ARTICLE IN PRESS

Journal of Econometrics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Econometrics

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

Efficient closed-form estimation of large spatial autoregressions☆

Abhimanyu Gupta[1]

Department of Economics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK

ARTICLE INFO

Article history:
Received 11 September 2020
Received in revised form 18 May 2021
Accepted 22 May 2021
Available online xxxx

JEL classification:
C21
C31
C33
C36

Keywords:
Spatial autoregression
Efficiency
Many parameters
Networks

ABSTRACT

Newton-step approximations to pseudo maximum likelihood estimates of spatial autoregressive models with a large number of parameters are examined, in the sense that the parameter space grows slowly as a function of sample size. These have the same asymptotic efficiency properties as maximum likelihood under Gaussianity but are of closed form. Hence they are computationally simple and free from compactness assumptions, thereby avoiding two notorious pitfalls of implicitly defined estimates of large spatial autoregressions. When commencing from an initial least squares estimate, the Newton step can also lead to weaker regularity conditions for a central limit theorem than some extant in the literature. A simulation study demonstrates excellent finite sample gains from Newton iterations, especially in large multiparameter models for which grid search is costly. A small empirical illustration shows improvements in estimation precision with real data.

1. Introduction

Spatial autoregressive (SAR) models, introduced by Cliff and Ord (1973), are popular tools for modelling cross-sectionally dependent economic data. The pre-eminent feature of such models is the presence of one or more 'spatial weight' matrices, which parsimoniously capture the dependence between units in the sample. Such dependence need not be geographic in nature, indeed the spatial weight matrix is known by other terms such as 'adjacency matrix', 'network link matrix' and 'sociomatrix'. For $n \times 1$ vectors y_n and u of responses and unobserved disturbances, respectively, and an $n \times k$ covariate matrix X_n, the SAR model is

$$y_n = \sum_{i=1}^{p} \lambda_{0in} W_{in} y_n + X_n \beta_{0n} + u, \tag{1.1}$$

where the elements of the $n \times n$ spatial weight matrices W_{in} are inverse economic distances and $\lambda_{0n} = \left(\lambda_{01n}, \ldots, \lambda_{0pn}\right)'$ and β_{0n} are unknown parameter vectors. Subscripting with n permits treatment of triangular arrays, an important issue for spatial models in general (see Robinson, 2011), and for SAR models even more so due to various normalizations of the

☆ I am grateful to the associate editor, three referees and Peter Robinson for very helpful comments that improved the paper significantly, and to Christos Kolympiris for permission to use the data in the empirical illustration.

E-mail address: a.gupta@essex.ac.uk.

[1] Research supported by Economic and Social Research Council grant ES/R006032/1.

https://doi.org/10.1016/j.jeconom.2021.05.005

W_{in} that make them n-dependent. This paper justifies computationally straightforward estimation for the parameters of (1.1) with the same asymptotic properties as pseudo maximum likelihood estimates.

SAR models allow dependence to occur across a very generalized notion of space: so long as a mapping exists between every pair of individuals to the real line a spatial weight matrix may be constructed. The flexible nature of the SAR model means that it may be used to model a very wide range of phenomena. Thus it has found application in many fields of economics such as development economics (Case, 1991; Helmers and Patnam, 2014), industrial organization (Pinkse et al., 2002), trade (Conley and Dupor, 2003) and peer effects (Hsieh and van Kippersluis, 2018), to name only a few examples. Another frequently used approach to model cross-sectional dependence is the 'common factor' technique, see e.g. Chudik and Pesaran (2015) for a review.

Estimation of SAR models has long been considered in the regional science literature, see e.g. Anselin (1988). Rigorous asymptotic theory for instrumental variables (IV) estimation was initially provided by Kelejian and Prucha (1998), leading to the present flourishing theoretical literature. Lee (2002) studied ordinary least squares (OLS) estimation of SAR models, stressing the need for lack of sparsity in the spatial weight matrix to establish desirable asymptotic properties such as consistency and efficiency. This was followed by Lee (2004), a seminal contribution that provided a taxonomical asymptotic theory for Gaussian pseudo maximum likelihood estimates (PMLE) of SAR models. Recently Kuersteiner and Prucha (2013, 2020) have provided general theory for such models in a panel data setting.

The flexible nature of SAR modelling is further embellished by the seamless ability to integrate more than one spatial weight matrix in the model (1.1), thus permitting simultaneous connections between units across a number of channels. This is an accurate representation of typical economic situations, e.g. countries are 'connected' by both geographical proximity as well as trade ties. Furthermore, in many economic settings the sample partitions naturally into p clusters or groups, leading to block diagonal structure for the spatial weight matrix $V_n = diag\left(V_{1n}, \ldots, V_{pn}\right)$, where V_{in} is $m_i \times m_i$ and $\sum_{i=1}^{p} m_i = n$. To permit the modelling of heterogeneous spillover effects across clusters, one may take W_{in} to be the $n \times n$ block diagonal matrix with the $m_i \times m_i$ dimensional ith diagonal block given by V_{in}. This approach has been suggested by Gupta and Robinson (2015, 2018). Here and more generally, the specification (1.1) is termed a 'higher-order' SAR model if $p > 1$, see e.g Blommestein (1983), Lee and Liu (2010), Li (2017), Han et al. (2017) and Kwok (2019).

In the study of higher-order SAR models, Gupta and Robinson (2015, 2018) have suggested that p, k be allowed to diverge slowly to infinity as functions of sample size. The motivation for such generality is typically threefold: first, it is desirable to permit a richer model as the sample size permits. Second, clustered data as mentioned in the previous paragraph naturally imply asymptotic regimes with increasing p. For instance, when $m_i = m$ for each $i = 1, \ldots, p$, we have $n = mp$ and the results of Lee (2004) imply that $p \to \infty$ is necessary for consistent estimation, analogous to the problems created in the spatial statistics literature by 'infill asymptotics', see e.g. Lahiri (1996). Finally, a theory that allows the model dimension to grow with sample size provides a more incisive analysis of large models in practice, much as typical asymptotic theory with a fixed parameter space itself can be thought as providing an approximation in finite samples.

The estimation of such increasing-order SAR models has been studied by Gupta and Robinson (2015, 2018) using IV, OLS and PMLE approaches. The first two methods have the advantage of being in closed-form, while even for $p = 1$ PMLE (in)famously requires grid search and the inversion of an $n \times n$ matrix in every iteration, leading to many ingenious solutions for faster computation, see e.g. Ord (1975) and Pace and Barry (1997). The computational cost of PMLE in SAR type models is particularly salient as data sets increase in size, as stressed by Zhu et al. (2020). Modern network data sets are amenable to modelling via SAR techniques and can feature, or accommodate, large parameter spaces but computation remains a serious challenge. Han et al. (2020) provide a discussion of the problems and propose a Bayesian solution.

These problems are naturally exacerbated if $p > 1$, with grid search requiring more iterations to converge and each iteration requiring inversion of an $n \times n$ matrix, as well as risk of convergence to local optima. Furthermore, the requirement of a compact parameter space for λ_{0n} can severely restrict the admissible parameter values (see Gupta and Robinson, 2018). On the other hand, under Gaussianity the PMLE becomes the MLE and is efficient. This property is shared by OLS, but under rather delicate and specific conditions even for $p = 1$ (see Lee, 2002). Thus the IV/OLS and PMLE approaches each have their advantages and it is desirable to combine the positive properties of both.

One method of obtaining closed-form estimates with the same asymptotic covariance matrix as a target estimate is to use Newton-type iterations commencing from an initial consistent estimator that is straightforward to compute. The approach dates back at least to Fisher (1925) and LeCam (1956). It enjoys the added attraction of avoiding a potentially complicated consistency proof for an implicitly defined estimate, as well as the compactness assumptions this typically entails. As a result, the technique has been used in a vast variety of settings, see e.g. Rothenberg and Leenders (1964) (simultaneous equations), Hartley and Booker (1965) (nonlinear least squares), Janssen et al. (1985) (M-estimation), Rothenberg (1984) (generalized least squares), Hualde and Robinson (2011), Kristensen and Linton (2006), Robinson (2005) (time series and adaptive estimation), Andrews (1997) (generalized method of moments), Kasahara and Shimotsu (2008), Kristensen and Salanié (2017) (structural estimation), De Luca et al. (2018) (generalized linear models) and Frazier and Renault (2017) (efficient two-step estimation), to name just a few.

In this paper we use IV and OLS estimates as initial estimates to form a single Newton-step asymptotic approximation to the Gaussian PMLE with $p = p_n$ and $k = k_n$ allowed to diverge as functions of $n \to \infty$. The approach has been studied in the case of fixed-dimensional SAR models by Robinson (2010) and Lee and Yu (2013), but the previous discussion hints at its particular usefulness when considering large models. One avoids grid search over a high-dimensional parameter space, compactness assumptions on this space and the inversion of large $(n \times n)$ matrix for every search iteration, as well as various headaches related to convergence and local optima. When commencing from IV estimates, this leads to closed-form efficient estimates under Gaussianity. As suggested by the results of Lee (2002) and Gupta and Robinson (2015), commencing iteration from OLS preserves the efficiency property. However, we show that the Newton step approach cancels out certain terms of large stochastic order that allows for weaker rate conditions than those imposed in these papers.

In a simulation study, we demonstrate that the Newton step can lead to much improved estimates in finite samples, both in terms of bias and efficiency. While a single step is sufficient to establish desirable asymptotic properties, in our simulation study we also explore the finite sample implications of additional Newton steps, reporting results with up to six iterations. We find large finite sample gains in both bias and mean squared error that are robust to heavy tailed error distributions. We also observe fast convergence of iterations, which conforms to extant theoretical observations. The gains are particularly notable when the parameter space and sample size is large, a situation in which PMLE becomes computationally onerous. In a small illustration with real world data, we show that the estimates work well in practice and lead to more precise results.

We collect some frequently used notation here for the convenience of the reader. For a generic matrix A denote $\|A\| = \left(\overline{\eta}\left(A'A\right)\right)^{\frac{1}{2}}$, with $\overline{\eta}(\cdot)$ and $\underline{\eta}(\cdot)$ denoting the largest and smallest eigenvalues, respectively, of a symmetric positive semidefinite matrix. Note that if A is a vector then $\|A\|$ is simply its Euclidean norm. Let $\|A\|_R$ denote the maximum absolute row sum norm of A. For any parameter τ, function $f(\tau)$ and generic estimate $\check{\tau}$, we will write $\check{f} \equiv f\left(\check{\tau}\right)$. We denote true parameter values with 0 subscript and suppress the argument for a quantity evaluated at a true parameter value, i.e. $f\left(\tau_0\right) \equiv f$.

2. Approximations to Gaussian PMLE

The ($-2/n$ times) log pseudo Gaussian likelihood function for model (1.1) at any admissible point $\theta = (\lambda', \beta')'$ is given by

$$\mathcal{Q}_n\left(\theta, \sigma^2\right) = \log\left(2\pi\sigma^2\right) - \frac{2}{n}\log\left|S_n\left(\lambda\right)\right| + \frac{1}{n\sigma^2}\left(S_n(\lambda)y_n - X_n\beta\right)'\left(S_n(\lambda)y_n - X_n\beta\right), \tag{2.1}$$

where $S_n(\lambda) = I_n - \sum_{i=1}^{p_n}\lambda_{in}W_{in}$, with I_n denoting the $n \times n$ identity matrix. If S_n is invertible, (1.1) admits the reduced form $y_n = S_n^{-1}X_n\beta_n + S_n^{-1}u$, and we define $R_n = A_n + B_n$, where $A_n = \left(G_{1n}X_n\beta_n, \ldots, G_{p_nn}X_n\beta_n\right)$, $B_n = \left(G_{1n}u, \ldots, G_{p_nn}u\right)$, $G_{in}(\lambda) = W_{in}S_n^{-1}(\lambda)$, $i = 1, \ldots, p_n$, and so $R_n = \left(W_{1n}y_n, \ldots, W_{p_nn}y_n\right)$.

Defining $\mathcal{R}_n^y\left(\theta\right) = R_n\lambda_n + X_n\beta_n - y_n$, the derivative of (2.1) at any admissible $\left(\theta, \sigma^2\right)$ is

$$\xi_n\left(\theta, \sigma^2\right) = \left(\varphi_n'(\theta, \sigma^2),\ 2\sigma^{-2}n^{-1}\mathcal{R}_n^{y\prime}\left(\theta\right)X_n\right)', \tag{2.2}$$

where $\varphi_n\left(\theta, \sigma^2\right) = 2\sigma^{-2}n^{-1}\left(\sigma^2 trG_{1n}(\lambda) + y_n'W_{1n}'\mathcal{R}_n^y\left(\theta\right), \ldots, \sigma^2 trG_{p_nn}(\lambda) + y_n'W_{p_nn}'\mathcal{R}_n^y\left(\theta\right)\right)'$. Because $\mathcal{R}_n^y = -u$, denoting $\phi_n = \sigma_0^{-2}n^{-1}\left(\sigma_0^2 trC_{1n} - u'C_{1n}u, \ldots, \sigma_0^2 trC_{p_n} - u'C_{p_nn}u\right)'$ with $C_{in} = G_{in} + G_{in}'$, we obtain

$$\xi_n \equiv \frac{\partial\mathcal{Q}_n}{\partial\theta} = \left(\phi_n', 0\right)' - 2\sigma_0^{-2}t_n, \tag{2.3}$$

with $t_n = n^{-1}\left[A_n, X_n\right]'u$. The Hessian at any admissible point in the parameter space is

$$H_n\left(\theta_n, \sigma^2\right) \equiv \frac{\partial^2\mathcal{Q}_n(\theta_n, \sigma^2)}{\partial\theta\partial\theta'} = \begin{pmatrix} \frac{2}{n}P_{ji,n}(\lambda_n) + \frac{2}{n\sigma^2}R_n'R_n & \frac{2}{n\sigma^2}R_n'X_n \\ \frac{2}{n\sigma^2}X_n'R_n & \frac{2}{n\sigma^2}X_n'X_n \end{pmatrix} \tag{2.4}$$

where $P_{ji,n}(\lambda_n)$ is the $p_n \times p_n$ matrix with (i, j)th element given by $tr\left(G_{jn}(\lambda_n)G_{in}(\lambda_n)\right)$.

Let Z_n be an $n \times r_n$ matrix of instruments, with $r_n \geq p_n$, and define the IV and OLS estimates as

$$\hat{\theta}_n = \hat{Q}_n^{-1}\hat{K}_n'J_n^{-1}\hat{k}_n, \qquad \hat{\sigma}_n^2 = n^{-1}\left\|y_n - \left(R_n, X_n\right)\hat{\theta}_n\right\|^2, \tag{2.5}$$

$$\tilde{\theta}_n = \hat{L}_n^{-1}\hat{l}_n, \qquad \tilde{\sigma}_n^2 = n^{-1}\left\|y_n - \left(R_n, X_n\right)\tilde{\theta}_n\right\|^2, \tag{2.6}$$

respectively, with $\hat{Q}_n = \hat{K}_n'J_n^{-1}\hat{K}_n$, $\hat{K}_n = n^{-1}\left[Z_n, X_n\right]'\left[R_n, X_n\right]$, $\hat{k}_n = n^{-1}\left[Z_n, X_n\right]'y_n$, $J_n = n^{-1}\left[Z_n, X_n\right]'\left[Z_n, X_n\right]$, and $\hat{L}_n = n^{-1}\left[R_n, X_n\right]'\left[R_n, X_n\right]$, $\hat{l}_n = n^{-1}\left[R_n, X_n\right]'y_n$. Define the respective 'one-step' estimates $\hat{\hat{\theta}}_n$ and $\tilde{\tilde{\theta}}_n$ by the following equations

$$\hat{\hat{\theta}}_n = \hat{\theta}_n - \hat{H}_n^{-1}\hat{\xi}_n, \tag{2.7}$$

$$\tilde{\tilde{\theta}}_n = \tilde{\theta}_n - \tilde{H}_n^{-1}\tilde{\xi}_n. \tag{2.8}$$

We observe that other initial estimates, such as the GMM estimates of Kelejian and Prucha (1999) and Lee (2007), can also be used. However we choose initial estimates that are available in closed form for computational ease. While consistent initial estimates are needed to obtain a desirable asymptotic theory, even in the fixed-dimension parametric case these are permitted to be n^{ψ}-consistent, where $\psi < 1/2$, see Robinson (1988) and references therein.

While our theorems below establish desired asymptotic properties for the one step estimates, from a practical point of view more iterations may be desirable. In fact, these also improve the statistical rate of convergence to the target PMLE, yielding an even faster statistical counterpart to the famous quadratic numerical rate of convergence of Newton estimates, see for example Theorem 2 of Robinson (1988) and pp. 312–313 of Ortega and Rheinboldt (1970). We examine this issue in more detail in the next section and also the Monte Carlo study.

3. Asymptotic properties

The following assumptions are discussed in Lee (2002, 2004), and Gupta and Robinson (2015, 2018), amongst other spatial papers in which they are routinely employed. These conditions are by no means the weakest possible set, but we opt for tractability to convey the main message especially in view of the large number of spatial parameters involved. For example, stochastic regressors can be easily accommodated but complicate the notation.

Assumption 1. $u = (u_1, \ldots, u_n)'$ has iid elements with zero mean and finite variance σ_0^2.

Assumption 2. For $i = 1, \ldots, p_n$, the elements of W_{in} are uniformly $\mathscr{O}(1/h_n)$, where h_n is some positive sequence which may be bounded or divergent, but always bounded away from zero and such that $n/h_n \to \infty$ as $n \to \infty$. The diagonal elements of each W_{in} are zero.

Assumption 3. S_n is non-singular for all sufficiently large n.

Assumption 4. $\left\|S_n^{-1}\right\|_R$, $\left\|S_n'^{-1}\right\|_R$, $\|W_{in}\|_R$ and $\left\|W_{in}'\right\|_R$ are uniformly bounded in n and i for all $i = 1, \ldots, p_n$ and sufficiently large n.

Assumption 5. The elements of X_n are constants and are uniformly bounded in n, in absolute value, for all sufficiently large n.

Assumption 6. The elements of Z_n are constants and are uniformly bounded, in absolute value, for all sufficiently large n.

Assumption 7. $\overline{\lim}_{n\to\infty} \overline{\eta}(J_n) < \infty$ and $\underline{\lim}_{n\to\infty} \underline{\eta}(K_n'K_n) > 0$.

Assumption 8. $\underline{\lim}_{n\to\infty} \underline{\eta}(J_n) > 0$ and $\overline{\lim}_{n\to\infty} \overline{\eta}(K_n'K_n) < \infty$.

Assumption 9. $\overline{\lim}_{n\to\infty} \overline{\eta}(L_n) < \infty$.

Assumption 10. $\underline{\lim}_{n\to\infty} \underline{\eta}(L_n) > 0$.

Assumption 11. $\mathbb{E}\left(u_i^4\right) \le C$ for $i = 1, \ldots, n$.

Let Ψ_n be an $s\times(p_n + k_n)$ matrix of constants with full row-rank. The claims of the following theorems also hold when p_n and k_n are fixed, but we state and prove the results for the more challenging case when these diverge.

Theorem 3.1.

(i) Let Assumptions 1–11 hold along with

$$\frac{1}{p_n} + \frac{1}{r_n} + \frac{1}{k_n} + \frac{p_n^3 k_n^4}{n} + \frac{p_n^{\frac{3}{2}} k_n}{h_n} \to 0 \text{ as } n \to \infty, \tag{3.1}$$

and

$$\frac{r_n^2}{n} \text{ bounded as } n \to \infty. \tag{3.2}$$

Then

$$\frac{n^{\frac{1}{2}}}{(p_n + k_n)^{\frac{1}{2}}} \Psi_n \left(\hat{\hat{\theta}}_n - \theta_{0n}\right) \xrightarrow{d} N\left(0, \lim_{n\to\infty} \frac{\sigma_0^2}{p_n + k_n} \Psi_n L_n^{-1} \Psi_n'\right),$$

where the asymptotic covariance matrix exists, and is positive definite, by Assumptions 9 and 10.

(ii) Let Assumptions 1–5 and 9–11 hold. Suppose also that

$$\frac{1}{p_n}+\frac{1}{k_n}+\frac{p_n^3k_n^4}{n}+\frac{p_n^{\frac{3}{2}}k_n}{h_n}+\frac{n^{\frac{1}{2}}p_n^{\frac{5}{2}}}{h_n^3}\to 0. \tag{3.3}$$

Then

$$\frac{n^{\frac{1}{2}}}{(p_n+k_n)^{\frac{1}{2}}}\Psi_n\left(\tilde{\tilde{\theta}}_n-\theta_{0n}\right)\xrightarrow{d}N\left(0,\lim_{n\to\infty}\frac{\sigma_0^2}{p_n+k_n}\Psi_nL_n^{-1}\Psi_n'\right),$$

where the asymptotic covariance matrix exists, and is positive definite, by Assumptions 9 and 10.

In the 'just identified' case $p_n=r_n$, condition (3.2) is implied by (3.1). Theorem 3.1(i) shows that the one-step estimate asymptotically achieves the efficiency bound noted by Lee (2002). On the other hand, Theorem 3.1(ii) yields the same distributional result as for the OLS estimate (Theorem 4.3 of Gupta and Robinson, 2015). This should come as no surprise since Lee (2002) has already established the efficiency of OLS under suitable conditions. Nevertheless, Theorem 3.1(ii) imposes weaker conditions on the relative rates of h_n and $n^{\frac{1}{2}}$ than those extant in the literature.

Indeed, for their result, Gupta and Robinson (2015) assumed $n^{\frac{1}{2}}p_n^{\frac{1}{2}}/h_n\to 0$ as compared to our $n^{\frac{1}{2}}p_n^{\frac{5}{2}}/h_n^3\to 0$. The latter is a quantity of smaller order as $\left(n^{\frac{1}{2}}p_n^{\frac{5}{2}}/h_n^3\right)/\left(n^{\frac{1}{2}}p_n^{\frac{1}{2}}/h_n\right)=(p_n/h_n)^2\to 0$. For fixed p_n and k_n, our asymptotic normality result relies only on $n^{\frac{1}{2}}/h_n^3\to 0$, as $n\to\infty$. This is a weaker requirement as compared to Lee (2002), who assumed $n^{\frac{1}{2}}/h_n\to 0$ as $n\to\infty$. The reason for these favourable outcomes is the cancellation of higher order terms when using the one-step approximation. The key difference is in the rates $\left\|n^{-1}[B_n,0]'u\right\|=\mathcal{O}_p\left(p_n^{\frac{1}{2}}/h_n\right)$ and $\|\phi_n\|=\mathcal{O}_p\left(p_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n^{\frac{1}{2}}\right)$, the latter being sharper since $n/h_n\to\infty$ as $n\to\infty$.

To more transparently illustrate the implications of our weaker rate conditions, consider data collected in a 'farmer-district' type of environment, such as in Case (1991). Suppose that there are D districts, each containing m farmers, so that $n=Dm$, and $D,m\to\infty$ simultaneously. There is independence across districts, but equal dependence within districts, yielding $h_n=m-1$ (see Lee, 2002, 2004 for a more detailed discussion). Then, with fixed p_n, Lee (2002) and Gupta and Robinson (2015) required $n^{\frac{1}{2}}/h_n=D^{\frac{1}{2}}/m^{\frac{1}{2}}=o(1)$, while our condition imposes $n^{\frac{1}{2}}/h_n^3=D^{\frac{1}{2}}/m^{\frac{5}{2}}=o(1)$. Thus, our condition permits D to grow much faster as we only need $D^{\frac{1}{5}}=o(m)$ as compared to $D=o(m)$. The author thanks an anonymous referee for suggesting this illustration. We note that Robinson (2010) obtained asymptotic normality, indeed efficiency, in a semiparametric setup with $p_n=1$ requiring only $h_n\to\infty$ if the disturbances are symmetrically distributed or the weight matrix is symmetric. This condition would likely need to be suitably amended as $p_n\to\infty$.

If h_n is bounded as $n\to\infty$, a more complicated analysis is required to establish that one-step estimates achieve the PMLE asymptotic covariance matrix, because the information equality does not hold asymptotically. Denote $\mu_l=\mathbb{E}\left(u_i^l\right)$ for natural numbers l, and introduce, with $i,j=1,\ldots,p_n$, the $p_n\times p_n$ matrix $\Omega_{\lambda\lambda,n}$ with (i,j)th element $\frac{4\mu_3}{n\sigma_0^4}\sum_{r=1}^n c_{rr,in}b_{r,jn}X_n\beta_{0n}+\frac{\left(\mu_4-3\sigma_0^4\right)}{n\sigma_0^4}\sum_{r=1}^n c_{rr,in}c_{rr,jn}$ and the $k_n\times p_n$ matrix $\Omega_{\lambda\beta,n}$ with ith column $\frac{2\mu_3}{n\sigma_0^4}\sum_{r=1}^n c_{rr,in}x_{r,n}$ where $c_{pq,in}$ is the (p,q)th element of C_{in}, $b_{jn}=G_{jn}X_n\beta_{0n}$ with tth element $b_{t,jn}$ ($j=1,\ldots,p_n$ and $t=1,\ldots,n$) and $x_{p,n}$ is the pth column of X_n'. Define

$$\Omega_n=\begin{pmatrix}\Omega_{\lambda\lambda,n} & \Omega_{\lambda\beta,n}'\\ \Omega_{\lambda\beta,n} & 0\end{pmatrix}. \tag{3.4}$$

Then $\mathbb{E}\left(\xi_n\xi_n'\right)=n^{-1}\left(2\Xi_n+\Omega_n\right)$, where

$$\Xi_n=\mathbb{E}\left(H_n\right)=\begin{pmatrix}\frac{2}{n}\left(P_{ji,n}+P_{j'i,n}+\frac{1}{\sigma_0^2}A_n'A_n\right) & \frac{2}{n\sigma_0^2}A_n'X_n\\ \frac{2}{n\sigma_0^2}X_n'A_n & \frac{2}{n\sigma_0^2}X_n'X_n\end{pmatrix}. \tag{3.5}$$

When h_n is bounded OLS cannot be consistent (see Lee, 2002), so the following theorem considers only initial IV estimates.

Theorem 3.2. *Let Assumptions 1–7 hold. Suppose that h_n is bounded away from zero and that there is a real number $\delta>0$ such that $\mathbb{E}\left|u_i\right|^{4+\delta}\le C$ for $i=1,\ldots,n$. In addition, assume that*

$$\overline{\lim_{n\to\infty}}\,\overline{\eta}\left(\Xi_n^{-1}\Omega_n\Xi_n^{-1}\right)<\infty,\ \ \underline{\lim}_{n\to\infty}\,\underline{\eta}\left(2\Xi_n^{-1}+\Xi_n^{-1}\Omega_n\Xi_n^{-1}\right)>0\text{ and }\underline{\lim}_{n\to\infty}\,\underline{\eta}\left(\Xi_n\right)>0. \tag{3.6}$$

Suppose also that

$$\frac{1}{p_n}+\frac{1}{r_n}+\frac{1}{k_n}+\frac{p_n^3k_n^2\left(p_n\left(r_n+k_n\right)+k_n^2\right)}{n}+\frac{(p_nk_n)^{2+\frac{8}{\delta}}}{n}\to 0\text{ as }n\to\infty. \tag{3.7}$$

Then

$$\frac{n^{\frac{1}{2}}}{(p_n+k_n)^{\frac{1}{2}}}\Psi_n\left(\hat{\hat{\theta}}_n-\theta_{0n}\right)\xrightarrow{d}N\left(0,\lim_{n\to\infty}\frac{\sigma_0^2}{p_n+k_n}\Psi_n\left(2\Xi_n^{-1}+\Xi_n^{-1}\Omega_n\Xi_n^{-1}\right)\Psi_n'\right),$$

where the asymptotic covariance matrix exists, and is positive definite, by (3.6).

The rate condition (3.7) can simplify depending on the value of δ, i.e. the order of the finite moments assumed for u_i. As δ grows larger, the last term in the rate condition becomes redundant, indeed the numerator therein tends to $p_n^2k_n^2$ as $\delta\to\infty$, which is evidently dominated by the numerator of the other rate restriction. In the 'farmer-district' setting discussed earlier, we have bounded $h_n=m-1$ in this case. To further illustrate the rate condition, suppose that we are in the just identified case $p_n=r_n$. Then (3.7) requires $p_n^5k_n^3+p_n^4k_n^4+p_n^3k_n^5+(p_nk_n)^{2+\frac{8}{\delta}}=o(n)$. Then the term involving δ dominates the other three if $\delta\le 8/3$.

As indicated earlier, further iterations on the Newton step can improve the rate of statistical convergence to the target as well as finite sample properties. To see this, let $\hat{\hat{\theta}}_n^{\ell}$ be the ℓth Newton iteration towards the PMLE $\check{\theta}_n$. By Theorem 2 of Robinson (1988), $\left\|\check{\theta}_n-\hat{\hat{\theta}}_n^{\ell+1}\right\|=\mathcal{O}_p\left(\left\|\check{\theta}_n-\hat{\hat{\theta}}_n\right\|^{2^{\ell}}\right)$, an identical bound holding also for $\tilde{\tilde{\theta}}_n^{\ell+1}$. A factor that depends on ℓ is suppressed in the stated stochastic bound, indicating that this is not uniform in ℓ. Because the results of Gupta and Robinson (2018) and this paper show that one-step Newton estimates and $\check{\theta}_n$ are $n^{1/2}/(p_n+k_n)^{1/2}$-consistent, we have

$$\left\|\check{\theta}_n-\hat{\hat{\theta}}_n^{\ell+1}\right\|=\mathcal{O}_p\left((n/(p_n+k_n))^{-2^{\ell-1}}\right),\qquad\left\|\check{\theta}_n-\tilde{\tilde{\theta}}_n^{\ell+1}\right\|=\mathcal{O}_p\left((n/(p_n+k_n))^{-2^{\ell-1}}\right),$$

thus yielding the rate at which the iterations approximate the target estimate in a statistical sense, pointwise in ℓ.

4. Finite-sample performance of Newton-step estimates

4.1. Fixed number of neighbours (bounded h_n)

We examine finite-sample performance of $\hat{\theta}_n$ in this section, since the IV case entails a change in limiting distribution due to the Newton step and OLS requires divergent h_n to be consistent. Following Das et al. (2003) and the design in Gupta and Robinson (2015), define W_{in}^* as the symmetric circulant matrix with first row

$$w_{1j,in}^*=\begin{cases}0 & \text{if } j=1 \text{ or } j=i+2,\ldots,n-i;\\ 1 & \text{if } j=2,\ldots,i+1 \text{ or } j=n-i+1,\ldots,n,\end{cases}\tag{4.1}$$

and take $W_{in}^c=\left\|W_{in}^*\right\|^{-1}W_{in}^*$, where $\left\|W_{in}^*\right\|=\overline{\eta}\left(W_{in}^*\right)=2i$, because W_{in}^* is a symmetric, circulant matrix (see e.g. Davis, 1979 p. 73). Thus W_{in}^c is also a symmetric circulant matrix with first row given by $w_{1j,in}^*/2i$. This is an example of spatial weight matrices with bounded h_n.

We now dispense with some n subscripts for brevity. Our design generates $y=S^{-1}(X\beta+u)$ for sample sizes $n=200,400,800$ and $k=2$, with elements x_{j1} and x_{j2} of X generated as iid replicates from a $U(0,1)$ distribution, $j=1,\ldots,n$. We generate the disturbance u using two different distributions: $N(0,1)$ and t_6. PMLE becomes MLE under the first, while the second has heavier tails. Our experiments take $p=2,4,6$ for each of the described designs. We use a design with weights matrices given by W_i^c, $i=1,\ldots,p$. Finally, we set $\beta_1=1$, $\beta_2=0.5$ and $p=2:\lambda_1=0.4,\lambda_2=0.5$; $p=4:\lambda_1=0.3,\lambda_i=0.2,i=2,3,4$; $p=6:\lambda_i=0.15,i=1,\ldots,6$. The choices of λ_i satisfy the sufficient condition $\sum_{i=1}^p|\lambda_i|<1$ for invertibility of S.

With the aim of comparing initial IV estimates and MLE to Newton-step estimates, we first report three statistics: Monte Carlo mean, Monte Carlo mean squared error, and relative root Monte Carlo mean squared error, the latter being a straightforward ratio of the root MSE for IV and the iterated estimate. We also examine the use of more than one iteration in finite samples, and for this recall the notation $\hat{\hat{\theta}}_n^{\ell}$ for the ℓth Newton iteration. Our results are reported for $\ell=1,3,6$. The set of instruments that we use for our initial estimates are the linearly independent columns of $Z=\left(W_1^cX,\ldots,W_p^cX,X\right)$.

In Tables 1 and 2, we report the Monte Carlo mean of our estimates for standard normal and t_6 errors, respectively. For standard normal errors, we notice that the initial IV estimate can be heavily biased but Newton iterations improve matters, sometimes spectacularly. Indeed, for $p=6$ and $n=200$ the performance of $\hat{\theta}_n$ can be appalling, with $\hat{\lambda}_5<0$. However after six Newton steps this has improved to 0.1216 and even three iterations lead to a significant improvement. The reduction of bias from Newton iterations is not a universal feature, however broadly speaking the Newton steps reduce bias in the estimates, even for smaller values of p. As the sample size increases the iterations converge substantially, with little to choose typically between $\hat{\hat{\theta}}_n^3$ and $\hat{\hat{\theta}}_n^6$ for $n=800$. However for $n<800$, we notice that three iterations usually do the job quite satisfactorily, especially when $p<6$.

For t_6 errors, Table 2 paints a similar picture to Table 1. Once again, the noticeable 'rogue' estimate is for λ_5 when $p=6$ and $n=200$. Considering that all our simulations start from the same seed, this outlier may possibly be attributed

Table 1
Monte Carlo mean of parameter estimates. IV and iterated Newton-step estimates with $N(0, 1)$ errors. Bounded h_n. The parameter values used in the DGPs are as follows. $\beta_1 = 1$, $\beta_2 = 0.5$ and $p = 2 : \lambda_1 = 0.4, \lambda_2 = 0.5$; $p = 4 : \lambda_1 = 0.3, \lambda_i = 0.2, i = 2, 3, 4$; $p = 6 : \lambda_i = 0.15, i = 1, \ldots, 6$.

	$n = 200$				$n = 400$				$n = 800$			
$p = 2$	$\hat{\theta}_n$	$\hat{\hat{\theta}}_n^1$	$\hat{\hat{\theta}}_n^3$	$\hat{\hat{\theta}}_n^6$	$\hat{\theta}_n$	$\hat{\hat{\theta}}_n^1$	$\hat{\hat{\theta}}_n^3$	$\hat{\hat{\theta}}_n^6$	$\hat{\theta}_n$	$\hat{\hat{\theta}}_n^1$	$\hat{\hat{\theta}}_n^3$	$\hat{\hat{\theta}}_n^6$
λ_1	0.4488	0.4381	0.4153	0.4148	0.4198	0.4144	0.4037	0.4037	0.4251	0.4061	0.3997	0.3997
λ_2	0.4582	0.4611	0.4812	0.4818	0.4834	0.4843	0.4942	0.4942	0.4769	0.4934	0.4992	0.4992
β_1	0.9265	0.9922	1.0164	1.0160	0.9612	1.0002	1.0083	1.0082	0.9832	1.0052	1.0100	1.0101
β_2	0.4586	0.5061	0.5218	0.5211	0.4858	0.5120	0.5168	0.5166	0.4862	0.5011	0.5044	0.5044
$p = 4$												
λ_1	0.3021	0.3006	0.3008	0.2999	0.3181	0.3085	0.3079	0.3110	0.3237	0.3010	0.2975	0.2975
λ_2	0.3512	0.2922	0.2219	0.2236	0.2469	0.2361	0.1888	0.1821	0.2209	0.2146	0.2042	0.2042
λ_3	0.1695	0.2729	0.2956	0.2905	0.2233	0.2254	0.2410	0.2469	0.2104	0.2145	0.2060	0.2059
λ_4	0.0937	0.0365	0.0799	0.0841	0.1207	0.1293	0.1603	0.1583	0.1499	0.1690	0.1908	0.1910
β_1	0.8371	0.9562	0.9988	0.9993	0.9036	0.9871	1.0033	1.0027	0.9524	1.0047	1.0113	1.0114
β_2	0.4064	0.4890	0.5101	0.5103	0.4574	0.5108	0.5159	0.5153	0.4723	0.5050	0.5068	0.5069
$p = 6$												
λ_1	0.1678	0.1759	0.1774	0.1769	0.1727	0.1628	0.1570	0.1566	0.1742	0.1507	0.1493	0.1495
λ_2	0.2756	0.2017	0.1373	0.1420	0.2225	0.1662	0.1383	0.1392	0.1878	0.1547	0.1525	0.1511
λ_3	0.1046	0.1401	0.1856	0.1815	0.1279	0.1467	0.1641	0.1629	0.1361	0.1531	0.1490	0.1515
λ_4	0.2741	0.2885	0.2149	0.2029	0.1587	0.2026	0.1564	0.1613	0.1488	0.1720	0.1487	0.1459
λ_5	−0.0390	0.0475	0.1085	0.1216	0.1522	0.1753	0.1870	0.1761	0.1700	0.1439	0.1613	0.1636
λ_6	0.1390	0.0496	0.0749	0.0736	0.0794	0.0456	0.0949	0.1016	0.0906	0.1242	0.1374	0.1366
β_1	0.7790	0.9346	0.9887	0.9893	0.8594	0.9802	1.0023	1.0034	0.9272	1.0064	1.0122	1.0124
β_2	0.3812	0.4893	0.5109	0.5109	0.4349	0.5143	0.5193	0.5196	0.4590	0.5093	0.5090	0.5091

Table 2
Monte Carlo mean of parameter estimates. IV and iterated Newton-step estimates with t_6 errors. Bounded h_n. The parameter values used in the DGPs are as follows. $\beta_1 = 1$, $\beta_2 = 0.5$ and $p = 2 : \lambda_1 = 0.4, \lambda_2 = 0.5$; $p = 4 : \lambda_1 = 0.3, \lambda_i = 0.2, i = 2, 3, 4$; $p = 6 : \lambda_i = 0.15, i = 1, \ldots, 6$.

	$n = 200$				$n = 400$				$n = 800$			
$p = 2$	$\hat{\theta}_n$	$\hat{\hat{\theta}}_n^1$	$\hat{\hat{\theta}}_n^3$	$\hat{\hat{\theta}}_n^6$	$\hat{\theta}_n$	$\hat{\hat{\theta}}_n^1$	$\hat{\hat{\theta}}_n^3$	$\hat{\hat{\theta}}_n^6$	$\hat{\theta}_n$	$\hat{\hat{\theta}}_n^1$	$\hat{\hat{\theta}}_n^3$	$\hat{\hat{\theta}}_n^6$
λ_1	0.4715	0.4620	0.4266	0.4216	0.4728	0.4558	0.4325	0.4305	0.4055	0.4044	0.3978	0.3978
λ_2	0.4378	0.4402	0.4712	0.4761	0.4332	0.4454	0.4665	0.4685	0.4978	0.4951	0.5009	0.5010
β_1	0.8999	0.9659	1.0062	1.0069	0.9413	0.9864	1.0050	1.0057	0.9699	1.0035	1.0107	1.0107
β_2	0.4507	0.4935	0.5174	0.5175	0.4591	0.4879	0.4998	0.5003	0.4770	0.4987	0.5023	0.5024
$p = 4$												
λ_1	0.3136	0.3104	0.3118	0.3063	0.3501	0.3297	0.3165	0.3146	0.2934	0.2999	0.3021	0.3021
λ_2	0.3967	0.3464	0.2487	0.2515	0.2559	0.2536	0.2177	0.2179	0.2745	0.2271	0.1872	0.1870
λ_3	0.1082	0.2043	0.2898	0.2916	0.1745	0.2041	0.2145	0.2115	0.1853	0.2053	0.2319	0.2320
λ_4	0.1021	0.0455	0.0509	0.0518	0.1334	0.1149	0.1508	0.1556	0.1548	0.1676	0.1781	0.1781
β_1	0.7846	0.9075	0.9656	0.9653	0.8622	0.9635	0.9941	0.9944	0.9228	0.9947	1.0053	1.0054
β_2	0.3961	0.4735	0.4984	0.4975	0.4221	0.4881	0.5010	0.5014	0.4534	0.4977	0.4991	0.4991
$p = 6$												
λ_1	0.1583	0.1606	0.1642	0.1625	0.1866	0.1712	0.1605	0.1592	0.1365	0.1397	0.1450	0.1447
λ_2	0.3500	0.2670	0.1761	0.1787	0.2425	0.2043	0.1655	0.1675	0.2425	0.1750	0.1522	0.1515
λ_3	0.0776	0.0968	0.1614	0.1574	0.0057	0.0449	0.0897	0.0822	0.1076	0.1191	0.1471	0.1478
λ_4	0.2629	0.3275	0.2592	0.2579	0.4181	0.3746	0.2792	0.2825	0.2095	0.2412	0.1740	0.1737
λ_5	−0.0027	0.0708	0.1706	0.1796	−0.0390	0.0799	0.1547	0.1633	0.0987	0.1363	0.1579	0.1574
λ_6	0.0791	−0.0166	−0.0337	−0.0385	0.1060	0.0285	0.0503	0.0450	0.1171	0.0885	0.1232	0.1243
β_1	0.7449	0.9029	0.9835	0.9852	0.8024	0.9418	0.9846	0.9855	0.8848	0.9911	1.0036	1.0038
β_2	0.3700	0.4786	0.5224	0.5215	0.3948	0.4887	0.5035	0.5048	0.4332	0.5000	0.4984	0.4983

to a bad draw. As in the normal errors case, results are quite stable for larger n and smaller p, and typically show bias reduction due to Newton steps and near convergence after three iterations.

Tables 3 and 4 report mean squared error (MSE) for the IV estimates and iterated estimates with $N(0, 1)$ and t_6 errors, respectively. As may be expected, MSE is very high for designs that combine the largest values of p with the smallest values of n. The efficiency improvement due to the Newton step is apparent, with iterations leading to very clear improvements (i.e. reductions) in MSE. These gains can be spectacular in many cases, for example for the λ_i estimates when $p = 6$ and $n = 800$. These patterns of improvement with iteration are similar for both error distributions but the magnitude of MSE is generally much larger for t_6 errors, which features heavier tails than the normal distribution.

In Tables 5 and 6, we report the ratio of the Monte Carlo root mean squared error of $\hat{\theta}_n$ to that of $\hat{\hat{\theta}}_n^{\ell}$, $\ell = 1, 3, 6$, abbreviating this quantity to RRMSE. An RRMSE of two indicates that the RMSE of the IV estimate is twice that of the Newton iteration it is being compared to. Our results in Table 5 show that Newton iterations can lead to tremendous

Table 3
Monte Carlo mean squared error of parameter estimates. IV and iterated Newton-step estimates with $N(0, 1)$ errors. Bounded h_n. The parameter values used in the DGPs are as follows. $\beta_1 = 1$, $\beta_2 = 0.5$ and $p = 2 : \lambda_1 = 0.4, \lambda_2 = 0.5$; $p = 4 : \lambda_1 = 0.3, \lambda_i = 0.2, i = 2, 3, 4$; $p = 6 : \lambda_i = 0.15, i = 1, \ldots, 6$.

	$n = 200$				$n = 400$				$n = 800$			
$p = 2$	$\hat{\theta}_n$	$\hat{\hat{\theta}}{}^1_n$	$\hat{\hat{\theta}}{}^3_n$	$\hat{\hat{\theta}}{}^6_n$	$\hat{\theta}_n$	$\hat{\hat{\theta}}{}^1_n$	$\hat{\hat{\theta}}{}^3_n$	$\hat{\hat{\theta}}{}^6_n$	$\hat{\theta}_n$	$\hat{\hat{\theta}}{}^1_n$	$\hat{\hat{\theta}}{}^3_n$	$\hat{\hat{\theta}}{}^6_n$
λ_1	0.2708	0.0920	0.0564	0.0581	0.1338	0.0232	0.0160	0.0162	0.0595	0.0041	0.0026	0.0026
λ_2	0.2688	0.0890	0.0541	0.0544	0.1315	0.0220	0.0147	0.0147	0.0585	0.0040	0.0027	0.0027
β_1	0.0848	0.0593	0.0541	0.0571	0.0394	0.0257	0.0268	0.0277	0.0198	0.0124	0.0119	0.0119
β_2	0.0659	0.0490	0.0483	0.0517	0.0332	0.0234	0.0240	0.0246	0.0168	0.0115	0.0112	0.0112
$p = 4$												
λ_1	0.4368	0.1998	0.1416	0.1418	0.2197	0.0788	0.1164	0.1895	0.0862	0.0065	0.0035	0.0035
λ_2	1.8844	0.9538	0.6557	0.6544	0.9329	0.3404	0.3776	0.5522	0.3945	0.0351	0.0147	0.0147
λ_3	3.3021	1.9049	1.4072	1.3875	1.6625	0.5826	0.4328	0.5056	0.8284	0.0990	0.0322	0.0322
λ_4	1.5280	0.9705	0.7526	0.7441	0.7306	0.2520	0.1699	0.1824	0.3374	0.0461	0.0166	0.0165
β_2	0.1169	0.0846	0.0789	0.0797	0.0542	0.0348	0.0393	0.0485	0.0231	0.0137	0.0126	0.0126
β_2	0.0771	0.0644	0.0686	0.0704	0.0352	0.0282	0.0308	0.0360	0.0168	0.0123	0.0120	0.0120
$p = 6$												
λ_1	0.4517	0.1945	0.1279	0.1242	0.2246	0.0622	0.0309	0.0287	0.0936	0.0082	0.0035	0.0036
λ_2	1.8930	0.9716	0.6408	0.6276	0.9502	0.3064	0.1582	0.1431	0.4118	0.0469	0.0211	0.0172
λ_3	3.6096	2.0848	1.4710	1.4689	2.0160	0.8173	0.4501	0.3993	0.9385	0.1365	0.0715	0.0515
λ_4	6.6252	4.0488	2.7290	2.7963	3.6232	1.5393	0.8754	0.8247	1.6112	0.2257	0.0968	0.0853
λ_5	9.7636	6.9952	5.3933	5.5213	5.4748	2.5178	1.4660	1.3992	2.3297	0.3602	0.1310	0.1482
λ_6	4.1165	3.1102	2.5321	2.5745	1.9934	0.8924	0.5186	0.4947	0.8629	0.1333	0.0503	0.0566
β_1	0.1418	0.0972	0.0905	0.0916	0.0662	0.0377	0.0318	0.0309	0.0278	0.0148	0.0132	0.0131
β_2	0.0769	0.0660	0.0744	0.0758	0.0378	0.0286	0.0271	0.0270	0.0175	0.0128	0.0125	0.0125

Table 4
Monte Carlo mean squared error of parameter estimates. IV and iterated Newton-step estimates with t_6 errors. Bounded h_n. The parameter values used in the DGPs are as follows. $\beta_1 = 1$, $\beta_2 = 0.5$ and $p = 2 : \lambda_1 = 0.4, \lambda_2 = 0.5$; $p = 4 : \lambda_1 = 0.3, \lambda_i = 0.2, i = 2, 3, 4$; $p = 6 : \lambda_i = 0.15, i = 1, \ldots, 6$.

	$n = 200$				$n = 400$				$n = 800$			
$p = 2$	$\hat{\theta}_n$	$\hat{\hat{\theta}}{}^1_n$	$\hat{\hat{\theta}}{}^3_n$	$\hat{\hat{\theta}}{}^6_n$	$\hat{\theta}_n$	$\hat{\hat{\theta}}{}^1_n$	$\hat{\hat{\theta}}{}^3_n$	$\hat{\hat{\theta}}{}^6_n$	$\hat{\theta}_n$	$\hat{\hat{\theta}}{}^1_n$	$\hat{\hat{\theta}}{}^3_n$	$\hat{\hat{\theta}}{}^6_n$
λ_1	0.4325	0.2335	0.1675	0.1642	0.2502	0.0999	0.0765	0.0755	0.1000	0.0095	0.0030	0.0030
λ_2	0.4266	0.2262	0.1596	0.1556	0.2427	0.0936	0.0707	0.0696	0.0992	0.0092	0.0031	0.0031
β_1	0.1355	0.1005	0.0966	0.0989	0.0650	0.0472	0.0477	0.0480	0.0295	0.0179	0.0168	0.0168
β_2	0.0989	0.0778	0.0826	0.0848	0.0539	0.0383	0.0387	0.0390	0.0251	0.0164	0.0156	0.0156
$p = 4$												
λ_1	0.6466	0.3928	0.3568	0.3974	0.3465	0.1219	0.0788	0.0788	0.1533	0.0393	0.0327	0.0328
λ_2	2.5815	1.6617	1.3660	1.4574	1.4694	0.6406	0.4449	0.4718	0.7129	0.1887	0.1406	0.1407
λ_3	5.4552	3.9905	3.4192	3.4337	3.0272	1.6428	1.2006	1.2038	1.5214	0.4213	0.2783	0.2785
λ_4	2.3061	1.7137	1.4596	1.4664	1.2423	0.6693	0.4696	0.4459	0.6005	0.1721	0.1123	0.1124
β_1	0.1956	0.1553	0.1612	0.1769	0.0924	0.0612	0.0557	0.0580	0.0402	0.0246	0.0260	0.0260
β_2	0.1105	0.1007	0.1250	0.1429	0.0579	0.0441	0.0460	0.0474	0.0270	0.0192	0.0209	0.0210
$p = 6$												
λ_1	0.6070	0.3161	0.2456	0.2536	0.3653	0.1450	0.0837	0.0876	0.1474	0.0221	0.0096	0.0095
λ_2	2.4683	1.4365	1.0475	1.0406	1.4230	0.6889	0.4757	0.4834	0.6844	0.1429	0.0713	0.0721
λ_3	5.0058	3.3374	2.5153	2.5688	3.5074	2.0217	1.5395	1.6457	1.5694	0.3715	0.1384	0.1338
λ_4	9.1724	6.6282	4.9930	5.0256	6.3975	3.7914	2.6633	2.7654	2.4876	0.6532	0.3418	0.3397
λ_5	12.9389	10.3626	8.4181	8.5079	9.7583	6.2570	4.3328	4.2820	3.6504	1.0522	0.5230	0.5185
λ_6	5.3635	4.4986	3.7806	3.8367	3.6693	2.4140	1.7002	1.6941	1.5609	0.5773	0.3346	0.3307
β_1	0.2043	0.1626	0.1647	0.1711	0.1229	0.0839	0.0866	0.0932	0.0480	0.0247	0.0219	0.0219
β_2	0.1222	0.1220	0.1547	0.1622	0.0657	0.0537	0.0687	0.0735	0.0285	0.0193	0.0193	0.0193

finite sample gains in MSE. These gains are present in 100% of the cases considered, but are generally larger for the spatial parameters λ_i than the regression parameters β_i.

We discuss the spatial parameter estimates first. Note that for greater sample sizes we have greater MSE gains, often the gains more than doubling from $n = 200$ to $n = 800$, and sometimes even tripling. As observed for the means in Table 1, there is usually not much to choose from between the third and sixth iterations. With and $n = 800$ we nearly always obtain Newton estimates with RMSE a quarter of that for IV, and occasionally even a fifth of the IV RMSE. In most cases three iterations are enough to achieve these superb gains.

These patterns for the λ_i qualitatively repeat themselves when the errors are t_6, as seen in Table 6. In this case when $n = 800$ we achieve RMSE improvements over IV of a factor of 2.15 always when three iterations are carried out, with factors of three commonly seen and one case with nearly a fourfold improvement. The factors of efficiency improvement that we observe in our results can dominate similar precedents in other settings. Indeed, the greatest relative root MSE

Table 5
Monte Carlo relative root MSE of IV estimates to iterated Newton-step estimates with $N(0,1)$ errors. Bounded h_n. The parameter values used in the DGPs are as follows. $\beta_1 = 1$, $\beta_2 = 0.5$ and $p = 2 : \lambda_1 = 0.4, \lambda_2 = 0.5$; $p = 4 : \lambda_1 = 0.3, \lambda_i = 0.2, i = 2, 3, 4$; $p = 6 : \lambda_i = 0.15, i = 1, \ldots, 6$.

	$n = 200$			$n = 400$			$n = 800$		
$p = 2$	$\frac{RMSE(\hat\theta_n)}{RMSE(\hat{\hat\theta}_n^1)}$	$\frac{RMSE(\hat\theta_n)}{RMSE(\hat{\hat\theta}_n^3)}$	$\frac{RMSE(\hat\theta_n)}{RMSE(\hat{\hat\theta}_n^6)}$	$\frac{RMSE(\hat\theta_n)}{RMSE(\hat{\hat\theta}_n^1)}$	$\frac{RMSE(\hat\theta_n)}{RMSE(\hat{\hat\theta}_n^3)}$	$\frac{RMSE(\hat\theta_n)}{RMSE(\hat{\hat\theta}_n^6)}$	$\frac{RMSE(\hat\theta_n)}{RMSE(\hat{\hat\theta}_n^1)}$	$\frac{RMSE(\hat\theta_n)}{RMSE(\hat{\hat\theta}_n^3)}$	$\frac{RMSE(\hat\theta_n)}{RMSE(\hat{\hat\theta}_n^6)}$
λ_1	1.7153	2.1714	2.1687	2.4010	2.9156	2.8782	3.7949	4.7436	4.7428
λ_2	1.7376	2.2054	2.2259	2.4464	2.9881	2.9891	3.8068	4.6181	4.6176
β_1	1.1957	1.2528	1.2263	1.2386	1.2355	1.1936	1.2631	1.2884	1.2882
β_2	1.1593	1.1720	1.1398	1.1907	1.1912	1.1630	1.2086	1.2257	1.2256
$p = 4$									
λ_1	1.4788	1.7373	1.7560	1.6691	1.5114	1.1593	3.6541	4.9734	4.9985
λ_2	1.4056	1.6729	1.6978	1.6553	1.6775	1.3799	3.3546	5.1613	5.1771
λ_3	1.3166	1.5146	1.5430	1.6893	1.9765	1.8648	2.8921	5.0092	5.0753
λ_4	1.2547	1.4118	1.4332	1.7027	2.0674	2.0299	2.7064	4.4333	4.5202
β_1	1.1754	1.2238	1.2112	1.2485	1.2159	1.0953	1.2991	1.3505	1.3530
β_2	1.0943	1.0783	1.0475	1.1169	1.0951	1.0150	1.1695	1.1853	1.1850
$p = 6$									
λ_1	1.5238	1.8631	1.8977	1.9000	2.5526	2.8017	3.3835	5.1829	5.1846
λ_2	1.3958	1.6874	1.7349	1.7611	2.3462	2.5433	2.9643	4.3448	4.7316
λ_3	1.3158	1.5418	1.5722	1.5706	2.0309	2.2114	2.6220	3.5252	4.0337
λ_4	1.2792	1.5372	1.5509	1.5342	1.9460	2.1012	2.6719	4.0257	4.2878
λ_5	1.1814	1.3301	1.3422	1.4746	1.8744	1.9866	2.5431	4.2191	4.0369
λ_6	1.1505	1.2624	1.2736	1.4946	1.9244	2.0053	2.5447	4.1692	3.9460
β_1	1.2082	1.2674	1.2433	1.3249	1.4373	1.4601	1.3712	1.4450	1.4542
β_2	1.0795	1.0387	1.0058	1.1487	1.1796	1.1850	1.1715	1.1854	1.1850

Table 6
Monte Carlo relative root MSE of IV estimates to iterated Newton-step estimates with t_6 errors. Bounded h_n. The parameter values used in the DGPs are as follows. $\beta_1 = 1$, $\beta_2 = 0.5$ and $p = 2 : \lambda_1 = 0.4, \lambda_2 = 0.5$; $p = 4 : \lambda_1 = 0.3, \lambda_i = 0.2, i = 2, 3, 4$; $p = 6 : \lambda_i = 0.15, i = 1, \ldots, 6$.

	$n = 200$			$n = 400$			$n = 800$		
$p = 2$	$\frac{RMSE(\hat\theta_n)}{RMSE(\hat{\hat\theta}_n^1)}$	$\frac{RMSE(\hat\theta_n)}{RMSE(\hat{\hat\theta}_n^3)}$	$\frac{RMSE(\hat\theta_n)}{RMSE(\hat{\hat\theta}_n^6)}$	$\frac{RMSE(\hat\theta_n)}{RMSE(\hat{\hat\theta}_n^1)}$	$\frac{RMSE(\hat\theta_n)}{RMSE(\hat{\hat\theta}_n^3)}$	$\frac{RMSE(\hat\theta_n)}{RMSE(\hat{\hat\theta}_n^6)}$	$\frac{RMSE(\hat\theta_n)}{RMSE(\hat{\hat\theta}_n^1)}$	$\frac{RMSE(\hat\theta_n)}{RMSE(\hat{\hat\theta}_n^3)}$	$\frac{RMSE(\hat\theta_n)}{RMSE(\hat{\hat\theta}_n^6)}$
λ_1	1.3611	1.5838	1.6234	1.5828	1.7797	1.8209	3.2483	5.7778	5.7832
λ_2	1.3732	1.6083	1.6552	1.6106	1.8208	1.8673	3.2901	5.6271	5.6293
β_1	1.1611	1.1917	1.1769	1.1740	1.1715	1.1650	1.2833	1.3258	1.3261
β_2	1.1274	1.1032	1.0848	1.1867	1.1876	1.1758	1.2365	1.2693	1.2701
$p = 4$									
λ_1	1.2830	1.3679	1.2971	1.6858	2.0797	2.0964	1.9760	2.1586	2.1630
λ_2	1.2464	1.3772	1.3447	1.5146	1.8212	1.7783	1.9437	2.2461	2.2509
λ_3	1.1692	1.2565	1.2630	1.3575	1.5761	1.5905	1.9002	2.3358	2.3374
λ_4	1.1600	1.2507	1.2569	1.3624	1.5991	1.6645	1.8678	2.3099	2.3115
β_1	1.1222	1.1241	1.0647	1.2284	1.2979	1.2687	1.2789	1.2516	1.2422
β_2	1.0478	0.9682	0.8944	1.1461	1.1318	1.1094	1.1873	1.1448	1.1334
$p = 6$									
λ_1	1.3856	1.5766	1.5560	1.5873	2.0536	2.0571	2.5801	3.8920	3.9340
λ_2	1.3108	1.5128	1.5439	1.4373	1.7032	1.7232	2.1881	3.0784	3.0942
λ_3	1.2247	1.3957	1.4010	1.3171	1.5076	1.4729	2.0554	3.2351	3.4299
λ_4	1.1764	1.3411	1.3528	1.2990	1.5323	1.5328	1.9515	2.6465	2.7122
λ_5	1.1174	1.2310	1.2361	1.2488	1.4732	1.5106	1.8626	2.5997	2.6616
λ_6	1.0919	1.1848	1.1859	1.2329	1.4490	1.4734	1.6442	2.1280	2.1812
β_1	1.1208	1.1325	1.0967	1.2099	1.2201	1.1582	1.3931	1.4778	1.4808
β_2	1.0007	0.9090	0.8726	1.1060	1.0196	0.9512	1.2142	1.2169	1.2145

improvement that Robinson (2005) finds in his fractional time series setting is $\sqrt{1/0.23} = 2.085$ (see Table 4 of that paper).

Moving to the estimates of the regression parameters β_1 and β_2, in both Tables 5 and 6 we see almost universal improvement over IV. The exceptions are four cases out of a total of 54 in Table 6, for the t_6 case. These RMSE gains are not as spectacular as for the λ_i, but are generally noticeably large as both n and p increase. Indeed, for $n = 800$ we observe that the RMSE for the IV estimate can sometimes be almost one and a half times as large as the Newton iterations when $p = 6$ and $n = 800$. For $n \geq 400$, IV performs worse than the Newton iterations almost uniformly (there are only two exceptions for t_6 errors) over both β_1 and β_2, the values of p, the number of iterations and the error distribution. Thus

Table 7
Monte Carlo relative root MSE of ML estimates to iterated Newton-step estimates with $N(0,1)$ errors. Bounded h_n. The parameter values used in the DGPs are as follows. $\beta_1 = 1$, $\beta_2 = 0.5$ and $p = 2 : \lambda_1 = 0.4, \lambda_2 = 0.5$; $p = 4 : \lambda_1 = 0.3, \lambda_i = 0.2, i = 2, 3, 4$; $p = 6 : \lambda_i = 0.15, i = 1, \ldots, 6$.

	$n = 200$			$n = 400$			$n = 800$		
$p = 2$	$\frac{\text{RMSE}(\mathring{\theta}_n)}{\text{RMSE}(\hat{\hat{\theta}}_n^1)}$	$\frac{\text{RMSE}(\mathring{\theta}_n)}{\text{RMSE}(\hat{\hat{\theta}}_n^3)}$	$\frac{\text{RMSE}(\mathring{\theta}_n)}{\text{RMSE}(\hat{\hat{\theta}}_n^6)}$	$\frac{\text{RMSE}(\mathring{\theta}_n)}{\text{RMSE}(\hat{\hat{\theta}}_n^1)}$	$\frac{\text{RMSE}(\mathring{\theta}_n)}{\text{RMSE}(\hat{\hat{\theta}}_n^3)}$	$\frac{\text{RMSE}(\mathring{\theta}_n)}{\text{RMSE}(\hat{\hat{\theta}}_n^6)}$	$\frac{\text{RMSE}(\mathring{\theta}_n)}{\text{RMSE}(\hat{\hat{\theta}}_n^1)}$	$\frac{\text{RMSE}(\mathring{\theta}_n)}{\text{RMSE}(\hat{\hat{\theta}}_n^3)}$	$\frac{\text{RMSE}(\mathring{\theta}_n)}{\text{RMSE}(\hat{\hat{\theta}}_n^6)}$
λ_1	0.3415	0.4322	0.4317	0.4811	0.5842	0.5767	0.8024	1.0030	1.0028
λ_2	0.3562	0.4521	0.4563	0.5027	0.6141	0.6143	0.8269	1.0031	1.0030
β_1	0.9036	0.9468	0.9267	0.9432	0.9409	0.9090	0.9788	0.9984	0.9983
β_2	0.9637	0.9743	0.9475	0.9725	0.9729	0.9499	0.9851	0.9991	0.9989
$p = 4$									
λ_1	0.2626	0.3085	0.3118	0.2962	0.2682	0.2057	0.7307	0.9945	0.9995
λ_2	0.2421	0.2881	0.2924	0.2966	0.3006	0.2472	0.6480	0.9970	1.0001
λ_3	0.2567	0.2954	0.3009	0.3378	0.3952	0.3729	0.5534	0.9585	0.9711
λ_4	0.2384	0.2683	0.2723	0.3377	0.4101	0.4026	0.5409	0.8861	0.9035
β_1	0.7688	0.8004	0.7922	0.8221	0.8006	0.7212	0.9547	0.9925	0.9944
β_2	0.8535	0.8410	0.8170	0.9085	0.8908	0.8256	0.9831	0.9964	0.9961
$p = 6$									
λ_1	0.2558	0.3127	0.3185	0.3193	0.4290	0.4708	0.6178	0.9464	0.9467
λ_2	0.2350	0.2841	0.2921	0.2992	0.3986	0.4321	0.5348	0.7838	0.8536
λ_3	0.2388	0.2798	0.2853	0.2777	0.3591	0.3911	0.4629	0.6223	0.7121
λ_4	0.2199	0.2643	0.2666	0.2582	0.3276	0.3537	0.4876	0.7347	0.7825
λ_5	0.1991	0.2242	0.2262	0.2373	0.3017	0.3198	0.4650	0.7715	0.7382
λ_6	0.1937	0.2126	0.2145	0.2602	0.3350	0.3491	0.4869	0.7977	0.7550
β_1	0.7325	0.7684	0.7538	0.8029	0.8711	0.8849	0.9384	0.9889	0.9952
β_2	0.8497	0.8176	0.7917	0.9267	0.9516	0.9560	0.9866	0.9982	0.9979

there is evidence of the usefulness of Newton iterations even for the regression parameters, albeit the gains are greater for the spatial parameters.

Finally, we also present the RRMSE of MLE (denoted $\mathring{\theta}_n$) to our proposed iterated estimates in Table 7 for $N(0,1)$ errors. Naturally, we anticipate MLE to outperform iterated IV estimates for smaller sample sizes and, because our iterations target the MLE limiting covariance matrix, a reasonable aim is to approach the RMSE of the MLE as n grows larger. Indeed, we find that this is the case. Recall that our estimates are designed to approximate but not outperform MLE: the main focus of the paper is computational simplicity. Our estimates are available in closed form and can be computed much faster than those requiring grid search and inversion of an $n \times n$ matrix. Thus, approaching the MLE in RMSE as n grows is an encouraging and desirable property of our estimates. Finally, we observe that the RMSE of $\mathring{\beta}_n$ is much closer to the iterated estimates than is the case for $\mathring{\lambda}_n$. For the latter, larger sample sizes are needed for the RRMSE to approach unity.

4.2. Growing number of neighbours (divergent h_n)

In this section we explore the performance of the Newton step estimates when the number of neighbours diverges with sample size, i.e. $h_n \to \infty$. This design, with diverging h_n, also allows us to study the performance of iterations on OLS starting values. For each $i = 1, \ldots, p$, we generate a $n \times n$ matrix W_{in}^* as $w_{rs,in}^* = \Phi\left(-d_{rs,i}\right) I\left(c_{rs,i} < n^{1/3}/100\right)$ if $r \neq s$, and $w_{rr,in}^* = 0$, where $\Phi(\cdot)$ is the standard normal cdf, $d_{rs,i} \sim$iid $U[-3,3]$, and $c_{rs,i} \sim$iid $U[0,1]$. This construction generates W_{in}^* with approximately $n^{1/3}$% (up to closest integer) nonzero elements. These W_{in}^* are then symmetrized and normalized by spectral norm to ensure stability, yielding the final set of W_{in} that we employ. The remaining design details are as in the previous subsection. To conserve space, we report results only for $N(0,1)$ errors.

Tables 8 and 9 display the Monte Carlo mean of $\hat{\theta}_n$, $\hat{\hat{\theta}}_n^1$, $\hat{\hat{\theta}}_n^3$, $\tilde{\theta}_n$, $\tilde{\tilde{\theta}}_n^1$ and $\tilde{\tilde{\theta}}_n^3$. Convergence of iterations is achieved after three Newton steps, so we do not report the sixth iteration as in the previous subsection. In fact, convergence is practically fully achieved by just a single iteration with the IV starting values $\hat{\theta}_n$, as Table 8 indicates. Examining Table 9 suggests that a third iteration has more influence for OLS starting values, but modestly so. Tables 10 and 11 report MSE for the same sets of estimates and we find a similar pattern: for IV starting values one iteration seems to do the job and reduces MSE. On the other hand, for OLS starting values the first iteration increases MSE but the third iteration reduces it, following which performance is stable and so we do not report further iterations.

In Table 12 we report RRMSE of the estimates studies above. We notice that IV estimates improve in MSE with a single Newton step, and subsequent iterations do not help much, because convergence is achieved. On the other hand, when starting with OLS values $\tilde{\theta}_n$, further iterations are beneficial and yield more efficient estimates. Convergence is completely achieved after three iterations in this case. We also find that Newton steps, whether they commence from $\hat{\theta}_n$ or $\tilde{\theta}_n$, give greater efficiency gains for the spatial parameters λ_i rather than the regression coefficients β_i. This matches the results in the previous subsection. Because the λ_i correspond to the potentially endogenous spatial lags $W_{in}y$, we might expect initial estimates of these to have greater potential for improvement compared to the β_i.

Table 8
Monte Carlo mean of parameter estimates. IV and iterated Newton-step estimates with $N(0, 1)$ errors. Divergent h_n. The parameter values used in the DGPs are as follows. $\beta_1 = 1$, $\beta_2 = 0.5$ and $p = 2 : \lambda_1 = 0.4, \lambda_2 = 0.5$; $p = 4 : \lambda_1 = 0.3, \lambda_i = 0.2, i = 2, 3, 4$; $p = 6 : \lambda_i = 0.15, i = 1, \ldots, 6$.

	$n = 200$			$n = 400$			$n = 800$		
$p = 2$	$\hat{\theta}_n$	$\hat{\hat{\theta}}_n^1$	$\hat{\hat{\theta}}_n^3$	$\hat{\theta}_n$	$\hat{\hat{\theta}}_n^1$	$\hat{\hat{\theta}}_n^3$	$\hat{\theta}_n$	$\hat{\hat{\theta}}_n^1$	$\hat{\hat{\theta}}_n^3$
λ_1	0.3992	0.3934	0.3933	0.3985	0.3952	0.3952	0.4012	0.3988	0.3988
λ_2	0.5019	0.4945	0.4944	0.5015	0.4984	0.4984	0.4974	0.4967	0.4967
β_1	0.9892	1.0241	1.0245	0.9958	1.0145	1.0145	1.0040	1.0137	1.0137
β_2	0.4875	0.5207	0.5210	0.4985	0.5171	0.5171	0.5027	0.5120	0.5120
$p = 4$									
λ_1	0.3002	0.2952	0.2952	0.2993	0.2966	0.2966	0.2981	0.2982	0.2982
λ_2	0.1990	0.1931	0.1930	0.1963	0.1927	0.1927	0.2059	0.2024	0.2024
λ_3	0.1952	0.1942	0.1941	0.2018	0.2032	0.2032	0.1983	0.1986	0.1986
λ_4	0.2023	0.1987	0.1987	0.2009	0.1985	0.1985	0.1980	0.1975	0.1975
β_1	0.9885	1.0262	1.0265	1.0019	1.0252	1.0253	0.9941	1.0051	1.0051
β_2	0.5048	0.5404	0.5407	0.5008	0.5198	0.5198	0.4977	0.5091	0.5091
$p = 6$									
λ_1	0.1469	0.1455	0.1454	0.1578	0.1537	0.1537	0.1542	0.1503	0.1503
λ_2	0.1436	0.1436	0.1436	0.1462	0.1458	0.1458	0.1504	0.1485	0.1485
λ_3	0.1548	0.1529	0.1529	0.1466	0.1462	0.1462	0.1436	0.1473	0.1473
λ_4	0.1486	0.1450	0.1449	0.1552	0.1532	0.1532	0.1507	0.1529	0.1529
λ_5	0.1519	0.1480	0.1479	0.1466	0.1434	0.1434	0.1536	0.1526	0.1526
λ_6	0.1501	0.1415	0.1415	0.1465	0.1496	0.1496	0.1462	0.1437	0.1437
β_1	1.0026	1.0518	1.0524	0.9977	1.0199	1.0199	1.0045	1.0152	1.0152
β_2	0.4966	0.5434	0.5440	0.4960	0.5153	0.5153	0.4950	0.5057	0.5057

Table 9
Monte Carlo mean of parameter estimates. OLS and iterated Newton-step estimates with $N(0, 1)$ errors. Divergent h_n. The parameter values used in the DGPs are as follows. $\beta_1 = 1$, $\beta_2 = 0.5$ and $p = 2 : \lambda_1 = 0.4, \lambda_2 = 0.5$; $p = 4 : \lambda_1 = 0.3, \lambda_i = 0.2, i = 2, 3, 4$; $p = 6 : \lambda_i = 0.15, i = 1, \ldots, 6$.

	$n = 200$			$n = 400$			$n = 800$		
$p = 2$	$\tilde{\theta}_n$	$\tilde{\tilde{\theta}}_n^1$	$\tilde{\tilde{\theta}}_n^3$	$\tilde{\theta}_n$	$\tilde{\tilde{\theta}}_n^1$	$\tilde{\tilde{\theta}}_n^3$	$\tilde{\theta}_n$	$\tilde{\tilde{\theta}}_n^1$	$\tilde{\tilde{\theta}}_n^3$
λ_1	0.4017	0.3956	0.3933	0.3967	0.3936	0.3952	0.3977	0.3956	0.3988
λ_2	0.5093	0.5014	0.4944	0.5067	0.5034	0.4984	0.5021	0.5012	0.4967
β_1	0.9643	1.0010	1.0245	0.9854	1.0045	1.0145	0.9998	1.0096	1.0137
β_2	0.4635	0.4983	0.5210	0.4889	0.5078	0.5171	0.4990	0.5083	0.5120
$p = 4$									
λ_1	0.3062	0.3008	0.2952	0.3035	0.3006	0.2966	0.3034	0.3032	0.2982
λ_2	0.1967	0.1910	0.1930	0.1941	0.1907	0.1927	0.2019	0.1987	0.2024
λ_3	0.1966	0.1954	0.1941	0.2034	0.2046	0.2032	0.1982	0.1986	0.1986
λ_4	0.2005	0.1972	0.1987	0.1980	0.1959	0.1985	0.1968	0.1964	0.1975
β_1	0.9827	1.0206	1.0265	0.9994	1.0228	1.0253	0.9943	1.0053	1.0051
β_2	0.4979	0.5338	0.5407	0.4997	0.5186	0.5198	0.4971	0.5084	0.5091
$p = 6$									
λ_1	0.1480	0.1464	0.1454	0.1556	0.1516	0.1537	0.1502	0.1465	0.1503
λ_2	0.1443	0.1443	0.1436	0.1478	0.1473	0.1458	0.1493	0.1475	0.1485
λ_3	0.1562	0.1543	0.1529	0.1468	0.1464	0.1462	0.1479	0.1513	0.1473
λ_4	0.1478	0.1442	0.1449	0.1541	0.1522	0.1532	0.1536	0.1556	0.1529
λ_5	0.1524	0.1483	0.1479	0.1439	0.1409	0.1434	0.1530	0.1521	0.1526
λ_6	0.1486	0.1402	0.1415	0.1506	0.1534	0.1496	0.1444	0.1420	0.1437
β_1	1.0012	1.0503	1.0524	0.9996	1.0215	1.0199	1.0056	1.0162	1.0152
β_2	0.4951	0.5418	0.5440	0.4959	0.5150	0.5153	0.4959	0.5066	0.5057

4.3. Heteroskedastic errors

In this design, we confirm the robustness of our findings to heteroskedasticity in the error distribution. We generate the errors using multiplicative heteroskedasticity via the regressors, and report only the bounded h_n weight matrices of Section 4.1 and designs with Gaussian errors to conserve space. Specifically, we employ a $N(0, h_{jn})$ distribution for the errors, where $h_{jn} = n\left(\sum_{r=1}^{n}(|x_{r1}| + |x_{r2}|)\right)^{-1}\left(|x_{j1}| + |x_{j2}|\right)$, see Liu and Yang (2015) and also Lin and Lee (2010). Monte Carlo mean, mean squared error and RRMSE of IV estimates to iterated Newton step estimates are presented in Tables 13–15. We find the same qualitative patterns as were observed for the homoskedastic designs presented earlier, with the 'rogue' IV estimate for λ_5 appearing again because we start our simulations from the same seed. As far as quantitative results are concerned, the improvements due to the Newton step are generally smaller than the homoskedastic case but still substantial.

Table 10
Monte Carlo mean squared error of parameter estimates. IV and iterated Newton-step estimates with $N(0, 1)$ errors. Divergent h_n. The parameter values used in the DGPs are as follows. $\beta_1 = 1$, $\beta_2 = 0.5$ and $p = 2 : \lambda_1 = 0.4, \lambda_2 = 0.5$; $p = 4 : \lambda_1 = 0.3, \lambda_i = 0.2, i = 2, 3, 4$; $p = 6 : \lambda_i = 0.15, i = 1, \ldots, 6$.

	$n = 200$			$n = 400$			$n = 800$		
$p = 2$	$\hat{\theta}_n$	$\hat{\hat{\theta}}_n^1$	$\hat{\hat{\theta}}_n^3$	$\hat{\theta}_n$	$\hat{\hat{\theta}}_n^1$	$\hat{\hat{\theta}}_n^3$	$\hat{\theta}_n$	$\hat{\hat{\theta}}_n^1$	$\hat{\hat{\theta}}_n^3$
λ_1	0.0051	0.0040	0.0040	0.0046	0.0037	0.0037	0.0063	0.0051	0.0051
λ_2	0.0053	0.0041	0.0041	0.0046	0.0037	0.0037	0.0062	0.0049	0.0049
β_1	0.0571	0.0541	0.0540	0.0290	0.0284	0.0284	0.0151	0.0151	0.0151
β_2	0.0532	0.0498	0.0497	0.0286	0.0280	0.0280	0.0145	0.0143	0.0143
$p = 4$									
λ_1	0.0087	0.0063	0.0063	0.0075	0.0060	0.0060	0.0082	0.0065	0.0065
λ_2	0.0091	0.0065	0.0065	0.0078	0.0059	0.0059	0.0093	0.0072	0.0072
λ_3	0.0083	0.0063	0.0063	0.0077	0.0058	0.0058	0.0076	0.0059	0.0059
λ_4	0.0091	0.0069	0.0069	0.0076	0.0058	0.0058	0.0080	0.0068	0.0068
β_1	0.0522	0.0500	0.0498	0.0335	0.0328	0.0328	0.0152	0.0149	0.0149
β_2	0.0528	0.0515	0.0514	0.0283	0.0281	0.0280	0.0149	0.0146	0.0146
$p = 6$									
λ_1	0.0093	0.0070	0.0070	0.0080	0.0063	0.0063	0.0099	0.0077	0.0077
λ_2	0.0090	0.0069	0.0069	0.0086	0.0068	0.0068	0.0098	0.0078	0.0078
λ_3	0.0074	0.0058	0.0058	0.0089	0.0069	0.0069	0.0086	0.0069	0.0069
λ_4	0.0095	0.0073	0.0073	0.0091	0.0069	0.0069	0.0084	0.0066	0.0066
λ_5	0.0083	0.0062	0.0062	0.0088	0.0068	0.0068	0.0088	0.0073	0.0073
λ_6	0.0092	0.0074	0.0074	0.0087	0.0066	0.0066	0.0079	0.0065	0.0065
β_1	0.0642	0.0628	0.0625	0.0283	0.0277	0.0277	0.0150	0.0149	0.0149
β_2	0.0610	0.0584	0.0582	0.0333	0.0328	0.0328	0.0163	0.0161	0.0161

Table 11
Monte Carlo mean squared error of parameter estimates. OLS and iterated Newton-step estimates with $N(0, 1)$ errors. Divergent h_n. The parameter values used in the DGPs are as follows. $\beta_1 = 1$, $\beta_2 = 0.5$ and $p = 2 : \lambda_1 = 0.4, \lambda_2 = 0.5$; $p = 4 : \lambda_1 = 0.3, \lambda_i = 0.2, i = 2, 3, 4$; $p = 6 : \lambda_i = 0.15, i = 1, \ldots, 6$.

	$n = 200$			$n = 400$			$n = 800$		
$p = 2$	$\tilde{\theta}_n$	$\tilde{\tilde{\theta}}_n^1$	$\tilde{\tilde{\theta}}_n^3$	$\tilde{\theta}_n$	$\tilde{\tilde{\theta}}_n^1$	$\tilde{\tilde{\theta}}_n^3$	$\tilde{\theta}_n$	$\tilde{\tilde{\theta}}_n^1$	$\tilde{\tilde{\theta}}_n^3$
λ_1	0.0045	0.0052	0.0040	0.0042	0.0051	0.0037	0.0057	0.0066	0.0051
λ_2	0.0046	0.0050	0.0041	0.0042	0.0051	0.0037	0.0055	0.0063	0.0049
β_1	0.0574	0.0547	0.0540	0.0291	0.0286	0.0284	0.0151	0.0151	0.0151
β_2	0.0527	0.0496	0.0497	0.0286	0.0281	0.0280	0.0144	0.0143	0.0143
$p = 4$									
λ_1	0.0074	0.0085	0.0063	0.0069	0.0085	0.0060	0.0074	0.0095	0.0065
λ_2	0.0077	0.0090	0.0065	0.0068	0.0081	0.0059	0.0082	0.0092	0.0072
λ_3	0.0075	0.0088	0.0063	0.0068	0.0084	0.0058	0.0066	0.0079	0.0059
λ_4	0.0082	0.0098	0.0069	0.0067	0.0081	0.0058	0.0076	0.0093	0.0068
β_1	0.0520	0.0507	0.0498	0.0331	0.0329	0.0328	0.0151	0.0150	0.0149
β_2	0.0522	0.0516	0.0514	0.0283	0.0283	0.0280	0.0148	0.0146	0.0146
$p = 6$									
λ_1	0.0083	0.0094	0.0070	0.0072	0.0085	0.0063	0.0088	0.0104	0.0077
λ_2	0.0081	0.0093	0.0069	0.0078	0.0093	0.0068	0.0088	0.0103	0.0078
λ_3	0.0068	0.0079	0.0058	0.0081	0.0098	0.0069	0.0077	0.0087	0.0069
λ_4	0.0086	0.0097	0.0073	0.0079	0.0093	0.0069	0.0074	0.0084	0.0066
λ_5	0.0073	0.0083	0.0062	0.0077	0.0091	0.0068	0.0082	0.0099	0.0073
λ_6	0.0088	0.0102	0.0074	0.0076	0.0089	0.0066	0.0072	0.0089	0.0065
β_1	0.0640	0.0641	0.0625	0.0280	0.0279	0.0277	0.0149	0.0150	0.0149
β_2	0.0604	0.0592	0.0582	0.0333	0.0331	0.0328	0.0163	0.0162	0.0161

5. Empirical illustration

In this small empirical illustration we show that the Newton step estimates perform well in practice and can lead to more precise estimation. The example is based on Kolympiris et al. (2011) (KKM), and is also studied in Gupta and Robinson (2015). KKM seek to model the venture capital funding (provided by venture capital firms (VCFs)) for dedicated biotechnology firms (DBFs) with a SAR model. The hypothesis is that the level of VC funding for a DBF increases with the number of VCFs located in close proximity. Denoting by d_{lk} the distance in miles between the lth and kth DBFs, we

Table 12
Monte Carlo relative root MSE of parameter estimates. IV, OLS and iterated Newton-step estimates with $N(0, 1)$ errors. Divergent h_n. The parameter values used in the DGPs are as follows. $\beta_1 = 1$, $\beta_2 = 0.5$ and $p = 2 : \lambda_1 = 0.4, \lambda_2 = 0.5$; $p = 4 : \lambda_1 = 0.3, \lambda_i = 0.2, i = 2, 3, 4$; $p = 6 : \lambda_i = 0.15, i = 1, \ldots, 6$.

	$n = 200$				$n = 400$				$n = 800$			
$p = 2$	$\frac{RMSE(\hat{\theta}_n)}{RMSE(\hat{\hat{\theta}}_n^1)}$	$\frac{RMSE(\hat{\theta}_n)}{RMSE(\hat{\hat{\theta}}_n^3)}$	$\frac{RMSE(\tilde{\theta}_n)}{RMSE(\tilde{\tilde{\theta}}_n^1)}$	$\frac{RMSE(\tilde{\theta}_n)}{RMSE(\tilde{\tilde{\theta}}_n^3)}$	$\frac{RMSE(\hat{\theta}_n)}{RMSE(\hat{\hat{\theta}}_n^1)}$	$\frac{RMSE(\hat{\theta}_n)}{RMSE(\hat{\hat{\theta}}_n^3)}$	$\frac{RMSE(\tilde{\theta}_n)}{RMSE(\tilde{\tilde{\theta}}_n^1)}$	$\frac{RMSE(\tilde{\theta}_n)}{RMSE(\tilde{\tilde{\theta}}_n^3)}$	$\frac{RMSE(\hat{\theta}_n)}{RMSE(\hat{\hat{\theta}}_n^1)}$	$\frac{RMSE(\hat{\theta}_n)}{RMSE(\hat{\hat{\theta}}_n^3)}$	$\frac{RMSE(\tilde{\theta}_n)}{RMSE(\tilde{\tilde{\theta}}_n^1)}$	$\frac{RMSE(\tilde{\theta}_n)}{RMSE(\tilde{\tilde{\theta}}_n^3)}$
λ_1	1.1211	1.1214	0.9365	1.0612	1.1141	1.1141	0.9056	1.0592	1.1172	1.1172	0.9312	1.0634
λ_2	1.1326	1.1331	0.9592	1.0643	1.1115	1.1115	0.9072	1.0646	1.1275	1.1275	0.9328	1.0629
β_1	1.0271	1.0282	1.0242	1.0310	1.0112	1.0113	1.0100	1.0132	1.0013	1.0013	1.0003	1.0008
β_2	1.0331	1.0342	1.0308	1.0299	1.0104	1.0105	1.0089	1.0101	1.0041	1.0041	1.0033	1.0019
$p = 4$												
λ_1	1.1765	1.1763	0.9364	1.0864	1.1122	1.1121	0.8988	1.0702	1.1222	1.1221	0.8853	1.0662
λ_2	1.1771	1.1772	0.9264	1.0885	1.1485	1.1484	0.9172	1.0707	1.1425	1.1424	0.9419	1.0672
λ_3	1.1434	1.1433	0.9192	1.0844	1.1519	1.1518	0.8974	1.0770	1.1334	1.1333	0.9147	1.0591
λ_4	1.1470	1.1469	0.9168	1.0944	1.1440	1.1439	0.9073	1.0710	1.0850	1.0849	0.9056	1.0591
β_1	1.0215	1.0232	1.0129	1.0218	1.0105	1.0106	1.0033	1.0056	1.0099	1.0099	1.0049	1.0083
β_2	1.0122	1.0132	1.0057	1.0073	1.0043	1.0044	1.0004	1.0042	1.0088	1.0088	1.0066	1.0055
$p = 6$												
λ_1	1.1473	1.1470	0.9384	1.0833	1.1320	1.1319	0.9184	1.0725	1.1304	1.1303	0.9189	1.0665
λ_2	1.1427	1.1424	0.9359	1.0870	1.1247	1.1245	0.9184	1.0733	1.1168	1.1167	0.9281	1.0627
λ_3	1.1256	1.1253	0.9254	1.0804	1.1299	1.1297	0.9093	1.0774	1.1222	1.1221	0.9380	1.0586
λ_4	1.1447	1.1446	0.9463	1.0891	1.1493	1.1491	0.9233	1.0698	1.1301	1.1301	0.9343	1.0568
λ_5	1.1591	1.1590	0.9399	1.0903	1.1418	1.1416	0.9199	1.0700	1.0964	1.0963	0.9104	1.0610
λ_6	1.1177	1.1176	0.9261	1.0896	1.1500	1.1498	0.9224	1.0741	1.1055	1.1054	0.9023	1.0576
β_1	1.0112	1.0132	0.9993	1.0119	1.0108	1.0109	1.0025	1.0058	1.0008	1.0008	0.9968	1.0001
β_2	1.0219	1.0238	1.0099	1.0187	1.0081	1.0082	1.0033	1.0084	1.0067	1.0067	1.0034	1.0070

Table 13
Monte Carlo mean of parameter estimates. IV and iterated Newton-step estimates with heteroskedastic zero mean Gaussian errors. Bounded h_n. The parameter values used in the DGPs are as follows. $\beta_1 = 1$, $\beta_2 = 0.5$ and $p = 2 : \lambda_1 = 0.4, \lambda_2 = 0.5$; $p = 4 : \lambda_1 = 0.3, \lambda_i = 0.2, i = 2, 3, 4$; $p = 6 : \lambda_i = 0.15, i = 1, \ldots, 6$.

	$n = 200$				$n = 400$				$n = 800$			
$p = 2$	$\hat{\theta}_n$	$\hat{\hat{\theta}}_n^1$	$\hat{\hat{\theta}}_n^3$	$\hat{\hat{\theta}}_n^6$	$\hat{\theta}_n$	$\hat{\hat{\theta}}_n^1$	$\hat{\hat{\theta}}_n^3$	$\hat{\hat{\theta}}_n^6$	$\hat{\theta}_n$	$\hat{\hat{\theta}}_n^1$	$\hat{\hat{\theta}}_n^3$	$\hat{\hat{\theta}}_n^6$
λ_1	0.4503	0.4406	0.4215	0.4191	0.4223	0.4171	0.4057	0.4053	0.4229	0.4070	0.4007	0.4005
λ_2	0.4572	0.4588	0.4757	0.4781	0.4815	0.4817	0.4923	0.4926	0.4793	0.4925	0.4983	0.4984
β_1	0.9308	0.9957	1.0149	1.0151	0.9592	1.0003	1.0084	1.0086	0.9835	1.0059	1.0101	1.0103
β_2	0.4583	0.5061	0.5176	0.5169	0.4850	0.5134	0.5177	0.5179	0.4872	0.5023	0.5050	0.5052
$p = 4$												
λ_1	0.3051	0.3085	0.3133	0.3118	0.3220	0.3063	0.3034	0.3016	0.3204	0.3044	0.3029	0.3027
λ_2	0.3539	0.2838	0.2138	0.2063	0.2422	0.2327	0.1913	0.1908	0.2210	0.2089	0.1947	0.1945
λ_3	0.1577	0.2529	0.2926	0.3013	0.2134	0.2291	0.2476	0.2447	0.2182	0.2196	0.2136	0.2133
λ_4	0.1001	0.0574	0.0794	0.0797	0.1318	0.1311	0.1557	0.1608	0.1454	0.1662	0.1876	0.1882
β_1	0.8399	0.9568	0.9929	0.9938	0.9024	0.9896	1.0047	1.0058	0.9532	1.0048	1.0104	1.0107
β_2	0.4096	0.4903	0.5073	0.5059	0.4574	0.5143	0.5185	0.5192	0.4735	0.5057	0.5068	0.5071
$p = 6$												
λ_1	0.1635	0.1682	0.1726	0.1763	0.1718	0.1623	0.1540	0.1505	0.1658	0.1491	0.1495	0.1486
λ_2	0.2858	0.1928	0.1270	0.1146	0.2232	0.1643	0.1384	0.1365	0.1962	0.1556	0.1503	0.1515
λ_3	0.0977	0.1651	0.2195	0.2208	0.1256	0.1444	0.1760	0.1788	0.1408	0.1515	0.1515	0.1527
λ_4	0.2516	0.2367	0.1627	0.1644	0.1704	0.2167	0.1525	0.1457	0.1467	0.1815	0.1520	0.1497
λ_5	−0.0140	0.0937	0.1403	0.1262	0.1349	0.1708	0.1933	0.1956	0.1572	0.1319	0.1544	0.1531
λ_6	0.1381	0.0472	0.0768	0.0964	0.0880	0.0410	0.0837	0.0908	0.1009	0.1290	0.1407	0.1428
β_1	0.7807	0.9349	0.9856	0.9888	0.8590	0.9813	1.0027	1.0049	0.9288	1.0082	1.0128	1.0130
β_2	0.3830	0.4900	0.5131	0.5135	0.4342	0.5150	0.5191	0.5198	0.4593	0.5106	0.5096	0.5096

estimate

$$y = \sum_{i=1}^{p} \lambda_i W_i^b y + X\beta + U, \tag{5.1}$$

where W_i^b is the (row-normalized) weight matrix having off-diagonal (l, k)th element equal to 1 if $i - 1 < d_{lk} \leq i$, $i = 1 \ldots, p$, and if $d_{lk} = 0$ for $i = 1$. Thus the matrices are based on each one of p sequential 1-mile rings from the origin DBF. y is the vector of natural logs of the amount of VC funding (million \$) received by each of $n = 816$ DBFs.

Table 14
Monte Carlo mean squared error of parameter estimates. IV and iterated Newton-step estimates with heteroskedastic zero mean Gaussian errors. Bounded h_n. The parameter values used in the DGPs are as follows. $\beta_1 = 1$, $\beta_2 = 0.5$ and $p = 2 : \lambda_1 = 0.4, \lambda_2 = 0.5$; $p = 4 : \lambda_1 = 0.3, \lambda_i = 0.2, i = 2, 3, 4$; $p = 6 : \lambda_i = 0.15, i = 1, \ldots, 6$.

	$n = 200$				$n = 400$				$n = 800$			
$p = 2$	$\hat{\theta}_n$	$\hat{\hat{\theta}}_n^1$	$\hat{\hat{\theta}}_n^3$	$\hat{\hat{\theta}}_n^6$	$\hat{\theta}_n$	$\hat{\hat{\theta}}_n^1$	$\hat{\hat{\theta}}_n^3$	$\hat{\hat{\theta}}_n^6$	$\hat{\theta}_n$	$\hat{\hat{\theta}}_n^1$	$\hat{\hat{\theta}}_n^3$	$\hat{\hat{\theta}}_n^6$
λ_1	0.2641	0.0959	0.0713	0.0719	0.1396	0.0290	0.0205	0.0206	0.0592	0.0050	0.0028	0.0028
λ_2	0.2612	0.0920	0.0656	0.0657	0.1368	0.0273	0.0188	0.0188	0.0583	0.0049	0.0029	0.0029
β_1	0.0854	0.0651	0.0665	0.0685	0.0401	0.0280	0.0295	0.0297	0.0193	0.0130	0.0126	0.0126
β_2	0.0664	0.0527	0.0592	0.0614	0.0338	0.0256	0.0263	0.0265	0.0166	0.0122	0.0119	0.0119
$p = 4$												
λ_1	0.4143	0.1888	0.1449	0.1442	0.2023	0.0471	0.0327	0.0328	0.0898	0.0144	0.0121	0.0121
λ_2	1.7785	0.9054	0.6786	0.6871	0.8824	0.2323	0.1385	0.1375	0.4095	0.0705	0.0525	0.0525
λ_3	3.3452	1.9580	1.4817	1.4957	1.6244	0.4871	0.2570	0.2458	0.8319	0.1259	0.0821	0.0822
λ_4	1.5165	0.9575	0.7425	0.7336	0.7135	0.2309	0.1303	0.1269	0.3389	0.0485	0.0283	0.0283
β_1	0.1172	0.0875	0.0831	0.0884	0.0527	0.0317	0.0295	0.0295	0.0228	0.0149	0.0139	0.0139
β_2	0.0742	0.0640	0.0712	0.0789	0.0345	0.0275	0.0263	0.0263	0.0165	0.0131	0.0128	0.0128
$p = 6$												
λ_1	0.4441	0.1949	0.1286	0.1234	0.2278	0.0603	0.0246	0.0229	0.0891	0.0084	0.0043	0.0042
λ_2	1.8396	0.9700	0.6390	0.6173	0.9514	0.2731	0.1095	0.0980	0.4013	0.0402	0.0218	0.0216
λ_3	3.6843	2.1334	1.5197	1.4967	2.0022	0.7566	0.3695	0.3415	0.8983	0.1051	0.0538	0.0529
λ_4	6.6772	4.0414	2.7525	2.7348	3.4930	1.4344	0.8111	0.7667	1.5677	0.2080	0.1123	0.1118
λ_5	9.8262	6.9996	5.2694	5.3615	5.3131	2.2573	1.1760	1.0838	2.3504	0.3432	0.1277	0.1246
λ_6	4.4893	3.5345	2.8173	2.8448	1.9724	0.8713	0.4730	0.4442	0.8862	0.1279	0.0392	0.0379
β_1	0.1451	0.1037	0.0949	0.0956	0.0657	0.0373	0.0303	0.0300	0.0265	0.0150	0.0137	0.0137
β_2	0.0773	0.0676	0.0778	0.0786	0.0379	0.0295	0.0280	0.0282	0.0174	0.0133	0.0130	0.0130

Table 15
Monte Carlo relative root MSE of IV estimates to iterated Newton-step estimates with heteroskedastic zero mean Gaussian errors. Bounded h_n. The parameter values used in the DGPs are as follows. $\beta_1 = 1$, $\beta_2 = 0.5$ and $p = 2 : \lambda_1 = 0.4, \lambda_2 = 0.5$; $p = 4 : \lambda_1 = 0.3, \lambda_i = 0.2, i = 2, 3, 4$; $p = 6 : \lambda_i = 0.15, i = 1, \ldots, 6$.

	$n = 200$			$n = 400$			$n = 800$		
$p = 2$	$\frac{\mathrm{RMSE}(\hat{\theta}_n)}{\mathrm{RMSE}(\hat{\hat{\theta}}_n^1)}$	$\frac{\mathrm{RMSE}(\hat{\theta}_n)}{\mathrm{RMSE}(\hat{\hat{\theta}}_n^3)}$	$\frac{\mathrm{RMSE}(\hat{\theta}_n)}{\mathrm{RMSE}(\hat{\hat{\theta}}_n^6)}$	$\frac{\mathrm{RMSE}(\hat{\theta}_n)}{\mathrm{RMSE}(\hat{\hat{\theta}}_n^1)}$	$\frac{\mathrm{RMSE}(\hat{\theta}_n)}{\mathrm{RMSE}(\hat{\hat{\theta}}_n^3)}$	$\frac{\mathrm{RMSE}(\hat{\theta}_n)}{\mathrm{RMSE}(\hat{\hat{\theta}}_n^6)}$	$\frac{\mathrm{RMSE}(\hat{\theta}_n)}{\mathrm{RMSE}(\hat{\hat{\theta}}_n^1)}$	$\frac{\mathrm{RMSE}(\hat{\theta}_n)}{\mathrm{RMSE}(\hat{\hat{\theta}}_n^3)}$	$\frac{\mathrm{RMSE}(\hat{\theta}_n)}{\mathrm{RMSE}(\hat{\hat{\theta}}_n^6)}$
λ_1	1.6599	1.9196	1.9171	2.1926	2.6162	2.6054	3.4381	4.5874	4.5872
λ_2	1.6846	1.9786	1.9944	2.2392	2.7016	2.7003	3.4509	4.4707	4.4704
β_1	1.1453	1.1525	1.1182	1.1963	1.1748	1.1621	1.2174	1.2386	1.2386
β_2	1.1222	1.0814	1.0416	1.1488	1.1403	1.1298	1.1666	1.1803	1.1801
$p = 4$									
λ_1	1.4814	1.6847	1.7005	2.0728	2.4833	2.4865	2.4979	2.7215	2.7236
λ_2	1.4015	1.6097	1.6175	1.9490	2.4974	2.5337	2.4102	2.7911	2.7917
λ_3	1.3071	1.4947	1.5028	1.8262	2.4549	2.5662	2.5708	3.1828	3.1829
λ_4	1.2585	1.4194	1.4414	1.7577	2.2906	2.3680	2.6443	3.4592	3.4601
β_1	1.1572	1.1994	1.1562	1.2886	1.3360	1.3374	1.2376	1.2778	1.2782
β_2	1.0766	1.0422	0.9760	1.1200	1.1455	1.1450	1.1231	1.1364	1.1362
$p = 6$									
λ_1	1.5096	1.8362	1.8823	1.9431	2.8899	3.1347	3.2481	4.4402	4.5830
λ_2	1.3771	1.6572	1.7221	1.8666	2.7634	3.1174	3.1612	4.2230	4.3131
λ_3	1.3141	1.5379	1.5681	1.6267	2.1914	2.4170	2.9232	4.0449	4.1204
λ_4	1.2854	1.5390	1.5668	1.5605	1.9983	2.1282	2.7456	3.6757	3.7439
λ_5	1.1848	1.3467	1.3641	1.5342	2.0565	2.2023	2.6171	4.1638	4.3423
λ_6	1.1270	1.2461	1.2625	1.5045	1.9893	2.1013	2.6321	4.6272	4.8341
β_1	1.1829	1.2454	1.2314	1.3269	1.4564	1.4791	1.3293	1.3890	1.3905
β_2	1.0693	1.0151	0.9924	1.1334	1.1671	1.1607	1.1424	1.1538	1.1536

We first focus on estimates of the main parameters of interest λ_i in (5.1). We estimate (5.1) with $p = 2, 4, 6$ using initial IV and the Newton-step estimates that we have justified theoretically. We only report the Newton-step for a single iteration as convergence is achieved. Like Gupta and Robinson (2015), we find that only λ_1 and λ_2 are statistically significant at the 1% level, and the magnitude of our parameter estimates is also close to their findings, with our results reported in Table 16. The table reports t statistics in parentheses. In square brackets we report for each parameter estimate the ratio of IV standard error to Newton-step standard error, and find that this difference can be as great as 12.53%. Thus the iteration scheme we propose can lead to more accurate inference in practice as the estimates are more precise.

Table 16
IV and single Newton-step estimates of λ_i in model (5.1). t-statistics are in parentheses and the ratio of IV standard errors to Newton-step standard errors are in square brackets.

$p = 6$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_4$	$\hat{\lambda}_5$	$\hat{\lambda}_6$
	0.1228 (3.0525)	0.0853 (2.1215)	0.0490 (1.1225)	0.0510 (1.0627)	0.0249 (0.6064)	0.0136 (0.3450)
	$\hat{\hat{\lambda}}_1$	$\hat{\hat{\lambda}}_2$	$\hat{\hat{\lambda}}_3$	$\hat{\hat{\lambda}}_4$	$\hat{\hat{\lambda}}_5$	$\hat{\hat{\lambda}}_6$
	0.0821 (3.3776) [1.1065]	0.0848 (2.3139) [1.0907]	0.0080 (1.2404) [1.1050]	0.0483 (1.1628) [1.0942]	0.0223 (0.6559) [1.0817]	0.0171 (0.3568) [1.0341]
$p = 4$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_4$		
	0.0982 (2.4240)	0.0844 (2.0840)	0.0581 (1.3298)	0.0516 (1.1102)		
	$\hat{\hat{\lambda}}_1$	$\hat{\hat{\lambda}}_2$	$\hat{\hat{\lambda}}_3$	$\hat{\hat{\lambda}}_4$		
	0.0844 (2.7276) [1.1253]	0.0875 (2.3051) [1.1061]	0.0049 (1.4859) [1.1174]	0.0555 (1.2116) [1.0913]		
$p = 2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$				
	0.0984 (2.4879)	0.0894 (2.1819)				
	$\hat{\hat{\lambda}}_1$	$\hat{\hat{\lambda}}_2$				
	0.0891 (2.7968) [1.1241]	0.0882 (2.4516) [1.1236]				

Table 17
IV and single Newton-step estimate properties of β_i in model (5.1): standard error ratios and absolute values of t-statistics for $H_0 : \beta_i = 0$, $i = 1, \ldots, 21$. β_1 is the intercept.

	$p = 2$			$p = 4$			$p = 6$		
	$\frac{se(\hat{\beta})}{se(\hat{\hat{\beta}})}$	$\left\lvert\frac{\hat{\beta}}{se(\hat{\beta})}\right\rvert$	$\left\lvert\frac{\hat{\hat{\beta}}}{se(\hat{\hat{\beta}})}\right\rvert$	$\frac{se(\hat{\beta})}{se(\hat{\hat{\beta}})}$	$\left\lvert\frac{\hat{\beta}}{se(\hat{\beta})}\right\rvert$	$\left\lvert\frac{\hat{\hat{\beta}}}{se(\hat{\hat{\beta}})}\right\rvert$	$\frac{se(\hat{\beta})}{se(\hat{\hat{\beta}})}$	$\left\lvert\frac{\hat{\beta}}{se(\hat{\beta})}\right\rvert$	$\left\lvert\frac{\hat{\hat{\beta}}}{se(\hat{\hat{\beta}})}\right\rvert$
β_1	1.0003	2.0982	2.0989	1.0021	2.1962	2.2008	1.0011	2.2533	2.2558
β_2	0.9998	5.4574	5.4566	1.0013	5.4328	5.4396	1.0016	5.3744	5.3828
β_3	0.9999	4.7903	4.7896	1.0013	4.8049	4.8112	1.0017	4.7425	4.7505
β_4	1.0001	1.9100	1.9102	1.0017	1.9084	1.9116	1.0022	1.9004	1.9047
β_5	0.9998	0.2659	0.2658	1.0025	0.5221	0.5233	1.0038	0.5368	0.5388
β_6	1.0005	1.5215	1.5222	1.0029	1.4942	1.4986	1.0028	1.4065	1.4105
β_7	1.0000	0.0523	0.0523	1.0022	0.2818	0.2824	1.0024	0.3466	0.3475
β_8	1.0014	0.5536	0.5544	1.0036	0.2482	0.2491	1.0037	0.2037	0.2044
β_9	0.9998	1.8588	1.8585	1.0015	1.9536	1.9565	1.0020	1.9164	1.9202
β_{10}	0.9999	0.6021	0.6020	1.0016	0.6283	0.6293	1.0017	0.6164	0.6174
β_{11}	1.0028	6.6865	6.7052	1.0037	6.5660	6.5903	1.0032	6.3962	6.4165
β_{12}	1.0005	10.7482	10.7532	1.0028	10.5510	10.5810	1.0025	10.5389	10.5654
β_{13}	1.0005	10.0880	10.0930	1.0019	9.9490	9.9674	1.0025	9.7908	9.8156
β_{14}	1.0099	0.1857	0.1876	1.0193	0.6536	0.6662	1.0169	0.4469	0.4544
β_{15}	1.0002	2.4057	2.4062	1.0017	2.4102	2.4143	1.0018	2.4498	2.4543
β_{16}	1.0039	1.6017	1.6080	1.0158	2.3354	2.3722	1.0119	2.4618	2.4912
β_{17}	1.0024	0.5792	0.5805	1.0043	0.9156	0.9196	1.0030	0.9163	0.9191
β_{18}	1.0007	1.9997	2.0011	1.0043	1.9866	1.9952	1.0036	2.0354	2.0428
β_{19}	1.0006	2.6259	2.6273	1.0020	2.4606	2.4655	1.0011	2.3586	2.3613
β_{20}	1.0006	1.9874	1.9886	1.0043	2.3170	2.3269	1.0042	2.1617	2.1708
β_{21}	1.0004	1.3650	1.3655	1.0037	1.4905	1.4959	1.0032	1.4793	1.4840

As far as the β_i are concerned, our simulations generally show that the efficiency gains are smaller for these as compared to the λ_i. Table 17 reports standard error ratios and absolute t-statistics for exclusion tests and confirms this. Indeed, all standard error ratios are very close to unity and the t-statistics are practically identical. We note that our proposed iteration does not make the estimation precision of the β_i worse and improves the estimation precision of the λ_i, leading to an overall improvement in estimation quality.

We give a very brief description of the explanatory variables in X_n and refer the reader to KKM for details. The covariates include the number of proximate VCFs and DBFs to capture the effects of being in areas of high VCF or DBF concentration. Firm-specific characteristics include the distance from each DBF to its funding VCFs, the average age of each funding VCF, exposure of VCFs through syndication and an indicator for foreign VCF investment. Variables controlling for DBF-specific factors include firm age, dummies for receiving a grant and being in an R&D tax credit state, a cost of business index for the DBF's home state, distance to the closest university and the number of non-biotech establishments in the DBF's zip code. Two further variables recognize that additional factors can affect the cost of doing business in ways that influence the VC funding levels of a given DBF.

Appendix A. Proofs of theorems

Write $a_n = p_n + k_n$, $b_n = r_n + k_n$, $c_n = p_n k_n^2 + k_n$ and $\tau_n = n^{\frac{1}{2}}/a_n^{\frac{1}{2}}$.

Proof of Theorem 3.1.

(i) By the mean value theorem (2.7) implies that

$$\hat{\hat{\theta}}_n - \theta_{0n} = \left(I_{a_n} - \hat{H}_n^{-1}\overline{H}_n\right)\left(\hat{\theta}_n - \theta_{0n}\right) - \hat{H}_n^{-1}\xi_n$$
$$= \hat{\theta}_n - \theta_{0n} - \hat{H}_n^{-1}\overline{H}_n\left(\hat{\theta}_n - \theta_{0n}\right) - \hat{H}_n^{-1}\xi_n \tag{A.1}$$

where $\overline{H}_n = \partial^2 \mathcal{Q}_n(\overline{\theta}_n, \hat{\sigma}_n^2)/\partial\theta\partial\theta'$ and $\left\|\overline{\theta}_n - \theta_{0n}\right\| \leq \left\|\hat{\theta}_n - \theta_{0n}\right\|$, with each row of the Hessian matrix evaluated at possibly different $\overline{\theta}_n$. The latter point is a technical comment that we take as given in the remainder of the paper whenever a mean-value theorem is applied to vector of values. For any $s\times 1$ vector α, we can use (A.1) to write

$$\tau_n\alpha'\Psi_n\left(\hat{\hat{\theta}}_n - \theta_{0n}\right) = \tau_n\alpha'\Psi_n\hat{H}_n^{-1}\left(\hat{H}_n - \overline{H}_n\right)\left(\hat{\theta}_n - \theta_{0n}\right)$$
$$- \tau_n\alpha'\Psi_n\hat{H}_n^{-1}\xi_n, \tag{A.2}$$

recalling that $\tau_n = n^{\frac{1}{2}}/a_n^{\frac{1}{2}}$. The first term on RHS above has modulus bounded by $\tau_n\left\|\alpha\right\|\left\|\Psi_n\right\|\left\|\hat{H}_n^{-1}\right\|\left\|\hat{H}_n - \overline{H}_n\right\|\left\|\hat{\theta}_n - \theta_{0n}\right\|$, where the second factor in norms is $\mathcal{O}\left(a_n^{\frac{1}{2}}\right)$, the third is bounded in probability for sufficiently large n by Lemma B.5, by Lemma B.3 the fourth is $\mathcal{O}_p\left(\max\left\{p_n^{\frac{3}{2}}b_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n, b_n^{\frac{1}{2}}c_n^{\frac{1}{2}}/n, p_n^{\frac{1}{2}}b_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n^{\frac{1}{2}}, b_n/n\right\}\right)$ and the fifth is $\mathcal{O}_p\left(b_n^{\frac{1}{2}}/n^{\frac{1}{2}}\right)$ by Theorem 3.1 of Gupta and Robinson (2015). We conclude that the first term on the RHS of (A.2) is $\mathcal{O}_p\left(\max\left\{p_n^{\frac{3}{2}}b/n^{\frac{1}{2}}h_n, b_nc_n^{\frac{1}{2}}/n, p_n^{\frac{1}{2}}b/n^{\frac{1}{2}}h_n^{\frac{1}{2}}, b_n^{\frac{3}{2}}/n\right\}\right)$, which is negligible by (3.1) and (3.2) because $\frac{p_n^3b_n^2}{nh_n^2} \leq C\left(\frac{p_n^3r_n^2+p_n^3k_n^2}{nh_n^2}\right) = \mathcal{O}\left(\frac{p_n^3}{h_n^2}\frac{r_n^2}{n} + \frac{p_n^3k_n^2}{nh_n^2}\right)$, $\frac{b_nc_n^{\frac{1}{2}}}{n} \leq C\left(\frac{r_np_n^{\frac{1}{2}}k_n+p_n^{\frac{1}{2}}k_n^2}{n^2}\right) = \mathcal{O}\left(\frac{p_n^{\frac{1}{2}}k_n}{n^{\frac{1}{2}}}\frac{r_n}{n^{\frac{1}{2}}} + \frac{p_n^{\frac{1}{2}}k_n^2}{n^2}\right)$, $\frac{p_nb_n^2}{nh_n} \leq C\left(\frac{p_nr_n^2+p_nk_n^2}{nh_n}\right) = \mathcal{O}\left(\frac{r_n^2}{n}\frac{p_n}{h_n} + \frac{p_nk_n^2}{n}\frac{1}{h_n}\right)$, $\frac{b_n^3}{n^2} \leq C\left(\frac{r_n^3+k_n^3}{n^2}\right)$. For the negligibility of the last term note that $\frac{r_n^{\frac{3}{2}}}{n} = \frac{r_n^2}{n}r_n^{-\frac{1}{2}}$.
Thus we only need to find the asymptotic distribution of $-\tau_n\alpha'\Psi_n\hat{H}_n^{-1}\xi_n$. We can write

$$-\tau_n\alpha'\Psi_n\hat{H}_n^{-1}\xi_n = \frac{2}{\sigma_0^2}\tau_n\alpha'\Psi_n\hat{H}_n^{-1}t_n - \tau_n\alpha'\Psi_n\hat{H}_n^{-1}\phi_n. \tag{A.3}$$

We have $\mathbb{E}\left\|\phi_n\right\|^2 \leq \sum_{i=1}^{p_n}\mathbb{E}\left(n^{-1}trC_{in} - n^{-1}\sigma_0^{-2}u'C_{in}u\right)^2 = \sum_{i=1}^{p_n}\mathrm{var}\left(n^{-1}u'C_{in}u\right) = \mathcal{O}\left(p_n/nh_n\right)$, (see (A.20) in the proof of Theorem 3.3 and Lemma B.2 in Gupta and Robinson, 2018), so that

$$\left\|\phi_n\right\| = \mathcal{O}_p\left(\frac{p_n^{\frac{1}{2}}}{n^{\frac{1}{2}}h_n^{\frac{1}{2}}}\right). \tag{A.4}$$

Therefore the second term on the right of (A.3) has modulus bounded by τ_n times

$$\left\|\alpha\right\|\left\|\Psi_n\right\|\left\|\hat{H}_n^{-1}\right\|\left\|\phi_n\right\|, \tag{A.5}$$

where the second factor is $\mathcal{O}\left(a_n^{\frac{1}{2}}\right)$, the third is bounded in probability for sufficiently large n by Lemma B.5 and the last is $\mathcal{O}_p\left(p_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n^{\frac{1}{2}}\right)$. Thus (A.5) is $\mathcal{O}_p\left(p_n^{\frac{1}{2}}a_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n^{\frac{1}{2}}\right)$ and the second term on the right of (A.3) is $\mathcal{O}_p\left(p_n^{\frac{1}{2}}/h_n^{\frac{1}{2}}\right)$

which is negligible by (3.1). Then the asymptotic distribution required is that of

$$\frac{2}{\sigma_0^2}\tau_n\alpha'\Psi_n\hat{H}_n^{-1}t_n = \sum_{i=1}^{3}\Upsilon_{in} + \tau_n\alpha'\Psi_n L_n^{-1}t_n, \tag{A.6}$$

$\Upsilon_{1n} = \frac{2}{\sigma_0^2}\tau_n\alpha'\Psi_n\hat{H}_n^{-1}\left(\hat{H}_n - H_n\right)H_n^{-1}t_n$, $\Upsilon_{2n} = \frac{2}{\sigma_0^2}\tau_n\alpha'\Psi_n\Xi_n^{-1}\left(H_n - \Xi_n\right)H_n^{-1}t_n$, $\Upsilon_{3n} = \tau_n\alpha'\Psi_n L_n^{-1}\left(\frac{\sigma_0^2}{2}\Xi_n - L_n\right)\left(\frac{\sigma_0^2}{2}\Xi_n\right)^{-1}t_n$. We will demonstrate that $|\Upsilon_{in}| = o_p(1)$, $i = 1, 2, 3$. First we observe that $|\Upsilon_{1n}| \leq \frac{2}{\sigma_0^2}\tau_n\|\alpha\|\|\Psi_n\|\left\|\hat{H}_n^{-1}\right\|\left\|\hat{H}_n - H_n\right\|\left\|H_n^{-1}\right\|\|t_n\|$, where the second factor in norms is $\mathcal{O}\left(a_n^{\frac{1}{2}}\right)$, the third and fifth are bounded (in probability) for sufficiently large n by Lemma B.5, the fourth is $\mathcal{O}_p\left(\max\left\{p_n^{\frac{3}{2}}b_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n, b_n^{\frac{1}{2}}c_n^{\frac{1}{2}}/n, p_n^{\frac{1}{2}}b_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n^{\frac{1}{2}}, b_n/n\right\}\right)$ by Lemma B.3, and by (A.13) of Gupta and Robinson (2015) the last is $\mathcal{O}_p\left(c_n^{\frac{1}{2}}/n^{\frac{1}{2}}\right)$.

Then $|\Upsilon_{1n}| = \mathcal{O}_p\left(\max\left\{p_n^{\frac{3}{2}}b_n^{\frac{1}{2}}c_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n, b_n^{\frac{1}{2}}c_n/n, p_n^{\frac{1}{2}}b_n^{\frac{1}{2}}c_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n^{\frac{1}{2}}, b_nc_n^{\frac{1}{2}}/n\right\}\right)$, which is negligible by (3.1) and (3.2) because $\frac{p_n^3b_nc_n}{nh_n^2} \leq C\left(\frac{p_n^4r_nk_n^2+p_n^4k_n^3}{nh_n^2}\right) = \mathcal{O}\left(\frac{r_n}{n^{\frac{1}{2}}}\frac{p_n^{\frac{3}{2}}k_n^2}{n^{\frac{1}{2}}}\frac{p_n^{\frac{5}{2}}}{h_n^2} + \frac{p_n^3k_n^3}{n}\frac{p_n}{h_n^2}\right)$, $\frac{b_nc_n^2}{n^2} \leq C\left(\frac{r_np_n^2k_n^4+p_n^2k_n^5}{n^2}\right) = \mathcal{O}\left(\frac{r_n}{n^{\frac{1}{2}}}\frac{p_n^2k_n^4}{n^{\frac{3}{2}}} + \frac{p_n^2k_n^4}{n}\frac{k_n}{n}\right)$, $\frac{p_nb_nc_n}{nh_n} \leq C\left(\frac{r_np_n^2k_n^2+p_n^2k_n^3}{nh_n}\right) = \mathcal{O}\left(\frac{r_n}{n^{\frac{1}{2}}}\frac{p_n^{\frac{3}{2}}k_n^2}{n^{\frac{1}{2}}}\frac{p_n^{\frac{1}{2}}}{h_n} + \frac{p_n^2k_n^3}{nh_n}\right)$ and $b_nc_n^{\frac{1}{2}}/n = o(1)$ has been shown earlier.

Next $|\Upsilon_{2n}| \leq 2\sigma_0^{-2}\tau_n\|\alpha\|\|\Psi_n\|\left\|H_n^{-1}\right\|\|H_n - \Xi_n\|\left\|\Xi_n^{-1}\right\|\|t_n\|$, where the second factor in norms is $\mathcal{O}\left(a_n^{\frac{1}{2}}\right)$, the third and fifth are bounded (in probability) for sufficiently large n by Lemma B.5, the fourth is $\mathcal{O}_p\left(p_nk_n/n^{\frac{1}{2}}\right)$ by Lemma B.4 and the last is $\mathcal{O}_p\left(c_n^{\frac{1}{2}}/n^{\frac{1}{2}}\right)$ as above. Then $|\Upsilon_{2n}| = \mathcal{O}_p\left(p_nk_nc_n^{\frac{1}{2}}/n^{\frac{1}{2}}\right)$ which is negligible by (3.1) because $p_n^2k_n^2c_n/n \leq Cp_n^3k_n^4/n$. Similarly $|\Upsilon_{3n}| = \mathcal{O}_p\left(p_nc_n^{\frac{1}{2}}/h_n\right)$ by Lemma B.4, which is negligible by (3.1) because $p_n^2c_n/h_n^2 \leq Cp_n^3k_n^2/h_n^2$. Then we only need to find the asymptotic distribution of the last term in (A.6), but this is precisely the proof of Theorem 3.3 of Gupta and Robinson (2015). Replicating those arguments leads to the theorem.

(ii) In view of Lemmas B.4, B.6 and B.7, the theorem is proved exactly like Theorem 3.1(*i*), except for different orders of magnitudes of various expressions. In this case two of the orders will be different from the analogous ones considered in the proof of Theorem 3.1(*i*). Indeed, the analogue of the bound for the first term in (A.2) is

$$\begin{aligned}&\mathcal{O}_p\left(n^{\frac{1}{2}}\max\left\{\frac{p_n^{\frac{3}{2}}c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}h_n}, \frac{p_n^2}{h_n^2}, \frac{p_n^{\frac{7}{4}}c_n^{\frac{1}{4}}}{n^{\frac{1}{4}}h_n^{\frac{3}{2}}}, \frac{c_n}{n}\right\}\max\left\{\frac{c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}, \frac{p_n^{\frac{1}{2}}}{h_n}\right\}\right)\\ &= \mathcal{O}_p\left(\max\left\{\pi_{1n}, \pi_{2n}, \pi_{3n}, \pi_{4n}, \pi_{5n}, \pi_{6n}\right\}\right),\end{aligned}$$

where $\pi_{1n} = p_n^{\frac{3}{2}}c_n/n^{\frac{1}{2}}h_n$, $\pi_{2n} = p_n^2c_n^{\frac{1}{2}}/h_n^2$, $\pi_{3n} = p_n^{\frac{7}{4}}c_n^{\frac{3}{4}}/n^{\frac{1}{4}}h_n^{\frac{3}{2}}$, $\pi_{4n} = c_n^{\frac{3}{2}}/n$, $\pi_{5n} = n^{\frac{1}{2}}p_n^{\frac{5}{2}}/h_n^3$, $\pi_{6n} = n^{\frac{1}{4}}p_n^{\frac{9}{4}}c_n^{\frac{1}{4}}/h_n^{\frac{5}{2}}$. Now π_{5n} is assumed to tend to zero by (3.3), while the remaining π_{in} terms are also negligible by (3.3) because $\pi_{1n}^2 = \mathcal{O}\left(p_n^5k_n^4/nh_n^2\right)$, $\pi_{2n}^2 = \mathcal{O}\left(p_n^5k_n^2/h_n^4\right)$, $\pi_{3n}^4 = \mathcal{O}\left(\left(p_n^4k_n^3/n\right)\left(p_n^6k_n^3/h_n^6\right)\right)$, $\pi_{4n}^2 = \mathcal{O}\left(p_n^3k_n^6/n^2\right)$, $\pi_{6n}^2 = \mathcal{O}\left(n^{\frac{1}{2}}p_n^5k_n/h_n^5\right) = \mathcal{O}\left(\pi_{5n}\left(p_n^{\frac{5}{2}}k_n/h_n^2\right)\right)$.

The analogue for the bound on Υ_{1n} is of order

$$\begin{aligned}&\mathcal{O}\left(n^{\frac{1}{2}}\right)\mathcal{O}_p\left(\max\left\{p_n^{\frac{3}{2}}c_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n, p_n^2/h_n^2, p_n^{\frac{7}{4}}c_n^{\frac{1}{4}}/n^{\frac{1}{4}}h_n^{\frac{3}{2}}, c_n/n\right\}\right)\mathcal{O}_p\left(c_n^{\frac{1}{2}}/n^{\frac{1}{2}}\right)\\ &= \mathcal{O}_p\left(\max\left\{\pi_{1n}, \pi_{2n}, \pi_{3n}, \pi_{4n}\right\}\right),\end{aligned}$$

which was shown to be negligible under the assumed conditions. All other bounds remain unchanged and will be also be negligible under (3.3), as in the proof of Theorem 3.1(*i*). □

Proof of Theorem 3.2. Proceeding as in the proof of Theorem 3.1(i), we can write

$$\begin{aligned}\tau_n\alpha'\Psi_n\left(\hat{\hat{\theta}}_n - \theta_{0n}\right) &= \tau_n\alpha'\Psi_n\hat{H}_n^{-1}\left(\hat{H}_n - \overline{H}_n\right)\left(\hat{\theta}_n - \theta_{0n}\right)\\ &\quad - \tau_n\alpha'\Psi_n\left(\hat{H}_n^{-1} - \Xi_n^{-1}\right)\xi_n - \tau_n\alpha'\Psi_n\Xi_n^{-1}\xi_n. \end{aligned}\tag{A.7}$$

As in the proof of Theorem 3.1(i), the first term on the RHS above is negligible by (3.7). Lemma B.5 (for bounded h_n) indicates that the second term on the RHS of (A.7) is $\mathscr{O}_p\left(\tau_n \|\Psi_n\| (\|t_n\| + \|\phi_n\|)\left(\left\|\hat{H}_n - H_n\right\| + \|H_n - \Xi_n\|\right)\right)$ which is $\mathscr{O}_p\left(n^{\frac{1}{2}} \max\left\{c_n^{\frac{1}{2}}/n^{\frac{1}{2}}, p_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n^{\frac{1}{2}}\right\} \max\left\{p_n^{\frac{3}{2}}b_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n, b_n^{\frac{1}{2}}c_n^{\frac{1}{2}}/n, p_n^{\frac{1}{2}}b_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n^{\frac{1}{2}}, b_n/n, p_n k_n/n^{\frac{1}{2}}\right\}\right)$, using (A.13) of Gupta and Robinson (2015), (A.4) and Lemmas B.3 and B.4(i). This is negligible by (3.7), in a similar way to the preceding proofs. Thus we need to establish the asymptotic distribution of $-\tau_n \alpha' \Psi_n \Xi_n^{-1} \xi_n$, which is established under the assumed conditions in Theorem 3.4 of Gupta and Robinson (2018). □

Appendix B. Lemmas

In the subsequent lemmas the assumptions of the theorems that these are used to prove are taken to hold.

Lemma B.1 (*Lemma LS.4 of Gupta and Robinson, 2018, Supplementary Material*). $\left\|B_n' A_n\right\| = \left\|A_n' B_n\right\| = \mathscr{O}_p\left(n^{\frac{1}{2}} p_n k_n\right)$.

Lemma B.2 (*Lemma LS.4 of Gupta and Robinson, 2018, Supplementary Material*). $\left\|X_n' B_n\right\| = \left\|B_n' X_n\right\| = \mathscr{O}_p\left(n^{\frac{1}{2}} p_n^{\frac{1}{2}} k_n^{\frac{1}{2}}\right)$.

Lemma B.3. $\left\|\hat{H}_n - \overline{H}_n\right\|$ *and* $\left\|\hat{H}_n - H_n\right\|$ *are* $\mathscr{O}_p\left(\max\left\{p_n^{\frac{3}{2}}b_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n, b_n^{\frac{1}{2}}c_n^{\frac{1}{2}}/n, p_n^{\frac{1}{2}}b_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n^{\frac{1}{2}}, b_n/n\right\}\right)$.

Proof. By the triangle inequality $\left\|\hat{H}_n - \overline{H}_n\right\| \leq \left\|\hat{H}_n - H_n\right\| + \left\|\overline{H}_n - H_n\right\|$, and again by the triangle inequality $\left\|\hat{H}_n - H_n\right\|$ is bounded by

$$\frac{2}{n}\left\|P_{ji,n}(\hat{\lambda}_n) - P_{ji,n}\right\| + \frac{2}{n}\left|\frac{1}{\hat{\sigma}_n^2} - \frac{1}{\sigma_0^2}\right|\left(\left\|R_n' R_n\right\| + 2\left\|X_n' R_n\right\| + \left\|X_n' X_n\right\|\right). \tag{B.8}$$

The first term in (B.8) is bounded by

$$\left\{\sum_{i,j=1}^{p_n}\left(\frac{2}{n}tr(G_{jn}(\hat{\lambda}_n)G_{in}(\hat{\lambda}_n)) - \frac{2}{n}tr(G_{jn}G_{in})\right)^2\right\}^{\frac{1}{2}} \tag{B.9}$$

By the MVT, we have $tr(G_{jn}(\hat{\lambda}_n)G_{in}(\hat{\lambda}_n)) = tr(G_{jn}G_{in}) + \overline{\overline{\zeta}}_{ij,n}'\left(\hat{\lambda}_n - \lambda_{0n}\right)$, where $\overline{\overline{\zeta}}_{ij,n}$ has elements $tr\left(G_{in}\left(\overline{\overline{\lambda}}_n\right)G_{sn}\left(\overline{\overline{\lambda}}_n\right)G_{jn}\left(\overline{\overline{\lambda}}_n\right) + G_{sn}\left(\overline{\overline{\lambda}}_n\right)G_{in}\left(\overline{\overline{\lambda}}_n\right)G_{jn}\left(\overline{\overline{\lambda}}_n\right)\right)$, $s = 1, \dots, p_n$, and $\left\|\overline{\overline{\lambda}}_n - \lambda_{0n}\right\| \leq \left\|\hat{\lambda}_n - \lambda_{0n}\right\|$. Thus, the summands in (B.9) are $4n^{-2}\left(\overline{\overline{\zeta}}_{ij,n}'\left(\hat{\lambda}_n - \lambda_{0n}\right)\right)^2 \leq 4n^{-2}\left\|\overline{\overline{\zeta}}_{ij,n}\right\|^2\left\|\hat{\lambda}_n - \lambda_{0n}\right\|^2$ by Cauchy–Schwarz inequality, where the first factor in norms on the RHS is $\mathscr{O}\left(p_n n^2/h_n^2\right)$ by Lemma LS.3 of Gupta and Robinson (2018), supplementary material. The second factor is bounded by $\left\|\hat{\theta}_n - \theta_{0n}\right\|^2 = \mathscr{O}_p\left(b_n/n\right)$ (see (A.6) of Gupta and Robinson, 2015), so we conclude that the summands in (B.9) are $\mathscr{O}_p\left(b_n p_n/nh_n^2\right)$ and therefore (B.9) is $\mathscr{O}_p\left(p_n^{\frac{3}{2}}b_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n\right)$ and it follows that so is the first term in (B.8). By (A.7) of Gupta and Robinson (2015),

$$\left|\frac{1}{\hat{\sigma}_n^2} - \frac{1}{\sigma_0^2}\right| = \mathscr{O}_p\left(\max\left\{\frac{b_n^{\frac{1}{2}}c_n^{\frac{1}{2}}}{n}, \frac{p_n^{\frac{1}{2}}b_n^{\frac{1}{2}}}{n^{\frac{1}{2}}h_n^{\frac{1}{2}}}, \frac{b_n}{n}\right\}\right), \tag{B.10}$$

which handles the second factor in the second term in (B.8). We shall now bound the terms inside the parentheses in the second term in (B.8). These are $\mathscr{O}_p(n)$ because $n^{-\frac{1}{2}}\|R_n\| = \mathscr{O}_p(1)$, $n^{-\frac{1}{2}}\|X_n\| = \mathscr{O}(1)$ and $n^{-1}\left\|X_n' R_n\right\| = \mathscr{O}_p(1)$, by Assumption 9. From (B.9), (B.10), we conclude that

$$\left\|\hat{H}_n - H_n\right\| = \mathscr{O}_p\left(\frac{p_n^{\frac{3}{2}}b_n^{\frac{1}{2}}}{n^{\frac{1}{2}}h_n}\right) + \mathscr{O}_p\left(\max\left\{\frac{b_n^{\frac{1}{2}}c_n^{\frac{1}{2}}}{n}, \frac{p_n^{\frac{1}{2}}b_n^{\frac{1}{2}}}{n^{\frac{1}{2}}h_n^{\frac{1}{2}}}, \frac{b_n}{n}\right\}\right)$$

Similarly, it may be shown that $\left\|\overline{H}_n - H_n\right\|$ has the same order, whence the lemma follows. □

Lemma B.4 (*Lemma B.2 of Gupta and Robinson, 2018*). *(i)* $\|H_n - \Xi_n\| = \mathscr{O}_p\left(p_n k_n/n^{\frac{1}{2}}\right)$ *if Assumption 11 holds, (ii)* $\left\|L_n - \left(\sigma_0^2/2\right)\Xi_n\right\| = \mathscr{O}\left(p_n/h_n\right)$.

Lemma B.5 (*Lemma B.3 of Gupta and Robinson, 2018*)**.** *The following inequalities are satisfied:* $\left\|\hat{H}_n^{-1}\right\| = \mathcal{O}_p\left(\left\|H_n^{-1}\right\|\right) = \mathcal{O}_p\left(\left\|\Xi_n^{-1}\right\|\right) = \mathcal{O}_p\left(\left(\underline{\eta}(L_n)\right)^{-1}\right) = \mathcal{O}_p(1)$. *If h_n does not diverge, the above result becomes* $\left\|\hat{H}_n^{-1}\right\| = \mathcal{O}_p\left(\left\|H_n^{-1}\right\|\right) = \mathcal{O}_p\left(\left(\underline{\eta}(\Xi_n)\right)^{-1}\right) = \mathcal{O}_p(1)$, *if also* $\underline{\lim}_{n\to\infty}\underline{\eta}(\Xi_n) > 0$.

Lemma B.6. $\left\|\tilde{H}_n - \overline{H}_n\right\|$ *and* $\left\|\tilde{H}_n - H_n\right\|$ *are* $\mathcal{O}_p\left(\max\left\{c_n/n, p_n^2/h_n^2, p_n^{\frac{3}{2}}c_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n, p_n^{\frac{7}{4}}c_n^{\frac{1}{4}}/n^{\frac{1}{4}}h_n^{\frac{3}{2}}\right\}\right)$.

Proof. The proof is similar to that of Lemma B.3 and we only elaborate on the differences from that proof. In this case we need to bound

$$\frac{2}{n}\left\|P_{ji,n}(\tilde{\lambda}_n) - P_{ji,n}\right\| + \frac{2}{n}\left|\frac{1}{\tilde{\sigma}_n^2} - \frac{1}{\sigma_0^2}\right|\left(\left\|R_n'R_n\right\| + 2\left\|X_n'R_n\right\| + \left\|X_n'X_n\right\|\right). \tag{B.11}$$

In the OLS case we have $\tilde{\sigma}_n^2 - \sigma_0^2 = \mathcal{O}_p\left(\max\left\{c_n/n, p_n/h_n^2, p_n^{\frac{1}{2}}c_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n\right\}\right)$ and $\left\|\tilde{\theta}_n - \theta_{0n}\right\| = \mathcal{O}_p\left(\max\left\{c_n^{\frac{1}{2}}/n^{\frac{1}{2}}, p_n^{\frac{1}{2}}/h_n\right\}\right)$, from (A.23) and (A.21) of Gupta and Robinson (2015), respectively. The first term in (B.11) is then $\mathcal{O}_p\left(\max\left\{p_n^{\frac{3}{2}}c_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n, p_n^2/h_n^2, p_n^{\frac{7}{4}}c_n^{\frac{1}{4}}/n^{\frac{1}{4}}h_n^{\frac{3}{2}}\right\}\right)$ while the second one is $\mathcal{O}_p\left(\max\left\{c_n/n, p_n/h_n^2, p_n^{\frac{1}{2}}c_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n\right\}\right)$. We may then argue in a similar way that the Hessian evaluated at the OLS estimate differs from its value at an intermediate point in norm by the same to conclude the proof. □

Lemma B.7 (*Lemma B.3 of Gupta and Robinson, 2018*)**.** $\left\|\tilde{H}_n^{-1}\right\| = \mathcal{O}_p\left(\left\|H_n^{-1}\right\|\right) = \mathcal{O}_p\left(\left\|\Xi_n^{-1}\right\|\right) = \mathcal{O}_p\left(\frac{\sigma_0^2}{2}\left(\underline{\eta}(L_n)\right)^{-1}\right) = \mathcal{O}_p(1)$.

References

Andrews, D.W.K., 1997. A stopping rule for the computation of generalized method of moments estimators. Econometrica 65, 913–931.
Anselin, L., 1988. Spatial Econometrics: Methods and Models, Vol. 4. Kluwer Academic Publishers, Boston.
Blommestein, H.J., 1983. Specification and estimation of spatial econometric models: *a* discussion of alternative strategies for spatial economic modelling. Reg. Sci. Urban Econ. 13, 251–270.
Case, A.C., 1991. Spatial patterns in household demand. Econometrica 59, 953–965.
Chudik, A., Pesaran, M.H., 2015. Large panel data models with cross-sectional dependence: A survey. In: Baltagi, B.H. (Ed.), The Oxford HandBook of Panel Data. Oxford University Press.
Cliff, A.D., Ord, J.K., 1973. Spatial Autocorrelation. Pion, London.
Conley, T.G., Dupor, B., 2003. A spatial analysis of sectoral complementarity. J. Polit. Economy 111, 311–352.
Das, D., Kelejian, H.H., Prucha, I.R., 2003. Finite sample properties of estimators of spatial autoregressive models with autoregressive disturbances. Pap. Reg. Sci. 82, 1–26.
Davis, P.J., 1979. Circulant Matrices. Wiley Interscience, New York.
De Luca, G., Magnus, J.R., Peracchi, F., 2018. Weighted-average least squares estimation of generalized linear models. J. Econometrics 204, 1–17.
Fisher, R.A., 1925. Theory of statistical estimation. Proc. Cambridge Phil. Soc. 22, 700–725.
Frazier, D.T., Renault, E., 2017. Efficient two-step estimation via targeting. J. Econometrics 201, 212–227.
Gupta, A., Robinson, P.M., 2015. Inference on higher-order spatial autoregressive models with increasingly many parameters. J. Econometrics 186, 19–31.
Gupta, A., Robinson, P.M., 2018. Pseudo maximum likelihood estimation of spatial autoregressive models with increasing dimension. J. Econometrics 202, 92–107.
Han, X., Hsieh, C.-s., Lee, L.F., 2017. Estimation and model selection of higher-order spatial autoregressive model: An efficient Bayesian approach. Reg. Sci. Urban Econ. 63, 97–120.
Han, X., Lee, L.F., Xu, X., 2020. Large sample properties of Bayesian estimation of spatial econometric models. Econometr. Theory Firstview 39.
Hartley, H.O., Booker, A., 1965. Nonlinear least squares estimation. Ann. Math. Stat. 36, 638–650.
Helmers, C., Patnam, M., 2014. Does the rotten child spoil his companion? Spatial peer effects among children in rural India. Quant. Econ. 5, 67–121.
Hsieh, C.-s., van Kippersluis, H., 2018. Smoking initiation: Peers and personality. Quant. Econ. 9, 825–863.
Hualde, J., Robinson, P.M., 2011. Gaussian pseudo-maximum likelihood estimation of fractional time series models. Ann. Statist. 39, 3152–3181.
Janssen, P., Jurečkova, J., Veraverbeke, N., 1985. Rate of convergence of one-and two-step M-estimators with applications to maximum likelihood and Pitman estimators. Ann. Statist. 13, 1222–1229.
Kasahara, H., Shimotsu, K., 2008. Pseudo-likelihood estimation and bootstrap inference for structural discrete Markov decision models. J. Econometrics 146, 92–106.
Kelejian, H.H., Prucha, I.R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. J. Real Estate Finance Econ. 17, 99–121.
Kelejian, H.H., Prucha, I.R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. Internat. Econom. Rev. 40, 509–533.
Kolympiris, C., Kalaitzandonakes, N., Miller, D., 2011. Spatial collocation and venture capital in the US biotechnology industry. Res. Policy 40, 1188–1199.

Kristensen, D., Linton, O., 2006. A closed-form estimator for the GARCH(1, 1) model. Econometr. Theory 22, 323–337.
Kristensen, D., Salanié, B., 2017. Higher-order properties of approximate estimators. J. Econometrics 198, 189–208.
Kuersteiner, G.M., Prucha, I.R., 2013. Limit theory for panel data models with cross sectional dependence and sequential exogeneity. J. Econometrics 174, 107–126.
Kuersteiner, G.M., Prucha, I.R., 2020. Dynamic spatial panel models: networks, common shocks, and sequential exogeneity. Econometrica 88, 2109–2146.
Kwok, H.H., 2019. Identification and estimation of linear social interaction models. J. Econometrics 210, 434–458.
Lahiri, S.N., 1996. On inconsistency of estimators based on spatial data under infill asymptotics. Sankhyā A 58, 403–417.
LeCam, L., On the asymptotic theory of estimation and testing hypotheses, in: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, 1956, pp. 129–156.
Lee, L.F., 2002. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. Econometr. Theory 18, 252–277.
Lee, L.F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72, 1899–1925.
Lee, L.F., 2007. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. J. Econometrics 137, 489–514.
Lee, L.F., Liu, X., 2010. Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. Econometr. Theory 26, 187–230.
Lee, L.F., Yu, J., 2013. Near unit root in the spatial autoregressive model. Spat. Econ. Anal. 8, 314–351.
Li, K., 2017. Fixed-effects dynamic spatial panel data models and impulse. J. Econometrics 198, 102–121.
Lin, X., Lee, L.F., 2010. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. J. Econometrics 157, 34–52.
Liu, S.F., Yang, Z., 2015. Modified QML estimation of spatial autoregressive models with unknown heteroskedasticity and nonnormality. Reg. Sci. Urban Econ. 52, 50–70.
Ord, K., 1975. Estimation methods for models of spatial interaction. J. Amer. Statist. Assoc. 70, 120–126.
Ortega, J., Rheinboldt, W., 1970. Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York.
Pace, R.K., Barry, R., 1997. Quick computation of spatial autoregressive estimators. Geogr. Anal. 29, 232–247.
Pinkse, J., Slade, M.E., Brett, C., 2002. Spatial price competition: A semiparametric approach. Econometrica 70, 1111–1153.
Robinson, P.M., 1988. The stochastic difference between econometric statistics. Econometrica 56, 531–548.
Robinson, P.M., 2005. Efficiency improvements in inference on stationary and nonstationary fractional time series. Ann. Statist. 33, 1800–1842.
Robinson, P.M., 2010. Efficient estimation of the semiparametric spatial autoregressive model. J. Econometrics 157, 6–17.
Robinson, P.M., 2011. Asymptotic theory for nonparametric regression with spatial data. J. Econometrics 165, 5–19.
Rothenberg, T.J., 1984. Approximate normality of generalized least squares estimates. Econometrica 52, 811–825.
Rothenberg, T.J., Leenders, C.T., 1964. Efficient estimation of simultaneous equation systems. Econometrica 32, 57–76.
Zhu, X., Huang, D., Wang, H., 2020. Multivariate spatial autoregressive model for large scale social networks. J. Econometrics 215, 591–606.