = K / 2$. We do not report the results for this prior as they were nearly identical to the $m = K / 2$ case.

we are able to obtain random draws of this model indicator from its probability distribution. It is then possible to construct a simulation-consistent estimate of $\Pr\left(M_j | Y\right)$ as the proportion of the random draws for which model M_j was drawn. In particular, we can construct the following estimate of $\Pr\left(M_j | Y\right)$:

$$p_j = \frac{\sum_{s=1}^{S} I_s}{S}, \tag{11}$$

where S is the number of random draws of the model indicator and I_s is an indicator function that is one if the s^{th} draw of the model indicator was j. Note that (11) will estimate $\Pr\left(M_j | Y\right)$ to be zero if M_j is never drawn. However, assuming a large number of simulations are conducted, it will be exactly these models that are likely to have very low posterior model probability. Thus, estimates of $\Pr\left(M_j | Y\right)$ constructed by simulating from the model space provide an efficient approach to identifying the set of models with relatively high posterior probability.

Note that if we condition on $\Pr\left(M_j | Y\right)$ equalling zero if M_j is never drawn, equation (2) suggests an alternative, approximation-free approach to evaluating the posterior model probabilities for the visited models:

$$p_j = \frac{f\left(Y|M_j\right)\Pr\left(M_j\right)}{\sum_{i \in \Delta} f\left(Y|M_i\right)\Pr\left(M_i\right)}, \ j \in \Delta, \tag{12}$$

where Δ denotes the set of models that are visited by the sampler. As this set of models will be feasible to consider individually, the summation in the denominator of (12) will be feasible, whereas the summation in the denominator of (2) was not. If the models never visited by the sampler are assumed to have zero probability, model probabilities based on (12) will be exact, while those based on (11) will contain estimation error. All results presented for our FDI determinants analysis use model probabilities based on (12).

To simulate from the model space, we use the Markov Chain Monte Carlo Model Composition (MC^3) algorithm of Madigan and York (1995). This approach relies on the Metropolis-Hastings algorithm, which can be used to provide random samples from any probability distribution provided it is known up to a proportionality constant, which, by inspection of (2), is true for $\Pr\left(M_j | Y\right)$. MC^3 was implemented by Raftery, Madigan and Hoeting (1997) for BMA in linear regression models and has been used in a number of economic applications involving linear regression (e.g., Fernandez, Ley and Steel 2001a, 2001b).[9]

The MC^3 algorithm requires an arbitrary model to initialize the sequence of model draws. Given this initial model, model draws obtained from the algorithm

9 For details of the implementation of MC^3 in the context of a linear regression model, see Koop (2003).

form a Markov chain that converges to draws from $\Pr\left(M_j|\,Y\right)$. An important issue with such Markov-chain based samplers is assessing the convergence of the chain. In producing the results described in section 5 below, we assume that 200,000 draws is sufficient to ensure convergence and then base our estimates of posterior model probabilities on 1 million additional draws. We performed three checks to ensure convergence of the sampling procedure. First, results from an independent simulation using a longer convergence sample of 400,000 draws were very similar to those based on the shorter convergence sample. Second, our results are insensitive to two widely dispersed initial models: one with no FDI determinants and one with all possible FDI determinants. This insensitivity of results to the size of the convergence sample and the initialization of the chain suggests the sampler has converged. Finally, FLS suggest using the correlation between the probability estimates based on (11) and (12) as a check on the convergence of the sampler. For all results we present, this correlation was above 0.99.

3.4. Calculating BMA posterior distributions

In this section, we describe calculation of the BMA posterior distributions for the various objects of interest, λ, that we will use in our analysis of FDI determinants. The primary BMA posterior distribution we construct is the so-called "posterior inclusion probability," which is the BMA posterior probability that a covariate belongs to in the true model. In this case, $\lambda = \tau_i$, and the model dependent posterior distribution, $p\left(\tau_i|\,Y, M_j\right)$, is simply an indicator variable that is one if the i^{th} variable is included in model M_j and is zero otherwise. From equation (4), the posterior inclusion probability is then:

$$p\left(\tau_i|\,Y\right) = \sum_{j=1}^{R} p\left(\tau_i|\,Y, M_j\right) \Pr\left(M_j|\,Y\right) = \sum_{j\in\omega} \Pr\left(M_j|\,Y\right), \tag{13}$$

where ω denotes the set of models that include the i^{th} covariate.

We are also interested in the BMA posterior distribution for the marginal effect of the i^{th} potential covariate. Denote the K x 1 vector of marginal effects for the K potential covariates as β. We then wish to construct the BMA posterior distribution:

$$p\left(\beta|\,Y\right) = \sum_{j=1}^{R} p\left(\beta|\,Y, M_j\right) \Pr\left(M_j|\,Y\right).$$

Define a K x k_j selection matrix, T_j, such that $\beta = T_j\beta_j$ is the K x 1 vector of marginal effects for model M_j. Here, the i^{th} element of β is the appropriate slope parameter from β_j if model M_j includes the i^{th} covariate and is zero otherwise. As discussed in Magnus, Powell and Prüfer (2010), the BMA posterior distribution for β then has the following moments:

$$E(\beta|Y) = \sum_{j=1}^{R} T_j E\left(\beta_j|Y, M_j\right) Pr\left(M_j|Y\right), \tag{14}$$

$$\begin{aligned} Var(\beta|Y) &= -E(\beta|Y)\,E(\beta|Y)' \\ &\quad + \sum_{j=1}^{R} Pr\left(M_j|Y\right) T_j \left(Var\left(\beta_j|Y, M_j\right) + E\left(\beta_j|Y, M_j\right) E\left(\beta_j|Y, M_j\right)' \right) T_j', \end{aligned} \tag{15}$$

where $E\left(\beta_j|Y, M_j\right)$ and $Var\left(\beta_j|Y, M_j\right)$ are the moments of the posterior distribution for β_j conditional on M_j. Given the linear regression model and natural conjugate parameter priors presented above, these moments of the conditional posterior distribution are given by:

$$E\left(\beta_j|Y, M_j\right) = \frac{1}{1+g}\left(X_j' P X_j\right)^{-1} X_j' P Y,$$

$$Var\left(\beta_j|Y, M_j\right) = \frac{Y' P A_j P Y}{(1+g)(N-3)}\left(X_j' P X_j\right)^{-1},$$

where $P = I_N - \frac{1}{N}\iota_N \iota_N'$ and $A_j = \frac{g}{1+g}P + \frac{1}{1+g}\left(P - PX_j\left(X_j'PX_j\right)^{-1} X_j' P\right)$.

4. Data

Measurement of FDI and related activity is far from ideal. Unlike for trade flows, reliable measures of FDI are unavailable for many countries. In addition, there is no common source for FDI data; prior studies have therefore employed a number of different measures of FDI. As we wish to compare our results to these prior studies, we have collected data on three different FDI measures typically used.

Our first source of cross-country FDI activity is bilateral FDI stocks reported by members of the Organization of Economic Cooperation and Development (OECD), which is the most comprehensive source of reliable data on total FDI stocks that we are aware of.[10] OECD provides excellent coverage of FDI activity between OECD countries. It also has some coverage of FDI between OECD and non-OECD countries, though many transactions with small non-OECD countries are missing. OECD does not report any observations of FDI between

10 These data can be obtained from SourceOECD, www.sourceoecd.org.

countries where they are both non-OECD. The FDI stock data will be the benchmark measure of FDI used in our study, but we will also compare and contrast our results when using two alternative measures of FDI activity, described next.

Some studies (e.g., CMM 2001; Bergstrand and Egger 2007) have stressed the use of affiliate sales as the most appropriate measure of actual multinational firm activity in a host country, as FDI stock data can be significantly affected by financial transactions of a firm not related to current productive activity. Unfortunately, affiliate sales data are much less available than FDI stock data. To our knowledge, Braconier, Norback and Urban (2005; BNU) have collected the most extensive database of cross-country affiliate sales and have graciously provided it to us. Their database provides information on outward affiliate sales involving 56 parent countries and 85 host countries over roughly four years from the late 1980s to 1998. Despite this, the number of observations is much smaller than with the FDI stock data.[11]

Finally, we employ data on cross-border mergers and acquisitions (M&A) that have been used in such studies as Rossi and Volpin (2004) and Head and Ries (2008). These data come from Thomsen's SDC Platinum database on M&A activity, meant to be a comprehensive census of worldwide M&A above the $1 million threshold since the early 1990s. While this level of country coverage in the M&A data clearly dominates the other two measures of FDI activity, the M&A measure also has relative disadvantages. First, it measures only one type of FDI, though M&A does account for the majority of worldwide FDI activity. Second, because many of the transactions are between private firms, over half of the M&A in the database do not have any recorded value. Thus, we rely on counts of the number of M&A occurring between country pairs.[12] More specifically, we use cumulated sums of counts of prior and current-year M&A by country pair to create a measure analogous to cross-country FDI stocks. Head and Ries (2008) also use cumulated measures of M&A activity and find a quite high correlation (greater than 0.80) between the FDI stock and M&A measures of FDI activity.

It is important to note that virtually all theory and empirics of worldwide FDI has focused on the (static) cross-country patterns rather than the dynamics of worldwide FDI flows. We follow this pattern and primarily focus on the year 2000, since it comes before the world recession following the events of 9/11 and most

11 We refer the reader to Braconier, Norback and Urban (2005) for further details on country coverage and data sources.

12 Prior studies, including Rossi and Volpin (2004) and Head and Ries (2008), assumed that the missing M&A transactions' values were random and summed up remaining observations of values to create their measure of cross-border M&A activity. There are some obvious advantages and disadvantages with using M&A count versus (non-missing) value data. One clear disadvantage for our purposes was how many missing observations are created when using the value data—many of the bilateral country pairings show M&A activity, but the value data for all the M&A transactions for that pairing are missing. For this reason, and because the correlation between the M&A counts and values by bilateral-country pairs is 0.96, we use the M&A count data.

closely matches the most recent data we have for the affiliate sales database.[13] For those FDI measures where it was available, we also collected data for 1990. This allows us to examine specifications where we first difference the data to control for country-pair fixed effects.

The set of potential covariates we consider is intended to be comprehensive and is listed in table 2. The variables in table 2 are grouped into broad categories of factors that plausibly determine FDI. We have included all covariates from previous studies listed in table 1 with only a few exceptions. First, we do not include exchange rate variables or changes in recent consumer prices, as we wish to examine the long-run determinants of FDI decisions, leaving examination of dynamic, short-run changes for other work. Second, bilateral trade flows are clearly endogenous and so we do not include this covariate as some studies have done. Finally, there are a few variables where available data are so limited (e.g., wage data) that we feel the cost in terms of reduced sample size is too great.

We also include a number of additional variables. First, a few recent studies have found that geographic spatial issues are important for understanding bilateral FDI patterns (see Baltagi, Egger and Pfaffermayr 2007, BBEP; Blonigen, Davies, Waddell and Naughton 2007, BDWN). To account for such spatial features of the data to some extent, we include a remoteness variable for both the host and parent country, constructed as the distance-weighted average of all other countries' GDP. Possible agglomeration effects within countries also led us to add a measure of urban concentration for both the host and parent country. Previous studies have hypothesized that endowments may matter, particularly if FDI is motivated to find lower cost locations (i.e., vertically motivated FDI). However, these studies have included only measures of relative labour and capital endowments. We include measures of land and oil as well. Business costs in the host country have been included in some previous studies, but they often use proxies that have limited country coverage, which we found significantly reduces the potential sample. Thus, we rely on relatively recent measures of host-country business costs collected by the World Bank that measure the average time it takes to enforce a contract, register property, start a business and resolve an insolvency. We also include measures from the World Bank's World Development Indicators on communications infrastructure, which previous studies have not included but plausibly could affect FDI decisions.

These additions and subtractions from the combined set of regressors from previous studies leave us with 56 variables to examine as potential covariates with FDI. The data sources for our variables are primarily the Penn World Tables, the World Development Indicators database and the Gravity database at CEPII (www.cepii.org). A full list of data sources is available from the authors upon request.

13 The most recent data we have available for the affiliate sales database is 1998. Our analysis will use FDI stock and M&A count data for the year 2000 and affiliate sales data for the year 1998.

TABLE 2
Variables

Variable	Definition	Included in previous study listed in table 1
Dependent variables		
FDI stock	FDI position of PARENT country in HOST country (in millions of U.S. dollars)	
Affiliate sales	Sales of PARENT-owned affiliates in HOST country	
M&A counts	Cumulated counts of PARENT country acquisitions of HOST country targets prior to year of observation	
Gravity measures		
1. PARENT real GDP	Real GDP of PARENT country (in trillions)	X
2. HOST real GDP	Real GDP of HOST country (in trillions)	X
3. Distance	Distance between the two most populous cities in the PARENT and HOST countries	X
Other GDP-related terms		
4. PARENT real GDP per capita	Real GDP per capita of PARENT country (constant price: Chain Series)	X
5. HOST real GDP per capita	Real GDP per capita of HOST country (constant price: Chain Series)	X
6. Sum of HOST and PARENT real GDP	Sum of HOST and PARENT real GDP	X
7. Similarity of HOST and PARENT real GDP	Share of HOST real GDP in the sum of HOST and PARENT GDP x Share of PARENT real GDP in the sum of HOST and PARENT GDP	X
8. Squared GDP difference	Squared real GDP difference between HOST and PARENT country	X
9. Squared GDP per capita difference	Squared real GDP per capita difference between HOST and PARENT countries	X
10. HOST urban concentration	Urban population (% of total) in HOST country	
11. PARENT urban concentration	Urban population (% of total) in PARENT country	
Geography measures other than distance		
12. Contiguous border	Dummy variable indicating PARENT and HOST countries are geographically contiguous	X
13. HOST remoteness	Distance of HOST country from all other countries in the world weighted by those other countries' share of world GDP (does not include host country in calculations)	
14. PARENT remoteness	Distance of PARENT country from all other countries in the world weighted by those other countries' share of world GDP (does not include host country in calculations)	
15. Time zone difference	Time zone difference between capital cities of HOST and PARENT countries	X
Relative labour endowments		
16. HOST education level	Average education years in HOST country	X
17. HOST skill level	Percent of employment by skilled labour in HOST country	X

(*Continued*)

TABLE 2
(*Continued*)

Variable	Definition	Included in previous study listed in table 1
18. PARENT education level	Average education years in PARENT country	
19.PARENT skill level	Percent of employment by skilled labour in PARENT country	
20. Squared education difference	Squared difference in average education years between PARENT and HOST countries (proxy for relative skilled labour endowments)	X
21. Squared skill difference	Squared difference in % of employment by skilled labour between PARENT and HOST countries (proxy for relative skilled labour endowments)	X
22. Interaction of GDP differences with education differences	Interaction of GDP differences with education differences	X
23. Interaction of GDP differences with skill differences	Interaction of GDP differences with skill differences	X
Other relative endowment measures		
24. HOST capital per worker	Capital per worker in HOST country	
25. PARENT capital per worker	Capital per worker in PARENT country	
26. Squared difference in capital per worker	Squared difference in capital per worker between HOST and PARENT countries	X
27. HOST land area	Land area (sq. km) in HOST country	
28. PARENT land area	Land area (sq. km) in PARENT country	
29. HOST population density	Population divided by land area in HOST country	X
30. HOST is oil country	Indicator variable that the HOST country is a top 10 producer or top 10 exporter of oil	
Cultural distance		
31. Common official language	Indicator variable that PARENT and HOST countries share a common official language	X
32. Common language overlap	Indicator variable that PARENT and HOST countries share a language that at least 9% speak in each country	
33. Colonial relationship	Dummy variable indicating PARENT and HOST countries have had (or do have) a colonial link	X
Multilateral trade openness		
34. HOST trade openness	HOST country openness (imports plus exports divided by GDP) in constant prices (constant prices, in %)	X
35. PARENT trade openness	PARENT country openness (imports plus exports divided by GDP) in constant prices (constant prices, in %)	X
36. Interaction of education differences with HOST trade openness	Interaction of education differences with HOST trade openness	X

(*Continued*)

TABLE 2
(*Continued*)

Variable	Definition	Included in previous study listed in table 1
37. Interaction of skill differences with HOST trade openness	Interaction of skill differences with HOST trade openness	X
Bilateral trade openness		
38. Regional trade agreement	Indicator variable for regional trade agreement between PARENT and HOST countries	X
39. Customs union	Indicator variable for customs union between PARENT and HOST countries	X
40. Service sector agreement	Indicator variable for economic integration agreement in services between PARENT and HOST countries	X
Host country FDI/business costs		
41. HOST time to enforce contract	Time required to enforce a contract (days) in HOST country	
42. HOST time to register property	Time required to register property (days) in HOST country	
43. HOST time to start business	Time required to start a business (days) in HOST country	
44. HOST time to resolve insolvency	Time to resolve insolvency (years) in HOST country	
Host country tax policies		
45. HOST corporate tax	Highest marginal tax rate, corporate rate (%) in HOST country	X
46. HOST is tax haven	Indicator variable that the HOST country is considered a tax haven by OECD	
Bilateral tax and investment agreements		
47. Bilateral investment treaty	Dummy variable indicating a bilateral investment treaty in place between HOST and PARENT countries before July 1 of year	X
48. Double taxation treaty	Dummy variable indicating a double taxation treaty governing "income and capital" in place between HOST and PARENT countries before July 1 of year	X
Host country communications infrastructure		
49. HOST telephones	Mobile and fixed-line telephone subscribers (per 100 people) in HOST country	
50. HOST Internet users	Internet users (per 100 people) in HOST country	
51. HOST computers	Personal computers (per 100 people) in HOST country	
Host country financial infrastructure		
52. HOST domestic credit	Domestic credit provided by banking sector in HOST country (% of GDP)	X

(*Continued*)

TABLE 2
(*Continued*)

Variable	Definition	Included in previous study listed in table 1
53. HOST market capitalization	Market capitalization of listed companies (% of GDP)	X
Political environment and institutions		
54. HOST legal institutions	Strength of legal rights index (0 = weak to 10 = strong) in HOST country	X
55. HOST political rights	Political rights index for HOST country (ranges from 1 to 7, with highest score indicating the lowest level of freedom)	X
56. HOST civil liberties	Civil liberties index for HOST country (ranges from 1 to 7, with highest score indicating the lowest level of freedom)	

5. Results

Because previous studies have employed a variety of FDI measures and specifications (e.g., logging variables or not), the reported results below proceed through a number of possible combinations of the FDI measure and variable transformation, before we compare our results to those in previous studies.

5.1. Base results

We begin with results using our benchmark measure of FDI (FDI stock) as our dependent variable, considering both a specification where all (non-binary) variables are logged and a specification where all variables are not logged. We refer to these as the "log-levels" and "levels" specifications, respectively. Note that interactive variables drop out of the log-levels specifications as they generate perfect collinearity in the regression upon taking logarithms. For each potential covariate, table 3 reports the posterior inclusion probability and the mean of the BMA posterior density for the covariate's slope coefficient for both the levels (columns 1 and 2) and log-levels (columns 3 and 4) specifications using our sample of 2000 data. Again, the posterior inclusion probability and mean of the BMA posterior distribution are computed as in equations (13) and (14), respectively.

One striking similarity between the levels and log-levels specifications is the relatively small set of variables out of the 56 potential covariates that have high inclusion probabilities. Only seven variables have inclusion probabilities at or above 50% in the levels specification, while the analogous number of variables is 16 in the log-levels specification. This suggests a fairly parsimonious specification is sufficient to explain cross-country FDI patterns. In the levels specification, only GDP-related variables, the colonial relationship variable and the bilateral

TABLE 3
Level and log-level regressions to explain FDI stocks in 2000

Variable	Levels: Inclusion probability	Levels: Posterior mean	Log-levels: Inclusion probability	Log-levels: Posterior mean
1. PARENT real GDP	100	6,322.22	100	1.40
2. HOST real GDP	100	6,606.47	100	1.74
3. Distance	39	−0.15	100	−0.94
4. PARENT real GDP per capita	9	0.05	100	2.31
5. HOST real GDP per capita	1	0.00	2	0.01
6. Sum of HOST and PARENT real GDP	0	0.00	0	0.00
7. Similarity of HOST and PARENT real GDP	86	21,097.71	2	0.00
8. Squared GDP difference	100	−326.09	1	0.00
9. Squared GDP per capita difference	8	0.00	1	0.00
10. HOST urban concentration	0	0.08	52	0.63
11. PARENT urban concentration	0	0.00	1	0.00
12. Contiguous border	3	158.15	1	0.00
13. HOST remoteness	0	0.00	100	2.29
14. PARENT remoteness	2	−0.01	30	0.27
15. Time zone differences	4	−15.68	5	0.01
16. HOST education level	0	−0.03	1	0.00
17. HOST skill level	39	6,090.09	97	1.94
18. PARENT education level	0	0.67	1	0.00
19. PARENT skill level	1	83.09	1	0.00
20. Squared education difference	0	−0.04	7	−0.01
21. Squared skill difference	1	0.00	89	1.11
22. Interaction of GDP differences with education differences	100	−3.53	NA	NA
23. Interaction of GDP differences with skill differences	0	−5.02	NA	NA
24. HOST capital per worker	4	−0.01	1	0.00
25. PARENT capital per worker	8	−0.02	36	0.25
26. Squared difference in capital per worker	1	0.01	4	0.00
27. HOST land area	7	0.00	3	0.00
28. PARENT land area	1	0.00	1	0.00
29. HOST population density	0	0.52	3	0.01
30. HOST is oil country	2	−67.25	92	−0.92
31. Common official language	39	2,613.06	92	1.08
32. Common language overlap	1	13.52	1	0.00
33. Colonial relationship	81	8,071.84	87	1.14
34. HOST trade openness	6	1.63	95	0.79
35. PARENT trade openness	0	0.05	1	0.00
36. Interaction of education differences with HOST trade openness	0	0.00	NA	NA
37. Interaction of skill differences with HOST trade openness	1	1.42	NA	NA
38. Regional trade agreement	0	8.80	100	1.47
39. Customs union	1	22.61	97	1.15
40. Service sector agreement	49	2,887.97	4	0.03
41. HOST time to enforce contract	0	0.01	1	0.00
42. HOST time to register property	3	−0.61	26	0.05
43. HOST time to start business	6	−2.27	3	−0.01
44. HOST time to resolve insolvency	0	−0.16	1	0.00
45. HOST corporate tax	0	0.07	67	−0.56

(*Continued*)

TABLE 3
(*Continued*)

Variable	Levels: Inclusion probability	Levels: Posterior mean	Log-levels: Inclusion probability	Log-levels: Posterior mean
46. HOST is tax haven	0	10.81	4	0.10
47. Bilateral investment treaty	50	−1,838.23	1	0.00
48. Double taxation treaty	0	−3.21	23	0.10
49. HOST telephones	2	−0.72	1	0.00
50. HOST Internet users	2	0.80	1	0.00
51. HOST computers	6	5.28	2	0.00
52. HOST domestic credit	3	0.63	1	0.00
53. HOST market capitalization	4	0.72	7	0.02
54. HOST legal institutions	1	5.72	85	−0.68
55. HOST political rights	0	−1.09	5	−0.02
56. HOST civil liberties	1	−9.07	1	0.00
Sample size	1,066		1,066	

NOTES: "Inclusion probability" refers to the posterior probability that the associated variable is in the true FDI determinants model. "NA" denotes "not applicable" when the variable is not included because it is perfectly collinear with other variables once logged.

investment treaty variable have high inclusion probabilities for explaining FDI stock.

Our preferred specification is the log-levels specification because of the substantial skewness in the dependent variable. In that specification, the evidence suggests that standard gravity variables with a few friction variables comprise the bulk of the variables with explanatory power for cross-country FDI patterns. The key gravity variables—real GDP for the host and parent countries, distance, common language and colonial relationships—all have inclusion probabilities above 85% in the log-levels specification. In addition, the trade openness variables indicating the presence of a custom union, the presence of a regional trade agreement and host-country country openness, all have inclusion values above 90%. There is also evidence that endowment differences across the host and parent country may matter, as predicted by some models of FDI, such as the knowledge-capital model of CMM (2001). The host-country skill level and the squared skill difference between the host and parent country have high inclusion probabilities, though all other endowment variables (including those capturing capital and land differences) have very low inclusion probabilities.[14] In general, other broad categories of variables receive little statistical support, particularly those related to business costs, infrastructure and institutions in the host country. The exception is some support for legal institutions (85%) and the corporate tax level (67%) in the host country. On the other hand, there are a few variables not

14 The exception is an indicator for whether the host country is an oil-producing country. However, as will be discussed in section 5.3, oil production in the host country is associated with reduced FDI rather than increased FDI.

TABLE 4
Inclusion probabilities above 50% using alternative measures of FDI (logged 2000 data)

Variable	FDI stock	Affiliate sales	Cross-border M&A
PARENT real GDP	**100**	**100**	**100**
HOST real GDP	**100**	**100**	**100**
Distance	**100**	**100**	**100**
PARENT real GDP per capita	**100**	**99**	**100**
HOST remoteness	**100**	**100**	**100**
Regional trade agreement	**100**	4	**100**
Customs union	**97**	1	**100**
HOST skill level	**97**	1	**100**
HOST trade openness	**95**	3	2
Common official language	**92**	1	**100**
HOST is oil country	**92**	1	**94**
Squared skill difference	**89**	2	10
Colonial relationship	**87**	1	**97**
HOST legal institutions	**85**	22	1
HOST corporate tax	**67**	**95**	3
HOST urban concentration	**52**	0	1
PARENT remoteness	30	0	**100**
Squared GDP per capita difference	1	**82**	2
PARENT urban concentration	1	0	**98**
PARENT skill level	1	1	**100**
HOST time to resolve insolvency	1	2	**91**
Sample size	1,066	395	1,066

NOTES: The table displays all variables that have at least a 50% inclusion probability in one of the listed specifications. Instances where the inclusion probability is 50% or higher are in **bold**.

typically included in empirical FDI studies that have very high inclusion probability in our log-levels specification. These are real GDP per capita in the parent country (100%), remoteness of the host country (100%) and, to a lesser extent, urban concentration of the host country (52%).

Our results to this point use FDI stock as our measure of cross-country FDI activity. Table 4 next compares results when we use two other measures of FDI that have been used by prior studies—affiliate sales and cross-border M&A activity. The table displays all variables that receive at least 50% in one of our three specifications (FDI stock, affiliate sales or M&A). For ease in reading the table, we bold the instances where the inclusion probability is 50% or higher. For comparison sake, we report only the results for the log-levels specification, and, for the M&A sample, we use only observations for the 1,066 country pairs for which we observe the FDI stock variable. (We have many more country-pair observations for the M&A sample that we will analyze and discuss below.) We use all observations available for affiliate sales, but this provides just 395 observations.

Despite these data issues, many of the patterns found in the FDI stock specification are also found when using these other FDI measures. First, the traditional gravity variables (real GDP of both countries and distance) all have inclusion probabilities of 100% across all three specifications. Parent-country real GDP per capita also has at least a 99% inclusion probability across all three, suggesting that the wealth of the source country is a key determinant of FDI. Interestingly, host country real GDP per capita does not have similarly high inclusion probabilities. There is a similar asymmetry in that host country remoteness generally garners high inclusion probabilities across all the measures of FDI activity, whereas parent country remoteness does not. It has a high inclusion probability only in the cross-border M&A specification. These asymmetric results are an example of empirical patterns our analysis finds that have not been examined by prior theory or empirical studies of FDI to our knowledge.

In general, the M&A and FDI stock samples share many variables with high inclusion probabilities beyond the ones we have mentioned, including common official language, colonial relationship, regional trade agreement, customs union, host oil country and host skill level. One interesting difference between the M&A and FDI stock results are that while legal institutions and corporate taxes in the host country have modestly high inclusion probabilities for FDI stock, they have very low ones in the M&A sample. Instead, days to resolve insolvencies in the host country is the only host country business cost variable to have a high inclusion variable in the M&A sample. One final notable difference is that *parent-country* remoteness and urban concentration have high inclusion probabilities in the M&A sample, but not in the other samples.

The FDI stock and affiliate sales specifications find less commonality in the variables that have high inclusion probabilities. We have also produced results for the FDI stock and affiliate sales specifications on a common, overlapping sample of 253 observations and found much more similarity in results that mirror those for affiliate sales in table 4. This suggests that the differences across the affiliate sales and FDI stock specifications in table 4 are due primarily to the relatively small sample available for the affiliate sales measure. Overall, the general patterns noted in earlier specifications reported above continue to hold—gravity finds very strong support, while cultural distance and endowment variables find support as well. In contrast, there continues to be much less support for variables capturing host country business costs, infrastructure or institutions.

As mentioned, the data on FDI stock and affiliate sales is limited primarily to OECD country pairs, though there is some information on FDI from OECD into less-developed countries, but not on FDI patterns between less-developed countries. On the one hand, this selection may not be a significant issue because the vast majority of FDI in the world economy is between the developed economies, which are well represented in our sample. On the other hand, it is useful to know how FDI determinants may differ when a more representative sample of countries is examined. Our M&A data source has the ability to address this as it is a census of worldwide M&A activity.

Table 5 lists all variables with inclusion variables above 50% for three specifications using logged data for the year 2000. The first two columns of inclusion probabilities are for comparison purposes and are for the FDI stock specification and the M&A specification when limited to the same observations as the

TABLE 5
Inclusion probabilities above 50% for OECD and worldwide samples (logged 2000 data)

	OECD sample		Worldwide sample
Variable	FDI stock	Cross-border M&A	Cross-border M&A
HOST real GDP	**100**	**100**	**100**
PARENT real GDP	**100**	**100**	**100**
Distance	**100**	**100**	**100**
PARENT real GDP per capita	**100**	**100**	**100**
HOST remoteness	**100**	**100**	**100**
Regional trade agreement	**100**	**100**	**100**
Customs union	**97**	**100**	**100**
HOST skill level	**97**	**100**	**71**
HOST country trade openness	**95**	2	2
Common official language	**92**	**100**	**99**
HOST is oil country	**92**	**94**	**92**
Squared skill difference	**89**	10	3
Colonial relationship	**87**	**97**	**100**
HOST legal institutions	**85**	1	1
HOST corporate tax	**67**	3	**99**
HOST urban concentration	**52**	1	1
PARENT remoteness	30	**100**	**100**
Double taxation treaty	23	2	**100**
Squared education difference	7	38	**97**
Service sector agreement	4	1	**97**
Similarity of HOST and PARENT real GDP	2	1	**54**
PARENT education level	1	1	**85**
PARENT urban concentration	1	**98**	**100**
PARENT skill level	1	**100**	**76**
Bilateral investment treaty	1	14	**100**
HOST education level	1	3	**100**
HOST years to resolve insolvency	1	**91**	**98**
Contiguous border	1	1	**95**
Observations	1,066	1,066	3,429

NOTES: The table displays all variables that have at least a 50% inclusion probability in one of the listed specifications. Instances where the inclusion probability is 50% or higher are in **bold**.

FDI stock sample. The third column is the M&A specification when we use all observations for which we have available data (we call this the "worldwide" sample), as opposed to the restricted sample (we call this the "OECD" sample). This more than triples the sample size over the other two listed specifications to 3,429 observations, adding many more observations involving non-OECD countries.[15]

The results from the worldwide M&A sample show a lot of commonalities with the previous results. Gravity variables, cultural distance and relative skilled labour variables all show very high inclusion probabilities. In fact, all of the variables that have high inclusion probabilities in the OECD M&A sample specification (column 2) also have high inclusion probabilities in the worldwide M&A sample

15 In the "OECD" sample, all country-pair observations involve at least one OECD country, and 40% of the country-pair observations are between OECD countries. In the "worldwide" sample, 32% of the country-pair observations do not involve at least one OECD country, and only 18% of the country-pair observations are between OECD countries.

specification. However, the worldwide M&A sample also shows high inclusion probabilities for a number of additional variables. These include a few more endowment variables (education levels in both the host and parent country as well as the squared difference in education levels between the two countries), as one might expect when one includes many more observations between relatively poor non-OECD countries and OECD countries. It also includes variables connected with bilateral treaties (bilateral investment treaty, double taxation treaty and service sector agreements) as well as the presence of a contiguous border. This suggests that these bilateral treaties may be much more important for spurring FDI into non-OECD countries than into OECD ones.

5.2. Implications for prior studies

With our BMA results in hand, we now turn to address the fundamental question of how our BMA results compare to those of previous studies. Virtually all of the prior studies include gravity-related variables, and, thus, our results confirm the inclusion of such variables. Common official language also finds robust support in our analysis and is included in five of the prior eight studies in table 1. Beyond this small set of variables, however, prior studies vary significantly in what they include, and what they include does not necessarily match very well with the variables our analysis finds to have high inclusion probabilities. For example, our analysis finds that parent country wealth (real per capita GDP) has strong and robust support, yet only one study (Head and Ries 2008) of the eight studies in table 1 includes this variable. In contrast, four of the studies in table 1 include host country wealth, yet we find this variable does not have strong support for inclusion. The reason for this asymmetry in wealth effects on FDI is also something that past theoretical papers, to our knowledge, do not address. Only four of the prior eight studies include variables related to relative skilled-labour endowment levels or differences, whereas our analysis finds that such variables should be included. There is little evidence that other relative endowments matter besides the presence of oil in the host country. Colonial relationships, host country remoteness, trade agreements and customs unions are additional variables that find strong support in our analysis but are rarely included in prior studies. On the other hand, a number of the prior studies include variables connected to host country business costs, infrastructure and institutions, but these do not find robust support in our analysis. Finally, the studies in table 1 whose main focus is on a particular hypothesized relationship between a potential covariate and FDI generally do not fare very well in terms of the inclusion probabilities we estimate for the same covariate. This includes Wei (2000), whose focus is on corruption; Stein and Daude (2007), whose focus is on time zone differences; and di Giovanni (2005), whose partial focus is on financial market institutions.

5.3. Slope coefficient magnitudes

To this point, we have focused only on inclusion probabilities. In table 6, we report estimates of the slope coefficient of the variables listed in table 5. In particular,

TABLE 6
Posterior mean and variance of slope coefficients for OECD and worldwide samples (logged 2000 data)

	OECD sample		Worldwide sample
Variable	FDI stock	Cross-border M&A	Cross-border M&A
HOST real GDP	**1.74 (0.02)**	**0.97 (0.00)**	**0.61 (0.00)**
PARENT real GDP	**1.40 (0.00)**	**1.02 (0.00)**	**0.79 (0.00)**
Distance	**−0.94 (0.02)**	**−0.63 (0.01)**	**−0.44 (0.00)**
PARENT real GDP per capita	**2.31 (0.14)**	**1.29 (0.01)**	**0.61 (0.01)**
HOST remoteness	**2.29 (0.24)**	**1.24 (0.05)**	**0.64 (0.01)**
Regional trade agreement	**1.47 (0.09)**	**1.37 (0.04)**	**1.19 (0.02)**
Customs union	**1.15 (0.11)**	**1.23 (0.04)**	**0.92 (0.04)**
HOST skill level	**1.94 (0.36)**	**1.40 (0.08)**	**0.26 (0.04)**
HOST country trade openness	**0.79 (0.07)**	0.00 (0.00)	0.00 (0.00)
Common official language	**1.08 (0.20)**	**1.08 (0.04)**	**0.45 (0.01)**
HOST is oil country	**−0.92 (0.13)**	**−0.56 (0.04)**	**−0.29 (0.01)**
Squared skill difference	**1.11 (0.25)**	0.05 (0.03)	0.01 (0.00)
Colonial relationship	**1.14 (0.31)**	**0.93 (0.08)**	**1.22 (0.02)**
HOST legal institutions	**−0.68 (0.12)**	0.00 (0.00)	0.00 (0.00)
HOST corporate tax	**−0.56 (0.19)**	−0.01 (0.00)	**−0.32 (0.01)**
HOST urban concentration	**0.63 (0.44)**	0.00 (0.00)	0.00 (0.00)
PARENT remoteness	0.27 (0.21)	**1.17 (0.05)**	**0.59 (0.01)**
Double taxation treaty	0.10 (0.04)	0.00 (0.00)	**0.42 (0.00)**
Squared education difference	−0.01 (0.00)	−0.03 (0.00)	**−0.06 (0.00)**
Service sector agreement	0.03 (0.03)	0.00 (0.00)	**0.68 (0.05)**
Similarity of HOST and PARENT real GDP	0.00 (0.00)	0.00 (0.00)	**0.13 (0.02)**
PARENT education level	0.00 (0.00)	0.00 (0.00)	**0.34 (0.03)**
PARENT urban concentration	0.00 (0.00)	**−0.76 (0.05)**	**−0.50 (0.00)**
PARENT skill level	0.00 (0.00)	**1.19 (0.06)**	**0.32 (0.05)**
Bilateral investment treaty	0.00 (0.00)	−0.04 (0.01)	**−0.36 (0.00)**
HOST education level	0.00 (0.00)	0.01 (0.00)	**0.69 (0.02)**
HOST years to resolve insolvency	0.00 (0.00)	**0.25 (0.01)**	**0.15 (0.00)**
Contiguous border	0.00 (0.00)	0.00 (0.00)	**0.21 (0.01)**
Observations	1,066	1,066	3,429

NOTES: The table displays the posterior mean and variance (in parentheses) of slope coefficient for all variables that have 50% or higher inclusion probability for at least one of the listed specifications. Coefficients where the associated inclusion probability is 50% or higher are in **bold**.

for the variables and specifications in table 5, we report the mean and variance of the BMA posterior density for the slope coefficient on each variable, calculated as inequations (14) and (15). With few exceptions, the coefficient signs are as one would expect and consistent with prior studies. This includes the gravity variables, cultural distance variables and bilateral trade openness variables. For many of the coefficients, the magnitude of the effect is smaller in the worldwide M&A sample than for the OECD sample, which suggests that FDI responds much less to economic forces for host countries that are less developed. A few of the coefficients have unexpected signs. One of the more intriguing results is that while the bilateral distance between country pairs lowers FDI (as expected), the remoteness of both the parent and host countries (that is, how far they are from the entire world's markets, not just the other country in the country pair) has positive coefficients. This distinction has not been made before to our knowledge but certainly deserves future investigation. Another surprising result is that the presence of oil in the host country is associated with lower FDI, as is the strength of host country legal institutions.

The posterior mean and variance of the slope coefficients can also be used to construct a "pseudo t-ratio" by dividing the posterior mean by the posterior standard deviation. This statistic is a natural candidate as a measure of the relative importance of potential covariates in a linear regression setting and is sometimes reported along with BMA inclusion probabilities. In our application, the pseudo t-ratios (not reported) provide very similar conclusions to those reached from the BMA inclusion probabilities regarding the relative importance of the candidate FDI determinants. As a specific example, for the log-levels specification using FDI stocks as the dependent variable, the ranking of FDI determinants by the two metrics is nearly identical, with the Spearman rank correlation statistic between the pseudo t-ratio and the BMA inclusion probability equal to 0.98.

5.4. Controlling for country-pair effects

Many prior studies of FDI determinants include country or country-pair effects. A simple way to control for such effects is to difference the data by country-pair combinations. Table 7 provides results from log-linear specifications for a 1990–2000 differenced sample for our FDI stock, OECD M&A and worldwide M&A samples. First-differencing in this manner eliminates a number of time-invariant variables, as is typical. It unfortunately also eliminates a very large portion of the observations, due to many more missing values for variables in 1990. This may be why the FDI stock and OECD M&A samples have only one variable between them that comes in with an inclusion probability over 50%, though a possible alternative explanation is that bilateral FDI patterns are largely driven by slow-moving or time-invariant factors that are then differenced out of these regressions. However, the worldwide M&A sample still has over 1,200 observations and finds 12 variables to have inclusion probabilities over 50%. What we find most important is that these high-inclusion probabilities in the first-differenced worldwide M&A sample are largely the same ones as we have found throughout the many varied permutations we have evaluated in this paper: GDP-related variables, skilled-labour variables and trade agreements. Distance and cultural distance factors do not show up in this table because first-differencing leaves no (or virtually no) variation from which to identify the impact of these factors.

TABLE 7
Inclusion probabilities above 50% for OECD and worldwide samples (logged and first-differenced 2000 data)

	OECD sample		Worldwide sample
Variable	FDI stock	Cross-border M&A	Cross-border M&A
PARENT real GDP per capita	**96**	17	2
PARENT real GDP	2	22	**100**
PARENT remoteness	1	0	**97**
PARENT urban concentration	0	44	**100**
HOST real GDP	0	8	**100**
PARENT education level	0	2	**100**
Regional trade agreement	0	1	**100**
Service sector agreement	0	0	**100**
Customs union	0	0	**97**
GDP similarity	0	6	**96**
HOST real GDP per capita	0	1	**97**
PARENT skill level	0	1	**92**
HOST skill level	0	17	**78**
Observations	244	244	1,246

NOTES: The table displays all variables that have at least a 50% inclusion probability in one of the listed specifications. Instances where the inclusion probability is 50% or higher are in **bold.**

5.5. Robustness to an alternative parameter prior

The results presented in 5.4 were generated for a specific choice of parameter prior distribution, namely those suggested in FLS, as described in section 3. It is well known that BMA results can be sensitive to parameter priors, although, for the relatively large sample sizes available in our application, this sensitivity should be muted. To verify this, we also present results from an alternative prior specification known as the Unit Information Prior (UIP). The UIP is designed to contain roughly the same amount of information as a typical single observation (Kass and Wasserman 1995). EPR (2011) argue for the UIP as a reasonable "default" prior based on evidence that it outperforms the prior of FLS for prediction. As discussed in Kass and Wasserman (1995) and Raftery (1995), the UIP suggests a convenient approximation to the marginal likelihood based on the Bayesian Information Criterion (BIC), which makes this prior simple to implement.

Table 8 compares results from the FLS parameter priors to those based on the UIP for the FDI stock measure of FDI and the log-levels specification. The table displays all variables that receive a 50% or higher inclusion probability for at least one of the alternative priors. For ease in reading the table, we bold the instances where the inclusion probability is 50% or higher. The inclusion probabilities suggest that the BMA results are not very sensitive to parameter priors, which again is what we might have expected given the relatively large sample size. In particular, the inclusion probabilities are generally close in magnitude for the two alternative priors, and there is no case where the

TABLE 8
Inclusion probabilities above 50% using alternative parameter priors (FDI stock – logged 2000 data)

Variable	FLS	UIP
PARENT real GDP	**100**	**100**
HOST real GDP	**100**	**100**
Distance	**100**	**100**
PARENT real GDP per capita	**100**	**100**
HOST remoteness	**100**	**100**
Regional trade agreement	**100**	**100**
Customs union	**97**	**97**
HOST skill level	**97**	**99**
HOST trade openness	**95**	**97**
Common official language	**92**	**93**
HOST is oil country	**92**	**97**
Squared skill difference	**89**	**97**
Colonial relationship	**87**	**95**
HOST legal institutions	**85**	**94**
HOST corporate tax	**67**	**84**
HOST urban concentration	**52**	**67**
PARENT capital per worker	36	**50**
Observations	1,066	1,066

NOTES: The table displays all variables that have 50% or higher inclusion probability for at least one of two alternative specifications for parameter priors. Instances where the inclusion probability is 50% or higher are in **bold**. Results are for the FDI stock dataset and log-levels specification. FLS refers to priors suggested by Fernandez, Ley and Steel (2001a), as described in section 3. UIP refers to the Unit Information Prior of Kass and Wasserman (1995), as described in section 5.

two priors yield radically different conclusions regarding the importance of a covariate.

5.6. *Robustness to a nonlinear specification*

Due to the computational intensity of the BMA approach, our analysis to this point was restricted to linear regression models. However, there are some potential issues with a linear specification for the measures of FDI used as the dependent variable. First, there are many country pairs for which the FDI measure is zero. This creates an issue in the log-level regressions, as the logarithm of these observations is undefined. In the results presented above, we retained these observations in the sample by adding a small constant to each FDI measure before taking logarithms. Alternatively, we could have eliminated these observations from the sample. As is discussed in Santos Silva and Tenreyro (2006), each of these solutions might distort inference from that produced by an appropriate nonlinear model estimated on the levels of the dependent variable. These authors argue for the use of Poisson regression methods to effectively deal with zero observations. Second, our measure of FDI based on M&A activity is a discrete count variable, a fact that is ignored when working in the linear regression framework.

In this section we evaluate the robustness of the conclusions regarding the determinants of FDI when a nonlinear model is used to link FDI to potential covariates. We focus on M&A counts as the FDI measure, as this data displays both of the features discussed above—zero observations and discreteness. We use a negative binomial regression to model the M&A counts. This framework models the level of the M&A counts directly, which eliminates any issues associated with the need to take logarithms of zero observations. Also, the negative binomial distribution is a discrete distribution with a natural interpretation for count data.[16]

Extending the MC^3 algorithm discussed in section 3 for linear regression models to conduct BMA for negative binomial regressions is conceptually straightforward. Specifically, the only change is that the marginal likelihood in (3) is replaced by the marginal likelihood for the negative binomial model. Unfortunately, unlike the case of the linear regression model with natural conjugate priors, the marginal likelihood for the negative binomial model is not available analytically and needs to be approximated. One approach would be to compute a simulation-consistent estimate of the marginal likelihood using Markov chain Monte Carlo techniques. However, when incorporated inside of the large number of simulations necessary for the MC^3 algorithm, this would be very computationally demanding. Instead, we use an asymptotic approximation to the marginal likelihood based on the BIC. This requires only the maximum likelihood estimates of the negative binomial regression and can be computed relatively quickly.[17]

Table 9 shows the posterior inclusion probabilities computed for the log-level linear regression model, along with those based on the negative binomial model, when we use the sample of cross-border M&A counts across OECD countries, i.e., the sample identical to the one used in the last column of table 4. We use the sample of OECD countries, rather than the larger worldwide sample, to reduce the computational time needed to calculate the SIC for the negative binomial specification. Results across the linear and negative binomial models are very similar, suggesting model misspecification bias from running linear models in this setting is small. Out of 52 potential covariates, there are only four instances where the inclusion probability of a covariate is very high in one specification but close to zero in the other. In particular, urban concentration of the parent country and host country oil production have very high inclusion probabilities in the linear specification but inclusion probabilities near zero in the negative binomial

16 A common starting point for modelling count data is the Poisson regression model. However, our sample of M&A count data has sample variance far greater than sample mean, suggesting a model that incorporates this overdispersion is better suited for M&A counts. The negative binomial regression model, which arises from a natural extension of the Poisson regression, is a popular choice for overdispersed count data in the applied literature. Indeed, we experienced substantial convergence issues when estimating a simple Poisson specification, which further indicated that it is important to model overdispersion in these data.

17 The BIC approximation to the marginal likelihood is a common choice in applied work. See, for example, Brock, Durlauf and West (2003) and Doppelhofer, Miller and Sala-i-Martin (2004). For additional discussion of the BIC-based approach to model averaging, see Raftery (1995).

TABLE 9
Inclusion probabilities for linear and negative binomial specifications (cross-border M&A for OECD sample – logged 2000 data)

Variable	Linear model	Negative binomial
HOST real GDP	**100**	**100**
PARENT real GDP	**100**	**100**
Distance	**100**	**100**
PARENT real GDP per capita	**100**	**100**
HOST remoteness	**100**	**100**
Regional trade agreement	**100**	**100**
Customs union	**100**	**100**
HOST skill level	**100**	**100**
Common official language	**100**	**100**
PARENT remoteness	**100**	**100**
PARENT skill level	**100**	**95**
PARENT urban concentration	**98**	2
Colonial relationship	**97**	**100**
HOST is oil country	**94**	0
HOST years to resolve insolvency	**91**	**100**
Squared education difference	38	**95**
HOST political rights	12	**100**
HOST country trade openness	2	**64**
Similarity of HOST and PARENT real GDP	1	**98**
HOST legal institutions	1	**53**
Observations	1,066	1,066

NOTES: The table displays the inclusion probability for every potential covariate listed in table 2, with the exception of the interaction terms, which become perfectly collinear when variables are logged.

specification, while GDP similarity and host country political rights are estimated to have high inclusion probabilities in the negative binomial specification but inclusion probabilities near zero in the linear model. Three other variables, the squared education difference, host country trade openness and host country legal institutions, receive more support in the negative binomial specification, with inclusion probabilities roughly 50 percentage points above those for the linear model. Outside of these seven exceptions, which is less than 15% of the covariates we consider, the average absolute difference in inclusion probabilities across the linear and negative binomial models is just 2.6 percentage points. Also, it is notable that of the five variables for which the negative binomial model provides more support than the linear model, four have high inclusion probabilities for the linear model applied to the FDI stock or worldwide sample M&A count data. Thus, the negative binomial specification is not revealing a substantial number of additional relevant covariates beyond those identified elsewhere in our analysis.

5.7. BMA on trade flows

A related BMA analysis we can perform using our covariates is an examination of the determinants of trade flows. This is an interesting litmus test for the BMA procedure, as we would be concerned, for example, if standard gravity variables did not have high inclusion probabilities for trade flows using our BMA

TABLE 10
Inclusion probabilities above 50% for explaining FDI vs. bilateral trade (worldwide sample – logged 2000 data)

Variable	Cross-border M&A	Bilateral trade
PARENT real GDP	**100**	**100**
HOST real GDP	**100**	**100**
Distance	**100**	**100**
PARENT real GDP per capita	**100**	**100**
HOST remoteness	**100**	**100**
PARENT remoteness	**100**	**100**
Regional trade agreement	**100**	**100**
Customs union	**100**	**100**
Colonial relationship	**100**	**94**
HOST education level	**100**	13
Double taxation treaty	**100**	**100**
Bilateral investment treaty	**100**	**100**
Parent urban concentration	**100**	1
Common official language	**99**	**96**
HOST corporate tax	**99**	20
HOST time to resolve insolvency	**98**	6
Squared education difference	**97**	1
Service sector agreement	**97**	2
Contiguous border	**95**	19
HOST is oil country	**92**	**62**
PARENT education level	**85**	**100**
PARENT skill level	**76**	1
HOST skill level	**71**	0
Similarity of HOST and PARENT real GDP	**54**	**100**
PARENT trade openness	4	**98**
Squared skill difference	3	**100**
HOST trade openness	2	**51**
HOST land area	2	**95**
HOST legal institutions	1	**100**
HOST urban concentration	1	**60**
HOST Internet users	1	**100**
HOST domestic credit	1	**70**
HOST time to start business	1	**99**
Observations	3,429	3,429

methods. The analysis also provides a comparison of the determinants of FDI and trade flows within the same framework. We gather data on bilateral trade flows from the dataset connected with Rose and Spiegel (2011) and made available online by Andrew Rose at **faculty.haas.berkeley.edu/arose/RecRes.htm#Software**. Specifically, the data are CIF imports measured in US$, taken from International Financial Statistics' Direction of Trade CD-ROM, deflated by U.S. CPI for All Urban Consumers (CPI-U), all items, 1982 to 1984 = 100. For comparison purposes, we sample the year 2000 for the same observations we use for our cross-border M&A results, which yielded the largest sample size out of all the FDI measures.

Table 10 provides inclusion probabilities for our BMA analysis of trade flows, as well as repeats the cross-border M&A inclusion probabilities for comparison purposes. The table displays all variables that receive a 50% or higher inclusion probability for explaining at least one of either trade flows or M&A counts. Reassuringly, our BMA analysis of trade flows yields results that are quite in line with accepted practice on how to specify trade flows. The gravity variables and typical frictions (including trade and FDI agreements) show very strong support. This also means that our analysis suggests very similar determinants for trade and cross-border M&A, though there is much less support generally for endowment terms with trade than for cross-border M&A. The biggest differences between the two come in which business cost, infrastructure and other host-country attributes matter for trade versus cross-border M&A. For example, the number of Internet users, legal institutions and domestic credit have high inclusion probabilities for trade but not for cross-border M&A.

This is not the first BMA analysis of trade flows. Eicher, Henn and Papageorgiou (2012) use BMA methods to examine the impact of preferential trade agreements on trade flows. They estimate very similar inclusion probabilities for the common overlap of variables between their study and ours, including strong support for GDP terms, geographic features (such as distance) and cultural distance terms.

6. Conclusion

The prior literature examining the determinants of FDI comprises a limited number of studies that typically propose fairly parsimonious specifications, but quite varied in their specifications and FDI measurement. This suggests significant uncertainty in the true model of bilateral cross-country FDI patterns. Our approach is to provide some needed systematic investigation of the determinants of FDI by using Bayesian Model Averaging. Our analysis does not support the inclusion of many variables found in prior FDI studies and suggests that the statistical importance of the main focus variables in many prior studies is not robust to considering a much wider set of covariates. The results also suggest a fairly parsimonious FDI specification comprised of mainly gravity variables, cultural distance factors, parent-country per capita GDP, relative labour endowments and trade agreements.

Of note, our results reflect much less support for government policies to encourage FDI, as there is little robust evidence in our analysis that policy variables controlled by the host country (such as multilateral trade costs, business costs, infrastructure or political institutions) have an effect on FDI. Exceptions include policies that are often negotiated bilateral agreements, including trade agreements, bilateral investment treaties, customs unions and service agreements in the case of M&A. However, we caution that exogeneity of these variables may be more in doubt than many of the other covariates we consider.

References

Baltagi, B.H., P. Egger and M. Pfaffermayr (2007) "Estimating models of complex FDI: Are there third-country effects," *Journal of Econometrics* 140, 260–81

Bergstrand, J.H., and P. Egger (2007) "A knowledge-and-physical-capital model of international trade flows, foreign direct investment, and multinational enterprises," *Journal of International Economics* 73, 278–308

Blonigen, B.A., R. B. Davies, G. R. Waddell and H. Naughton (2007) "FDI in space: Spatial autoregressive relationships in foreign direct investment," *European Economic Review* 51, 1303–25

Braconier, H., P.-J. Norbäck and D. Urban (2005) "Reconciling the evidence on the knowledge-capital model," *Review of International Economics* 13, 770–86

Brock, W.A., S. N. Durlauf and K. D. West (2003) "Policy evaluation in uncertain economic environments," *Brookings Papers on Economic Activity* 1, 235–301

Carr, D., J. R. Markusen and K. E. Maskus (2001) "Estimating the knowledge–capital model of the multinational enterprise," *American Economic Review* 91, 693–708

Chakrabarti, A. (2001) "The determinants of foreign direct investment: Sensitivity analyses of cross-country regressions," *Kyklos* 54, 89–114

di Giovanni, J. (2005) "What drives capital flows? The case of cross-border M&A activity and financial deepening," *Journal of International Economics* 65, 127–49

Doppelhofer, G., R. I. Miller and X. Sala-i-Martin (2004) "Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach," *American Economic Review* 94, 813–35

Draper, D. (1995) "Assessment and propagation of model uncertainty (with discussion)," *Journal of the Royal Statistical Society, Series B* 57, 45–97

Eaton, J., and A. Tamura (1994) "Bilateralism and regionalism in Japanese and U.S. trade and direct foreign investment patterns," *Journal of the Japanese and International Economies* 8, 478–510

Eicher, T.S., L. Helfman and A. Lenkoski (2010) "Robust FDI determinants," *Journal of Macroeconomics*, forthcoming

Eicher, T.S., C. Henn and C. Papageorgiou. (2012) "Trade creation and diversion revisited: Accounting for model uncertainty and natural trading partner effects," *Journal of Applied Econometrics* 27, 296–321

Eicher, T.S., C. Papageorgiou and A. E. Raftery (2011) "Default priors and predictive performance in Bayesian model averaging, with application to growth determinants," *Journal of Applied Econometrics* 26, 30–55

Fernandez, C., E. Ley and M. F. J. Steel (2001a) "Benchmark priors for Bayesian model averaging," *Journal of Econometrics* 100, 381–427

——— (2001b) "Model uncertainty in cross country growth regressions," *Journal of Applied Econometrics* 16, 563–76

Head, K., and J. Ries (2008) "FDI as an outcome of the market for corporate control: Theory and evidence," *Journal of International Economics* 74, 2–20

Hodges, J.S. (1987) "Uncertainty, policy analysis, and statistics," *Statistical Science* 2, 259–91

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

Jordan, A., and A. Lenkoski (2012), "Tobit Bayesian model averaging and the determinants of foreign direct investment," mimeo

Kass, R.E., and A. E. Raftery (1995) "Bayes factors," *Journal of the American Statistical Association* 90, 773–95

Kass, R.E., and L. Wasserman (1995) “A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion,” *Journal of the American Statistical Association* 90, 928–34

Koop, G. (2003) *Bayesian Econometrics.* Hoboken, NJ: J. Wiley

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

Ley, E., and M. F. J. Steel (2009) “On the effect of prior assumptions in Bayesian model averaging with applications to growth regression,” *Journal of Applied Econometrics* 24, 651–74

Madigan, D., and J. York (1995) “Bayesian Graphical Models for Discrete Data,” *International Statistical Review* 63, 215–32

Magnus, J.R., O. Powell and P. Prüfer (2010) “A comparison of two model averaging techniques with an application to growth empirics,” *Journal of Econometrics* 154, 139–53

Moulton, B.R. (1991) “A Bayesian approach to regression selection and estimation with application to a price index for radio services,” *Journal of Econometrics* 49, 169–93

Raftery, A.E. (1995) “Bayesian model selection for social research,” *Sociological Methodology* 25, 111–63

——— (1996) “Approximate Bayes factors and accounting for model uncertainty in generalised linear models,” *Biometrika* 83, 251–66

Raftery, A.E., D. Madigan and J. A. Hoeting (1997) “Bayesian model averaging for linear regression models,” *Journal of the American Statistical Association* 92, 179–91

Rose, A.K., and M. M. Spiegel. (2011) “The Olympic effect,” *Economic Journal* 121, 652–77

Rossi, S., and P. F. Volpin (2004) “Cross-country determinants of mergers and acquisitions,” *Journal of International Economics* 74, 277–304

Santos Silva, J.M.C., and S. Tenreyro (2006) “The log of gravity,” *Review of Economics and Statistics* 88, 641–58

Stein, E., and C. Daude (2007) “Longitude matters: Time zones and the location of foreign direct investment,” *Journal of International Economics* 71, 96–122

Wei, S-J. (2000) “How taxing is corruption on international business,” *Review of Economics and Statistics* 82, 1–11

Zellner, A. (1986) “On assessing prior distributions and Bayesian regression analysis with g-prior distributions,” in *Bayesian Inference and Decision Techniques: Essays in Honour of Bruno de Finetti.*, eds. P.K. Goel and A. Zellner. Amsterdam: North Holland