

**Bayesian Estimation of Multi-
variate Panel Probits with
Higher-Order Network Inter-
dependence and an Applica-
tion to Firms' Global Market
Participation in Guangdong**

Badi H. Baltagi, Peter H. Egger, Michaela Kesina

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: <https://www.cesifo.org/en/wp>

Bayesian Estimation of Multivariate Panel Probits with Higher-Order Network Interdependence and an Application to Firms' Global Market Participation in Guangdong

Abstract

This paper proposes a Bayesian estimation framework for panel-data sets with binary dependent variables where a large number of cross-sectional units is observed over a short period of time, and cross-sectional units are interdependent in more than a single network domain. The latter provides for a substantial degree of flexibility towards modelling the decay function in network neighbourliness (e.g., by disentangling the importance of rings of neighbors) or towards allowing for several channels of interdependence whose relative importance is unknown *ex ante*. Besides the flexible parameterization of cross-sectional dependence, the approach allows for simultaneity of the equations. These features should make the approach interesting for applications in a host of contexts involving structural and reduced-form models of multivariate choice problems at micro-, meso-, and macroeconomic levels. The paper outlines the estimation approach, illustrates its suitability by simulation examples, and provides an application to study exporting and foreign ownership among potentially interdependent firms in the specialized and transport machinery sector in the province of Guangdong.

JEL-Codes: C110, C310, C350, F140, F230, L220, R100.

Keywords: network models, spatial models, higher-order network interdependence, multivariate panel probit, Bayesian estimation, firm-level data, Chinese firms.

Badi Baltagi

*Department of Economics & Center for Policy Research
Syracuse University / NY 13244-1020 / USA
bbaltagi@maxwell.syr.edu*

*Peter H. Egger
ETH Zurich / Switzerland
egger@kof.ethz.ch*

*Michaela Kesina
University of Groningen / The Netherlands
m.kesina@rug.nl*

February 2022

The authors gratefully acknowledge numerous helpful comments by three anonymous reviewers and the editor in charge (Herman K. van Dijk) on an earlier version of the manuscript. Egger gratefully acknowledges funding from SNF under grant no. 100018-169537.

1 Introduction

The past two decades have witnessed major developments in the analysis of economic, econometric, and statistical models with data-generating processes that involve cross-sectionally dependent outcome variables. The economic foundations of such processes include structural models where vectors of outcomes are generated through general-equilibrium models, through strategic interaction, or through non-strategic network interactions (e.g., through social networks). In econometrics, the estimators devised for the analysis of such processes range from maximum likelihood estimation to generalized method of moments to Bayesian Markov chain Monte Carlo simulation.

Three observations regarding the development of estimation methods are important. First, the range of estimators available with processes, where the structural equations are linear in parameters is ample, but this is not the case with processes for nonlinear (e.g., probability) models. Second, research on multivariate equation models, e.g., seemingly unrelated or simultaneous equation models, allowing for cross-sectional dependence is scarce for linear structural-equations models and even more so for nonlinear models. Third, as we explain below, currently available estimators address situations where not only the source or channel but also the degree is fully known up to scale to the researcher. We give two examples to clarify the background for these observations, one related to corporate tax competition between jurisdictions and one to firms' market participation.

With corporate tax competition, one typically estimates a model, where a vector of corporate tax rates (across countries or subnational units) is determined by a weighted average of those tax rates in other jurisdictions and a linear index of shifters of those tax rates (see Devereux, Lockwood, and Redoano, 2008). In this example, the tax rates of other jurisdictions would be weighted either by a matrix with row-sum or maximum-row-sum normalized inverse distances, the coefficient on the weighted tax rate would measure the nature and degree of tax competition and parameterize together with the entries in the weights matrix the slope of the reaction function (e.g., whether tax rates are strategic substitutes or complements), and the linear index would be a parameterization of the shifters of equilibrium tax rates. Three issues are worth noting in this regard. First, corporate tax rates are assumed to be independent and not a function of other tax instruments (e.g., personal income tax rates, sales tax rates, value added tax rates, etc.). Second, the weights used to aggregate other jurisdictions' tax rates are all known to the researcher and only the one parameter on the weighted average tax rates of other jurisdictions is estimated. Hence, the (tax) reaction function is fully postulated up to one unknown scalar. Third, the structural form of this tax-competition problem – where one jurisdiction's tax rate depends on the tax rates of the others – is linear in parameters. For this example, a host of estimators for network problems along the mentioned lines could be used for estimation.

Modern trade models of the firm motivate empirical models where the participation of a firm in the export market depends on the firms' latent profits associated with that participation (see Melitz, 2003). Since a firm's export market participation is captured by a binary indicator variable and net profits associated with exporting are only observed for exporters, the problem is typically analyzed

by means of a probit or other nonlinear probability model and net profits from exporting are treated as latent (see Bernard and Wagner, 2001). Again, three issues are worth noting. First, export market participation is assumed as to be independent and not a function of other globalization strategies (e.g., foreign direct investment, etc.; see Helpman, Melitz, and Yeaple, 2004, for a model where firms consider the choice of exporting and/or going multinational simultaneously). Second, export entry is strategically independent from other firms' choices (i.e., the network weights are set to zero between firms and only a firm's own characteristics and general-equilibrium-related, common aggregate market characteristics matter for that choice). Third, the structural form of this choice problem is nonlinear in parameters. For this example, estimators for single-equation choice problems without network interactions are commonly postulated and used.

However, researchers are faced with processes that do not adhere to such structures. E.g., individuals, firms, or jurisdictions may be faced with a choice between several correlated alternatives. Moreover, the intensity of the network may not be known up to just one single parameter. In this paper, we will focus on the scenario where firms may participate in the global market not only by trading but also by being part of a multinational company with foreign establishments. In doing so, we will permit the choices of export and/or multinational organization participation to be interrelated between firms, and we will also permit the decay of interdependencies to change at a rate that is more flexible than what is customarily assumed.²

In choice problems with cross-sectionally independent units, maximum likelihood estimation is the leading approach. This approach maximizes a nonlinear transform of an index function that is linear in parameters. With, cross-sectional dependence of outcomes, the reduced form of a structural model that is linear in parameters is nonlinear in those parameters, and the associated likelihood function with N cross-sectional units involves evaluating an N -dimensional integral, which is cumbersome already with a single-equation model.³ Therefore, over the past years, Bayesian Markov chain Monte Carlo methods became a popular approach for estimating nonlinear choice models with network interdependencies (see LeSage, 2000; LeSage and Pace, 2009; Wang and Kockelmann, 2009, Zhou, Wang, and Holguin-Veras, 2016, to mention a few).

While many economic models involve frictions which parameterize the decay of strategic or non-strategic interdependencies (in the above examples: the weights used to aggregate other jurisdictions' tax rates or other firms' latent profits), they do not inform us about the mapping of observable friction-related variables (e.g., geographical distance, trade links, etc.) into the theoretical concept of frictions. The latter suggests that an approach, where the mapping of interactions-decay-generating variables is parameterized more flexibly may be preferable over one, where the decay function is fully specified. One approach to permit for a greater flexibility of the decay function considers the weights

²Recall that the customary approach with spatial models is to postulate a network decay of one of two forms: either network effects are identical for all neighbors and zero for nonneighbors (that is the case with a single contiguity matrix) or they decline with distance in a pre-specified, known form.

³Clearly, the maximization problem becomes even more cumbersome with multiple equations at hand.

matrices aggregating other units' outcome and other variables to be split into several matrices to allow for interactions-distance piece-wise slopes (see Lee and Liu, 2010; Badinger and Egger, 2011, 2013, 2015; Gupta and Robinson, 2015; or Han, Hsieh, and Lee, 2017; to mention a few). The latter is fit to provide a more flexible parameterization of the interactions-decay function than an approach with just one fully-specified weights matrix.

This paper proposes an estimation framework for a multivariate binary choice panel-data model that features higher-order network or spatial interdependence in the latent outcomes. Such a model is well suited for analyzing structural systems of binary choices, where the cross-sectional units make their choices in an interdependent way, and where interdependence materializes in a way that cannot be captured by a single weights matrix that is scaled by one unknown parameter.

The merit of such a model can best be described by way of the empirical example we use in this paper. We employ census-type firm-level panel data covering four years. The firms are located in the province of Guangdong in China, and they all belong to the specialized and transport facilities manufacturing sector. The outcomes of interest are the time-variant binary choices about exporting and multinational-network participation. The empirical model can be viewed as a representation of a structural-model framework of exporting and multinational-firm set-up. Economic theory suggests that these decisions are not independent of each other, and multinational firms likely engage in exporting, more so than the average firm. Key parameters in such a set-up are the specific fixed costs to these choices. Traditionally, the economic literature assumes that these fixed costs are independent of the latent outcomes of other firms with regard to the same decisions. We challenge this assumption by allowing fixed costs to depend on latent outcomes of neighboring firms. In doing so, we distinguish between close and less-close neighbors. The results indicate that the (latent) net benefits from exporting depend positively on the net benefits of multinational-network membership. Net benefits of either choice depend on those of neighboring firms. However, the interdependencies between firms do not only vanish with distance, but they vanish *relatively* more strongly for closer than for less-close neighbors. Unveiling such insights together requires the use of a novel model as the one proposed in this paper.

The potential applications of the model we outline in the paper are wide. A fixed number of outcomes and a fixed number of rings of neighbors or different channels of interdependence could be considered.⁴ Examples of other applications could include the following. With regard to firms, one could study the determinants of binary incorporation and different binary managerial leadership styles, when such choices are allowed to be interdependent between firms. With regard to households and individuals, one could study binary labor market participation and fertility choices, again assuming that households make such decisions not independently from each other. At a more macroeconomic level,

⁴It should be noted that the consideration of rings of neighbors, where the weights matrices pertaining to different rings are not correlated should be considered as less problematic than a case where several correlated weights matrices (e.g., parameterizing different spillover channels) are considered. See both Section 2.2. and the simulation results in Debarys and LeSage (2020) for evidence on problems associated with the use of several correlated weights matrices.

one could study the determinants of binary memberships in trade, investment, and tax agreements, allowing for such memberships not to happen independently across countries and country pairs. Economic theory provides structural guidance for modelling all such choices. However, the opportunity to specify latent outcomes as to influence each other and each outcome for any cross-sectional unit to depend flexibly on latent outcome of other units is important.

The proposed model promises two benefits. On the one hand, it permits treating interrelated choices which happen under network interdependencies simultaneously. Moreover, it permits a more flexible specification of the decay function which parameterizes the cross-sectional interdependencies. The latter is important, if one is interested in quantifying the marginal effects of shocks across the agents or units of observation in the data, as these marginal effects do not only depend on the magnitude of the shock and the level of latent outcome in the outset but also on the positioning of the units of observation in the network. Clearly, postulating a decay function of interdependencies which is globally too flat or too steep in some space (e.g., geographical distance) entails that the degree of heterogeneity in the marginal effects of said shock may have a large bias. This can be avoided when allowing for a more flexible decay function for interdependencies.

The remainder of the paper is organized as follows. The subsequent section provides a literature review. Section 3 outlines the econometric model. Section 4 proposes the estimation of this model using Markov Chain Monte Carlo simulation. Simulation results on the performance of the estimation approach are reported in an online appendix and Section 5 presents the empirical application. The last section concludes with a brief summary of the results. We provide further details on the Monte Carlo simulations, the empirical implementation and results, and the theoretically admissible parameter space in an online appendix to the paper.

2 Review of the relevant literature

Models for the analysis of network or spatial data have become increasingly important in recent years. The reason for this is that data at all levels of aggregation – country, region, city, firm, household, etc. – have become available which enable one to study the determinants of outcomes of interest of these aggregates as cross-sectionally dependent processes. Cross-sectional units can depend on each other in geographical space, input-output linkages, factor flows, and market interaction as well as collaboration or membership in the same digital or physical networks. The computing power available in modern times makes it possible to allow for such processes and links in a much wider way than was possible even a decade ago.

Most of the models of interdependence – network or spatial – are available for cross-sectional models with continuous dependent variables, where the structural form of the model is linear in parameters (see Cliff and Ord, 1973, 1981, and Anselin, 1988, Kelejian and Prucha, 1998, 1999, 2004, 2010, Lee, 2003, 2004, 2007, or Lin and Lee, 2010, to mention a few). Much less research is available for the case of limited-dependent, in particular, binary-dependent, variables in the cross section (see McMillen,

1992, Pinkse and Slade, 1998, Klier and McMillen, 2008, LeSage, 2000, Beron and Vijverberg, 2004, Wang, Iglesias, and Wooldridge, 2013), not to mention the panel data case (see Pinkse, Slade, and Shen, 2006, Lei, 2013; Arduini, 2016, or Baltagi, Egger, and Kesina, 2018a).⁵

Two relatively new developments in the literature on network- or spatial-interaction models are particularly relevant for the present paper. The first one is the consideration of multiple equations or systems of equations. Such an approach is desirable, e.g., with structural models, where several outcomes influence each other, and where some or each of these outcomes feature network or spatial interdependence. E.g., Kelejian and Prucha (2004) studied such a case with continuous outcomes in a cross-sectional model, and Drukker, Egger, and Prucha (2021) generalized it and derived the joint variance-covariance matrix of all model estimates. Other studies analyzing systems of – simultaneous as well as seemingly unrelated – equations with continuous outcomes include Cohen-Cole, Liu, and Zenou (2018), Liu (2014), and Yang and Lee (2017) for cross-sectional data and Baltagi and Bresson (2011), Baltagi and Pirotte (2011), and Baltagi and Deng (2015) for panel data. In general, network- or spatial-interdependence multiple-equations models are rarely considered for the case of binary or limited-dependent-variable outcomes (see LeSage and Pace, 2009, Carrión-Flores, Flores-Lagunes, and Guci, 2009, or Baltagi, Egger, and Kesina, 2019, for exceptions in the case of cross-sectional data; also Wang and Kockelmann, 2009, and Baltagi, Egger, and Kesina, 2016, 2017, for exceptions using panel data).

The second new development relates to the parameterization of the degree of network or spatial interdependence towards higher-order social-network or spatial models. These models are particularly important for two situations: (i) when the researcher knows that the channel of interdependence is, e.g., geography but feels uncomfortable to fully parameterize the decay of interdependence up to one normalizing scalar; (ii) when the researcher has a number of possible channels at hand but does not know which ones are more or less relevant. Higher-order network or spatial models allow for the consideration of not just one but several weights matrices in the model and to estimate parameters on each and every interdependence term based on these matrices. Examples of such models have been proposed, again for the continuous-outcome and linear-structural-form case, by Badinger and Egger (2011) and Lee and Liu (2010) for cross-sectional data and by Badinger and Egger (2013, 2015), Gupta and Robinson (2015), Han, Hsieh, and Lee (2017), and Baltagi, Ding, and Egger (2020) for panel data. To the best of our knowledge, higher-order network- or spatial-interdependence models have not been considered so far in the case of limited-dependent and, particularly, binary outcomes.

⁵Even with continuous outcome variables, the literature on network- or spatial-interaction panel data models is much scarcer than the one on cross sections (see Kapoor, Kelejian, and Prucha, 2007, or Lee and Yu 2010a, 2010b, 2012).

3 Model

3.1 Outline

In this section, we outline a panel-data model for a system of equations of binary outcomes which features three generic forms of interdependence: (i) cross-sectional (network or spatial) correlation among the individual units $i = 1, \dots, N$, (ii) equi-correlation of the cross-sectional units over time $t = 1, \dots, T$, and (iii) correlation among the various outcomes as is customary with seemingly-unrelated or systems of equations.

We permit multiple sources of interdependence of the cross-sectional units. For instance, the various sources could reflect different rings (or orders) of neighbors, or they could reflect conceptually different sources of interdependence (such as geography, input-output relationships, etc.); see, e.g., Badinger and Egger, 2013, Prucha, Drukker and Egger, 2017, for such considerations in the context of continuous outcomes with single-equation panel and systems-cross-sectional data, respectively. We will introduce the weights matrices W_{st} to capture interdependence with respect to the source s at time t , whose characteristics will be specified below.

Regarding equi-correlation, we will assume the presence of time-invariant individual-specific characteristics in the stochastic process, denoted by α_{gi} for individual i and outcome (equation) g . The latter may be correlated across equations for a given individual i .

Regarding correlation among the outcomes, we will consider the presence of a latent outcome $h \neq g$ in the process for outcome g . Furthermore, we will consider the correlation of the time-variant error component across the equations for a given individual i at time t .

In general, we will not require the number of individuals to be identical over time, and we will use N_t to refer to the number of cross-sectional units observed at time t , and T_i to refer to the number of time periods we observe unit i , respectively. We use N to denote the unique number of cross-sectional units over all time periods, while n denotes the total number of observations in the data, i.e., $n = \sum_{i=1}^N T_i = \sum_{t=1}^T N_t$.

We outline the model generically here for G outcomes and equations and S sources of interdependence, but we will focus on the case of a bivariate process and a few rings of neighbors in the application and the model simulations in the rest of the paper.

From the viewpoint of a specific unit of observation it and outcome g , we consider the following process:

$$y_{git} = 1(y_{git}^* > 0) \tag{1}$$

$$y_{git}^* = \sum_{h \neq g} \gamma_{gh} y_{hit}^* + \sum_{s=1}^S \lambda_{gs} \bar{y}_{gsit}^* + x_{git} \delta_g + u_{git} \tag{2}$$

$$\bar{y}_{gsit}^* = \sum_{j=1}^{N_t} w_{sijt} y_{gjt}^*, \tag{3}$$

where the parameters γ_{gh} and λ_{gs} could be thought of as structural-equation parameters. In particular,

γ_{gh} governs the structural dependence of *one outcome on the other ones for the same unit it* , while λ_{gs} governs the structural cross-sectional dependence of *the same outcome for individual i on that outcome of other individuals j in the same period*. While γ_{gh} naturally emerges in systems of equations, λ_{gs} is specific to network or spatial models of cross-sectional dependence. δ_g are the structural behavioral parameters on the k_g exogenous variables x_{git} . Clearly, for structural identification, some exclusion restrictions have to be met, of which a necessary one is that the unique number of exogenous variables in the system, K , has to be at least as large as the number of outcomes, G . We implicitly assume that this is the case in all that follows. The disturbance term is denoted by u_{git} and consists of two components:

$$u_{git} = \alpha_{gi} + \nu_{git}.$$

The parameter α_{gi} is a time-invariant, gi -specific component generating equi-correlation, and ν_{git} is a git -specific error component. Both unobserved components are assumed to be correlated across equations for individual i .

We specify the $G \times G$ symmetric variance-covariance matrix of $\alpha_i = (\alpha_{1i}, \dots, \alpha_{Gi})'$ as

$$E(\alpha_i \alpha_i') = (\sigma_{\alpha_g \alpha_h}). \quad (4)$$

Similarly, we specify the $G \times G$ symmetric variance-covariance matrix of $\nu_{it} = (\nu_{1it}, \dots, \nu_{Git})'$ as

$$\Psi = E(\nu_{it} \nu_{it}') = (\sigma_{\nu_g \nu_h}), \quad (5)$$

which we normalize such that $\sigma_{\nu_g \nu_h} = 1$ whenever $g = h$.

Stacking the observations for equation g such that i is the fast index and t is the slow index, as is customary with network or spatial panel models (see Kapoor, Kelejian, and Prucha, 2007), and after collecting the elements w_{sijt} into the $N_t \times N_t$ matrices and $n \times n$ matrices W_{st} and $W_s = \text{diag}_t(W_{st})$, respectively, we obtain

$$y_g = 1(y_g^* > 0) \quad (6)$$

$$y_g^* = \sum_{h \neq g} \gamma_{gh} y_h^* + \sum_s \lambda_{gs} W_s y_g^* + X_g \delta_g + \dot{A} \alpha_g + \nu_g, \quad (7)$$

where y_g , y_g^* , y_h^* , and ν_g denote vectors of dimension $n \times 1$. The regressor matrix X_g is of dimension $n \times k_g$ with the conformable $k_g \times 1$ parameter vector δ_g . \dot{A} is an $n \times N$ matrix which assigns the N elements in $\alpha_g = (\alpha_{g1}, \dots, \alpha_{gN})'$ to the n units of observations.

We may further stack the system of observations to obtain

$$y = 1(y^* > 0), \quad (8)$$

$$y^* = B y^* + X \delta + A \alpha + \nu, \quad (9)$$

$$B = (\Gamma \otimes I_n) + (\Lambda \otimes I_n)(I_G \otimes W), \quad (10)$$

where y , y^* , and ν , are $Gn \times 1$ vectors. $X = \text{diag}(X_g)$ is a $Gn \times k$ matrix with $k = \sum_{g=1}^G k_g$ containing the equation-specific regressors on the diagonal and $\delta = (\delta'_1, \dots, \delta'_G)'$ is its conformable

parameter vector. $A = I_G \otimes \dot{A}$ is the $Gn \times GN$ assignment matrix for the elements in the $GN \times 1$ vector $\alpha = (\alpha'_1, \dots, \alpha'_G)'$. $\Gamma = (\gamma_{gh})$ is a $G \times G$ parameter matrix (upon which suitable restrictions have to be placed to ensure identifiability). $\Lambda = \text{diag}_g(\lambda_g)$ is a $G \times GS$ matrix of parameters with the equation-specific $1 \times S$ vectors of network or spatial autocorrelation parameters $\lambda_g = (\lambda_{g1}, \dots, \lambda_{gS})$ on the diagonal, and the matrix W is defined as $W = (W'_1, \dots, W'_S)'$. I_n and I_G denote identity matrices of dimensions n and G . After defining $L \equiv I_{Gn} - B \equiv I_{Gn} - (\Gamma \otimes I_n) - (\Lambda \otimes I_n)(I_G \otimes W)$, the reduced form of the above model can be written as

$$y^* = L^{-1}(X\delta + A\alpha + \nu). \quad (11)$$

Obviously, this model accommodates two cases of systems of equations with higher-order network or spatial dependence that may emerge. The first one – which we might refer to as the seemingly-unrelated-equations type – would be represented by a situation where $\gamma_{gh} = 0$ for all g, h . The second one – which we might refer to as the simultaneous-equations type – is represented by a situation where $\gamma_{gh} \neq 0$ at least for some g, h .

Towards a specialization of the system for a case where $G = S = 2$, and $g, h \in \{e, m\}$, L reduces to:

$$L = \begin{pmatrix} I_n - \lambda_{e1}W_1 - \lambda_{e2}W_2 & -\gamma_{em}I_n \\ -\gamma_{me}I_n & I_n - \lambda_{m1}W_1 - \lambda_{m2}W_2 \end{pmatrix} \quad (12)$$

and Ψ reduces to

$$\Psi = \begin{pmatrix} 1 & \tau \\ \tau & 1 \end{pmatrix}.$$

3.2 Basic assumptions

We wish to ascertain that the system of equations in this section captures a process that ensures the following. First, finite shocks should result in finite responses of latent outcomes. This should materialize in interior solutions for the associated choice probabilities in each equation of the system. In this section, we need to invoke a customary set of assumptions. Second, the unknown parameters should be identifiable. The set of assumptions we wish to invoke is four-prong, and it addresses requirements for the network-weights matrices W_s , for the parameters scaling the network effects for each neighborhood bracket (or ring of neighbors) captured by the matrix L , for the regressors in X , and for the variance-covariance matrix of the stochastic variables in the model. The following four assumptions summarize conditions, which help in establishing the desired properties.

Assumption 1 - Spatial weights matrices: For all $s = 1, \dots, S$, the diagonal elements of W_s are zero. $\|W_s\|_1$ is smaller than or equal to some finite constant which does not depend on the sample size, and $\|W_s\|_\infty = 1$. The elements in W_s are row-sum normalized.

Specifically, in this paper we will work with matrices W_s which are uncorrelated among each other. Under Assumption 1, any W_s -weighted variable has the same support as its unweighted counterpart.

Assumption 2 - Stability of the system: The matrix L is nonsingular. See the online appendix for a more formal treatment of the conditions underlying the invertibility of L .

Assumption 2 makes sure that finite shocks induce finite responses in the system.

Assumption 3 - Regressors: The columns of X are exogenous and finite, and X has full column rank.

Assumption 4 - Variance-covariance matrix of the random variables: The elements of α are independent of those in ν , and $\sigma_{\alpha_g \alpha_h}$ as well as $\sigma_{\nu_g \nu_h}$ are finite for all gh .

Assumption 3 is standard also for ordinary least squares, and Assumption 4 ensures that not only deterministic but also stochastic shocks in the system are bounded.

4 Estimation

The maximum likelihood estimator is the leading approach to assess univariate and multivariate processes with binary dependent variables with either cross-section or panel data without cross-sectional dependence. However, its suitability for processes with cross-sectionally interdependent latent variables due to network or spatial interactions is limited. The reason is that the likelihood of the estimator becomes cluttered with the absence of independence of the cross-sectional units.

Alternative procedures exist, like the Bayesian Markov Chain Monte Carlo (MCMC) simulation (see LeSage, 2000, LeSage and Pace, 2009) which has become a leading approach for analyzing interdependent processes for binary variables. However, to date most of the available models assume that the data are cross sectional, and that the process of network or spatial interdependence is of a first-order nature, where only a single process and one unknown parameter characterizes the strength of interdependence among the cross-sectional units.

In this paper, we formulate a multivariate binary choice (probit) model with higher-order network or spatial interdependence in panel data. The key merit of this approach is that it permits the consideration of cross-sectional interdependence in a system of binary-choice dependent variables, while at the same time allowing for a relatively flexible parameterization of the nature and decay of the function governing the strength of interdependence.

4.1 Parameters, priors, and likelihood

All parameters of our model can be summarized by $\theta = \{\delta, \lambda_{gs}, \gamma_{gh}, \tau, y^*, \alpha, \mu_\alpha, V_\alpha\}$ for $g, h \in \{e, m\}$ and $s \in \{1, 2\}$, with $y^* = (y_e^*, y_m^*)'$ and $\delta = (\delta_e', \delta_m')'$. μ_α and V_α are hyperparameters. The joint posterior distribution is given by

$$p(\theta|y, X, W) \propto p(y|y^*, X, W)p(y^*|\delta, \lambda_{gs}, \gamma_{gh}, \tau, \alpha, \mu_\alpha, V_\alpha, X, W) \\ p(\delta)p(\gamma_{gh}|\lambda_{gs})p(\lambda_{gs})p(\tau)p(\alpha|\mu_\alpha, V_\alpha)p(\mu_\alpha)p(V_\alpha),$$

where y denotes the observed binary variables. The expression consists of three parts. The first expression shows the relationship between the latent variable and the observed binary variable. Following the approach suggested by Albert and Chib (1993), the latent variables are treated as parameters, which are estimated. The second term describes the likelihood in terms of the latent variables. The

last line contains the priors. Due to intractability of the joint posterior, the conditional distributions for each element in θ , denoted by θ_ℓ , given the observed data (y, X, W) and the other elements in θ , denoted by $\theta_{-\ell}$, are derived and simulated from. In the remainder of this section, an overline and an underline describe posterior and prior distributions, respectively.

4.2 Prior distributions and likelihood

We rely on uninformative priors, which means that we place little weight on the prior beliefs and more on the data. Using \mathcal{N} , \mathcal{W} , and \mathcal{U} to denote the normal, Wishart and uniform distribution, our priors are as follows:

$$\delta \sim \mathcal{N}(\underline{\delta}, \underline{V}_\delta) \quad \text{where} \quad \underline{\delta} = 0_{k \times 1} \quad \text{and} \quad \underline{V}_\delta = I_k \cdot 1e^{12}, \quad (13)$$

$$\mu_\alpha \sim \mathcal{N}(\underline{\mu}_{\mu\alpha}, \underline{V}_{\mu\alpha}) \quad \text{where} \quad \underline{\mu}_{\mu\alpha} = 0_{2 \times 1} \quad \text{and} \quad \underline{V}_{\mu\alpha} = I_2 \quad (14)$$

$$V_\alpha^{-1} \sim \mathcal{W}(\underline{V}_{V\alpha}^{-1}, \underline{v}_{V\alpha}) \quad \text{where} \quad \underline{V}_{V\alpha}^{-1} = I_2 \quad \text{and} \quad \underline{v}_{V\alpha} = 2. \quad (15)$$

$$p \sim \mathcal{U}(-1, 1), \quad (16)$$

where p summarizes the scalars λ_{gs} , γ_{gh} , and τ . These priors are standard in the literature. We apply a hierarchical structure to estimate $\alpha = (\alpha'_e, \alpha'_m)'$ using the *hyper-parameters* μ_α and V_α . More details will be given below. Utilizing the joint distribution of y^*

$$y^* \sim \mathcal{N}(\mu_{y^*}, \Omega_{y^*}) \quad \text{with} \quad \mu_{y^*} = L^{-1}(X\delta + A\alpha) \quad \Omega_{y^*} = L^{-1}(\Psi \otimes I_n)L^{-1'}, \quad (17)$$

yields the following likelihood

$$p(y^*|\theta, X, W) = \frac{|L|}{(2\pi)^n |\Psi|^{n/2}} \exp \left[-\frac{1}{2} \text{tr} (D\Psi^{-1}) \right],$$

where tr denotes the trace and the $G \times G$ matrix $D = (d_{gh})$ with

$$d_{gh} = (L_g y^* - X_g \delta_g - \dot{A} \alpha_g)' (L_h y^* - X_h \delta_h - \dot{A} \alpha_h),$$

where L_g and L_h are the g th and h th $n \times Gn$ blocks of L .

4.3 Posterior distributions

4.3.1 Posterior of δ

Given the other parameters, the conditional distribution of δ is

$$\delta|\theta_{-\delta} \sim \mathcal{N}(\bar{\delta}, \bar{V}_\delta), \quad (18)$$

where

$$\bar{\delta} = \bar{V}_\delta (X' (\Psi^{-1} \otimes I_n) (Ly^* - A\alpha) + \underline{V}_\delta^{-1} \underline{\delta}) \quad \text{and} \quad \bar{V}_\delta = (X' (\Psi^{-1} \otimes I_n) X + \underline{V}_\delta^{-1})^{-1}.$$

Since this is a standard distribution, Gibbs sampling can be applied.

4.3.2 Posterior of λ_{gs} and γ_{gh}

The conditional distribution for λ_{gs} and γ_{gh} for $g, h \in \{e, m\}$ and $s \in \{1, 2\}$ are as follows:

$$\lambda_{gs}|\theta_{-\lambda_{gs}} \propto |L| \exp \left[-\frac{1}{2} \text{tr} (D\Psi^{-1}) \right] \quad \text{and} \quad \gamma_{gh}|\theta_{-\gamma_{gh}} \propto |L| \exp \left[-\frac{1}{2} \text{tr} (D\Psi^{-1}) \right]. \quad (19)$$

The distributions in (19) do not have a known form. Therefore we utilize a Metropolis Hastings approach, where we draw a proposal candidate and evaluate it and the previous value at the conditional distribution. Based on the acceptance probability, we either take the new proposal candidate or keep the previous value. We draw proposal candidates as in LeSage and Pace (2009) using $\lambda_{gsc} = \lambda_{gs} + c_{\lambda_{gs}} \cdot N(0, 1)$ where λ_{gsc} denotes the proposal candidate, λ_{gs} the previous value, and $c_{\lambda_{gs}}$ a tuning parameter. We only use proposal candidates that are in the admissible parameter range. The tuning parameter $c_{\lambda_{gs}}$ is adapted to ensure an acceptance probability between 40% and 60%.

4.3.3 Posterior of τ

The conditional distribution of τ , which is the off-diagonal element of Ψ is given by

$$\tau|\theta_{-\tau} \propto (1 - \tau^2)^{-n/2} \exp \left[-\frac{1}{2} \text{tr} \left(D \begin{pmatrix} 1 & \tau \\ \tau & 1 \end{pmatrix}^{-1} \right) \right] \quad (20)$$

Since this distribution also takes an unknown form, we draw τ with a Metropolis Hastings approach.

4.3.4 Posterior of α

Since we model α hierarchically, we first draw the hyperparameters μ_α and V_α from the following distributions

$$\mu_\alpha|\theta_{-\alpha} \sim \mathcal{N}(\bar{\mu}_\alpha, \bar{V}_{\mu_\alpha}),$$

with

$$\bar{\mu}_\alpha = \bar{V}_{\mu_\alpha} \left((V_\alpha^{-1} \otimes \iota'_N) \alpha + \underline{V}_{\mu_\alpha}^{-1} \underline{\mu}_\alpha \right) \quad \text{and} \quad \bar{V}_{\mu_\alpha} = \left(NV_\alpha^{-1} + \underline{V}_{\mu_\alpha}^{-1} \right)^{-1},$$

and

$$V_\alpha^{-1}|\theta_{-\alpha} \sim \mathcal{W}(\bar{V}_{V_\alpha}, \bar{v}_{V_\alpha}),$$

with

$$\bar{v}_{V_\alpha} = \underline{v} + N \quad \text{and} \quad \bar{V}_{V_\alpha} = (Q + \underline{V}_{V_\alpha})^{-1},$$

and the $G \times G$ matrix $Q = (q_{gh})$ with $q_{gh} = (\alpha_g - \iota_N \mu_{\alpha g})' (\alpha_h - \iota_N \mu_{\alpha h})$ using $\mu_\alpha = (\mu'_{\alpha 1}, \dots, \mu'_{\alpha G})'$.

Then the $2N \times 1$ vector α is simulated from

$$\alpha|\theta_{-\alpha} \sim \mathcal{N}(\bar{\alpha}, \bar{V}_\alpha),$$

with

$$\bar{\alpha} = \bar{V}_\alpha \left(A' (\Psi^{-1} \otimes I_n) (Ly^* - X\beta) + (V_\alpha^{-1} \otimes I_N) (\mu_\alpha \otimes \iota_N) \right) \quad \text{and} \quad \bar{V}_\alpha = \left(A' (\Psi^{-1} \otimes I_n) A + V_\alpha^{-1} \otimes I_N \right)^{-1}.$$

We apply Gibbs sampling since all three posteriors have standard distributions.

4.3.5 Conditional distribution of y^*

When drawing the latent variables, we use the joint distribution y^* in (17) together with the observed values y . This means we take draws from left-truncated or right-truncated normals depending on whether the observed y is 1 or 0 using the procedure by Geweke (1991).

5 Empirical analysis of export-market and multinational-firm-network participation of firms in Guangdong

5.1 Theoretical background and empirical model specification

The participation in export markets and in global multinational enterprise (MNE) networks are key indicators not only for firms' market success in general but, in the aggregate – for a sector, region, or country – for a firm's degree of participation in the global market. Theoretical research in economics provides guidance regarding the determinants of the participation decision in exports and MNE networks.

The seminal paper by Melitz (2003) derives a specification of the key latent variable behind the export-market participation decision, namely the profitability of exporting, governing a firm's export-participation status. The main determinants of latent export profitability in logs are (i) log firm productivity which is proportional to overall firm size, (ii) log factor costs, (iii) log price-cost mark-ups, and (iv) log variable as well as log fixed costs of exporting. In related work, Helpman, Melitz, and Yeaple (2004) derive the specification of the key latent variable behind the MNE-network participation decision, the profitability of a production unit from being part of an MNE, governing a firm's MNE-participation status. The key determinants of latent profitability in logs from being part of an MNE are (i) log firm productivity which is proportional to overall log firm size, (ii) log factor costs, (iii) log price-cost mark-ups, and (iv) log variable as well as log fixed costs of participating in an MNE. In particular, Helpman, Melitz, and Yeaple (2004) in an appendix elaborate on the fact that exporting and MNE activity – and profitability – may be interdependent, as foreign affiliates may serve as export platforms. Hence, the potential theoretical interdependence between exporting and MNE-network participation – which is an empirical stylized fact – is acknowledged.

However, the aforementioned literature does not consider the fixed costs of either activity as to depend on other firms' profitability from export-market versus MNE-network participation. Yet, in the presence of knowledge spillovers, learning, and mimicking, such interdependence is fundamental and could be empirically relevant.

In light of this literature and the above arguments, the empirical model specification employs two equations and two sets of explanatory variables. The two equations pertain to export-market and MNE-network participation of firm i in year t . We refer to the latent log profitability of exporting of firm i in year t as $y_{e,it}^*$, and to the log profitability of MNE-network participation of firm i in year t as $y_{m,it}^*$. The observable indicator variables pertaining to these latent outcomes will be denoted by $y_{e,it}$

and $y_{m,it}$, respectively.

Variables that appear in both equations are log employment (for firm size), log labor productivity (sales per employee), a second-order polynomial function of log sales over operating profits (capturing log-nonlinear price-mark-up effects in the model of Melitz, 2003), and time-invariant unobservable factors (denoted as α_i in the econometric model above). Control variables that appear only in the export-market-participation equation are log minimum fixed costs (fixed assets) of all exporters in a four-digit sub-sector. Control variables that appear only in the MNE-network-participation equation are log minimum fixed costs (fixed assets) of all firms in an MNE network in a four-digit sub-sector.

Moreover, we include three variables that relate to firm fixed costs – namely a firm’s capital intensity, its intangible assets ratio (intangibles as a share of total assets), and the expenses on management in the total wage bill – as well as two variables that relate to domestic- and global-market access costs – namely an indicator variable for firms that are located in the Pearl River Delta and one for firms that are located in a Special Economic Zone. In general, we would expect exporting and even more so MNE-network participation to be intensive in fixed costs. A location in the Pearl River Delta enables easier access to the Chinese as well as the global market, whereas a location in a Special Economic Zone enables easier access to global export markets. Hence, we would expect the former to be relatively more relevant for non-exporting firms and the latter for exporting firms.

Our key focus is on the cross-sectional dependence of firms’ profitability associated with one or the other decision – exporting or participating in MNE networks – and the interdependence of the profitability of the two decisions within the firm. Therefore, we include the following variables in only the export-participation equation: the latent MNE-network-participation profitability of firm i in year t , and the latent export profitability of other firms – namely the one of firms within a radius of at most 10 miles of firm i in year t ($W_1y_e^*$), and the one of firms within a radius of more than 10 and at most 20 miles of firm i in year t ($W_2y_e^*$). Similarly, we include the following variables in only the MNE-network participation equation: the latent export-participation profitability of firm i in year t , and the latent MNE-network-participation profitability of other firms – namely the one of firms within a radius of at most 10 miles of firm i in year t ($W_1y_m^*$), and the one of firms within a radius of more than 10 to at most 20 miles of firm i in year t ($W_2y_m^*$). The parameters on the spatial lags will be reflective of the degree to which the interdependence between firms decays in space. In an extension, we will consider a specification with an even more fine-grained portrait of the decay in a third-order specification of spatial lags, where W_1 will aggregate the latent profitability of firms within a radius of 5 miles around i , W_2 will aggregate firms within a radius of more than 5 and at most 10 miles around i , and W_3 will aggregate firms within a radius of more than 10 and at most 20 miles around i . The parameters on y_m^* in the model for y_e^* and on y_e^* in the model for y_m^* capture the degree to which the profitability of the two participation decisions are directly interdependent within the firm.

Clearly, this model is a bivariate specification of the above econometric approach with second-order (and, in the extension, with third-order) spatial dependence in the latent dependent variables. Assuming that the disturbances are bivariate normal leads to a set-up which conforms with the model

in Section 3. Specifically, we estimate variants of the following equation system for the two binary dependent variables for export market participation y_e and multinational activity y_m

$$y_e = 1(y_e^* > 0) \quad (21)$$

$$y_m = 1(y_m^* > 0), \quad (22)$$

where the underlying latent variables are modeled as

$$y_e^* = \gamma_{em}y_m^* + \sum_{s=1}^S \lambda_{es}W_s y_e^* + X_e \delta_e + \dot{A}\alpha_e + \nu_e, \quad (23)$$

$$y_m^* = \gamma_{me}y_e^* + \sum_{s=1}^S \lambda_{ms}W_s y_m^* + X_m \delta_m + \dot{A}\alpha_m + \nu_m. \quad (24)$$

In all estimations, we allow for cross-equation correlation between the unobserved error components $E(\alpha_{ei}, \alpha_{mi}) \neq 0$ and $E(\nu_{eit}, \nu_{mit}) \neq 0$. Moreover, we consider second-order, alternatively, third-order spatial processes, so that $S = 2$ in the former and $S = 3$ in the latter case. For comparison, we also estimate a more restrictive model, where $\gamma_{em} = \gamma_{me} = 0$, and $\lambda_{es} = \lambda_{ms} = 0$ for all $s \in \{1, 2, (3)\}$.

5.2 Firm-level census panel data on specialized manufacturers in Guangdong

We employ census-type panel data on export-market and MNE-network participation of producers of specialized and transport facility manufactures in the province of Guangdong. The data come from the National Bureau of Statistics of China (NBS), over the period 2005 to 2007, and cover all firms with total annual sales of at least 5 mn. RMB (about 700,000 US dollars). The number of firms active in that sector varies over the years: 1,361 in 2005, 1,614 in 2006, and 1,436 in 2007. Hence, the size of the cross section has to be indexed by t , N_t , as is accommodated by the model outlined above. The total number of observations, i.e., the number of rows of the stacked vectors y_e and y_m , is given by $n = \sum_{t=2005}^{2007} N_t = 4,411$. We summarize descriptive statistics for all observable right-hand-side variables involved in the empirical analysis in Table 1.

In Guangdong, for the producers in the specialized and transport facility manufactures sector, over the period 2005 to 2007, about 44% of the firms participate in the export market in an average year, and about 39% of the firms participate in MNE networks. The two activities are apparently not independent of each other: of the exporters, about 63% participate in MNE networks; and of the MNEs, 71% participate in exporting. As expected, the minimum (log) fixed costs of participating in an MNE network are somewhat higher (about 8.377) than those of participating in the export market (about 8.158), which is consistent with the assumptions in Helpman, Melitz, and Yeaple (2004). Consistent with economic theory, individual MNE-network participants tend to be smaller in log employment (5.188 versus 5.324) but have a higher log labor productivity (5.360 versus 5.308) than export-market participants, on average.⁶

⁶Economic theory predicts that MNEs as a whole – i.e., when aggregating over all subsidiaries – are larger than exporters, but individual production sites are smaller on average, since non-MNE exporters tend to concentrate production at one location, which is not necessary for MNEs (see Markusen, 2002).

In general, we employ all elements in the exogenous determinants of exporting, X_e and MNE-network participation, X_m , with a one-year lag. In the tables, we will refer to the elements of X_e and X_m by Employment_{it-1} (log firm size), $\text{Productivity}_{it-1}$ (log output per worker), $\text{Competition}_{it-1}$ (log sales over operating profits), $\text{Competition}_{it-1}^2$, $\text{Fixed export costs}_{it-1}$ (log minimum fixed assets of exporters in the same four-digit sub-sector), $\text{Fixed MNE costs}_{it-1}$ (log minimum fixed assets of MNE-network members in the same four-digit sub-sector), $\text{Capital intensity}_{it-1}$ (capital divided by total sales), $\text{Intangible-assets ratio}_{it-1}$ (intangible assets over total assets), $\text{Management fee to wage ratio}_{it-1}$ (the expenses on management as a share of the total wage bill), $\text{Pearl River Delta}_{it-1}$ (a binary indicator variable for shipping-transport-network access costs), and $\text{Special Economic Zone}_{it-1}$ (a binary indicator variable for the location-related tariff-deductability and global-market-access costs).

For modelling the decay of network interdependencies with distance, we consider a second-order and, alternatively, a third-order spatial-lag process. Specifically, we specify the network-weights matrices in a way that they contain normalized inverse-distance entries but only for specific distance brackets. For each pair of firms, we determine the inverse distance between them based on the geographical information about the longitude and latitude associated with their administrative code. Based on this information, we postulate two (with a second-order spatial-lag model) or three (with a second-order spatial-lag model) matrices $W_{s,0}$ with $s \in \{1, 2, 3\}$. Any such matrix has zero diagonal entries. We use $d_1^\#$, $d_2^\#$, and $d_3^\#$ to denote three distances, where $d_1^\# < d_2^\# < d_3^\#$. Let us describe the construction of the matrices W_s for the case of a third-order process with $S = 3$. In that case, the off-diagonal entries in W_1 are zero whenever $d_{ij} \geq d_1^\#$ and they are $(d_{ij})^{-1} \geq 0$ otherwise. Those in W_2 are zero whenever $d_{ij} < d_1^\#$ or $d_{ij} > d_2^\#$ and they are $(d_{ij})^{-1} \geq 0$ otherwise. And the entries in W_3 are zero whenever $d_{ij} < d_2^\#$ or $d_{ij} > d_3^\#$ and $(d_{ij})^{-1} \geq 0$ otherwise.⁷

This approach permits estimating a separate parameter on each of the (two or three) spatial lags of latent outcome. Hence, the decay of network interdependencies may then vary between the two to three considered distance brackets. Overall, the decay function of network interdependencies is parameterized in a more flexible way by this token than when considering a model with just a single spatial lag.

Of course, one would ideally like to specify the decay function even more flexibly in a nonparametric way. However, the use of three spatial lags and associated weights matrices rather than a single one serves as an approximation of the true decay function governing spillovers in space. Allowing for a much more flexible decay function would not be feasible at this point for computational reasons. In order to provide a sensible approximation, we chose the cutoffs separating the rings of neighbors with the preferred third-order model, $(d_1^\#, d_2^\#, d_3^\#)$, as follows. First, we fixed $d_1^\#$ and $d_3^\#$ at a pre-specified level, and then we chose $d_2^\#$ on an integer scale of miles between $d_1^\#$ and $d_3^\#$ so as to maximize the

⁷In the case of a second-order process, we define the off-diagonal entries in W_1 to be zero whenever $d_{ij} \geq d_1^\#$ and to be $(d_{ij})^{-1} \geq 0$ otherwise. Those in W_2 are zero whenever $d_{ij} < d_1^\#$ or $d_{ij} > d_2^\#$ and they are $(d_{ij})^{-1} \geq 0$ otherwise.

correlations between the predicted vectors of choice probabilities jointly between the two equations e and m .⁸

Specifically, we chose $d_1^\# = 5$ miles and $d_3^\# = 20$ miles. Regarding $d_2^\#$ we selected it from the set of integer distance values in miles between 7 and 18 miles, and we identified $d_2^\# = 10$ based on this procedure. Table A.5 in the Online Appendix reports on the correlation coefficients for the cutoff values of $d_2^\# \in \{7, \dots, 18\}$. In principal, one could also choose $d_1^\#$ and $d_2^\#$ simultaneously. However, we refrained from doing so, as the computational burden associated with determining several cutoff values simultaneously even with integer distance cutoffs is significant with the empirical model at hand.

The matrices W_s are obtained by normalizing the entries of $W_{s,0}$ with the respective maximum row sum of the latter.

5.3 Estimation results

In this subsection, we present results based on two different econometric models. One model makes two restrictions: $\gamma_{gh} = 0$ for all $g, h \in \{e, m\}$ and $\lambda_{gs} = 0$ for all $g \in \{e, m\}$ and $s \in \{1, 2, 3\}$. Otherwise, the model conforms with the assumptions in Section 3. Hence, this model is a standard bivariate panel probit model. We present the estimation results for this model based on Markov Chain Monte Carlo simulation in Table 2. A second model relaxes the restrictions on $\gamma_{gh} = 0$ and $\lambda_{gs} = 0$ and fully conforms to the set-up outlined in Section 3. We present the estimation results for the second-order-spatial-lag version of this model in Table 7 and for the third-order-spatial-lag version in Table 8. Each of the three tables is organized in four columns and three vertical blocks. The first column lists the covariate or parameter considered, the second and third columns provide the average and standard deviation of the chain of parameters (which can be interpreted as point estimates and standard errors, respectively), and the fourth column gives the p-value of the Geweke (1992) test. The vertical blocks refer to parameters in the export-profitability equation, the MNE-network-participation-profitability equation, and parameters relating to cross-equation correlation of the remainder disturbances (τ), as well as the systems (γ_{gh}) and the spatial-lag parameters (λ_{gs}).

– Tables 2-4 about here –

The results in Table 2 are largely consistent with the hypotheses from economic theory. Larger firms (in terms of log employment) tend to engage more likely in exporting as well as in MNE networks. Firms, where fixed costs are more important (through a higher capital intensity) tend to be exporters with greater likelihood. Higher specific exporting fixed costs reduces the profitability of exporting, and higher MNE-related fixed costs reduce the profitability of participating in an MNE. For exporting, four of the ten parameters estimated on X_e are statistically significant at conventional levels. Those are the ones on log employment, the linear term in the quadratic competition function, the fixed exporting

⁸More precisely, we determined the predicted binary outcomes obtained from the procedure and consistent with the chosen distance threshold and then computed the correlations between the observed and predicted binary choice indicators. See the Online Appendix for details.

costs term, and the indicator for a location in the Pearl River Delta. For MNE-network participation, all but the parameters on the quadratic Competition term, the Management-fee-to-wage-ratio, and the indicator for a firm's location in a Special Economic Zone are statistically significant at conventional levels.

What stands out in terms of location indicators is that the profitability of both exporting and MNE-network participation is increased for firms which are located in the Pearl River Delta, which permits an easy access to sea transport and other, on-land transport infrastructure. According to the estimate of τ , we could have estimated these two equations independently of each other without any loss of efficiency.

Except for the parameter on productivity, whose sign changes in the profitability-of-exporting equation, the findings in Tables 3 and 4 are qualitatively – in terms of parameter signs – fully aligned with the ones in Table 2. The magnitudes of the (scaled) parameters change to some extent. However, the parameters cannot be interpreted at face value, because of the presence of two spatial lags in both equations underlying Table 3 and three spatial lags in both equations underlying Table 4. This makes the marginal effects of changes in variables on latent outcome heterogeneous among firms and years and different on average from their counterparts in Table 2.

We abstain from a detailed discussion of the results in Table 3, as the parameters in Table 4 suggest that a third-order network- or spatial-lag model should be estimated rather than a second-order one with the data at hand. The most important changes on the parameters on X_e and X_m are that the parameter on the Intangible asset ratio was insignificant in the exporting equation in Table 2 but is significant in Table 4 (as well as in Table 3), and that not only being located in the Pearl River Delta but also being located in a Special Economic Zone raises the profitability of exporting. Interestingly, while the profitability of MNE-network participation appears to depend crucially on the competitive environment, the statistical support for this on exporting is more modest. However, nothing strong can be concluded from the parameter point estimates, as the reduced form even of the latent process of the model is nonlinear in parameters.

Clearly, the estimates of the parameters pertaining to the simultaneous-equations part of the model, $(\gamma_{em}, \gamma_{me})$, the spatial lags, $(\lambda_{es}, \lambda_{ms})$ for $s \in \{1, 2, 3\}$, and the correlation of the cross-equation correlation of the idiosyncratic errors, τ , are of particular interest. It turns out that there is a positive feedback of the profitability of exporting or MNE-network participation on the profitability of the yet other activity. Hence, the impact of covariates, which raise either type of activity is reinforced by the complementarity of the two profitabilities, as $(\hat{\gamma}_{em} > 0, \hat{\gamma}_{me} > 0)$. The corresponding estimates are statistically significantly different from zero at conventional levels. Moreover, except for $\hat{\lambda}_{e3}$, all of the spatial-lag parameters are statistically different from zero at conventional levels, and the magnitude of the point estimates tends to decline as we move from closer to less-close rings of neighbors within each type of profitability, $\{e, m\}$. Finally, $\hat{\tau}$ is statistically different from zero at conventional levels and the point estimate is negative, suggesting that there is a negative correlation of stochastic shocks on the profitability of either type of activity considered. The presence of significant spatial lag terms in

Tables 3-4 means that estimating the empirical model ignoring these spatial effects will lead to biased and misleading results.

The presence of $W_s y_g^*$ (in conjunction or without any presence of y_h^*) on the right-hand side of the model for the g th latent profitability variable leads to a nonlinear form of the reduced form of latent outcome as is indicated by equation (11). Clearly, there is a further nonlinearity involved, when mapping latent outcome, y_g^* , to the observable choice, y_g . However, from a theoretical standpoint both effects of covariates on the profitability of either activity – exporting and MNE-network participation – as well as the choice of the activity are of interest. We will report the magnitude of either type of estimated effects in the subsequent subsection based on the results from the third-order network or spatial model in Table 4.

5.4 Effect estimates on latent continuous profitability outcomes

Equations (11) and (12) suggest that the computation of marginal effects of the elements of the covariates in X involves the estimate of a $2n \times 2n$ inverse, L^{-1} . Let us refer to the gh -th block of L^{-1} by L_{gh}^{-1} and to an $n \times 1$ vector of ones by ι_n so as to define the $n \times 1$ vectors

$$l_{gh}^d = \text{vecdiag}(L_{gh}^{-1}), \quad l_{gh}^o = L_{gh}^{-1} \iota_n - l_{gh}^d, \quad l_{gh}^f = L_{gh}^{-1} \iota_n - l_{gh}^d, \quad (25)$$

$$l_{gh}^t = l_{gh}^d + l_{gh}^f. \quad (26)$$

Note that superscript $\{t\}$ here refers to total effects, whereas superscripts $\{d, o, f\}$ index direct effects and effects on and from other firms (in any year).

Let us use the scalar Δ_{hk} to refer to the change in a generic covariate k – for this consider a one-standard-deviation increase in covariate k that is scaled by the parameter estimate $\hat{\delta}_{hk}$ for each draw of the Markov chain in equation h , where this parameter would be zero if the covariate is absent from that equation.⁹

Accordingly, we can define for each $a \in \{d, o, f, t\}$ the $n \times 1$ vector

$$\hat{d}_{ghk}^a = \hat{l}_{gh}^a \hat{\Delta}_{hk}. \quad (27)$$

for each draw of the parameter chain.

This is the panel systems counterpart to the single-equation and cross-section definition of effect estimates on latent outcomes in spatial models in LeSage and Pace (2009), and it permits a quantitative decomposition of the effects of changes in the determinants of the profitability of export–market or MNE-network participation of firms with panel-data spatial systems. Clearly, as \hat{d}_{ghk}^a depends on the tuple $\{gh\}$ of which there are four as well as on a which can take on four values as well, there are eight effects on the profitability of exporting and MNE-network participation each for every regressor

⁹We report effects of one-standard-deviation changes for all variables, be they continuous or binary. This is done to enable a comparison of their relative scaled importance. Clearly, one could alternatively do experiments by switching on and off binary variables rather than changing them in a way akin to continuous covariates. However, doing so does not permit any straightforward conclusions about the relative importance of the covariates for latent outcomes.

in X_e and X_m . Moreover, \hat{d}_{ghk}^a is an $n \times 1$ vector so that there is a distribution of effect estimates that depends on the location of firm i as well as the time period t (since the number of firms varies over time). We report moments – the minimum, 25th percentile, median, 75th percentile, maximum, and average – of the distribution of effect estimates on the latent profitability of either choice for each variable of interest in an online appendix. We do so, because we view the corresponding results as an intermediate step for determining binary participation outcomes. We focus on the latter in the presentation, here.

5.5 Effect estimates on binary participation outcomes

Besides the effects on latent profitabilities, the effect on the resulting binary outcomes are of interest. Using the estimate of \hat{y}_g^* for each draw in the Markov chain and the previously calculated \hat{d}_{ghk}^a in (27), we can calculate counterfactual binary outcomes \tilde{y}_{ghk}^a to assess the effects of a one standard deviation change of the k th regressor on the respective choice. Clearly, one could do so for each $a \in \{d, o, f, t\}$. We refrain from doing so to economize on space, as responses will be mostly interesting for total effects, where $a = t$, which we will focus on below. Again, we use means and standard deviations from the chain of draws to estimate parameter point estimates and standard errors.

For any generic variable in X , let us define the vectors of estimated counterfactual binary outcome g associated with the total change in latent profitability as

$$\tilde{y}_g^t = 1 \left(\left(\hat{y}_g^* + \sum_{h \in \{e, m\}} \hat{d}_{gh}^t \right) > 0 \right). \quad (28)$$

Moreover, let us define the following estimates of unconditional probabilities:

$$\hat{P}_{g+} = n^{-1} y_g' \iota_n \quad (29)$$

$$\hat{P}_{g-} = n^{-1} (\iota_n - y_g)' \iota_n \quad (30)$$

$$\hat{P}^{\tilde{g}+} = n^{-1} \tilde{y}_g^t' \iota_n, \quad (31)$$

$$\hat{P}^{\tilde{g}-} = n^{-1} (\iota_n - \tilde{y}_g^t)' \iota_n, \quad (32)$$

and analogously for $\{\hat{P}_{h+}, \hat{P}_{h-}, \hat{P}^{\tilde{h}+}, \hat{P}^{\tilde{h}-}\}$.

We can define the following conditional probabilities using similar notation:

$$\hat{P}_{g+}^{\tilde{g}+} = n^{-1} \tilde{y}_g^t' y_g \quad (33)$$

$$\hat{P}_{g+}^{\tilde{g}-} = n^{-1} (\iota_n - \tilde{y}_g^t)' y_g \quad (34)$$

$$\hat{P}_{g-}^{\tilde{g}+} = n^{-1} \tilde{y}_g^t' (\iota_n - y_g) \quad (35)$$

$$\hat{P}_{g-}^{\tilde{g}-} = n^{-1} (\iota_n - \tilde{y}_g^t)' (\iota_n - y_g), \quad (36)$$

where, e.g., $\hat{P}^{\tilde{g}+} = \hat{P}_{g+}^{\tilde{g}+} + \hat{P}_{g-}^{\tilde{g}+}$, and analogous sets of conditional probabilities $\{\hat{P}_{h+}^{\tilde{h}+}, \hat{P}_{h-}^{\tilde{h}+}, \hat{P}_{h+}^{\tilde{h}-}, \hat{P}_{h-}^{\tilde{h}-}\}$, $\{\hat{P}_{h+}^{\tilde{g}+}, \hat{P}_{h+}^{\tilde{g}-}, \hat{P}_{h-}^{\tilde{g}+}, \hat{P}_{h-}^{\tilde{g}-}\}$, and $\{\hat{P}_{g+}^{\tilde{h}+}, \hat{P}_{g+}^{\tilde{h}-}, \hat{P}_{g-}^{\tilde{h}+}, \hat{P}_{g-}^{\tilde{h}-}\}$ are straightforwardly defined based on this notation.

We can also define another set of conditional probabilities to compute the entries of a full “transition matrix” between the events in the benchmark and a counterfactual situation as

$$\widehat{P}_{g_+,h_+}^{\tilde{g}_+,\tilde{h}_+} = n^{-1}(\tilde{y}_g^t \circ \tilde{y}_h^t)'(y_g \circ y_h),$$

with all other permutations being straightforwardly defined using the above notation. Clearly, the latter suggests that $\widehat{P}_{g_+}^{\tilde{g}_+} = \widehat{P}_{g_+,h_+}^{\tilde{g}_+,\tilde{h}_+} + \widehat{P}_{g_+,h_+}^{\tilde{g}_+,\tilde{h}_-} + \widehat{P}_{g_+,h_-}^{\tilde{g}_+,\tilde{h}_+} + \widehat{P}_{g_+,h_-}^{\tilde{g}_+,\tilde{h}_-}$.

Based on those definitions, unconditional as well as conditional probabilities of exporting or MNE-network participation – as well as their changes – are clearly defined. With a suitable series of Markov Chain Monte Carlo draws for parameters and benchmark as well as counterfactual (estimated) binary outcomes, we can determine a series of draws of the aforementioned propensities – and, hence, first and second moments thereof. We will report on those moments in this subsection for the case where $\{g, h\} = \{e, m\}$

Tables A.6 and A.7 suggest some sizable effects of one-standard-deviation changes in the explanatory variables on the latent profitability of the choices considered. However, whether these changes are large enough for a sizable portion of firms to actually change the probability of choosing to export or participate in an MNE network depends on the shape of the empirical distribution of the respective profitabilities and their conditional mean. Clearly, this is evident from the presence of benchmark latent-outcome estimates, \widehat{y}_g^* , in the expression for counterfactual outcome, \tilde{y}_g^t , in (27).

– Table 5 about here –

In Table 5, we report unconditional changes of exporting and MNE-network participation. For this, let us define any generic change in probabilities of taking on activity $g \in \{e, m\}$ in response to the changes $(\widehat{\Delta}_{ek}, \widehat{\Delta}_{mk})$ as

$$\Delta \widehat{P}_{g_+} = \widehat{P}_{g_+}^{\tilde{g}_+} - \widehat{P}_{g_+}.$$

In the table, we report point estimates as well as standard errors based on the Markov chain of draws for both $\Delta \widehat{P}_{e_+}$ and $\Delta \widehat{P}_{m_+}$ based on the non-spatial-bivariate-panel-probit-model parameters in the left block and on the third-order spatial-bivariate-panel-probit-model parameters in the right block.

The results in the table indicate that changes in probability tend to be significantly smaller with the non-spatial than with the third-order spatial-lag model. Moreover, the scaling of $(\widehat{\Delta}_{ek}, \widehat{\Delta}_{mk})$ by considering one-standard-deviation changes throughout permits insights into the relative importance of the individual covariates for changes in activity. The results for the spatial model suggest that firm size (Employment) and the location in the Pearl River Delta are the relatively most important drivers of both exporting and MNE-network participation in Guangdong province. However, Competition and Fixed export costs as well as Fixed MNE costs are also relatively important in deterring these activities.

We summarize the entries in the “transition matrix” in Table 6 – as, e.g., $\widehat{P}_{g_+,h_+}^{\tilde{g}_+,\tilde{h}_+}$ with $g, h \in \{e, m\}$ – for the non-spatial bivariate panel probit model and for the third-order spatial-lag model, again providing point estimates and standard errors based on the Markov chain of draws.

– Table 6 about here –

In general, the means in a row add up to unity in this matrix. The entries in the table suggest that the propensity of not changing any activity is generally relatively high, as can be seen from the values on the diagonal of the matrix, which are close to unity. Moreover, with those constellations where both changes $\widehat{\Delta}_{ek} > 0$ and $\widehat{\Delta}_{mk} > 0$, the means for (e_+, m_+) , (e_+, m_-) , and (e_-, m_+) in the column denoted $(\tilde{e}_+, \tilde{m}_+)$ tend to be higher than (e_-, m_-) . Hence, firms tend to switch to doing both at higher likelihood in these cases. We use colors to highlight quantitatively large changes. We chose red to denote large marginal effects which led to changes in predicted state choices and blue to denote large marginal effects which did not lead to changes in choices (indicating stayers). In line with the previous findings, the main shifters of exports and MNE participation, e.g. employment, location in the Pearl River Delta, fixed costs, and capital intensity induce the quantitatively the largest changes. Among the status switchers, the majority of them switches to doing both exporting and MNE participation from some other state. As higher associated fixed costs reduce the profitability of participation, increasing them for exporting or MNE activity lead to a reduction of those activities.

6 Conclusions

This paper proposes an estimation approach which accommodates the following features of panel data. First, there are several binary outcomes, where the latent processes are interdependent between outcomes as is the case with simultaneous systems of equations. Second, the latent outcomes, and, hence, the resulting binary outcomes, are generated in a dependent way between cross-sectional units. The latter is motivated as to flow from spatial or network externalities between the cross-sectional units. Third, the degree of interdependence cannot be parameterized fully ex ante by a single network- or spatial-interaction weights matrix. In this regard, the paper proposes a higher-order modelling approach, where an arbitrary, fixed number of such interdependence matrices can be included and used to parameterize the decay of interdependence in some space, be it geography, social interaction, or other observable dimensions. The decay of interdependence can then be estimated in terms of a set of parameters that is of arbitrary, fixed size that is sufficiently smaller than the sample size.

The paper applies this model to data on firms in a particular, high-technology sector, namely specialized and transport facilities manufacturing, in Guangdong province in China, which are observed over the years 2004-2007. What the papers aims at answering with this data-set is the question to which extent the participation of firms in export markets and in multinational-firm networks depend on each other. Also, to what extent these two choices are made in an interdependent way among close and somewhat less-close geographically neighboring firms. This analysis brought to light that

the participation in an MNE network makes exporting more likely, while there is no statistically significant support for the opposite. Moreover, the analysis uncovered that these choices are made indeed in an interdependent way, and that cross-firm spillovers decline more rapidly in the close neighborhood than in the less-close neighborhood.

We envisage two directions for future research. First, from a methodological point of view, the modelling of patterns in the disturbances could be interesting. With panel data, such patterns could relate to the time-series or the cross-sectional correlation of the disturbances. To the extent that at least two components could be distinguished in the disturbances, a host of different patterns of cross-sectional dependence could be allowed for. Moreover, a host of patterns of serial correlation could be considered with regards to the double-indexed error component. Second, from an applied point of view, it would be interesting to apply such models for systems of binary dependent variables at different levels of aggregation. The latter could help shed light on novel patterns in the data that are not typically addressed in economic theory motivating structural models of individuals', households', firms', micro-regional jurisdictions', country governments', or supranational organizations' choices. We mentioned some specific examples of applications that could be interesting in this context, but this list is by no means exhaustive.

References

- Albert, J.H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 88(422), 669–679.
- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Boston, USA: Kluwer Academic Publishers.
- Arduini, T. (2016). Distribution free estimation of spatial autoregressive binary choice panel data models. Quaderni - Working Paper DSE No. 1052.
- Badinger, H. and Egger, P.H. (2011). GM estimation of higher-order spatial autoregressive processes in cross-section models with heteroskedastic disturbances. *Papers in Regional Science* 90, 213–235.
- Badinger, H. and Egger, P.H. (2013). Estimation and testing of higher-order spatial autoregressive panel data error component models. *Journal of Geographical Systems* 15, 453–489.
- Badinger, H. and Egger, P.H. (2015). Fixed effects and random effects estimation of higher-order spatial autoregressive models with spatial autoregressive and heteroscedastic disturbances. *Spatial Economic Analysis* 10, 11–35.
- Baltagi, B.H. and Bresson, G. (2011). Maximum likelihood estimation and lagrange multiplier tests for panel seemingly unrelated regressions with spatial lag and spatial errors: an application to hedonic housing prices in Paris. *Journal of Urban Economics* 69, 24–42.

- Baltagi, B.H. and Deng, Y. (2015). EC3SLS estimator for a simultaneous system of spatial autoregressive equations with random effects. *Econometric Reviews* 34, 659–694.
- Baltagi, B.H., Ding, S., and Egger, P.H., (2020). A panel data model with generalized higher-order network Effects. In: Chudik, A., Hsiao, C., and Timmermann, A. (eds.), *Festschrift in Honor of Hashem Pesaran (Advances in Econometrics)*. Emerald Group Publishing Limited, Bingley, UK, forthcoming.
- Baltagi, B.H., Egger, P.H., and Kesina, M. (2016). Bayesian spatial bivariate panel probit estimation. In: Baltagi, B.H., LeSage, J.P., Pace, R.K. (eds.), *Spatial Econometrics: Qualitative and Limited Dependent Variables (Advances in Econometrics, Volume 37)*. Emerald Group Publishing Limited, Bingley, UK, 119–144.
- Baltagi, B.H., Egger, P.H., and Kesina, M. (2017). Determinants of firm-level domestic sales and exports with spillovers: evidence from China. *Journal of Econometrics* 199(2), 184–201.
- Baltagi, B.H., Egger, P.H., and Kesina, M. (2019). Contagious exporting and foreign ownership: evidence from firms in Shanghai using a Bayesian spatial bivariate probit model. *Regional Science and Urban Economics* 76, 125–146.
- Baltagi, B.H. and Pirotte, A. (2011). Seemingly unrelated regressions with spatial error components. *Empirical Economics* 40, 5–49.
- Bernard, A.B., Wagner, J. (2001). Export entry and exit by German firms. *Weltwirtschaftliches Archiv* 137(1), 105–123.
- Beron, K.J. and Vijverberg, W.P. (2004). Probit in a spatial context: a Monte Carlo approach. In: Anselin, L., Florax, R.J.G.M., and Rey, S.J. (eds.), *Advances in Spatial Econometrics. Methodology, Tools and Applications*. Springer, Berlin, Germany, 169–195.
- Blonigen B.A., L. Fontagné, N. Sly, and F. Toubal (2014). Cherries for sale: the incidence and timing of cross-border M&A. *Journal of International Economics* 94(2), 341–357.
- Carrión-Flores, C.E., Flores-Lagunes, A., and Guci, L. (2009). Land use change: a spatial multinomial choice analysis. Working paper.
- Cohen-Cole, E., Liu, X., and Zenou, Y. (2018). Multivariate choice and identification of social interactions. *Journal of Applied Econometrics* 33(2), 165–178.
- Cliff, A. and Ord, J. (1973). *Spatial Autocorrelation*. London, UK: Pion.
- Cliff, A. and Ord, J. (1981). *Spatial Processes, Models and Applications*. London, UK: Pion.
- Debarsy, N. and LeSage, J.P. (2020). Bayesian Model Averaging for Spatial Autoregressive Models Based on Convex Combinations of Different Types of Connectivity Matrices. *Journal of Business & Economic Statistics* DOI: 10.1080/07350015.2020.1840993.

- Devereux, M., Lockwood, B., and Redoano, M. (2008). Do Countries Compete over Corporate Tax Rates? *Journal of Public Economics* 92(5), 1210–1235.
- Drukker, D.M., Egger, P.H., and Prucha, I.R. (2021). Simultaneous equations models with higher-order spatial or social network interactions. University of Maryland.
- Geweke, J. (1991). Efficient simulation from the multivariate normal and Student t distributions subject to linear constraints. In: Keramidas, E. (ed.), *Computer Science and Statistics Proceedings of the Twenty-Third Symposium on the Interface*, 571–578.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (eds.), *Bayesian Statistics 4*. Clarendon Press, Oxford, 641–649.
- Gupta, A. and Robinson, P.M. (2015). Inference on higher-order spatial autoregressive models with increasingly many parameters. *Journal of Econometrics* 186, 19–31.
- Han, X., Hsieh, C.S., and Lee, L.-F. (2017). Estimation and model selection of higher-order spatial autoregressive model: An efficient Bayesian approach. *Regional Science and Urban Economics* 63, 97–120.
- Helpman, E., Melitz, M.J., and Yeaple, S.R. (2004). Export versus FDI with heterogeneous firms. *American Economic Review* 94, 300–316.
- Kapoor, M., Kelejian, H.H., and Prucha, I.R. (2007). Panel data models with spatially correlated error components. *Journal of Econometrics* 140, 97–130.
- Kelejian, H.H. and Prucha, I.R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics* 17, 99–121.
- Kelejian, H.H. and Prucha, I.R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40, 509–533.
- Kelejian, H.H. and Prucha, I.R. (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations. *Journal of Econometrics* 118, 27–50.
- Kelejian, H.H. and Prucha, I.R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157, 53–67.
- Klier, T. and McMillen, D.P. (2008). Clustering of auto supplier plants in the United States. *Journal of Business & Economic Statistics* 26, 460–471.
- Lei, J. (2013). Smoothed spatial maximum score estimation of spatial autoregressive binary choice panel models. CentER Discussion Paper Series No. 2013-061.

- Lee, L.-F. (2003). Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews* 22, 307–335.
- Lee, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72, 1899–1925.
- Lee, L.-F. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* 137, 489–514.
- Lee, L.-F. and Yu, J. (2010a). Some recent developments in spatial panel data models. *Regional Science and Urban Economics* 40, 255–271.
- Lee, L.-F. and Yu, J. (2010b). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics* 154, 165–185.
- Lee, L.-F. and Yu, J. (2012) Spatial panels: random components versus fixed effects. *International Economic Review* 53, 1369–1412.
- Lee, L.-F. and Liu, X. (2010). Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. *Econometric Theory* 26, 187–230.
- LeSage, J.P. (2000). Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis* 32, 19–35.
- LeSage, J.P. and Pace, R.K. (2009). *Introduction to Spatial Econometrics*. CRC Press/Taylor & Francis, Boca Raton.
- Lin, X. and Lee, L.-F. (2010). GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics* 157, 34–52.
- Liu, X. (2014). Identification and Efficient Estimation of Simultaneous Equations Network Models, *Journal of Business & Economic Statistics* 32, 516–536.
- Markusen, J.R. (2002). *Multinational Firms and the Theory of International Trade*. MIT Press. Cambridge, MA.
- McMillen, D.P. (1992). Probit with spatial autocorrelation. *Journal of Regional Science* 32, 335–348.
- Melitz, M.J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica* 71, 1695–1725.
- Pinkse, J. and Slade, M.E. (1998). Contracting in space: an application of spatial statistics to discrete-choice models. *Journal of Econometrics* 85, 125–154.
- Pinkse, J., Slade, M.E., and Shen, L. (2006) Dynamic spatial discrete choice using one-step GMM: an application to mine operating decisions, *Spatial Economic Analysis* 1, 53–99.

- Wang, H., Iglesias, E.M., and Wooldridge, J.M. (2013). Partial maximum likelihood estimation of spatial probit models. *Journal of Econometrics* 172, 77–89.
- Wang, X. and Kockelmann K. M. (2009). Bayesian inference for ordered response data with a dynamic spatial-ordered probit model. *Journal of Regional Science* 49, 877–913.
- Yang, K. and Lee, L.-F. (2017). Identification and QML estimation of multivariate and simultaneous equations spatial autoregressive models. *Journal of Econometrics* 196, 196–214.
- Zhou, Y., Wang, X., and Holguin-Veras, J. (2016). Discrete choice with spatial correlation: A spatial autoregressive binary probit model with endogenous weight matrix (SARBP-EWM). *Transportation Research Part B: Methodological* 94, 440-455.

Tables

Table 1: Descriptive statistics

	All firms		Exporters		MNEs	
	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.
Exporting (binary)	0.437	0.496	1	0	0.714	0.452
MNE (binary)	0.386	0.487	0.631	0.483	1	0
Employment (in logs)	4.905	1.033	5.324	1.046	5.188	1.068
Productivity (in logs)	5.322	0.896	5.308	0.869	5.360	0.930
Competition (in logs)	2.987	0.760	2.964	0.739	2.935	0.711
Fixed export costs (in logs)	8.158	0.899	8.058	0.887	–	–
Fixed MNE costs (in logs)	8.377	0.904	–	–	8.210	0.828
Capital intensity (in logs)	5.140	0.990	5.204	0.988	5.333	1.040
Intangible asset ratio	0.025	0.069	0.024	0.055	0.020	0.046
Management-fee-to-wage-ratio	1.068	1.110	1.012	1.013	1.114	1.327
Pearl River Delta (binary)	0.871	0.335	0.910	0.286	0.941	0.235
Special Economic Zone (binary)	0.129	0.336	0.138	0.345	0.108	0.311

Notes: All regressors are one period lagged. The total number of observations is 4,411. The total number of exporters is 1,927, while the total number of MNEs is 1,703.

Table 2: Exporting and MNE participation - bivariate (non-spatial) panel probit results

	Mean	Std.err.	GT p-value
X_e			
Employment (in logs)	1.036	0.084	0.861
Productivity (in logs)	-0.039	0.088	0.412
Competition (in logs)	-0.422	0.230	0.265
Competition ²	0.035	0.025	0.394
Fixed export costs (in logs)	-0.329	0.068	0.905
Capital intensity (in logs)	0.135	0.093	0.659
Intangible asset ratio	1.188	1.052	0.107
Management-fee-to-wage-ratio	0.034	0.050	0.185
Pearl River Delta (binary)	0.910	0.283	0.585
Special Economic Zone (binary)	0.395	0.304	0.366
X_m			
Employment (in logs)	0.771	0.108	0.264
Productivity (in logs)	-0.267	0.125	0.973
Competition (in logs)	-0.650	0.292	0.509
Competition ²	0.055	0.034	0.653
Fixed MNE costs (in logs)	-0.878	0.100	0.746
Capital intensity (in logs)	0.884	0.137	0.314
Intangible asset ratio	-3.185	1.686	0.819
Management-fee-to-wage-ratio	0.063	0.077	0.489
Pearl River Delta (binary)	2.652	0.493	0.139
Special Economic Zone (binary)	-0.577	0.444	0.651
τ	0.056	0.114	0.111
m_{α_e}	-3.503	0.794	0.467
m_{α_m}	-2.101	0.575	0.303
$V_{\alpha_e \alpha_e}$	10.284	1.218	0.579
$V_{\alpha_e \alpha_m}$	9.359	0.888	0.105
$V_{\alpha_m \alpha_m}$	22.830	2.649	0.145

Notes: Year dummies are included. All regressors are one period lagged. We ran the MCMC procedure 26,000 times and applied a burn-in of 5,500 and a thinning of 2. The last column contains the p-value of the Geweke (1992) test.

Table 3: Exporting and MNE participation - second-order-spatial-lag simultaneous-equations bivariate panel probit results

	Mean	Std.err.	GT p-value
<hr/>			
X_e			
Employment (in logs)	0.914	0.093	0.462
Productivity (in logs)	0.102	0.099	0.828
Competition (in logs)	-0.287	0.249	0.549
Competition ²	0.022	0.028	0.568
Fixed export costs (in logs)	-0.173	0.069	0.442
Capital intensity (in logs)	-0.083	0.106	1.000
Intangible asset ratio	2.375	1.170	0.405
Management-fee-to-wage-ratio	0.011	0.055	0.429
Pearl River Delta (binary)	0.273	0.298	0.342
Special Economic Zone (binary)	0.580	0.286	0.517
<hr/>			
X_m			
Employment (in logs)	0.574	0.134	0.106
Productivity (in logs)	-0.246	0.131	0.433
Competition (in logs)	-0.671	0.337	0.959
Competition ²	0.055	0.038	0.976
Fixed MNE costs (in logs)	-0.764	0.111	0.575
Capital intensity (in logs)	0.889	0.142	0.172
Intangible asset ratio	-3.755	1.565	0.165
Management-fee-to-wage-ratio	0.095	0.084	0.653
Pearl River Delta (binary)	2.674	0.517	0.325
Special Economic Zone (binary)	-0.719	0.451	0.101
<hr/>			
γ_{em}	0.269	0.067	0.465
γ_{me}	0.265	0.092	0.934
λ_{e1}	0.635	0.109	0.948
λ_{e2}	0.095	0.170	0.593
λ_{m1}	0.530	0.110	0.236
λ_{m2}	0.378	0.120	0.573
τ	-0.256	0.111	0.373
<hr/>			
m_{α_e}	-3.059	0.586	0.989
m_{α_m}	-1.845	0.762	0.116
$V_{\alpha_e\alpha_e}$	7.846	1.003	0.547
$V_{\alpha_e\alpha_m}$	0.431	1.447	0.490
$V_{\alpha_m\alpha_m}$	19.931	3.131	0.118
<hr/>			
Notes: Year dummies are included. All regressors are one period lagged. We ran the MCMC procedure 38,000 times and applied a burn-in of 9,000 and a thinning of 2. The last column contains the p-value of the Geweke (1992) test.			
<hr/>			

Table 4: Exporting and MNE participation - third-order-spatial-lag simultaneous-equations bivariate panel probit results

	Mean	Std.err.	GT p-value
X_e			
Employment (in logs)	0.968	0.085	0.117
Productivity (in logs)	0.075	0.098	0.957
Competition (in logs)	-0.316	0.227	0.711
Competition ²	0.026	0.025	0.565
Fixed export costs (in logs)	-0.206	0.068	0.687
Capital intensity (in logs)	-0.019	0.099	0.192
Intangible asset ratio	1.972	1.123	0.570
Management-fee-to-wage-ratio	0.024	0.052	0.327
Pearl River Delta (binary)	0.472	0.264	0.535
Special Economic Zone (binary)	0.517	0.263	0.599
X_m			
Employment (in logs)	0.442	0.118	0.129
Productivity (in logs)	-0.228	0.123	0.386
Competition (in logs)	-0.537	0.302	0.399
Competition ²	0.042	0.036	0.442
Fixed MNE costs (in logs)	-0.753	0.093	0.714
Capital intensity (in logs)	0.892	0.138	0.184
Intangible asset ratio	-3.771	1.523	0.377
Management-fee-to-wage-ratio	0.087	0.079	0.753
Pearl River Delta (binary)	2.555	0.462	0.142
Special Economic Zone (binary)	-0.705	0.415	0.611
γ_{em}	0.191	0.036	0.114
γ_{me}	0.390	0.039	0.178
λ_{e1}	0.581	0.098	0.105
λ_{e2}	0.225	0.119	0.335
λ_{e3}	0.085	0.112	0.172
λ_{m1}	0.409	0.091	0.163
λ_{m2}	0.285	0.144	0.973
λ_{m3}	0.239	0.134	0.664
τ	-0.307	0.094	0.866
m_{α_e}	-3.482	0.417	0.416
m_{α_m}	-1.733	0.364	0.118
$V_{\alpha_e\alpha_e}$	8.030	1.049	0.116
$V_{\alpha_e\alpha_m}$	1.286	0.699	0.783
$V_{\alpha_m\alpha_m}$	18.951	2.377	0.142

Notes: Year dummies are included. All regressors are one period lagged. We ran the MCMC procedure 43,000 times and applied a burn-in of 28,000 and a thinning of 2. The last column contains the p-value of the Geweke (1992) test.

Table 5: Effects of a one-std.dev. increase in explanatory variables on exporting and MNE probabilities in the bivariate (non-spatial) panel probit and the third-order spatial lag simultaneous equations bivariate panel probit

		Non-spatial		Spatial	
		$\Delta \hat{P}_{e+}^d$	$\Delta \hat{P}_{m+}^d$	$\Delta \hat{P}_{e+}^t$	$\Delta \hat{P}_{m+}^t$
Employment	Mean	0.1081	0.0520	0.1493	0.0822
	Std.err.	0.0098	0.0079	0.0125	0.0113
Productivity	Mean	-0.0035	-0.0157	0.0018	-0.0163
	Std.err.	0.0079	0.0076	0.0112	0.0100
Competition	Mean	-0.0159	-0.0159	-0.0243	-0.0264
	Std.err.	0.0078	0.0065	0.0098	0.0088
Fixed export costs	Mean	-0.0292		-0.0252	-0.0088
	Std.err.	0.0064		0.0085	0.0036
Fixed MNE costs	Mean		-0.0523	-0.0242	-0.0642
	Std.err.		0.0065	0.0062	0.0086
Capital intensity	Mean	0.0134	0.0572	0.0282	0.0772
	Std.err.	0.0095	0.0095	0.0126	0.0120
Intangible asset ratio	Mean	0.0081	-0.0143	0.0093	-0.0182
	Std.err.	0.0073	0.0078	0.0097	0.0096
Management-fee-to-wage -ratio	Mean	0.0038	0.0046	0.0071	0.0102
	Std.err.	0.0057	0.0057	0.0073	0.0075
Pearl River Delta	Mean	0.0306	0.0581	0.0515	0.0832
	Std.err.	0.0098	0.0112	0.0128	0.0145
Special Economic Zone	Mean	0.0133	-0.0127	0.0153	-0.0144
	Std.err.	0.0104	0.0099	0.0126	0.0135

Table 6: Transition matrix: Effects of a one-std.dev. increase in the covariates on exporting and MNE status for the bivariate non-spatial panel probit and the third-order spatial lag simultaneous equations bivariate panel probit

		Bivariate non-spatial panel probit				Third-order spatial lag simultaneous equations bivariate panel probit				
			$(\tilde{e}_+, \tilde{m}_+)$	$(\tilde{e}_+, \tilde{m}_-)$	$(\tilde{e}_-, \tilde{m}_+)$	$(\tilde{e}_-, \tilde{m}_-)$	$(\tilde{e}_+, \tilde{m}_+)$	$(\tilde{e}_+, \tilde{m}_-)$	$(\tilde{e}_-, \tilde{m}_+)$	$(\tilde{e}_-, \tilde{m}_-)$
Employment	(e_+, m_+)	Mean	1.0000	0	0	0	1.0000	0	0	0
		Std.err.	0	0	0	0	0	0	0	0
	(e_+, m_-)	Mean	0.1688	0.8312	0	0	0.2548	0.7452	0	0
		Std.err.	0.0263	0.0263	0	0	0.0325	0.0325	0	0
	(e_-, m_+)	Mean	0.3507	0	0.6493	0	0.4749	0	0.5251	0
		Std.err.	0.0332	0	0.0332	0	0.0371	0	0.0371	0
	(e_-, m_-)	Mean	0.0156	0.1376	0.0392	0.8076	0.0372	0.1767	0.0537	0.7324
		Std.err.	0.0039	0.0143	0.0073	0.0181	0.0083	0.0167	0.0094	0.0235
Productivity	(e_+, m_+)	Mean	0.9599	0.0323	0.0074	0.0004	0.9603	0.0340	0.0052	0.0004
		Std.err.	0.0190	0.0158	0.0085	0.0008	0.0242	0.0200	0.0086	0.0010
	(e_+, m_-)	Mean	0.0001	0.9815	0.0000	0.0184	0.0007	0.9867	0.0000	0.0126
		Std.err.	0.0013	0.0208	0.0000	0.0208	0.0043	0.0207	0.0001	0.0207
	(e_-, m_+)	Mean	0.0058	0.0003	0.9338	0.0601	0.0191	0.0008	0.9170	0.0631
		Std.err.	0.0122	0.0010	0.0308	0.0296	0.0262	0.0018	0.0377	0.0374
	(e_-, m_-)	Mean	0.0000	0.0022	0.0000	0.9977	0.0000	0.0070	0.0002	0.9928
		Std.err.	0.0000	0.0047	0.0004	0.0047	0.0001	0.0095	0.0011	0.0097
Competition	(e_+, m_+)	Mean	0.9434	0.0319	0.0234	0.0013	0.9095	0.0516	0.0350	0.0039
		Std.err.	0.0185	0.0135	0.0120	0.0013	0.0273	0.0177	0.0146	0.0030
	(e_+, m_-)	Mean	0.0000	0.9432	0.0000	0.0568	0.0000	0.9155	0.0000	0.0845
		Std.err.	0.0011	0.0276	0.0001	0.0276	0.0006	0.0338	0.0000	0.0338
	(e_-, m_+)	Mean	0.0002	0.0000	0.9385	0.0613	0.0000	0.0000	0.9000	0.1000
		Std.err.	0.0018	0.0001	0.0258	0.0258	0.0009	0.0001	0.0341	0.0342
	(e_-, m_-)	Mean	0	0.0001	0.0000	0.9999	0	0.0000	0.0000	1.0000
		Std.err.	0	0.0007	0.0003	0.0008	0	0.0003	0.0002	0.0003
Fixed export costs	(e_+, m_+)	Mean	0.9543	0	0.0457	0	0.9432	0.0170	0.0383	0.0015
		Std.err.	0.0112	0	0.0112	0	0.0202	0.0071	0.0136	0.0016
	(e_+, m_-)	Mean	0	0.8970	0	0.1030	0.0000	0.9116	0	0.0884
		Std.err.	0	0.0233	0	0.0233	0.0001	0.0298	0	0.0298
	(e_-, m_+)	Mean	0	0	1.0000	0	0.0000	0	0.9667	0.0333
		Std.err.	0	0	0	0	0.0005	0	0.0148	0.0149
	(e_-, m_-)	Mean	0	0	0	1.0000	0	0.0000	0.0000	1.0000
		Std.err.	0	0	0	0	0	0.0001	0.0000	0.0002
Fixed MNE costs	(e_+, m_+)	Mean	0.8893	0.1107	0	0	0.8330	0.1276	0.0299	0.0094
		Std.err.	0.0153	0.0153	0	0	0.0247	0.0178	0.0082	0.0045
	(e_+, m_-)	Mean	0	1.0000	0	0	0	0.9172	0	0.0828
		Std.err.	0	0	0	0	0	0.0215	0	0.0215
	(e_-, m_+)	Mean	0	0	0.8031	0.1969	0	0	0.7609	0.2391
		Std.err.	0	0	0.0278	0.0278	0	0	0.0342	0.0342
	(e_-, m_-)	Mean	0	0	0	1.0000	0	0	0	1.0000
		Std.err.	0	0	0	0	0	0	0	0
Capital intensity	(e_+, m_+)	Mean	0.9996	0	0.0004	0	0.9999	0	0.0001	0
		Std.err.	0.0021	0	0.0021	0	0.0008	0	0.0008	0
	(e_+, m_-)	Mean	0.1837	0.8153	0.0001	0.0009	0.2421	0.7576	0.0000	0.0002
		Std.err.	0.0298	0.0297	0.0008	0.0044	0.0348	0.0347	0.0004	0.0018
	(e_-, m_+)	Mean	0.0493	0	0.9507	0	0.1076	0	0.8924	0
		Std.err.	0.0328	0	0.0328	0	0.0463	0	0.0463	0
	(e_-, m_-)	Mean	0.0024	0.0159	0.0584	0.9233	0.0072	0.0290	0.0771	0.8867
		Std.err.	0.0019	0.0106	0.0112	0.0171	0.0041	0.0134	0.0138	0.0234
Intangible asset ratio	(e_+, m_+)	Mean	0.9694	0.0299	0.0007	0.0000	0.9605	0.0380	0.0014	0.0002
		Std.err.	0.0169	0.0164	0.0026	0.0002	0.0215	0.0200	0.0037	0.0006
	(e_+, m_-)	Mean	0.0002	0.9981	0	0.0017	0.0004	0.9964	0	0.0032
		Std.err.	0.0017	0.0065	0	0.0064	0.0025	0.0093	0	0.0091
	(e_-, m_+)	Mean	0.0297	0.0013	0.9150	0.0539	0.0360	0.0020	0.8932	0.0688
		Std.err.	0.0233	0.0021	0.0342	0.0297	0.0303	0.0027	0.0387	0.0361
	(e_-, m_-)	Mean	0.0000	0.0114	0.0001	0.9885	0.0000	0.0133	0.0001	0.9866
		Std.err.	0.0000	0.0088	0.0005	0.0089	0.0001	0.0110	0.0005	0.0111

Continued on next page

Table 6: Transition matrix: Effects of a one-std.dev. increase in the covariates on exporting and MNE status for the bivariate non-spatial panel probit and the third-order spatial lag simultaneous equations bivariate panel probit (cont.)

		Bivariate non-spatial panel probit				Third-order spatial lag simultaneous equations bivariate panel probit				
			$(\tilde{e}_+, \tilde{m}_+)$	$(\tilde{e}_+, \tilde{m}_-)$	$(\tilde{e}_-, \tilde{m}_+)$	$(\tilde{e}_-, \tilde{m}_-)$	$(\tilde{e}_+, \tilde{m}_+)$	$(\tilde{e}_+, \tilde{m}_-)$	$(\tilde{e}_-, \tilde{m}_+)$	$(\tilde{e}_-, \tilde{m}_-)$
Management-fee-to-wage ratio	(e_+, m_+)	Mean	0.9975	0.0013	0.0012	0.0000	0.9984	0.0006	0.0009	0.0000
		Std.err.	0.0049	0.0036	0.0032	0.0001	0.0044	0.0028	0.0030	0.0000
	(e_+, m_-)	Mean	0.0180	0.9790	0.0000	0.0029	0.0373	0.9605	0.0000	0.0022
		Std.err.	0.0166	0.0175	0.0002	0.0073	0.0245	0.0244	0.0003	0.0068
	(e_-, m_+)	Mean	0.0166	0.0000	0.9810	0.0024	0.0300	0.0000	0.9687	0.0013
		Std.err.	0.0170	0.0003	0.0178	0.0068	0.0247	0.0002	0.0247	0.0058
	(e_-, m_-)	Mean	0.0001	0.0060	0.0051	0.9889	0.0003	0.0096	0.0097	0.9805
		Std.err.	0.0002	0.0061	0.0048	0.0082	0.0005	0.0080	0.0067	0.0117
Pearl River Delta	(e_+, m_+)	Mean	1.0000	0	0.0000	0	1.0000	0	0	0
		Std.err.	0.0001	0	0.0001	0	0	0	0	0
	(e_+, m_-)	Mean	0.1862	0.8138	0.0000	0.0000	0.2580	0.7420	0	0
		Std.err.	0.0338	0.0338	0.0000	0.0004	0.0411	0.0411	0	0
	(e_-, m_+)	Mean	0.1080	0	0.8920	0	0.1881	0	0.8119	0
		Std.err.	0.0355	0	0.0355	0	0.0456	0	0.0456	0
	(e_-, m_-)	Mean	0.0053	0.0359	0.0567	0.9022	0.0138	0.0541	0.0781	0.8540
		Std.err.	0.0026	0.0119	0.0126	0.0202	0.0057	0.0139	0.0151	0.0264
Special Economic Zone	(e_+, m_+)	Mean	0.9719	0.0274	0.0007	0.0001	0.9671	0.0318	0.0010	0.0001
		Std.err.	0.0204	0.0194	0.0028	0.0003	0.0272	0.0257	0.0034	0.0006
	(e_+, m_-)	Mean	0.0015	0.9967	0	0.0018	0.0035	0.9942	0.0000	0.0023
		Std.err.	0.0063	0.0093	0	0.0072	0.0099	0.0121	0.0000	0.0081
	(e_-, m_+)	Mean	0.0478	0.0017	0.9018	0.0488	0.0573	0.0021	0.8837	0.0569
		Std.err.	0.0338	0.0025	0.0371	0.0348	0.0426	0.0029	0.0422	0.0461
	(e_-, m_-)	Mean	0.0000	0.0184	0.0004	0.9812	0.0001	0.0207	0.0007	0.9785
		Std.err.	0.0002	0.0129	0.0016	0.0133	0.0003	0.0150	0.0022	0.0160

Notes: Colors denote quantitatively large changes. Red denotes effects leading to changes in state choices and blue denotes effects that do not lead to changes (indicating stayers).