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

Emmanuel S. Tsyawo‡

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 it allows asymmetry and time-variation in the spatial matrix. On firm innovation, we find strategic substitutability in firm R&D effort, negative spillover effects, and positive private effects.

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 Research Grant.

‡Department of Economics, Temple University. Email: estsyawo@temple.edu

1 Introduction

Quantifying the spillover effects of research and development (R&D) investment 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 (Grillitsch and Nilsson, 2015; Bloom, Schankerman, and Van Reenen, 2013; Cincera, 1997). Firms engage in R&D not only to create knowledge but also to boost their capacity 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 spillovers are varied. Knowledge spillovers have policy implications and quantifying them aids to fine-tune policy,[1] 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 quantifying R&D spillover effects.

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 time-varying spatial matrices that can allow several sources of spillovers simultaneously. This avoids making an arbitrary choice of metric, spatial matrix, or both

[1] 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.

and does not impose symmetry.[2] We are able to identify sources of spillovers and conduct hypothesis tests on their significance in order to establish their relevance. Identification of spillover sources 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)).[3] A second approach employs industry distance to the frontier[4] 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,

[2] Bhattacharjee and Jensen-Butler (2013) imposes time-invariance and symmetry on the spatial matrix for identification. By flexibly modelling the elements of the spatial matrix, we avoid imposing symmetry and time-invariance on the spatial matrix.

[3] Cincera (1997) pools R&D investment at the manufacturing sector level and only allows intra-sector spillovers.

[4] 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; Tsyawo, 2019). 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[5] 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

[5] 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.[6]

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) using the SAR framework on limited dependent variables. These approaches, however, suppose foreknowledge of the spatial matrix and lack direct generalisability[7] to other non-linear models. Innovation is measured using citation-weighted patent counts (e.g., Hall, Jaffe, and Trajtenberg (2005) and Aghion, Van Reenen, and Zingales (2013)), 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). Except TFP, non-linear models viz. the Poisson, negative binomial, and logit models are required for estimation.

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 sources of spillovers, our framework avoids endogeneity in the spatial matrix.[8] Our model adopts a semi-parametric approach to estimating spatial matrices. Semi-parametric estimation is often a good choice in the presence of metric ambiguity or where the notion of "metric" is not guided by theory.[9]

Our empirical application studies spillover effects of R&D investment on firm innovation. Firm innovation is measured by citation-weighted patent counts and R&D is measured by the firm's research and development stock. Due to the discreteness of the outcome variable,

[6] 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.

[7] 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.

[8] 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.

[9] See, for example, Pinkse, Slade, and Brett (2002) and Pinkse and Slade (2004).

the Poisson model is used in estimation. We find that innovation mediates substitutability in firm R&D efforts. We also find statistically positive private effects and negative spillover effects of R&D on firm innovation. Like Bloom, Schankerman, and Van Reenen (2013), we identify a strong persistence in firm patenting behaviour.

The rest of the paper is organised as follows. Section 2 describes the model and presents a general identification result. Section 3 covers estimation algorithms and section 4 presents asymptotic results. The empirical study is conducted in section 5, and section 6 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 including linear, quantile, logit, and probit regressions. Conditional functionals of interest include conditional means, conditional quantile (quantile regression), conditional probability (distribution regression), and conditional variance (ARCH-type models).

2.1 Scope of model

To set the scope of the model, consider a baseline functional of the distribution of outcome y_{it} conditional on observables

$$\nu(y_{it}|\mathbf{x}_t, \mathbf{z}_{it}, \boldsymbol{\xi}_{t-1}) = g(\rho_0 + \rho_1 x_{it} + \rho_2 \sum_{j\neq i} x_{jt} w_{ij}(\phi(\xi_{i,t-1}, \xi_{j,t-1})) + \mathbf{z}_{it}\boldsymbol{\gamma}) \tag{2.1}$$

where $g : \mathbb{R} \to \mathbb{R}$ is known and differentiable, $w_{ij}(\phi(\xi_{i,t-1}, \xi_{j,t-1})) = \dfrac{\exp(\phi(\xi_{i,t-1}, \xi_{j,t-1}))}{\sum_{j\neq i}^{N} \exp(\phi(\xi_{i,t-1}, \xi_{j,t-1}))}$, $\sum_{j\neq i}^{N} w_{ij}(\phi(\xi_{i,t-1}, \xi_{j,t-1})) = 1 \ \forall\ i = 1, \ldots, N,\ t = 1, \ldots, T,\ \phi : \mathbb{R}^2 \to \mathbb{R}$ may be unknown.[10]

[10]Unknown $g(\cdot)$ is permissible in our framework, and this suggests applying single-index methods. We, however, do not pursue it. The function $\exp(\cdot)$ can be replaced by any twice differentiable function $f : \mathbb{R} \to \mathbb{R}_+$.

The following variables are observed: outcome y_{it}, covariates $\mathbf{z}_{it}$, $\mathbf{x}_t = [x_{1t}, \ldots, x_{Nt}]'$ generates spillovers (e.g., R&D), and $\boldsymbol{\xi}_{t-1} = [\xi_{1,t-1}, \ldots, \xi_{N,t-1}]$ constitute sources of spillovers (e.g. proximity in technology space). ξ_{it} can be considered as a source of spillovers or spatial dependence in general. For instance, innovation of firms may benefit from past patents of other firms. In our empirical application (section 5), $\phi(\cdot)$ is estimated semi-parametrically using polynomial series methods. The formulation eq. (2.1) allows the outcome of agent i, to be impacted, not only by own characteristics $\mathbf{z}_{it}$ and x_{it}, but also by the characteristics of others x_{jt}, $j \neq i$.

For subsequent reference, let us write out the design matrix in eq. (2.1). Define the following: $\mathbf{z}$ is an $NT \times k-1$ whose rows comprise $\mathbf{z}_{11}, \ldots, \mathbf{z}_{N1}, \ldots, \mathbf{z}_{N1}, \ldots, \mathbf{z}_{NT}$, an $NT \times NT$ block diagonal matrix $\boldsymbol{w} = diag[\boldsymbol{w}_1(\boldsymbol{\delta}), \ldots, \boldsymbol{w}_T(\boldsymbol{\delta})]$ where $\boldsymbol{w}_t(\boldsymbol{\delta}) = [w_{ij}(\phi(\xi_{i,t-1}, \xi_{j,t-1}; \boldsymbol{\delta}))]$, $i \neq j$ and zeroes on the diagonal is an $N \times N$ matrix, and $\mathbf{x} = [x_{11}, \ldots, x_{N1}, \ldots, x_{N1}, \ldots, x_{NT}]'$. The design matrix associated with eq. (2.1) is given by $\boldsymbol{m} = [\mathbf{1}_{NT}, \mathbf{x}, \boldsymbol{w}\mathbf{x}, \mathbf{z}]$ where $\mathbf{1}_{NT}$ is an NT vector of ones. Note that $\boldsymbol{w}\mathbf{x}$ is $NT \times 1$. We suppress the dependence of $\boldsymbol{w}$ on $\boldsymbol{\delta}$ for notational convenience.

Generally, the function $\phi(\cdot, \cdot)$ is not assumed known. The only requirement is that it be expressible in terms of observables $\boldsymbol{\xi}$. Notice that our formulation of the spatial function $w_{ij}(\cdot)$ does admit a number of cases. For instance, $\phi(\xi_{i,t-1}, \xi_{j,t-1}) = -\delta d(\xi_{i,t-1}, \xi_{j,t-1})$ for a metric $d(\cdot, \cdot)$ obtains the formulation (equation 22) in Kapetanios, Mitchell, and Shin (2014). Using observed links in network data, for example in the peer effects literature (equation (1) of Bramoullé, Djebbari, and Fortin (2009)) with $w_{ij} = 1/n_i$ if j is a friend of i and zero otherwise. Suppose i and j have several characteristics ($\boldsymbol{\xi}_{i,t-1}$ and $\boldsymbol{\xi}_{j,t-1}$) whose interaction drive spatial dependence in outcomes, then, $\phi : \mathbb{R}^{2k_\xi} \rightarrow \mathbb{R}$ defined on $\boldsymbol{\xi}_i \in \mathbb{R}^{k_\xi}$ accommodates multiple sources of spatial dependence.

Models we focus on admit linear predictor functions as in eq. (2.1). Examples include linear regression, quantile regression, logit, Poisson, and negative binomial models. As and when model specificity becomes necessary, we illustrate results using specific models. In general, we assume a generic objective function $\mathcal{Q}_n(\cdot)$ that spans the class of models under consideration.

2.2 The objective function

In the rest of the paper, we partition the parameter space $\boldsymbol{\Theta} \subset \mathbb{R}^{k_\beta+k_\delta}$ into $\boldsymbol{B} \subset \mathbb{R}^{k_\beta}$ and $\boldsymbol{\Delta} \subset \mathbb{R}^{k_\delta}$. $\boldsymbol{\beta} = [\rho_0, \rho_1, \rho_2, \boldsymbol{\gamma}']' \in \boldsymbol{B}$, $\boldsymbol{\delta} \in \boldsymbol{\Delta}$, $\boldsymbol{\Theta} \equiv \boldsymbol{B} \times \boldsymbol{\Delta}$, and $\boldsymbol{\theta} \equiv [\boldsymbol{\beta}', \boldsymbol{\delta}']'$, where $\boldsymbol{\delta}$ comprises parameters associated with $\phi(\cdot,\cdot)$ and $\boldsymbol{\beta}$ includes all other parameters in the model. 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

$$[\hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\delta}}']' = \underset{[\boldsymbol{\beta}', \boldsymbol{\delta}']' \in \boldsymbol{\Theta}}{\arg\min}\, \mathcal{Q}_n(\boldsymbol{\beta}, \boldsymbol{\delta}) \tag{2.2}$$

$q_{it}(\boldsymbol{\beta}, \boldsymbol{\delta})$ is the contribution of unit i at time t to the objective function $\mathcal{Q}_n(\cdot,\cdot)$.

Examples - models: Examples of models in the class under consideration include linear regression $q_{it}(\boldsymbol{\beta}, \boldsymbol{\delta}) = (y_{it} - \rho_0 - \rho_1 x_{it} - \rho_2 \sum_{j\neq i} x_{jt} w_{ij}(\phi(\xi_{i,t-1}, \xi_{j,t-1}; \boldsymbol{\delta})) - \mathbf{z}_{it}\boldsymbol{\gamma})^2$, the logit model $q_{it}(\boldsymbol{\beta}, \boldsymbol{\delta}) = -(\rho_0 + \rho_1 x_{it} + \rho_2 \sum_{j\neq i} x_{jt} w_{ij}(\phi(\xi_{i,t-1}, \xi_{j,t-1}; \boldsymbol{\delta})) + \mathbf{z}_{it}\boldsymbol{\gamma})y_{it} - log(1 + \exp(\rho_0 + \rho_1 x_{it} + \rho_2 \sum_{j\neq i} x_{jt} w_{ij}(\phi(\xi_{i,t-1}, \xi_{j,t-1}; \boldsymbol{\delta})) + \mathbf{z}_{it}\boldsymbol{\gamma}))$, and the Poisson model $q_{it}(\boldsymbol{\beta}, \boldsymbol{\delta}) = -y_{it}(\rho_0 + \rho_1 x_{it} + \rho_2 \sum_{j\neq i} x_{jt} w_{ij}(\phi(\xi_{i,t-1}, \xi_{j,t-1}; \boldsymbol{\delta})) + \mathbf{z}_{it}\boldsymbol{\gamma}) + \exp(\rho_0 + \rho_1 x_{it} + \rho_2 \sum_{j\neq i} x_{jt} w_{ij}(\phi(\xi_{i,t-1}, \xi_{j,t-1}; \boldsymbol{\delta})) + \mathbf{z}_{it}\boldsymbol{\gamma})$. The inclusion of $\boldsymbol{\delta}$ in $\phi(\cdot)$ allows flexible modelling of the weight function $w_{ij}(\cdot)$ across observations and through time up to a parameter vector of finite length k_δ.

If $\boldsymbol{\xi}_{t-1}$ contains at least one time-varying variable, $\boldsymbol{w}_t$ is time-varying and the structure of spatial dependence need not be stable through time. A feature like this one can be assessed statistically by testing the significance of elements in $\boldsymbol{\delta}$ corresponding to time-varying $\boldsymbol{\xi}_{t-1}$.

2.3 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 sources of spillovers as well as an assessment of time-variation in the spatial matrix $\boldsymbol{w}$. 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 spillovers 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 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.

Spillovers in demand behaviour: Spillovers in demand due to geographically varying prices, e.g., regional housing demand (Bhattacharjee and Jensen-Butler, 2013), rice demand in Indonesia (Case, 1991), cigarette demand Kelejian and Piras, 2014 among other forms of spatial dependence in demand behaviour have been identified and widely studied. Markets are not perfectly segregated and shocks in any spatial unit are transmitted and impact demand behaviour in adjacent areas. Our model can contribute to this literature by allowing flexibility (e.g. time-variation, asymmetry, and multiple metrics) in the determination of spatial matrices.

2.4 Identification

Full column rank of the design matrix is not sufficient for identification across the entire class of models of interest. We provide it as a necessary condition for all models under consideration. We make the following assumptions.

Assumption 1. *(a)* $\boldsymbol{z} = [\mathbf{1}_{NT}, \mathbf{x}, \mathbf{z}]$ *is full column rank. (b) Each spatial matrix* $\boldsymbol{w}_t(\boldsymbol{\delta})$ *is non-singular,* $t = 1, .., T$ *and for all* $\boldsymbol{\delta} \in \boldsymbol{\Delta}$.[11] *(c) The diagonal elements of* $\boldsymbol{w}_t(\boldsymbol{\delta})$ *are zeroes,* $t = 1, .., T$ *and for all* $\boldsymbol{\delta} \in \boldsymbol{\Delta}$

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

Proposition 1 (Full column rank of design matrix $\boldsymbol{m}$)**.** *Under assumption 1(a-c), the design matrix* $\boldsymbol{m}'\boldsymbol{m}$ *is positive definite.*

Examples - models continued: In the case of linear regression, proposition 1 is necessary but not sufficient. Consider the linear model $y_{it} = \rho_0 + \rho_1 x_{it} + \rho_2 \sum_{j \neq i} x_{jt} w_{ij}(\phi(\xi_{i,t-1}, \xi_{j,t-1})) + \mathbf{z}_{it}\boldsymbol{\gamma} + u_{it}$. For a given $\boldsymbol{\delta}$ that satisfies assumption 1(b), $\hat{\boldsymbol{\beta}}(\boldsymbol{\delta}) = [\boldsymbol{m}(\boldsymbol{\delta})'\boldsymbol{m}(\boldsymbol{\delta})]^{-1}\boldsymbol{m}(\boldsymbol{\delta})'\mathbf{y}$ with emphasis on the dependence on $\boldsymbol{\delta}$. In addition to proposition 1, identification requires $E(\mathbf{u}|\mathbf{x}, \mathbf{z}, \boldsymbol{\xi_{t-1}}) = \mathbf{0}$, and that $E[\mathbf{u}(\boldsymbol{\delta})'\mathbf{u}(\boldsymbol{\delta})]$ be minimised uniquely by $\boldsymbol{\delta}_o$. The Poisson model, in our context, is not as analytical as linear regression. Express the objective function as $\mathcal{Q}_n(\boldsymbol{\beta}, \boldsymbol{\delta}) = \mathbf{y}'\boldsymbol{m}(\boldsymbol{\delta})\boldsymbol{\beta} - \exp(\boldsymbol{m}(\boldsymbol{\delta})\boldsymbol{\beta})$. $\hat{\boldsymbol{\beta}}$ solves the first order condition $\frac{\partial \mathcal{Q}_n(\boldsymbol{\beta}, \boldsymbol{\delta})}{\partial \boldsymbol{\beta}} := \frac{1}{NT}\boldsymbol{m}(\boldsymbol{\delta})'[\mathbf{y} - \exp(\boldsymbol{m}(\boldsymbol{\delta})\boldsymbol{\beta})] = \mathbf{0}$. $\boldsymbol{m}(\boldsymbol{\delta})$ is column full rank (proposition 1), and it implies that for $\boldsymbol{\beta} \neq \boldsymbol{\beta}_o$, $\boldsymbol{m}(\boldsymbol{\delta})\boldsymbol{\beta} \neq \boldsymbol{m}(\boldsymbol{\delta})\boldsymbol{\beta}_o$, and by the strict monotonicity of the expo-

[11] This assumption is not strong because $\boldsymbol{\Delta}$ can be defined as small as possible such that the sequence of optimal $\boldsymbol{\delta}_m$, $m \geq n$ is contained and converges in $\boldsymbol{\Delta}$.

nential function $\exp(\cdot)$, $\exp(\boldsymbol{m}(\boldsymbol{\delta})\boldsymbol{\beta}) \neq \exp(\boldsymbol{m}(\boldsymbol{\delta})\boldsymbol{\beta}_o)$. With $\boldsymbol{\beta}$ defined as a function of $\boldsymbol{\delta}$ and under proposition 1, identification requires that $\mathcal{Q}_o(\boldsymbol{\beta}(\boldsymbol{\delta}), \boldsymbol{\delta})$ be minimised at a unique $\boldsymbol{\delta}_o$.

2.5 Parameters of interest

We explore parameters that are interesting from a policy perspective. Though these parameters can have varied interpretation by application (see section 2.3 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 investment in year t.

Private effect: The private effect of $R\&D$ measures the impact of a firm's R&D investment on its innovation. $\mathcal{PE}_i = T^{-1}\sum_{t=1}^{T} \frac{\partial \nu(y_{it}|\cdot)}{\partial x_{it}} = T^{-1}\sum_{t=1}^{T} \mathcal{PE}_{it} = \rho_1 T^{-1}\sum_{t=1}^{T} g'(\boldsymbol{m}_{it}(\boldsymbol{\delta})\boldsymbol{\beta})$ where $g'(\cdot)$ denotes the derivative of $g(\cdot)$. 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 ρ_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. Note that for linear $g(\cdot)$, $\mathcal{PE} = \mathcal{PE}_i$ for all $i = 1, \ldots, N$.

Spillover effect: The spillover effect measures the impact of firm j's R&D investment 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 $\mathcal{S}p\mathcal{E}_{ijt} = \frac{\partial \nu(y_{it}|\cdot)}{\partial x_{jt}} = \rho_2 w_{ijt} g'(\boldsymbol{m}_{it}(\boldsymbol{\delta})\boldsymbol{\beta})$ where $w_{ijt} \equiv w_{ij}(\phi(\xi_{i,t-1}, \xi_{j,t-1}; \boldsymbol{\delta}))$. 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 $\mathcal{S}p\mathcal{E}_{ij.} = T^{-1}\sum_{t=1}^{T} \mathcal{S}p\mathcal{E}_{ijt} = \rho_2 T^{-1}\sum_{t=1}^{T} w_{ijt} g'(\boldsymbol{m}_{it}(\boldsymbol{\delta})\boldsymbol{\beta})$. The spillover effect of other firms' R&D on firm i's innovation obtains as $\mathcal{S}p\mathcal{E}_{i..} = \sum_{j\neq i} \mathcal{S}p\mathcal{E}_{ij.}$, the spillover effect exerted by firm j on other firms' innovation is $\mathcal{S}p\mathcal{E}_{.j.} = \sum_{i\neq j} \mathcal{S}p\mathcal{E}_{ij.}$, and the average spillover effects across all firms is $\mathcal{S}p\mathcal{E} = N^{-1}\sum_{i=1}^{N} \mathcal{S}p\mathcal{E}_{i..} = N^{-1}\sum_{j=1}^{N} \mathcal{S}p\mathcal{E}_{.j.}$. A researcher interested in time-variation of average spillover effects can compute $\mathcal{S}p\mathcal{E}_{..t} = N^{-1}\sum_{i=1}^{N}\sum_{j\neq i} \mathcal{S}p\mathcal{E}_{ijt}$ for each $t \in \{1, \ldots, T\}$.

Social effects: Social (marginal) effects sum private and spillover effects. The social marginal effects generated by firm j's R&D investment at time t is $\mathcal{SE}_{.jt} = \sum_{i=1}^{N} \frac{\partial \nu(y_{it}|\cdot)}{\partial x_{jt}} = \frac{\partial \nu(y_{jt}|\cdot)}{\partial x_{jt}} + \sum_{i \neq j} \frac{\partial \nu(y_{it}|\cdot)}{\partial x_{jt}} = \mathcal{PE}_{jt} + \sum_{i \neq j} \mathcal{S}p\mathcal{E}_{ijt} = \mathcal{PE}_{jt} + \mathcal{S}p\mathcal{E}_{.jt}$. In a similar vein, the social effects of R&D (from all firms) received by firm i at time t is defined as $\mathcal{SE}_{i.t} = \sum_{j=1}^{N} \frac{\partial \nu(y_{it}|\cdot)}{\partial x_{jt}} = \frac{\partial \nu(y_{it}|\cdot)}{\partial x_{jt}} + \sum_{j \neq i} \frac{\partial \nu(y_{it}|\cdot)}{\partial x_{jt}} = \mathcal{PE}_{it} + \sum_{j \neq i} \mathcal{S}p\mathcal{E}_{ijt} = \mathcal{PE}_{it} + \mathcal{S}p\mathcal{E}_{i.t}$. Average social effects across firms and time obtains as $\mathcal{SE} = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathcal{SE}_{i.t} = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathcal{SE}_{.jt}$.[12]

Strategic complementarity/substitutability of R&D With respect to the outcome y_{it} (innovation), we are able to evaluate and test the relationship between R&D efforts of firms. For any pair of firms i, j at time t,

$$\frac{\partial x_{it}}{\partial x_{jt}} = \frac{\partial x_{it}}{\partial \nu(y_{it}|\cdot)} \frac{\partial \nu(y_{it}|\cdot)}{\partial x_{jt}} = \frac{\mathcal{S}p\mathcal{E}_{ijt}}{\mathcal{PE}_{it}}$$

For a given firm i, vis-à-vis all other firms, the relationship is summarised by $\sum_{j \neq i} \frac{\partial x_{it}}{\partial x_{jt}} = \mathcal{PE}_{it}^{-1} \sum_{j \neq i} \mathcal{S}p\mathcal{E}_{ijt}$. As a summary measure for the entire set of firms $j = 1, \ldots, N$, across time $t = 1, \ldots, T$, the parameter of interest is $\mathcal{SCS} = (NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j \neq i} \frac{\partial x_{it}}{\partial x_{jt}} = (NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} \mathcal{PE}_{it}^{-1} \sum_{j \neq i} \mathcal{S}p\mathcal{E}_{ijt}$. With respect to innovation, $\mathcal{SCS} < 0$ means firm R&D efforts are strategic substitutes whereas $\mathcal{SCS} > 0$ means firm R&D efforts are strategic complements.

Strength of spatial dependence: ρ_2 controls the strength of the spatial dependence in our model; it can accommodate both weak and strong forms of spatial dependence[13] depending on the magnitude of ρ. 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.

[12] Comparable approaches to computing structural parameters viz. private, spillover, and social effects are available in Bloom, Schankerman, and Van Reenen (2013, Sect. 6.5.1) and Manresa (2013, Sect. 5.5).

[13] This feature is also present in Kapetanios, Mitchell, and Shin (2014)'s model.

Sources of R&D spillovers: By modelling the spatial matrix $\boldsymbol{w}$ as a function of observable sources of spillovers $\boldsymbol{\xi}$ up to a parameter vector $\boldsymbol{\delta}$ of finite length k_δ, our approach allows the evaluation of factors that drive spatial dependence i.e., test elements of $\boldsymbol{\delta}$ that are statistically significant. In a similar vein, the strength of time-variation in the interaction structure is assessable by testing the significance (jointly or individually) of elements in $\boldsymbol{\delta}$ that correspond to time-varying elements in $\boldsymbol{\xi}$.[14]

2.6 Extensions

The methodology we propose is not without drawbacks. Our approach requires metrics or sources of spillovers and spatial dependence. The elements of the spatial matrix are non-negative.[15] Also, the number of elements in the spatial matrix to be modelled grows in the order of $O(N^2T)$. Without further adjustment, our baseline model does not accommodate sparse interaction structures well. We consider solutions to these challenges in extensions to the baseline model eq. (2.1). Our approach to modelling spatial dependence offers a gamut of extensions that tackle drawbacks of the baseline model. Though we do not explore the computational and inferential challenges associated with extensions to the baseline model eq. (2.1), we state the following.

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

$$\nu(y_{it}|\mathbf{z}_i, \boldsymbol{\xi}_{t-1}, \mathbf{x}_t) = g(\rho_0 + \rho_1 y_{i,t-1} + \rho_2 \sum_{j\neq i} y_{j,t-1} w_{ij}(\phi(\xi_{i,t-1}, \xi_{j,t-1}; \boldsymbol{\delta})) + \mathbf{z}_{it}\boldsymbol{\gamma})$$

$\nu(y_{it}|\mathbf{z}_i, \boldsymbol{\xi}_{t-1}, \mathbf{x}_t)$ above generalises Kapetanios, Mitchell, and Shin (2014)'s model. Like Cincera (1997, Sect. 4), dynamics in the spillover parameter can be accommodated by including

[14] A work that conducts a similar exercise is Lychagin, Pinkse, Slade, and Reenen (2016) which finds that R&D spillovers via geographic and technological proximity matter.

[15] Soale and Tsyawo (2019), Pinkse, Slade, and Brett (2002), Bhattacharjee and Jensen-Butler (2013), and Manresa (2013) allow for both negative and positive weights.

lagged x_{it}, x_{jt} and lagged spatial matrices in order to capture dynamics in, for example, the patenting process.

$$\nu(y_{it}|\mathbf{z}_i, \boldsymbol{\xi}, \mathbf{x}) = g(\rho_0 + \rho_1 x_{it} + \rho_2 \sum_{j \neq i} x_{jt} w_{ij}(\phi(y_{i,t-1}, y_{j,t-1}; \boldsymbol{\delta})) + \mathbf{z}_{it}\boldsymbol{\gamma})$$

Where the lag in impact goes beyond one, more lags can be included as long as there are sufficient degrees of freedom to estimate the model.

Multiple metrics: A particularly useful feature of our model lies in its ability to handle metric ambiguity by incorporating several metrics in $\phi(\cdot,\cdot)$ at once.

$$w_{ij}(\xi_{jt}) = \frac{\exp(\xi_{jt})}{\sum_{j=1}^{N} \exp(\xi_{jt})} \tag{2.3}$$

where $\xi_{jt} = \boldsymbol{d}_{ij}'\boldsymbol{\delta}$ and $\boldsymbol{d}_{ij}$ is a vector of metrics (or transformations thereof) between i and j. 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 the relevant metrics are uncertain, a formulation like eq. (2.3) allows for a data-driven determination of the relevant subset of metrics via hypothesis tests.

Prior information in spatial matrix: Where sparsity in the spatial matrix is theoretically justifiable, binaries on neighbours, or geographic distance can be incorporated in the spatial matrix to induce sparsity. For instance, w_{ij} can be weighted by $C(d_{ij}) = \exp(-\delta_d d_{ij})$ where d_{ij} denotes the distance between i and j and δ_d can be estimated from the data (see Lychagin, Pinkse, Slade, and Reenen (2016, Sect. IV)). Alternatively, one can consider

$C(d_{ij}) = 1$ if i is in the same region as j and zero otherwise.[16]

Group heterogeneity: In some cases, it is plausible to assume spillovers among economic agents at a higher level of aggregation. This approach becomes natural if spillover generating policies are implemented at that level and it allows both within and between-group spillovers. Group heterogeneity is particularly helpful in micro data that have thousands of units and few time periods. Interactions can be allowed at group levels, say, peer groups (eg. Bramoullé, Djebbari, and Fortin (2009)), county level, state level, and industry level (Manresa, 2013). The conditional functional eq. (2.1), under group heterogeneity, is

$$\nu(y_{it}|\mathbf{z}_i, \boldsymbol{\xi}, \mathbf{x}] = g(\rho_0 + \rho_1 x_{it} + \rho_2 \sum_{j \neq i} x_{jt} w_{ij}(\phi(\boldsymbol{\xi}_{l_i,t-1}, \boldsymbol{\xi}_{l_j,t-1}; \boldsymbol{\delta})) + \mathbf{z}_{it}\boldsymbol{\gamma}) \tag{2.4}$$

where $\boldsymbol{\xi}_{l_j,t-1}$ denotes a vector of spillover sources related to agent j's group l at time $t-1$. The baseline case obtains when the number of groups equals $N-1$. This, however, leaves group membership across time unrestricted as long as group membership is observed. Formulation 2.4 has a major advantage of not allowing the dimension of $\boldsymbol{w}$ to grow *ad infinitum* as $N, T \to \infty$.

3 Estimation

The model eq. (2.1) cannot be estimated in two steps because of an intrinsic latency. Assume $\phi(\cdot,\cdot)$ is known. Since $\boldsymbol{\delta}$ is an unknown parameter vector, the $N \times N$ spatial matrices $\boldsymbol{w}_t$, $t = 1, \dots, T$ are unknown ex-ante. With $\phi(\cdot,\cdot)$ unknown, $\phi(\cdot,\cdot)$ needs to be estimated alongside $\boldsymbol{\delta}$. The foregoing suggests a joint determination of $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, and $\phi(\cdot,\cdot)$. Before proceeding, let us make the following assumption concerning the estimation of $\phi(\cdot,\cdot)$ using a series method for non-discrete elements in $\boldsymbol{\xi}$.

[16] The binary metric can be extended to capture proximity in other settings like industry, product market among others.

Assumption 2 (Series estimation of $\phi(\cdot,\cdot)$)**.** *There exists a finite $k_n \in \mathbb{N}$ such that (a) $\phi(\xi_{i,t-1}, \xi_{j,t-1}) = \sum_{l=1}^{\infty} \psi_l(\xi_{i,t-1}, \xi_{j,t-1})\delta_l \approx \sum_{l=1}^{k_n} \psi_l(\xi_{i,t-1}, \xi_{j,t-1})\delta_l \equiv \hat{\phi}(\xi_{i,t-1}, \xi_{j,t-1})$ where $\psi_l(\cdot,\cdot)$ forms a basis for $\phi(\cdot,\cdot)$'s function space. (b) $|\phi(\xi_{i,t-1}, \xi_{j,t-1}) - \hat{\phi}(\xi_{i,t-1}, \xi_{j,t-1})| = o_p(1)$*

Assumption 2 is fairly standard (see Pinkse, Slade, and Brett (2002, theorem 1), for example).

Estimation of eq. (2.2) 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, direct minimisation of $\mathcal{Q}_n(\boldsymbol{\beta}, \boldsymbol{\delta})$ becomes very slow and risks getting caught in local minima. Third, direct optimisation fails to use efficient available routines for common models like linear, logit, or Poisson regression. We propose two practical estimation techniques.

3.1 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 software.

Algorithm 1.

(a) Initialise counter $l = 0$ and starting values $\hat{\boldsymbol{\delta}}^{(l)}$

(b) Construct design matrix $\boldsymbol{m}^{(l)} = [\mathbf{1}_{NT}, \mathbf{x}, \boldsymbol{w}(\hat{\boldsymbol{\delta}}^{(l)})\mathbf{x}, \mathbf{z}]$ and update counter $l \leftarrow l+1$

(c) Estimate $\hat{\boldsymbol{\beta}}^{(l)} = \underset{\boldsymbol{\beta} \in \boldsymbol{B}}{\arg\min}\, \mathcal{Q}_n(\boldsymbol{\beta}, \hat{\boldsymbol{\delta}}^{(l-1)})$, i.e. regress $\mathbf{y}$ on $\boldsymbol{m}^{(l)}$

(d) Solve $\hat{\boldsymbol{\delta}}^{(l)} = \underset{\boldsymbol{\delta} \in \boldsymbol{\Delta}}{\arg\min}\, \mathcal{Q}_n(\hat{\boldsymbol{\beta}}^{(l)}, \boldsymbol{\delta})$

(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)

Constructing the design matrix involves plugging $\hat{\boldsymbol{\delta}}^{(l)}$ into $\hat{\phi}(\cdot,\cdot)$ for $j \neq i$. 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).

3.2 Metropolis-Hastings Algorithms

The independence and random-walk Metropolis-Hastings algorithms are useful and commonly used estimation methods in Bayesian econometrics. Bayesian inference is not crucial to our model, and we do not dwell much on it. It is, however, a desirable alternative when the presence of several nuisance parameters induces near-collinearity in the hessian matrix thereby undermining the robustness of the estimates.[17] For maximum likelihood methods, applying Bayesian inference is simple and only requires a specification of the prior density over the parameter vector $\boldsymbol{\theta}$. In applications where specifying a prior density may be challenging, a non-informative prior can be used. For other models including non-linear least squares and GMM where the criterion function $\mathcal{Q}_n(\boldsymbol{\beta}, \boldsymbol{\delta})$ in eq. (2.2) is not a log-likelihood function, quasi-bayesian methods can be used (see, for instance, Chernozhukov and Hong (2003)).

4 Inference

This section concerns shows the asymptotic properties of our model and presents a note on Bayesian inference. The properties we focus on are consistency and asymptotic normality.

4.1 Consistency and Asymptotic normality

Our model eq. (2.2) is standard and the arguments closely follow established results.

Theorem 1 (Consistency)**.** *Under standard assumptions,* $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_o$

Theorem 2 (Asymptotic Normality)**.** *Under standard assumptions,* $\sqrt{NT}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V}_\theta)$

Parameters of interest - continued: Inference on parameters of interest viz. private, spillover, and social effects is important from a policy analysis perspective. Let $\eta : \boldsymbol{\Theta} \to$

[17] This problem crops up in section 5 when we include firm and year fixed effects. Bayesian Metropolis-Hastings MCMC algorithms prove more robust.

$\mathbb{R}$, $\eta(\boldsymbol{\theta}) \equiv \lim_{N,T\to\infty} \eta_n(\boldsymbol{\theta})$ 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 $\eta(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and theorem 2, $\sqrt{NT}(\eta_n(\hat{\boldsymbol{\theta}}) - \eta(\boldsymbol{\theta}_o)) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \eta'(\boldsymbol{\theta}_o)'\mathbf{V}_\theta \eta'(\boldsymbol{\theta}_o))$ where $\eta'(\cdot)$ denotes $k \times 1$ the Jacobian matrix.*

5 Empirical Application

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

5.1 Data

Data for our empirical analyses are taken from Bloom, Schankerman, and Van Reenen (2013).[18] A balanced panel is constructed out of the original data set hence we do not recover 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 and Allied Products (34 firms), Industrial Machinery & Equipment (33 firms), and Instruments & Related Products (31 firms).

Table 2 presents summary statistics on the main variables. The outcome variable of interest 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

[18]The data and codes are accessible at `https://nbloom.people.stanford.edu/research`.

Table 1: Firms by 2-digit SIC

SIC2	**Industry**	**no. of firms**
13	Oil and Gas Extraction	2
14	Nonmetalic 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: Summary Statistics

	Pat_Cite_t	$ln(R\&D)_{t-1}$	$ln(SpTECH)_{t-1}$	$ln(SpSIC)_{t-1}$	$ln(Sales)_{t-1}$	$PreSFE$
Min.	0	-2.818	11.484	-4.377	-0.061	-0.916
1st Qu.	0	3.580	15.097	3.536	5.893	2.639
Median	6	5.160	15.725	4.607	7.496	3.984
Mean	453.641	5.234	15.650	4.485	7.313	4.071
3rd Qu.	158	7.141	16.251	5.66	8.909	5.929
Max.	5714	10.764	17.511	7.863	12.242	8.788
Std. Dev.	1218.84	2.632	0.886	1.700	2.182	2.281

Notes: The outcome variable is citation-weighted patent counts. Covariates include log stock of R&D ($ln(R\&D)_{t-1}$), the SPILLTECH measure $ln(SpTECH)_{t-1}$, the SPILLSIC measure $ln(SpSIC)_{t-1}$, log of real sales revenue $ln(Sales)_{t-1}$, Pre-sample mean scaling $PreSFE$ of Blundell, Griffith, and Van Reenen (1999). SPILLTECH and SPILLSIC are Bloom, Schankerman, and Van Reenen (2013)'s exogenously constructed spillover measures in technology and product market spaces respectively.

at zero as about 43% of the observations have zero citation-weighted patent counts. The main covariate of interest is $ln(R\&D)_{t-1}$; it generates spillovers. Bloom, Schankerman, and Van Reenen (2013)'s exogenously constructed measures of spillovers, $ln(SpTECH)_{t-1}$ (in technology space) and $ln(SpSIC)_{t-1}$ (in product market space), are also included. To control for firm size, we include log real sales revenue[19] $ln(Sales)_{t-1}$. Pre-sample mean scaling approach $PreSFE$, 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 firm and year fixed effects in all specifications to control for unobserved heterogeneity.

5.2 Empirical model

The model in our empirical model is a special case of the general model eq. (2.1) with $g(\cdot) = \exp(\cdot)$. We consider the Poisson model because the outcome variable (citation-weighted patent counts) is discrete. While the propriety of the Poisson for count data with mass points is debatable (see discussion in Berger, Stocker, and Zeileis (2017) and Huang and Tsyawo (2018)), the crux of our empirical application lies in endogenously eliciting spillover interactions and estimating spillover effects in a possibly non-linear model.[20] Our model is applicable to a wide class of models which admit linear predictor functions, and since this class encompasses a number of methods used in the literature viz. negative binomial, hurdle models, Tobit, Heckman selection, logit, distribution regression, and quantile regression, our method remains applicable.

The conditional mean of the model taken to the data is

$$E[y_{it}|\mathbf{z}_{it}, \boldsymbol{\xi}, \mathbf{x}] = \exp(\rho_0 + \rho_1 x_{it} + \rho_2 \sum_{j \neq i} x_{jt} w_{ij}(\phi(\xi_{i,t-1}, \xi_{j,t-1}; \boldsymbol{\delta})) + \alpha_i + \gamma_t + \mathbf{z}_{it}\boldsymbol{\gamma}) \tag{5.1}$$

where the exponential link function is used for the Poisson model, ρ_1 and ρ_2 respectively

[19] Nominal sales are deflated by industry price indices to obtain real sales.

[20] Cincera (1997) devotes space for an econometric treatment of models that account for the discreteness of the outcome variable – patent applications.

denote the private and spillover elasticity of research and development ($R\&D$) investment on firm innovation, α_i is a firm-specific intercept, and γ_t is a year fixed effect. x_{it}, $i \in \{1, ..., N\}$ is firm i's log R&D stock at time $t-1$ taken as the generator of spillovers. Sources of R&D spillovers $\boldsymbol{\xi}$ considered are $ln(R\&D)_{t-2}$ and industry proximity $SIC3$.[21] We use polynomial series expansion with $k(k+1)/2$ terms for $ln(R\&D)_{t-2}$ in $\hat{\phi}(\xi_{i,t-1}, \xi_{j,t-1}; \boldsymbol{\delta}) = \sum_{l_i=0}^{k_n-1} \sum_{l_j=1}^{k_n-i} \delta_{l_i l_j} \xi_{i,t-1}^{l_i} \xi_{j,t-1}^{l_j}$ and k_n is determined using the Schwarz Bayesian Criterion (SBC) subject to non-singularity of the spatial matrix $\boldsymbol{w}$.

5.3 Results

Results comparable to table 3 can be found in Bloom, Schankerman, and Van Reenen (2013, table IV, column (3)). Column (1) replicates Bloom, Schankerman, and Van Reenen (2013, table IV, column (3)) using the Poisson model on our balanced panel. The requirement of non-singularity is relatively higher in our model, and MLE in the presence of both firm and year fixed effects lacks robustness. We circumvent this hurdle by implementing the Independence Metropolis-Hastings algorithm for Bayesian estimation and inference in table 3 columns (2)-(5). When the 95% posterior interval on a coefficient excludes zero, we loosely term the coefficient as significant.

The coefficient ρ_1 on $ln(R\&D)_{t-1}$ denotes the elasticity of own R&D investment to innovation whereas spillover elasticity of R&D using our model is captured by the coefficient ρ_2. $ln(SpTECH)_{t-1}$ and $ln(SpSIC)_{t-1}$ are measures constructed by Bloom, Schankerman, and Van Reenen (2013) as spillover measures of R&D in technological and product market spaces respectively. Besides $ln(R\&D)_{t-2}$ as a source of spillovers, we include industry proximity measure $SIC3$.

Considering results in table 3, one observes that the elasticity of own R&D on innovation is positive and significant across all model specifications save specification (1). Beginning

[21] Industry proximity of any two firms i and j equals 1 if both firms belong to the same 3-digit industry and zero otherwise.

with Bloom, Schankerman, and Van Reenen (2013)'s specification (column (1)), the spillover elasticities of R&D on innovation in both technological and product market spaces are negative though comparable coefficients in Bloom, Schankerman, and Van Reenen (2013, table IV) are positive.[22]

In Bloom, Schankerman, and Van Reenen (2013, Sect. 2)'s formal theoretical analyses, the relation between firms' R&D can be positive (negative), reflecting whether R&D efforts are strategic complements (substitutes) (see also Cohen and Levinthal (1989, Sect. II.2)). Thus, negative spillovers cannot be formally ruled out.[23] To build intuition, let the key goal of firms be profit maximisation. To the extent that a firm invests enough in R&D in order to boost its absorptive capacity, contribute to its own innovation, and appropriate technology created by other firms, the R&D spillover effect on profits is positive. Because R&D investment of other firms constitutes a substitute to the firm's own R&D, the spillover effect on innovation is negative because innovation undertaken by firm j no longer counts as innovation when appropriated by firm i.

The coefficient ρ_2 in specifications (2) through (5) gives the spillover elasticity of $R\&D$ on firm innovation using our model. Notice that in all specifications, the 95% posterior intervals exclude zero thus confirming the presence of spillovers and generally, a strong spatial dependence in firm innovation with respect to research and development. This measure captures R&D spillover effects generated via sources $ln(R\&D)_{t-2}$ and $SIC3$ and provides evidence of support of our model by the data vis-à-vis a model that uses exogenously constructed spillover measures.

In specifications (1) and (3)-(5) where the Bloom, Schankerman, and Van Reenen (2013)'s exogenously constructed spillover measures ($ln(SpTECH)_{t-1}$ and $ln(SpSIC)_{t-1}$) are included, neither coefficient is anywhere significant. Coefficients on the lagged outcome variable $ln(Pat)_{t-1}$ are positive and significant across all specifications which confirms strong

[22] We reckon this is due to creating a balanced panel out of the original unbalanced sample.

[23] For example, Cincera (1997, Table II, columns (1) through (3)) finds elasticities of the lagged spillover measure at $t-1$ ranging from -0.55 to -0.11.

Table 3: Coefficients - Cite-Weighted Patent Counts

	Pois (1)	Pois (2)	Pois (3)	Pois (4)	Pois (5)
ρ_1	-0.039	0.191	0.230	0.197	0.202
	(-0.166, 0.090)	(0.104, 0.276)	(0.158, 0.298)	(0.125, 0.272)	(0.135, 0.270)
ρ_2		-0.438	-0.359	-0.422	-0.398
		(-0.632, -0.248)	(-0.544, -0.167)	(-0.571, -0.279)	(-0.537, -0.248)
$ln(SpTECH)_{t-1}$	-0.048		-0.041	-0.041	0.001
	(-0.497, 0.403)		(-0.162, 0.078)	(-0.147, 0.067)	(-0.099, 0.111)
$ln(SpSIC)_{t-1}$	-0.038		-0.040		-0.033
	(-0.124, 0.047)		(-0.095, 0.008)		(-0.094, 0.026)
$ln(Pat)_{t-1}$	0.560	0.846	0.859	0.857	0.829
	(0.510, 0.609)	(0.795, 0.897)	(0.814, 0.908)	(0.810, 0.905)	(0.784, 0.871)
$Pre-SFE$	1.249	-0.094	-0.086	-0.100	-0.122
	(0.532, 1.966)	(-0.164, -0.03)	(-0.138, -0.029)	(-0.151, -0.049)	(-0.179, -0.063)
Spillover sources					
$2\ln(BF)_{R\&D}$		5.539	3.525	5.360	7.958
df, $\hat{p}$		6,(0.000)	6,(0.533)	6,(0.000)	6,(0.000)
$2\ln(BF)_{SIC3}$					0.328
df, $\hat{p}$					1,(0.000)
$2\ln(BF)_{\boldsymbol{\delta}}$		5.539	3.525	5.360	4.132
df, $\hat{p}$		6,(0.000)	6,(0.533)	6,(0.000)	7,(0.000)
$2\ln(BF)_{\boldsymbol{\theta}}$	42.448	33.957	33.462	34.456	33.285
df, $\hat{p}$	10, (0.000)	12,(0.000)	17,(0.00)	16,(0.00)	18,(0.000)

Notes: All specifications above (Pois (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. 10 500 MCMC simulations are drawn with 500 as burn-in. The spillover-generating variable $\mathbf{x}$ is lagged log R&D investment $ln(R\&D)_{t-1}$. The uninformative uniform prior over the entire parameter vector is used. $2\ln(BF)_z$ denotes twice the logarithm of the Bayes Factor according to Kass and Raftery (1995) by setting the corresponding parameter vector in $\boldsymbol{\delta}$ to $\mathbf{0}$ for the null hypothesis. Sources of spillovers (with respective Bayes factors) considered: $ln(R\&D)_{t-2}$ ($2\ln(BF)_{R\&D}$) and $SIC3$ ($2\ln(BF)_{SIC3}$). $2\ln(BF)_{\boldsymbol{\delta}}$ corresponds to the parameter vector on all spillover source terms. $\hat{p} \equiv \min_{\alpha\in(0,1)}\{\alpha : \mathbf{0} \notin \boldsymbol{I}_{1-\alpha}\}$ where $\boldsymbol{I}_{1-\alpha}$ denotes the $1-\alpha$ simultaneous confidence bands (see Montiel Olea and Plagborg-Møller (2019, algorithm 2), Huang and Tsyawo (2018, algorithm 2), and references therein) and df equals the length of the corresponding parameter vector. R implementation of the simultaneous confidence bands are available in the R package `bayesdistreg` (see Tsyawo and Huang (2019)). Column (1) implements specification (3) in Bloom, Schankerman, and Van Reenen (2013, table IV) on our balanced panel using the Bayesian Poisson model. Columns (2) through (5) are estimated using the (Bayesian) Independence Metropolis Hastings MCMC algorithm. The lower and upper 95% posterior intervals are provided in parentheses for inference.

persistence in firm patenting behaviour.[24]

For tests and comparisons on sources of R&D spillovers, we generate null hypotheses on parameter vectors $\boldsymbol{\delta}$ and $\boldsymbol{\theta}$ by setting them to $\mathbf{0}$. Two measures are considered for hypothesis tests: twice the logarithm of the Bayes Factor (Kass and Raftery, 1995) and a scalar measure we introduce $\hat{p} \equiv \min_{\alpha\in(0,1)}\{\alpha : \mathbf{0} \notin \boldsymbol{I}_{1-\alpha}\}$ where $\boldsymbol{I}_{1-\alpha}$ denotes the $1-\alpha$ simultaneous confidence bands (see Montiel Olea and Plagborg-Møller (2019, algorithm 2), Huang and Tsyawo (2018, algorithm 2), and references therein). $\hat{p}$ gives a scalar summary for a hypothesis test that can involve more than one parameter, and it can be considered as a Bayesian analogue of the classical p-value for joint hypothesis tests. $\hat{p}$ gives the minimum $\alpha \in (0, 1)$ such that the $(1 - \alpha)$ simultaneous Bayesian confidence bands (see Montiel Olea and Plagborg-Møller (2019) and Huang and Tsyawo (2018)) on the vector estimates exclude the vector $\mathbf{0}$. One ought to note, however, that the Bayes Factor measure $(2\ln(BF))$ enables us to select a model over another and test hypotheses. $\hat{p}$, on the other hand, provides evidence (or otherwise) against the null conditional on the estimated model.

Following the interpretation of Kass and Raftery (1995, sect. 3.2), except $SIC3$ in specification (5) where the evidence against the null is not worth more than a bare mention, the evidence against the null for $ln(R\&D)_{t-2}$, $\boldsymbol{\delta}$, and $\boldsymbol{\theta}$ is at least positive. The evidence against the null positive for $ln(R\&D)_{t-2}$ is positive in specifications (2)-(5). It is very strong for covariates (save fixed effects) and spillover sources. The spillover source terms are jointly positive in (2)-(5). The foregoing are largely corroborated by the values $\hat{p}$ we provide alongside the Bayes factors. Save specification (3), $\hat{p}$ equals 0 for $ln(R\&D)_{t-2}$ and $SIC3$ across all specifications. That elements in $\boldsymbol{\delta}$ corresponding to the time-varying $ln(R\&D)_{t-2}$ confirms the support of the data for a time-varying interaction structure through time. Coupled with ρ_2 estimates, whose 95% posterior intervals exclude 0 in all specifications, we see that our model is strongly supported by the data.

[24]Bloom, Schankerman, and Van Reenen (2013) finds strong persistence in firm patenting behaviour.

6 Conclusion

In this text, we address the problem of unknown spatial matrices in a study of R&D spillover effects on firm innovation. The study extends as well to other studies that quantify spillover effects. As a key contribution of the study, the proposed model determines spatial matrix, allows time variation, and asymmetry in the spatial matrix. The paper proposes a flexible approach to quantifying spillover effects or spatial dependence in a general class of linear and non-linear that admit linear predictor functions. Our approach tackles the problem of metric ambiguity by accommodating several sources of spillovers and assesses their relevance via hypothesis tests. Also, because our weights are modelled as functions of exogenous observables, we are able to handle latency and avoid endogeneity in spatial matrices. The estimator is consistent and asymptotically normal. We apply our method to study R&D spillovers on innovation in a knowledge production framework. We find a negative statistically significant spillover effect and a positive statistically significant private effect of R&D investment on firm innovation. The foregoing provide evidence that while own R&D investment boots innovation, R&D investment by other firms engender substitutability (in terms of innovation) in own R&D and hence has a negative impact on innovation. Exogenously constructed measures of spillovers lose explanatory power when included in our model. Our results also confirm strong persistence in firm patenting behaviour.

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] Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. "Identification of peer effects through social networks". *Journal of Econometrics* 150.1 (2009), pp. 41–55.

[11] Case, Anne C. "Spatial patterns in household demand". *Econometrica: Journal of the Econometric Society* (1991), pp. 953–965.

[12] Chernozhukov, Victor and Han Hong. "An MCMC approach to classical estimation". *Journal of Econometrics* 115.2 (2003), pp. 293–346.

[13] 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.

[14] Coe, David T, Elhanan Helpman, and Alexander W Hoffmaister. "International R&D spillovers and institutions". *European Economic Review* 53.7 (2009), pp. 723–741.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] Hall, Bronwyn H, Adam Jaffe, and Manuel Trajtenberg. "Market value and patent citations". *RAND Journal of Economics* (2005), pp. 16–38.

[20] Hoshino, Tadao. “Semiparametric Estimation of Censored Spatial Autoregressive Models”. *Econometric Theory* (2017), pp. 1–38.

[21] Huang, Weige and Emmanuel Tsyawo. “Bayesian Distribution Regression”. *Available at SSRN 3048658* (2018).

[22] Jaffe, Adam B. *Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market value*. 1986.

[23] 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.

[24] Kass, Robert E and Adrian E Raftery. “Bayes factors”. *Journal of the American Statistical Association* 90.430 (1995), pp. 773–795.

[25] 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.

[26] 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.

[27] Manresa, Elena. “Estimating the structure of social interactions using panel data”. *Unpublished Manuscript. CEMFI, Madrid* (2013).

[28] Montiel Olea, José Luis and Mikkel Plagborg-Møller. “Simultaneous confidence bands: Theory, implementation, and an application to SVARs”. *Journal of Applied Econometrics* 34.1 (2019), pp. 1–17.

[29] Newey, Whitney K and Daniel McFadden. “Large sample estimation and hypothesis testing”. *Handbook of Econometrics* 4 (1994), pp. 2111–2245.

[30] Pinkse, Joris and Margaret E Slade. "Mergers, brand competition, and the price of a pint". *European Economic Review* 48.3 (2004), pp. 617–643.

[31] Pinkse, Joris, Margaret E Slade, and Craig Brett. "Spatial price competition: a semi-parametric approach". *Econometrica* 70.3 (2002), pp. 1111–1153.

[32] 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.

[33] Soale, Abdul-Nasah and Emmanuel Tsyawo. "Clustered Covariate Regression". *Available at SSRN: https://ssrn.com/abstract=3394012* (2019).

[34] Souza, PC. "Estimating network effects without network data". *PUC-Rio Working Paper* (2014).

[35] Syverson, Chad. "What determines productivity?" *Journal of Economic Literature* 49.2 (2011), pp. 326–65.

[36] Tsyawo, Emmanuel. "An SAR Spatial Matrix Estimator". *Available at SSRN: https://ssrn.com/abstr* (2019).

[37] Tsyawo, Emmanuel and Weige Huang. *bayesdistreg: Bayesian Distribution Regression.* R package version 0.1.0. 2019. URL: `https://CRAN.R-project.org/package=bayesdistreg`.

[38] 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 of lemmata and theorems

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

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

Note that under assumption 1(a), $\boldsymbol{z}'\boldsymbol{z}$ is invertible. By assumption 1(b), the block diagonal $\boldsymbol{w}$ 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 $\boldsymbol{z}'\boldsymbol{z}$ in $\boldsymbol{m}'\boldsymbol{m}$, $S = \mathbf{x}'\boldsymbol{w}'\boldsymbol{w}\mathbf{x} - \mathbf{x}'\boldsymbol{w}'\boldsymbol{z}(\boldsymbol{z}'\boldsymbol{z})^{-1}\boldsymbol{z}'\boldsymbol{w}\mathbf{x} > 0$.

Let us factor S as $S = \mathbf{x}'\boldsymbol{w}'[\mathbf{I}_{NT} - \boldsymbol{z}(\boldsymbol{z}'\boldsymbol{z})^{-1}\boldsymbol{z}']\boldsymbol{w}\mathbf{x}$. Note that $\boldsymbol{z}(\boldsymbol{z}'\boldsymbol{z})^{-1}\boldsymbol{z}'$ is symmetric and idempotent; $[\mathbf{I}_{NT} - \boldsymbol{z}(\boldsymbol{z}'\boldsymbol{z})^{-1}\boldsymbol{z}']$ is also symmetric and idempotent. Since $\mathbf{x} \neq \mathbf{0}$ (assumption 1(c)), it implies S can be expressed as $S = \mathbf{x}'\boldsymbol{w}'[\mathbf{I}_{NT} - \boldsymbol{z}(\boldsymbol{z}'\boldsymbol{z})^{-1}\boldsymbol{z}'][\mathbf{I}_{NT} - \boldsymbol{z}(\boldsymbol{z}'\boldsymbol{z})^{-1}\boldsymbol{z}']\boldsymbol{w}\mathbf{x} = \sum_{i=1}^{N}\sum_{t=1}^{T}\zeta_{it}^2 > 0$ where $\mathbf{I}_{NT}$ denotes an $NT \times NT$ identity matrix and $\zeta_{it},\ i = 1, \ldots, N,\ t = 1, \ldots, T$ are residuals obtained by regressing $\boldsymbol{w}\mathbf{x}$ on $\boldsymbol{z}$, i.e., S is the sum of squared residuals. □

Proof of theorem 1.

The consistency result for the entire class of models under consideration is standard. See theorem 2.1 in Newey and McFadden (1994). □

Proof of theorem 2. Like in the proof of theorem 1, the asymptotic normality result is fairly standard. See theorem 3.1 in Newey and McFadden (1994). □

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