

# R&D spillover effects on firm innovation - A spatial approach \*

Emmanuel S. Tsyawo<sup>†</sup>

[Latest version here](#)

September 2019

## Abstract

Quantifying R&D spillover effects requires a spatial matrix that characterises the strength of connectivity between firms. In practice, the spatial matrix is often unknown. This paper proposes a parsimonious approach to estimating the spatial matrix alongside parameters and quantifies R&D spillovers on innovation. The approach generalises to a class of linear and non-linear models and allows asymmetry and time-variation in the spatial matrix. On firm innovation, we find positive spillover and private effects of R&D on innovation. We confirm time-variation and asymmetry in the interaction structure of firms and find that geographic and product market proximity are relevant. Moreover, the strength of connectivity between firms is not limited to often-assumed notions of closeness.

*Keywords: R&D spillovers, innovation, spatial dependence, spatial matrix*

*JEL classification: C21, L25*

---

\*Our gratitude goes to Brantly Callaway, Oleg Rytchkov, Charles E. Swanson, Catherine Maclean, Austin Bean, Pedro Silos, and all participants at the Temple University Econ Summer Job Market Boot-camp for helpful comments. This project is supported by the Temple University Economics Department Summer (2019) Research Grant.

<sup>†</sup>Department of Economics, Temple University. Email: estsyawo@temple.edu

# 1 Introduction

Quantifying the spillover effects of research and development (R&D) on firm innovation, productivity, costs, and growth is a major pursuit in the empirical industrial organisation literature. Firms’ research and development (R&D) efforts are known to generate spillover effects on other firms’ outcomes (e.g., Grillitsch and Nilsson (2015), Bloom, Schankerman, and Van Reenen (2013), and Cincera (1997)). Firms engage in R&D not only to create knowledge but also to boost their capacity in order to use knowledge created by other firms (Cohen and Levinthal, 1989). While there appears to be a consensus on the existence and importance of R&D spillovers, existing approaches to quantifying R&D spillover effects remain varied. Knowledge spillovers have policy implications and quantifying them aids to fine-tune policy,<sup>1</sup> promote innovation, and accelerate technological progress.

A crucial component to quantifying spillover effects is the spillover measure. The spillover measure corresponding to a firm is obtained by weighting the R&D of other firms using a spatial matrix that defines the strength of connectivity between firms. In the literature, there are different ways to constructing the spatial matrix leading to different spillover measures. We note that different measures of spillovers lead to different conclusions and policy recommendations (see, e.g., Bhattacharjee and Jensen-Butler (2013)) whence this paper’s introduction of a flexible methodology used in modelling and estimating the spatial matrix in order to quantify R&D spillover effects on innovation.

The method we propose models the structure of interactions between firms across time using spatial matrices. The framework is applicable to a wide class of linear and non-linear models. This is particularly crucial for quantifying spillovers and spatial dependence when the appropriate model is not linear. Our approach is parsimonious, data-driven, and testable. It allows a time-varying and asymmetric spatial matrix that admits several spillover covari-

---

<sup>1</sup>Spillovers such as positive externalities imply under-investment in R&D from a social perspective and a need to subsidise firm R&D effort via policies such as R&D tax credits. In assessing knowledge spillovers of firms, Bloom, Schankerman, and Van Reenen (2013) finds limited knowledge spillovers of small firms and recommends a reconsideration of higher R&D tax credits given to small firms.

ates simultaneously. The relevance spatial covariates in the spatial matrix is statistically assessable and hence obviates the arbitrary choice of metric, spatial matrix, or both. We are able to identify spillover covariates and conduct hypothesis tests on their significance in order to establish their relevance. Identification of spillover sources (measured by spillover covariates, e.g., geographic distance) is particularly suitable for studies quantifying spillover effects as it sheds light on the channels of transmission of spillovers between firms. Our method is easy to implement and standard tools of M-estimation and inference are applicable. We contribute to the literature in two ways. First, we quantify spillover effects when the interaction structure is unknown. Second, we propose a method generalisable to a class of linear and non-linear models that admit linear predictor functions.

Works in the literature construct R&D spillover measures in a number of ways; we identify five of them. A simple approach pools the R&D or knowledge stock generated by other firms in the same industry (see, for example, Bernstein and Nadiri (1989) and Cincera (1997)).<sup>2</sup> A second approach employs industry distance to the frontier<sup>3</sup> or a measure of potential of knowledge transfer as a proxy for technological spillovers (see Grillitsch and Nilsson (2015), Acemoglu et al. (2007), and Griffith, Redding, and Reenen (2004)). A third approach constructs spillover measures by using exports, imports, and foreign direct investment to weight domestic and foreign R&D stocks (e.g., Coe, Helpman, and Hoffmaister (2009)). Bloom, Schankerman, and Van Reenen (2013) uses Jaffe (1986)'s approach to construct exogenous spatial matrices based on distances in technology and product market spaces, from which R&D spillover measures obtain. Last, a recent approach estimates the spatial matrix from data (e.g., Manresa (2013) and Soale and Tsyawo (2019)).

At the econometric level, much progress is realised in the last two decades with respect to modelling spillover effects and spatial dependence in general. A first category spatially lags the outcome variable, namely, the Spatial Autoregressive (SAR) model (Anselin, Le Gallo,

---

<sup>2</sup>Cincera (1997) pools R&D investment at the manufacturing sector level and only allows intra-sector spillovers.

<sup>3</sup>Griffith, Redding, and Reenen (2004) defines an industry's frontier as the country with the largest total factor productivity (TFP).

and Jayet, 2008). A second category estimates the spatial matrix as a set of parameters. The second category requires high-dimensional estimation techniques whenever the number of time periods does not substantially exceed the number of firms (Manresa, 2013; Soale and Tsyawo, 2019). A third models individual elements of the spatial matrix directly up to a finite number of parameters and jointly estimates all parameters using panel data (Kapetanios, Mitchell, and Shin, 2014; Pinkse, Slade, and Brett, 2002). Our approach falls in this last category. We employ a flexible approach to estimating the spatial matrix and allow a generalisation to a class of models with linear predictor functions.

As regards constructing spillover measures and quantifying R&D spillovers, we note a number of shortcomings of existing approaches. Exogenously constructing spillover measures from pre-specified metrics can lead to misleading results if the choice of metric is not guided by theory. On the Jaffe (1986) measure, note the assumption that interaction between firms  $i$  and  $j$  is proportional to a time-invariant and symmetric weight  $w_{ij}$ . The interaction between firms of different sizes and research capacities is generally not symmetric. In settings where the structure of interactions evolves through time, time-invariant spatial matrices are inadequate.

Direct application of existing spatial panel models to quantify R&D spillovers presents some challenges. First, spatial matrices are often not known with certainty. The choice of pre-specified spatial matrices in the spatial econometrics literature is often arbitrary<sup>4</sup> and empirical results vary substantially by choice of metric (e.g., Bhattacharjee and Jensen-Butler, 2013). In the presence of multiple sources of R&D spillovers, a researcher faces a *metric ambiguity* problem. Besides the often documented sources of knowledge transfers viz. geographic (Grillitsch and Nilsson, 2015; Lychagin, Pinkse, Slade, and Reenen, 2016), technology (Bloom, Schankerman, and Van Reenen, 2013; Manresa, 2013), product market (Bloom, Schankerman, and Van Reenen, 2013), or input market factors, knowledge spillovers

---

<sup>4</sup>A case in point, in a study that analyses the monthly change rates of the consumer price index (CPI) in the EU, Dou, Parrella, and Yao (2016) uses a normalised sample correlation matrix of monthly CPI as spatial matrix.

can be explained by social networks, labour mobility *inter alia*.<sup>5</sup>

Second, modelling spatial dependence is largely confined to linear panel data models. Recent contributions to the literature on non-linear spatial models include Xu and Lee (2015) and Hoshino (2017). These approaches, however, suppose foreknowledge of the spatial matrix and lack direct generalisability<sup>6</sup> to other non-linear models. In this paper, innovation is measured using citation-weighted patent counts (see also Hall, Jaffe, and Trajtenberg (2005), Aghion, Van Reenen, and Zingales (2013), and Huang and Tsyawo (2018)). Besides citation-weighted patent counts, patent applications (e.g., Cincera (1997)), total factor productivity, TFP, (Bloom, Schankerman, and Van Reenen (2013)), and a binary on collaboration in innovative activity (Grillitsch and Nilsson, 2015) are also used as measures of innovation. Except TFP, non-linear models viz. negative binomial, Poisson, and logit models are required for estimation because the outcome variables are discrete.

In light of the preceding challenges, we propose a framework that quantifies spillovers effects in both linear and non-linear models when the spatial matrix is unknown. Beyond the spillover framework, our model is applicable to modelling networks using panel data (see Manresa (2013), Souza (2014), and Soale and Tsyawo (2019)). By flexibly modelling weights as functions of observed exogenous spillover covariates, our framework avoids endogeneity in the spatial matrix.<sup>7</sup> A flexible approach that incorporates several metrics among other spatial covariates is a good choice in the presence of metric ambiguity or where the notion of “metric” is not guided by theory.<sup>8</sup>

Our empirical application studies spillover effects of R&D on firm innovation. Firm innovation is measured by citation-weighted patent counts and R&D is measured by the

---

<sup>5</sup>Grillitsch and Nilsson (2015) finds that firms in peripheral regions with less access to local (geographical) knowledge spillovers compensate for the lack via collaborations with non-local firms. Bhattacharjee and Jensen-Butler (2013, p. 630) notes a number of non-significant weights between contiguous region pairs. These findings support the assertion that sources of knowledge spillovers are sometimes less obvious.

<sup>6</sup>Hoshino (2017)’s method estimates a censored SAR by interacting the latent uncensored propensity variable. This framework is limited to models with continuous latent outcome representations.

<sup>7</sup>Some works in the literature focus on endogeneity in spatial matrices. Kelejian and Piras (2014) uses an IV whereas Qu and Lee (2015) uses a control function approach to deal with endogeneity of spatial matrices.

<sup>8</sup>See, for example, Pinkse, Slade, and Brett (2002) and Pinkse and Slade (2004).

firm’s research and development stock. Due to the discreteness of the outcome variable, the negative binomial model is used in estimation. Our empirical results confirm the sensitivity of spillover measures to the choice of spatial covariate (and spatial matrix, by extension). We confirm the presence of positive spillover effects of R&D on firm innovation. Though the private effect of R&D on innovation is positive and statistically significant, it is dominated by the spillover effect. Besides geographic and product market proximity, the other relevant spillover covariates are associated with past R&D and patenting behaviour of firms. The results on relevant spatial covariates enable us to confirm asymmetry and time-variation in the strength of connectivity between firms with respect to innovation. Also, the results confirm that the strength of connectivity between firms is not only tied to commonly used notions of proximity viz. geographic and product market.

The rest of the paper is organised as follows. Section 2 describes the model and presents a general identification result. Section 4 covers estimation algorithms and section 5 presents asymptotic results. The empirical study is conducted in section 6, and section 7 concludes. All proofs are relegated to appendix A.

## 2 The model

Modelling spatial dependence using our approach spans a class of models that admit linear predictor functions, e.g., linear, quantile, logit, and probit regressions. In this paper, we focus on the negative binomial model because the outcome variable (Citation-weighted patent counts) is discrete.

## 2.1 The conditional expectation

The conditional expectation is the functional of interest for our model (the negative binomial).

$$(2.1) \quad E(y_{it} | \mathbf{x}_t, \mathbf{z}_{it}, \mathbf{D}_t) = \exp(\rho_0 + \rho_1 x_{it} + \rho_2 \sum_{j \neq i} x_{jt} w_{ijt}(\boldsymbol{\delta}) + \mathbf{z}_{it} \boldsymbol{\gamma})$$

where  $w_{ijt}(\boldsymbol{\delta}) \equiv \frac{\exp(\mathbf{d}'_{ijt} \boldsymbol{\delta})}{\sum_{j \neq i} \exp(\mathbf{d}'_{ijt} \boldsymbol{\delta})}$ ,  $\forall i = 1, \dots, N$ ,  $t = 1, \dots, T$ . The function  $\exp(\cdot)$  can be replaced by any twice differentiable function  $f : \mathbb{R} \rightarrow \mathbb{R}_+$ .

We partition the parameter space  $\boldsymbol{\Theta} \subset \mathbb{R}^{k_\theta}$ ,  $k_\theta = k_\beta + k_\delta$  into  $\mathbf{B} \subset \mathbb{R}^{k_\beta}$  and  $\boldsymbol{\Delta} \subset \mathbb{R}^{k_\delta}$  where  $\boldsymbol{\beta} = [\rho_0, \rho_1, \rho_2, \boldsymbol{\gamma}']' \in \mathbf{B}$ ,  $\boldsymbol{\delta} \in \boldsymbol{\Delta}$ ,  $\boldsymbol{\Theta} \equiv \mathbf{B} \times \boldsymbol{\Delta}$ , and  $\boldsymbol{\theta} \equiv [\boldsymbol{\beta}', \boldsymbol{\delta}']'$ . The following variables are observed: (1) outcome  $y_{it}$ , (2) covariates  $\mathbf{z}_{it}$  and  $\mathbf{x}_t = [x_{1t}, \dots, x_{Nt}]'$  (where  $\mathbf{x}_t$  generates spillovers, e.g., R&D), and (3) spillover covariates  $\mathbf{d}_{ijt}$  (e.g. proximity in technology space).  $\mathbf{D}_t \equiv \{\mathbf{d}_{ijt} : i = 1, \dots, N, j \neq i\}$  and let  $\mathbf{D} \equiv \bigcup_{t=1}^T \mathbf{D}_t$ . Define an  $NT \times (k_\beta - 1)$  matrix  $\mathbf{z}$  whose  $((t-1)N + i)$ 'th row is  $\mathbf{z}_{it}$  and an  $NT \times 1$  vector  $\mathbf{x}$  whose  $((t-1)N + i)$ 'th element is  $x_{it}$ . Define an  $NT \times NT$  block diagonal matrix  $\mathbf{w}(\boldsymbol{\delta}) = \text{diag}[\mathbf{w}_1(\boldsymbol{\delta}), \dots, \mathbf{w}_T(\boldsymbol{\delta})]$  where the  $(i, j)$ 'th element of  $\mathbf{w}_t(\boldsymbol{\delta})$ ,  $t \in \{1, \dots, T\}$  is  $w_{ijt}(\boldsymbol{\delta})$  if  $j \neq i$  and zero otherwise. The  $NT \times (k_\beta + 1)$  design matrix associated with the conditional expectation eq. (2.1) is given by  $\mathbf{m} = [\mathbf{1}_{NT}, \mathbf{x}, \mathbf{w}\mathbf{x}, \mathbf{z}]$  where  $\mathbf{1}_{NT}$  denotes an  $NT \times 1$  vector of ones, and the dependence of  $\mathbf{w}$  and  $\mathbf{m}$  on  $\boldsymbol{\delta}$  is suppressed for notational convenience.

Spillover covariates, denoted by the  $k_\delta \times 1$  vector  $\mathbf{d}_{ijt}$ , are indicators of connectivity between firms and can contain several metrics (e.g. proximity in technological space and geographical proximity) and flexible functional forms of variables related to R&D, firm characteristics, or innovation but not necessarily tied to, for instance, geography or industry (e.g. lagged  $x_{it}$ ,  $x_{jt}$ ,  $j \neq i$  in, for example, Kapetanios, Mitchell, and Shin (2014)).  $\mathbf{d}_{ijt} \in \mathbf{D}_t$  allows a flexible modelling of the weight function across observations and through time up to a parameter vector  $\boldsymbol{\delta}$  of finite length  $k_\delta$ . The formulation eq. (2.1) allows innovation of firm  $i$  to be impacted not only by own R&D  $x_{it}$  and characteristics  $\mathbf{z}_{it}$ , but also by the R&D

$x_{jt}$ ,  $j \neq i$  of other firms.

The negative binomial is not the only model to which our model is applicable. Our approach applies to a class of models that admit linear predictor functions as in eq. (2.1). Examples include linear regression, quantile regression, logit, and Poisson models. To set the scope of the model, consider a baseline functional of the distribution of outcome  $y_{it}$  conditional on observables

$$(2.2) \quad \nu(y_{it}|\mathbf{x}_t, \mathbf{z}_{it}, \mathbf{D}_t) = g(\rho_0 + \rho_1 x_{it} + \rho_2 \sum_{j \neq i} x_{jt} w_{ijt}(\boldsymbol{\delta}) + \mathbf{z}_{it} \boldsymbol{\gamma})$$

where  $g : \mathbb{R} \rightarrow \mathbb{R}$  is known and differentiable, and the conditional functional can be the conditional expectation (e.g., the negative binomial model eq. (2.1)), the conditional quantile (quantile regression), or the conditional probability (distribution regression).

## 2.2 The objective function

The objective function has a sample average representation  $\mathcal{Q}_n(\boldsymbol{\beta}, \boldsymbol{\delta}) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T q_{it}(\boldsymbol{\beta}, \boldsymbol{\delta})$  and the parameter set of interest solves the following minimisation problem

$$(2.3) \quad [\hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\delta}}']' = \arg \min_{[\boldsymbol{\beta}', \boldsymbol{\delta}']' \in \boldsymbol{\Theta}} \mathcal{Q}_n(\boldsymbol{\beta}, \boldsymbol{\delta})$$

$q_{it}(\boldsymbol{\beta}, \boldsymbol{\delta})$  is the contribution of unit  $i$  at time  $t$  to the objective function  $\mathcal{Q}_n(\cdot, \cdot)$  and  $q_{it}(\boldsymbol{\beta}, \boldsymbol{\delta}) = -\eta^{-2} \log[\eta^{-2}/(\eta^{-2} + \exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}))] - y_{it} \log[\exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta})/(\eta^{-2} + \exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}))]$  for a given finite  $\eta^2 > 0$  where  $\exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}) = \exp(\rho_0 + \rho_1 x_{it} + \rho_2 \sum_{j \neq i} x_{jt} w_{ijt}(\boldsymbol{\delta}) + \mathbf{z}_{it} \boldsymbol{\gamma})$  and  $m_{it}(\boldsymbol{\delta})$  is the  $((t-1)N + i)$ 'th row of  $\mathbf{m}(\boldsymbol{\delta})$ .

$q_{it}(\boldsymbol{\beta}, \boldsymbol{\delta})$  of other models includes  $q_{it}(\boldsymbol{\beta}, \boldsymbol{\delta}) = (y_{it} - m_{it}(\boldsymbol{\delta})\boldsymbol{\beta})^2$  for linear regression and  $q_{it}(\boldsymbol{\beta}, \boldsymbol{\delta}) = -(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta})y_{it} - \log(1 + \exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}))$  for the logit model.

If  $\mathbf{D}$  contains at least one time-varying element,  $\mathbf{w}_t(\boldsymbol{\delta})$  for  $\boldsymbol{\delta} \in \boldsymbol{\Delta}$ ,  $\boldsymbol{\delta} \neq \mathbf{0}$  varies through time and the structure of spatial dependence need not be stable through time. Also, we do



not assume  $\mathbf{d}_{ijt} = \mathbf{d}_{ijt}$ ,  $j \neq i$  whence the possible asymmetry of  $\mathbf{w}(\boldsymbol{\delta})$ . Features of our model like time-variation and asymmetry can be assessed statistically by testing the significance of elements in  $\boldsymbol{\delta}$  corresponding to time-varying and asymmetric elements in  $\mathbf{D}$ .

## 2.3 Identification

We make the following assumptions.

**Assumption 1.** (a)  $\mathbf{z} \equiv [\mathbf{1}_{NT}, \mathbf{x}, \mathbf{z}]$  is full column rank. (b) Each spatial matrix  $\mathbf{w}_t(\boldsymbol{\delta})$  is non-singular,  $t = 1, \dots, T$  and for all  $\boldsymbol{\delta} \in \Delta$ . (c) The diagonal elements of  $\mathbf{w}_t(\boldsymbol{\delta})$  are zeroes,  $t = 1, \dots, T$  and for all  $\boldsymbol{\delta} \in \Delta$

**Assumption 2.** The parameter space  $\Theta$  of  $\boldsymbol{\theta} = [\boldsymbol{\beta}', \boldsymbol{\delta}']'$  is a compact subset of  $\mathbb{R}^{k_\theta}$ .

Assumption 1(a) is a standard assumption. Assumption 1(b) holds if there is enough variation in  $\mathbf{D}_t$  for  $\boldsymbol{\delta} \in \Delta$  and  $t = 1, \dots, T$ . Note that  $\mathbf{x} \neq \mathbf{0}$  follows from assumption 1(a), and it is necessary (coupled with assumption 1(b)) for  $\mathbf{w}\mathbf{x}$  to possess independent variation from  $\mathbf{z}$ . These assumptions are verifiable. In the following, we show the full column rank condition of the design matrix  $\mathbf{m}$  holds for all  $\boldsymbol{\delta} \in \Delta$  under assumption 1.

**Lemma 1** (Full column rank of design matrix  $\mathbf{m}$ ). *Under assumption 1(a-c),  $\mathbf{m}$  is column full rank for all  $\boldsymbol{\delta} \in \Delta$ .*

The following assumptions are useful in bounding the design matrix  $\mathbf{m}(\boldsymbol{\delta})$  for all  $\boldsymbol{\delta} \in \Delta$

**Assumption 3.** (a) There exists a positive constant  $\kappa_w < \infty$  such that  $\sup_{\boldsymbol{\delta} \in \Delta} \exp((\mathbf{d}_{ijt} - \bar{\mathbf{d}}_{i.t})'\boldsymbol{\delta}) \leq \kappa_w$ ,  $\bar{\mathbf{d}}_{i.t} \equiv N^{-1} \sum_{j \neq i} \mathbf{d}_{ijt}$  for all  $\mathbf{d}_{ijt} \in \mathbf{D}_t$  and  $t = 1, \dots, T$ . (b) There exists positive constant  $\kappa_y$ ,  $1 \leq \kappa_y < \infty$  such that  $\sup_{\boldsymbol{\beta} \in \mathbf{B}, \boldsymbol{\delta} \in \Delta} |m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}| \leq 1/2 \log \kappa_y$  for all  $i = 1, \dots, N$  and  $t = 1, \dots, T$  (c) The elements of  $[\mathbf{x}, \mathbf{z}]$  are bounded in absolute value.

By definition, the row sums of  $\mathbf{w}(\boldsymbol{\delta})$  are bounded at 1. In the following lemma, we bound the column sums of  $\mathbf{w}(\boldsymbol{\delta})$ .

**Lemma 2.** *Under assumption 3(a), the column sums of  $\mathbf{w}(\boldsymbol{\delta})$  are bounded in absolute value for all  $\boldsymbol{\delta} \in \Delta$ .*

**Lemma 3.** *Under assumption 3(b),  $E|\mathcal{Q}_n(\boldsymbol{\beta}, \boldsymbol{\delta})| < \infty$  for all  $\boldsymbol{\beta} \in \mathbf{B}$  and  $\boldsymbol{\delta} \in \Delta$ .*

**Lemma 4** (Identification). *Under the assumptions of lemma 1,  $\boldsymbol{\theta}_o = [\boldsymbol{\beta}'_o, \boldsymbol{\delta}'_o]'$  is identified.*

**Theorem 1** (Information Inequality). *Under the assumptions of lemmas 1 to 4  $\mathcal{Q}_o(\boldsymbol{\beta}, \boldsymbol{\delta}) \equiv E[\mathcal{Q}_n(\boldsymbol{\beta}, \boldsymbol{\delta})]$  has a unique minimum at  $\boldsymbol{\theta}_o = [\boldsymbol{\beta}'_o, \boldsymbol{\delta}'_o]'$ .*

## 2.4 Parameters of interest

We explore parameters that are interesting from a policy perspective. Though these parameters can have varied interpretation by application (see section 3.1 for examples), we interpret them in the context of R&D spillover effects on innovation. In what follows, let  $x_{it}$  denote firm  $i$ 's R&D in year  $t$ .

**Private effect:** The private effect of R&D measures the impact of a firm's R&D on its innovation.  $\mathcal{PE}_i = T^{-1} \sum_{t=1}^T \frac{\partial E(y_{it}|\cdot)}{\partial x_{it}} = T^{-1} \sum_{t=1}^T \mathcal{PE}_{it} = \rho_1 T^{-1} \sum_{t=1}^T E(y_{it}|\cdot)$ . The average private effect is given by  $\mathcal{PE} = N^{-1} \sum_{i=1}^N \mathcal{PE}_i$ . The parsimony of our model does not allow heterogeneity in  $\rho_1$  across firms unlike, Manresa (2013) and Soale and Tsyawo (2019, sect. 6). It does, however, allow variation through time and does not require a static interaction structure.

**Spillover effect:** The spillover effect measures the impact of firm  $j$ 's R&D on  $i$ 's innovation. The building block for computing spillover effects, taking account of pairwise interaction of firms  $i, j, i \neq j$  and time  $t$  is  $\mathcal{SpE}_{ijt} = \frac{\partial E(y_{it}|\cdot)}{\partial x_{jt}} = \rho_2 w_{ijt} E(y_{it}|\cdot)$  where  $w_{ijt} \equiv w_{ijt}(\boldsymbol{\delta})$ . The spillover effect of a firm  $j$ 's R&D is the effect it has on another firm  $i$ 's,  $j \neq i$ , innovation. It is given by  $\mathcal{SpE}_{ij.} = T^{-1} \sum_{t=1}^T \mathcal{SpE}_{ijt} = \rho_2 T^{-1} \sum_{t=1}^T w_{ijt} E(y_{it}|\cdot)$ . The spillover effect of other firms' R&D on firm  $i$ 's innovation obtains as  $\mathcal{SpE}_{i..} = \sum_{j \neq i} \mathcal{SpE}_{ij.}$ , the spillover effect exerted by firm  $j$  on other firms' innovation is  $\mathcal{SpE}_{.j.} = \sum_{i \neq j} \mathcal{SpE}_{ij.}$ , and the average spillover

effects across all firms is  $\mathcal{Sp}\mathcal{E} = N^{-1} \sum_{i=1}^N \mathcal{Sp}\mathcal{E}_{i..} = N^{-1} \sum_{j=1}^N \mathcal{Sp}\mathcal{E}_{.j.}$ . A researcher interested in time-variation of average spillover effects can compute  $\mathcal{Sp}\mathcal{E}_{..t} = N^{-1} \sum_{i=1}^N \sum_{j \neq i} \mathcal{Sp}\mathcal{E}_{ijt}$  for each  $t \in \{1, \dots, T\}$ .

**Social effects:** Social effects of R&D on innovation sum private and spillover effects. The social effect generated by firm  $j$ 's R&D at time  $t$  is  $\mathcal{SE}_{.jt} = \sum_{i=1}^N \frac{\partial E(y_{it}|\cdot)}{\partial x_{jt}} = \frac{\partial E(y_{jt}|\cdot)}{\partial x_{jt}} + \sum_{i \neq j} \frac{\partial E(y_{it}|\cdot)}{\partial x_{jt}} = \mathcal{PE}_{jt} + \sum_{i \neq j} \mathcal{Sp}\mathcal{E}_{ijt} = \mathcal{PE}_{jt} + \mathcal{Sp}\mathcal{E}_{.jt}$ . In a similar vein, the social effects of R&D (from all firms) received by firm  $i$  at time  $t$  is defined as  $\mathcal{SE}_{i.t} = \sum_{j=1}^N \frac{\partial E(y_{it}|\cdot)}{\partial x_{jt}} = \frac{\partial E(y_{it}|\cdot)}{\partial x_{jt}} + \sum_{j \neq i} \frac{\partial E(y_{it}|\cdot)}{\partial x_{jt}} = \mathcal{PE}_{it} + \sum_{j \neq i} \mathcal{Sp}\mathcal{E}_{ijt} = \mathcal{PE}_{it} + \mathcal{Sp}\mathcal{E}_{i.t}$ . Average social effects across firms and time obtains as  $\mathcal{SE} = (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \mathcal{SE}_{i.t} = (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \mathcal{SE}_{.jt}$ .<sup>9</sup>

**Strength of spatial dependence:**  $\rho_2$  controls the strength of the spatial dependence in our model; it can accommodate both weak and strong forms of spatial dependence<sup>10</sup> depending on the magnitude of  $\rho$ . A null hypothesis  $H_o : \rho_2 = 0$  can be used to assess the strength of spatial dependence. This approach provides a way of testing the strength of spatial dependence in a specific model. Failing to reject  $H_o$  is indicative of a weak spatial dependence.

**Relevant spatial covariates:** By modelling the spatial matrix  $\mathbf{w}$  as a function of observable spillover covariates in  $\mathbf{D}$  up to a parameter vector  $\boldsymbol{\delta}$  of finite length  $k_\delta$ , our approach allows the evaluation of factors that drive spatial dependence i.e., test elements of  $\boldsymbol{\delta}$  that are statistically significant.<sup>11</sup> In a similar vein, the strength of time-variation or asymmetry in  $\mathbf{w}$  is assessable by testing the significance (jointly or individually) of elements in  $\boldsymbol{\delta}$  that correspond to time-varying or asymmetric elements in  $\mathbf{D}$ .

<sup>9</sup>Comparable approaches to computing structural parameters viz. private, spillover, and social effects are given in Bloom, Schankerman, and Van Reenen (2013, Sect. 6.5.1) and Manresa (2013, Sect. 5.5).

<sup>10</sup>This feature is also present in Kapetanios, Mitchell, and Shin (2014)'s model.

<sup>11</sup>Lychagin, Pinkse, Slade, and Reenen (2016), for example, finds that R&D spillover effects on firm productivity via geographic and technological proximity (but not product market) are important.

## 3 Extensions

### 3.1 Possible empirical applications of the model

A crucial aspect to our model, from the econometric perspective, is the ability to model spatial dependence in both linear and non-linear models. Also, it allows empirical determination of observable spillover covariates as well as an assessment of time-variation and asymmetry in the spatial matrix  $\mathbf{w}$ . Empirically, our method has applications beyond the study of R&D spillover effects on innovation. In the following, we illustrate, with empirical examples, other possible applications of our model.

**R&D Spillovers on firm productivity:** A major challenge with estimating R&D spillover effects on firm productivity is that spillovers may not be tied to any plausible metric, for example, geographic distance (Syverson, 2011). Like Lychagin, Pinkse, Slade, and Reenen (2016), our framework is applicable to study this problem by jointly accounting for all plausible sources of R&D spillovers (measured by the respective spatial covariates) in a production function framework. Unlike Bloom, Schankerman, and Van Reenen (2013), the spatial matrix is determined within the model and the interaction between firms over time can vary through time and be asymmetric. Our proposed framework can also be used to study R&D spillover effects on other interesting firm outcomes viz. market value and R&D factor demand.

**Spillovers in demand behaviour:** Spillovers in demand due to geographically varying prices, e.g., regional housing demand (Bhattacharjee and Jensen-Butler, 2013), rice demand (Case, 1991), and cigarette demand Kelejian and Piras, 2014 receive much attention in the literature. Markets are not perfectly segregated, and shocks to demand in a region impact demand behaviour in other regions. Our model can contribute to this literature by allowing time-variation, asymmetry, and other metrics (besides geographical contiguity) in determining the spatial matrix.

### 3.2 Extensions of the model

The approach we propose is not without drawbacks. Our approach requires metrics or spillover covariates. The elements of the spatial matrix are non-negative.<sup>12</sup> Also, the number of elements in the spatial matrix to be modelled grows in the order of  $O(N^2T)$  which may require the dimension of  $\boldsymbol{\delta}$  to grow. Because of the exponential function in  $w_{ijt}(\boldsymbol{\delta})$ , our approach may not accommodate sparse interaction structures well without further adjustment. Consistent estimation of the spatial matrix requires a balanced panel, and this can result in the loss of some firms. We propose extensions to the baseline model eq. (2.1) to deal with these challenges. Though we do not explore the computational and inferential challenges associated with the extensions, we present them.

**Dynamic spatial dependence:** Past outcomes of other firms may affect future outcomes. Capturing such dynamic spatial dependence is quite straightforward.

$$E(y_{it}|\mathbf{x}_t, \mathbf{z}_{it}, \mathbf{D}_t) = \exp(\rho_0 + \rho_1 y_{i,t-1} + \rho_2 \sum_{j \neq i} y_{j,t-1} w_{ijt}(\boldsymbol{\delta}) + \mathbf{z}_{it} \boldsymbol{\gamma})$$

$E(y_{it}|\mathbf{x}_t, \mathbf{z}_{it}, \mathbf{D}_t)$  above generalises Kapetanios, Mitchell, and Shin (2014)’s model. Like Cincera (1997, Sect. 4), dynamics in the spillover parameter can be accommodated by including lagged  $x_{it}, x_{jt}$  and lagged spatial matrices in order to capture dynamics in, for example, the patenting process. In some applications, lagged outcome drives spatial dependence and can be used as a spatial covariate (e.g. Kapetanios, Mitchell, and Shin (2014)). The spatial covariate vector  $\mathbf{d}_{ijt}$  in  $w_{ijt}(\boldsymbol{\delta})$  can contain functions of  $y_{i,t-1}$  and  $y_{j,t-1}$ , e.g.,  $(y_{i,t-1} - y_{j,t-1})$ ,  $|y_{i,t-1} - y_{j,t-1}|$ , polynomial functions of  $(y_{i,t-1} - y_{j,t-1})$ , or a bivariate polynomial series expansion of  $y_{i,t-1}$  and  $y_{j,t-1}$ . More lags of the outcome can be included in the conditional expectation as covariates or spatial covariates provided there are sufficient degrees of freedom to estimate the model.

---

<sup>12</sup>Soale and Tsyawo (2019), Pinkse, Slade, and Brett (2002), Bhattacharjee and Jensen-Butler (2013), and Manresa (2013) allow for both negative and positive weights.

**Multiple metrics:** A particularly useful feature of our model lies in its ability to handle metric ambiguity by incorporating several metrics in  $\mathbf{D}$ . Spatial dependence in outcomes can emerge as an interplay of several metrics. For example, production-related spillovers from R&D can be simultaneously impacted by a number of factors linked to geographical, output market, input market, and technological distance. For example, Lychagin, Pinkse, Slade, and Reenen (2016) simultaneously assesses geographic, technological, and product market spaces as sources of R&D spillovers. In the case where there is uncertainty over the choice of relevant metrics, our formulation allows for a data-driven determination of the relevant subset of metrics and spatial covariates (or functions thereof) via hypothesis tests.

**Prior information in spatial matrix:** Where sparsity in the spatial matrix is justifiable, binaries on contiguity can be incorporated in the spatial matrix to induce sparsity. For instance,  $w_{ijt}(\boldsymbol{\delta})$  can be weighted by  $C(d_{ijt}) = 1$  if  $i$  and  $j$  are contiguous at time  $t$  and zero otherwise in order to reduce the number of elements in the spatial matrix to model.<sup>13</sup>

**Group heterogeneity:** In some cases, it is plausible to assume that spillovers are specific to pairs of groups of firms (see, e.g., Manresa (2013, p. 5.2.1)). Group heterogeneity implicitly assumes that firms within the same group (e.g. by industrial classification) are equally impacted by R&D of firms in other groups. Group heterogeneity is particularly helpful in dealing with unbalanced panel data with thousands of firms and few time periods. The conditional expectation eq. (2.1), under group heterogeneity, is

$$(3.1) \quad E(y_{it}|\mathbf{x}_t, \mathbf{z}_{it}, \mathbf{D}_t) = \exp(\rho_0 + \rho_1 x_{it} + \rho_2 \sum_{L \neq i, K \ni i} w_{KLt}(\boldsymbol{\delta}) \sum_{j \in L} x_{jt} + \mathbf{z}_{it} \boldsymbol{\gamma})$$

where  $w_{KLt}(\boldsymbol{\delta}) \equiv \frac{\exp(\mathbf{d}'_{KLt} \boldsymbol{\delta})}{\sum_{j \neq i}^N \exp(\mathbf{d}'_{KLt} \boldsymbol{\delta})}$  and  $\mathbf{d}_{KLt} \in \mathbf{D}_t$  is specific to groups  $K$  and  $L$ . The baseline case obtains when there are  $N - 1$  groups. Group heterogeneity leaves group mem-

---

<sup>13</sup>See Lychagin, Pinkse, Slade, and Reenen (2016, Sect. IV) a discussion on this. The binary metric can be extended to capture proximity in other settings like industry, product market among others.

bership across time unrestricted as long as group membership is observed. Formulation 3.1 has a major advantage of not allowing the dimension of  $\mathbf{w}_t(\boldsymbol{\delta})$  to remain fixed as  $T$  grows *ad infinitum*.

## 4 Estimation

The model corresponding to eq. (2.1) cannot be estimated in two steps because of an intrinsic latency. Since  $\boldsymbol{\delta}$  is an unknown parameter vector, the  $N \times N$  spatial matrices  $\mathbf{w}_t(\boldsymbol{\delta})$ ,  $t = 1, \dots, T$  are unknown ex-ante. The foregoing suggests a joint determination of  $\boldsymbol{\beta}$  and  $\boldsymbol{\delta}$ . Estimation of eq. (2.3) is fairly straightforward using built-in optimisers in available software. Proceeding thus, however, raises a number of issues. First, not all models have smooth objective functions  $\mathcal{Q}_n(\boldsymbol{\beta}, \boldsymbol{\delta})$ , e.g., quantile regression. Second, where the dimensionality of  $\boldsymbol{\beta}$  is high (e.g., firm and year fixed effects with large  $N$  and  $T$ ), direct minimisation of  $\mathcal{Q}_n(\boldsymbol{\beta}, \boldsymbol{\delta})$  with respect to  $\boldsymbol{\beta}$  and  $\boldsymbol{\delta}$  becomes very slow and risks getting trapped in local minima. Third, direct optimisation fails to use efficient available routines for common models like the negative binomial, linear, logit, or Poisson regression. We propose the following estimation algorithm.

### An Iterated Minimisation Scheme

In the following algorithm, we propose an iterative scheme that reduces direct optimisation to  $\boldsymbol{\delta}$  while minimisation with respect to  $\boldsymbol{\beta}$  becomes a standard regression problem that is easily handled using available routines.

**Algorithm 1.**

- (a) *Initialise counter  $l = 0$  and starting values  $\hat{\boldsymbol{\delta}}^{(l)}$*
- (b) *Construct design matrix  $\mathbf{m}^{(l)} = [\mathbf{1}_{NT}, \mathbf{x}, \mathbf{w}(\hat{\boldsymbol{\delta}}^{(l)})\mathbf{x}, \mathbf{z}]$  and update counter  $l \leftarrow l + 1$*
- (c) *Estimate  $\hat{\boldsymbol{\beta}}^{(l)} = \arg \min_{\boldsymbol{\beta} \in \mathcal{B}} \mathcal{Q}_n(\boldsymbol{\beta}, \hat{\boldsymbol{\delta}}^{(l-1)})$ , i.e. regress  $\mathbf{y}$  on  $\mathbf{m}^{(l)}$*

- (d) Solve  $\hat{\boldsymbol{\delta}}^{(l)} = \arg \min_{\boldsymbol{\delta} \in \boldsymbol{\Delta}} \mathcal{Q}_n(\hat{\boldsymbol{\beta}}^{(l)}, \boldsymbol{\delta})$   
(e) If  $\mathcal{Q}_n(\hat{\boldsymbol{\beta}}^{(l)}, \hat{\boldsymbol{\delta}}^{(l-1)}) - \mathcal{Q}_n(\hat{\boldsymbol{\beta}}^{(l)}, \hat{\boldsymbol{\delta}}^{(l)}) \leq \epsilon$ , stop else return to step (b)

$\epsilon$  is a small user-specified number, e.g.,  $\epsilon = 10^{-6}$ . Constructing the design matrix involves plugging  $\hat{\boldsymbol{\delta}}^{(l)}$  into  $\boldsymbol{w}(\cdot)$  with a pre-specified  $\boldsymbol{D}$  and updating the column  $\boldsymbol{w}(\hat{\boldsymbol{\delta}}^{(l)})\mathbf{x}$  in  $\boldsymbol{m}^{(l)}$ . The ease of algorithm 1 is at step (c) where regression of the outcome is run on the design matrix obtained in step (b) using available routines. The crux of Algorithm 1 reduces to the solving for  $\hat{\boldsymbol{\delta}}^{(l)}$  at step (d) at each iteration.

## 5 Inference

This section concerns the asymptotic properties of our model. The asymptotic properties of concern are consistency and asymptotic normality. The following assumptions are introduced.

**Assumption 4.**  $\{y_{it} : 1 \leq i \leq N\}$  for  $t = 1, \dots, T$  are *i.i.d.*

Note that assumption 4 does not assume serial independence for each  $y_{it}$ ,  $t = 1, \dots, T$ , and it allows the scores to be arbitrarily serially correlated. In the following lemma, we show that  $\mathcal{Q}_n(\boldsymbol{\beta}, \boldsymbol{\delta})$  converges uniformly in probability to  $\mathcal{Q}_o(\boldsymbol{\beta}, \boldsymbol{\delta})$ .

**Lemma 5.** *Under the assumptions of lemma 3 and assumption 4,  $\mathcal{Q}_o(\boldsymbol{\beta}, \boldsymbol{\delta})$  is continuous, and  $\sup_{\boldsymbol{\beta} \in \boldsymbol{B}, \boldsymbol{\delta} \in \boldsymbol{\Delta}} |\mathcal{Q}_n(\boldsymbol{\beta}, \boldsymbol{\delta}) - \mathcal{Q}_o(\boldsymbol{\beta}, \boldsymbol{\delta})| \xrightarrow{p} 0$ .*

With the uniform convergence in probability result (lemma 5) in hand, we are now able to show consistency of  $\hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\delta}}']'$ .

**Theorem 2** (Consistency). *Under the assumptions of lemmas 1 to 5,  $[\hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\delta}}']' \xrightarrow{p} [\boldsymbol{\beta}_o', \boldsymbol{\delta}_o']'$*

Asymptotic normality relies on consistency (theorem 2) and a set of standard assumptions (see assumption 5). The following theorem establishes the asymptotic normality.



**Theorem 3** (Asymptotic Normality). *Under the assumptions of theorem 2 and the set of assumptions 5 (see appendix A),  $\sqrt{NT}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V}_\theta)$  where  $\mathbf{V}_\theta$  is the variance-covariance matrix.*

**Parameters of interest - continued:** Inference on parameters of interest viz. private, spillover, and social effects is important from a policy analysis perspective. Let  $\varrho : \boldsymbol{\Theta} \rightarrow \mathbb{R}$ ,  $\varrho(\boldsymbol{\theta}) \equiv \lim_{N,T \rightarrow \infty} \varrho_n(\boldsymbol{\theta})$  denote an aforementioned parameter of interest. The following corollary shows asymptotic normality of the parameters of interest.

**Corollary 1.** *Suppose  $\boldsymbol{\theta}_o$  is in the interior of  $\boldsymbol{\Theta}$ . By the continuous differentiability of  $\varrho(\boldsymbol{\theta})$  with respect to  $\boldsymbol{\theta} \in \boldsymbol{\Theta}$  and theorem 3,  $\sqrt{NT}(\varrho_n(\hat{\boldsymbol{\theta}}) - \varrho(\boldsymbol{\theta}_o)) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \varrho'(\boldsymbol{\theta}_o)' \mathbf{V}_\theta \varrho'(\boldsymbol{\theta}_o))$  where  $\varrho'(\cdot)$  denotes  $k_\theta \times 1$  the Jacobian matrix.*

## 6 Empirical Application

Quantifying knowledge spillovers of R&D is an important research pursuit because of its policy implications. In this section, we estimate private and spillover effects of R&D on firm innovation. This problem is studied in Bloom, Schankerman, and Van Reenen (2013), Cincera (1997), and Grillitsch and Nilsson (2015).

### 6.1 Data

Data for our empirical analyses are an updated version of data used in Bloom, Schankerman, and Van Reenen (2013).<sup>14</sup> A balanced panel is constructed out of the original data set hence we do not replicate the exact results in Bloom, Schankerman, and Van Reenen (2013). The balanced panel comprises 217 firms in 21 two-digit SIC industries from 1985 to 2011. Table 1 presents the number of firms by 2-digit industry classification (SIC2). Prominent among the industries are the Electronic & Other Electric Equipment (38 firms), Chemicals

---

<sup>14</sup>The updated version are used in Lucking, Bloom, and Van Reenen (2018). The data and codes are accessible at <https://nbloom.people.stanford.edu/research>.

and Allied Products (34 firms), Industrial Machinery & Equipment (33 firms), and Instruments & Related Products (31 firms). Majority of the firms (95%) are in the manufacturing sector.

Table 1: Firms by 2-digit SIC

<b>SIC2</b>	<b>Industry</b>	<b>no. of firms</b>
13	Oil and Gas Extraction	2
14	Nonmetallic minerals, except fuels	1
20	Food and Kindred Products	6
24	Lumber and Wood Products	1
25	Furniture and Fixtures	7
26	Paper and Allied Products	5
27	Printing and Publishing	1
28	Chemicals and Allied Products	34
29	Petroleum and Coal Products	1
30	Rubber & Misc. Plastics Products	4
32	Stone, Clay, and Glass Products	3
33	Primary Metal Industries	4
34	Fabricated Metal Products	13
35	Industrial Machinery & Equipment	33
36	Electronic & Other Electric Equipment	38
37	Transport Equipment	23
38	Instruments & Related Products	31
39	Misc. Manufacturing Industries	3
50	Wholesale Trade - Durable Goods	1
73	Business Services	5
99	Unclassified	1

Table 2 presents summary statistics of the main variables. The outcome variable is citation-weighted patent counts *Pat\_Cite*. As Aghion, Van Reenen, and Zingales (2013) argued, citation-weighted patent counts not only capture the quantity of R&D output (patent counts) but also the acknowledged relevance (citations). *Pat\_Cite* has a mass point at zero as about 43% of the observations have zero citation-weighted patent counts. Patent counts (Patent flow) is zero for about 38% of the firm-year observations. The main covariate of interest is R&D (R&D stock); it generates spillovers. Bloom, Schankerman, and Van Reenen (2013)'s exogenously constructed measures of spillovers,  $\ln(SpTECH)$  (in technology

Table 2: Summary Statistics

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std. Dev.
Pat_Cite	0	0	6	453.641	158	5714	1218.84
Patent flow	0	0	3	80.113	30	4344	267.668
R&D stock	0	37.513	181.609	2253.656	1302.627	47296.664	5663.882
$\ln(SpTECH)$	3.649	5.538	5.919	5.95	6.305	7.933	0.65
$\ln(SpSIC)$	4.813	5.193	5.495	5.449	5.709	5.977	0.308
$\ln(SpGEOG)$	1.625	4.748	5.232	5.364	5.757	9.256	1.005
Sales	0.827	356.386	1514.999	11131.897	6748.933	1176922.25	51194.824

*Notes:* All values above are based on firm-year observations. The outcome variable is citation-weighted patent counts (Pat.Cite). Other variables include patent counts (Patent flow), stock of R&D (R&D stock), spillover measure using technology proximity  $\ln(SpTECH)$ , spillover measure using product market proximity  $\ln(SpSIC)$ , spillover measure using geographical proximity  $\ln(SpGEOG)$ , real sales revenue (Sales).  $\ln(SpTECH)$ ,  $\ln(SpSIC)$ , and  $\ln(SpGEOG)$  are constructed à la Bloom, Schankerman, and Van Reenen (2013) using the Jaffe (1986) measures related to technology, product market, and geographic distances with the log of lagged R&D stock for the 217 firms in the balanced panel. Prices are measured in millions of 1996 dollars.

space) and  $\ln(SpSIC)$  (in product market space), and  $\ln(SpGEOG)$  (using geographical location) are included for comparison. To control for firm size, we include log real sales revenue<sup>15</sup>  $\ln(Sales)$ . Pre-sample mean scaling approach  $Pre - SFE$ , which conditions on pre-sample citation-weighted patents, is included in our specifications to estimate fixed effects à la Blundell, Griffith, and Van Reenen (1999). We include 3-digit SIC industry and year dummies in all specifications to control for unobserved heterogeneity.

## 6.2 Empirical model

We consider the negative binomial model because the outcome variable (citation-weighted patent counts) is discrete. While there exist alternatives to the negative binomial model for handling outcome variables with mass points at zero (see discussion in Berger, Stocker, and Zeileis (2017) and Huang and Tsyawo (2018)), the crux of our empirical application lies in eliciting spatial interactions and estimating spillover effects in a possibly non-linear model. Our model is applicable to a wide class of models which admit linear predictor functions, and

<sup>15</sup>Nominal sales are deflated by industry price indices to obtain real sales.

since this class encompasses a number of methods used in the literature viz. zero-inflated negative binomial, zero-inflated Poisson, hurdle models, Tobit, Heckman selection, logit, distribution regression, and quantile regression, our approach remains applicable.

The conditional mean is given by

$$(6.1) \quad E(y_{it}|\mathbf{x}_t, \mathbf{z}_{it}, \mathbf{D}_t) = \exp(\rho_0 + \rho_1 x_{it} + \rho_2 \sum_{j \neq i} x_{jt} w_{ijt}(\boldsymbol{\delta}) + \mathbf{z}_{it} \boldsymbol{\gamma})$$

where the exponential link function is used for the negative binomial model,  $\rho_1$  and  $\rho_2$  respectively denote the private and spillover elasticities of research and development (*R&D*) on firm innovation.  $x_{it}$  denotes firm  $i$ 's *R&D* (lagged log *R&D* stock) and  $\mathbf{z}_{it}$  contains firm characteristics viz. lagged log real sales, a dummy variable for observations with zero lagged *R&D* stock, a dummy for observations with lagged patent stock equal to zero, a “pre-sample mean-scaling approach” to estimate fixed effects à la Blundell, Griffith, and Van Reenen (1999), 3-digit industry dummies and year dummies. In some specifications, spillover measures  $\ln(SpTECH)$ ,  $\ln(SpSIC)$ , and  $\ln(SpGEOG)$  are included. Standard errors are clustered at the firm level.

Spillover covariates in  $\mathbf{d}_{ijt} \in \mathbf{D}_t$  considered are the Jaffe measures in technology ( $dTECH$ ), product market ( $dSIC$ ), and geographical proximity ( $dGEOG$ ), 6 bivariate polynomial expansion terms of  $R\&D_{i,t-2}, (dR\&D_{j,t-2})$ ,  $j \neq i$  up to the third order, a dummy equal to one if firms  $i, j$  are in the same 3-digit industry classification ( $dI(SIC3)$ ), a dummy equal to one if one (but not both firms) has zero lagged *R&D* stock  $dI(R\&D)$ , and a dummy equal to one if one (but not both firms) has a lagged patent stock equal to zero  $dI(Cite)$ . The polynomial series terms on  $R\&D_{i,t-2}$  and  $R\&D_{j,t-2}$  is given by a vector comprising  $\{R\&D_{i,t-2}^{l_i} R\&D_{j,t-2}^{l_j} : l_i = 0, \dots, k-1, l_j = 1, \dots, k-l_i\}$  and  $k = 3$ .

### 6.3 Results

The main results are summarised in tables 3 and 4. Since the main covariate  $\mathbf{x}$  of interest is in logarithms, the coefficients  $\rho_1$  and  $\rho_2$  are interpretable as elasticities. The coefficient  $\rho_1$  on  $\ln(R\&D)_{t-1}$  denotes the elasticity of own R&D to innovation whereas spillover elasticity of R&D using our model is captured by the coefficient  $\rho_2$ . Bloom, Schankerman, and Van Reenen (2013)’s exogenously constructed spillover measures are included in specifications (1) and (7) in table 3 and in specification (5) in table 4. Columns (2)-(7) in table 3 and columns (1)-(5) in table 4 are specifications using our approach with different combinations of spatial covariates  $\mathbf{d}_{ijt}$  in the weight function  $w_{ijt}(\cdot)$ . The main specifications of interest are columns (6) and (7) in table 3; these combine all spatial covariates in the weight function and allows us to assess all spatial covariates simultaneously.

In each table, panel A reports coefficients with standard errors in parentheses. The standard errors are clustered by firm to allow serial correlation. Panel B reports Wald-statistics (chi-square statistic) corresponding to the spatial covariate (subscript to  $\chi^2$ ), the degree of freedom  $df$ , and the p-value  $P(> \chi^2)$  of the chi-square statistic (in parentheses).

Results comparable to tables 3 and 4 can be found in Bloom, Schankerman, and Van Reenen (2013, table IV) and Lucking, Bloom, and Van Reenen (2018, table 3). Specification (1) includes Bloom, Schankerman, and Van Reenen (2013)’s exogenously constructed spillover measures, namely,  $\ln(SpTECH)_{t-1}$ ,  $\ln(SpSIC)_{t-1}$ , and  $\ln(SpGEOG)_{t-1}$  using the balanced panel. Out of the three spillover measures,  $\ln(SpGEOG)_{t-1}$  is positive and significant at the 5% level whereas  $\ln(SpTECH)_{t-1}$  and  $\ln(SpSIC)_{t-1}$  are not significant at any of the conventional levels. One makes a similar observation in specification (7) of table 3 and specification (5) of table 4 where the coefficient on  $\ln(SpGEOG)$  is significant at the 1% level (in both specifications). Observe that the coefficient on  $\ln(SpTECH)$  is negative (but not significant) in specification (7) of table 3 while the coefficient on  $\ln(SpSIC)$  is consistently negative.

Considering results in tables 3 and 4, one observes that the elasticity of own R&D on in-

Table 3: Coefficients - Citation-Weighted Patent Counts I

	Neg. Bin (1)	Neg. Bin (2)	Neg. Bin (3)	Neg. Bin (4)	Neg. Bin (5)	Neg. Bin (6)	Neg. Bin (7)
Panel A: ( $\beta$ )							
$\rho_1$	0.155 (0.044)	0.187 (0.042)	0.176 (0.043)	0.186 (0.042)	0.181 (0.041)	0.170 (0.041)	0.157 (0.045)
$\rho_2$		0.048 (0.402)	-1.331 (1.173)	1.566 (0.534)	0.213 (0.070)	0.319 (0.067)	0.325 (0.067)
$\ln(SpTECH)_{t-1}$	0.059 (0.126)						-0.016 (0.124)
$\ln(SpSIC)_{t-1}$	-2.818 (2.084)						-0.939 (2.225)
$\ln(SpGEOG)_{t-1}$	0.068 (0.029)						0.085 (0.031)
$\ln(Pat)_{t-1}$	0.551 (0.034)	0.560 (0.035)	0.555 (0.033)	0.552 (0.035)	0.584 (0.034)	0.563 (0.033)	0.559 (0.033)
$Pre - SFE$	0.172 (0.038)	0.147 (0.038)	0.153 (0.039)	0.164 (0.040)	0.164 (0.039)	0.180 (0.039)	0.195 (0.040)
Panel B: ( $\delta$ )							
$\chi^2_{dTECH}$		0.210				0.94	1.000
df, $P(> \chi^2)$		1,(0.650)				1,(0.330)	1,(0.310)
$\chi^2_{dSIC}$			122.200			10.600	13.400
df, $P(> \chi^2)$			1,(0.000)			1,(0.001)	1,(0.000)
$\chi^2_{dGEOG}$				34.100		14.000	12.000
df, $P(> \chi^2)$				1,(0.000)		1,(0.000)	1,(0.001)
$\chi^2_{dR\&D_{t-2}}$					190.700	50.000	49.100
df, $P(> \chi^2)$					6,(0.000)	6,(0.000)	6,(0.000)
$\chi^2_{dI(SIC3)}$						1.000	0.910
df, $P(> \chi^2)$						1,(0.310)	1,(0.340)
$\chi^2_{dI(R\&D)}$						0.76	0.620
df, $P(> \chi^2)$						1,(0.380)	1,(0.430)
$\chi^2_{dI(Cite)}$						5.000	5.200
df, $P(> \chi^2)$						1,(0.025)	1,(0.022)

Notes: All specifications above (Neg. Bin (1)-(5)) contain firm and year fixed effects. The outcome variable is citation-weighted patent counts ( $Pat\_Cite_t$ ). Number of firm-year observations: 5859. The spillover-generating variable  $\mathbf{x}$  is lagged log R&D  $\ln(R\&D)_{t-1}$ . Spillover covariates (with respective Wald-statistics) considered: Jaffe measures in technology ( $dTECH$ ), product market ( $dSIC$ ), and geographical proximity ( $dGEOG$ ), 6 bivariate polynomial expansion terms of  $R\&D_{i,t-2}$ , ( $dR\&D_{j,t-2}$ ),  $j \neq i$  up to the third order, a dummy equal to one if firms  $i, j$  are in the same 3-digit industry classification ( $dI(SIC3)$ ), a dummy equal to one if one (but not both firms) has zero lagged R&D stock  $dI(R\&D)$ , and a dummy equal to one if one (but not both firms) has a lagged patent stock equal to zero  $dI(Cite)$ .  $P(> \chi^2)$  denotes the p-value and df equals the length of the corresponding parameter vector.

novation is positive and significant at the 1% level across all model specifications. Relative to specification (1), the elasticity of own R&D ( $\rho_1$ ) in other specifications (where our approach is used) is slightly higher in magnitude save specification (5) in table 4. The corresponding standard errors do not change much across specifications in both tables.

The coefficient  $\rho_2$  in specifications (2) through (5) in table 3 and specifications (2)-(5) in table 4 gives the spillover elasticity of *R&D* on firm innovation using our model. Notice that the sign and magnitude vary by spatial covariate(s) included in the weight function.  $\rho_2$  is negative in specifications (3) of table 3 and (2)-(5) of table 4 where *dSIC*, *dI(SIC3)*, *dI(R&D)*, and *dI(Cite)* are spatial covariates included in the weight function. *dTECH*, *dGEOG*, *dR&D<sub>t-2</sub>* in specifications (2), (3)-(5) spatial covariates associated with positive  $\rho_2$ . Save specifications (2)-(3) in table 3 and (2)-(3) in table 4,  $\rho_2$  is statistically significant at the 1% level across specifications. One also notes variation in the standard errors of  $\rho_2$  by the spatial covariate included in the weight function. The significance of  $\rho_2$  is also indicative of the presence of spillovers and cross-sectional dependence in firm patenting behaviour. Coefficients on the lagged outcome variable  $\ln(Pat)_{t-1}$  are positive and significant at the 1% level across all specifications which confirms strong persistence in firm patenting behaviour as in Bloom, Schankerman, and Van Reenen (2013) and Lucking, Bloom, and Van Reenen (2018).

The significance of coefficients  $\boldsymbol{\delta}$  on the spatial covariates  $\mathbf{d}_{ijt}$  are indicative of support against the a null hypothesis that each element of  $\mathbf{w}(\boldsymbol{\delta})$  is equal to  $1/(N - 1)$  where  $N$  is the number of firms.<sup>16</sup> Rejecting this hypothesis does, however, not indicate the relevance of the spatial covariate as long as the outcome is concerned. It ought to be coupled with  $\rho_2$  for the relevance of the spatial covariate to be confirmed by the data. In this vein, one notes that while  $\boldsymbol{\delta}$  on *dSIC* in table 3(3), *dI(SIC3)* in table 4(2), and *dI(R&D)* table 4(3) are statistically significant at the 1% level, the corresponding  $\rho_2$  are not significant at any conventional level.

---

<sup>16</sup> $w_{ijt}(\boldsymbol{\delta}) \equiv \frac{\exp(\mathbf{d}'_{ijt}\boldsymbol{\delta})}{\sum_{j \neq i}^N \exp(\mathbf{d}'_{ijt}\boldsymbol{\delta})} = 1/(N - 1)$  if  $\boldsymbol{\delta} = \mathbf{0}$ ,  $\forall i = 1, \dots, N$ ,  $t = 1, \dots, T$

Table 4: Coefficients - Citation-Weighted Patent Counts II

	Neg. Bin (1)	Neg. Bin (2)	Neg. Bin (3)	Neg. Bin (4)	Neg. Bin (5)
Panel A: ( $\beta$ )					
$\rho_1$	0.141 (0.045)	0.167 (0.046)	0.187 (0.042)	0.175 (0.042)	0.151 (0.045)
$\rho_2$	-0.235 (0.045)	-1.518 (1.169)	-0.056 (0.191)	-0.227 (0.047)	-0.228 (0.046)
$\ln(SpTECH)_{t-1}$	0.044 (0.132)				0.053 (0.129)
$\ln(SpSIC)_{t-1}$	-0.567 (2.328)				-0.993 (2.290)
$\ln(SpGEOG)_{t-1}$	0.077 (0.029)				0.077 (0.028)
$\ln(Pat)_{t-1}$	0.556 (0.034)	0.561 (0.035)	0.560 (0.035)	0.557 (0.033)	0.550 (0.033)
$Pre - SFE$	0.151 (0.036)	0.143 (0.039)	0.147 (0.039)	0.132 (0.036)	0.147 (0.036)
Panel B: ( $\delta$ )					
$\chi^2_{dI(SIC3)}$	12.400	92.500			
df, $P(> \chi^2)$	1,(0.000)	1,(0.000)			
$\chi^2_{dI(R\&D)}$	81.100		17.400		
df, $P(> \chi^2)$	1,(0.000)		1,(0.000)		
$\chi^2_{dI(Cite)}$	9.200			19.000	21.600
df, $P(> \chi^2)$	1,(0.002)			1,(0.000)	1,(0.000)

Notes: All specifications above (Neg. Bin (1)-(5)) contain firm and year fixed effects. The outcome variable is citation-weighted patent counts ( $Pat\_Cite_t$ ). Number of firm-year observations: 5859. The spillover-generating variable  $\mathbf{x}$  is lagged log R&D  $\ln(R\&D)_{t-1}$ . Spillover covariates (with respective Wald-statistics) considered: a dummy equal to one if firms  $i, j$  are in the same 3-digit industry classification ( $dI(SIC3)$ ), a dummy equal to one if one (but not both firms) has zero lagged R&D stock  $dI(R\&D)$ , and a dummy equal to one if one (but not both firms) has a lagged patent stock equal to zero  $dI(Cite)$ .  $P(> \chi^2)$  denotes the p-value and df equals the length of the corresponding parameter vector.



Specifications (6) and (7) in table 3 allows a comparison of all spatial covariates simultaneously. One notes that elements of  $\delta$  corresponding to  $dTECH$ ,  $dSIC3$ , and  $dI(R\&D)$  are not significant. Those on  $dSIC$ ,  $dGEOG$ ,  $dR\&D_{t-2}$  are significant at the 1% level while the coefficient in  $\delta$  on  $dI(Cite)$  is significant at the 5% level.  $\rho_2$  in specifications (6) and (7) are positive and significant at the 1% level, confirming that R&D spillover effects on firm innovation are positive and the relevance of  $dSIC$ ,  $dGEOG$ ,  $dR\&D_{t-2}$  and  $dI(Cite)$  as spatial covariates. In the relevant set of spatial covariates,  $dR\&D_{t-2}$  and  $dI(Cite)$  are time-varying and  $dR\&D_{t-2}$  is asymmetric.

In sum, our estimation (specification (6) in table 3) confirms positive private spillover effect of R&D on innovation. A 1% increase in own R&D is associated with a 0.170% increase in innovation. The private effect is dominated by the spillover effect of innovation. If every other firm increases R&D by 1%, innovation is expected to increase by 0.319%. In addition to the significance of the spillover parameter  $\rho_2$ , the relevant set of spatial covariates are associated with proximity in the product market space ( $dSIC$ ), geographic proximity ( $dGEOG$ ), past R&D ( $dR\&D_{t-2}$ ), and past patent citations  $dI(Cite)$ . ( $dR\&D_{t-2}$ ) and  $dI(Cite)$  are not tied to any measure of proximity (e.g. geographic contiguity) and are time-varying.

## 7 Conclusion

The problem of quantifying R&D spillover effects on firm innovation among other outcomes crucially depends on the spatial matrix that is used to construct the R&D spillover measure. In this paper, we address the problem of unknown spatial matrices in a study of R&D spillover effects on firm innovation using the negative binomial model. The study extends to other studies that quantify spillover effects. As a key contribution of the paper, the proposed approach parsimoniously models and estimates the spatial matrix from panel data. The approach allows time variation and asymmetry in the spatial matrix and extends

to a general class of linear and non-linear models that admit linear predictor functions. Our approach tackles the problem of metric ambiguity by accommodating several spillover covariates simultaneously and assesses their relevance via hypothesis tests. We provide identification results of the model and establish consistency and asymptotic normality of the estimator. We apply our method to study R&D spillovers on innovation in a knowledge production framework. The importance of our approach is confirmed by the sensitivity of spillover effect estimates to the choice of spatial covariate and spatial matrix by extension. Our results confirm the presence of positive and statistically significant spillover effects of R&D on innovation. Though the private effects of R&D on innovation are positive and statistically significant, they are dominated by the spillover effects of R&D. In a specification that assesses the relevance of several spatial covariates, we find that product market proximity, geographic proximity, past R&D, and past patent citations are relevant sources of connectivity between firms with respect to innovation. The result on relevant spillover covariates confirm time-variation and asymmetry in the interaction structure between firms and that the strength of connectivity between firms is not only tied to commonly assumed notions of closeness viz. technological, product market, industry, and geographic proximity.

## References

- [1] Acemoglu, Daron, Philippe Aghion, Claire Lelarge, John Van Reenen, and Fabrizio Zilibotti. “Technology, Information, and the Decentralization of the Firm”. *The Quarterly Journal of Economics* 122.4 (2007), pp. 1759–1799.
- [2] Aghion, Philippe, John Van Reenen, and Luigi Zingales. “Innovation and institutional ownership”. *American economic review* 103.1 (2013), pp. 277–304.
- [3] Anselin, Luc, Julie Le Gallo, and Hubert Jayet. “Spatial panel econometrics”. *The econometrics of panel data*. Springer, 2008, pp. 625–660.
- [4] Berger, Susanne, Herbert Stocker, and Achim Zeileis. “Innovation and institutional ownership revisited: an empirical investigation with count data models”. *Empirical Economics* 52.4 (2017), pp. 1675–1688.
- [5] Bernstein, Jeffrey I and M Ishaq Nadiri. “Research and development and intra-industry spillovers: an empirical application of dynamic duality”. *The Review of Economic Studies* 56.2 (1989), pp. 249–267.
- [6] Bhattacharjee, Arnab and Chris Jensen-Butler. “Estimation of the spatial weights matrix under structural constraints”. *Regional Science and Urban Economics* 43.4 (2013), pp. 617–634.
- [7] Bloom, Nicholas, Mark Schankerman, and John Van Reenen. “Identifying technology spillovers and product market rivalry”. *Econometrica* 81.4 (2013), pp. 1347–1393.
- [8] Blundell, Richard, Rachel Griffith, and John Van Reenen. “Market share, market value and innovation in a panel of British manufacturing firms”. *The Review of Economic Studies* 66.3 (1999), pp. 529–554.
- [9] Boyd, Stephen and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

- [10] Case, Anne C. “Spatial patterns in household demand”. *Econometrica: Journal of the Econometric Society* (1991), pp. 953–965.
- [11] Cincera, Michele. “Patents, R&D, and technological spillovers at the firm level: some evidence from econometric count models for panel data”. *Journal of Applied Econometrics* 12.3 (1997), pp. 265–280.
- [12] Coe, David T, Elhanan Helpman, and Alexander W Hoffmaister. “International R&D spillovers and institutions”. *European Economic Review* 53.7 (2009), pp. 723–741.
- [13] Cohen, Wesley M and Daniel A Levinthal. “Innovation and learning: the two faces of R & D”. *The economic journal* 99.397 (1989), pp. 569–596.
- [14] Dou, Baojun, Maria Lucia Parrella, and Qiwei Yao. “Generalized Yule–Walker estimation for spatio-temporal models with unknown diagonal coefficients”. *Journal of Econometrics* 194.2 (2016), pp. 369–382.
- [15] Griffith, Rachel, Stephen Redding, and John Van Reenen. “Mapping the two faces of R&D: Productivity growth in a panel of OECD industries”. *Review of economics and statistics* 86.4 (2004), pp. 883–895.
- [16] Grillitsch, Markus and Magnus Nilsson. “Innovation in peripheral regions: Do collaborations compensate for a lack of local knowledge spillovers?” *The Annals of Regional Science* 54.1 (2015), pp. 299–321.
- [17] Hall, Bronwyn H, Adam Jaffe, and Manuel Trajtenberg. “Market value and patent citations”. *RAND Journal of Economics* (2005), pp. 16–38.
- [18] Hoshino, Tadao. “Semiparametric Estimation of Censored Spatial Autoregressive Models”. *Econometric Theory* (2017), pp. 1–38.
- [19] Huang, Weige and Emmanuel Tsyawo. “Bayesian Distribution Regression”. *Available at SSRN 3048658* (2018).

- [20] Jaffe, Adam B. *Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market value*. 1986.
- [21] Kapetanios, George, James Mitchell, and Yongcheol Shin. "A nonlinear panel data model of cross-sectional dependence". *Journal of Econometrics* 179.2 (2014), pp. 134–157.
- [22] Kelejian, Harry H and Gianfranco Piras. "Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes". *Regional Science and Urban Economics* 46 (2014), pp. 140–149.
- [23] Lucking, Brian, Nicholas Bloom, and John Van Reenen. *Have R&D Spillovers Changed?* Tech. rep. National Bureau of Economic Research, 2018.
- [24] Lychagin, Sergey, Joris Pinkse, Margaret E Slade, and John Van Reenen. "Spillovers in space: Does geography matter?" *The Journal of Industrial Economics* 64.2 (2016), pp. 295–335.
- [25] Manresa, Elena. "Estimating the structure of social interactions using panel data". *Unpublished Manuscript. CEMFI, Madrid* (2013).
- [26] Newey, Whitney K and Daniel McFadden. "Large sample estimation and hypothesis testing". *Handbook of Econometrics* 4 (1994), pp. 2111–2245.
- [27] Pinkse, Joris and Margaret E Slade. "Mergers, brand competition, and the price of a pint". *European Economic Review* 48.3 (2004), pp. 617–643.
- [28] Pinkse, Joris, Margaret E Slade, and Craig Brett. "Spatial price competition: a semi-parametric approach". *Econometrica* 70.3 (2002), pp. 1111–1153.
- [29] Qu, Xi and Lung-fei Lee. "Estimating a spatial autoregressive model with an endogenous spatial weight matrix". *Journal of Econometrics* 184.2 (2015), pp. 209–232.
- [30] Soale, Abdul-Nasah and Emmanuel Tsyawo. "Clustered Covariate Regression". *Available at SSRN: <https://ssrn.com/abstract=3394012>* (2019).

- [31] Souza, PC. “Estimating network effects without network data”. *PUC-Rio Working Paper* (2014).
- [32] Syverson, Chad. “What determines productivity?” *Journal of Economic Literature* 49.2 (2011), pp. 326–65.
- [33] Wooldridge, Jeffrey M. *Econometric analysis of cross section and panel data*. MIT press, 2010.
- [34] Xu, Xingbai and Lung-fei Lee. “A spatial autoregressive model with a nonlinear transformation of the dependent variable”. *Journal of Econometrics* 186.1 (2015), pp. 1–18.

## A Proofs

**Proof of lemma 1.** Without loss of generality, and for ease of exposition, use the column-permuted  $\mathbf{m}$  as  $\mathbf{m} = [\mathbf{z}, \mathbf{w}\mathbf{x}]$ . The Gram matrix of the design matrix  $\mathbf{m} = [\mathbf{z}, \mathbf{w}\mathbf{x}]$  is

$$\mathbf{m}'\mathbf{m} = \begin{bmatrix} \mathbf{z}'\mathbf{z} & \mathbf{z}'\mathbf{w}\mathbf{x} \\ \mathbf{x}'\mathbf{w}'\mathbf{z} & \mathbf{x}'\mathbf{w}'\mathbf{w}\mathbf{x} \end{bmatrix}$$

Note that under assumption 1(a),  $\mathbf{z}'\mathbf{z}$  is invertible. By assumption 1(b), the block diagonal matrix  $\mathbf{w}$  is full rank. Using the result in Boyd and Vandenberghe (2004, sect A.5.5) on the positive definiteness of symmetry matrices, we only need to show that the Shur complement of  $\mathbf{z}'\mathbf{z}$  in  $\mathbf{m}'\mathbf{m}$ ,  $S = \mathbf{x}'\mathbf{w}'\mathbf{w}\mathbf{x} - \mathbf{x}'\mathbf{w}'\mathbf{z}(\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}'\mathbf{w}\mathbf{x} > 0$ .

Let us factor  $S$  as  $S = \mathbf{x}'\mathbf{w}'[\mathbf{I}_{NT} - \mathbf{z}(\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}']\mathbf{w}\mathbf{x}$ . Note that  $\mathbf{z}(\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}'$  is symmetric and idempotent;  $[\mathbf{I}_{NT} - \mathbf{z}(\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}']$  is also symmetric and idempotent. Since  $\mathbf{x} \neq \mathbf{0}$  (assumption 1(a)), it implies  $S$  can be expressed as  $S = \mathbf{x}'\mathbf{w}'[\mathbf{I}_{NT} - \mathbf{z}(\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}'][\mathbf{I}_{NT} - \mathbf{z}(\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}']\mathbf{w}\mathbf{x} = \sum_{i=1}^N \sum_{t=1}^T \zeta_{it}^2 > 0$  where  $\zeta_{it}$ ,  $i = 1, \dots, N$ ,  $t = 1, \dots, T$  are residuals obtained by regressing  $\mathbf{w}\mathbf{x}$  on  $\mathbf{z}$ , i.e.,  $S$  is the sum of squared residuals. The positive definiteness of  $\mathbf{m}'\mathbf{m}$  implies  $\mathbf{m}$  is full column rank.  $\square$

**Proof of lemma 2.**

First, we need to show that  $Nw_{ijt}(\boldsymbol{\delta}) = \frac{N \exp(\mathbf{d}'_{ijt}\boldsymbol{\delta})}{\sum_{j \neq i}^N \exp(\mathbf{d}'_{ijt}\boldsymbol{\delta})}$  is bounded for all  $\boldsymbol{\delta} \in \boldsymbol{\Delta}$ .

$$\begin{aligned} Nw_{ijt}(\boldsymbol{\delta}) &= \frac{\exp(\mathbf{d}'_{ijt}\boldsymbol{\delta})}{N^{-1} \sum_{j \neq i}^N \exp(\mathbf{d}'_{ijt}\boldsymbol{\delta})} \leq \frac{\exp(\mathbf{d}'_{ijt}\boldsymbol{\delta})}{\exp(N^{-1} \sum_{j \neq i}^N \mathbf{d}'_{ijt}\boldsymbol{\delta})} \\ &= \exp((\mathbf{d}_{ijt} - \bar{\mathbf{d}}_{i,t})'\boldsymbol{\delta}) \leq \sup_{\boldsymbol{\delta} \in \boldsymbol{\Delta}} \exp((\mathbf{d}_{ijt} - \bar{\mathbf{d}}_{i,t})'\boldsymbol{\delta}) \leq \kappa_w \end{aligned}$$

The first equality re-expresses  $Nw_{ijt}(\boldsymbol{\delta})$ . The first inequality follows by Jensen's inequality since the  $\exp(\cdot)$  is convex, and  $N^{-1} \sum_{j \neq i}^N \exp(\mathbf{d}'_{ijt}\boldsymbol{\delta}) \geq \exp(N^{-1} \sum_{j \neq i}^N \mathbf{d}'_{ijt}\boldsymbol{\delta})$ . The last two inequalities follows from the sup operator and assumption 3(a).

Second, we know from the first part that  $Nw_{ijt}(\boldsymbol{\delta}) \leq \kappa_w$ . This implies that for  $j \in \{1, \dots, N\}$  and  $t \in \{1, \dots, T\}$ ,

$$\sum_{i=1}^N w_{ijt}(\boldsymbol{\delta}) \leq \max_{\substack{j \in \{1, \dots, N\} \\ t \in \{1, \dots, T\}}} \sum_{i=1}^N w_{ijt}(\boldsymbol{\delta}) \equiv \|\mathbf{w}(\boldsymbol{\delta})\|_1 \leq N \sup_{\boldsymbol{\delta} \in \boldsymbol{\Delta}} \max_{\substack{j \in \{1, \dots, N\} \\ t \in \{1, \dots, T\}}} w_{ijt}(\boldsymbol{\delta}) \leq N \frac{\kappa_w}{N} = \kappa_w$$

Note that by definition,  $\|\mathbf{w}(\boldsymbol{\delta})\|_1 \equiv \max_{\substack{j \in \{1, \dots, N\} \\ t \in \{1, \dots, T\}}} \sum_{i=1}^N w_{ijt}(\boldsymbol{\delta})$  because  $\mathbf{w}(\boldsymbol{\delta})$  is block-diagonal. □

### ***Proof of lemma 3.***

For a given  $\eta^2 > 0$ , write

$$\begin{aligned} |q_{it}(\boldsymbol{\beta}, \boldsymbol{\delta})| &= |-\eta^{-2} \log[\eta^{-2}/(\eta^{-2} + \exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}))] - y_{it} \log[\exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta})/(\eta^{-2} + \exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}))]| \\ &\leq |\eta^{-2} \log \eta^{-2}| + |(\eta^{-2} + y_{it}) \log(\eta^{-2} + \exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}))| + |y_{it} m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}| \\ &\leq |\eta^{-2} \log \eta^{-2}| + |(\eta^{-2} + y_{it})(\eta^{-2} - 1 + \exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}))| + |y_{it} m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}| \\ &\leq |\eta^{-2} \log \eta^{-2}| + |(\eta^{-2} + y_{it})(\eta^{-2} - 1)| + |(\eta^{-2} + y_{it}) \exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta})| + |y_{it} m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}| \\ &\equiv q_1 + q_2 + q_3 + q_4 \end{aligned}$$

The first and third inequalities follow from the triangle inequality, and the second follows from the inequality on the natural logarithm  $\log u \leq u - 1$ .

Note that under assumption 3(b), the conditional mean is bounded, i.e.,  $E(y_{it}|\cdot) = \exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}) \leq \exp(|m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}|) \leq \sqrt{\kappa_y} < \infty$

Now the following bounds obtain on the terms  $q_1$ ,  $q_2$ ,  $q_3$ , and  $q_4$ .

$q_1 \equiv |\eta^{-2} \log \eta^{-2}|$  is bounded because  $\eta^2$  is finite.

$$\begin{aligned} E(q_2) &\equiv |(\eta^{-2} + y_{it})(\eta^{-2} - 1)| = |(\eta^{-2} - 1)|E(\eta^{-2} + y_{it}) = |(\eta^{-2} - 1)|(\eta^{-2} + E(y_{it})) = \\ &|(\eta^{-2} - 1)|(\eta^{-2} + E[E(y_{it}|\cdot)]) = |(\eta^{-2} - 1)|(\eta^{-2} + E \exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta})) \leq |(\eta^{-2} - 1)|(\eta^{-2} + \sqrt{\kappa_y}) < \infty \end{aligned}$$

$$\begin{aligned} E(q_3) &= E|(\eta^{-2} + y_{it}) \exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta})| \leq \eta^{-2} E \exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}) + E(y_{it} \exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta})) = \eta^{-2} E \exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}) + \\ &E(y_{it} E(y_{it}|\cdot)) = \eta^{-2} E \exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}) + E(E(y_{it} E(y_{it}|\cdot))|\cdot) = \eta^{-2} E \exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}) + E(E(y_{it}|\cdot)^2) = \end{aligned}$$



$$\eta^{-2}E \exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}) + E \exp(2m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}) \leq \eta^{-2}\sqrt{\kappa_y} + \kappa_y < \infty$$

$$E(q_4) = E|y_{it}m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}| = E(E(|y_{it}m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}||\cdot)) = E(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}E(y_{it}|\cdot)) = (E(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta})^2)^{1/2}(E(E(y_{it}|\cdot))^2)^{1/2} \\ 1/2\sqrt{\kappa_y} \log \kappa_y < \infty$$

Combining terms shows that  $E|q_{it}(\boldsymbol{\beta}, \boldsymbol{\delta})| < \infty$  from which it follows that  $E|\mathcal{Q}_n(\boldsymbol{\beta}, \boldsymbol{\delta})| \leq \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T E|q_{it}(\boldsymbol{\beta}, \boldsymbol{\delta})| \leq |\eta^{-2} \log \eta^{-2}| + |(\eta^{-2} - 1)|(\eta^{-2} + \sqrt{\kappa_y}) + \eta^{-2}\sqrt{\kappa_y} + \kappa_y + 1/2\sqrt{\kappa_y} \log \kappa_y < \infty$   $\square$

**Proof of lemma 4.**  $\mathbf{w}(\boldsymbol{\delta})$  is non-singular (assumption 1(b)) which implies that  $\mathbf{w}(\boldsymbol{\delta}) \neq \mathbf{w}(\boldsymbol{\delta}_o)$  if  $\boldsymbol{\delta} \neq \boldsymbol{\delta}_o$ . Since  $\mathbf{x} \neq \mathbf{0}$  (assumption 1(a)),  $\mathbf{m}(\boldsymbol{\delta}) \neq \mathbf{m}(\boldsymbol{\delta}_o)$  if  $\boldsymbol{\delta} \neq \boldsymbol{\delta}_o$ . By lemma 1,  $\mathbf{m}(\boldsymbol{\delta})$  is full rank which implies  $\mathbf{m}(\boldsymbol{\delta})\boldsymbol{\beta} \neq \mathbf{m}(\boldsymbol{\delta}_o)\boldsymbol{\beta}_o$  if  $\boldsymbol{\delta} \neq \boldsymbol{\delta}_o$  and  $\boldsymbol{\beta} \neq \boldsymbol{\beta}_o$ .

Note that  $\mathcal{Q}_n(\boldsymbol{\beta}, \boldsymbol{\delta}) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T q_{it}(\boldsymbol{\beta}, \boldsymbol{\delta}) = \eta^{-2} \log \eta^{-2} + \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T ((\eta^{-2} + y_{it}) \log(\eta^{-2} + \exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta})) - y_{it}m_{it}(\boldsymbol{\delta})\boldsymbol{\beta})$ . Because the function  $f(u) = \log(\eta^{-2} + \exp(u))$  is monotone in  $u$ ,  $f(\mathbf{m}(\boldsymbol{\delta})\boldsymbol{\beta}) \neq f(\mathbf{m}(\boldsymbol{\delta}_o)\boldsymbol{\beta}_o)$  if  $\boldsymbol{\delta} \neq \boldsymbol{\delta}_o$  and  $\boldsymbol{\beta} \neq \boldsymbol{\beta}_o$ . From the foregoing,  $\mathcal{Q}_n(\boldsymbol{\beta}, \boldsymbol{\delta}) \neq \mathcal{Q}_n(\boldsymbol{\beta}_o, \boldsymbol{\delta}_o)$  if  $\boldsymbol{\delta} \neq \boldsymbol{\delta}_o$  and  $\boldsymbol{\beta} \neq \boldsymbol{\beta}_o$  and  $\boldsymbol{\theta}_o = [\boldsymbol{\beta}'_o, \boldsymbol{\delta}'_o]'$  is identified.  $\square$

**Proof of theorem 1.** Under the assumptions of lemma 1,  $\boldsymbol{\theta}_o = [\boldsymbol{\beta}'_o, \boldsymbol{\delta}'_o]'$  is identified. Also, under the assumptions of lemma 3,  $E|\mathcal{Q}_n(\boldsymbol{\beta}, \boldsymbol{\delta})| < \infty$ . The conclusion follows from Newey and McFadden (1994, lemma 2.2).  $\square$

**Proof of lemma 5.** Proving uniform convergence in probability of  $\mathcal{Q}_n(\boldsymbol{\beta}, \boldsymbol{\delta})$  to  $\mathcal{Q}_o(\boldsymbol{\beta}, \boldsymbol{\delta})$  requires the verification of the following conditions.

First, the continuity of  $q_{it}(\boldsymbol{\beta}, \boldsymbol{\delta}) = -\eta^{-2} \log[\eta^{-2}/(\eta^{-2} + \exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}))] - y_{it} \log[\exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta})/(\eta^{-2} + \exp(m_{it}(\boldsymbol{\delta})\boldsymbol{\beta}))]$  at each  $[\boldsymbol{\beta}', \boldsymbol{\delta}']' \in \mathbf{B} \times \boldsymbol{\Delta}$  holds by inspection.

Second, from lemma 3(see proof),  $E|q_{it}(\boldsymbol{\beta}, \boldsymbol{\delta})| \leq |\eta^{-2} \log \eta^{-2}| + |(\eta^{-2} - 1)|(\eta^{-2} + \sqrt{\kappa_y}) + \eta^{-2}\sqrt{\kappa_y} + \kappa_y + 1/2\sqrt{\kappa_y} \log \kappa_y < \infty$  for all  $[\boldsymbol{\beta}', \boldsymbol{\delta}']' \in \mathbf{B} \times \boldsymbol{\Delta}$

Coupled with assumptions 2 and 4, the conclusion follows from Newey and McFadden (1994, lemma 2.4).  $\square$

**Proof of theorem 2.** The proof of consistency requires the verification of the following conditions.

First, under the assumptions of theorem 1,  $\mathcal{Q}_o(\boldsymbol{\beta}, \boldsymbol{\delta})$  is uniquely minimised at  $\boldsymbol{\theta}_o = [\boldsymbol{\beta}'_o, \boldsymbol{\delta}'_o]'$ .

Second, under the assumptions of lemma 5,  $\mathcal{Q}_o(\boldsymbol{\beta}, \boldsymbol{\delta})$  is continuous and  $\sup_{\boldsymbol{\beta} \in \mathbf{B}, \boldsymbol{\delta} \in \boldsymbol{\Delta}} |\mathcal{Q}_n(\boldsymbol{\beta}, \boldsymbol{\delta}) - \mathcal{Q}_o(\boldsymbol{\beta}, \boldsymbol{\delta})| \xrightarrow{p} 0$ .

Coupled with assumption 2, the conclusion follows from Newey and McFadden (1994, theorem 2.1).  $\square$

**Proof of theorem 3.** The following additional assumptions are needed for asymptotic normality.

**Assumption 5.** (a)  $\boldsymbol{\theta}_o$  is in the interior of  $\boldsymbol{\Theta}$  (b) The score function  $s_n(\boldsymbol{\beta}, \boldsymbol{\delta}) \equiv \frac{\partial \mathcal{Q}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$  is continuously differentiable on the interior of  $\boldsymbol{\Theta}$ . (c) Each element of  $H_n(\boldsymbol{\theta})$  is bounded in absolute value. (d)  $\mathbf{A}_o \equiv E[H_n(\boldsymbol{\theta})]$  is positive definite (e)  $E[s_n(\boldsymbol{\beta}, \boldsymbol{\delta})] = \mathbf{0}$  (f)  $\mathbf{B}_o \equiv E[s_{it}(\boldsymbol{\beta}, \boldsymbol{\delta})s_{it}(\boldsymbol{\beta}, \boldsymbol{\delta})']$  is positive definite.

Under the assumptions of theorem 2 and assumption 5, the conclusion follows from Wooldridge (2010, theorem 12.3) where  $\mathbf{V}_\theta \equiv \mathbf{A}_o^{-1}\mathbf{B}_o\mathbf{A}_o^{-1}$ .  $\square$

**Proof of corollary 1.**  $\sqrt{NT}(\varrho_n(\hat{\boldsymbol{\theta}}) - \varrho(\boldsymbol{\theta}_o)) = \sqrt{NT}(\varrho_n(\hat{\boldsymbol{\theta}}) - \varrho(\hat{\boldsymbol{\theta}})) + \sqrt{NT}(\varrho(\hat{\boldsymbol{\theta}}) - \varrho(\boldsymbol{\theta}_o))$ . The first term converges to zero in probability. Continuous differentiability allows a Taylor expansion of the second term and applying the delta method  $\sqrt{NT}(\varrho(\hat{\boldsymbol{\theta}}) - \varrho(\boldsymbol{\theta}_o)) = \varrho'(\boldsymbol{\theta}_o)\sqrt{NT}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o) + o_p(1) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \varrho'(\boldsymbol{\theta}_o)'\mathbf{V}_\theta\varrho'(\boldsymbol{\theta}_o))$   $\square$