Cambridge Working Papers in Economics

Cambridge Working Papers in Economics: 2012

# ESTIMATION OF SPATIAL SAMPLE SELECTION MODELS: A PARTIAL MAXIMUM LIKELIHOOD APPROACH

Renata Rabovic

Pavel Cizek

27 February 2020

Estimation of a sample selection model with a spatial lag of a latent dependent variable or a spatial error in both the selection and outcome equations is considered in the presence of cross-sectional dependence. Since there is no estimation framework for the spatial lag model and the existing estimators for the spatial error model are either computationally demanding or have poor small sample properties, we suggest to estimate these models by the partial maximum likelihood estimator, following Wang et al. (2013)'s framework for a spatial error probit model. We show that the estimator is consistent and asymptotically normally distributed. To facilitate easy and precise estimation of the variance matrix without requiring the spatial stationarity of errors, we propose the parametric bootstrap method. Monte Carlo simulations demonstrate the advantages of the estimators.

# Estimation of Spatial Sample Selection Models: A Partial Maximum Likelihood Approach

Renata Rabovič* Pavel Čížek†

February 27, 2020

**Abstract**

Estimation of a sample selection model with a spatial lag of a latent dependent variable or a spatial error in both the selection and outcome equations is considered in the presence of cross-sectional dependence. Since there is no estimation framework for the spatial lag model and the existing estimators for the spatial error model are either computationally demanding or have poor small sample properties, we suggest to estimate these models by the partial maximum likelihood estimator, following Wang et al. (2013)'s framework for a spatial error probit model. We show that the estimator is consistent and asymptotically normally distributed. To facilitate easy and precise estimation of the variance matrix without requiring the spatial stationarity of errors, we propose the parametric bootstrap method. Monte Carlo simulations demonstrate the advantages of the estimators.

*JEL codes:* C13, C31, C34

*Keywords:* asymptotic distribution, maximum likelihood, near epoch dependence, sample selection model, spatial autoregressive model

*Corresponding author. Faculty of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD, United Kingdom. `Email:` rr574@cam.ac.uk, Phone: +44 122 333 5283.

†CentER, Dept. of Econometrics & Operations Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. `Email:` P.Cizek@tilburguniversity.edu, Phone: +31 13 466 8723.

# 1 Introduction

The assumption about independent observations is often not met even in the analysis of cross-sectional data. Since cross-sectional dependence can be captured by a certain spatial or economic ordering in many economic applications, spatial models have become an extensively used tool in applied econometrics. In this paper, we propose spatial extensions of sample selection models. We introduce spatial dependence into a sample selection model via a spatial lag of a latent dependent variable or a spatial error in both the selection and outcome equations. To our best knowledge, this is the first paper which analyzes a sample selection model with a spatial lag of a latent dependent variable, facilitating easy estimation in applications such as peer effects in education with non-randomly missing data (see Section 2 for more details). A spatial sample selection model with a spatial error, which can be used, for instance, in agricultural yield studies, has been analyzed before, but the proposed estimators are either computationally demanding or they do not have desirable small sample performance.

The computational difficulties in the spatial sample selection models stem from the (spatially) correlated errors: their joint density function cannot be expressed as a product of the density functions for each observation, and the full maximum likelihood estimator (MLE) becomes computationally very demanding as it involves high dimensional integration. It is possible to overcome this obstacle by using the heteroskedastic maximum likelihood estimator (HMLE), which takes into account only heteroskedasticity stemming from spatial correlation while neglecting the corresponding spatial autocorrelation to obtain consistent but inefficient estimates.[1] Based on this idea, Flores-Lagunes and Schnier (2012) in the context of a sample selection model with a spatial error in both the selection and outcome equations proposed to use the generalized method of moments (GMM) estimator.[2] The estimator however has poor small sample properties (see Section 4 in their paper and Section 5). Doğan and Taşpınar (2018) suggest to estimate the same model using the Markov chain Monte Carlo approach in the context of Bayesian estimation, whereas earlier McMillen (1995) suggested to use the Expectation Maximization algorithm. Both of these methods are however computationally demanding in larger samples due to the necessity to invert the spatial weight matrices numerous times, and moreover, a rigorous theory is not developed for either of them.

In the closely related context of binary choice models with spatial errors, Wang et al. (2013) therefore suggested an intermediate approach between the full MLE and HMLE that is based on the idea that all

[1] Poirier and Ruud (1988) developed the result under fairly general conditions for a probit model with serial correlation in a time series setting, whereas Robinson (1982) established the same result for a Tobit model.

[2] For empirical studies that use the estimator suggested by Flores-Lagunes and Schnier (2012), see Section 5 of Flores-Lagunes and Schnier (2012), Mukherjee and Singer (2010), and Ward et al. (2014).

observations are divided into clusters of two observations and the dependence within clusters is taken into account, whereas the dependence between clusters is not employed in the estimation. This approach avoids the computationally demanding full MLE (at the cost of losing efficiency), while it facilitates the estimation of the spatial error structure by taking the correlation within clusters into account. Wang et al. (2013) apply the partial maximum likelihood estimator (PMLE) to a spatial error probit model. In this paper, the PMLE approach is generalized to sample selection models with a spatial lag of a latent dependent variable or a spatial error and their special cases.

Since the special cases of the considered sample selection models include probit and Tobit models (see Section 2), this paper also extends Wang et al. (2013)'s results to the probit and Tobit models with a spatial lag of the latent dependent variable and to the Tobit model with a spatial error.[3] We analyze the asymptotic properties of the proposed PMLE using the near epoch dependent random fields framework introduced by Jenish and Prucha (2012). Note that the asymptotic results derived for a spatial error probit model in Wang et al. (2013) cannot be directly applied to our models because the structure of the spatial sample selection models is more complicated and requires additional treatment. For example, the uniform $L_p$-boundedness of the (bivariate) likelihood scores cannot be established by simply assuming that the support of exogeneous regressors is bounded since the observed dependent variables also enter the cumulative distribution function of the bivariate normal distribution. Moreover, Wang et al. (2013) base their analysis on $\alpha$-mixing processes and make assumptions about dependence based on the observed responses instead of deriving more primitive conditions.[4] They also impose a strong assumption on the expansion speed of the sampling region[5] and suggest to estimate the variance matrix of the proposed estimator based on the approach proposed by Conley (1999), who explicitly models the sampling process from a regular lattice and assumes that the data generating process is strongly spatially stationary. This assumption is in general not satisfied, for example, for the Cliff-Ord type models (see Kelejian and Prucha, 2007, for a further discussion). We relax these assumptions and suggest to estimate the asymptotic variance matrix using parametric bootstrap.

The paper is organized as follows. In Section 2, the sample selection models are defined, whereas the PMLE is introduced in Section 3. In Section 4, its consistency and asymptotic normality are established, and the estimator of the asymptotic variance matrix is discussed. In Section 5, we study the finite sample properties, while Section 6 concludes. Proofs are provided in Appendices.

It proves helpful to introduce the following notation. Let $A_n$, $n \in \mathbb{N}$, be some matrix indexed by $n$; we

[3]The Tobit model with a spatial lag of the observed dependent variable has been recently analyzed by Xu and Lee (2015a).
[4]See conditions (vii) and (i) of Theorems 1 and 2 by Wang et al. (2013), respectively.
[5]See condition (ii) of Theorem 2 by Wang et al. (2013).

denote the $ij$th element, the $i$th row, and the $j$th column of $A_n$ by $A_{ij,n}$, $A_{i\cdot,n}$, and $A_{\cdot j,n}$, respectively.

Similarly, if $v_n$ is a vector, then $v_{i,n}$ denotes the $i$th element of $v_n$. (The same notation applies for vectors and matrices that are not indexed by $n$.) If $A$ is a $2 \times 2$ matrix, then $A = [A_{11}\ A_{12};\ A_{21}\ A_{22}]$ is a matrix with the vectors $(A_{11}, A_{12})'$ and $(A_{21}, A_{22})'$ in the first and second row, respectively. Further, let $g = (i, j)'$ and $\dot{g} = (k, l)'$. Then $A_{g\cdot,n} = (A'_{i\cdot,n}, A'_{j\cdot,n})'$ with its $qm$th element, $q$th row, and $m$th column denoted by $A_{gqm,n}$, $A_{gq\cdot,n}$, and $A_{g\cdot m,n}$, respectively; $A_{g\dot{g},n} = [A_{ik,n}\ A_{il,n};\ A_{jk,n}\ A_{jl,n}]$ and $A_{g,n} = A_{gg,n}$ with $A_{g,n} = [A_{g11,n}\ A_{g12,n};\ A_{g21,n}\ A_{g22,n}]$. Similarly, $v_{g,n} = (v_{i,n}, v_{j,n})'$ with its $q$th element denoted by $v_{gq,n}$. Furthermore, for any random vector $Y$, let $\|Y\|_p = [E\|Y\|^p]^{1/p}$, $p \geq 1$, denote its $L_p$-norm, where $\|\cdot\|$ is the Euclidean norm. For an $n \times n$ matrix $A$, the Euclidean, row sum, and column sum matrix norms are defined as $\|A\| = \left(\sum_{i=1}^n \sum_{j=1}^n |A_{ij}|^2\right)^{1/2}$, $\|A\|_\infty = \max_{i=1,\ldots,n} \sum_{j=1}^n |A_{ij}|$, and $\|A\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^n |A_{ij}|$, respectively.[6] Note that these norms are sub-multiplicative: $\|AB\|_a \leq \|A\|_a \|B\|_a$, where $\|\cdot\|_a$ denotes one of the mentioned norms.

# 2 Model

To define the sample selection model, consider first the following latent selection ($s$) and outcome ($o$) equations with spatial lags of the latent dependent variable:

$$
\begin{aligned}
y^{*s}_{i,n} &= \lambda^s_0 W^s_{i\cdot,n} y^{*s}_n + X^s_{i\cdot,n}\beta^s_0 + u^s_{i,n} \\
y^{*o}_{i,n} &= \lambda^o_0 W^o_{i\cdot,n} y^{*o}_n + X^o_{i\cdot,n}\beta^o_0 + u^o_{i,n}
\end{aligned} \tag{1}
$$

for $i = 1, \ldots, 2n$,[7] where $2n$ represents the actual sample size and $n$ serves as the sample-size index, $y^{*s}_{i,n}$ and $y^{*o}_{i,n}$ are latent variables, $X^s_{i\cdot,n}$ and $X^o_{i\cdot,n}$ are $1 \times L^s$ and $1 \times L^o$ dimensional vectors of exogenous variables, and $u^s_{i,n}$ and $u^o_{i,n}$ are the error terms for the selection and outcome equations, respectively; the corresponding vectors and matrices of all observations are denoted by $y^{*s}_n = (y^{*s}_{i,n})^{2n}_{i=1}$, $W^s_n = (W^s_{i\cdot,n})^{2n}_{i=1}$, $X^s_n = (X^s_{i\cdot,n})^{2n}_{i=1}$, $u^s_n = (u^s_{i,n})^{2n}_{i=1}$ and analogously for the outcome equation. The spatial nonstochastic weight matrices $W^s_n$ and $W^o_n$ are assumed to be known, contain nonnegative elements, and have zero elements on the main diagonal. For example, the elements of $W^s_n$ and $W^o_n$ can be indirectly proportional to the strength of an economic relationship or distance between two observations, or they can be equal to 0 or 1, indicating unrelated or related (neighboring) observations (e.g., see LeSage and Pace, 2009). If the $ij$th element of the spatial weight matrix is nonzero, there is a direct dependence between the latent variables of observations $i$ and $j$. If the

[6] See Horn and Johnson (1985, pp. 291, 294-295) for more details.
[7] For notational convenience, we assume that the number of observations is even.

$ij$th element of the spatial weight matrix is zero, it does not mean that observations $i$ and $j$ are independent because there might exist an observation $k$ that has an effect on the latent variables of both observations $i$ and $j$.

The relation between the observed outcomes and latent variables in (1) is defined as $y^s_{i,n} = \mathbb{1}(y^{*s}_{i,n} > 0)$ and $y^o_{i,n} = y^s_{i,n} y^{*o}_{i,n}$, so that the selection equation determines which cases are observed, while the outcome equation determines the magnitude of the observed responses. (In general, $y^o_{i,n}$ is missing for observation $i$ rather than being zero if $y^s_{i,n} = 0$, but the definition is made for the simplicity of notation similarly to Chen and Zhou, 2010, among others and does not affect the likelihood function.) This version of a sample selection model is chosen because it is used in many empirical applications and includes other important models. For example, under normality of errors, modelling just $y^s_{i,n}$ leads to probit, and taking equations in (1) identical results in Tobit. Generalizations to other sample selection models might also be considered. For instance, a binary sample selection model with $y^s_{i,n} = \mathbb{1}(y^{*s}_{i,n} > 0)$ and $y^o_{i,n} = y^s_{i,n}\mathbb{1}(y^{*o}_{i,n} > 0)$ or a model with a Tobit selection equation defined as $y^s_{i,n} = \max\{0, y^{*s}_{i,n}\}$ and $y^o_{i,n} = \mathbb{1}(y^s_{i,n} > 0) y^{*o}_{i,n}$. Finally, the latent model (1) can be easily adapted to include spatial errors instead of spatial lags:

$$
\begin{aligned}
y^{*s}_{i,n} &= X^s_{i\cdot,n}\beta^s_0 + \varepsilon^s_{i,n}(\lambda^s_0) \\
y^{*o}_{i,n} &= X^o_{i\cdot,n}\beta^o_0 + \varepsilon^o_{i,n}(\lambda^o_0),
\end{aligned}
\tag{2}
$$

where $\varepsilon^s_{i,n}(\lambda^s_0) = \lambda^s_0 W^s_{i\cdot,n}\varepsilon^s_n(\lambda^s_0) + u^s_{i,n}$ and $\varepsilon^o_{i,n}(\lambda^o_0) = \lambda^o_0 W^o_{i\cdot,n}\varepsilon^o_n(\lambda^o_0) + u^o_{i,n}$ similarly to $y^{*s}_{i,n}$ and $y^{*o}_{i,n}$ in (1) with $\varepsilon^s_n(\lambda^s_0) = (\varepsilon^s_{i,n}(\lambda^s_0))^{2n}_{i=1}$ and $\varepsilon^o_n(\lambda^o_0) = (\varepsilon^o_{i,n}(\lambda^o_0))^{2n}_{i=1}$; the observed variables are defined in the same way as before. The results presented in the paper are also derived and hold for this sample selection model with spatial errors. Adjusting for different spatial error structures such as $\varepsilon^s_{i,n}(\lambda^s_0) = \lambda^s_0 W^s_{i\cdot,n} u^s_n + u^s_{i,n}$ with $u^s_n = (u^s_{i,n})^{2n}_{i=1}$ is straightforward.

An important feature of the latent model in (1) is that spatial lags of the latent instead of observed variables are included in both the selection and outcome equations. For the outcome equation, it is true that $y^o_{j,n} = y^{*o}_{j,n}$ if $y^s_{j,n} = 1$. Thus, the outcome equations with a lag of the latent variable and a lag of the observed variable differ primarily by the presence of $y^{*o}_{j,n}$ with $y^s_{j,n} = 0$ on the right hand side of the equation. Note also that $\varepsilon^o_{i,n}(\lambda^o_0)$ in (2) in general depends on $u^o_{j,n}$ with $y^s_{j,n} = 0$. For the selection equation, $y^s_{j,n}$ and $y^{*s}_{j,n}$ differ though, and the choice of the model depends on whether only individuals' decisions or also their motives are observable to others. By means of two empirical examples we will illustrate where models (1) and (2) are plausible specifications; see also a related discussion in Qu and Lee (2012) for the censored model.

**Example 1** (Peer effects in education with non-randomly missing data)**.** The peers effects in education literature investigates whether students are affected by behaviours of their peers. For example, students could increase their effort level if their peers study hard. In such situations, typically test scores are used to measure the unobserved effort level and, if test scores are observed for all students, the second equation in (1) is employed with $y_{i,n}^{*o}$ being replaced with $y_{i,n}^{o}$ (e.g., see Lin, 2010). Unfortunately, in some cases test scores are missing for some students. For example, Booij et al. (2016) found that, in the department of economics and business of the University of Amsterdam, only 46% of the students take all the first-year exams during the first year of their study. The sample selection problem arises if a student's decision to attend the exam and his score depend on the student's ability to succeed in the subject. Such a situation can be handled using the model in (1) with $y_{i,n}^{s}$ and $y_{i,n}^{*o}$ being a student's decision to take an exam and his (potential) score from that exam, respectively.

This model does not only allow for the effort level of students to be affected by their peers but also incorporates peer interactions in the decision to attend the exam. These interactions could be modelled either by a spatial lag of the latent dependent variable or a spatial lag of the observed dependent variable. Since students decide whether to attend the exam simultaneously, their decisions are likely to be affected by their beliefs about their peers' attendance rather than actual choices. Therefore, this situation might be better approximated by a spatial lag of the latent dependent variable as in equation (1) rather than the observed one. Likewise, we also include a lag of the latent rather than observed dependent variable in the outcome equation. Students who did not attend the exam could had influenced their peers by solving assignments together and attending the same tutorial classes. Therefore, the outcome equation should also capture these interactions.

Model (1) can also be used to study cases when a missing data problem arises due to non-responsiveness to a survey. Consider, for example, the National Longitudinal Study of Adolescent Health data, which has been extensively used to study peer effects in education (e.g., Calvò-Armengol et al., 2009; Lin, 2010). In this data, Hoshino (2019) found that information on GPA is missing for 11% of the respondents (after taking into account missing data on the exogeneous variables used in his study). It might be the case that unobserved abilities of a student affect both his decision to reveal his GPA and his GPA itself. Model (1) can thus be used to account for this kind of sample selection, where a lag of the latent variable is included in the outcome equation and is not included in the selection equation as the students filled-in the questionnaire independently.

**Example 2** (Agricultural yield)**.** Ward et al. (2014) apply the GMM estimator proposed by Flores-Lagunes and Schnier (2012) to estimate a cereal yield response function taking into account potential sample selection bias due to farmers' endogenous decision about whether to plant cereals. Flores-Lagunes and Schnier (2012) consider the latent model in (2) with spatially correlated errors and $W_n^s = W_n^o$ (see, for instance, equations (1) and (2) in their paper), but for the estimation, observations with $y_{i,n}^s = 0$ are omitted from the weight matrix in the outcome equation (see footnote 16 of their paper). The estimator is thus inconsistent with the model. Ward et al. (2014) overcame this issue by choosing $W_n^o \neq W_n^s$ in such a way that $W_{ij,n}^o$ can have positive values only if $y_{i,n}^s = y_{j,n}^s = 1$. In this case, the weight matrix depends on potentially endogenous farmers' decisions whether to plant cereals. This approach however requires further research as neither PMLE nor the GMM estimator proposed by Flores-Lagunes and Schnier (2012) are designed for the cases when the weight matrix in the outcome equation depends on the outcomes in the selection equation. On the one hand, if the correlation among unobservables in the outcome equation is driven by production technology or knowledge spillovers, then the farmers who decided not to plant a field do not likely have a lot of influence on those who decided to plant a field, and the weight matrix $W_n^o \neq W_n^s$ considered in Ward et al. (2014) should be chosen. On the other hand, if the correlation among unobservables is mainly driven by unobserved geographical and meteorological characteristics, then both planted and not planted fields are affected similarly if they are close to each other. Since the unobserved geographical and meteorological characteristics affect both the decision to plant a field and a cereal yield response function, a nonstochastic weight matrix which captures the closeness of fields can be used, and the specification in (2) with $W_n^o = W_n^s$ should be considered.

Model (1) can be written in a reduced form, provided that the respective inverses exist, as

$$\begin{aligned} y_n^{*s} &= S_n^s(\lambda_0^s) X_n^s \beta_0^s + \varepsilon_n^s(\lambda_0^s) \\ y_n^{*o} &= S_n^o(\lambda_0^o) X_n^o \beta_0^o + \varepsilon_n^o(\lambda_0^o), \end{aligned} \tag{3}$$

where the observed responses $y_{i,n}^s = \mathbb{1}(y_{i,n}^{*s} > 0)$ and $y_{i,n}^o = y_{i,n}^s y_{i,n}^{*o}$, and for $b \in \{s, o\}$, matrices $S_n^b(\lambda) = (I_{2n} - \lambda W_n^b)^{-1}$ and errors $\varepsilon_n^b(\lambda) = S_n^b(\lambda) u_n^b$. These definitions of $\varepsilon_n^s(\lambda_0^s)$ and $\varepsilon_n^o(\lambda_0^o)$ are equivalent to those in the spatial error model (2), and models (2) and (3) thus differ only by the presence of $S_n^s(\lambda_0^s)$ and $S_n^o(\lambda_0^o)$ in the latter model.[8]

The spatial weight matrices in the original and reduced form models have to satisfy the following assumption.

**Assumption 1.** *(i) The matrices $I_{2n} - \lambda^s W_n^s$ and $I_{2n} - \lambda^o W_n^o$ are nonsingular for all $(\lambda^s, \lambda^o)' \in \Lambda$, where $\Lambda$ is the space of the spatial parameters. (ii) The row and column sum matrix norms of matrices $W_n^s$, $W_n^o$,*

[8]For the simplicity of notation, we do not consider models with both spatial lags and spatial errors in both the selection and outcome equations. These models can however be analyzed in a similar way as the spatial lag model.

*$S^s_n(\lambda^s)$, and $S^o_n(\lambda^o)$ are bounded uniformly in $n \in \mathbb{N}$ and $(\lambda^s, \lambda^o)' \in \Lambda$.*

The first condition implies that there is a unique solution to $y^{*s}_n$ and $y^{*o}_n$ in (1) as well as to $\varepsilon^s_n(\lambda^s_0)$ and $\varepsilon^o_n(\lambda^o_0)$ in (2). Since there is no natural parameter space for spatial parameters, this condition is usually ensured by normalizing spatial weight matrices and bounding the parameter space. In applications, the weight matrices are typically normalized in such a way that the sum of each row is equal to 1 and the parameter space of $(\lambda^s, \lambda^o)$ is chosen to be $(-1, 1) \times (-1, 1)$. However, if there is no theoretical reason for the row normalization, this might lead to misspecification. Kelejian and Prucha (2010) instead suggest to normalize the weight matrices by their largest absolute eigenvalues. The second condition restricts dependence to a manageable degree. This is a classical assumption in the spatial econometrics literature (e.g., see Kelejian and Prucha, 1998, 1999, 2010).

# 3 Partial Maximum Likelihood Estimator

The (partial) maximum likelihood estimator requires a parametric specification of the distribution of the error terms. Although we could consider a general elliptically contoured distribution of $(u^s_{i,n}, u^o_{i,n})'$, we restrict our attention to the Gaussian case as it turns out to be not only the most frequently used one, but also the most complicated one (relative to heavier-tailed distributions) due to the necessity to study and bound the moments of the logarithm of the bivariate normal cumulate distribution function and their derivatives. Let $\theta = (\beta^{s'}, \beta^{o'}, \lambda^s, \lambda^o, \rho, \sigma^2)'$.

**Assumption 2.** *(i) The error terms $(u^s_{i,n}, u^o_{i,n})' \overset{iid}{\sim} \mathcal{N}(0, \Sigma(\theta_0))$, where $\Sigma(\theta_0) = [1\ \ \rho_0\sigma_0;\ \ \rho_0\sigma_0\ \ \sigma^2_0]$ is a positive definite matrix. (ii) $(X^s_n, X^o_n)$ and $(u^s_n, u^o_n)$ are independent. (iii) $X^s_{i\cdot,n}, X^o_{i\cdot,n}, W^s_{i\cdot,n}$, and $W^o_{i\cdot,n}$ are always observed.*

Assumption 2$(i)$ is strong but standard in the literature that analyzes parametric sample selection models (see Heckman, 1974, 1979). The variance of $u^s_{i,n}$ is normalized to 1 in order to ensure identification. The correlation coefficient $\rho_0$ controls the selection bias; if $\rho_0$ is zero, the outcome equation can be estimated independently of the selection equation. Even in that case, standard estimators for spatial linear models, for example, MLE (see Lee, 2004) or GMM (see Kelejian and Prucha, 1998), cannot be applied in order to estimate the outcome equation in (1) for a subsample of observations with $y^s_{i,n} = 1$ because $y^{*o}_{j,n}$ on the right hand side of the outcome equation is missing if $y^s_{j,n} = 0$. If data are missing at random, the methods developed by Wang and Lee (2013) for estimation of spatial autoregressive models are applicable to model

(1) and a version of the MLE estimator[9] can be applied to the outcome equation in (2) for a subsample of observations with $y^s_{i,n} = 1$. Neither method is applicable if $\rho_0 \neq 0$ though.

Further, in the standard sample selection model, it is assumed that $(X^s_{i\cdot,n}, X^o_{i\cdot,n})$ and $(u^s_{i,n}, u^o_{i,n})$ are independent. Due to spatially correlated errors $\varepsilon^s_{i,n}(\lambda^s)$ and $\varepsilon^o_{i,n}(\lambda^o)$, which are present in both the spatial lag and spatial error models in (3) and (2), respectively, we need to make an assumption that the exogenous variables and the error terms are mutually independent as in Assumption 2($ii$). Finally, Assumption 2($iii$) states that the exogenous variables and spatial weights have to be observed even for missing observations. The observability of the exogenous variables in the outcome equation for missing observations is not required neither in the standard parametric sample selection model nor in model (2) with spatial errors. If the model contains a spatial lag, the partial maximum likelihood estimator is however based on the reduced form in (3), which requires full observability. This assumption is not too strong since $X^s_{i\cdot,n}$ contains variables in $X^o_{i\cdot,n}$ in many empirical applications (e.g., Buchinsky, 1998; Vella, 1998; Sharma et al., 2013).

The assumption about observability of spatial weights is not very restrictive either, at least if the spatial weights are based on distances between observations, which are typically not difficult to obtain even for missing observations. For example, both exogenous variables and spatial weights are available for missing observations in Examples 1 and 2. In studies on peer effects in education, weight matrices are typically constructed by assigning value 1 to the $ij$th element of the weight matrix if students $i$ and $j$ are in the same classroom and value zero otherwise. As studies on peer effects in education usually use educational registers, both classroom compositions and background variables are also available for students who skipped a few exams. Likewise, in agricultural yield studies, spatial weight matrices are usually constructed based on the proximity of fields, hence also available for missing observations. The meteorological variables considered in Ward et al. (2014) are also available for fields that were not planted.

Due to the spatial dependence, the error terms $\varepsilon^s_{i,n}(\lambda^s)$ and $\varepsilon^o_{i,n}(\lambda^o)$ in latent models (3) and (2) are heteroskedastic and cross-correlated. Hence, the full MLE is computationally demanding in this setting. Based on the idea introduced by Wang et al. (2013), we therefore suggest to estimate the models by applying the partial maximum likelihood estimator. Specifically, we divide $2n$ observations into $n$ mutually disjoint pairs based on the idea that the internal correlation between two observations in a pair is more important than the external correlation with observations in the other pairs, at least if observations within a pair are "close" to each other. If only very weakly correlated observations are paired, the estimator will be similar to HMLE and there will be no gains from forming the pairs. The way how the observations are paired

[9] If $u^o_{i,n}|X^o_{i\cdot,n} \sim \mathcal{N}(0, \sigma^2_0)$, then $y^{*o}_{i,n}|X^o_{i\cdot,n} \sim \mathcal{N}(X^o_{i\cdot,n}\beta^o_0, S^o_{i\cdot,n}(\lambda^o_0)S^o_{\cdot i,n}(\lambda^o_0)\sigma^2_0)$. If the dependent variable is missing randomly, $y^o_{i,n}|X^o_{i\cdot,n}$ follows the same distribution.

thus has an effect on the estimation precision, and it is desirable to group observations in such a way that the variance of the estimator is minimized. Given that the asymptotic variance is a function of unknown parameters, a two-step procedure might be considered, where the initial estimates based on some primitive grouping are used to construct an optimal grouping. It is unfortunately very difficult, if not impossible, to construct the asymptotic distribution of such a two-step estimator because the grouping becomes data dependent. Moreover, it is not clear how to obtain an optimal grouping practically as it would involve huge computational costs unless very rough approximations of the asymptotic variance are used. For these reasons, we suggest to group observations based on deterministic variables that potentially capture the strength of dependence between observations, for example, the Euclidean distance between observations (see Section 5.1 for details). As discussed in Wang et al. (2013), it is also possible to try a finite number of different grouping schemes and to choose the one which delivers the smallest standard errors.

Let a grouping of observations be described by an index set $\mathcal{G}_n$ containing $n$ pairs $g = (i, j)'$ of observations $i$ and $j$; $\cup_{g\in\mathcal{G}_n}\{g_1, g_2\} = \{1, \ldots, 2n\}$. Let $y^{*s}_{g,n} = (y^{*s}_{g1,n}, y^{*s}_{g2,n})'$ and $y^{*o}_{g,n} = (y^{*o}_{g1,n}, y^{*o}_{g2,n})'$ be 2-dimensional vectors of latent variables in a group $g \in \mathcal{G}_n$. The latent processes for a group $g$ from the reduced form in (3) can be then written as

$$y^{*s}_{g,n} = S^s_{g\cdot,n}(\lambda^s_0)X^s_n\beta^s_0 + \varepsilon^s_{g,n}(\lambda^s_0)$$
$$y^{*o}_{g,n} = S^o_{g\cdot,n}(\lambda^o_0)X^o_n\beta^o_0 + \varepsilon^o_{g,n}(\lambda^o_0)$$

with observed responses $y^s_{g,n}$ and $y^o_{g,n}$, where all variables are defined in the same way as in Section 2 except that now they are defined for a group $g$ instead of an individual $i$; that is, $S^s_{g\cdot,n}(\lambda^s_0) = (S^{s'}_{g1\cdot,n}(\lambda^s_0), S^{s'}_{g2\cdot,n}(\lambda^s_0))'$, $\varepsilon^s_{g,n}(\lambda^s_0) = (\varepsilon^s_{g1,n}(\lambda^s_0), \varepsilon^s_{g2,n}(\lambda^s_0))'$, $y^s_{g,n} = (y^s_{g1,n}, y^s_{g2,n})'$, and so on. The grouped spatial error model can be defined analogously.

Before constructing the log-likelihood function $Q_n(\theta)$ and its population counterpart in the limit $Q_0(\theta)$, note that the log-likelihood function will be composed of four parts because there are four scenarios: one observation in a pair is missing ($y^s_{g1,n} = 1$ and $y^s_{g2,n} = 0$ or vice versa), no observations are missing ($y^s_{g1,n} = y^s_{g2,n} = 1$), and two observations are missing ($y^s_{g1,n} = y^s_{g2,n} = 0$). To simplify notation, we therefore define an index set $\mathcal{A} = \{10, 01, 11, 00\}$ based on the values that $y^s_{g1,n}$ and $y^s_{g2,n}$ take and the corresponding indicator functions $d^a_{g,n} = \mathbb{1}(10y^s_{g1,n} + y^s_{g2,n} = a)$. In order to construct the likelihood function, we also need to introduce some additional notation. Let $\zeta_{g,n} = 2y^s_{g,n} - \iota_2$, where $\iota_2$ is the 2-dimensional vector of ones,

$\tilde{S}^s_{g\cdot,n}(\lambda) = \left(\zeta_{g1,n} S^{s'}_{g1\cdot,n}(\lambda), \quad \zeta_{g2,n} S^{s'}_{g2\cdot,n}(\lambda)\right)'$, and

$$\tilde{\Omega}^{ss}_{g,n}(\theta) = \begin{pmatrix} \Omega^{ss}_{g11,n}(\theta) & \zeta_{g1,n}\zeta_{g2,n}\Omega^{ss}_{g12,n}(\theta) \\ \zeta_{g1,n}\zeta_{g2,n}\Omega^{ss}_{g21,n}(\theta) & \Omega^{ss}_{g22,n}(\theta) \end{pmatrix}, \ \tilde{\Omega}^{so}_{g,n}(\theta) = -\begin{pmatrix} \zeta_{g1,n}\Omega^{so}_{g11,n}(\theta) & \zeta_{g1,n}\Omega^{so}_{g12,n}(\theta) \\ \zeta_{g2,n}\Omega^{so}_{g21,n}(\theta) & \zeta_{g2,n}\Omega^{so}_{g22,n}(\theta) \end{pmatrix}$$

with $\Omega^{ss}_n(\theta) = S^s_n(\lambda^s)S^{s'}_n(\lambda^s)$, $\Omega^{so}_n(\theta) = S^s_n(\lambda^s)S^{o'}_n(\lambda^o)\rho\sigma$, and $\Omega^{oo}_n(\theta) = S^o_n(\lambda^o)S^{o'}_n(\lambda^o)\sigma^2$. Further, let $z_{g,n}(\theta) = y^o_{g,n} - S^o_{g\cdot,n}(\lambda^o)X^o_n\beta^o$, and for any $a \in \mathcal{A}$, $R^a_{g,n}(\theta)$ be the correlation matrix obtained from $\Sigma^a_{g,n}(\theta)$, $v^a_{g,n}(\theta) = (\mathrm{Diag}(\Sigma^a_{g,n}(\theta)))^{-1/2}\left(q_{g,n}(\theta) - \mu^a_{g,n}(\theta)\right)$ with $q_{g,n}(\theta) = \tilde{S}^s_{g\cdot,n}(\lambda^s)X^s_n\beta^s$, and

$$\begin{aligned}
&\mu^{10}_{g,n}(\theta) = \tilde{\Omega}^{so}_{g\cdot 1,n}(\theta)z_{g1,n}(\theta)/\Omega^{oo}_{g11,n}(\theta), && \Sigma^{10}_{g,n}(\theta) = \tilde{\Omega}^{ss}_{g,n}(\theta) - \tilde{\Omega}^{so}_{g\cdot 1,n}(\theta)\tilde{\Omega}^{so'}_{g\cdot 1,n}(\theta)/\Omega^{oo}_{g11,n}(\theta), \\
&\mu^{01}_{g,n}(\theta) = \tilde{\Omega}^{so}_{g\cdot 2,n}(\theta)z_{g2,n}(\theta)/\Omega^{oo}_{g22,n}(\theta), && \Sigma^{01}_{g,n}(\theta) = \tilde{\Omega}^{ss}_{g,n}(\theta) - \tilde{\Omega}^{so}_{g\cdot 2,n}(\theta)\tilde{\Omega}^{so'}_{g\cdot 2,n}(\theta)/\Omega^{oo}_{g22,n}(\theta), \\
&\mu^{11}_{g,n}(\theta) = \tilde{\Omega}^{so}_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)z_{g,n}(\theta), && \Sigma^{11}_{g,n}(\theta) = \tilde{\Omega}^{ss}_{g,n}(\theta) - \tilde{\Omega}^{so}_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)\tilde{\Omega}^{so'}_{g,n}(\theta), \\
&\mu^{00}_{g,n}(\theta) = 0, && \Sigma^{00}_{g,n}(\theta) = \tilde{\Omega}^{ss}_{g,n}(\theta).
\end{aligned}$$

Then the log-likelihood function based on a grouping $\mathcal{G}_n$ is defined by (see the derivation in Appendix G.1)

$$\begin{aligned}
Q_n(\theta) = \frac{1}{n}\sum_{g\in\mathcal{G}_n}\Bigg\{ &\mathbb{1}(y^s_{g1,n}=1, y^s_{g2,n}=0)\ln\left[\frac{1}{\sqrt{\Omega^{oo}_{g11,n}(\theta)}}\phi\left(\frac{z_{g1,n}(\theta)}{\sqrt{\Omega^{oo}_{g11,n}(\theta)}}\right)\Phi_2\left(v^{10}_{g,n}(\theta), R^{10}_{g,n}(\theta)\right)\right] \\
&+ \mathbb{1}(y^s_{g1,n}=0, y^s_{g2,n}=1)\ln\left[\frac{1}{\sqrt{\Omega^{oo}_{g22,n}(\theta)}}\phi\left(\frac{z_{g2,n}(\theta)}{\sqrt{\Omega^{oo}_{g22,n}(\theta)}}\right)\Phi_2\left(v^{01}_{g,n}(\theta), R^{01}_{g,n}(\theta)\right)\right] \\
&+ \mathbb{1}(y^s_{g1,n}=1, y^s_{g2,n}=1)\ln\left[\phi_2\left(z_{g,n}(\theta), \Omega^{oo}_{g,n}(\theta)\right)\Phi_2\left(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta)\right)\right] \\
&+ \mathbb{1}(y^s_{g1,n}=0, y^s_{g2,n}=0)\ln\left[\Phi_2\left(v^{00}_{g,n}(\theta), R^{00}_{g,n}(\theta)\right)\right]\Bigg\} \\
= \frac{1}{n}\sum_{g\in\mathcal{G}_n}\sum_{a\in\mathcal{A}} &d^a_{g,n}f^a_{g,n}(\theta),
\end{aligned} \tag{4}$$

where $f^a_{g,n}(\theta)$, $a \in \mathcal{A}$, represent the log-density functions, $\phi(\cdot)$ is the standard normal density function, and $\phi_2(\cdot,\Sigma)$ and $\Phi_2(\cdot,\Sigma)$ are the bivariate normal density and distribution functions, respectively, with zero mean and variance matrix $\Sigma$.

Although the log-likelihood function looks complicated, it is not difficult to implement and to maximize. If there is the spatial error instead of the spatial lag in the selection or outcome equations, $z_{g,n}(\theta)$ and $q_{g,n}(\theta)$ have to be replaced with $z^e_{g,n}(\theta) = y^o_{g,n} - X^o_{g\cdot,n}\beta^o$ and $q^e_{g,n}(\theta) = \tilde{X}^s_{g\cdot,n}\beta^s$, respectively, where $\tilde{X}^s_{g\cdot,n}$ is constructed in the same way as $\tilde{S}^s_{g\cdot,n}(\lambda)$.

# 4 Asymptotic Properties of Partial Maximum Likelihood Estimator

The main difficulty in proving asymptotic properties of PMLE stems from analyzing the nonlinear objective function based on heterogeneous and spatially dependent processes. Hence, this dependence has to be restricted to a manageable degree. We do so by employing the near epoch dependent (NED) random fields framework developed by Jenish and Prucha (2012). We consider a topological structure proposed in their paper. Let the location of an observation $i$ be defined by $l_i \in \widetilde{\mathfrak{D}}_n$, where $\widetilde{\mathfrak{D}}_n$ is a finite sample region of a $\tilde{d}$-dimensional lattice $\widetilde{\mathfrak{D}} \subset \mathbb{R}^{\tilde{d}}$, $\tilde{d} > 1$, equipped with the Euclidean metric. Since the likelihood function in (4) is in terms of likelihood contributions for pairs, let a group $g = (i,j)'$ be assigned a location $l_g = (l'_{g_1}, l'_{g_2})' = (l'_i, l'_j)' \in \widetilde{\mathfrak{D}}_n \times \widetilde{\mathfrak{D}}_n = \mathfrak{D}_n$, which is a finite sample region of a $2\tilde{d}$-dimensional lattice $\mathfrak{D} = \widetilde{\mathfrak{D}} \times \widetilde{\mathfrak{D}} \subset \mathbb{R}^{2\tilde{d}}$. Given this definition, the distance between two groups $g$ and $\dot{g}$ depends on configurations of four points in $\mathbb{R}^{\tilde{d}}$. We consider a distance metric between two points in $\mathbb{R}^{2\tilde{d}}$ defined by $d(l_g, l_{\dot{g}}) = \min\{\|(l'_{g_1}, l'_{g_2})' - (l'_{\dot{g}_1}, l'_{\dot{g}_2})'\|, \|(l'_{g_1}, l'_{g_2})' - (l'_{\dot{g}_2}, l'_{\dot{g}_1})'\|\}$. The distance between any two subsets $\mathfrak{A}, \mathfrak{B} \subseteq \mathfrak{D}$ is defined as $d(\mathfrak{A}, \mathfrak{B}) = \inf\{d(g, \dot{g}) : l_g \in \mathfrak{A}, l_{\dot{g}} \in \mathfrak{B}\}$, where the fact that the observations are indexed by natural numbers allows us to write $d(g, \dot{g}) \equiv d(l_g, l_{\dot{g}})$ for two groups $g$ and $\dot{g}$ with locations $l_g, l_{\dot{g}} \in \mathbb{R}^{2\tilde{d}}$.

**Assumption 3.** *Individual units in the economy are located or living in a region $\widetilde{\mathfrak{D}}_n \subset \widetilde{\mathfrak{D}} \subset \mathbb{R}^{\tilde{d}}$. The cardinality of $\mathfrak{D}_n = \widetilde{\mathfrak{D}}_n \times \widetilde{\mathfrak{D}}_n$ satisfies $\lim_{n\to\infty} |\mathfrak{D}_n| = \infty$. The distance $d(g, \dot{g})$ between any two different groups $g$ and $\dot{g}$ is larger than or equal to a specific positive constant, which we normalize to 1.*

Region $\mathfrak{D}$ corresponds to a space of economic or geographic characteristics or a mixture of them. In Example 2, a geographical space can simply be used. Although there is no natural location for an observation $i$ in Example 1, a location can be constructed. Assume that there are $t = 1, \ldots, T$ tutorial groups with at most $\bar{S}$ students in each group. Let $t_i$ be the tutorial group of student $i$ and $a_i$ be his rank in tutorial group $t_i$ based on the alphabetical ordering. Then student $i$'s location can be given by $l_i = (t_i\bar{S}, a_i)'$. Assumption 3 implies that the increasing domain asymptotics is used (as an alternative to the infill domain asymptotics): the distance restriction in Assumption 3 implies that there is a finite number of units in any bounded region and that the sample region $\mathfrak{D}_n$ has to expand when the sample grows.

For reference, the definitions of $\alpha$-mixing and NED properties presented in Jenish and Prucha (2009, 2012) are reviewed first.

**Definition 1.** *Let $\{\eta_{g,n}\}_{g\in\mathcal{G}_n}$ be a triangular array of real random variables defined on a probability space $(\Omega,\mathcal{F},P)$. Moreover, let $\mathcal{A}$ and $\mathcal{B}$ be two $\sigma$-algebras of $\mathcal{F}$ and*

$$\alpha(\mathcal{A},\mathcal{B})=\sup(|P(A\cap B)-P(A)P(B)|, A\in\mathcal{A}, B\in\mathcal{B}).$$

*For $\mathfrak{A}\subseteq\mathfrak{D}_n$ and $\mathfrak{B}\subseteq\mathfrak{D}_n$, let $\sigma_n(\mathfrak{A})=\sigma(\eta_{g,n}:l_g\in\mathfrak{A})$ and $\alpha_n(\mathfrak{A},\mathfrak{B})=\alpha(\sigma_n(\mathfrak{A}),\sigma_n(\mathfrak{B}))$. Then the $\alpha$-mixing coefficients for the random fields $\{\eta_{g,n}\}_{g\in\mathcal{G}_n}$ are defined as*

$$\bar{\alpha}(k,m,s)=\sup_n\sup_{\mathfrak{A},\mathfrak{B}}(\alpha_n(\mathfrak{A},\mathfrak{B}), |\mathfrak{A}|\le k, |\mathfrak{B}|\le m, d(\mathfrak{A},\mathfrak{B})\ge s),$$

*where $|\cdot|$ denotes the cardinality of a set.*

This definition is similar to the time series literature. The major difference is that, in the random fields setting, the $\alpha$-mixing coefficients do not only depend on the distance between two sets but also on the sizes of the sets. The definition of the NED property follows.

**Definition 2.** *Let $\{Z_{g,n}\}_{g\in\mathcal{G}_n}$ and $\{\eta_{g,n}\}_{g\in\mathcal{G}_n}$ be random fields located on $\mathfrak{D}_n$, and additionally, $\{Z_{g,n}\}_{g\in\mathcal{G}_n}$ satisfy $\|Z_{g,n}\|_p<\infty$, $p\ge 1$. Moreover, let $\{t_{g,n}\}_{g\in\mathcal{G}_n}$ be an array of positive constants. Then the random field $\{Z_{g,n}\}_{g\in\mathcal{G}_n}$ is said to be $L_p$-near epoch dependent on the random field $\{\eta_{g,n}\}_{g\in\mathcal{G}_n}$ if*

$$\|Z_{g,n}-E[Z_{g,n}|\mathcal{F}_{g,n}(s)]\|_p\le t_{g,n}\psi(s)$$

*for some sequence $\psi(s)\ge 0$ with $\lim_{s\to\infty}\psi(s)=0$, where $\mathcal{F}_{g,n}(s)=\sigma(\eta_{\dot{g},n}:d(g,\dot{g})\le s), s\in\mathbb{N}$. The NED random field is uniform if and only if $\sup_{n,g} t_{g,n}<\infty$.*

In Definition 2, the term $Z_{g,n}-E[Z_{g,n}|\mathcal{F}_{g,n}(s)]$ measures the prediction error of $Z_{g,n}$ based on the information contained in $\{\eta_{\dot{g},n}:d(g,\dot{g})\le s\}$. The NED property then states that the prediction error converges to zero as $s$ increases. Note that NED is not a property of a random variable itself as $\alpha$-mixing is, but it is a property of a mapping.

## 4.1 Consistency

To prove consistency, we need to introduce additional assumptions.

**Assumption 4.** *(i) $\{(X^s_{g\cdot,n},\ X^o_{g\cdot,n})\}_{g\in\mathcal{G}_n}$ is an $\alpha$-mixing random field with $\alpha$-mixing coefficients $\bar{\alpha}(k,m,s)\le(k+m)^\tau\hat{\alpha}(s)$, $\tau\ge 0$, for some $\hat{\alpha}(s)\to 0$ as $s\to\infty$ such that $\sum_{s=1}^{\infty}s^{2\tilde{d}-1}\hat{\alpha}(s)<\infty$. (ii) $\sup_{n,i}E\|X^b_{i\cdot,n}\|^p<\infty$*

*and* $\sup_{n,i} E\left[\|X^b_{i\cdot,n}\|^p | y^s_{i,n} = 1\right] < \infty$ *for any* $1 \le p \le \bar{p}$*, where* $\bar{p} > 18$ *and* $b \in \{s, o\}$.

Assumption $4(i)$ states that the exogenous variables may be cross-sectionally dependent under some restrictions. Assumption $4(ii)$ implies that $\bar{p}$ moments, $\bar{p} > 18$, of the exogenous variables exist. This is less restrictive than the assumption that the support of the exogenous variables is bounded, which is usual in the (spatial) literature studying discrete choice or limited dependent variable models.[10] Moreover, the large number $p$ of finite moments is related to the assumed normality of the errors and the need to bound the moments of the logarithm of the bivariate normal cumulative distribution function and their derivatives (cf. Lemma C.5). It is likely that for heavier-tailed error distributions (e.g., the Laplace distribution), a substantially smaller number of moments would have to exist. The derivatives of the log-likelihood function are inversely proportional to the values of the distribution function, and therefore, are likely to be increasing slower for heavier-tailed distributions. Thus less moments of the explanatory variables could potentially be required.

The elements of the spatial weight matrices determine the strength of dependence between observations. An important question is under which structures of spatial weight matrices the limit laws based on the NED framework hold. Given that the likelihood function is specified in terms of (the inverses $S^s_n(\lambda^s)$ and $S^o_n(\lambda^o)$ of) $I_{2n} - \lambda^s W^s_n$ and $I_{2n} - \lambda^o W^o_n$ and the grouping is determined by the user of the method, we impose restrictions on the weight matrices indirectly by the following assumption.

**Assumption 5.** $\lim_{s\to\infty} \psi(s) = 0$ *with* $\psi(s) = \max\{\psi^s(s), \psi^o(s)\}$*, where* $\psi^b(s) = \sup_{n,g} \sup_{\theta\in\Theta} \sum_{\dot{g}:d(g,\dot{g})>s} \|S^b_{g\dot{g},n}(\lambda^b)\| / \sup_{n,g} \sup_{\theta\in\Theta} \sum_{\dot{g}\in\mathcal{G}_n} \|S^b_{g\dot{g},n}(\lambda^b)\|$*,* $b \in \{s, o\}$.

As shown in the proof of Theorem 1, Assumption 5 is needed to show that $\{\sum_{a\in\mathcal{A}} d^a_{g,n} f^a_{g,n}(\theta)\}_{g\in\mathcal{G}_n}$ is a NED random field on the $\alpha$-mixing random field $\{\eta_{g,n} = (X^s_{g\cdot,n}, X^o_{g\cdot,n}, u^s_{g,n}, u^o_{g,n})\}_{g\in\mathcal{G}_n}$. The main implication of Assumption 5 is that dependence between pairs should decrease to zero with an increasing distance between the pairs. Given the Euclidean norm used in Assumption 5, this means that the weights $|S^b_{ij,n}(\lambda)|$ between observations $i \in g$ and $j \in \dot{g}$ have to decrease with their distance. For example, it follows similarly to Qu and Lee (2015) that Assumption 5 and also Assumption $1(ii)$ hold if the spatial weights $W^b_{ij,n}$, $b \in \{s, o\}$, can be bounded by $C_c\|l_i - l_j\|^{-C_p\tilde{d}}$ for some $C_c > 0$ and $C_p > 2$, and additionally in the case of an asymmetric $W^b_n$, if the number of columns with their absolute sums exceeding the row norm $\|W^b_n\|_\infty$ is bounded uniformly in $n$ (see Appendix H). Assumption 5 is also satisfied if the observations with their distance above a certain threshold have zero spatial weights. For instance, if tutorial groups are analyzed in Example 1, it is typically assumed that the $ij$th element of the weight matrix is equal to zero if students $i$ and $j$ are from different

[10] E.g., see condition (v) of Theorem 1 by Pinkse and Slade (1998) and condition (vi) of Theorem 1 by Wang et al. (2013).

tutorial groups. Thus assuming that the number of students in each tutorial group is even, it is beneficial to form only pairs of students who are in the same tutorial group. Given the definition of locations for Example 1 presented below Assumption 3, if $d(g,\dot{g}) \geq \bar{S}$, students in pairs $g$ and $\dot{g}$ are from different tutorial groups implying that $\|S^b_{g\dot{g},n}(\lambda^b)\| = 0$ and Assumption 5 is trivially satisfied.

Next, we make an assumption about $2\times 2$ submatrices of matrices $\Omega^{ss}_n(\theta)$ and $\Omega^{oo}_n(\theta)$ and $\Sigma^a_{g,n}(\theta)$, $a \in \mathcal{A}$, defined in Section 3.

**Assumption 6.** *The minimum eigenvalues of matrices $\Omega^{ss}_{g,n}(\theta)$, $\Omega^{oo}_{g,n}(\theta)$, and $\Sigma^a_{g,n}(\theta)$, $a \in \mathcal{A}$, are bounded away from zero uniformly in $n \in \mathbb{N}$, $g \in \mathcal{G}_n$, and $\theta \in \Theta$, where $\Theta$ is the parameter space of $\theta$.*

Assumption 6 ensures that the above mentioned $2\times 2$ (sub)matrices are invertible for each pair. Thus, the observations should be grouped in such a way that this assumption is not violated. Its validity can be checked using a grid covering possible values of the spatial parameters and the correlation coefficient since matrices $\Omega^{ss}_{g,n}(\theta)$, $\Omega^{oo}_{g,n}(\theta)/\sigma^2$, and $\Sigma^a_{g,n}(\theta)$, $a \in \mathcal{A}$, do not depend on regression parameters $\beta^s$, $\beta^o$, and variance $\sigma^2$.

**Assumption 7.** *The parameter space $\Theta$ is a compact subset of $\mathbb{R}^L$.*

**Assumption 8.** *The population log-likelihood functions $E[Q_n(\theta)]$ are uniquely maximized at $\theta_0$ for $n \geq n_0$ and some $n_0 \in \mathbb{N}$: $\liminf_{n\to\infty}(E[Q_n(\theta_0)] - \sup_{\|\theta-\theta_0\|>\epsilon} E[Q_n(\theta)]) > 0$ for any $\epsilon > 0$.*

Whereas Assumption 7 is a standard assumption for nonlinear extremum estimators, Assumption 8 is the identification condition for PMLE, allowing the non-existence of the limit of $E[Q_n(\theta)]$.[11] If $Q_0(\theta) = \lim_{n\to\infty} E[Q_n(\theta)]$ exists, $Q_0(\theta)$ is simply required to have a unique maximum at $\theta_0$. This PMLE likelihood is based on the correctly specified distribution for each pair of observations, and therefore, it attains its maximum at the true parameter values by the Kullback-Leibler information inequality (see also Appendix G for the validity of the population first-order conditions). We assume rather than prove identification because it is a very challenging task. Wooldridge (1994) claims that usually some additional knowledge about the distribution of conditioning variables is needed to establish identification, and therefore, it is assumed rather than proved in many cases. For this reason, many other papers analyzing nonlinear spatial models make assumptions similar to Assumption 8, for example, see Wang et al. (2013) and Xu and Lee (2015a,b, 2018).

Finally, the consistency of the proposed PMLE follows.

**Theorem 1.** *Under Assumptions 1–8, $\hat{\theta}_n - \theta_0 = o_p(1)$ as $n \to \infty$.*

[11] We thank an anonymous referee for suggesting this generalization allowing for very general forms of heterogeneity defined by the spatial weight matrices.

### 4.2 Asymptotic Normality

In order to establish asymptotic normality, we need to strengthen the assumptions regarding the dependence structure. In particular, the NED coefficients of random field $\{\sum_{a\in\mathcal{A}} d^a_{g,n}\partial f^a_{g,n}(\theta_0)/\partial\theta\}_{g\in\mathcal{G}_n}$ have to decrease to zero at a certain relatively fast rate, see Assumption 10 below, because the likelihood function is not only nonlinear but contains indicator functions as well. Additionally, some standard regularity conditions are required.

**Assumption 9.** *(i)* $\{(X^s_{g\cdot,n},\ X^o_{g\cdot,n})\}_{g\in\mathcal{G}_n}$ *is an* $\alpha$*-mixing random field with* $\alpha$*-mixing coefficients* $\bar{\alpha}(k,m,s)\leq(k+m)^\tau\hat{\alpha}(s)$ *with some* $\hat{\alpha}(s)\to 0$ *as* $s\to\infty$ *such that for some* $\delta>0$, $\sum_{s=1}^{\infty} s^{2\tilde{d}(\tau_*+1)-1}\hat{\alpha}^{\delta/(4+2\delta)}(s)<\infty$, *where* $\tau_*=\delta\tau/(2+\delta)$, $\tau\geq 0$. *(ii)* $\sup_{n,i} E\|X^b_{i\cdot,n}\|^p<\infty$ *and* $\sup_{n,i} E\left[\|X^b_{i\cdot,n}\|^p|y^s_{i,n}=1\right]<\infty$ *for any* $p\geq 1$.

**Assumption 10.** *The maximum* $\psi(s)$ *of the NED coefficients* $\psi^s(s)$ *and* $\psi^o(s)$ *defined in Assumption 5 satisfies* $\sum_{s=1}^{\infty} s^{2\tilde{d}-1}\psi^{(r-2)/(12r-12)}(s)<\infty$ *for some* $r>2$.

**Assumption 11.** $\theta_0$ *is in the interior of the parameter space* $\Theta$.

**Assumption 12.** *Let* $\min\operatorname{eig} A$ *represents the smallest eigenvalue of matrix* $A$ *and assume for some* $n_0\in\mathbb{N}$ *that (i)* $H_n(\theta_0)=E\left[\frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'}\right]$ *exists, is finite, and non-singular; (ii)* $J_n(\theta_0)=E\left[n\frac{\partial Q_n(\theta_0)}{\partial\theta}\frac{\partial Q_n(\theta_0)}{\partial\theta'}\right]$ *exists, is finite, and* $\inf_{n\geq n_0}\min\operatorname{eig} J_n(\theta_0)>0$.

Assumption 9$(ii)$ implies that infinitely many moments of the exogenous variables exist. The proof of asymptotic normality requires only finitely many moments but finding the exact $p$, which is much larger than 18 imposed in Assumption 4$(ii)$ and thus practically irrelevant, would require a lot of effort such as calculating the third order derivatives of the bivariate normal distribution functions in (4). Some sufficient conditions on the spatial weight matrices for Assumption 10 are provided in Appendix H. Assumption 12 admits again that the limits $H(\theta_0)=\lim_{n\to\infty}H_n(\theta_0)$ and $J(\theta_0)=\lim_{n\to\infty}J_n(\theta_0)$ do not necessarily exist due to practically unrestricted heterogeneity defined by the spatial weight matrices. If these limits are well defined, Assumption 12 becomes the standard requirement that the Jacobian and Hessian matrices corresponding to the limit of the population partial maximum likelihood function exist and are non-singular. Note that the non-singular Hessian matrix is not necessary for the asymptotic normality formulated in the following theorem, but for its use in real data analysis.

**Theorem 2.** *Under Assumptions 1–3, 5–12,* $\sqrt{n}J_n^{-1/2}(\theta_0)H_n(\theta_0)(\hat{\theta}_n-\theta_0)\xrightarrow{d}\mathcal{N}(0,I_L)$ *as* $n\to\infty$.

Finally, since the likelihood function does not account for the dependence between groups, the variance matrix of PMLE is not equal to the inverse of the Fisher information matrix. Thus, PMLE is in general not efficient as the full MLE is.

### 4.3 Estimation of the Variance Matrix

Although we do not model the dependence between pairs in the likelihood function, it has to be accounted for when the variance matrix is estimated. On the one hand, it is relatively easy to estimate the Hessian matrix $H_n(\theta_0) = E[\partial^2 Q_n(\theta_0)/\partial\theta\partial\theta']$ as it can be obtained using its sample analog and a consistent estimate of $\theta$: $\hat{H}_n(\hat{\theta}_n) = \partial^2 Q_n(\hat{\theta}_n)/\partial\theta\partial\theta'$. On the other hand, estimation of the variance matrix $J_n(\theta_0)$ and its limit $J(\theta_0)$ is complicated due to the dependence between groups. It is theoretically possible to consider a spatial analog of a heteroskedasticity and autocorrelation consistent (HAC) estimator of the variance matrix that has been extensively analyzed in the time series literature (i.e., Newey and West, 1987, and Andrews, 1991). Conley (1999) adapted the HAC estimator for the spatially stationary observations. Noting that the Cliff-Ord type models are in general not spatially stationary, Kelejian and Prucha (2007) and Kim and Sun (2011) relaxed the stationarity assumption, but considered only processes linear in error terms. This is not the case for $\partial Q_n(\theta_0)/\partial\theta$ here, and therefore, the HAC estimator is not easily applicable in the present setting.

On the other hand, it is not uncommon to estimate the variance of an estimator of a spatial model using the bootstrap when it is very difficult or practically impossible to obtain a closed form expression of the variance matrix (e.g., a residual based bootstrap method is used by Su and Yang, 2015, in spatial dynamic panel data models). The bootstrap standard errors are also often recommended in the linear regression models with clustered errors (Cameron et al., 2008), which are closely related to model (2), although stronger dependence (larger clusters) can adversely affect traditional boostrap procedures such as the wild bootstrap (MacKinnon and Webb, 2017). However given that the considered sample selection models are completely parametrically specified, it is possible to use the parametric bootstrap to estimate $J_n(\theta_0)$. Note that we suggest to bootstrap $J_n(\theta_0)$ instead of the complete variance matrix of the estimator to guarantee good computational speed. The bootstrap procedure for estimating $J_n(\theta_0)$ can be described for the spatial lag model as follows (the spatial error model can be dealt with analogously).

1. Obtain the partial maximum likelihood estimate $\hat{\theta}_n = (\hat{\beta}_n^{s\prime}, \hat{\beta}_n^{o\prime}, \hat{\lambda}_n^s, \hat{\lambda}_n^o, \hat{\rho}_n, \hat{\sigma}_n^2)'$.
2. For every $b = 1, \ldots, B$, generate a random sample $(u_{i,n}^{s(b)}, u_{i,n}^{o(b)})'$ of size $2n$ from the distribution $\mathcal{N}(0, \Sigma(\hat{\theta}_n))$, where $\Sigma(\hat{\theta}_n) = [1\ \hat{\rho}_n\hat{\sigma}_n;\ \hat{\rho}_n\hat{\sigma}_n\ \hat{\sigma}_n^2]$.

3. Given $W_n^s$, $W_n^o$, $X_n^s$, $X_n^o$, and $\hat{\theta}_n$, generate the bootstrap data (indexed by $b$) according to

$$y_n^{*s(b)} = S_n^s(\hat{\lambda}_n^s)X_n^s\hat{\beta}_n^s + \varepsilon_n^{s(b)}(\hat{\lambda}_n^s)$$
$$y_n^{*o(b)} = S_n^o(\hat{\lambda}_n^o)X_n^o\hat{\beta}_n^o + \varepsilon_n^{o(b)}(\hat{\lambda}_n^o)$$

with observed responses $y_{i,n}^{s(b)} = \mathbb{1}(y_{i,n}^{*s(b)} > 0)$ and $y_{i,n}^{o(b)} = y_{i,n}^{s(b)}y_{i,n}^{*o(b)}$, where $\varepsilon_n^{s(b)}(\hat{\lambda}_n^s) = S_n^s(\hat{\lambda}_n^s)u_n^{s(b)}$ and $\varepsilon_n^{o(b)}(\hat{\lambda}_n^o) = S_n^o(\hat{\lambda}_n^o)u_n^{o(b)}$.

4. Compute the score $\Gamma_n^{(b)}(\hat{\theta}_n) = \partial Q_n^{(b)}(\hat{\theta}_n)/\partial\theta$ for $B$ bootstrap samples; $Q_n^{(b)}(\hat{\theta}_n)$ is thus obtained using $y_{i,n}^{s(b)}$, $y_{i,n}^{o(b)}$, $W_n^s$, $W_n^o$, $X_n^s$, and $X_n^o$, where $b = 1, \ldots, B$.

5. Finally, the bootstrap estimate of $J_n(\theta_0)$ is given by

$$\hat{J}_n(\hat{\theta}_n) = \frac{n}{B-1}\sum_{b=1}^{B}\left(\Gamma_n^{(b)}(\hat{\theta}_n) - \frac{1}{B}\sum_{b=1}^{B}\Gamma_n^{(b)}(\hat{\theta}_n)\right)\left(\Gamma_n^{(b)}(\hat{\theta}_n) - \frac{1}{B}\sum_{b=1}^{B}\Gamma_n^{(b)}(\hat{\theta}_n)\right)'.$$

Since the functional form of the derivatives of the likelihood function is rather complicated, we suggest to use numerical differentiation to evaluate $\Gamma_n^{(b)}$ in step 4. To achieve consistent estimation, the step size of the numerical derivatives has to decrease to zero as the sample size increases (Hong et al., 2015), although no particular rate is required as just averages of the numerically differentiated functions are evaluated. In practice, the step size should not be too small because this might yield large rounding errors. In our simulation study (see Section 5), we used a widely adopted rule of thumb and calculated the first-order derivatives using the step size $\sqrt{\texttt{eps}} \cdot \operatorname{sgn}(\theta_j) \cdot \|\theta\|_\infty$, $j = 1, \ldots L$, where `eps` is a machine epsilon and $\operatorname{sgn}(\theta_j)$ is equal to -1 if $\theta_j$ is negative and is equal to 1 otherwise.[12] Additionally, the bootstrapped first-order conditions are now shown to consistently approximate the asymptotic distribution, and due to their uniform integrability verified in the proof of Theorem 3, also the moments of the first-order conditions. The result however requires an analog of Assumption 10 for a product of the spatial matrices.

**Assumption 13.** $\lim_{s\to\infty}\tilde{\psi}^s(s) = 0$, *where* $\tilde{\psi}^s(s) = \sup_{n,g}\sup_{\theta\in\Theta}\sum_{\dot{g}:d(g,\dot{g})>s}\|S_{g\dot{g},n}^{s3}(\lambda^s)\|/\sup_{n,g}\sup_{\theta\in\Theta}\sum_{\dot{g}\in\mathcal{G}_n}\|S_{g\dot{g},n}^{s3}(\lambda^s)\|$, $S_n^{s3}(\lambda) = S_n^s(\lambda)W_n^sS_n^s(\lambda)$, *and additionally, the maximum* $\psi(s)$ *of NED coefficients* $\psi^s(s)$, $\psi^o(s)$, *and* $\tilde{\psi}^s(s)$ *satisfies* $\sum_{s=1}^{\infty}s^{2\tilde{d}-1}\psi^{(r-2)/(24r-24)}(s) < \infty$, *for some* $r > 2$.

**Theorem 3.** *Under Assumptions 1–13,* $\sup_{c\in\mathbb{R}}|P^b(\sqrt{n}J_n^{-1/2}(\theta_0)\Gamma_n^{(b)}(\hat{\theta}_n) \le c) - P(\mathcal{N}(0, I_L) \le c)| = o_p(1)$ *as* $n \to \infty$, *and*

$$\sup_{c\in\mathbb{R}}|P^b(\sqrt{n}\Gamma_n^{(b)}(\hat{\theta}_n) \le c) - P(\sqrt{n}\Gamma_n(\theta_0) \le c)| \xrightarrow{p} 0,$$

[12] For the second-order derivatives, we used $\sqrt[4]{\texttt{eps}} \cdot \operatorname{sgn}(\theta_j) \cdot \|\theta\|_\infty$.

*as $n \to \infty$, where $P^b$ denotes the probability measure induced by the parametric bootstrap.*

# 5 Monte Carlo Simulations

## 5.1 Experimental Design

We consider the following data generating process:

$$y_n^{*s} = (I_{2n} - \lambda^s W_n^s)^{-1}(X_n^s \beta^s + u_n^s)$$
$$y_n^{*o} = (I_{2n} - \lambda^o W_n^o)^{-1}(X_n^o \beta^o + u_n^o)$$

for the spatial lag model and

$$y_n^{*s} = X_n^s \beta^s + (I_{2n} - \lambda^s W_n^s)^{-1} u_n^s$$
$$y_n^{*o} = X_n^o \beta^o + (I_{2n} - \lambda^o W_n^o)^{-1} u_n^o$$

for the spatial error model, where $X_{i\cdot,n}^s = (X_{i1,n}^s, X_{i2,n}^s, X_{i3,n}^s)$ and $X_{i\cdot,n}^o = (X_{i1,n}^o, X_{i2,n}^o, X_{i3,n}^o)$ with $X_{i1,n}^s = X_{i1,n}^o = 1$, $X_{i2,n}^s = X_{i2,n}^o \overset{iid}{\sim} \mathcal{N}(0,1)$, $X_{i3,n}^s \overset{iid}{\sim} \chi^2(1)$, and $X_{i3,n}^o \overset{iid}{\sim} \chi^2(1)$. The error terms $(u_{i,n}^s, u_{i,n}^o)' \overset{iid}{\sim} \mathcal{N}(0,\Sigma)$, where $\Sigma = [1\ 0.5;\ 0.5\ 1]$. The parameters $(\beta_2^s, \beta_3^s, \beta_1^o, \beta_2^o, \beta_3^o) = (1, -1, 1, 1, -1)$, while $\beta_1^s$ is chosen such that $P[y_{i,n}^s = 1 | W_n^s; \lambda^s] = 2/3$. We analyze all the possible combinations of the spatial parameters $\lambda^s$ and $\lambda^o$ taking values from the set $\{0, 0.4, 0.85\}$.

Let $D_n$ represent great-cycle distances in miles between counties in the US.[13] Similarly to Xu and Lee (2015a), we use counties in the 10 Upper Great Plains States for $2n = 760$.[14] For $2n = 344$, counties in Nebraska, South Dakota, Minessota, and Iowa are used, whereas only the first two states are utilized for $2n = 158$.[15] The weight matrices are generated as follows: $\widetilde{W}_{ij,n}^s = \widetilde{W}_{ij,n}^o = \mathbb{1}(D_{ij,n} \leq 50) \cdot 1/D_{ij,n}$. We row-normalize $\widetilde{W}_n^s$ and $\widetilde{W}_n^o$ in order to get $W_n^s$ and $W_n^o$. Wang et al. (2013) propose to group adjacent observations, for instance, using the Euclidean distance. Based on this idea, we formulate the following

[13] The data is available at http://data.nber.org/data/county-distance-database.html.

[14] The ten states include Nebraska, South Dakota, Minnesota, Iowa, Colorado, Kansas, Missouri, Montana, North Dakota, and Wyoming.

[15] To obtain a sample size that is an even number, Adams county in Nebraska is excluded for $2n \in \{158, 344\}$.

integer linear programing (ILP) problem:[16]

$$\min_{e_{ij,n}:i=1,\ldots,2n,j=1,\ldots,2n} \sum_{i=1}^{2n}\sum_{j=1}^{2n} e_{ij,n}\widetilde{D}_{ij,n}$$

$$\text{s.t. } \sum_{j=1}^{2n} e_{ij,n} = 1,\; e_{ii,n} = 0, \quad e_{ij,n} = e_{ji,n},$$

where $e_{ij,n}$ is a binary variable, which is equal to 1 if observations $i$ and $j$ form a pair and to zero otherwise; $\widetilde{D}_{ij,n} = \mathbb{1}(D_{ij,n} \leq 50)D_{ij,n}$. We use $\widetilde{D}_n$ instead of $D_n$ in the ILP problem in order to reduce the burden of computation. The effect on the grouping is small since the algorithm groups nearby observations.

The partial maximum likelihood estimator (PMLE) is compared to the heteroskedastic maximum likelihood estimator (HMLE), which has a likelihood function of a form similar to (4) but for univariate rather than bivariate observations. The model with the spatial error is also estimated by the GMM estimator proposed by Flores-Lagunes and Schnier (2012). Two versions of the GMM estimator are explored: with the identity weight matrix (GMM) and with the optimal weight matrix (GMM2).

Given that $\beta^s$ and $\beta^o$ are 3-dimensional vectors, bias, standard deviation, and root mean squared error (RMSE) are calculated for each element of $\beta^s$ and $\beta^o$ separately. In Tables 1, 2, 4, 6, 7, 10, and 11 in Appendix A, however, only the Euclidean norms of the vectors of the corresponding statistics are reported, i.e. $\text{stat}(\hat{\beta}_n^b) = \|(\text{stat}(\hat{\beta}_{1n}^b), \text{stat}(\hat{\beta}_{2n}^b), \text{stat}(\hat{\beta}_{3n}^b))\|$, where $b \in \{s, o\}$ and $\text{stat} \in \{\text{bias}, \text{s.d.}, \text{rmse}\}$.

In many empirical applications, marginal effects play a crucial role. For this reason, we also consider the following marginal effects: $mfx_1 = \partial P(y_{i,n}^s = 1|X_n^s)/\partial X_{j2,n}^s$, $mfx_2 = \partial P(y_{i,n}^s = 1|X_n^s)/\partial X_{j3,n}^s$, $mfx_3 = \partial E\left[y_{i,n}^{*o}|y_{i,n}^s = 1, X_n^s, X_n^o\right]/\partial X_{j2,n}^o$, and $mfx_4 = \partial E\left[y_{i,n}^{*o}|y_{i,n}^s = 1, X_n^s, X_n^o\right]/\partial X_{j3,n}^o$; the formulas are presented in Appendix G.3. (In the spatial error model, the marginal effects are conditional on $X_{i\cdot,n}^s$ and $X_{i\cdot,n}^o$ instead of $X_n^s$ and $X_n^o$.) For spatial lag models, three types of marginal effects might be considered – total, direct, and indirect – as discussed by LeSage and Pace (2009). In this paper, we discuss only total marginal effects. Since the marginal effects are different for each individual, we use the average marginal effects to calculate bias, standard deviation, and RMSE, e.g.,

$$\text{bias} = H^{-1}\sum_{h=1}^{H}\left(\overline{mfx^{(h)}}(\hat{\theta}_n^{(h)}) - \overline{mfx^{(h)}}(\theta_0^{(h)})\right),$$

where $\overline{mfx^{(h)}}(\theta) = (2n)^{-1}\sum_{i=1}^{2n} mfx^{(h)}(i,\theta)$ with $mfx^{(h)}(i,\theta)$ being one of the four marginal effects evaluated for an individual $i$ using parameter $\theta$ at iteration $h = 1, \ldots, H$. The true marginal effects in Tables 3

[16]We solve this problem in Matlab using the IBM ILOG CPLEX optimizer, which is free of charge for academics.

and 5 (Appendix A) are obtained as follows:

$$mfx(\theta_0) = H^{-1} \sum_{h=1}^{H} \overline{mfx^{(h)}}(\theta_0^{(h)}).$$

Knowing that the estimator presented in this paper is based on parametric assumptions, we examine its robustness to distributional misspecification. In particular, we consider the spatial lag model with the error terms being drawn from $t$- and Wishart distributions with 10 degrees of freedom. For the $t$-distribution, we draw bivariate random numbers from $t_{10}(0, \Sigma)$ and normalize them to have unit variance. For the Wishart distribution, we draw $2 \times 2$ matrices from the Wishart distribution with 10 degrees of freedom and variance matrix $\Sigma$, take the two diagonal elements and normalize them to have zero mean and unit variance. In both distributions, the resulting correlation parameter is equal to 0.5. It is interesting to compare the performance of PMLE under these two designs with the design where the error terms are obtained from the normal distribution because the $t$-distribution has heavier tails whereas the Wishart distribution generates errors that are asymmetric and not elliptically contoured.

To investigate the finite sample performance of standard errors estimates obtained by the parametric bootstrap method defined in Section 4.3, we compare these estimates with the corresponding standard deviations obtained from the Monte Carlo experiments. Moreover, we perform $z$-tests of the nominal significance level of 5% for the null hypotheses that the respective parameters are equal to their true values and report the empirical sizes of the tests.

The PMLE estimator does not take the entire structure of the variance matrix of the error terms into account. Therefore, PMLE is not efficient. To investigate the efficiency loss, we would like to compare PMLE with the full maximum likelihood estimator (FMLE). Given the structure of our weight matrices, the computation of FMLE would require us to reliably compute $2n$-dimensional integrals. Since this computation is very demanding – perhaps even not feasible at all – we overcome this difficulty by using block diagonal weight matrices with blocks of size 4 and 8 instead of the weight matrix based on distances between counties discussed above. To compute the integrals of the normal probability distribution functions needed for FMLE, we use the Geweke-Hajivassiliou-Keane (GHK) recursive algorithm with 500 draws.

Finally, note that the empirical means, standard deviations, and RMSEs are based on 1000 replications of each experiment. For bootstrapping standard errors, the number of bootstrap samples is chosen to be $B = 100$.

## 5.2 Monte Carlo Results

**Spatial lag model.** First of all, let us discuss the sample selection model with the spatial lag in both the selection and outcome equations. Tables 1 and 2 in Appendix A report biases, standard deviations, and RMSEs of HMLE and PMLE for sample sizes 158, 344, and 760. In general, the results show superiority of the former estimator.

Even though the estimates of $\beta^s$ when $\lambda^s = 0.85$ and $2n = 158$ obtained using both HMLE and PMLE are severely biased, the bias is smaller when the latter estimator is used. Specifically, the bias of HMLE is 19%, 18%, and 39% larger compared to PMLE when $\lambda^o$ is equal to 0, 0.40, and 0.85, respectively. Additionally, when $\lambda^o = 0.85$, the bias of the estimates of $\sigma^2$ obtained by HMLE is on average 103%, 70%, and 78% higher when the sample size is equal to 158, 344, and 760, respectively. Furthermore, in almost all the cases, the standard deviation of HMLE is higher than the standard deviation of PMLE. The difference is especially pronounced when $\lambda^o = 0.85$ with the standard deviations of the HMLE estimates for $\lambda^o$, $\rho$, and $\sigma^2$ being on average 147%, 35%, and 69% higher than the ones obtained using PMLE. Likewise, in terms of RMSE, PMLE outperforms HMLE in the great majority of the cases.

Note that the biases, standard deviations, and RMSEs of $\hat{\beta}_1^s$ obtained using both estimators for the sample size equal to 158 and $\lambda^s = \lambda^o = 0.85$ are very high. This result is driven by one Monte Carlo iteration: Figure 1 in Appendix I reports the estimates of $\beta_1^s$ obtained in all the iterations, whereas Figure 2 in Appendix I reports the same estimates excluding the Monte Carlo iteration prominent in the former figure. After the exclusion of the prominent iteration, the bias, standard deviation, and RMSE of HMLE and PMLE are equal to, respectively, 0.545, 1.045, 1.178 and 0.408, 0.771, 0.872 implying that the qualitative implications about the superiority of PMLE remain.

Further, as the sample size increases, the bias, standard deviation, and RMSE of both estimators decrease. Despite this fact, PMLE still outperforms HMLE even in the largest sample. For the rest of our analysis, we consider only samples with 344 observations.

Regarding the marginal effects, the magnitude of the bias obtained by PMLE is in the great majority of the cases lower than the one of HMLE, whereas the standard deviation and RMSE are always lower. The difference for the bias is especially pronounced for $mfx_3$ and $mfx_4$ when $\lambda^o = 0.85$, whereas the difference for the standard deviation and RMSE is especially noticeable for $mfx_1$ and $mfx_2$ when $\lambda^s = 0.85$ and for $mfx_3$ and $mfx_4$ when $\lambda^o = 0.85$. This result is consistent with our findings above about the parameter estimates. Note that both $mfx_1$ and $mfx_2$ depend only on the parameters of the selection equation and that the relative difference in the performance of HMLE and PMLE in the estimation of the selection equation parameters is especially pronounced when $\lambda^s = 0.85$. Furthermore, both $mfx_3$ and $mfx_4$ depend on $\lambda^o$, which is noticeably better estimated by PMLE than HMLE.

**Spatial error model.** Next, we discuss the spatial error sample selection model (Tables 4–5 in Appendix A). The HMLE estimator in the spatial error case performs much worse compared to the spatial lag case: there are large biases in the estimates of $\beta^s$, $\lambda^s$, and $\lambda^o$ when the spatial parameters in the respective equations are not equal to zero, although severe biases are not present in the estimates of the correlation coefficient. The estimates of $\beta^s$ and $\lambda^s$ obtained by GMM and GMM2 are in most of the cases severely biased. These results are consistent with the simulation results in Flores-Lagunes and Schnier (2012). Finally, there are biases in the estimates of $\beta^s$ obtained by PMLE, which increase with the magnitude of $\lambda^s$, whereas the other parameters are estimated well.

In most of the cases, PMLE outperforms HMLE with respect to both standard deviation and RMSE. Although the estimates of $\lambda^o$ (and in some cases $\sigma^2$) obtained by the GMM estimators have lower standard deviations and RMSEs than the ones obtained by PMLE when $\lambda^o \in \{0, 0.40\}$, the rest of the parameters are estimated noticeably better by PMLE.

Table 5 in Appendix A shows that the bias in the parameter estimates does not have a substantial influence on the marginal effects: for all the estimators, the bias is very small. The PMLE estimator, however, outperforms the remaining estimators in terms of both standard deviation and RMSE. When $\lambda^o = 0.85$, the standard deviation (and RMSE) of the estimates of $mfx_3$ and $mfx_4$ obtained by HMLE is on average 15% and 22% higher compared to PMLE. The superiority of PMLE compared to the GMM estimator is especially pronounced for $mfx_3$ and $mfx_4$ when $\lambda^s = 0.85$, where the standard deviation (and RMSE) obtained by the GMM estimator is, respectively, on average 11 and 4 times higher than the ones obtained by PMLE. Although the marginal effects are estimated better when the GMM2 estimator is employed, the standard deviation (RMSE) obtained by the GMM2 estimator is on average 55% (56%), 61% (61%), 21% (24%), and 12% (12%) higher than the ones obtained by PMLE for, respectively, $mfx_1$, $mfx_2$, $mfx_3$, and $mfx_4$.

**Distributional misspecification.** Tables 6 and 7 in Appendix A report the performance of PMLE and HMLE under distributional misspecification in the context of the spatial lag model. Similarly to the case of the normally distributed errors, PMLE outperforms HMLE in terms of bias when the errors are obtained from the $t$-distribution, whereas the performance of the two estimators under the Weibull distribution is similar. The standard deviations and RMSEs of HMLE are almost always higher than the ones obtained by PMLE in all the three cases. The largest relative advantage is observed for the estimate of $\lambda^o$ when $\lambda^o = 0.85$, where the standard deviation (RMSE) obtained by HMLE is on average 2.8 (2.7), 1.8 (1.8), and 2.2 (2.1) higher than the ones obtained by HMLE when the errors are drawn from the normal, $t$-, and Weibull distributions, respectively.

**Estimates of standard errors.** The performance of the parametric bootstrap is investigated in the context of the spatial lag model. First, the standard errors (SEs) obtained using the parametric bootstrap are very close to the corresponding standard deviations obtained from the Monte Carlo simulation, see Table 8 in Appendix A. Next, Table 9 in Appendix A reports the empirical sizes of the $z$-tests for the null hypotheses that the parameters are equal to the true values. Although the empirical sizes of most of the tests are close to the nominal 5% level, there are a few exceptions. The empirical size of the test that $\rho = 0.5$ is on average 4.7pp higher than the nominal 5% level, and the empirical size of the test that $\sigma^2 = 1$ is on average 4.9pp higher when $\lambda^o = 0.85$. Nevertheless, the Monte Carlo study shows that the parametric bootstrap works reasonably well in finite samples.

**Comparison with FMLE.** Tables 10 and 11 report the performance of PMLE and FMLE in the context of the spatial lag model with a block diagonal weight matrix. In terms of bias, PMLE and FMLE perform similarly except for the estimates of $\beta^s$ when $\lambda^s = 0.85$ where FMLE clearly outperforms PMLE. In almost all the cases, the standard deviation and RMSE of PMLE are higher compared with FMLE. The difference is the largest when at least one of the spatial parameters is equal to 0.85. For example, when $\lambda^s = 0.85$ and block size is 4, the standard deviation obtained by PMLE of $\hat{\beta}^s_n$, $\hat{\lambda}^s_n$, and $\hat{\rho}_n$ is 6.7 – 13.4% higher than the one obtained by FMLE. When $\lambda^o = 0.85$ and block size is 4, the standard deviation obtained by PMLE of $\hat{\beta}^o_n$, $\hat{\lambda}^o_n$, $\hat{\rho}_n$, and $\hat{\sigma}^2_n$ is 6.7 – 27.7% higher than the one obtained by FMLE. The disadvantage of PMLE relative to FMLE is more pronounced when block size is equal to 8: the respective ratios are equal to 12.0 – 31.3% and 9.4 – 53.6%. The results in terms of RMSE are similar.

Hence, implementing PMLE instead of FMLE can lead to quite substantial efficiency losses, especially when weight matrices are composed of larger blocks. However, FMLE is often unfeasible when more realistic weight matrices are considered, for instance, the one in our main Monte Carlo design. In these cases, PMLE seems to be a good alternative.

# 6 Conclusion

This paper examines the sample selection model with a spatial lag of a latent dependent variable or a spatial error in both the selection and outcome equations. We propose to estimate this model by the partial maximum likelihood estimator which is based on the idea that all observations are divided into pairs in such a way that dependence within a pair is more important than dependence between pairs; the likelihood function is constructed as a product of marginal likelihood contributions for these pairs. Since the likelihood function does not capture the dependence between pairs, complexity is reduced and the model can be easily estimated. Using the limit laws for the NED random fields, we establish consistency and asymptotic normality of the PMLE. Our simulation study shows that the proposed estimator performs quite well in small samples, and in most cases, outperforms HMLE and the GMM estimator proposed by Flores-Lagunes and Schnier (2012). Moreover, PMLE and the developed asymptotic theory can be easily applied to other limited dependent variable models, that is, probit and Tobit models, because the sample selection model has all the components of the former models and they are thus special cases of the sample selection model.

The studied model can be extended in several ways. The asymptotic distribution of the proposed estimator depends on the way how the observations are divided into groups. It is desirable to find an optimal grouping scheme based on some criterion, for example, such that the sum of variances of parameters of interest is minimized. Given the complexity of the variance matrix of PMLE, this is a very difficult task. Nevertheless, as our simulation shows, PMLE performs quite well even with a non-optimal grouping.

# Appendix A Results of the Monte Carlo Experiments

**Table 1:** Biases and standard deviations of parameter estimates in the context of the sample selection model with a spatial lag in both the selection and outcome equations.

| | | | bias | | | | | | s.d. | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\lambda^s$ | $\lambda^o$ | | $2n = 158$ | | $2n = 344$ | | $2n = 760$ | | $2n = 158$ | | $2n = 344$ | | $2n = 760$ | |
| | | | HMLE | PMLE | HMLE | PMLE | HMLE | PMLE | HMLE | PMLE | HMLE | PMLE | HMLE | PMLE |
| 0.00 | 0.00 | $\hat{\beta}^s$ | 0.123 | 0.123 | 0.044 | 0.043 | 0.019 | 0.019 | 0.721 | 0.719 | 0.449 | 0.446 | 0.281 | 0.281 |
| | | $\hat{\beta}^o$ | 0.001 | 0.000 | 0.006 | 0.005 | 0.002 | 0.002 | 0.411 | 0.412 | 0.255 | 0.255 | 0.165 | 0.165 |
| | | $\hat{\lambda}^s$ | -0.010 | -0.012 | -0.012 | -0.014 | -0.006 | -0.006 | 0.200 | 0.190 | 0.147 | 0.141 | 0.084 | 0.080 |
| | | $\hat{\lambda}^o$ | -0.009 | -0.010 | -0.002 | -0.004 | -0.005 | -0.005 | 0.140 | 0.125 | 0.099 | 0.092 | 0.060 | 0.055 |
| | | $\hat{\rho}$ | -0.035 | -0.040 | -0.004 | -0.005 | 0.001 | 0.000 | 0.367 | 0.370 | 0.204 | 0.204 | 0.128 | 0.128 |
| | | $\hat{\sigma}^2$ | -0.020 | -0.017 | -0.006 | -0.005 | -0.004 | -0.003 | 0.174 | 0.175 | 0.111 | 0.111 | 0.072 | 0.072 |
| 0.00 | 0.40 | $\hat{\beta}^s$ | 0.080 | 0.082 | 0.041 | 0.040 | 0.024 | 0.024 | 0.736 | 0.734 | 0.442 | 0.441 | 0.283 | 0.281 |
| | | $\hat{\beta}^o$ | 0.008 | 0.006 | 0.002 | 0.001 | 0.000 | 0.001 | 0.387 | 0.376 | 0.234 | 0.230 | 0.152 | 0.148 |
| | | $\hat{\lambda}^s$ | -0.004 | -0.008 | -0.012 | -0.012 | -0.004 | -0.005 | 0.200 | 0.187 | 0.141 | 0.135 | 0.079 | 0.076 |
| | | $\hat{\lambda}^o$ | -0.009 | -0.011 | -0.008 | -0.008 | -0.001 | -0.002 | 0.109 | 0.097 | 0.078 | 0.070 | 0.046 | 0.042 |
| | | $\hat{\rho}$ | -0.025 | -0.023 | -0.014 | -0.011 | 0.002 | 0.002 | 0.373 | 0.358 | 0.224 | 0.217 | 0.135 | 0.131 |
| | | $\hat{\sigma}^2$ | -0.020 | -0.012 | -0.012 | -0.008 | -0.007 | -0.005 | 0.194 | 0.183 | 0.123 | 0.119 | 0.080 | 0.076 |
| 0.00 | 0.85 | $\hat{\beta}^s$ | 0.093 | 0.098 | 0.045 | 0.044 | 0.018 | 0.019 | 0.709 | 0.700 | 0.459 | 0.452 | 0.289 | 0.285 |
| | | $\hat{\beta}^o$ | 0.021 | 0.008 | 0.009 | 0.005 | 0.003 | 0.005 | 0.517 | 0.428 | 0.368 | 0.284 | 0.234 | 0.190 |
| | | $\hat{\lambda}^s$ | -0.007 | -0.014 | -0.004 | -0.003 | 0.001 | 0.000 | 0.208 | 0.191 | 0.142 | 0.131 | 0.084 | 0.080 |
| | | $\hat{\lambda}^o$ | -0.005 | -0.012 | -0.004 | -0.005 | -0.003 | -0.003 | 0.081 | 0.044 | 0.086 | 0.030 | 0.061 | 0.019 |
| | | $\hat{\rho}$ | -0.037 | -0.032 | -0.012 | -0.011 | 0.001 | 0.003 | 0.482 | 0.372 | 0.305 | 0.230 | 0.190 | 0.147 |
| | | $\hat{\sigma}^2$ | -0.182 | -0.101 | -0.109 | -0.064 | -0.040 | -0.024 | 0.411 | 0.260 | 0.280 | 0.174 | 0.238 | 0.128 |
| 0.40 | 0.00 | $\hat{\beta}^s$ | 0.101 | 0.099 | 0.048 | 0.047 | 0.016 | 0.016 | 0.719 | 0.712 | 0.413 | 0.411 | 0.284 | 0.283 |
| | | $\hat{\beta}^o$ | 0.004 | 0.005 | 0.003 | 0.003 | 0.001 | 0.001 | 0.394 | 0.397 | 0.252 | 0.252 | 0.162 | 0.162 |
| | | $\hat{\lambda}^s$ | -0.011 | -0.013 | -0.008 | -0.010 | -0.002 | -0.002 | 0.146 | 0.138 | 0.102 | 0.097 | 0.065 | 0.063 |
| | | $\hat{\lambda}^o$ | -0.002 | -0.006 | -0.007 | -0.007 | -0.002 | -0.002 | 0.134 | 0.126 | 0.095 | 0.089 | 0.058 | 0.053 |
| | | $\hat{\rho}$ | -0.027 | -0.024 | -0.002 | -0.001 | 0.001 | 0.001 | 0.357 | 0.355 | 0.209 | 0.208 | 0.133 | 0.132 |
| | | $\hat{\sigma}^2$ | -0.023 | -0.020 | -0.012 | -0.010 | -0.005 | -0.004 | 0.158 | 0.157 | 0.105 | 0.104 | 0.068 | 0.068 |
| 0.40 | 0.40 | $\hat{\beta}^s$ | 0.097 | 0.099 | 0.034 | 0.034 | 0.024 | 0.023 | 0.713 | 0.714 | 0.426 | 0.424 | 0.273 | 0.270 |
| | | $\hat{\beta}^o$ | 0.004 | 0.002 | 0.004 | 0.006 | 0.001 | 0.001 | 0.364 | 0.352 | 0.237 | 0.231 | 0.158 | 0.155 |
| | | $\hat{\lambda}^s$ | -0.007 | -0.012 | -0.004 | -0.005 | -0.004 | -0.004 | 0.151 | 0.146 | 0.099 | 0.092 | 0.062 | 0.059 |
| | | $\hat{\lambda}^o$ | -0.003 | -0.007 | -0.003 | -0.004 | -0.003 | -0.003 | 0.108 | 0.097 | 0.074 | 0.068 | 0.045 | 0.042 |
| | | $\hat{\rho}$ | -0.026 | -0.019 | -0.004 | -0.002 | -0.002 | -0.001 | 0.372 | 0.355 | 0.204 | 0.197 | 0.133 | 0.129 |
| | | $\hat{\sigma}^2$ | -0.039 | -0.030 | -0.018 | -0.014 | -0.010 | -0.008 | 0.184 | 0.173 | 0.123 | 0.116 | 0.081 | 0.077 |
| 0.40 | 0.85 | $\hat{\beta}^s$ | 0.100 | 0.101 | 0.036 | 0.038 | 0.023 | 0.023 | 0.702 | 0.695 | 0.422 | 0.415 | 0.279 | 0.277 |
| | | $\hat{\beta}^o$ | 0.007 | 0.023 | 0.012 | 0.004 | 0.009 | 0.000 | 0.543 | 0.415 | 0.357 | 0.282 | 0.237 | 0.191 |
| | | $\hat{\lambda}^s$ | 0.004 | -0.005 | -0.000 | -0.004 | -0.001 | -0.003 | 0.157 | 0.145 | 0.105 | 0.097 | 0.064 | 0.060 |
| | | $\hat{\lambda}^o$ | -0.006 | -0.013 | -0.005 | -0.005 | 0.001 | -0.002 | 0.080 | 0.041 | 0.101 | 0.029 | 0.022 | 0.018 |
| | | $\hat{\rho}$ | -0.040 | 0.009 | -0.020 | -0.001 | -0.020 | -0.003 | 0.453 | 0.337 | 0.292 | 0.207 | 0.183 | 0.134 |
| | | $\hat{\sigma}^2$ | -0.161 | -0.076 | -0.101 | -0.061 | -0.060 | -0.032 | 0.395 | 0.252 | 0.327 | 0.171 | 0.174 | 0.116 |
| 0.85 | 0.00 | $\hat{\beta}^s$ | 0.384 | 0.322 | 0.125 | 0.111 | 0.049 | 0.045 | 1.797 | 1.570 | 0.693 | 0.641 | 0.407 | 0.389 |
| | | $\hat{\beta}^o$ | 0.014 | 0.013 | 0.003 | 0.002 | 0.000 | 0.000 | 0.376 | 0.376 | 0.229 | 0.227 | 0.154 | 0.152 |
| | | $\hat{\lambda}^s$ | -0.004 | -0.008 | -0.002 | -0.004 | -0.001 | -0.001 | 0.078 | 0.068 | 0.053 | 0.047 | 0.032 | 0.029 |
| | | $\hat{\lambda}^o$ | -0.005 | -0.009 | -0.005 | -0.005 | -0.000 | -0.001 | 0.152 | 0.137 | 0.106 | 0.097 | 0.060 | 0.055 |
| | | $\hat{\rho}$ | 0.017 | 0.011 | 0.037 | 0.027 | 0.004 | 0.003 | 0.423 | 0.402 | 0.242 | 0.226 | 0.155 | 0.147 |
| | | $\hat{\sigma}^2$ | -0.037 | -0.033 | -0.012 | -0.011 | -0.008 | -0.007 | 0.153 | 0.153 | 0.099 | 0.100 | 0.068 | 0.068 |
| 0.85 | 0.40 | $\hat{\beta}^s$ | 0.356 | 0.303 | 0.121 | 0.108 | 0.047 | 0.044 | 1.938 | 1.463 | 0.697 | 0.646 | 0.410 | 0.387 |
| | | $\hat{\beta}^o$ | 0.002 | 0.004 | 0.001 | 0.000 | 0.001 | 0.001 | 0.352 | 0.340 | 0.238 | 0.229 | 0.154 | 0.149 |
| | | $\hat{\lambda}^s$ | -0.010 | -0.012 | -0.002 | -0.004 | 0.001 | 0.001 | 0.098 | 0.088 | 0.051 | 0.045 | 0.032 | 0.029 |
| | | $\hat{\lambda}^o$ | -0.008 | -0.010 | -0.006 | -0.005 | -0.004 | -0.004 | 0.114 | 0.098 | 0.081 | 0.072 | 0.048 | 0.042 |
| | | $\hat{\rho}$ | -0.009 | -0.010 | 0.009 | 0.006 | 0.001 | -0.001 | 0.434 | 0.398 | 0.252 | 0.231 | 0.161 | 0.146 |
| | | $\hat{\sigma}^2$ | -0.044 | -0.035 | -0.021 | -0.019 | -0.009 | -0.009 | 0.178 | 0.166 | 0.116 | 0.109 | 0.078 | 0.072 |
| 0.85 | 0.85 | $\hat{\beta}^s$ | 14.816 | 10.665 | 0.117 | 0.102 | 0.050 | 0.046 | 1367.749 | 1357.736 | 0.706 | 0.625 | 0.413 | 0.389 |
| | | $\hat{\beta}^o$ | 0.027 | 0.009 | 0.013 | 0.004 | 0.002 | 0.005 | 0.551 | 0.411 | 0.418 | 0.273 | 0.241 | 0.185 |
| | | $\hat{\lambda}^s$ | 0.003 | -0.007 | 0.002 | -0.003 | 0.001 | -0.000 | 0.074 | 0.068 | 0.049 | 0.044 | 0.030 | 0.027 |
| | | $\hat{\lambda}^o$ | 0.003 | -0.009 | -0.012 | -0.007 | -0.001 | -0.003 | 0.053 | 0.039 | 0.140 | 0.028 | 0.023 | 0.018 |
| | | $\hat{\rho}$ | -0.102 | -0.004 | -0.075 | -0.014 | -0.031 | -0.007 | 0.510 | 0.375 | 0.322 | 0.232 | 0.214 | 0.151 |
| | | $\hat{\sigma}^2$ | -0.222 | -0.102 | -0.111 | -0.065 | -0.049 | -0.027 | 0.364 | 0.230 | 0.323 | 0.155 | 0.179 | 0.115 |

**Table 2:** RMSEs of parameter estimates in the context of the sample selection model with a spatial lag in both the selection and outcome equations.

| $\lambda^s$ | $\lambda^o$ | | $2n = 158$ | | $2n = 344$ | | $2n = 760$ | |
|---|---|---|---|---|---|---|---|---|
| | | | HMLE | PMLE | HMLE | PMLE | HMLE | PMLE |
| 0.00 | 0.00 | $\hat{\beta}^s$ | 0.764 | 0.761 | 0.459 | 0.456 | 0.286 | 0.285 |
| | | $\hat{\beta}^o$ | 0.411 | 0.412 | 0.255 | 0.255 | 0.165 | 0.165 |
| | | $\hat{\lambda}^s$ | 0.200 | 0.190 | 0.147 | 0.141 | 0.085 | 0.080 |
| | | $\hat{\lambda}^o$ | 0.140 | 0.126 | 0.099 | 0.092 | 0.060 | 0.055 |
| | | $\hat{\rho}$ | 0.369 | 0.373 | 0.204 | 0.204 | 0.128 | 0.128 |
| | | $\hat{\sigma}^2$ | 0.176 | 0.176 | 0.112 | 0.111 | 0.072 | 0.072 |
| 0.00 | 0.40 | $\hat{\beta}^s$ | 0.764 | 0.762 | 0.451 | 0.449 | 0.288 | 0.286 |
| | | $\hat{\beta}^o$ | 0.387 | 0.376 | 0.234 | 0.230 | 0.152 | 0.148 |
| | | $\hat{\lambda}^s$ | 0.200 | 0.187 | 0.141 | 0.135 | 0.079 | 0.076 |
| | | $\hat{\lambda}^o$ | 0.109 | 0.098 | 0.078 | 0.071 | 0.046 | 0.042 |
| | | $\hat{\rho}$ | 0.374 | 0.359 | 0.224 | 0.217 | 0.135 | 0.131 |
| | | $\hat{\sigma}^2$ | 0.195 | 0.184 | 0.123 | 0.119 | 0.080 | 0.076 |
| 0.00 | 0.85 | $\hat{\beta}^s$ | 0.744 | 0.740 | 0.469 | 0.462 | 0.292 | 0.288 |
| | | $\hat{\beta}^o$ | 0.521 | 0.429 | 0.370 | 0.284 | 0.234 | 0.190 |
| | | $\hat{\lambda}^s$ | 0.209 | 0.192 | 0.142 | 0.131 | 0.084 | 0.080 |
| | | $\hat{\lambda}^o$ | 0.081 | 0.045 | 0.086 | 0.031 | 0.061 | 0.019 |
| | | $\hat{\rho}$ | 0.484 | 0.374 | 0.305 | 0.230 | 0.190 | 0.147 |
| | | $\hat{\sigma}^2$ | 0.450 | 0.278 | 0.300 | 0.185 | 0.241 | 0.131 |
| 0.40 | 0.00 | $\hat{\beta}^s$ | 0.764 | 0.755 | 0.427 | 0.425 | 0.287 | 0.286 |
| | | $\hat{\beta}^o$ | 0.394 | 0.397 | 0.252 | 0.252 | 0.162 | 0.162 |
| | | $\hat{\lambda}^s$ | 0.147 | 0.139 | 0.102 | 0.098 | 0.065 | 0.063 |
| | | $\hat{\lambda}^o$ | 0.134 | 0.127 | 0.095 | 0.089 | 0.058 | 0.053 |
| | | $\hat{\rho}$ | 0.358 | 0.355 | 0.209 | 0.208 | 0.133 | 0.132 |
| | | $\hat{\sigma}^2$ | 0.159 | 0.158 | 0.105 | 0.105 | 0.068 | 0.068 |
| 0.40 | 0.40 | $\hat{\beta}^s$ | 0.754 | 0.755 | 0.436 | 0.433 | 0.278 | 0.275 |
| | | $\hat{\beta}^o$ | 0.365 | 0.352 | 0.238 | 0.231 | 0.158 | 0.155 |
| | | $\hat{\lambda}^s$ | 0.151 | 0.146 | 0.099 | 0.092 | 0.062 | 0.059 |
| | | $\hat{\lambda}^o$ | 0.108 | 0.097 | 0.074 | 0.068 | 0.045 | 0.042 |
| | | $\hat{\rho}$ | 0.373 | 0.355 | 0.204 | 0.197 | 0.133 | 0.129 |
| | | $\hat{\sigma}^2$ | 0.188 | 0.176 | 0.124 | 0.117 | 0.081 | 0.077 |
| 0.40 | 0.85 | $\hat{\beta}^s$ | 0.744 | 0.737 | 0.436 | 0.431 | 0.283 | 0.281 |
| | | $\hat{\beta}^o$ | 0.545 | 0.417 | 0.358 | 0.283 | 0.238 | 0.191 |
| | | $\hat{\lambda}^s$ | 0.157 | 0.145 | 0.105 | 0.097 | 0.064 | 0.060 |
| | | $\hat{\lambda}^o$ | 0.081 | 0.043 | 0.101 | 0.029 | 0.022 | 0.018 |
| | | $\hat{\rho}$ | 0.455 | 0.337 | 0.292 | 0.207 | 0.184 | 0.134 |
| | | $\hat{\sigma}^2$ | 0.427 | 0.263 | 0.342 | 0.181 | 0.184 | 0.120 |
| 0.85 | 0.00 | $\hat{\beta}^s$ | 2.090 | 1.800 | 0.788 | 0.723 | 0.432 | 0.412 |
| | | $\hat{\beta}^o$ | 0.376 | 0.376 | 0.229 | 0.227 | 0.154 | 0.152 |
| | | $\hat{\lambda}^s$ | 0.078 | 0.068 | 0.053 | 0.047 | 0.032 | 0.029 |
| | | $\hat{\lambda}^o$ | 0.152 | 0.137 | 0.106 | 0.097 | 0.060 | 0.055 |
| | | $\hat{\rho}$ | 0.424 | 0.402 | 0.244 | 0.227 | 0.155 | 0.147 |
| | | $\hat{\sigma}^2$ | 0.158 | 0.156 | 0.100 | 0.100 | 0.068 | 0.068 |
| 0.85 | 0.40 | $\hat{\beta}^s$ | 2.184 | 1.681 | 0.783 | 0.721 | 0.434 | 0.409 |
| | | $\hat{\beta}^o$ | 0.353 | 0.340 | 0.238 | 0.230 | 0.154 | 0.149 |
| | | $\hat{\lambda}^s$ | 0.098 | 0.089 | 0.051 | 0.045 | 0.032 | 0.029 |
| | | $\hat{\lambda}^o$ | 0.114 | 0.099 | 0.081 | 0.072 | 0.048 | 0.042 |
| | | $\hat{\rho}$ | 0.434 | 0.398 | 0.252 | 0.231 | 0.161 | 0.146 |
| | | $\hat{\sigma}^2$ | 0.183 | 0.169 | 0.117 | 0.111 | 0.078 | 0.073 |
| 0.85 | 0.85 | $\hat{\beta}^s$ | 1368.463 | 1358.438 | 0.788 | 0.692 | 0.433 | 0.407 |
| | | $\hat{\beta}^o$ | 0.555 | 0.411 | 0.420 | 0.274 | 0.241 | 0.185 |
| | | $\hat{\lambda}^s$ | 0.074 | 0.068 | 0.050 | 0.044 | 0.030 | 0.027 |
| | | $\hat{\lambda}^o$ | 0.053 | 0.040 | 0.141 | 0.029 | 0.023 | 0.019 |
| | | $\hat{\rho}$ | 0.520 | 0.375 | 0.331 | 0.232 | 0.216 | 0.151 |
| | | $\hat{\sigma}^2$ | 0.426 | 0.252 | 0.342 | 0.168 | 0.186 | 0.118 |

**Table 3:** Biases, standard deviations, and RMSEs of total marginal effects estimates in the context of the sample selection model with a spatial lag in both the selection and outcome equations ($2n = 344$).

| $\lambda^s$ | $\lambda^o$ | | $mfx(\theta_0)$ | HMLE | | | PMLE | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | bias | s.d. | rmse | bias | s.d. | rmse |
| 0.00 | 0.00 | $mfx_1$ | 0.193 | 0.002 | 0.032 | 0.032 | 0.001 | 0.031 | 0.031 |
| | | $mfx_2$ | -0.193 | -0.001 | 0.030 | 0.030 | -0.000 | 0.029 | 0.029 |
| | | $mfx_3$ | 0.782 | -0.002 | 0.131 | 0.131 | -0.004 | 0.126 | 0.126 |
| | | $mfx_4$ | -1.000 | -0.006 | 0.112 | 0.112 | -0.003 | 0.104 | 0.104 |
| 0.00 | 0.40 | $mfx_1$ | 0.193 | 0.002 | 0.032 | 0.032 | 0.001 | 0.031 | 0.031 |
| | | $mfx_2$ | -0.193 | -0.001 | 0.031 | 0.031 | -0.001 | 0.029 | 0.029 |
| | | $mfx_3$ | 1.433 | -0.000 | 0.229 | 0.229 | -0.005 | 0.210 | 0.210 |
| | | $mfx_4$ | -1.657 | -0.006 | 0.209 | 0.209 | -0.001 | 0.189 | 0.189 |
| 0.00 | 0.85 | $mfx_1$ | 0.193 | 0.004 | 0.033 | 0.033 | 0.004 | 0.031 | 0.031 |
| | | $mfx_2$ | -0.193 | -0.002 | 0.032 | 0.032 | -0.002 | 0.030 | 0.030 |
| | | $mfx_3$ | 6.306 | 0.241 | 1.771 | 1.787 | 0.009 | 1.345 | 1.344 |
| | | $mfx_4$ | -6.584 | -0.295 | 1.666 | 1.691 | -0.030 | 1.236 | 1.236 |
| 0.40 | 0.00 | $mfx_1$ | 0.303 | 0.005 | 0.054 | 0.054 | 0.004 | 0.052 | 0.052 |
| | | $mfx_2$ | -0.303 | -0.002 | 0.048 | 0.048 | -0.001 | 0.047 | 0.047 |
| | | $mfx_3$ | 0.663 | -0.010 | 0.162 | 0.162 | -0.009 | 0.158 | 0.158 |
| | | $mfx_4$ | -1.000 | 0.000 | 0.101 | 0.101 | 0.001 | 0.096 | 0.096 |
| 0.40 | 0.40 | $mfx_1$ | 0.303 | 0.003 | 0.052 | 0.052 | 0.002 | 0.049 | 0.049 |
| | | $mfx_2$ | -0.303 | -0.004 | 0.047 | 0.048 | -0.003 | 0.045 | 0.045 |
| | | $mfx_3$ | 1.298 | 0.011 | 0.236 | 0.236 | 0.006 | 0.221 | 0.221 |
| | | $mfx_4$ | -1.657 | -0.014 | 0.202 | 0.202 | -0.008 | 0.188 | 0.188 |
| 0.40 | 0.85 | $mfx_1$ | 0.303 | 0.005 | 0.054 | 0.054 | 0.003 | 0.050 | 0.050 |
| | | $mfx_2$ | -0.303 | -0.007 | 0.049 | 0.049 | -0.005 | 0.046 | 0.046 |
| | | $mfx_3$ | 6.093 | 0.232 | 1.666 | 1.681 | 0.016 | 1.291 | 1.290 |
| | | $mfx_4$ | -6.584 | -0.287 | 1.604 | 1.629 | -0.035 | 1.179 | 1.179 |
| 0.85 | 0.00 | $mfx_1$ | 0.757 | 0.052 | 0.282 | 0.287 | 0.028 | 0.231 | 0.232 |
| | | $mfx_2$ | -0.757 | -0.056 | 0.273 | 0.279 | -0.034 | 0.204 | 0.207 |
| | | $mfx_3$ | 0.304 | -0.130 | 0.469 | 0.486 | -0.093 | 0.404 | 0.415 |
| | | $mfx_4$ | -1.000 | -0.005 | 0.117 | 0.117 | -0.004 | 0.108 | 0.108 |
| 0.85 | 0.40 | $mfx_1$ | 0.757 | 0.057 | 0.278 | 0.283 | 0.033 | 0.221 | 0.223 |
| | | $mfx_2$ | -0.757 | -0.060 | 0.260 | 0.267 | -0.036 | 0.196 | 0.199 |
| | | $mfx_3$ | 0.849 | -0.105 | 0.543 | 0.552 | -0.078 | 0.471 | 0.477 |
| | | $mfx_4$ | -1.657 | -0.012 | 0.226 | 0.226 | -0.009 | 0.208 | 0.208 |
| 0.85 | 0.85 | $mfx_1$ | 0.757 | 0.061 | 0.312 | 0.318 | 0.027 | 0.210 | 0.212 |
| | | $mfx_2$ | -0.757 | -0.070 | 0.363 | 0.370 | -0.032 | 0.218 | 0.220 |
| | | $mfx_3$ | 5.196 | 0.299 | 1.829 | 1.852 | -0.059 | 1.298 | 1.298 |
| | | $mfx_4$ | -6.584 | -0.222 | 1.619 | 1.633 | 0.017 | 1.183 | 1.182 |

**Table 4:** Biases, standard deviations, and RMSEs of parameter estimates in the context of the sample selection model with a spatial error in both the selection and outcome equations ($2n = 344$).

| $\lambda^s$ | $\lambda^o$ | | HMLE | | | GMM | | | GMM2 | | | PMLE | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | bias | s.d. | rmse | bias | s.d. | rmse | bias | s.d. | rmse | bias | s.d. | rmse |
| 0.00 | 0.00 | $\hat{\beta}^s$ | 0.286 | 2.617 | 2.698 | 0.824 | 3.580 | 4.044 | 0.728 | 3.234 | 3.649 | 0.100 | 0.511 | 0.559 |
| | | $\hat{\beta}^o$ | 0.006 | 0.254 | 0.254 | 0.001 | 0.284 | 0.285 | 0.003 | 0.271 | 0.272 | 0.005 | 0.254 | 0.254 |
| | | $\hat{\lambda}^s$ | 0.029 | 0.418 | 0.419 | -0.038 | 0.593 | 0.594 | -0.079 | 0.592 | 0.597 | -0.041 | 0.396 | 0.398 |
| | | $\hat{\lambda}^o$ | 0.006 | 0.323 | 0.323 | 0.023 | 0.113 | 0.115 | 0.027 | 0.113 | 0.116 | -0.020 | 0.241 | 0.242 |
| | | $\hat{\rho}$ | 0.026 | 0.218 | 0.219 | 0.040 | 0.320 | 0.323 | 0.010 | 0.269 | 0.269 | 0.004 | 0.202 | 0.202 |
| | | $\hat{\sigma}^2$ | -0.053 | 0.137 | 0.147 | -0.082 | 0.090 | 0.122 | -0.019 | 0.100 | 0.102 | -0.024 | 0.112 | 0.114 |
| 0.00 | 0.40 | $\hat{\beta}^s$ | 0.299 | 2.854 | 2.939 | 0.749 | 2.960 | 3.419 | 0.787 | 3.223 | 3.707 | 0.108 | 1.108 | 1.138 |
| | | $\hat{\beta}^o$ | 0.000 | 0.279 | 0.280 | 0.009 | 0.303 | 0.304 | 0.011 | 0.299 | 0.301 | 0.001 | 0.276 | 0.276 |
| | | $\hat{\lambda}^s$ | 0.038 | 0.434 | 0.436 | -0.151 | 0.641 | 0.659 | -0.091 | 0.655 | 0.661 | -0.039 | 0.395 | 0.397 |
| | | $\hat{\lambda}^o$ | -0.366 | 0.417 | 0.555 | -0.075 | 0.094 | 0.120 | -0.068 | 0.096 | 0.118 | -0.045 | 0.192 | 0.198 |
| | | $\hat{\rho}$ | 0.016 | 0.229 | 0.229 | 0.023 | 0.307 | 0.308 | -0.005 | 0.273 | 0.273 | -0.001 | 0.208 | 0.208 |
| | | $\hat{\sigma}^2$ | -0.014 | 0.170 | 0.171 | -0.058 | 0.094 | 0.110 | 0.002 | 0.102 | 0.102 | -0.015 | 0.132 | 0.132 |
| 0.00 | 0.85 | $\hat{\beta}^s$ | 0.547 | 4.552 | 4.726 | 0.836 | 3.303 | 3.806 | 0.920 | 3.630 | 4.196 | 0.088 | 0.491 | 0.535 |
| | | $\hat{\beta}^o$ | 0.020 | 0.623 | 0.623 | 0.050 | 0.622 | 0.635 | 0.061 | 0.628 | 0.638 | 0.020 | 0.555 | 0.555 |
| | | $\hat{\lambda}^s$ | 0.068 | 0.491 | 0.496 | -0.253 | 0.684 | 0.729 | -0.206 | 0.706 | 0.735 | -0.048 | 0.382 | 0.385 |
| | | $\hat{\lambda}^o$ | -0.484 | 0.733 | 0.879 | -0.089 | 0.054 | 0.104 | -0.087 | 0.058 | 0.104 | -0.017 | 0.046 | 0.049 |
| | | $\hat{\rho}$ | 0.014 | 0.339 | 0.339 | -0.129 | 0.321 | 0.346 | -0.142 | 0.296 | 0.328 | 0.003 | 0.241 | 0.241 |
| | | $\hat{\sigma}^2$ | 0.206 | 0.677 | 0.707 | 0.156 | 0.177 | 0.236 | 0.216 | 0.199 | 0.294 | -0.010 | 0.180 | 0.180 |
| 0.40 | 0.00 | $\hat{\beta}^s$ | 0.264 | 2.847 | 2.913 | 0.767 | 3.550 | 3.969 | 0.733 | 3.516 | 3.926 | 0.084 | 0.626 | 0.660 |
| | | $\hat{\beta}^o$ | 0.002 | 0.246 | 0.247 | 0.007 | 0.272 | 0.273 | 0.011 | 0.267 | 0.269 | 0.002 | 0.248 | 0.248 |
| | | $\hat{\lambda}^s$ | -0.394 | 0.474 | 0.616 | -0.441 | 0.621 | 0.762 | -0.434 | 0.623 | 0.759 | -0.103 | 0.357 | 0.372 |
| | | $\hat{\lambda}^o$ | 0.018 | 0.308 | 0.308 | 0.005 | 0.115 | 0.115 | 0.008 | 0.116 | 0.117 | -0.008 | 0.233 | 0.233 |
| | | $\hat{\rho}$ | 0.017 | 0.229 | 0.230 | 0.048 | 0.330 | 0.334 | 0.005 | 0.289 | 0.289 | -0.002 | 0.211 | 0.211 |
| | | $\hat{\sigma}^2$ | -0.051 | 0.135 | 0.144 | -0.084 | 0.085 | 0.119 | -0.021 | 0.094 | 0.096 | -0.025 | 0.113 | 0.115 |
| 0.40 | 0.40 | $\hat{\beta}^s$ | 0.322 | 2.968 | 3.058 | 0.739 | 3.219 | 3.633 | 0.876 | 3.733 | 4.244 | 0.100 | 0.617 | 0.659 |
| | | $\hat{\beta}^o$ | 0.006 | 0.286 | 0.287 | 0.014 | 0.309 | 0.311 | 0.021 | 0.305 | 0.310 | 0.006 | 0.281 | 0.281 |
| | | $\hat{\lambda}^s$ | -0.352 | 0.512 | 0.622 | -0.577 | 0.656 | 0.874 | -0.489 | 0.682 | 0.839 | -0.110 | 0.368 | 0.384 |
| | | $\hat{\lambda}^o$ | -0.367 | 0.420 | 0.558 | -0.088 | 0.100 | 0.133 | -0.083 | 0.098 | 0.129 | -0.046 | 0.188 | 0.194 |
| | | $\hat{\rho}$ | 0.028 | 0.238 | 0.240 | 0.039 | 0.312 | 0.314 | -0.007 | 0.288 | 0.288 | -0.000 | 0.208 | 0.208 |
| | | $\hat{\sigma}^2$ | -0.022 | 0.171 | 0.173 | -0.063 | 0.095 | 0.114 | -0.001 | 0.101 | 0.101 | -0.018 | 0.130 | 0.131 |
| 0.40 | 0.85 | $\hat{\beta}^s$ | 0.608 | 4.673 | 4.875 | 0.883 | 3.550 | 4.075 | 0.874 | 3.332 | 3.877 | 0.093 | 0.624 | 0.658 |
| | | $\hat{\beta}^o$ | 0.007 | 0.626 | 0.626 | 0.035 | 0.643 | 0.657 | 0.051 | 0.636 | 0.646 | 0.005 | 0.551 | 0.551 |
| | | $\hat{\lambda}^s$ | -0.308 | 0.559 | 0.638 | -0.667 | 0.715 | 0.978 | -0.638 | 0.720 | 0.962 | -0.081 | 0.342 | 0.352 |
| | | $\hat{\lambda}^o$ | -0.458 | 0.705 | 0.841 | -0.090 | 0.052 | 0.104 | -0.085 | 0.052 | 0.100 | -0.018 | 0.047 | 0.050 |
| | | $\hat{\rho}$ | 0.086 | 0.320 | 0.332 | -0.110 | 0.323 | 0.341 | -0.125 | 0.301 | 0.326 | 0.014 | 0.210 | 0.211 |
| | | $\hat{\sigma}^2$ | 0.256 | 0.719 | 0.763 | 0.147 | 0.160 | 0.217 | 0.203 | 0.188 | 0.277 | 0.008 | 0.184 | 0.184 |
| 0.85 | 0.00 | $\hat{\beta}^s$ | 0.545 | 4.112 | 4.275 | 0.041 | 2.987 | 2.989 | 0.061 | 3.352 | 3.358 | 0.227 | 1.859 | 1.921 |
| | | $\hat{\beta}^o$ | 0.001 | 0.277 | 0.277 | 0.012 | 0.538 | 0.539 | 0.000 | 0.354 | 0.355 | 0.000 | 0.271 | 0.271 |
| | | $\hat{\lambda}^s$ | -0.648 | 0.747 | 0.989 | -0.903 | 0.645 | 1.110 | -0.903 | 0.653 | 1.114 | -0.026 | 0.112 | 0.115 |
| | | $\hat{\lambda}^o$ | 0.014 | 0.330 | 0.330 | -0.017 | 0.120 | 0.121 | -0.011 | 0.119 | 0.120 | -0.001 | 0.239 | 0.239 |
| | | $\hat{\rho}$ | 0.010 | 0.363 | 0.363 | -0.050 | 0.439 | 0.441 | -0.041 | 0.366 | 0.368 | -0.004 | 0.294 | 0.294 |
| | | $\hat{\sigma}^2$ | -0.051 | 0.142 | 0.151 | 9.925 | 316.227 | 316.383 | -0.011 | 0.142 | 0.142 | -0.021 | 0.120 | 0.122 |
| 0.85 | 0.40 | $\hat{\beta}^s$ | 0.482 | 4.006 | 4.145 | 0.034 | 2.880 | 2.880 | 0.022 | 2.899 | 2.900 | 0.232 | 1.925 | 1.988 |
| | | $\hat{\beta}^o$ | 0.002 | 0.305 | 0.305 | 0.005 | 0.572 | 0.572 | 0.004 | 0.446 | 0.447 | 0.002 | 0.286 | 0.286 |
| | | $\hat{\lambda}^s$ | -0.691 | 0.767 | 1.033 | -1.040 | 0.672 | 1.238 | -0.951 | 0.692 | 1.176 | -0.024 | 0.101 | 0.104 |
| | | $\hat{\lambda}^o$ | -0.355 | 0.442 | 0.567 | -0.102 | 0.105 | 0.147 | -0.096 | 0.107 | 0.143 | -0.028 | 0.177 | 0.179 |
| | | $\hat{\rho}$ | 0.053 | 0.339 | 0.343 | -0.009 | 0.437 | 0.437 | -0.014 | 0.380 | 0.380 | 0.002 | 0.253 | 0.253 |
| | | $\hat{\sigma}^2$ | -0.018 | 0.179 | 0.180 | 5.337 | 169.987 | 170.071 | 0.022 | 0.441 | 0.442 | -0.016 | 0.127 | 0.128 |
| 0.85 | 0.85 | $\hat{\beta}^s$ | 0.470 | 4.158 | 4.276 | 0.002 | 2.997 | 2.997 | 0.046 | 2.933 | 2.934 | 0.171 | 1.574 | 1.617 |
| | | $\hat{\beta}^o$ | 0.014 | 0.639 | 0.639 | 0.093 | 0.976 | 1.010 | 0.105 | 0.671 | 0.703 | 0.011 | 0.554 | 0.554 |
| | | $\hat{\lambda}^s$ | -0.709 | 0.785 | 1.058 | -1.182 | 0.698 | 1.373 | -1.130 | 0.709 | 1.334 | -0.024 | 0.091 | 0.094 |
| | | $\hat{\lambda}^o$ | -0.517 | 0.723 | 0.888 | -0.092 | 0.062 | 0.111 | -0.083 | 0.054 | 0.099 | -0.016 | 0.047 | 0.050 |
| | | $\hat{\rho}$ | 0.132 | 0.355 | 0.379 | -0.088 | 0.424 | 0.433 | -0.117 | 0.407 | 0.424 | -0.001 | 0.208 | 0.208 |
| | | $\hat{\sigma}^2$ | 0.311 | 0.720 | 0.784 | 0.574 | 14.469 | 14.480 | 0.162 | 0.153 | 0.223 | -0.012 | 0.167 | 0.167 |

**Table 5:** Biases, standard deviations, and RMSEs of marginal effects estimates in the context of the sample selection model with a spatial error in both the selection and outcome equations ($2n = 344$).

| $\lambda^s$ | $\lambda^o$ | | $mfx(\theta_0)$ | HMLE | | | GMM | | | GMM2 | | | PMLE | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | bias | sd | rmse | bias | sd | rmse | bias | sd | rmse | bias | sd | rmse |
| 0.00 | 0.00 | $mfx_1$ | 0.193 | 0.001 | 0.016 | 0.016 | 0.004 | 0.020 | 0.021 | 0.005 | 0.022 | 0.022 | 0.001 | 0.016 | 0.016 |
| | | $mfx_2$ | -0.193 | 0.000 | 0.014 | 0.014 | -0.001 | 0.019 | 0.019 | -0.002 | 0.021 | 0.021 | 0.000 | 0.014 | 0.014 |
| | | $mfx_3$ | 0.782 | -0.010 | 0.086 | 0.087 | -0.013 | 0.099 | 0.100 | -0.011 | 0.097 | 0.098 | -0.008 | 0.087 | 0.087 |
| | | $mfx_4$ | -1.000 | -0.000 | 0.047 | 0.047 | 0.000 | 0.047 | 0.047 | -0.000 | 0.048 | 0.048 | 0.000 | 0.047 | 0.047 |
| 0.00 | 0.40 | $mfx_1$ | 0.193 | -0.001 | 0.016 | 0.016 | 0.002 | 0.023 | 0.023 | 0.004 | 0.024 | 0.025 | -0.001 | 0.016 | 0.016 |
| | | $mfx_2$ | -0.193 | -0.000 | 0.014 | 0.014 | -0.001 | 0.018 | 0.018 | -0.003 | 0.019 | 0.020 | -0.000 | 0.014 | 0.014 |
| | | $mfx_3$ | 0.776 | -0.002 | 0.092 | 0.092 | -0.002 | 0.193 | 0.193 | -0.004 | 0.109 | 0.109 | -0.001 | 0.089 | 0.089 |
| | | $mfx_4$ | -1.000 | -0.001 | 0.048 | 0.048 | -0.001 | 0.048 | 0.048 | -0.001 | 0.049 | 0.049 | -0.001 | 0.047 | 0.047 |
| 0.00 | 0.85 | $mfx_1$ | 0.193 | 0.001 | 0.017 | 0.017 | 0.003 | 0.025 | 0.025 | 0.004 | 0.028 | 0.028 | 0.001 | 0.017 | 0.017 |
| | | $mfx_2$ | -0.193 | -0.000 | 0.014 | 0.014 | -0.001 | 0.025 | 0.025 | -0.002 | 0.026 | 0.026 | -0.000 | 0.014 | 0.014 |
| | | $mfx_3$ | 0.722 | 0.006 | 0.140 | 0.140 | 0.038 | 0.145 | 0.149 | 0.053 | 0.164 | 0.172 | 0.006 | 0.118 | 0.118 |
| | | $mfx_4$ | -1.000 | -0.001 | 0.072 | 0.072 | -0.001 | 0.077 | 0.077 | 0.000 | 0.087 | 0.087 | 0.000 | 0.059 | 0.059 |
| 0.40 | 0.00 | $mfx_1$ | 0.190 | -0.001 | 0.017 | 0.017 | 0.001 | 0.023 | 0.023 | 0.002 | 0.024 | 0.024 | -0.001 | 0.017 | 0.017 |
| | | $mfx_2$ | -0.190 | -0.000 | 0.016 | 0.016 | -0.002 | 0.021 | 0.021 | -0.003 | 0.023 | 0.023 | -0.000 | 0.016 | 0.016 |
| | | $mfx_3$ | 0.794 | -0.002 | 0.083 | 0.083 | -0.005 | 0.109 | 0.109 | -0.002 | 0.091 | 0.091 | -0.000 | 0.083 | 0.083 |
| | | $mfx_4$ | -1.000 | -0.000 | 0.046 | 0.046 | -0.000 | 0.047 | 0.047 | -0.000 | 0.048 | 0.048 | -0.000 | 0.047 | 0.047 |
| 0.40 | 0.40 | $mfx_1$ | 0.190 | -0.001 | 0.018 | 0.018 | 0.002 | 0.026 | 0.026 | 0.004 | 0.027 | 0.027 | -0.000 | 0.018 | 0.018 |
| | | $mfx_2$ | -0.190 | -0.000 | 0.015 | 0.015 | -0.001 | 0.025 | 0.025 | -0.002 | 0.027 | 0.027 | -0.000 | 0.015 | 0.015 |
| | | $mfx_3$ | 0.780 | -0.002 | 0.088 | 0.088 | -0.008 | 0.097 | 0.098 | -0.003 | 0.096 | 0.096 | -0.001 | 0.088 | 0.088 |
| | | $mfx_4$ | -1.000 | -0.001 | 0.047 | 0.047 | -0.001 | 0.048 | 0.048 | -0.001 | 0.048 | 0.048 | -0.001 | 0.046 | 0.046 |
| 0.40 | 0.85 | $mfx_1$ | 0.190 | -0.001 | 0.018 | 0.018 | 0.004 | 0.029 | 0.029 | 0.004 | 0.028 | 0.029 | -0.001 | 0.018 | 0.018 |
| | | $mfx_2$ | -0.190 | 0.001 | 0.016 | 0.016 | -0.000 | 0.026 | 0.026 | -0.001 | 0.025 | 0.025 | 0.001 | 0.015 | 0.015 |
| | | $mfx_3$ | 0.700 | 0.009 | 0.139 | 0.140 | 0.043 | 0.212 | 0.216 | 0.067 | 0.132 | 0.148 | 0.009 | 0.122 | 0.122 |
| | | $mfx_4$ | -1.000 | -0.000 | 0.070 | 0.070 | -0.001 | 0.074 | 0.074 | -0.001 | 0.068 | 0.068 | -0.001 | 0.058 | 0.058 |
| 0.85 | 0.00 | $mfx_1$ | 0.158 | 0.001 | 0.024 | 0.024 | 0.001 | 0.034 | 0.034 | 0.001 | 0.037 | 0.037 | 0.001 | 0.024 | 0.024 |
| | | $mfx_2$ | -0.158 | -0.001 | 0.023 | 0.023 | 0.000 | 0.033 | 0.033 | -0.000 | 0.034 | 0.034 | -0.001 | 0.022 | 0.022 |
| | | $mfx_3$ | 0.881 | -0.007 | 0.076 | 0.077 | -0.052 | 1.551 | 1.551 | -0.003 | 0.105 | 0.105 | -0.006 | 0.076 | 0.076 |
| | | $mfx_4$ | -1.000 | -0.002 | 0.047 | 0.047 | 0.005 | 0.240 | 0.240 | -0.002 | 0.051 | 0.051 | -0.002 | 0.047 | 0.047 |
| 0.85 | 0.40 | $mfx_1$ | 0.158 | -0.001 | 0.025 | 0.025 | -0.000 | 0.039 | 0.039 | 0.002 | 0.038 | 0.038 | -0.001 | 0.024 | 0.024 |
| | | $mfx_2$ | -0.158 | -0.000 | 0.023 | 0.023 | 0.002 | 0.035 | 0.035 | -0.000 | 0.034 | 0.034 | -0.000 | 0.023 | 0.023 |
| | | $mfx_3$ | 0.862 | -0.004 | 0.079 | 0.079 | -0.030 | 0.855 | 0.856 | -0.005 | 0.111 | 0.111 | -0.003 | 0.076 | 0.076 |
| | | $mfx_4$ | -1.000 | 0.002 | 0.047 | 0.047 | 0.007 | 0.178 | 0.178 | 0.001 | 0.048 | 0.048 | 0.002 | 0.046 | 0.046 |
| 0.85 | 0.85 | $mfx_1$ | 0.158 | 0.001 | 0.025 | 0.025 | 0.001 | 0.041 | 0.041 | 0.002 | 0.042 | 0.042 | 0.001 | 0.024 | 0.024 |
| | | $mfx_2$ | -0.158 | -0.000 | 0.022 | 0.022 | 0.002 | 0.045 | 0.045 | 0.001 | 0.039 | 0.039 | -0.001 | 0.022 | 0.022 |
| | | $mfx_3$ | 0.764 | 0.002 | 0.126 | 0.126 | 0.045 | 0.494 | 0.496 | 0.046 | 0.118 | 0.127 | -0.002 | 0.112 | 0.112 |
| | | $mfx_4$ | -1.000 | 0.002 | 0.073 | 0.073 | 0.013 | 0.369 | 0.369 | 0.001 | 0.070 | 0.070 | 0.001 | 0.060 | 0.060 |

**Table 6:** Biases and standard deviations of parameter estimates in the context of the sample selection model with a spatial lag in both the selection and outcome equations under misspecification of the distribution of the error terms ($2n = 344$).

| | | | normal | | | | t | | | | Weibull | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\lambda^s$ | $\lambda^o$ | | bias | | s.d. | | bias | | s.d. | | bias | | s.d. | |
| | | | HMLE | PMLE | HMLE | PMLE | HMLE | PMLE | HMLE | PMLE | HMLE | PMLE | HMLE | PMLE |
| 0.00 | 0.00 | $\hat{\beta}^s$ | 0.044 | 0.043 | 0.449 | 0.446 | 0.061 | 0.061 | 0.458 | 0.458 | 0.196 | 0.196 | 0.545 | 0.543 |
| | | $\hat{\beta}^o$ | 0.006 | 0.005 | 0.255 | 0.255 | 0.003 | 0.003 | 0.251 | 0.251 | 0.026 | 0.026 | 0.292 | 0.291 |
| | | $\hat{\lambda}^s$ | -0.012 | -0.014 | 0.147 | 0.141 | -0.006 | -0.006 | 0.137 | 0.134 | -0.008 | -0.010 | 0.134 | 0.129 |
| | | $\hat{\lambda}^o$ | -0.002 | -0.004 | 0.099 | 0.092 | -0.001 | -0.002 | 0.098 | 0.091 | -0.006 | -0.007 | 0.104 | 0.096 |
| | | $\hat{\rho}$ | -0.004 | -0.005 | 0.204 | 0.204 | -0.013 | -0.012 | 0.214 | 0.213 | 0.114 | 0.114 | 0.232 | 0.230 |
| | | $\hat{\sigma}^2$ | -0.006 | -0.005 | 0.111 | 0.111 | -0.023 | -0.022 | 0.125 | 0.125 | 0.128 | 0.129 | 0.156 | 0.155 |
| 0.00 | 0.40 | $\hat{\beta}^s$ | 0.041 | 0.040 | 0.442 | 0.441 | 0.063 | 0.063 | 0.465 | 0.462 | 0.189 | 0.188 | 0.523 | 0.521 |
| | | $\hat{\beta}^o$ | 0.002 | 0.001 | 0.234 | 0.230 | 0.001 | 0.002 | 0.237 | 0.232 | 0.007 | 0.006 | 0.272 | 0.270 |
| | | $\hat{\lambda}^s$ | -0.012 | -0.012 | 0.141 | 0.135 | -0.009 | -0.009 | 0.141 | 0.133 | -0.006 | -0.006 | 0.126 | 0.122 |
| | | $\hat{\lambda}^o$ | -0.008 | -0.008 | 0.078 | 0.070 | -0.000 | -0.002 | 0.075 | 0.069 | -0.012 | -0.013 | 0.079 | 0.070 |
| | | $\hat{\rho}$ | -0.014 | -0.011 | 0.224 | 0.217 | -0.010 | -0.010 | 0.220 | 0.218 | 0.106 | 0.104 | 0.238 | 0.236 |
| | | $\hat{\sigma}^2$ | -0.012 | -0.008 | 0.123 | 0.119 | -0.029 | -0.025 | 0.136 | 0.132 | 0.117 | 0.120 | 0.170 | 0.166 |
| 0.00 | 0.85 | $\hat{\beta}^s$ | 0.045 | 0.044 | 0.459 | 0.452 | 0.089 | 0.093 | 0.475 | 0.470 | 0.214 | 0.215 | 0.520 | 0.518 |
| | | $\hat{\beta}^o$ | 0.009 | 0.005 | 0.368 | 0.284 | 0.006 | 0.008 | 0.417 | 0.286 | 0.009 | 0.004 | 0.396 | 0.302 |
| | | $\hat{\lambda}^s$ | -0.004 | -0.003 | 0.142 | 0.131 | -0.008 | -0.010 | 0.139 | 0.133 | -0.003 | -0.006 | 0.135 | 0.128 |
| | | $\hat{\lambda}^o$ | -0.004 | -0.005 | 0.086 | 0.030 | -0.008 | -0.006 | 0.127 | 0.057 | -0.007 | -0.008 | 0.088 | 0.031 |
| | | $\hat{\rho}$ | -0.012 | -0.011 | 0.305 | 0.230 | 0.001 | -0.002 | 0.301 | 0.238 | 0.043 | 0.035 | 0.298 | 0.256 |
| | | $\hat{\sigma}^2$ | -0.109 | -0.064 | 0.280 | 0.174 | -0.106 | -0.073 | 0.343 | 0.232 | -0.019 | 0.027 | 0.335 | 0.210 |
| 0.40 | 0.00 | $\hat{\beta}^s$ | 0.048 | 0.047 | 0.413 | 0.411 | 0.064 | 0.062 | 0.426 | 0.425 | 0.159 | 0.159 | 0.502 | 0.500 |
| | | $\hat{\beta}^o$ | 0.003 | 0.003 | 0.252 | 0.252 | 0.002 | 0.002 | 0.243 | 0.243 | 0.021 | 0.020 | 0.290 | 0.287 |
| | | $\hat{\lambda}^s$ | -0.008 | -0.010 | 0.102 | 0.097 | -0.009 | -0.009 | 0.105 | 0.100 | 0.002 | 0.001 | 0.097 | 0.092 |
| | | $\hat{\lambda}^o$ | -0.007 | -0.007 | 0.095 | 0.089 | -0.004 | -0.003 | 0.098 | 0.090 | 0.006 | 0.005 | 0.103 | 0.096 |
| | | $\hat{\rho}$ | -0.002 | -0.001 | 0.209 | 0.208 | -0.004 | -0.004 | 0.217 | 0.216 | 0.118 | 0.117 | 0.236 | 0.233 |
| | | $\hat{\sigma}^2$ | -0.012 | -0.010 | 0.105 | 0.104 | -0.025 | -0.024 | 0.124 | 0.124 | 0.123 | 0.123 | 0.158 | 0.158 |
| 0.40 | 0.40 | $\hat{\beta}^s$ | 0.034 | 0.034 | 0.426 | 0.424 | 0.071 | 0.071 | 0.451 | 0.447 | 0.165 | 0.165 | 0.505 | 0.503 |
| | | $\hat{\beta}^o$ | 0.004 | 0.006 | 0.237 | 0.231 | 0.003 | 0.003 | 0.234 | 0.227 | 0.003 | 0.003 | 0.269 | 0.263 |
| | | $\hat{\lambda}^s$ | -0.004 | -0.005 | 0.099 | 0.092 | -0.006 | -0.008 | 0.099 | 0.096 | 0.001 | -0.000 | 0.090 | 0.087 |
| | | $\hat{\lambda}^o$ | -0.003 | -0.004 | 0.074 | 0.068 | -0.002 | -0.003 | 0.073 | 0.065 | -0.004 | -0.005 | 0.077 | 0.069 |
| | | $\hat{\rho}$ | -0.004 | -0.002 | 0.204 | 0.197 | -0.008 | -0.003 | 0.212 | 0.200 | 0.105 | 0.104 | 0.220 | 0.214 |
| | | $\hat{\sigma}^2$ | -0.018 | -0.014 | 0.123 | 0.116 | -0.034 | -0.031 | 0.130 | 0.124 | 0.100 | 0.104 | 0.157 | 0.150 |
| 0.40 | 0.85 | $\hat{\beta}^s$ | 0.036 | 0.038 | 0.422 | 0.415 | 0.062 | 0.064 | 0.436 | 0.428 | 0.183 | 0.182 | 0.499 | 0.492 |
| | | $\hat{\beta}^o$ | 0.012 | 0.004 | 0.357 | 0.282 | 0.004 | 0.015 | 0.374 | 0.288 | 0.007 | 0.016 | 0.397 | 0.304 |
| | | $\hat{\lambda}^s$ | -0.000 | -0.004 | 0.105 | 0.097 | -0.004 | -0.009 | 0.099 | 0.094 | 0.007 | 0.003 | 0.093 | 0.089 |
| | | $\hat{\lambda}^o$ | -0.005 | -0.005 | 0.101 | 0.029 | -0.002 | -0.004 | 0.086 | 0.029 | -0.007 | -0.008 | 0.102 | 0.031 |
| | | $\hat{\rho}$ | -0.020 | -0.001 | 0.292 | 0.207 | -0.029 | 0.002 | 0.314 | 0.228 | 0.050 | 0.062 | 0.285 | 0.222 |
| | | $\hat{\sigma}^2$ | -0.101 | -0.061 | 0.327 | 0.171 | -0.114 | -0.065 | 0.316 | 0.175 | -0.024 | 0.033 | 0.368 | 0.203 |
| 0.85 | 0.00 | $\hat{\beta}^s$ | 0.125 | 0.111 | 0.693 | 0.641 | 0.141 | 0.128 | 0.671 | 0.634 | 0.176 | 0.168 | 0.724 | 0.685 |
| | | $\hat{\beta}^o$ | 0.003 | 0.002 | 0.229 | 0.227 | 0.004 | 0.004 | 0.228 | 0.226 | 0.019 | 0.017 | 0.264 | 0.259 |
| | | $\hat{\lambda}^s$ | -0.002 | -0.004 | 0.053 | 0.047 | -0.002 | -0.005 | 0.051 | 0.045 | -0.004 | -0.005 | 0.050 | 0.045 |
| | | $\hat{\lambda}^o$ | -0.005 | -0.005 | 0.106 | 0.097 | -0.006 | -0.006 | 0.104 | 0.094 | 0.004 | 0.002 | 0.110 | 0.097 |
| | | $\hat{\rho}$ | 0.037 | 0.027 | 0.242 | 0.226 | 0.025 | 0.020 | 0.256 | 0.239 | 0.100 | 0.090 | 0.275 | 0.255 |
| | | $\hat{\sigma}^2$ | -0.012 | -0.011 | 0.099 | 0.100 | -0.016 | -0.015 | 0.123 | 0.122 | 0.082 | 0.083 | 0.143 | 0.142 |
| 0.85 | 0.40 | $\hat{\beta}^s$ | 0.121 | 0.108 | 0.697 | 0.646 | 0.142 | 0.128 | 0.707 | 0.668 | 0.189 | 0.183 | 0.759 | 0.722 |
| | | $\hat{\beta}^o$ | 0.001 | 0.000 | 0.238 | 0.229 | 0.002 | 0.003 | 0.225 | 0.219 | 0.008 | 0.007 | 0.256 | 0.244 |
| | | $\hat{\lambda}^s$ | -0.002 | -0.004 | 0.051 | 0.045 | -0.000 | -0.003 | 0.051 | 0.046 | -0.002 | -0.004 | 0.050 | 0.045 |
| | | $\hat{\lambda}^o$ | -0.006 | -0.005 | 0.081 | 0.072 | -0.002 | -0.003 | 0.075 | 0.067 | -0.000 | -0.001 | 0.082 | 0.071 |
| | | $\hat{\rho}$ | 0.009 | 0.006 | 0.252 | 0.231 | 0.009 | 0.006 | 0.252 | 0.232 | 0.077 | 0.073 | 0.277 | 0.250 |
| | | $\hat{\sigma}^2$ | -0.021 | -0.019 | 0.116 | 0.109 | -0.030 | -0.027 | 0.135 | 0.129 | 0.075 | 0.077 | 0.149 | 0.142 |
| 0.85 | 0.85 | $\hat{\beta}^s$ | 0.117 | 0.102 | 0.706 | 0.625 | 0.137 | 0.121 | 0.701 | 0.655 | 0.190 | 0.181 | 0.741 | 0.693 |
| | | $\hat{\beta}^o$ | 0.013 | 0.004 | 0.418 | 0.273 | 0.000 | 0.014 | 0.399 | 0.269 | 0.011 | 0.010 | 0.372 | 0.281 |
| | | $\hat{\lambda}^s$ | 0.002 | -0.003 | 0.049 | 0.044 | -0.001 | -0.005 | 0.051 | 0.045 | 0.004 | -0.000 | 0.048 | 0.044 |
| | | $\hat{\lambda}^o$ | -0.012 | -0.007 | 0.140 | 0.028 | -0.008 | -0.006 | 0.101 | 0.031 | -0.006 | -0.006 | 0.103 | 0.030 |
| | | $\hat{\rho}$ | -0.075 | -0.014 | 0.322 | 0.232 | -0.053 | -0.002 | 0.320 | 0.227 | -0.000 | 0.041 | 0.325 | 0.240 |
| | | $\hat{\sigma}^2$ | -0.111 | -0.065 | 0.323 | 0.155 | -0.095 | -0.057 | 0.356 | 0.173 | -0.055 | 0.002 | 0.329 | 0.187 |

**Table 7:** RMSEs of parameter estimates in the context of the sample selection model with a spatial lag in both the selection and outcome equations under misspecification of the distribution of the error terms ($2n = 344$).

| $\lambda^s$ | $\lambda^o$ | | normal | | t | | Weibull | |
|---|---|---|---|---|---|---|---|---|
| | | | HMLE | PMLE | HMLE | PMLE | HMLE | PMLE |
| 0.00 | 0.00 | $\hat{\beta}^s$ | 0.459 | 0.456 | 0.483 | 0.482 | 0.658 | 0.656 |
| | | $\hat{\beta}^o$ | 0.255 | 0.255 | 0.251 | 0.251 | 0.313 | 0.312 |
| | | $\hat{\lambda}^s$ | 0.147 | 0.141 | 0.137 | 0.135 | 0.134 | 0.130 |
| | | $\hat{\lambda}^o$ | 0.099 | 0.092 | 0.098 | 0.091 | 0.104 | 0.097 |
| | | $\hat{\rho}$ | 0.204 | 0.204 | 0.214 | 0.213 | 0.259 | 0.257 |
| | | $\hat{\sigma}^2$ | 0.112 | 0.111 | 0.127 | 0.127 | 0.202 | 0.202 |
| 0.00 | 0.40 | $\hat{\beta}^s$ | 0.451 | 0.449 | 0.494 | 0.491 | 0.632 | 0.630 |
| | | $\hat{\beta}^o$ | 0.234 | 0.230 | 0.237 | 0.232 | 0.285 | 0.281 |
| | | $\hat{\lambda}^s$ | 0.141 | 0.135 | 0.141 | 0.133 | 0.126 | 0.122 |
| | | $\hat{\lambda}^o$ | 0.078 | 0.071 | 0.075 | 0.069 | 0.080 | 0.072 |
| | | $\hat{\rho}$ | 0.224 | 0.217 | 0.221 | 0.219 | 0.260 | 0.258 |
| | | $\hat{\sigma}^2$ | 0.123 | 0.119 | 0.139 | 0.135 | 0.206 | 0.205 |
| 0.00 | 0.85 | $\hat{\beta}^s$ | 0.469 | 0.462 | 0.510 | 0.509 | 0.679 | 0.677 |
| | | $\hat{\beta}^o$ | 0.370 | 0.284 | 0.418 | 0.286 | 0.397 | 0.303 |
| | | $\hat{\lambda}^s$ | 0.142 | 0.131 | 0.139 | 0.133 | 0.135 | 0.128 |
| | | $\hat{\lambda}^o$ | 0.086 | 0.031 | 0.127 | 0.057 | 0.088 | 0.032 |
| | | $\hat{\rho}$ | 0.305 | 0.230 | 0.301 | 0.238 | 0.301 | 0.258 |
| | | $\hat{\sigma}^2$ | 0.300 | 0.185 | 0.359 | 0.243 | 0.335 | 0.212 |
| 0.40 | 0.00 | $\hat{\beta}^s$ | 0.427 | 0.425 | 0.460 | 0.457 | 0.598 | 0.595 |
| | | $\hat{\beta}^o$ | 0.252 | 0.252 | 0.243 | 0.244 | 0.312 | 0.310 |
| | | $\hat{\lambda}^s$ | 0.102 | 0.098 | 0.105 | 0.100 | 0.097 | 0.092 |
| | | $\hat{\lambda}^o$ | 0.095 | 0.089 | 0.098 | 0.090 | 0.103 | 0.097 |
| | | $\hat{\rho}$ | 0.209 | 0.208 | 0.217 | 0.216 | 0.264 | 0.261 |
| | | $\hat{\sigma}^2$ | 0.105 | 0.105 | 0.126 | 0.126 | 0.200 | 0.200 |
| 0.40 | 0.40 | $\hat{\beta}^s$ | 0.436 | 0.433 | 0.482 | 0.479 | 0.614 | 0.612 |
| | | $\hat{\beta}^o$ | 0.238 | 0.231 | 0.234 | 0.227 | 0.283 | 0.277 |
| | | $\hat{\lambda}^s$ | 0.099 | 0.092 | 0.099 | 0.096 | 0.090 | 0.087 |
| | | $\hat{\lambda}^o$ | 0.074 | 0.068 | 0.073 | 0.065 | 0.077 | 0.069 |
| | | $\hat{\rho}$ | 0.204 | 0.197 | 0.213 | 0.200 | 0.243 | 0.237 |
| | | $\hat{\sigma}^2$ | 0.124 | 0.117 | 0.134 | 0.128 | 0.186 | 0.183 |
| 0.40 | 0.85 | $\hat{\beta}^s$ | 0.436 | 0.431 | 0.466 | 0.459 | 0.643 | 0.637 |
| | | $\hat{\beta}^o$ | 0.358 | 0.283 | 0.375 | 0.289 | 0.398 | 0.307 |
| | | $\hat{\lambda}^s$ | 0.105 | 0.097 | 0.099 | 0.094 | 0.093 | 0.089 |
| | | $\hat{\lambda}^o$ | 0.101 | 0.029 | 0.086 | 0.029 | 0.102 | 0.031 |
| | | $\hat{\rho}$ | 0.292 | 0.207 | 0.316 | 0.228 | 0.289 | 0.231 |
| | | $\hat{\sigma}^2$ | 0.342 | 0.181 | 0.336 | 0.187 | 0.368 | 0.206 |
| 0.85 | 0.00 | $\hat{\beta}^s$ | 0.788 | 0.723 | 0.768 | 0.717 | 0.873 | 0.825 |
| | | $\hat{\beta}^o$ | 0.229 | 0.227 | 0.228 | 0.226 | 0.271 | 0.265 |
| | | $\hat{\lambda}^s$ | 0.053 | 0.047 | 0.051 | 0.045 | 0.050 | 0.045 |
| | | $\hat{\lambda}^o$ | 0.106 | 0.097 | 0.104 | 0.094 | 0.110 | 0.097 |
| | | $\hat{\rho}$ | 0.244 | 0.227 | 0.257 | 0.240 | 0.292 | 0.270 |
| | | $\hat{\sigma}^2$ | 0.100 | 0.100 | 0.124 | 0.123 | 0.165 | 0.165 |
| 0.85 | 0.40 | $\hat{\beta}^s$ | 0.783 | 0.721 | 0.802 | 0.750 | 0.905 | 0.863 |
| | | $\hat{\beta}^o$ | 0.238 | 0.230 | 0.225 | 0.219 | 0.258 | 0.247 |
| | | $\hat{\lambda}^s$ | 0.051 | 0.045 | 0.051 | 0.046 | 0.050 | 0.046 |
| | | $\hat{\lambda}^o$ | 0.081 | 0.072 | 0.075 | 0.067 | 0.082 | 0.071 |
| | | $\hat{\rho}$ | 0.252 | 0.231 | 0.252 | 0.232 | 0.287 | 0.260 |
| | | $\hat{\sigma}^2$ | 0.117 | 0.111 | 0.138 | 0.131 | 0.167 | 0.162 |
| 0.85 | 0.85 | $\hat{\beta}^s$ | 0.788 | 0.692 | 0.791 | 0.729 | 0.905 | 0.846 |
| | | $\hat{\beta}^o$ | 0.420 | 0.274 | 0.400 | 0.270 | 0.373 | 0.282 |
| | | $\hat{\lambda}^s$ | 0.050 | 0.044 | 0.051 | 0.046 | 0.048 | 0.044 |
| | | $\hat{\lambda}^o$ | 0.141 | 0.029 | 0.101 | 0.031 | 0.104 | 0.031 |
| | | $\hat{\rho}$ | 0.331 | 0.232 | 0.324 | 0.227 | 0.325 | 0.244 |
| | | $\hat{\sigma}^2$ | 0.342 | 0.168 | 0.369 | 0.182 | 0.333 | 0.187 |

**Table 8:** Standard deviations and standard errors of parameter estimates in the context of the sample selection model with a spatial lag in both the selection and outcome equations ($2n = 344$).

| $\lambda^s$ | $\lambda^o$ | | $\hat{\beta}^s_1$ | $\hat{\beta}^s_2$ | $\hat{\beta}^s_3$ | $\hat{\beta}^o_1$ | $\hat{\beta}^o_2$ | $\hat{\beta}^o_3$ | $\hat{\lambda}^s$ | $\hat{\lambda}^o$ | $\hat{\rho}$ | $\hat{\sigma}^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.00 | 0.00 | SD | 0.190 | 0.131 | 0.125 | 0.116 | 0.091 | 0.047 | 0.141 | 0.092 | 0.204 | 0.111 |
| | | SE | 0.181 | 0.129 | 0.123 | 0.118 | 0.091 | 0.047 | 0.132 | 0.091 | 0.205 | 0.111 |
| 0.00 | 0.40 | SD | 0.186 | 0.129 | 0.125 | 0.091 | 0.090 | 0.049 | 0.135 | 0.070 | 0.217 | 0.119 |
| | | SE | 0.181 | 0.129 | 0.124 | 0.094 | 0.093 | 0.048 | 0.132 | 0.068 | 0.207 | 0.119 |
| 0.00 | 0.85 | SD | 0.190 | 0.132 | 0.130 | 0.092 | 0.127 | 0.065 | 0.131 | 0.030 | 0.230 | 0.174 |
| | | SE | 0.182 | 0.130 | 0.124 | 0.088 | 0.124 | 0.063 | 0.134 | 0.028 | 0.232 | 0.179 |
| 0.40 | 0.00 | SD | 0.155 | 0.133 | 0.123 | 0.116 | 0.089 | 0.047 | 0.097 | 0.089 | 0.208 | 0.104 |
| | | SE | 0.158 | 0.132 | 0.125 | 0.117 | 0.089 | 0.047 | 0.095 | 0.092 | 0.205 | 0.110 |
| 0.40 | 0.40 | SD | 0.161 | 0.137 | 0.125 | 0.091 | 0.091 | 0.050 | 0.092 | 0.068 | 0.197 | 0.116 |
| | | SE | 0.157 | 0.132 | 0.125 | 0.094 | 0.090 | 0.048 | 0.095 | 0.067 | 0.201 | 0.115 |
| 0.40 | 0.85 | SD | 0.159 | 0.133 | 0.123 | 0.093 | 0.121 | 0.068 | 0.097 | 0.029 | 0.207 | 0.171 |
| | | SE | 0.161 | 0.134 | 0.127 | 0.090 | 0.120 | 0.064 | 0.095 | 0.027 | 0.219 | 0.176 |
| 0.85 | 0.00 | SD | 0.215 | 0.229 | 0.197 | 0.105 | 0.074 | 0.048 | 0.047 | 0.097 | 0.226 | 0.100 |
| | | SE | 0.216 | 0.221 | 0.198 | 0.107 | 0.077 | 0.048 | 0.043 | 0.094 | 0.242 | 0.103 |
| 0.85 | 0.40 | SD | 0.221 | 0.225 | 0.200 | 0.098 | 0.081 | 0.051 | 0.045 | 0.072 | 0.231 | 0.109 |
| | | SE | 0.216 | 0.221 | 0.198 | 0.096 | 0.081 | 0.049 | 0.043 | 0.070 | 0.241 | 0.109 |
| 0.85 | 0.85 | SD | 0.215 | 0.214 | 0.196 | 0.096 | 0.109 | 0.068 | 0.044 | 0.028 | 0.232 | 0.155 |
| | | SE | 0.219 | 0.220 | 0.198 | 0.092 | 0.107 | 0.064 | 0.043 | 0.028 | 0.255 | 0.164 |

**Table 9:** Empirical sizes of the $z$-tests with the null hypotheses that the parameters are equal to the corresponding true values in the context of the sample selection model with a spatial lag in both the selection and outcome equations ($2n = 344$).

| $\lambda^s$ | $\lambda^o$ | $\hat{\beta}^s_1$ | $\hat{\beta}^s_2$ | $\hat{\beta}^s_3$ | $\hat{\beta}^o_1$ | $\hat{\beta}^o_2$ | $\hat{\beta}^o_3$ | $\hat{\lambda}^s$ | $\hat{\lambda}^o$ | $\hat{\rho}$ | $\hat{\sigma}^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.00 | 0.00 | 5.00 | 5.30 | 3.80 | 5.30 | 6.30 | 4.80 | 7.10 | 5.40 | 10.30 | 6.50 |
| 0.00 | 0.40 | 5.20 | 5.70 | 5.50 | 5.10 | 5.40 | 5.30 | 5.10 | 5.90 | 11.00 | 7.00 |
| 0.00 | 0.85 | 5.20 | 6.20 | 5.60 | 5.50 | 5.60 | 5.30 | 5.50 | 7.70 | 9.50 | 10.40 |
| 0.40 | 0.00 | 3.80 | 4.00 | 4.50 | 5.10 | 6.10 | 5.60 | 6.50 | 4.70 | 11.20 | 6.40 |
| 0.40 | 0.40 | 4.50 | 5.10 | 4.90 | 4.90 | 5.70 | 5.60 | 3.90 | 5.60 | 10.30 | 7.00 |
| 0.40 | 0.85 | 4.70 | 4.60 | 4.10 | 6.50 | 6.50 | 7.40 | 5.70 | 6.00 | 8.10 | 10.00 |
| 0.85 | 0.00 | 6.80 | 5.80 | 5.70 | 5.00 | 4.10 | 5.40 | 9.40 | 5.10 | 8.00 | 5.10 |
| 0.85 | 0.40 | 5.70 | 4.80 | 6.00 | 5.60 | 4.90 | 5.80 | 7.10 | 5.70 | 10.50 | 6.30 |
| 0.85 | 0.85 | 4.30 | 4.30 | 4.10 | 6.40 | 5.70 | 6.60 | 6.30 | 5.40 | 8.30 | 9.30 |

**Table 10:** Biases and standard deviations of parameter estimates in the context of the sample selection model with a spatial lag in both the selection and outcome equations and a block diagonal weight matrix ($2n = 344$).

| | | | Block size = 4 | | | | | Block size = 8 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\lambda^s$ | $\lambda^o$ | | bias | | s.d. | | | bias | | s.d. | | |
| | | | PMLE | FMLE | PMLE | FMLE | ratio | PMLE | FMLE | PMLE | FMLE | ratio |
| 0.00 | 0.00 | $\hat{\beta}^s$ | 0.051 | 0.051 | 0.437 | 0.437 | 1.000 | 0.059 | 0.061 | 0.451 | 0.449 | 1.005 |
| | | $\hat{\beta}^o$ | 0.006 | 0.005 | 0.250 | 0.250 | 1.003 | 0.005 | 0.004 | 0.252 | 0.251 | 1.003 |
| | | $\hat{\lambda}^s$ | -0.007 | -0.008 | 0.090 | 0.083 | 1.085 | -0.019 | -0.023 | 0.151 | 0.136 | 1.106 |
| | | $\hat{\lambda}^o$ | -0.007 | -0.008 | 0.063 | 0.056 | 1.128 | -0.014 | -0.019 | 0.105 | 0.089 | 1.179 |
| | | $\hat{\rho}$ | 0.002 | 0.001 | 0.214 | 0.212 | 1.008 | 0.001 | 0.001 | 0.210 | 0.208 | 1.008 |
| | | $\hat{\sigma}^2$ | -0.007 | -0.006 | 0.111 | 0.111 | 1.000 | -0.009 | -0.007 | 0.111 | 0.111 | 1.000 |
| 0.00 | 0.40 | $\hat{\beta}^s$ | 0.027 | 0.027 | 0.435 | 0.436 | 0.998 | 0.055 | 0.058 | 0.449 | 0.443 | 1.012 |
| | | $\hat{\beta}^o$ | 0.005 | 0.004 | 0.235 | 0.217 | 1.078 | 0.000 | 0.000 | 0.230 | 0.221 | 1.041 |
| | | $\hat{\lambda}^s$ | -0.008 | -0.010 | 0.092 | 0.088 | 1.046 | -0.021 | -0.028 | 0.152 | 0.140 | 1.086 |
| | | $\hat{\lambda}^o$ | -0.005 | -0.005 | 0.038 | 0.034 | 1.105 | -0.005 | -0.007 | 0.056 | 0.048 | 1.185 |
| | | $\hat{\rho}$ | -0.001 | 0.002 | 0.207 | 0.187 | 1.112 | -0.004 | -0.002 | 0.209 | 0.195 | 1.074 |
| | | $\hat{\sigma}^2$ | -0.006 | -0.006 | 0.125 | 0.112 | 1.119 | -0.014 | -0.006 | 0.114 | 0.105 | 1.084 |
| 0.00 | 0.85 | $\hat{\beta}^s$ | 0.057 | 0.059 | 0.430 | 0.429 | 1.002 | 0.051 | 0.049 | 0.449 | 0.445 | 1.009 |
| | | $\hat{\beta}^o$ | 0.007 | 0.009 | 0.253 | 0.209 | 1.212 | 0.009 | 0.004 | 0.283 | 0.197 | 1.434 |
| | | $\hat{\lambda}^s$ | -0.008 | -0.010 | 0.089 | 0.084 | 1.054 | -0.027 | -0.026 | 0.156 | 0.145 | 1.076 |
| | | $\hat{\lambda}^o$ | -0.001 | -0.001 | 0.011 | 0.009 | 1.162 | -0.002 | -0.002 | 0.013 | 0.011 | 1.148 |
| | | $\hat{\rho}$ | 0.004 | 0.008 | 0.190 | 0.152 | 1.253 | -0.002 | 0.013 | 0.238 | 0.167 | 1.431 |
| | | $\hat{\sigma}^2$ | -0.010 | -0.003 | 0.140 | 0.111 | 1.263 | -0.025 | -0.006 | 0.166 | 0.108 | 1.536 |
| 0.40 | 0.00 | $\hat{\beta}^s$ | 0.043 | 0.041 | 0.432 | 0.427 | 1.012 | 0.052 | 0.045 | 0.430 | 0.421 | 1.021 |
| | | $\hat{\beta}^o$ | 0.007 | 0.007 | 0.242 | 0.239 | 1.012 | 0.002 | 0.003 | 0.249 | 0.250 | 0.997 |
| | | $\hat{\lambda}^s$ | -0.005 | -0.007 | 0.055 | 0.051 | 1.073 | -0.010 | -0.013 | 0.078 | 0.067 | 1.151 |
| | | $\hat{\lambda}^o$ | -0.004 | -0.004 | 0.062 | 0.055 | 1.124 | -0.010 | -0.013 | 0.104 | 0.089 | 1.172 |
| | | $\hat{\rho}$ | -0.002 | -0.004 | 0.210 | 0.202 | 1.040 | 0.002 | -0.000 | 0.205 | 0.202 | 1.014 |
| | | $\hat{\sigma}^2$ | -0.009 | -0.009 | 0.106 | 0.106 | 1.007 | -0.010 | -0.008 | 0.108 | 0.108 | 0.996 |
| 0.40 | 0.40 | $\hat{\beta}^s$ | 0.038 | 0.035 | 0.436 | 0.428 | 1.019 | 0.039 | 0.034 | 0.437 | 0.427 | 1.025 |
| | | $\hat{\beta}^o$ | 0.002 | 0.002 | 0.233 | 0.215 | 1.081 | 0.001 | 0.000 | 0.235 | 0.221 | 1.062 |
| | | $\hat{\lambda}^s$ | -0.004 | -0.004 | 0.054 | 0.050 | 1.078 | -0.006 | -0.010 | 0.070 | 0.063 | 1.112 |
| | | $\hat{\lambda}^o$ | -0.003 | -0.003 | 0.039 | 0.036 | 1.093 | -0.005 | -0.007 | 0.056 | 0.049 | 1.146 |
| | | $\hat{\rho}$ | -0.004 | 0.005 | 0.195 | 0.175 | 1.115 | -0.014 | -0.005 | 0.208 | 0.193 | 1.076 |
| | | $\hat{\sigma}^2$ | -0.010 | -0.008 | 0.115 | 0.104 | 1.108 | -0.017 | -0.008 | 0.115 | 0.104 | 1.106 |
| 0.40 | 0.85 | $\hat{\beta}^s$ | 0.038 | 0.035 | 0.422 | 0.413 | 1.021 | 0.045 | 0.041 | 0.433 | 0.420 | 1.032 |
| | | $\hat{\beta}^o$ | 0.002 | 0.003 | 0.248 | 0.201 | 1.234 | 0.003 | 0.000 | 0.278 | 0.197 | 1.411 |
| | | $\hat{\lambda}^s$ | -0.004 | -0.006 | 0.055 | 0.052 | 1.058 | -0.005 | -0.010 | 0.075 | 0.067 | 1.120 |
| | | $\hat{\lambda}^o$ | -0.001 | -0.001 | 0.010 | 0.009 | 1.144 | -0.002 | -0.002 | 0.013 | 0.011 | 1.153 |
| | | $\hat{\rho}$ | 0.004 | 0.007 | 0.166 | 0.138 | 1.201 | -0.007 | 0.000 | 0.211 | 0.155 | 1.366 |
| | | $\hat{\sigma}^2$ | -0.009 | -0.007 | 0.139 | 0.109 | 1.277 | -0.026 | -0.011 | 0.150 | 0.103 | 1.458 |
| 0.85 | 0.00 | $\hat{\beta}^s$ | 0.068 | 0.056 | 0.646 | 0.606 | 1.067 | 0.142 | 0.080 | 0.812 | 0.618 | 1.313 |
| | | $\hat{\beta}^o$ | 0.005 | 0.003 | 0.222 | 0.219 | 1.013 | 0.003 | 0.001 | 0.217 | 0.209 | 1.038 |
| | | $\hat{\lambda}^s$ | -0.001 | -0.002 | 0.036 | 0.034 | 1.074 | -0.001 | -0.003 | 0.032 | 0.028 | 1.120 |
| | | $\hat{\lambda}^o$ | -0.005 | -0.006 | 0.064 | 0.055 | 1.175 | -0.012 | -0.013 | 0.105 | 0.084 | 1.244 |
| | | $\hat{\rho}$ | -0.016 | -0.009 | 0.275 | 0.243 | 1.134 | 0.003 | 0.010 | 0.278 | 0.216 | 1.285 |
| | | $\hat{\sigma}^2$ | -0.020 | -0.018 | 0.100 | 0.100 | 0.997 | -0.015 | -0.013 | 0.096 | 0.095 | 1.004 |
| 0.85 | 0.40 | $\hat{\beta}^s$ | 0.077 | 0.064 | 0.680 | 0.628 | 1.084 | 0.122 | 0.071 | 0.788 | 0.621 | 1.268 |
| | | $\hat{\beta}^o$ | 0.000 | 0.003 | 0.224 | 0.214 | 1.047 | 0.000 | 0.002 | 0.224 | 0.212 | 1.056 |
| | | $\hat{\lambda}^s$ | -0.002 | -0.002 | 0.037 | 0.034 | 1.092 | 0.000 | -0.001 | 0.032 | 0.028 | 1.136 |
| | | $\hat{\lambda}^o$ | -0.002 | -0.002 | 0.040 | 0.037 | 1.084 | -0.007 | -0.007 | 0.059 | 0.052 | 1.129 |
| | | $\hat{\rho}$ | 0.000 | 0.015 | 0.247 | 0.219 | 1.125 | -0.001 | 0.016 | 0.271 | 0.229 | 1.186 |
| | | $\hat{\sigma}^2$ | -0.018 | -0.015 | 0.107 | 0.095 | 1.124 | -0.014 | -0.007 | 0.104 | 0.095 | 1.102 |
| 0.85 | 0.85 | $\hat{\beta}^s$ | 0.063 | 0.052 | 0.648 | 0.598 | 1.083 | 0.094 | 0.055 | 0.755 | 0.614 | 1.230 |
| | | $\hat{\beta}^o$ | 0.007 | 0.003 | 0.246 | 0.211 | 1.166 | 0.001 | 0.001 | 0.280 | 0.206 | 1.364 |
| | | $\hat{\lambda}^s$ | -0.003 | -0.003 | 0.036 | 0.032 | 1.097 | -0.000 | -0.002 | 0.031 | 0.027 | 1.129 |
| | | $\hat{\lambda}^o$ | -0.002 | -0.002 | 0.011 | 0.010 | 1.067 | -0.003 | -0.002 | 0.014 | 0.013 | 1.094 |
| | | $\hat{\rho}$ | 0.006 | 0.010 | 0.213 | 0.189 | 1.127 | -0.006 | 0.012 | 0.251 | 0.202 | 1.244 |
| | | $\hat{\sigma}^2$ | -0.016 | -0.011 | 0.120 | 0.099 | 1.215 | -0.020 | -0.006 | 0.140 | 0.099 | 1.421 |

*Note.* The columns with the label *ratio* provide ratios of the standard deviations of PMLE and FMLE.

**Table 11:** RMSEs of parameter estimates in the context of the sample selection model with a spatial lag in both the selection and outcome equations and a block diagonal weight matrix ($2n = 344$).

| $\lambda^s$ | $\lambda^o$ | | Block size = 4 | | | Block size = 8 | | |
|---|---|---|---|---|---|---|---|---|
| | | | PMLE | FMLE | ratio | PMLE | FMLE | ratio |
| 0.00 | 0.00 | $\hat{\beta}^s$ | 0.452 | 0.451 | 1.001 | 0.467 | 0.465 | 1.004 |
| | | $\hat{\beta}^o$ | 0.250 | 0.250 | 1.003 | 0.252 | 0.251 | 1.004 |
| | | $\hat{\lambda}^s$ | 0.091 | 0.084 | 1.084 | 0.152 | 0.138 | 1.099 |
| | | $\hat{\lambda}^o$ | 0.063 | 0.056 | 1.124 | 0.106 | 0.091 | 1.164 |
| | | $\hat{\rho}$ | 0.214 | 0.212 | 1.008 | 0.210 | 0.208 | 1.008 |
| | | $\hat{\sigma}^2$ | 0.111 | 0.111 | 1.001 | 0.111 | 0.111 | 1.001 |
| 0.00 | 0.40 | $\hat{\beta}^s$ | 0.442 | 0.443 | 0.998 | 0.465 | 0.460 | 1.010 |
| | | $\hat{\beta}^o$ | 0.235 | 0.218 | 1.078 | 0.230 | 0.221 | 1.041 |
| | | $\hat{\lambda}^s$ | 0.092 | 0.088 | 1.043 | 0.154 | 0.143 | 1.074 |
| | | $\hat{\lambda}^o$ | 0.038 | 0.034 | 1.104 | 0.057 | 0.048 | 1.175 |
| | | $\hat{\rho}$ | 0.207 | 0.187 | 1.112 | 0.209 | 0.195 | 1.074 |
| | | $\hat{\sigma}^2$ | 0.125 | 0.112 | 1.119 | 0.115 | 0.105 | 1.090 |
| 0.00 | 0.85 | $\hat{\beta}^s$ | 0.445 | 0.445 | 1.001 | 0.461 | 0.456 | 1.011 |
| | | $\hat{\beta}^o$ | 0.254 | 0.209 | 1.212 | 0.284 | 0.198 | 1.438 |
| | | $\hat{\lambda}^s$ | 0.089 | 0.085 | 1.051 | 0.158 | 0.147 | 1.074 |
| | | $\hat{\lambda}^o$ | 0.011 | 0.009 | 1.158 | 0.013 | 0.011 | 1.143 |
| | | $\hat{\rho}$ | 0.190 | 0.152 | 1.251 | 0.238 | 0.167 | 1.427 |
| | | $\hat{\sigma}^2$ | 0.141 | 0.111 | 1.265 | 0.168 | 0.108 | 1.552 |
| 0.40 | 0.00 | $\hat{\beta}^s$ | 0.446 | 0.440 | 1.013 | 0.450 | 0.437 | 1.029 |
| | | $\hat{\beta}^o$ | 0.242 | 0.239 | 1.012 | 0.250 | 0.250 | 0.998 |
| | | $\hat{\lambda}^s$ | 0.055 | 0.052 | 1.069 | 0.078 | 0.069 | 1.140 |
| | | $\hat{\lambda}^o$ | 0.062 | 0.056 | 1.124 | 0.104 | 0.090 | 1.165 |
| | | $\hat{\rho}$ | 0.210 | 0.202 | 1.040 | 0.205 | 0.202 | 1.014 |
| | | $\hat{\sigma}^2$ | 0.107 | 0.106 | 1.008 | 0.108 | 0.109 | 0.998 |
| 0.40 | 0.40 | $\hat{\beta}^s$ | 0.450 | 0.440 | 1.023 | 0.450 | 0.436 | 1.032 |
| | | $\hat{\beta}^o$ | 0.233 | 0.216 | 1.082 | 0.236 | 0.222 | 1.063 |
| | | $\hat{\lambda}^s$ | 0.054 | 0.051 | 1.077 | 0.071 | 0.064 | 1.102 |
| | | $\hat{\lambda}^o$ | 0.039 | 0.036 | 1.093 | 0.056 | 0.049 | 1.138 |
| | | $\hat{\rho}$ | 0.195 | 0.175 | 1.115 | 0.208 | 0.193 | 1.078 |
| | | $\hat{\sigma}^2$ | 0.115 | 0.104 | 1.109 | 0.117 | 0.105 | 1.114 |
| 0.40 | 0.85 | $\hat{\beta}^s$ | 0.432 | 0.422 | 1.025 | 0.448 | 0.432 | 1.035 |
| | | $\hat{\beta}^o$ | 0.248 | 0.201 | 1.234 | 0.278 | 0.197 | 1.412 |
| | | $\hat{\lambda}^s$ | 0.055 | 0.052 | 1.054 | 0.075 | 0.068 | 1.111 |
| | | $\hat{\lambda}^o$ | 0.010 | 0.009 | 1.142 | 0.013 | 0.012 | 1.149 |
| | | $\hat{\rho}$ | 0.166 | 0.138 | 1.200 | 0.211 | 0.155 | 1.366 |
| | | $\hat{\sigma}^2$ | 0.140 | 0.109 | 1.277 | 0.152 | 0.104 | 1.470 |
| 0.85 | 0.00 | $\hat{\beta}^s$ | 0.674 | 0.626 | 1.075 | 0.885 | 0.651 | 1.361 |
| | | $\hat{\beta}^o$ | 0.222 | 0.219 | 1.013 | 0.218 | 0.210 | 1.038 |
| | | $\hat{\lambda}^s$ | 0.036 | 0.034 | 1.073 | 0.032 | 0.029 | 1.116 |
| | | $\hat{\lambda}^o$ | 0.065 | 0.055 | 1.172 | 0.106 | 0.085 | 1.237 |
| | | $\hat{\rho}$ | 0.276 | 0.243 | 1.135 | 0.278 | 0.217 | 1.283 |
| | | $\hat{\sigma}^2$ | 0.102 | 0.102 | 1.001 | 0.097 | 0.096 | 1.008 |
| 0.85 | 0.40 | $\hat{\beta}^s$ | 0.712 | 0.651 | 1.093 | 0.844 | 0.646 | 1.308 |
| | | $\hat{\beta}^o$ | 0.224 | 0.214 | 1.047 | 0.224 | 0.212 | 1.055 |
| | | $\hat{\lambda}^s$ | 0.037 | 0.034 | 1.092 | 0.032 | 0.028 | 1.135 |
| | | $\hat{\lambda}^o$ | 0.040 | 0.037 | 1.085 | 0.059 | 0.053 | 1.127 |
| | | $\hat{\rho}$ | 0.247 | 0.220 | 1.123 | 0.271 | 0.229 | 1.183 |
| | | $\hat{\sigma}^2$ | 0.109 | 0.096 | 1.125 | 0.105 | 0.095 | 1.110 |
| 0.85 | 0.85 | $\hat{\beta}^s$ | 0.672 | 0.615 | 1.092 | 0.803 | 0.637 | 1.260 |
| | | $\hat{\beta}^o$ | 0.247 | 0.212 | 1.167 | 0.281 | 0.206 | 1.367 |
| | | $\hat{\lambda}^s$ | 0.036 | 0.033 | 1.097 | 0.031 | 0.027 | 1.127 |
| | | $\hat{\lambda}^o$ | 0.011 | 0.010 | 1.068 | 0.015 | 0.013 | 1.096 |
| | | $\hat{\rho}$ | 0.213 | 0.189 | 1.126 | 0.251 | 0.202 | 1.242 |
| | | $\hat{\sigma}^2$ | 0.121 | 0.099 | 1.218 | 0.142 | 0.099 | 1.432 |

*Note.* The columns with the label *ratio* provide ratios of the RMSEs of PMLE and FMLE.

## Appendix B Some Additional Notation

If $A$ is a matrix, $\mathrm{Diag}(A)$ indicates a square diagonal matrix with the diagonal elements of $A$ on the main diagonal of $\mathrm{Diag}(A)$, while $\mathrm{diag}(A)$ denotes a vector of the diagonal elements in $A$. If $a$ is a vector, then $\mathrm{Diag}(a)$ indicates a square diagonal matrix with the elements of vector $a$ on the main diagonal. If $\tau_1$ and $\tau_2$ are scalars, then $\mathrm{Diag}\{\tau_1, \tau_2\}$ denotes a diagonal matrix with $\tau_1$, $\tau_2$ on the main diagonal. For some constant $k$, $\mathrm{Diag}(\cdot)^k := (\mathrm{Diag}(\cdot))^k$. For some matrix A, $\mathrm{maxeig}(A)$ and $\mathrm{mineig}(A)$ denote the maximum and minimum eigenvalue of $A$, respectively. If $R$ and $R^{11}_{g,n}(\theta)$ are correlation matrices, then for notational convenience $\rho := R_{12}$ and $\rho^{11}_{g,n}(\theta) := R^{11}_{g12,n}(\theta)$. For any function $f(\theta)$, $\partial f(\theta)/\partial\theta\big|_{\theta=\theta_0}$ denotes the derivative of $f(\theta)$ evaluated at $\theta = \theta_0$. We use constants $C_1, C_2, \ldots$, which can be different in different places.

## Appendix C Some Theorems and Technical Lemmas

The appendix contains important theorems as well as several technical lemmas, which will be used later to prove lemmas and theorems in Appendices D and E. The proofs of Lemmas C.4–C.9 are provided in supplementary Appendix K.

**Theorem C.1** (Follows from Theorem 1 of Jenish and Prucha, 2012)**.** *Under Assumption 3, if*

*(i)* $\{Z_{g,n}\}_{g\in\mathcal{G}_n}$ *is uniformly* $L_1$*-NED on an* $\alpha$*-mixing random field* $\{\eta_{g,n}\}_{g\in\mathcal{G}_n}$*,*

*(ii)* $Z_{g,n}$ *is* $L_p$*-bounded uniformly in* $n \in \mathbb{N}$ *and* $g \in \mathcal{G}_n$*, for some* $p > 1$*,*

*(iii) the* $\alpha$*-mixing coefficients of the input process* $\{\eta_{g,n}\}_{g\in\mathcal{G}_n}$ *satisfy* $\bar{\alpha}(k, m, s) \leq (k+m)^\tau \hat{\alpha}(s)$, $\tau \geq 0$*, with some* $\hat{\alpha}(s) \to 0$ *as* $s \to \infty$*, such that* $\sum_{s=1}^{\infty} s^{2\tilde{d}-1}\hat{\alpha}(s) < \infty$*,*

*then* $n^{-1}\sum_{g\in\mathcal{G}_n}(Z_{g,n} - E[Z_{g,n}]) \xrightarrow{L_1} 0$.

**Theorem C.2** (Follows from Proposition 3 of Jenish and Prucha, 2012)**.** *Consider an* $\alpha$*-mixing random field* $\{\eta_{g,n}\}_{g\in\mathcal{G}_n}$, $\mathcal{F}_{g,n}(s) = \sigma(\{\eta_{\dot{g},n}\}_{\dot{g}\in\mathcal{G}_n} : d(g,\dot{g}) \leq s)$ *for* $s \in \mathbb{N}$*, and random vectors* $Z_{g,n}$ *and their transformations given by a family of functions* $h_{g,n} : \mathbb{R}^{K_Z} \to \mathbb{R}^{K_h}$*. Suppose that, for all* $(z, z^\bullet) \in \mathbb{R}^{K_Z} \times \mathbb{R}^{K_Z}$ *and all* $g \in \mathcal{G}_n$ *and* $n \in \mathbb{N}$*,*

*(i)* $\|h_{g,n}(z) - h_{g,n}(z^\bullet)\| \leq B_{g,n}(z, z^\bullet)\|z - z^\bullet\|$*, where* $B_{g,n}(z, z^\bullet) : \mathbb{R}^{K_Z} \times \mathbb{R}^{K_Z} \to \mathbb{R}_+$*,*

*(ii)* $\sup_s \|B^{(s)}_{g,n}\|_2 < \infty$*, where* $B^{(s)}_{g,n} = B_{g,n}(Z_{g,n}, Z^{(s)}_{g,n})$ *and* $Z^{(s)}_{g,n} = E[Z_{g,n}|\mathcal{F}_{g,n}(s)]$*,*

*(iii)* $\sup_s \left\| B^{(s)}_{g,n}\|Z_{g,n} - Z^{(s)}_{g,n}\| \right\|_r < \infty$ *for some* $r > 2$*,*

*(iv)* $\|h_{g,n}(Z_{g,n})\|_2 < \infty$*,*

*(v)* $\{Z_{g,n}\}_{g\in\mathcal{G}_n}$ *is* $L_2$*-NED of size* $-\lambda$ *on* $\{\eta_{g,n}\}_{g\in\mathcal{G}_n}$ *with scaling factors* $\{t_{g,n}\}_{g\in\mathcal{G}_n}$.

*Then* $h_{g,n}(Z_{g,n})$ *is* $L_2$*-NED of size* $-\lambda(r-2)/(2r-2)$ *on* $\{\eta_{g,n}\}_{g\in\mathcal{G}_n}$ *with scaling factors*

$$t_{g,n}^{(r-2)/(2r-2)}\sup_s \|B_{g,n}^{(s)}\|_2^{(r-2)/(2r-2)}\left\|B_{g,n}^{(s)}\|Z_{g,n}-Z_{g,n}^{(s)}\|\right\|_r^{r/(2r-2)}.$$

**Theorem C.3** (Follows from Corollary 1 of Jenish and Prucha, 2012)**.** *For* $K_Z$*-dimensional random vectors* $Z_{g,n}$, $g\in\mathcal{G}_n$, *let* $S_n=\sum_{g\in\mathcal{G}_n}Z_{g,n}$ *and* $\Psi_n=\mathrm{Var}\left(\sum_{g\in\mathcal{G}_n}Z_{g,n}\right)$. *Under Assumption 3, if*

*(i)* $\{Z_{g,n}\}_{g\in\mathcal{G}_n}$ *is a zero mean random field,*

*(ii)* $Z_{g,n}$ *is uniformly* $L_{2+\delta}$*-bounded, for some* $\delta>0$,

*(iii)* $\{Z_{g,n}\}_{g\in\mathcal{G}_n}$ *is* $L_2$*-NED random field on an* $\alpha$*-mixing random field* $\{\eta_{g,n}\}_{g\in\mathcal{G}_n}$ *with NED coefficients* $\psi(s)$ *and NED scaling factors* $\{t_{g,n}\}_{g\in\mathcal{G}_n}$,

*(iv) NED coefficients satisfy* $\sum_{s=1}^{\infty}s^{2\tilde{d}-1}\psi(s)<\infty$,

*(v) NED scaling factors satisfy* $\sup\limits_{n,g}t_{g,n}<\infty$,

*(vi) the* $\alpha$*-mixing coefficients of* $\{\eta_{g,n}\}_{g\in\mathcal{G}_n}$ *satisfy* $\bar{\alpha}(k,m,s)\le(k+m)^{\tau}\hat{\alpha}(s)$, *for some* $\tau\ge 0$ *and* $\hat{\alpha}(s)\to 0$ *as* $s\to\infty$, *such that for some* $\delta>0$, $\sum_{s=1}^{\infty}s^{2\tilde{d}(\tau_*+1)-1}\hat{\alpha}^{\delta/(4+2\delta)}(s)<\infty$, *where* $\tau_*=\delta\tau/(2+\delta)$,

*(vii)* $\inf_n \frac{1}{n}\mathrm{mineig}(\Psi_n)>0$,

*then* $\Psi_n^{-1/2}S_n\xrightarrow{d}\mathcal{N}(0,I_{K_Z})$ *as* $n\to\infty$.

**Lemma C.4.** *Let* $v:=v(\theta)$ *be a 2-dimensional vector,* $R:=R(\theta)$ *be a* $2\times 2$ *dimensional correlation matrix with the off-diagonal element* $\rho:=\rho(\theta)$, $|\rho|<1$, *and* $P(z,R)=z'R^{-1}(\theta)z$, *where* $z$ *is a 2-dimensional vector. Then*

$$\begin{aligned}
\frac{\partial\ln\Phi_2(v,R)}{\partial v}&=\xi(v,R),\\
\frac{\partial\ln\Phi_2(v,R)}{\partial\theta}&=\xi_1(v,R)\frac{\partial v_1}{\partial\theta}+\xi_2(v,R)\frac{\partial v_2}{\partial\theta}-\frac{1}{2}\left(\frac{\partial\ln|R|}{\partial\theta}+E_V\left[\frac{\partial P(V,R)}{\partial\theta}\Big|V\le v\right]\right),\\
\frac{\partial^2\ln\Phi_2(v,R)}{\partial\theta\partial\theta'}&=\xi_1(v,R)\frac{\partial^2 v_1}{\partial\theta\partial\theta'}+\xi_2(v,R)\frac{\partial^2 v_2}{\partial\theta\partial\theta'}+\kappa(v,R)\left(\frac{\partial v_1}{\partial\theta}\frac{\partial v_2}{\partial\theta'}+\frac{\partial v_2}{\partial\theta}\frac{\partial v_1}{\partial\theta'}\right)\\
&\quad-\frac{1}{2}\left(A(v,R)+B(v,R)+A'(v,R)+B'(v,R)+E_V[G(V,R)|V\le v]\right)\\
&\quad-\frac{1}{\Phi_2^2(v,R)}\frac{\partial\Phi_2(v,R)}{\partial\theta}\frac{\partial\Phi_2(v,R)}{\partial\theta'},
\end{aligned}$$

*where* $E_V$ *denotes expectation taken only with respect to the variable* $V$ *in its subscript and (keeping the* *dependence on $\theta$ implicit)*

$$\xi(v,R) = (\xi_1(v,R), \xi_2(v,R))' = \left( \frac{\phi(v_1)\Phi\left(\frac{v_2-\rho v_1}{\sqrt{1-\rho^2}}\right)}{\Phi_2(v,R)}, \frac{\phi(v_2)\Phi\left(\frac{v_1-\rho v_2}{\sqrt{1-\rho^2}}\right)}{\Phi_2(v,R)} \right)', \tag{C.1}$$

$$\kappa(v,R) = \frac{\phi_2(v,R)}{\Phi_2(v,R)}, \tag{C.2}$$

$$A(v,R) = \xi_1(v,R)\frac{\partial v_1}{\partial\theta}\left(\frac{\partial \ln|R|}{\partial\theta'} + E_{\widetilde{V}_2}\left[\frac{\partial P((v_1,\widetilde{V}_2)',R)}{\partial\theta'}\Big|\widetilde{V}_2 \le v_2\right]\right),$$

$$B(v,R) = \xi_2(v,R)\frac{\partial v_2}{\partial\theta}\left(\frac{\partial \ln|R|}{\partial\theta'} + E_{\widetilde{V}_1}\left[\frac{\partial P((\widetilde{V}_1,v_2)',R)}{\partial\theta'}\Big|\widetilde{V}_1 \le v_1\right]\right),$$

$$G(z,R) = \frac{\partial^2 \ln|R|}{\partial\theta\partial\theta'} + \frac{\partial^2 P(z,R)}{\partial\theta\partial\theta'} - \frac{1}{2}\left(\frac{\partial \ln|R|}{\partial\theta} + \frac{\partial P(z,R)}{\partial\theta}\right)\left(\frac{\partial \ln|R|}{\partial\theta'} + \frac{\partial P(z,R)}{\partial\theta'}\right) \text{ and} \tag{C.3}$$

$$V \sim \mathcal{N}(0,R),\ \widetilde{V}_1 \sim \mathcal{N}(\rho v_2, 1-\rho^2),\ \widetilde{V}_2 \sim \mathcal{N}(\rho v_1, 1-\rho^2).$$

**Lemma C.5.** *Let $v$ and $R$ be a 2-dimensional vector and a $2\times 2$ dimensional correlation matrix with the off-diagonal element $\rho$, $|\rho|<1$, respectively. Then for some constants $C_1$, $C_2$, $C_3 > 0$,*

$$\frac{\phi(v_1)\Phi\left(\frac{v_2-\rho v_1}{\sqrt{1-\rho^2}}\right)}{\Phi_2(v,R)} \le C_1(1-|\rho|)^{-7}\left((|v_1|+|v_2|+C_2)^8 + \left(1-\Phi\left((1-|\rho|)^{-1/2}\right)\right)^{-2}\right),$$

$$\frac{\phi_2(v,R)}{\Phi_2(v,R)} \le C_1(1-|\rho|)^{-3}\left((|v_1|+|v_2|+C_2)^2 + \left(1-\Phi\left((1-|\rho|)^{-1/2}\right)\right)^{-2}\right),$$

$$\Phi_2(0,R) \ge C_3(1-|\rho|)^{1/2}.$$

**Lemma C.6.** *Let $\theta$ be a $p\times 1$ vector, $f(\theta)$ an $n\times 1$ vector, and $F(\theta)$ an $n\times n$ symmetric matrix. Then*

$$\frac{\partial|F(\theta)|}{\partial\theta} = |F(\theta)|K(\theta)\operatorname{vec}F^{-1}(\theta),$$

$$\frac{\partial^2|F(\theta)|}{\partial\theta\partial\theta'} = K(\theta)\operatorname{vec}F^{-1}(\theta)\frac{\partial|F(\theta)|}{\partial\theta'} + |F(\theta)|\left((\operatorname{vec}F^{-1}(\theta))'\otimes I_p\right)\frac{\partial \operatorname{vec}K(\theta)}{\partial\theta'} + |F(\theta)|K(\theta)\frac{\partial \operatorname{vec}F^{-1}(\theta)}{\partial\theta'},$$

$$\frac{\partial(f'(\theta)F^{-1}(\theta)f(\theta))}{\partial\theta} = 2L(\theta)F^{-1}(\theta)f(\theta) + M(\theta)(f(\theta)\otimes f(\theta)),$$

$$\frac{\partial^2(f'(\theta)F^{-1}(\theta)f(\theta))}{\partial\theta\partial\theta'} = 2(f'(\theta)F^{-1}(\theta)\otimes I_p)\frac{\partial \operatorname{vec}L(\theta)}{\partial\theta'} + 2(f'(\theta)\otimes L(\theta))\frac{\partial \operatorname{vec}F^{-1}(\theta)}{\partial\theta'} + (f'(\theta)\otimes f'(\theta)\otimes I_p)\frac{\partial \operatorname{vec}M(\theta)}{\partial\theta'} + (2L(\theta)F^{-1}(\theta) + M(\theta)(\mathrm{K}_{1n}\otimes I_n)[(I_n\otimes f(\theta)) + (f(\theta)\otimes I_n)])\frac{\partial f(\theta)}{\partial\theta'},$$

*where*

$$K(\theta) = \left(\frac{\partial \operatorname{vec} F(\theta)}{\partial \theta'}\right)', \; L(\theta) = \left(\frac{\partial f(\theta)}{\partial \theta'}\right)', \; M(\theta) = \left(\frac{\partial \operatorname{vec} F^{-1}(\theta)}{\partial \theta'}\right)',$$

*and* $\mathrm{K}_{1n}$ *is the commutation matrix.*[17]

**Lemma C.7.** *Let* $A$ *and* $B$ *be* $m \times n$ *and* $p \times q$ *matrices. Then* $\|A \otimes B\| = \|A\|\|B\|$.

**Lemma C.8.** *Let* $X \sim \mathcal{N}(0, R)$*, where* $R$ *is a* $2 \times 2$ *dimensional correlation matrix with the off-diagonal element* $\rho$, $|\rho| < 1$. *Then for a* 2*-dimensional vector of constants* $v = (v_1, v_2)'$,

$$E[XX'|X \le v] = R - v_1 \xi_1(v, R) A_1(R) - v_2 \xi_2(v, R) A_2(R) + (1 - \rho^2)\kappa(v, R) A_3(R),$$

*where* $\xi(v, R)$ *and* $\kappa(v, R)$ *are defined in* (C.1) *and* (C.2)*, respectively, and*

$$A_1(R) = \begin{pmatrix} 1 & \rho \\ \rho & \rho^2 \end{pmatrix}, \; A_2(R) = \begin{pmatrix} \rho^2 & \rho \\ \rho & 1 \end{pmatrix}, \; A_3(R) = \begin{pmatrix} \rho & 1 \\ 1 & \rho \end{pmatrix}. \tag{C.4}$$

**Lemma C.9.** *If for some* $p \ge 1$, $\|X_{i,n} - E[X_{i,n}|\mathcal{F}_{i,n}(s)]\|_{2p} \le t^X_{i,n}\psi^X(s)$ *and* $\|Y_{i,n} - E[Y_{i,n}|\mathcal{F}_{i,n}(s)]\|_{2p} \le t^Y_{i,n}\psi^Y(s)$*, then* $\|X_{i,n}Y_{i,n} - E[X_{i,n}Y_{i,n}|\mathcal{F}_{i,n}(s)]\|_p \le t_{i,n}\psi(s)$*, where* $t_{i,n} = \max\{\|X_{i,n}\|_{2p} t^Y_{i,n},\ \|Y_{i,n}\|_{2p} t^X_{i,n},\ t^X_{i,n} t^Y_{i,n}\}$ *and* $\psi(s) = \psi^X(s) + \psi^Y(s) + \psi^X(s)\psi^Y(s)$*. Specifically, if* $\{X_{i,n}\}_{i=1}^n$ *and* $\{Y_{i,n}\}_{i=1}^n$ *are uniformly* $L_{2p}$*-NED, then* $\{X_{i,n}Y_{i,n}\}_{i=1}^n$ *is uniformly* $L_p$*-NED.*

# Appendix D Some Useful Lemmas

The appendix contains several lemmas that establish the uniform ($L_p$-) bounds and the NED property of the random variables in the studied sample selection models. The proofs of Lemmas D.1–D.6 are provided in supplementary Appendix L.

**Lemma D.1.**

*(i) Under Assumptions 1(ii), 2(i), 6, and 7,* $\inf_{n,g}\inf_{\theta\in\Theta} |\Omega^b_{g,n}(\theta)| > 0$ *and* $\inf_{n,i}\inf_{\theta\in\Theta} \Omega^b_{ii,n}(\theta) > 0$*, where* $b \in \{ss, oo\}$.

*(ii) Under Assumptions 1(ii) and 7,* $\|\Omega^c_{g,n}(\theta)\|$ *and* $\left\|\partial \operatorname{vec} \Omega^c_{g,n}(\theta)/\partial\theta'\right\|$ *are uniformly bounded in* $n \in \mathbb{N}$, $g \in \mathcal{G}_n$*, and* $\theta \in \Theta$*, where* $c \in \{ss, so, oo\}$.

[17] Let $A$ be an $m \times n$ matrix. Then there exists a unique $mn \times mn$ permutation matrix which transforms $\operatorname{vec} A$ into $\operatorname{vec} A'$, i.e. $\mathrm{K}_{mn} \operatorname{vec} A = \operatorname{vec} A'$.

*(iii) Under Assumptions 1(ii), 6, and 7,* $\|\Omega_{g,n}^{b-1}(\theta)\|$, $\|\partial \operatorname{vec} \Omega_{g,n}^{b-1}(\theta)/\partial\theta'\|$, *and* $\|\partial|\Omega_{g,n}^{b}(\theta)|/\partial\theta\|$ *are uniformly bounded in* $n \in \mathbb{N}$, $g \in \mathcal{G}_n$, *and* $\theta \in \Theta$, *where* $b \in \{ss, oo\}$.

**Lemma D.2.**

*(i) Under Assumption 6,* $\inf_{n,g} \inf_{\theta\in\Theta} \min_{j=1,2} \Sigma_{gjj,n}^{11}(\theta) > 0$ *and* $\sup_{n,g} \sup_{\theta\in\Theta} |\rho_{g,n}^{11}(\theta)| < 1$.

*(ii) Under Assumptions 1(ii), 6, and 7,* $\|R_{g,n}^{11}(\theta)\|$, $\|\partial|R_{g,n}^{11}(\theta)|/\partial\theta\|$, *and* $\left\|\partial \operatorname{vec} R_{g,n}^{11^{-1}}(\theta)/\partial\theta'\right\|$ *are uniformly bounded in* $n \in \mathbb{N}$, $g \in \mathcal{G}_n$, *and* $\theta \in \Theta$.

**Lemma D.3.** *Under Assumptions 1(ii), 2(i), 4(ii) or 9(ii), 6, and 7,* $\sup_{\theta\in\Theta} \|S_{g\cdot,n}^{b}(\lambda^b) X_n^b \beta^b\|$, $b \in \{s, o\}$, $\sup_{\theta\in\Theta} \|z_{g,n}(\theta)\|$, $\sup_{\theta\in\Theta} \|\mu_{g,n}^{11}(\theta)\|$, *and* $\sup_{\theta\in\Theta} \|v_{g,n}^{11}(\theta)\|$ *are* $L_p$*-bounded uniformly in* $n \in \mathbb{N}$ *and* $g \in \mathcal{G}_n$ *for any* $p$ *in Assumption 4(ii) or 9(ii).*

**Lemma D.4.** *Under Assumptions 1(ii), 2(i), 4(ii) or 9(ii), 6, and 7,* $\sup_{\theta\in\Theta} \|\partial z_{g,n}(\theta)/\partial\theta'\|$ *and* $\sup_{\theta\in\Theta} \|\partial v_{g,n}^{11}(\theta)/\partial\theta'\|$ *are* $L_p$*-bounded uniformly in* $n \in \mathbb{N}$ *and* $g \in \mathcal{G}_n$ *for any* $p$ *in Assumption 4(ii) or 9(ii).*

**Lemma D.5.** *Under Assumptions 1(ii), 2(i), 4(ii), 5, and 7,* $\{d_{g,n}^{11}\}_{g\in\mathcal{G}_n}$, $\{z_{g,n}(\theta_n)\}_{g\in\mathcal{G}_n}$, *and* $\{v_{g,n}^{11}(\theta_n)\}_{g\in\mathcal{G}_n}$ *are uniformly* $L_2$*-NED on random field* $\{\eta_{g,n} = (X_{g\cdot,n}^s, X_{g\cdot,n}^o, u_{g,n}^s, u_{g,n}^o)\}_{g\in\mathcal{G}_n}$ *with NED coefficients bounded by* $\psi^{1/6}(s)$, *where* $\psi(s)$ *is defined in Assumption 5 and* $\theta_n \to \theta$, $\theta_n, \theta \in \Theta$.

**Lemma D.6.** *Let sequence* $\theta_n \in \{\theta : \|\theta - \theta_0\| < n^{-1/2}M\}$ *for some* $M > 0$ *and* $n \geq n_0$, $n_0 \in \mathbb{N}$. *If responses* $y_{g,n}^{*s}$ *and* $\hat{y}_{g,n}^{*s}$ *follow the data generating process* (1) *with the parameter vector* $\theta_0$ *and with the parameter vector* $\theta_n$, *respectively, then under Assumptions 1(ii), 2(i), 4(ii), 5, 7, 11, and 13 it holds that* $\|\hat{y}_{g,n}^{*s} - y_{g,n}^{*s}\|_2 = O(n^{-1/2})$ *as* $n \to \infty$ *and* $\{n^{1/2}(\hat{y}_{g,n}^{*s} - y_{g,n}^{*s})\}_{g\in\mathcal{G}_n}$ *is* $L_2$*-NED on random field* $\{\eta_{g,n} = (X_{g\cdot,n}^s, X_{g\cdot,n}^o, u_{g,n}^s, u_{g,n}^o)\}_{g\in\mathcal{G}_n}$ *with NED coefficients bounded by* $\max\{\psi^s(s), \tilde{\psi}^s(s)\}$, *where* $\psi^s(s)$ *and* $\tilde{\psi}^s(s)$ *are defined in Assumptions 5 and 13, respectively.*

# Appendix E Proofs of the Asymptotic Results

In the following proofs of the main theorems, we verify various regularity conditions such as $L_p$-boundedness of the likelihood function $\sum_{a\in\mathcal{A}} d_{g,n}^a f_{g,n}^a(\theta)$ and its derivatives only for the terms corresponding to $a = 11$. The terms corresponding to the other values of $a$ can be verified analogously since they are special cases of $a = 11$: the inspection of the likelihood function (4) as well as the verification of the first-order conditions in Appendix G reveal that the forms of terms with $a = 10$, $a = 01$, or $a = 00$ correspond to the form of $a = 11$ with the bivariate density function replaced by the corresponding marginal densities or one.

***Proof of Theorem 1:*** General consistency results are given, for example, in Theorem 2.1 of Newey and McFadden (1994) or Lemma 3.1 of Pötscher and Prucha (1997) in the case when $Q_0(\theta) = \lim_{n\to\infty} E[Q_n(\theta)]$ exists or not, respectively. By the latter result, it is sufficient to verify that (i) $E[Q_n(\theta)]$ is uniquely maximized at $\theta_0$ (at least for a sufficiently large $n$) and (ii) $Q_n(\theta)$ converges uniformly in probability to $E[Q_n(\theta)]$ to prove the claim of the theorem. Since we have already assumed the first condition in Assumption 8, it thus remains to show that the last condition is satisfied. In order to prove uniform convergence in (ii), we apply Theorem 2 of Jenish and Prucha (2009), which requires the uniform $L_p$-boundedness (LB), $p > 1$, and $L_0$-stochastic equicontinuity (SE) of the individual likelihood terms as well as the pointwise convergence (PC) in probability; see the following paragraphs. As the bounds constructed to verify the uniform $L_p$-boundedness and $L_0$-stochastic equicontinuity are uniform in $g \in \mathcal{G}_n$, $n \in \mathbb{N}$, and $\theta \in \Theta$, it follows that the whole likelihood function $Q_n(\theta)$ also satisfies these conditions (once they are verified) and that PMLE is consistent.

LB: proof of $\sup\limits_{n,g} E\left[\sup\limits_{\theta\in\Theta} |\sum_{a\in\mathcal{A}} d^a_{g,n} f^a_{g,n}(\theta)|\right]^p < \infty$, for some $p > 1$

$$E\left[\sup_{\theta\in\Theta} |\sum_{a\in\mathcal{A}} d^a_{g,n} f^a_{g,n}(\theta)|\right]^p \leq E\left[\sum_{a\in\mathcal{A}} \sup_{\theta\in\Theta} |d^a_{g,n} f^a_{g,n}(\theta)|\right]^p \leq 4^{p-1} \sum_{a\in\mathcal{A}} E\left[\sup_{\theta\in\Theta} |d^a_{g,n} f^a_{g,n}(\theta)|\right]^p \\ \leq 4^{p-1} \sum_{a\in\mathcal{A}} E\left[\sup_{\theta\in\Theta} |f^a_{g,n}(\theta)|\right]^p \tag{E.1}$$

uniformly in $n \in \mathbb{N}$ and $g \in \mathcal{G}_n$, where the first and second inequalities follow by the triangle and Loève's $c_r$-inequalities, respectively, whereas the last inequality follows by noting that $d^a_{g,n} \in \{0, 1\}$.

We will show that $\sup\limits_{n,g} E[\sup_{\theta\in\Theta} |f^{11}_{g,n}(\theta)|]^p < \infty$, while the boundedness of the other terms can be proven in a similar way. By the definitions of $f^{11}_{g,n}(\theta)$ in (4) and of the multivariate normal density function,

$$E\left[\sup_{\theta\in\Theta} |f^{11}_{g,n}(\theta)|\right]^p \leq 4^{p-1} \left( |\ln 2\pi|^p + \sup_{\theta\in\Theta} |\ln |\Omega^{oo}_{g,n}(\theta)||^p + \sup_{\theta\in\Theta} \|\Omega^{oo-1}_{g,n}(\theta)\|^p E\left[\sup_{\theta\in\Theta} \|z_{g,n}(\theta)\|\right]^{2p} \right. \\ \left. + E\left[\sup_{\theta\in\Theta} |\ln \Phi_2(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta))|\right]^p \right), \tag{E.2}$$

where the result follows by the triangle and Loève's $c_r$-inequalities. The second and third terms are uniformly bounded by Lemmas D.1 and D.3. Hence, only the last term has to be shown to be uniformly bounded. Let $\xi(\cdot)$ be defined in the same way as in (C.1), Lemma C.4, with correlation matrix $R^{11}_{g,n}(\theta)$ and correlation coefficient $\rho^{11}_{g,n}(\theta)$, which is the off-diagonal element of $R^{11}_{g,n}(\theta)$. Then by the elementwise mean value theorem, there exists $\tilde{v}^{11}_{g,n}(\theta)$ with elements between 0 and $v^{11}_{g,n}(\theta)$ and constants $C_1, \ldots, C_8 > 0$ with $C_3, C_6 \geq 1$ such that

$$
\begin{aligned}
&|\ln\Phi_2(v^{11}_{g,n}(\theta),R^{11}_{g,n}(\theta))| \\
&\le |\ln\Phi_2(0,R^{11}_{g,n}(\theta))| + \left|\frac{\partial \ln\Phi_2(\tilde v^{11}_{g,n}(\theta),R^{11}_{g,n}(\theta))}{\partial v'} v^{11}_{g,n}(\theta)\right| = |\ln\Phi_2(0,R^{11}_{g,n}(\theta))| + |\xi^{'}(\tilde v^{11}_{g,n}(\theta),R^{11}_{g,n}(\theta))v^{11}_{g,n}(\theta)| \\
&\le |\ln\Phi_2(0,R^{11}_{g,n}(\theta))| + \sum_{j=1}^{2}\xi_j(\tilde v^{11}_{g,n}(\theta),R^{11}_{g,n}(\theta))|v^{11}_{gj,n}(\theta)| \qquad \text{(E.3)} \\
&\le |\ln(C_1(1-|\rho^{11}_{g,n}(\theta)|)^{1/2})| \\
&\quad + C_2(1-|\rho^{11}_{g,n}(\theta)|)^{-7}\left((|\tilde v^{11}_{g1,n}(\theta)|+|\tilde v^{11}_{g2,n}(\theta)|+C_3)^8 + \left(1-\Phi\left((1-|\rho^{11}_{g,n}(\theta)|)^{-1/2}\right)\right)^{-2}\right)\sum_{j=1}^{2}|v^{11}_{gj,n}(\theta)| \\
&\le C_4 + C_5\left((|\tilde v^{11}_{g1,n}(\theta)|+|\tilde v^{11}_{g2,n}(\theta)|+C_3)^8 + C_6\right)\sum_{j=1}^{2}|v^{11}_{gj,n}(\theta)| \\
&\le C_4 + C_5\left((|v^{11}_{g1,n}(\theta)|+|v^{11}_{g2,n}(\theta)|+C_3)^8 + C_6\right)\sum_{j=1}^{2}|v^{11}_{gj,n}(\theta)| \\
&\le C_4 + 2C_5C_6(|v^{11}_{g1,n}(\theta)|+|v^{11}_{g2,n}(\theta)|+C_3)^9 = C_4 + 2C_5C_6(\|v^{11}_{g,n}(\theta)\|_1 + C_3)^9 \\
&\le C_4 + 2^9C_5C_6(\|v^{11}_{g,n}(\theta)\|_1^9 + C_3^9) \le C_7 + C_8\|v^{11}_{g,n}(\theta)\|^9,
\end{aligned}
$$

where Lemma C.4 implies the first equality, the third inequality is implied by Lemma C.5, whereas the fourth inequality follows from Lemma D.2. The conclusion follows by the equivalence of vector norms on finite dimensional vector spaces. Given this result,

$$
E\left[\sup_{\theta\in\Theta}|\ln\Phi_2(v^{11}_{g,n}(\theta),R^{11}_{g,n}(\theta))|\right]^p \le E\left[\sup_{\theta\in\Theta}\left(C_7 + C_8\|v^{11}_{g,n}(\theta)\|^9\right)\right]^p \le 2^{p-1}\left(C_7^p + C_8^pE\left[\sup_{\theta\in\Theta}\|v^{11}_{g,n}(\theta)\|\right]^{9p}\right) < \infty \qquad \text{(E.4)}
$$

uniformly in $n\in\mathbb{N}$ and $g\in\mathcal{G}_n$ by Lemma D.3.

SE: proof that $\sum_{a\in\mathcal{A}} d^a_{g,n}f^a_{g,n}(\theta)$ is $L_0$-stochastically equicontinuous

The $L_0$-stochastic equicontinuity will be verified using Proposition 1 of Jenish and Prucha (2009). To apply it, we have to show that the individual likelihood terms are Lipschitz functions in parameters: for any $\theta,\theta^\bullet\in\Theta$,

$$
\begin{aligned}
\left|\sum_{a\in\mathcal{A}} d^a_{g,n}f^a_{g,n}(\theta) - \sum_{a\in\mathcal{A}} d^a_{g,n}f^a_{g,n}(\theta^\bullet)\right| &\le \sum_{a\in\mathcal{A}} d^a_{g,n}|f^a_{g,n}(\theta)-f^a_{g,n}(\theta^\bullet)| \le \sum_{a\in\mathcal{A}}|f^a_{g,n}(\theta)-f^a_{g,n}(\theta^\bullet)| \\
&\le \sum_{a\in\mathcal{A}}\left\|\frac{\partial f^a_{g,n}(\tilde\theta)}{\partial\theta}\right\|\|\theta-\theta^\bullet\| \le \sum_{a\in\mathcal{A}}\sup_{\theta\in\Theta}\left\|\frac{\partial f^a_{g,n}(\theta)}{\partial\theta}\right\|\cdot\|\theta-\theta^\bullet\|,
\end{aligned}
$$

where we used the elementwise mean value theorem with elements of $\tilde{\theta}$ being between elements of $\theta$ and $\theta^{\bullet}$. It thus suffices to show that $\sup_{n,g} E\left[\sum_{a\in\mathcal{A}} \sup_{\theta\in\Theta} \|\partial f^a_{g,n}(\theta)/\partial\theta\|\right]^p < \infty$ for some $p \geq 1$. Similarly to (E.1), Loève's $c_r$-inequality implies that it is enough to prove that all individual terms are bounded, that is, $\sup_{n,g} E\left[\sup_{\theta\in\Theta} \|\partial f^a_{g,n}(\theta)/\partial\theta\|\right]^p < \infty$ for all $a \in \mathcal{A}$. As before, we establish this result for $\partial f^{11}_{g,n}(\theta)/\partial\theta$, while the boundedness of the other terms can be proven in a similar way:

$$
\begin{aligned}
&E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial f^{11}_{g,n}(\theta)}{\partial\theta}\right\|\right]^p \\
&= E\left[\sup_{\theta\in\Theta}\left\|-\frac{1}{2}\left(\frac{1}{|\Omega^{oo}_{g,n}(\theta)|}\frac{\partial|\Omega^{oo}_{g,n}(\theta)|}{\partial\theta} + \frac{\partial(z'_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)z_{g,n}(\theta))}{\partial\theta}\right) + \frac{\partial \ln\Phi_2(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta))}{\partial\theta}\right\|\right]^p \\
&\leq 3^{p-1}\left(\sup_{\theta\in\Theta}\left(\frac{1}{|\Omega^{oo}_{g,n}(\theta)|}\left\|\frac{\partial|\Omega^{oo}_{g,n}(\theta)|}{\partial\theta}\right\|\right)^p + E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial(z'_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)z_{g,n}(\theta))}{\partial\theta}\right\|\right]^p\right. \\
&\qquad\left. + E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial\ln\Phi_2(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta))}{\partial\theta}\right\|\right]^p\right).
\end{aligned} \tag{E.5}
$$

The first term on the right hand side is uniformly bounded by Lemma D.1. To bound the second term on the right hand side of (E.5), we apply Lemma C.6:

$$
\begin{aligned}
E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial(z'_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)z_{g,n}(\theta))}{\partial\theta}\right\|\right]^p &= E\left[\sup_{\theta\in\Theta}\left\|2\left(\frac{\partial z_{g,n}(\theta)}{\partial\theta'}\right)'\Omega^{oo-1}_{g,n}(\theta)z_{g,n}(\theta)\right.\right. \\
&\qquad\left.\left. + \left(\frac{\partial\operatorname{vec}\Omega^{oo-1}_{g,n}(\theta)}{\partial\theta'}\right)'(z_{g,n}(\theta)\otimes z_{g,n}(\theta))\right\|\right]^p \\
&\leq 2^{2p-1}\left(\sup_{\theta\in\Theta}\|\Omega^{oo-1}_{g,n}(\theta)\|^p E\left[\sup_{\theta\in\Theta}\left(\left\|\frac{\partial z_{g,n}(\theta)}{\partial\theta'}\right\|\|z_{g,n}(\theta)\|\right)\right]^p\right. \\
&\qquad\left. + \sup_{\theta\in\Theta}\left\|\frac{\partial\operatorname{vec}\Omega^{oo-1}_{g,n}(\theta)}{\partial\theta'}\right\|^p E\left[\sup_{\theta\in\Theta}\|z_{g,n}(\theta)\|\right]^{2p}\right) < \infty
\end{aligned} \tag{E.6}
$$

uniformly in $n \in \mathbb{N}$ and $g \in \mathcal{G}_n$, where the first inequality is implied by Lemma C.7 and Loève's $c_r$-inequality. The conclusion follows by applying the Cauchy-Schwartz inequality to $E\left[\sup_{\theta\in\Theta}(\|\partial z_{g,n}(\theta)/\partial\theta'\|\|z_{g,n}(\theta)\|)\right]^p$ and observing that $E\left[\sup_{\theta\in\Theta}\|\partial z_{g,n}(\theta)/\partial\theta'\|\right]^{2p}$ and $E\left[\sup_{\theta\in\Theta}\|z_{g,n}(\theta)\|\right]^{2p}$ are uniformly bounded by Lemmas D.4 and D.3, respectively, while the norms of $\Omega^{oo-1}_{g,n}(\theta)$ and $\partial\operatorname{vec}\Omega^{oo-1}_{g,n}(\theta)/\partial\theta'$ are uniformly bounded by Lemma D.1.

Finally, by Lemma C.4, the last term in (E.5) can be bounded (all symbols defined in Lemma C.4 are again indexed by the subscripts $g$ and $n$ and superscript 11):

$$
\begin{aligned}
E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial \ln \Phi_2(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta))}{\partial\theta}\right\|\right]^p &= E\left[\sup_{\theta\in\Theta}\left\|\sum_{j=1}^{2}\xi_j(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta))\frac{\partial v^{11}_{gj,n}(\theta)}{\partial\theta}\right.\right. \\
&\quad \left.\left. -\frac{1}{2}\left(\frac{\partial \ln|R^{11}_{g,n}(\theta)|}{\partial\theta} + E_V\left[\frac{\partial P(V, R^{11}_{g,n}(\theta))}{\partial\theta}|V \le v^{11}_{g,n}(\theta)\right]\right)\right\|\right]^p \\
&\le 4^{p-1}\left(\sum_{j=1}^{2} E\left[\sup_{\theta\in\Theta}\left\|\xi_j(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta))\frac{\partial v^{11}_{gj,n}(\theta)}{\partial\theta}\right\|\right]^p\right. \\
&\quad \left. + \sup_{\theta\in\Theta}\left\|\frac{1}{|R^{11}_{g,n}(\theta)|}\frac{\partial|R^{11}_{g,n}(\theta)|}{\partial\theta}\right\|^p + E\left[\sup_{\theta\in\Theta}\left\|E_V\left[\frac{\partial P(V, R^{11}_{g,n}(\theta))}{\partial\theta}|V \le v^{11}_{g,n}(\theta)\right]\right\|\right]^p\right)
\end{aligned} \tag{E.7}
$$

uniformly in $n \in \mathbb{N}$ and $g \in \mathcal{G}_n$, where $P(v, R^{11}_{g,n}(\theta)) = v' R^{11^{-1}}_{g,n}(\theta) v$ and $V \sim \mathcal{N}(0, R^{11}_{g,n}(\theta))$.[18]

Now we will prove that each term in (E.7) is uniformly bounded. By the Cauchy-Schwartz inequality,

$$
\begin{aligned}
E\left[\sup_{\theta\in\Theta}\left\|\xi_j(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta))\frac{\partial v^{11}_{gj,n}(\theta)}{\partial\theta}\right\|\right]^p &\le \sqrt{E\left[\sup_{\theta\in\Theta}\xi_j(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta))\right]^{2p} E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial v^{11}_{gj,n}(\theta)}{\partial\theta'}\right\|\right]^{2p}} \\
&\le \sqrt{E\left[\sup_{\theta\in\Theta}\xi_j(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta))\right]^{2p} E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial v^{11}_{g,n}(\theta)}{\partial\theta'}\right\|\right]^{2p}},
\end{aligned} \tag{E.8}
$$

for $j = 1, 2$. In the same way as in (E.3), $\xi_j(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta)) \le C_1 + C_2\|v^{11}_{g,n}(\theta)\|^8$ for some constants $C_1, C_2 > 0$.[19] Thus for (E.8) to be uniformly bounded, it is enough to show that $E[\sup_{\theta\in\Theta}\|v^{11}_{g,n}(\theta)\|]^{16p}$ and $E[\sup_{\theta\in\Theta}\|\partial v^{11}_{g,n}(\theta)/\partial\theta'\|]^{2p}$ are uniformly bounded, which is the case by Lemmas D.3 and D.4, respectively. The second term on the right hand side of (E.7) is uniformly bounded because $\inf_{n,g}|R^{11}_{g,n}(\theta)| = \inf_{n,g}(1 - \rho^{11^2}_{g,n}(\theta)) > 0$ and $\sup_{n,g}\sup_{\theta\in\Theta}\|\partial|R^{11}_{g,n}(\theta)|/\partial\theta\| < \infty$ by Lemma D.2. Regarding the last term in (E.7), it is not difficult to see from Lemma C.6 on the first order derivative of a quadratic form and from Lemma D.2

[18]We use $V$ instead of $V^{11}_{g,n}(\theta)$ in order to simplify the notation.

[19]Note that the numbering of constants is renewed for each part of the proof.

uniformly bounding $\partial \operatorname{vec} R_{g,n}^{11^{-1}}(\theta)/\partial\theta'$ that

$$\begin{aligned}
&\left\| E_V\left[\frac{\partial P(V,R_{g,n}^{11}(\theta))}{\partial\theta}|V \le v_{g,n}^{11}(\theta)\right]\right\| = \left\| E_V\left[\left(\frac{\partial \operatorname{vec} R_{g,n}^{11^{-1}}(\theta)}{\partial\theta'}\right)'(V\otimes V)\big|V \le v_{g,n}^{11}(\theta)\right]\right\| \\
&\le \left\|\frac{\partial \operatorname{vec} R_{g,n}^{11^{-1}}(\theta)}{\partial\theta'}\right\| \left\|E_V[V\otimes V|V\le v_{g,n}^{11}(\theta)]\right\| = \left\|\frac{\partial \operatorname{vec} R_{g,n}^{11^{-1}}(\theta)}{\partial\theta'}\right\| \left\|E_V[VV'|V\le v_{g,n}^{11}(\theta)]\right\| \\
&\le C_3\left\|E_V[VV'|V\le v_{g,n}^{11}(\theta)]\right\| \qquad\qquad \text{(E.9)} \\
&= C_3\|R_{g,n}^{11}(\theta) - v_{g1,n}^{11}(\theta)\xi_1(v_{g,n}^{11}(\theta),R_{g,n}^{11}(\theta))A_1(R_{g,n}^{11}(\theta)) - v_{g2,n}^{11}(\theta)\xi_2(v_{g,n}^{11}(\theta),R_{g,n}^{11}(\theta))A_2(R_{g,n}^{11}(\theta)) \\
&\qquad + (1-\rho_{g,n}^{11^2}(\theta))\kappa(v_{g,n}^{11}(\theta),R_{g,n}^{11}(\theta))A_3(R_{g,n}^{11}(\theta))\| \\
&\le C_4\left[C_5 + |v_{g1,n}^{11}(\theta)\xi_1(v_{g,n}^{11}(\theta),R_{g,n}^{11}(\theta))| + |v_{g2,n}^{11}(\theta)\xi_2(v_{g,n}^{11}(\theta),R_{g,n}^{11}(\theta))| + \kappa(v_{g,n}^{11}(\theta),R_{g,n}^{11}(\theta))\right]
\end{aligned}$$

for some constants $C_3, C_4, C_5 > 0$, where the last equality follows by Lemma C.8 and its notation: recall that $V \sim \mathcal{N}(0, R_{g,n}^{11}(\theta))$ and $2\times 2$ matrices $A_1(R_{g,n}^{11}(\theta))$, $A_2(R_{g,n}^{11}(\theta))$, and $A_3(R_{g,n}^{11}(\theta))$ are functions of $\rho_{g,n}^{11}(\theta)$ defined in the same way as in (C.4) and $\kappa(v_{g,n}^{11}(\theta),R_{g,n}^{11}(\theta)) = \phi_2(v_{g,n}^{11}(\theta),R_{g,n}^{11}(\theta))/\Phi_2(v_{g,n}^{11}(\theta),R_{g,n}^{11}(\theta))$. It remains to show that the supremum of the last expression in (E.9) with respect to $\theta\in\Theta$ is uniformly $L_p$-bounded. By Loève's $c_r$-inequality, it suffices to show that $\sup_{n,g} E\left[\sup_{\theta\in\Theta}|v_{gj,n}^{11}(\theta)\xi_j(v_{g,n}^{11}(\theta),R_{g,n}^{11}(\theta))|\right]^p < \infty$, $j=1,2$, which is however implied by results in (E.3) and (E.4), and $\sup_{n,g} E\left[\sup_{\theta\in\Theta}\kappa(v_{g,n}^{11}(\theta),R_{g,n}^{11}(\theta))\right]^p < \infty$. For this last term, Lemmas C.5 and D.2 imply there are constants $C_6,\ldots,C_9 > 0$ such that

$$\begin{aligned}
&E\left[\sup_{\theta\in\Theta}\kappa(v_{g,n}^{11}(\theta),R_{g,n}^{11}(\theta))\right]^p \\
&\le E\left[\sup_{\theta\in\Theta}\left(C_6(1-|\rho_{g,n}^{11}(\theta)|)^{-3}\left((|v_{g1,n}^{11}(\theta)|+|v_{g2,n}^{11}(\theta)|+C_7)^2 + \left(1-\Phi\left((1-|\rho_{g,n}^{11}(\theta)|)^{-1/2}\right)\right)^{-2}\right)\right)\right]^p \\
&\le C_8 E\left[\sup_{\theta\in\Theta}(|v_{g1,n}^{11}(\theta)|+|v_{g2,n}^{11}(\theta)|+C_7)^2 + C_9\right]^p < \infty
\end{aligned}$$

uniformly in $n\in\mathbb{N}$ and $g\in\mathcal{G}_n$, where the conclusion follows from Lemma D.3 in the same way as in (E.3). This concludes the proof that (E.7) and thus (E.5) are uniformly bounded. The SE property thus follows from Proposition 1 of Jenish and Prucha (2009).

PC: proof of $\frac{1}{n}\sum_{g\in\mathcal{G}_n}\left(\sum_{a\in\mathcal{A}}[d_{g,n}^a f_{g,n}^a(\theta) - E[d_{g,n}^a f_{g,n}^a(\theta)]]\right) \xrightarrow{p} 0$ as $n\to\infty$ for $\theta\in\Theta$

In order to establish the pointwise convergence, we apply Theorem C.1. As before, we will establish the result only for $d_{g,n}^{11} f_{g,n}^{11}(\theta)$; the remaining terms can be analyzed analogously. We start by proving that the individual likelihood terms are $L_1$-NED on $\{\eta_{g,n}\}_{g\in\mathcal{G}_n}$ with $\eta_{g,n} = (X^s_{g\cdot,n}, X^o_{g\cdot,n}, u^s_{g,n}, u^o_{g,n})$. Note that

$$d^{11}_{g,n} f^{11}_{g,n}(\theta) = d^{11}_{g,n} \left( -\ln 2\pi - \frac{1}{2}\left(\ln|\Omega^{oo}_{g,n}(\theta)| + z'_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta) z_{g,n}(\theta)\right) + \ln \Phi_2(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta)) \right).$$

Given that $\Omega^{oo}_{g,n}(\theta)$ is non-stochastic and its determinant is uniformly bounded away from zero by Lemma D.1, it suffices to establish the $L_2$-NED property for $\{d^{11}_{g,n}\}_{g\in\mathcal{G}_n}$, $\{z'_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)z_{g,n}(\theta)\}_{g\in\mathcal{G}_n}$, and $\{\ln\Phi_2(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta))\}_{g\in\mathcal{G}_n}$ and apply Lemma C.9. In Lemma D.5, we have shown that $\{d^{11}_{g,n}\}_{g\in\mathcal{G}_n}$ is a uniform $L_2$-NED random field. Now we will apply Theorem C.2 to the remaining two random fields.

Let $z^{(s)}_{g,n}(\theta) = E[z_{g,n}(\theta)|\mathcal{F}_{g,n}(s)]$. Then by the elementwise mean value theorem, there exists $\tilde{z}^{(s)}_{g,n}(\theta)$ with elements between $z_{g,n}(\theta)$ and $z^{(s)}_{g,n}(\theta)$ such that

$$|z'_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)z_{g,n}(\theta) - z^{(s)'}_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)z^{(s)}_{g,n}(\theta)| \le \|2\Omega^{oo-1}_{g,n}(\theta)\tilde{z}^{(s)}_{g,n}(\theta)\|\|z_{g,n}(\theta) - z^{(s)}_{g,n}(\theta)\|.$$

In order to verify conditions (ii) and (iii) of Theorem C.2, we have to show that $\sup_s \|2\Omega^{oo-1}_{g,n}(\theta)\tilde{z}^{(s)}_{g,n}(\theta)\|_2$ and $\sup_s \|2\Omega^{oo-1}_{g,n}(\theta)\tilde{z}^{(s)}_{g,n}(\theta)\|z_{g,n}(\theta) - z^{(s)}_{g,n}(\theta)\|\|_r$ are uniformly bounded, for some $r > 2$. Since the elements of $\tilde{z}^{(s)}_{g,n}(\theta)$ lie between the elements of $z_{g,n}(\theta)$ and $z^{(s)}_{g,n}(\theta)$, let $C^{(s)}_{1g,n}(\theta)$ and $C^{(s)}_{2g,n}(\theta)$ be $2\times 2$ diagonal matrices with elements in $[0,1]$ such that $\tilde{z}^{(s)}_{g,n}(\theta) = C^{(s)}_{1g,n}(\theta)z_{g,n}(\theta) + C^{(s)}_{2g,n}(\theta)z^{(s)}_{g,n}(\theta)$. Given that $C^{(s)}_{1g,n}(\theta)$ and $C^{(s)}_{2g,n}(\theta)$ have all elements in $[0,1]$ irrespectively of $s, g, n$, and $\theta$, it holds that

$$\begin{aligned}
\sup_s E\|2\Omega^{oo-1}_{g,n}(\theta)\tilde{z}^{(s)}_{g,n}(\theta)\|^2 &\le 4\|\Omega^{oo-1}_{g,n}(\theta)\|^2 \sup_s E\|C^{(s)}_{1g,n}(\theta)z_{g,n}(\theta) + C^{(s)}_{2g,n}(\theta)z^{(s)}_{g,n}(\theta)\|^2 \\
&\le C_3\|\Omega^{oo-1}_{g,n}(\theta)\|^2 \left(E\|z_{g,n}(\theta)\|^2 + \sup_s E\|z^{(s)}_{g,n}(\theta)\|^2\right) \\
&\le 2C_3\|\Omega^{oo-1}_{g,n}(\theta)\|^2 E\|z_{g,n}(\theta)\|^2 \le 2C_3 \sup_{\theta\in\Theta}\|\Omega^{oo-1}_{g,n}(\theta)\|^2 E\left[\sup_{\theta\in\Theta}\|z_{g,n}(\theta)\|\right]^2 < \infty
\end{aligned}$$

for some constant $C_3 > 0$ uniformly in $n \in \mathbb{N}$ and $g \in \mathcal{G}_n$ by Lemmas D.1 and D.3, where the third inequality follows from the conditional Jensen's inequality. Next,

$$\begin{aligned}
&\sup_s E\|2\Omega^{oo-1}_{g,n}(\theta)\tilde{z}^{(s)}_{g,n}(\theta)\|z_{g,n}(\theta) - z^{(s)}_{g,n}(\theta)\|\|^r \\
&\le 2^r\|\Omega^{oo-1}_{g,n}(\theta)\|^r \sup_s E\|(C^{(s)}_{1g,n}(\theta)z_{g,n}(\theta) + C^{(s)}_{2g,n}(\theta)z^{(s)}_{g,n}(\theta))\|z_{g,n}(\theta) - z^{(s)}_{g,n}(\theta)\|\|^r \\
&\le C_4\|\Omega^{oo-1}_{g,n}(\theta)\|^r \sup_s E\left[\|z_{g,n}(\theta)\| + \|z^{(s)}_{g,n}(\theta)\|\right]^{2r} \le 2^{2r-1}C_4\|\Omega^{oo-1}_{g,n}(\theta)\|^r \left(E\|z_{g,n}(\theta)\|^{2r} + \sup_s E\|z^{(s)}_{g,n}(\theta)\|^{2r}\right) \\
&\le 2^{2r}C_4\|\Omega^{oo-1}_{g,n}(\theta)\|^r E\|z_{g,n}(\theta)\|^{2r} \le 2^{2r}C_4 \sup_{\theta\in\Theta}\|\Omega^{oo-1}_{g,n}(\theta)\|^r E\left[\sup_{\theta\in\Theta}\|z_{g,n}(\theta)\|\right]^{2r} < \infty
\end{aligned}$$

for some constant $C_4 > 0$. Since $E\|z'_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)z_{g,n}(\theta)\|^2 \leq \sup_{\theta\in\Theta}\|\Omega^{oo-1}_{g,n}(\theta)\|^2 E\left[\sup_{\theta\in\Theta}\|z_{g,n}(\theta)\|\right]^4 < \infty$ uniformly in $n \in \mathbb{N}$ and $g \in \mathcal{G}_n$ by Lemmas D.1 and D.3, condition (iv) of Theorem C.2 is fulfilled. As $\{z_{g,n}(\theta)\}_{g\in\mathcal{G}_n}$ is a uniform $L_2$-NED random field by Lemma D.5, it follows that $\{z'_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)z_{g,n}(\theta)\}_{g\in\mathcal{G}_n}$ is a uniform $L_2$-NED random field as well.

Similarly, let $v^{11(s)}_{g,n}(\theta) = E[v^{11}_{g,n}(\theta)|\mathcal{F}_{g,n}(s)]$ and verify the conditions of Theorem C.2 again. By the elementwise mean value theorem, there exists $\tilde{v}^{11(s)}_{g,n}(\theta)$ between $v^{11}_{g,n}(\theta)$ and $v^{11(s)}_{g,n}(\theta)$ such that

$$
\begin{aligned}
|\ln\Phi_2(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta)) - \ln\Phi_2(v^{11(s)}_{g,n}(\theta), R^{11}_{g,n}(\theta))| &\leq \left|\frac{\partial \ln\Phi_2(\tilde{v}^{11(s)}_{g,n}(\theta), R^{11}_{g,n}(\theta))}{\partial v'}(v^{11}_{g,n}(\theta) - v^{11(s)}_{g,n}(\theta))\right| \\
&\leq C_5\left(\|\tilde{v}^{11(s)}_{g,n}(\theta)\|^8 + C_6\right)\|v^{11}_{g,n}(\theta) - v^{11(s)}_{g,n}(\theta)\|
\end{aligned}
$$

for some constants $C_5, C_6 > 0$, where the last inequality follows from Lemmas C.4, C.5, and D.2 by the same argument as in (E.3). To bound $\|\tilde{v}^{11(s)}_{g,n}(\theta)\|^8$ to verify condition (ii) of Theorem C.2, denote $C^{(s)}_{7g,n}(\theta)$ and $C^{(s)}_{8g,n}(\theta)$ the $2\times 2$ diagonal matrices with elements in $[0,1]$ such that $\tilde{v}^{11(s)}_{g,n}(\theta) = C^{(s)}_{7g,n}(\theta)v^{11}_{g,n}(\theta) + C^{(s)}_{8g,n}(\theta)v^{11(s)}_{g,n}(\theta)$. Since $C^{(s)}_{7g,n}(\theta)$ and $C^{(s)}_{8g,n}(\theta)$ have all elements in $[0,1]$ irrespectively of $s, n, g$, and $\theta$, it holds that

$$
\begin{aligned}
\sup_s E\|\tilde{v}^{11(s)}_{g,n}(\theta)\|^{16} &= \sup_s E\|C^{(s)}_{7g,n}(\theta)v^{11}_{g,n}(\theta) + C^{(s)}_{8g,n}(\theta)v^{11(s)}_{g,n}(\theta)\|^{16} \\
&\leq \sup_s C_9\left(E\|v^{11}_{g,n}(\theta)\|^{16} + E\|v^{11(s)}_{g,n}(\theta)\|^{16}\right) \\
&\leq 2C_9 E\|v^{11}_{g,n}(\theta)\|^{16} \leq 2C_9 E\left[\sup_{\theta\in\Theta}\|v^{11}_{g,n}(\theta)\|\right]^{16}
\end{aligned}
$$

is bounded uniformly in $n \in \mathbb{N}$ and $g \in \mathcal{G}_n$ for some constant $C_9 > 0$ by Lemma D.3. Next, condition (iii) of Theorem C.2 can be verified for some $r > 2$ and some constant $C_{10} > 0$ by

$$
\begin{aligned}
&\sup_s E\left[\left(\|\tilde{v}^{11(s)}_{g,n}(\theta)\|^8 + C_6\right)\|v^{11}_{g,n}(\theta) - v^{11(s)}_{g,n}(\theta)\|\right]^r \\
&= \sup_s E\left[\left(\|C^{(s)}_{7g,n}(\theta)v^{11}_{g,n}(\theta) + C^{(s)}_{8g,n}(\theta)v^{11(s)}_{g,n}(\theta)\|^8 + C_6\right)\|v^{11}_{g,n}(\theta) - v^{11(s)}_{g,n}(\theta)\|\right]^r \\
&\leq \sup_s E\left[\left(C_{10}\left(\|v^{11}_{g,n}(\theta)\| + \|v^{11(s)}_{g,n}(\theta)\|\right)^8 + C_6\right)\left(\|v^{11}_{g,n}(\theta)\| + \|v^{11(s)}_{g,n}(\theta)\|\right)\right]^r \\
&\leq \sup_s 2^{r-1}\left(C^r_{10}E\left[\|v^{11}_{g,n}(\theta)\| + \|v^{11(s)}_{g,n}(\theta)\|\right]^{9r} + C^r_6 E\left[\|v^{11}_{g,n}(\theta)\| + \|v^{11(s)}_{g,n}(\theta)\|\right]^r\right) \\
&\leq \sup_s 2^{10r-2}C^r_{10}\left(E\|v^{11}_{g,n}(\theta)\|^{9r} + E\|v^{11(s)}_{g,n}(\theta)\|^{9r}\right) + \sup_s 2^{2r-2}C^r_6\left(E\|v^{11}_{g,n}(\theta)\|^r + E\|v^{11(s)}_{g,n}(\theta)\|^r\right)
\end{aligned}
$$

$$\leq 2^{10r-1}C_{10}^{r}E\|v_{g,n}^{11}(\theta)\|^{9r}+2^{2r-1}C_{6}^{r}E\|v_{g,n}^{11}(\theta)\|^{r}$$
$$\leq 2^{10r-1}C_{10}^{r}E\left[\sup_{\theta\in\Theta}\|v_{g,n}^{11}(\theta)\|\right]^{9r}+2^{2r-1}C_{6}^{r}E\left[\sup_{\theta\in\Theta}\|v_{g,n}^{11}(\theta)\|\right]^{r}<\infty$$

uniformly in $n\in\mathbb{N}$ and $g\in\mathcal{G}_n$ by Lemma D.3, where the second and third inequalities are implied by Loève's $c_r$-inequality, while the fourth inequality follows by the conditional Jensen's inequality. Finally, condition (iv) of Theorem C.2 can be verified in the same way as in (E.4) with the requirement that $E\left[\sup_{\theta\in\Theta}\|v_{g,n}^{11}(\theta)\|\right]^{18}<\infty$, which is the case by Lemma D.3. As we have shown in Lemma D.5 that $\{v_{g,n}^{11}(\theta)\}_{g\in\mathcal{G}_n}$ is a uniform $L_2$-NED random field, $\{\ln\Phi_2(v_{g,n}^{11}(\theta),R_{g,n}^{11}(\theta))\}_{g\in\mathcal{G}_n}$ is a uniform $L_2$-NED random field as well. Thus, by Lemma C.9, it follows that $\{d_{g,n}^{11}f_{g,n}^{11}(\theta)\}_{g\in\mathcal{G}_n}$ is a uniform $L_1$-NED random field.

Hence, condition (i) of Theorem C.1 is satisfied, whereas condition (ii) is already verified in the beginning of the proof; condition (iii) is implied by Assumptions $2(i)$, $2(ii)$, and $4(i)$. Since convergence in probability follows from convergence in $L_1$-norm, Theorem C.1 thus implies the pointwise convergence result. □

***Proof of Theorem 2:*** By the elementwise mean value theorem, there exists $\tilde{\theta}_n$ with elements between elements of $\theta_0$ and $\hat{\theta}_n$ such that

$$0=\frac{\partial Q_n(\hat{\theta}_n)}{\partial\theta}=\frac{\partial Q_n(\theta_0)}{\partial\theta}+\frac{\partial^2 Q_n(\tilde{\theta}_n)}{\partial\theta\partial\theta'}(\hat{\theta}_n-\theta_0).$$

Assumption 12 now implies that matrices $J_n(\theta_0)$ are non-singular for a sufficiently large $n$ and it holds that

$$\sqrt{n}\frac{\partial^2 Q_n(\tilde{\theta}_n)}{\partial\theta\partial\theta'}(\hat{\theta}_n-\theta_0)=-\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial\theta}$$

and

$$\sqrt{n}J_n^{-1/2}(\theta_0)\frac{\partial^2 Q_n(\tilde{\theta}_n)}{\partial\theta\partial\theta'}(\hat{\theta}_n-\theta_0)=-\sqrt{n}J_n^{-1/2}(\theta_0)\frac{\partial Q_n(\theta_0)}{\partial\theta}.$$

To prove the claim of the theorem, we first establish that the term $\sqrt{n}J_n^{-1/2}(\theta_0)\partial Q_n(\theta_0)/\partial\theta$ converges in distribution to $\mathcal{N}(0,I_L)$ as $n\to\infty$, and we show later that $\partial^2 Q_n(\tilde{\theta}_n)/\partial\theta\partial\theta'-H_n(\theta_0)\xrightarrow{p}0$ as $n\to\infty$.

Proof of $\sqrt{n}J_n^{-1/2}(\theta_0)\frac{\partial Q_n(\theta_0)}{\partial\theta}=\frac{1}{\sqrt{n}}J_n^{-1/2}(\theta_0)\sum_{g\in\mathcal{G}_n}\sum_{a\in\mathcal{A}}d_{g,n}^{a}\frac{\partial f_{g,n}^{a}(\theta_0)}{\partial\theta}\xrightarrow{d}\mathcal{N}(0,I_L)$ as $n\to\infty$

We apply Theorem C.3. The individual score components have mean zero because the marginal likelihood contributions for each group are correctly specified. The remaining assumptions of Theorem C.3 concerning $L_{2+\delta}$-boundedness ($\delta>0$) and NED properties are verified at a general $\theta\in\Theta$, but they are applied at $\theta=\theta_0$.

By Loève's $c_r$-inequality and $d^a_{g,n}$, $a \in \mathcal{A}$, being an indicator function, the individual score contributions are uniformly $L_{2+\delta}$-bounded if $\sup_{n,g} E\left[\|\partial f^a_{g,n}(\theta)/\partial\theta\|\right]^{2+\delta} \leq \sup_{n,g} E\left[\sup_{\theta\in\Theta}\|\partial f^a_{g,n}(\theta)/\partial\theta\|\right]^{2+\delta} < \infty$ for some $\delta > 0$. The result for $a = 11$ is a special case of the uniform boundedness of (E.5) verified in the proof of Theorem 1, property SE; the boundedness of the other terms can be proven in a similar way.

Now we establish that $\{d^{11}_{g,n}\partial f^{11}_{g,n}(\theta)/\partial\theta\}_{g\in\mathcal{G}_n}$ is a uniform $L_2$-NED random field on the $\alpha$-mixing random field $\{\eta_{g,n} = (X^s_{g\cdot,n}, X^o_{g\cdot,n}, u^s_{g,n}, u^o_{g,n})\}_{g\in\mathcal{G}_n}$. Recall that

$$d^{11}_{g,n}\frac{\partial f^{11}_{g,n}(\theta)}{\partial\theta} = -\frac{1}{2}d^{11}_{g,n}\frac{1}{|\Omega^{oo}_{g,n}(\theta)|}\frac{\partial|\Omega^{oo}_{g,n}(\theta)|}{\partial\theta} - \frac{1}{2}d^{11}_{g,n}\frac{\partial(z'_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)z_{g,n}(\theta))}{\partial\theta} + d^{11}_{g,n}\frac{\partial\ln\Phi_2(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta))}{\partial\theta}. \tag{E.10}$$

It thus suffices to show that each term of the summation is uniformly $L_2$-NED and to find their NED coefficients. We have already established in Lemma D.5 that $\{d^{11}_{g,n}\}_{g\in\mathcal{G}_n}$ is a uniform $L_2$-NED random field with NED coefficients bounded by $\psi^{1/6}(s)$, where $\psi(s)$ is defined in Assumption 5. Since $|\Omega^{oo}_{g,n}(\theta)|$ is uniformly bounded away from zero and the norm of $\partial|\Omega^{oo}_{g,n}(\theta)|/\partial\theta$ is uniformly bounded by Lemma D.1, the first term in (E.10) is uniformly $L_2$-NED with NED coefficients bounded by $\psi^{1/6}(s)$. For the second and third terms in (E.10), we apply Theorem C.2. Let $d^{11(s)}_{g,n} = E[d^{11}_{g,n}|\mathcal{F}_{g,n}(s)]$ and $z^{(s)}_{g,n}(\theta) = E[z_{g,n}(\theta)|\mathcal{F}_{g,n}(s)]$. Then

$$\begin{aligned}
&\left\|d^{11}_{g,n}\frac{\partial(z'_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)z_{g,n}(\theta))}{\partial\theta} - d^{11(s)}_{g,n}\frac{\partial(z^{(s)'}_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)z^{(s)}_{g,n}(\theta))}{\partial\theta}\right\| \\
&\leq \left(d^{11}_{g,n} + \left\|\frac{\partial z^{(s)'}_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)z^{(s)}_{g,n}(\theta)}{\partial\theta}\right\|\right)\left(|d^{11}_{g,n} - d^{11(s)}_{g,n}| + \left\|\frac{\partial z'_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)z_{g,n}(\theta)}{\partial\theta} - \frac{\partial z^{(s)'}_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)z^{(s)}_{g,n}(\theta)}{\partial\theta}\right\|\right) \\
&\leq \left(1 + \left\|\frac{\partial(z^{(s)'}_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)z^{(s)}_{g,n}(\theta))}{\partial\theta}\right\|\right)\left(|d^{11}_{g,n} - d^{11(s)}_{g,n}| + \left\|\frac{\partial^2(\tilde{z}^{(s)'}_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)\tilde{z}^{(s)}_{g,n}(\theta))}{\partial\theta\partial z'}\right\|\|z_{g,n}(\theta) - z^{(s)}_{g,n}(\theta)\|\right) \\
&\leq \left(1 + \left\|\frac{\partial(z^{(s)'}_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)z^{(s)}_{g,n}(\theta))}{\partial\theta}\right\|\right)\left(1 + \left\|\frac{\partial^2(\tilde{z}^{(s)'}_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)\tilde{z}^{(s)}_{g,n}(\theta))}{\partial\theta\partial z'}\right\|\right)\left(|d^{11}_{g,n} - d^{11(s)}_{g,n}| + \|z_{g,n}(\theta) - z^{(s)}_{g,n}(\theta)\|\right),
\end{aligned}$$

where the second inequality follows by the elementwise mean value theorem with elements of $\tilde{z}^{(s)}_{g,n}(\theta)$ being between elements of $z_{g,n}(\theta)$ and $z^{(s)}_{g,n}(\theta)$.

By the Cauchy-Schwartz, Minkowski's, and Liapunov's inequalities, conditions (ii) and (iii) of Theorem C.2 are fulfilled if $\|\partial(z^{(s)'}_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)z^{(s)}_{g,n}(\theta))/\partial\theta\|_{4r}$, $\|\partial^2(\tilde{z}^{(s)'}_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)\tilde{z}^{(s)}_{g,n}(\theta))/\partial\theta\partial z'\|_{4r}$, $\|d^{11}_{g,n} - d^{11(s)}_{g,n}\|_{2r}$, and $\|z_{g,n}(\theta) - z^{(s)}_{g,n}(\theta)\|_{2r}$ are uniformly bounded for some $r > 2$. The boundedness of the first term can be proven in the same way as in (E.6) with an additional application of the conditional Jensen's inequality. Given the second order derivative of a quadratic form in Lemma C.6, it is not difficult to prove that

$\partial^2(\tilde{z}_{g,n}^{(s)'}(\theta)\Omega_{g,n}^{oo-1}(\theta)\tilde{z}_{g,n}^{(s)}(\theta))/\partial\theta\partial z'$ is uniformly $L_{4r}$-bounded. Trivially, $d_{g,n}^{11} - d_{g,n}^{11(s)}$ is uniformly $L_{2r}$-bounded as well, while the $L_{2r}$-boundedness of $z_{g,n}(\theta) - z_{g,n}^{(s)}(\theta)$ follows from Minkowski's and the conditional Jensen's inequalities and Lemma D.3. Since the uniform $L_2$-boundedness of $\partial(z'_{g,n}(\theta)\Omega_{g,n}^{oo-1}(\theta)z_{g,n}(\theta))/\partial\theta$ follows in the same way as in (E.6), condition (iv) of Theorem C.2 is fulfilled. Furthermore, since $\{d_{g,n}^{11}\}_{g\in\mathcal{G}_n}$ and $\{z_{g,n}(\theta)\}_{g\in\mathcal{G}_n}$ are uniform $L_2$-NED random fields with NED coefficients bounded by $\psi^{1/6}(s)$ by Lemma D.5, Theorem C.2 implies that the second term in (E.10) is uniformly $L_2$-NED with NED coefficients bounded by $\psi^{(r-2)/(12r-12)}(s)$ for some $r > 2$.

Regarding the last term in (E.10), it follows similarly with $v_{g,n}^{11(s)}(\theta) = E[v_{g,n}^{11}(\theta)|\mathcal{F}_{g,n}(s)]$ that

$$\left\| d_{g,n}^{11}\frac{\partial \ln \Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))}{\partial\theta} - d_{g,n}^{11(s)}\frac{\partial \ln \Phi_2(v_{g,n}^{11(s)}(\theta), R_{g,n}^{11}(\theta))}{\partial\theta} \right\|$$
$$\leq \left(1 + \left\| \frac{\partial \ln \Phi_2(v_{g,n}^{11(s)}(\theta), R_{g,n}^{11}(\theta))}{\partial\theta} \right\| \right) \left(1 + \left\| \frac{\partial^2 \ln \Phi_2(\tilde{v}_{g,n}^{11(s)}(\theta), R_{g,n}^{11}(\theta))}{\partial\theta\partial v'} \right\| \right) (|d_{g,n}^{11} - d_{g,n}^{11(s)}| + \|v_{g,n}^{11}(\theta) - v_{g,n}^{11(s)}(\theta)\|)$$

with elements of $\tilde{v}_{g,n}^{11(s)}(\theta)$ lying between elements of $v_{g,n}^{11}(\theta)$ and $v_{g,n}^{11(s)}(\theta)$. Analogously to the previous case, applying Theorem C.2 requires us to check that $\partial \ln \Phi_2(v_{g,n}^{11(s)}(\theta), R_{g,n}^{11}(\theta))/\partial\theta$, $\partial^2 \ln \Phi_2(\tilde{v}_{g,n}^{11(s)}(\theta), R_{g,n}^{11}(\theta))/\partial\theta\partial v'$ and $d_{g,n}^{11} - d_{g,n}^{11(s)}$, $v_{g,n}^{11}(\theta) - v_{g,n}^{11(s)}(\theta)$ are uniformly $L_{4r}$- and $L_{2r}$-bounded, respectively. Given Lemmas C.4 and C.5, the boundedness of the first term has been established in the proof of Theorem 1, property SE, and the boundedness of the second term can be established analogously. The third term is obviously uniformly $L_{2r}$-bounded, while the uniform $L_{2r}$-boundedness of the fourth term again follows from Minkowski's and the conditional Jensen's inequalities and Lemma D.3. Given that $\{d_{g,n}\}_{g\in\mathcal{G}_n}$ and $\{v_{g,n}^{11}(\theta)\}_{g\in\mathcal{G}_n}$ are uniformly $L_2$-NED with NED coefficients bounded by $\psi^{1/6}(s)$ by Lemma D.5, $\{d_{g,n}^{11}\partial \ln \Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))/\partial\theta\}_{g\in\mathcal{G}_n}$ is a uniform $L_2$-NED random field with NED coefficients bounded by $\psi^{(r-2)/(12r-12)}(s)$ by Theorem C.2 for some $r > 2$, and consequently, $\{d_{g,n}^{11}\partial f_{g,n}^{11}(\theta)/\partial\theta\}_{g\in\mathcal{G}_n}$ is a uniform $L_2$-NED random field with NED coefficients bounded by $\psi^{(r-2)/(12r-12)}(s)$. Hence, conditions (iii) and (v) of Theorem C.3 are fulfilled, whereas conditions (iv) and (vi) are satisfied by Assumptions 10 and 2($i$), 2($ii$), and 9($i$), respectively; condition (vii) is assumed in Assumption 12($ii$). The asymptotic normality result follows from Theorem C.3.

Proof of $\frac{\partial^2 Q_n(\tilde{\theta}_n)}{\partial\theta\partial\theta'} - H_n(\theta_0) = \frac{1}{n}\sum_{g\in\mathcal{G}_n}\sum_{a\in\mathcal{A}} d_{g,n}^a \frac{\partial^2 f_{g,n}^a(\tilde{\theta}_n)}{\partial\theta\partial\theta'} - H_n(\theta_0) \xrightarrow{p} 0$ as $n \to \infty$

We can establish this result by showing that, for $n \to \infty$,

$$\frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'} - E\left[\frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'}\right] \xrightarrow{p} 0 \text{ and } \frac{\partial^2 Q_n(\tilde{\theta}_n)}{\partial\theta\partial\theta'} - \frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'} \xrightarrow{p} 0.$$

For the first claim, we apply Theorem C.1. As before, we establish the results for the part of the objective function term corresponding to index $a = 11$ and a general $\theta \in \Theta$ and apply it at $\theta = \theta_0$; the results for the other terms can be proven in a similar way.

Note that

$$d_{g,n}^{11}\frac{\partial^2 f_{g,n}^{11}(\theta)}{\partial\theta\partial\theta'} = d_{g,n}^{11}\left[-\frac{1}{2}\left(-\frac{1}{|\Omega_{g,n}^{oo}(\theta)|^2}\frac{\partial|\Omega_{g,n}^{oo}(\theta)|}{\partial\theta}\frac{\partial|\Omega_{g,n}^{oo}(\theta)|}{\partial\theta'} + \frac{1}{|\Omega_{g,n}^{oo}(\theta)|}\frac{\partial^2|\Omega_{g,n}^{oo}(\theta)|}{\partial\theta\partial\theta'}\right.\right.$$
$$\left.\left. + \frac{\partial^2(z_{g,n}'(\theta)\Omega_{g,n}^{oo-1}(\theta)z_{g,n}(\theta))}{\partial\theta\partial\theta'}\right) + \frac{\partial^2 \ln\Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))}{\partial\theta\partial\theta'}\right].$$

We start by showing that $d_{g,n}^{11}\partial^2 f_{g,n}^{11}(\theta)/\partial\theta\partial\theta'$ is uniformly $L_p$-bounded for some $p > 1$. Note that $d_{g,n}^{11}$ is an indicator function, whereas $|\Omega_{g,n}^{oo}(\theta)|$ is uniformly bounded away from zero and the norm of $\partial|\Omega_{g,n}^{oo}(\theta)|/\partial\theta$ is uniformly bounded by Lemma D.1. It can be shown using the second order derivative of a determinant in Lemma C.6 that the norm of $\partial^2|\Omega_{g,n}^{oo}(\theta)|/\partial\theta\partial\theta'$ is uniformly bounded as well. Given the second order derivative of a quadratic form calculated in Lemma C.6, it is not difficult to see that the uniform $L_p$-boundedness of $\partial^2(z_{g,n}'(\theta)\Omega_{g,n}^{oo-1}(\theta)z_{g,n}(\theta))/\partial\theta\partial\theta'$ can be easily established with the help of Lemmas D.1, D.3, and D.4 (analogously to the $L_p$-boundedness of the first derivative in (E.6)). Given the third result of Lemma C.4 and Lemma C.5, it can be also shown that $\partial^2 \ln\Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))/\partial\theta\partial\theta'$ is $L_p$-bounded in a similar way as it was done for the first order derivative in the proof of Theorem 1, property SE.

In order to show that $\{d_{g,n}^{11}\partial^2 f_{g,n}^{11}(\theta)/\partial\theta\partial\theta'\}_{g\in\mathcal{G}_n}$ is a uniform $L_1$-NED random field on the $\alpha$-mixing random field $\{\eta_{g,n} = (X_{g\cdot,n}^s, X_{g\cdot,n}^o, u_{g,n}^s, u_{g,n}^o)\}_{g\in\mathcal{G}_n}$, we have to establish the uniform $L_2$-NED property for $\{d_{g,n}^{11}\}_{g\in\mathcal{G}_n}$, as is already done in Lemma D.5, and for the second order derivatives $\{\partial^2(z_{g,n}'(\theta)\Omega_{g,n}^{oo-1}(\theta)z_{g,n}(\theta))/\partial\theta\partial\theta'\}_{g\in\mathcal{G}_n}$ and $\{\partial^2 \ln\Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))/\partial\theta\partial\theta'\}_{g\in\mathcal{G}_n}$ and apply Lemma C.9. It can be done in a similar way as is done for $\{z_{g,n}'(\theta)\Omega_{g,n}^{oo-1}(\theta)z_{g,n}(\theta)\}_{g\in\mathcal{G}_n}$ and $\{\ln\Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))\}_{g\in\mathcal{G}_n}$ in the proof of Theorem 1.

Finally, condition (iii) of Theorem C.1 is fulfilled by Assumptions 2$(i)$, 2$(ii)$, and 9$(i)$. The convergence of the second order derivative of $Q_n(\theta)$ at $\theta_0$ to $H(\theta_0)$ follows from Theorem C.1 and Assumption 12$(i)$ and the fact that convergence in $L_1$-norm implies convergence in probability.

We continue by proving that, for $n \to \infty$,

$$\frac{\partial^2 Q_n(\tilde{\theta}_n)}{\partial\theta\partial\theta'} - \frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'} \xrightarrow{p} 0.$$

We apply the strategy used in the proof of Theorem 2 by Xu and Lee (2015a) and show that $\{\partial^2 Q_n(\theta)/\partial\theta\partial\theta'\}_{g\in\mathcal{G}_n}$ is $L_0$-stochastically equicontinuous because the claim then follows directly from the proposition concerning the $L_0$-stochastic equicontinuity given in Andrews (1994). Since the objective function $Q_n(\theta)$ as well as its second order derivative are continuously differentiable, the stochastic equicontinuity of $\partial^2 Q_n(\theta)/\partial\theta\partial\theta'$ at $\theta = \theta_0$ can however be established in a similar way as we have verified it for $Q_n(\theta)$ in the proof of Theorem 1, property SE, which thus concludes the proof. $\square$

***Proof of Theorem 3:*** Let us recall and define the following notation. For each bootstrap sample $b = 1, \ldots, B$, $\Gamma_n^{(b,\theta')}(\theta) = \partial Q_n^{(b,\theta')}(\theta)/\partial\theta$ and $\bar{\Gamma}_n^{(b,\theta')}(\theta) = \partial Q_n^{(b,\theta')}(\theta)/\partial\theta - E\left[\partial Q_n^{(b,\theta')}(\theta)/\partial\theta\right]$, where $\theta$ in the parentheses denotes the parameter vector at which the objective function $Q_n$ and its derivatives are evaluated and where $\theta'$ in the superscripts indicates the parameter vector used to generate the bootstrap sample data. Next, $P^b$ denotes the probability measure induced by the bootstrap conditional on the original sample and thus conditional on the estimates $\hat{\theta}_n$. Since it was shown in the proof of Theorem 2 that $\sqrt{n}J_n(\theta_0)^{-1/2}\Gamma_n(\theta_0) \xrightarrow{d} \mathcal{N}(0, I_L)$ and $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$ as $n \to \infty$, the claims of the theorem follow for example by the Cramer-Wold device and Cavaliere and Georgiev (2017, Theorem 1) once we prove $\sqrt{n}J_n(\theta_0)^{-1/2}\Gamma_n^{(b,\hat{\theta}_n)}(\hat{\theta}_n) \xrightarrow{d} \mathcal{N}(0, I_L)$ conditional on $X_n^s$, $X_n^o$, and any sequence of estimates $\hat{\theta}_n = \theta_n = \theta_0 + \Delta_n n^{-1/2}$ such that $\|\Delta_n\| < M$ if $n > n_0$, $\Delta_n \in \mathbb{R}^L$, for any large $M > 0$ and $n_0 \in \mathbb{N}$.[20]

Furthermore, consider a particular bootstrap sample based on the $n$ draws $(u_{i,n}^{s0}, u_{i,n}^{o0})'$ from $\mathcal{N}(0, I_2)$, that is, $(u_{i,n}^{s(b,\theta_n)}, u_{i,n}^{o(b,\theta_n)})' = \Sigma(\theta_n)^{1/2}(u_{i,n}^{s0}, u_{i,n}^{o0})'$ and $y_{i,n}^{s(b,\theta_n)}$ and $y_{i,n}^{o(b,\theta_n)}$ defined as in Section 4.3. If the bootstrap sample is based on the same $n$ draws $(u_{i,n}^{s0}, u_{i,n}^{o0})'$ but relies on the true parameter $\theta_0$ instead of $\theta_n$, the bootstrap data $y_{i,n}^{s(b,\theta_0)}$ and $y_{i,n}^{o(b,\theta_0)}$ follow the same data generating process as the original data, $\Gamma_n^{(b,\theta_0)}(\theta_0) = \Gamma_n(\theta_0) = \partial Q_n(\theta_0)/\partial\theta$, and $\bar{\Gamma}_n^{(b,\theta_0)}(\theta_0) = \bar{\Gamma}_n(\theta_0) = \partial Q_n(\theta_0)/\partial\theta - E\left[\partial Q_n(\theta_0)/\partial\theta\right]$. Then by the proof of Theorem 2, $\sqrt{n}J_n(\theta_0)^{-1/2}\Gamma_n^{(b,\theta_0)}(\theta_0) = \sqrt{n}J_n(\theta_0)^{-1/2}\Gamma_n(\theta_0)$ converges in distribution to $\mathcal{N}(0, I_L)$ as $n \to \infty$. By Assumption 12, the claims of the theorem thus follow if we prove for $n \to \infty$ that $\sqrt{n}\{\Gamma_n^{(b,\theta_n)}(\theta_n) - \Gamma_n^{(b,\theta_0)}(\theta_0)\} = \sqrt{n}\{\bar{\Gamma}_n^{(b,\theta_n)}(\theta_n) - \bar{\Gamma}_n^{(b,\theta_0)}(\theta_0)\} \xrightarrow{p} 0$. To verify this claim, the following decomposition is used (noting $\bar{\Gamma}_n^{(b,\theta_0)}(\theta) \equiv \bar{\Gamma}_n(\theta)$):

$$\begin{aligned} \sqrt{n}\{\bar{\Gamma}_n^{(b,\theta_n)}(\theta_n) - \bar{\Gamma}_n^{(b,\theta_0)}(\theta_0)\} &= \sqrt{n}\{\bar{\Gamma}_n^{(b,\theta_n)}(\theta_n) - \bar{\Gamma}_n^{(b,\theta_0)}(\theta_n)\} && \text{(E.11)} \\ &+ \sqrt{n}\{\bar{\Gamma}_n(\theta_n) - \bar{\Gamma}_n(\theta_0)\}. && \text{(E.12)} \end{aligned}$$

The last term (E.12) can be shown to be asymptotically negligible in probability by proving that the

[20] To simplify the notation, we will keep this conditioning implicit.

difference

$$\sqrt{n}\{\bar{\Gamma}_n(\theta_n)-\bar{\Gamma}_n(\theta_0)\} = \frac{1}{\sqrt{n}} \sum_{g\in\mathcal{G}_n} \sum_{a\in\mathcal{A}} \left\{ d^a_{g,n} \frac{\partial f^a_{g,n}(\theta_n)}{\partial\theta} - d^a_{g,n} \frac{\partial f^a_{g,n}(\theta_0)}{\partial\theta} - E\left[ d^a_{g,n} \frac{\partial f^a_{g,n}(\theta_n)}{\partial\theta} - d^a_{g,n} \frac{\partial f^a_{g,n}(\theta_0)}{\partial\theta} \right] \right\}$$

converges to zero for any sequence $\theta_n \in U(\theta_0, n^{-1/2}M)$, where $U(\theta_0,\delta) = \{\theta : \|\theta-\theta_0\| < \delta\}$. Since the log-likelihood scores are differentiable, the $(1+\delta)$ moments of the scores as well as their derivatives are bounded (see the proof of Theorem 2), and even have the integrable majorants (see the proof of Theorem 1, part SE), the mean value theorem implies for some convex combination $\xi_n$ of $\theta_n$ and $\theta_0$ that

$$\begin{aligned}
\sqrt{n}\{\bar{\Gamma}_n(\theta_n) - \bar{\Gamma}_n(\theta_0)\} &= \frac{1}{\sqrt{n}} \sum_{g\in\mathcal{G}_n} \sum_{a\in\mathcal{A}} \left\{ d^a_{g,n} \frac{\partial^2 f^a_{g,n}(\xi_n)}{\partial\theta\partial\theta'} (\theta_n-\theta_0) - E\left[ d^a_{g,n} \frac{\partial^2 f^a_{g,n}(\xi_n)}{\partial\theta\partial\theta'} (\theta_n-\theta_0) \right] \right\} \\
&= \frac{\sqrt{n}(\theta_n-\theta_0)}{n} \sum_{g\in\mathcal{G}_n} \sum_{a\in\mathcal{A}} \left\{ d^a_{g,n} \frac{\partial^2 f^a_{g,n}(\xi_n)}{\partial\theta\partial\theta'} - E\left[ d^a_{g,n} \frac{\partial^2 f^a_{g,n}(\xi_n)}{\partial\theta\partial\theta'} \right] \right\} \\
&= \sqrt{n}(\theta_n-\theta_0) \left( \frac{\partial^2 Q_n(\xi_n)}{\partial\theta\partial\theta'} - E\left[ \frac{\partial^2 Q_n(\xi_n)}{\partial\theta\partial\theta'} \right] \right) \leq n^{1/2}\cdot n^{-1/2}M \cdot o_p(1) = o_p(1),
\end{aligned}$$

where the inequality follows by the definition of $\theta_n$ and the last part of the proof of Theorem 2 showing that $\partial^2 Q_n(\theta)/\partial\theta\partial\theta' - E[\partial^2 Q_n(\theta)/\partial\theta\partial\theta'] \xrightarrow{p} 0$ for any $\theta \in \Theta$.

To analyze the other term (E.11), we decompose (E.11) futher as

$$\begin{aligned}
&\sqrt{n}\{\bar{\Gamma}^{(b,\theta_n)}_n(\theta_n) - \bar{\Gamma}^{(b,\theta_0)}_n(\theta_n)\} \\
= & \frac{1}{\sqrt{n}} \sum_{g\in\mathcal{G}_n} \sum_{a\in\mathcal{A}} \left\{ d^{a(b,\theta_n)}_{g,n} \frac{\partial f^{a(b,\theta_n)}_{g,n}(\theta_n)}{\partial\theta} - d^{a(b,\theta_0)}_{g,n} \frac{\partial f^{a(b,\theta_0)}_{g,n}(\theta_n)}{\partial\theta} - E\left[ d^{a(b,\theta_n)}_{g,n} \frac{\partial f^{a(b,\theta_n)}_{g,n}(\theta_n)}{\partial\theta} - d^{a(b,\theta_0)}_{g,n} \frac{\partial f^{a(b,\theta_0)}_{g,n}(\theta_n)}{\partial\theta} \right] \right\} \\
= & \frac{1}{n^\alpha n^{1/2}} \sum_{g\in\mathcal{G}_n} \sum_{a\in\mathcal{A}} \left\{ n^\alpha \mathbb{1}(d^{a(b,\theta_n)}_{g,n} \neq d^{a(b,\theta_0)}_{g,n}) \left( d^{a(b,\theta_n)}_{g,n} \frac{\partial f^{a(b,\theta_n)}_{g,n}(\theta_n)}{\partial\theta} - d^{a(b,\theta_0)}_{g,n} \frac{\partial f^{a(b,\theta_0)}_{g,n}(\theta_n)}{\partial\theta} \right) \right. \qquad \text{(E.13)} \\
& \left. \qquad -E\left[ n^\alpha \mathbb{1}(d^{a(b,\theta_n)}_{g,n} \neq d^{a(b,\theta_0)}_{g,n}) \left( d^{a(b,\theta_n)}_{g,n} \frac{\partial f^{a(b,\theta_n)}_{g,n}(\theta_n)}{\partial\theta} - d^{a(b,\theta_0)}_{g,n} \frac{\partial f^{a(b,\theta_0)}_{g,n}(\theta_n)}{\partial\theta} \right) \right] \right\} \\
+ & \frac{1}{\sqrt{n}} \sum_{g\in\mathcal{G}_n} \sum_{a\in\mathcal{A}} \left\{ \mathbb{1}(d^{a(b,\theta_n)}_{g,n} = d^{a(b,\theta_0)}_{g,n}) d^{a(b,\theta_0)}_{g,n} \left( \frac{\partial f^{a(b,\theta_n)}_{g,n}(\theta_n)}{\partial\theta} - \frac{\partial f^{a(b,\theta_0)}_{g,n}(\theta_n)}{\partial\theta} \right) \right. \qquad \text{(E.14)} \\
& \left. \qquad -E\left[ \mathbb{1}(d^{a(b,\theta_n)}_{g,n} = d^{a(b,\theta_0)}_{g,n}) d^{a(b,\theta_0)}_{g,n} \left( \frac{\partial f^{a(b,\theta_n)}_{g,n}(\theta_n)}{\partial\theta} - \frac{\partial f^{a(b,\theta_0)}_{g,n}(\theta_n)}{\partial\theta} \right) \right] \right\}.
\end{aligned}$$

To show that (E.13) is asymptotically negligible in probability, we will first verify that the triangular array of elements in (E.13) satisfies the assumptions of the central limit Theorem C.3. Then we can employ the Chebyshev inequality along with the result shown in the proof of Theorem C.3 that variance of the sum in (E.13) is bounded by $C_1 \cdot n$, $C_1 > 0$, under Assumptions (i)–(vi) of Theorem C.3 (see the proof of Theorem 2, part 3, of Jenish and Prucha (2012)). To prove the convergence of (E.13) to zero in probability, we thus have to make sure that the elements of the sum (E.13) are $L_2$-NED on $\{\eta_{g,n} = (X^s_{g\cdot,n}, X^o_{g\cdot,n}, u^s_{g,n}, u^o_{g,n})\}_{g\in\mathcal{G}_n}$ with some NED coefficients $\psi(s)$ satisfying $\sum_{s=1}^{\infty} s^{2\tilde{d}-1}\psi(s) < \infty$ and uniformly bounded scaling coefficients, $\sup_{n,g} t_{g,n} < \infty$. We also have to prove that the elements of the sum (E.13) are uniformly $L_{2+\delta}$-bounded for some $\delta > 0$. The remaining assumptions (i) and (vi) follow by definition and by Assumption 4($i$). Before verifying these conditions, let us recall that data following the sample-selection data generating process with some parameter $\theta \in U(\theta_0, \delta)$ satisfy Assumptions 1–11 for a sufficiently small $\delta$, and therefore, the previously derived results assuming that $\theta_0$ represents the true parameter vector apply to the data generating process based on the parameter vector $\theta \neq \theta_0, \theta \in U(\theta_0, \delta)$, as well. Therefore, we will assume from now on that $n_0$ is sufficiently large so that $n_0^{-1/2}M < \delta$ and $\theta_n \in U(\theta_0, n^{-1/2}M)$.

Similarly to the previous analysis, we will establish the $L_2$-NED property for $a = 11$; the remaining terms can be analyzed similarly. Let $\tilde{\psi}(s) = \max\{\psi^s(s), \tilde{\psi}^s(s), \psi^o(s)\}$ and note that we have already verified in the proof of Theorem 2 the $L_2$-NED property of $\{d^{11}_{g,n}\partial f^{11}_{g,n}(\theta)/\partial\theta\}_{g\in\mathcal{G}_n}$. The $L_2$-NED property for (E.13) can be verified analogously once we show that $\{n^{\alpha}\mathbb{1}(d^{11(b,\theta_n)}_{g,n} \neq d^{11(b,\theta_0)}_{g,n})d^{11(b,\theta)}_{g,n}\}_{g\in\mathcal{G}_n}$, $\theta \in \{\theta_0, \theta_n\}$, is uniform $L_2$-NED and find its NED coefficients. As $d^{11(b,\theta)}_{g,n}$, $\theta \in \{\theta_n, \theta_0\}$, is $L_4$-NED with the NED coefficients bounded by $\tilde{\psi}^{1/6}(s)$ by the proof of Lemma D.5 for any $\theta_n \in U(\theta_0, n^{-1/2}M)$, the product rule in Lemma C.9 implies we only have to show that the indicator functions in (E.13) are $L_4$-NED on $\{\eta_{g,n} = (X^s_{g\cdot,n}, X^o_{g\cdot,n}, u^s_{g,n}, u^o_{g,n})\}_{g\in\mathcal{G}_n}$ with the NED scaling factors $t_{g,n} = O(n^{-\alpha})$ to compensate the multiplication by $n^{\alpha}$ in (E.13). To analyze the indicators, we split them first as

$$\mathbb{1}(d^{11(b,\theta_n)}_{g,n} \neq d^{11(b,\theta_0)}_{g,n}) = \mathbb{1}(d^{11(b,\theta_n)}_{g1,n} \neq d^{11(b,\theta_0)}_{g1,n}) + \mathbb{1}(d^{11(b,\theta_n)}_{g2,n} \neq d^{11(b,\theta_0)}_{g2,n}) - \mathbb{1}(d^{11(b,\theta_n)}_{g1,n} \neq d^{11(b,\theta_0)}_{g1,n})\mathbb{1}(d^{11(b,\theta_n)}_{g2,n} \neq d^{11(b,\theta_0)}_{g2,n}). \tag{E.15}$$

We will establish the $L_4$-NED property for $\{\mathbb{1}(d^{11(b,\theta_n)}_{g,n} \neq d^{11(b,\theta_0)}_{g,n})\}_{g\in\mathcal{G}_n}$ by showing that $\{\mathbb{1}(d^{11(b,\theta_n)}_{gi,n} \neq d^{11(b,\theta_0)}_{gi,n})\}_{g\in\mathcal{G}_n}$, $i \in \{1, 2\}$, is $L_8$-NED, applying the triangle inequality and the product rule in Lemma C.9, and noting that the $L_8$-NED property implies the $L_4$-NED property. Note that each indicator equals to the sum of two indicators:

$$\mathbb{1}(d^{11(b,\theta_n)}_{g1,n} \neq d^{11(b,\theta_0)}_{g1,n}) = \mathbb{1}(y^{*s(b,\theta_n)}_{g1,n} > 0 > y^{*s(b,\theta_0)}_{g1,n}) + \mathbb{1}(y^{*s(b,\theta_n)}_{g1,n} < 0 < y^{*s(b,\theta_0)}_{g1,n}). \tag{E.16}$$

We will thus verify the $L_8$-NED property for $\{\mathbb{1}(y^{*s(b,\theta_n)}_{g1,n} > 0 > y^{*s(b,\theta_0)}_{g1,n})\}_{g\in\mathcal{G}_n}$, noting that the other term can be analyzed analogously. Since $|\mathbb{1}(X) - E[\mathbb{1}(X)|\mathcal{F}]|^8 \leq |\mathbb{1}(X) - E[\mathbb{1}(X)|\mathcal{F}]|^2$, $\|\mathbb{1}(X) - E[\mathbb{1}(X)|\mathcal{F}]\|_8 \leq$

$\|\mathbb{1}(X) - E[\mathbb{1}(X)|\mathcal{F}]\|_2^{1/4}$. Therefore, it is sufficient to establish that $\{\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} > 0 > y_{g1,n}^{*s(b,\theta_0)})\}_{g\in\mathcal{G}_n}$ is $L_2$-NED.

We now proceed similarly to the proof of Proposition 2 by Xu and Lee (2015a). Due to the uniform boundedness of the indicator functions, the triangle inequality implies that

$$
\begin{aligned}
& \|\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} > 0 > y_{g1,n}^{*s(b,\theta_0)}) - E[\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} > 0 > y_{g1,n}^{*s(b,\theta_0)})|\mathcal{F}_{g,n}(s)]\|_2 \\
\leq\ & \|\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} > \epsilon > -\epsilon > y_{g1,n}^{*s(b,\theta_0)}) - E[\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} > \epsilon > -\epsilon > y_{g1,n}^{*s(b,\theta_0)})|\mathcal{F}_{g,n}(s)]\|_2 \\
& + \|\mathbb{1}(0 \leq y_{g1,n}^{*s(b,\theta_n)} \leq \epsilon) - E[\mathbb{1}(0 \leq y_{g1,n}^{*s(b,\theta_n)} \leq \epsilon)|\mathcal{F}_{g,n}(s)]\|_2 \\
& + \|\mathbb{1}(0 \geq y_{g1,n}^{*s(b,\theta_0)} \geq -\epsilon) - E[\mathbb{1}(0 \geq y_{g1,n}^{*s(b,\theta_0)} \geq -\epsilon)|\mathcal{F}_{g,n}(s)]\|_2 \\
\leq\ & \|\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} > \epsilon > -\epsilon > y_{g1,n}^{*s(b,\theta_0)}) - E[\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} > \epsilon > -\epsilon > y_{g1,n}^{*s(b,\theta_0)})|\mathcal{F}_{g,n}(s)]\|_2 \\
& + 2\left(\sqrt{P(0 \leq y_{g1,n}^{*s(b,\theta_n)} \leq \epsilon)} + \sqrt{P(0 \geq y_{g1,n}^{*s(b,\theta_0)} \geq -\epsilon)}\right) \\
\leq\ & \|\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} > \epsilon > -\epsilon > y_{g1,n}^{*s(b,\theta_0)}) - E[\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} > \epsilon > -\epsilon > y_{g1,n}^{*s(b,\theta_0)})|\mathcal{F}_{g,n}(s)]\|_2 + (C_2 + C_3)\sqrt{\epsilon},
\end{aligned}
$$

where the second inequality follows by the conditional Jensen's and Minkowski's inequalities. The existence of finite constants $C_2 > 0$ and $C_3 > 0$ follows from the uniform boundedness of the density function of $y_{g1,n}^{*s(b,\theta)}$, $\theta \in \{\theta_n, \theta_0\}$: Lemma D.1 guarantees the uniform lower and upper bounds of the variances of the latent variables. Considering $\epsilon > 0$ and a sufficiently large $n$ such that $\limsup_{\tilde{s}\to\infty} E[\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} > \epsilon > -\epsilon > y_{g1,n}^{*s(b,\theta_0)})|\mathcal{F}(\tilde{s})] < \limsup_{\tilde{s}\to\infty} E[\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} - y_{g1,n}^{*s(b,\theta_0)} > 2\epsilon)|\mathcal{F}(\tilde{s})] < 1/2$, the last expression can be bounded by

$$
\begin{aligned}
& \|\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} > \epsilon > -\epsilon > y_{g1,n}^{*s(b,\theta_0)}) - E[\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} > \epsilon > -\epsilon > y_{g1,n}^{*s(b,\theta_0)})|\mathcal{F}_{g,n}(s)]\|_2 + (C_2 + C_3)\sqrt{\epsilon}, \\
\leq\ & \|\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} - y_{g1,n}^{*s(b,\theta_0)} > 2\epsilon) - E[\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} - y_{g1,n}^{*s(b,\theta_0)} > 2\epsilon)|\mathcal{F}_{g,n}(s)]\|_2 + (C_2 + C_3)\sqrt{\epsilon},
\end{aligned}
$$

where the inequality follows by noting that $\|\mathbb{1}(X) - E[\mathbb{1}(X)|\mathcal{F}]\|^2 = E[\mathrm{Var}(\mathbb{1}(X)|\mathcal{F})] = E[\mathbb{1}(X)|\mathcal{F}](1 - E[\mathbb{1}(X)|\mathcal{F}])$ is monotonically increasing in $E[\mathbb{1}(X)|\mathcal{F}]$ on the interval $(0, 1/2)$. Let us now denote events $B_\epsilon = \{y_{g1,n}^{*s(b,\theta_n)} - y_{g1,n}^{*s(b,\theta_0)} \in (\epsilon, 3\epsilon), E[y_{g1,n}^{*s(b,\theta_n)} - y_{g1,n}^{*s(b,\theta_0)}|\mathcal{F}_{g,n}(s)] \in (\epsilon, 3\epsilon)\}$ for any $\epsilon > 0$ and $B_\epsilon^C$ its complement. By Theorem 10.12 in Davidson (1994), the last expression can be further bounded as in Proposition 2 in Xu and Lee (2015a) by

$$
\begin{aligned}
& \|\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} - y_{g1,n}^{*s(b,\theta_0)} > 2\epsilon) - \mathbb{1}(E[y_{g1,n}^{*s(b,\theta_n)} - y_{g1,n}^{*s(b,\theta_0)} | \mathcal{F}_{g,n}(s)] > 2\epsilon)\|_2 + (C_2 + C_3)\sqrt{\epsilon} \\
\leq \; & \left\{ \int_{B_\epsilon^C} (y_{g1,n}^{*s(b,\theta_n)} - y_{g1,n}^{*s(b,\theta_0)} - E[y_{g1,n}^{*s(b,\theta_n)} - y_{g1,n}^{*s(b,\theta_0)} | \mathcal{F}_{g,n}(s)])^2 dP/\epsilon^2 + P(B_\epsilon) \right\}^{1/2} + (C_2 + C_3)\sqrt{\epsilon} \\
\leq \; & \left\| y_{g1,n}^{*s(b,\theta_n)} - y_{g1,n}^{*s(b,\theta_0)} - E[y_{g1,n}^{*s(b,\theta_n)} - y_{g1,n}^{*s(b,\theta_0)} | \mathcal{F}_{g,n}(s)] \right\|_2 /\epsilon + (C_2 + C_3 + C_4)\sqrt{\epsilon}
\end{aligned}
$$

for some constant $C_4 > 0$, where the first inequality follows from $|\mathbb{1}(y_1 > 2\epsilon) - \mathbb{1}(y_2 > 2\epsilon)| \leq |y_1 - y_2|/\epsilon \mathbb{1}(y_1 \notin (\epsilon, 3\epsilon)$ or $y_2 \notin (\epsilon, 3\epsilon)) + \mathbb{1}(y_1 \in (\epsilon, 3\epsilon)$ and $y_2 \in (\epsilon, 3\epsilon))$ and the second inequality from $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ and from $P(B_\epsilon) \leq P(y_{g1,n}^{*s(b,\theta_n)} - y_{g1,n}^{*s(b,\theta_0)} \in (\epsilon, 3\epsilon)) \leq C_4\epsilon$, which follows again as above from Lemma D.1.

Finally, setting $\epsilon = \left\| y_{g1,n}^{*s(b,\theta_n)} - y_{g1,n}^{*s(b,\theta_0)} - E[y_{g1,n}^{*s(b,\theta_n)} - y_{g1,n}^{*s(b,\theta_0)} | \mathcal{F}_{g,n}(s)] \right\|_2^{2/3}$ and Lemma D.6 imply for uniformly bounded NED scaling factors

$$
t_{g,n} = 2 \sup_{n,g} \sup_{\theta \in \Theta} \sum_{\dot{g} \in \mathcal{G}_n} \left[ \left\| \frac{\partial S_{g\dot{g},n}^s(\lambda^s)}{\partial \lambda} \right\| + \|S_{g\dot{g},n}^s(\lambda^s)\| \right] \left( \sup_{n,g} \|X_{g\cdot,n}^s\|_2 (1 + \|\beta_0^s\|) + \sup_{n,g} \|u_{g,n}^s\|_2 \right)
$$

(derived in the proof of Lemma D.6) that

$$
\begin{aligned}
& \|\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} > 0 > y_{g1,n}^{*s(b,\theta_0)}) - E[\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} > 0 > y_{g1,n}^{*s(b,\theta_0)}) | \mathcal{F}_{g,n}(s)]\|_2 \\
\leq \; & \left\| y_{g1,n}^{*s(b,\theta_n)} - y_{g1,n}^{*s(b,\theta_0)} - E[y_{g1,n}^{*s(b,\theta_n)} - y_{g1,n}^{*s(b,\theta_0)} | \mathcal{F}_{g,n}(s)] \right\|_2^{1/3} \cdot (1 + C_2 + C_3 + C_4) \\
\leq \; & O(n^{-1/6}) t_{g,n}^{1/3} \tilde{\psi}^{1/3}(s)
\end{aligned}
$$

for any $s \in \mathbb{N}$. It follows that $\{\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} > 0 > y_{g1,n}^{*s(b,\theta_0)})\}_{g \in \mathcal{G}_n}$ is $L_8$-NED with the NED coefficients bounded by $\tilde{\psi}^{1/12}(s)$ and the scaling factors $O(n^{-1/24}) t_{g,n}^{1/12}$. The product rule in Lemma C.9 then implies that $\{n^\alpha \mathbb{1}(d_{g,n}^{11(b,\theta_n)} \neq d_{g,n}^{11(b,\theta_0)})\}_{g \in \mathcal{G}_n}$ is $L_4$-NED with NED coefficients bounded by $\tilde{\psi}^{1/12}(s)$ and uniformly bounded NED scaling factors provided that $\alpha \leq 1/24$. Given that $\{d_{g,n}^{11(b,\theta)}\}_{g \in \mathcal{G}_n}$, $\theta \in \{\theta_0, \theta_n\}$, is uniformly $L_4$-NED with NED coefficients bounded by $\tilde{\psi}^{1/6}(s)$ as is shown in the proof of Theorem 2, $\{n^\alpha \mathbb{1}(d_{g,n}^{11(b,\theta_n)} \neq d_{g,n}^{11(b,\theta_0)}) d_{g,n}^{11(b,\theta)}\}_{g \in \mathcal{G}_n}$, $\theta \in \{\theta_0, \theta_n\}$, is uniformly $L_2$-NED with NED coefficients bounded by $\tilde{\psi}^{1/12}(s)$. Similarly to the verification of the NED property in the proof of Theorem 2, we can show that $\{n^\alpha \mathbb{1}(d_{g,n}^{11(b,\theta_n)} \neq d_{g,n}^{11(b,\theta_0)}) d_{g,n}^{11(b,\theta)} \partial f_{g,n}^{11(b,\theta)}(\theta_n)/\partial\theta\}_{g \in \mathcal{G}_n}$ is uniformly $L_2$-NED with NED coefficients $\tilde{\psi}^{(r-2)/(24r-24)}(s)$. The result then follows by the triangle inequality.

To verify the $L_{2+\delta}$-boundedness, $\delta > 0$, of the elements of (E.13) with $a = 11$, note that the likelihood scores have been shown to have finite $p$th moments in the proof of Theorem 1 (property SE) for $p > 4$. By the

Cauchy-Schwartz and Loève's $c_r$-inequalities, it is thus sufficient to verify that $n^{\alpha p} E[\mathbb{1}(d_{g,n}^{11(b,\theta_n)} \neq d_{g,n}^{11(b,\theta_0)})]$ is bounded uniformly in $g$ and $n$ for some $\alpha > 0$ and $p > 4$. The decomposition in (E.15) and the Cauchy-Schwartz inequality imply that

$$E[\mathbb{1}(d_{g,n}^{11(b,\theta_n)} \neq d_{g,n}^{11(b,\theta_0)})] \leq E[\mathbb{1}(d_{g1,n}^{11(b,\theta_n)} \neq d_{g1,n}^{11(b,\theta_0)})] + E[\mathbb{1}(d_{g2,n}^{11(b,\theta_n)} \neq d_{g2,n}^{11(b,\theta_0)})] + \sqrt{E[\mathbb{1}(d_{g1,n}^{11(b,\theta_n)} \neq d_{g1,n}^{11(b,\theta_0)})] E[\mathbb{1}(d_{g2,n}^{11(b,\theta_n)} \neq d_{g2,n}^{11(b,\theta_0)})]}.$$

Similarly to (E.16), we can decompose $\mathbb{1}(d_{gj,n}^{11(b,\theta_n)} \neq d_{gj,n}^{11(b,\theta_0)})$, $j \in \{1,2\}$, into a summation of two terms and calculate the expectation for each of them separately. For example for $\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} > 0 > y_{g1,n}^{*s(b,\theta_0)})$, we can again write for any $\epsilon > 0$ that

$$\begin{aligned}
& E[\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} > 0 > y_{g1,n}^{*s(b,\theta_0)})] \\
\leq\ & E[\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} > \epsilon > -\epsilon > y_{g1,n}^{*s(b,\theta_0)})] + E[\mathbb{1}(0 \leq y_{g1,n}^{*s(b,\theta_n)} \leq \epsilon)] + E[\mathbb{1}(0 \geq y_{g1,n}^{*s(b,\theta_0)} \geq -\epsilon)] \\
=\ & E[\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} > \epsilon > -\epsilon > y_{g1,n}^{*s(b,\theta_0)})] + P(0 \leq y_{g1,n}^{*s(b,\theta_n)} \leq \epsilon) + P(0 \geq y_{g1,n}^{*s(b,\theta_0)} \geq -\epsilon) \\
\leq\ & E[\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} > \epsilon > -\epsilon > y_{g1,n}^{*s(b,\theta_0)})] + (C_5 + C_6)\epsilon \\
\leq\ & E[|y_{g1,n}^{*s(b,\theta_n)} - y_{g1,n}^{*s(b,\theta_0)}|/\epsilon] + (C_5 + C_6)\epsilon,
\end{aligned}$$

where the existence of finite constants $C_5 > 0$ and $C_6 > 0$ follows from Lemma D.1, which guarantees the uniform lower and upper bounds of the variances of the latent variables. After setting $\epsilon = \{E|y_{g1,n}^{*s(b,\theta_n)} - y_{g1,n}^{*s(b,\theta_0)}|\}^{1/2}$, the last term is bounded by $\{E|y_{g1,n}^{*s(b,\theta_n)} - y_{g1,n}^{*s(b,\theta_0)}|\}^{1/2}(1 + C_5 + C_6)$. Lemma D.6 along with Liapunov's inequality implies that $E|y_{g1,n}^{*s(b,\theta_n)} - y_{g1,n}^{*s(b,\theta_0)}| = O(n^{-1/2})$ and thus $E[\mathbb{1}(y_{g1,n}^{*s(b,\theta_n)} > 0 > y_{g1,n}^{*s(b,\theta_0)})] = O(n^{-1/4})$. It can be shown that $n^{\alpha p} E[\mathbb{1}(d_{g,n}^{11(b,\theta_n)} \neq d_{g,n}^{11(b,\theta_0)})] = n^{\alpha p} O(n^{-1/4}) = O(1)$ as $n \to \infty$ if $\alpha p \leq 1/4$ and thus $\alpha \leq 1/(4p)$.

For $p > 4$, we have thus verified that (E.13) satisfies the $L_2$-NED property and the $L_{p/2}$-boundedness conditions in Theorem C.3 if $0 < \alpha \leq \min\{1/24, 1/(4p)\}$. The proof of Theorem C.3, that is, of Corrolary 1 and Theorem 2 of Jenish and Prucha (2012) thus implies that the variance of (E.13) behaves as $n^{-2\alpha} n^{-1} O(n) = O(n^{-2\alpha})$ as $n \to \infty$. By the Chebyshev inequality, $P(|(\text{E.13})| > \epsilon) \leq \mathrm{var}((\text{E.13}))/\epsilon^2 = O(n^{-2\alpha})/\epsilon^2$ for any $\epsilon > 0$, and therefore, (E.13) in asymptotically negligible in probability.

Finally, the second term (E.14) is a difference of two score functions continuous and differentiable in data $y_{i,n}^{*s}$ and $y_{i,n}^{*o}$ generated using $\theta_n \in U(\theta_0, n^{-1/2}M)$ and $\theta_0$ (see the proof of Lemma D.6), and it can therefore be shown to be asymptotically negligible in probability similarly to (E.12).

Hence, it follows that $J(\theta_0)^{-1/2}\Gamma_n^{(b,\hat{\theta}_n)}(\hat{\theta}_n)$ converges in distribution to $\mathcal{N}(0, I_L)$ as $n \to \infty$, which concludes the proof. $\square$


## Acknowledgements

Authors thank Jaap Abbring, Otilia Boldea, Giuseppe Cavaliere, Raymond Florax, Bas van der Klaauw, Áureo de Paula, Ingmar Prucha, Arthur van Soest, Vaidotas Zemlys, the editor Han Hong, an associate editor, and three anonymous referees for valuable comments and discussions and Alfonso Flores-Lagunes for providing the code for the GMM estimator used in the Monte Carlo experiments. We also thank seminar participants at Tilburg University and VU Amsterdam and participants at the 14th International Workshop on Spatial Econometrics and Statistics (Paris, 2015), the 2nd Annual Conference of the International Association for Applied Econometrics (Thessaloniki, 2015), the 10th Meeting of the Netherlands Econometrics Study Group (Maastricht, 2015), the 4th Annual Lithuanian Conference on Economic Research (Kaunas, 2015), the 26th $(EC)^2$ Conference (Edinburgh, 2015), and the 1st Southampton Workshop in Econometrics & Statistics (Southampton, 2018) for helpful comments.

## References


Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*, 817–858.

Andrews, D. (1994). Empirical process method in econometrics. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics, Vol. 4*, pp. 2247–2294. Elsevier.

Booij, A. S., E. Leuven, and H. Oosterbeek (2016). Ability Peer Effects in University: Evidence from a Randomized Experiment. *The Review of Economic Studies 84*(2), 547–578.

Buchinsky, M. (1998). The dynamics of changes in the female wage distribution in the USA: A quantile regression approach. *Journal of Applied Econometrics 13*(1), 1–30.

Calvò-Armengol, A., E. Patacchini, and Y. Zenou (2009). Peer effects and social networks in education. *Review of Economic Studies 76*(4), 1239–1267.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics 90*, 414–427.

Cavaliere, G. and I. Georgiev (2017). Bootstrap inference under random distributional limits. *Discussion paper 1/17, Research Institute for Econometrics, Bar-Ilan University*.

Chen, S. and Y. Zhou (2010). Semiparametric and nonparametric estimation of sample selection models under symmetry. *Journal of Econometrics 157*(1), 143–150.

Conley, T. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics 92*(1), 1–45.

Davidson, J. (1994). *Stochastic Limit Theory.* Oxford University Press.

Doğan, O. and S. Taşpınar (2018). Bayesian inference in spatial sample selection models. *Oxford Bulletin of Economics and Statistics 80*(1), 90–121.

Flores-Lagunes, A. and K. Schnier (2012). Estimation of sample selection models with spatial dependence. *Journal of Applied Econometrics 27*(2), 173–204.

Heckman, J. (1974). Shadow prices, market wages, and labor supply. *Econometrica 42*(4), 679–694.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161.

Hong, H., A. Mahajan, and D. Nekipelov (2015). Extremum estimation and numerical derivatives. *Journal of Econometrics 188*, 250–263.

Horn, R. and C. Johnson (1985). *Matrix Analysis*. Cambridge University Press.

Hoshino, T. (2019). Two-step estimation of incomplete information social interaction models with sample selection. *Journal of Business & Economic Statistics 37*(4), 598–612.

Jenish, N. and I. R. Prucha (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics 150*(1), 86–98.

Jenish, N. and I. R. Prucha (2012). On spatial processes and asymptotic inference under near-epoch dependence. *Journal of Econometrics 170*(1), 178–190.

Kelejian, H. and I. R. Prucha (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics 17*(1), 99–121.

Kelejian, H. and I. R. Prucha (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review 40*(2), 509–533.

Kelejian, H. and I. R. Prucha (2007). HAC estimation in a spatial framework. *Journal of Econometrics 140*(1), 131–154.

Kelejian, H. and I. R. Prucha (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics 157*(1), 53–67.

Kim, M. and Y. Sun (2011). Spatial heteroskedasticity and autocorrelation consistent estimation of covariance matrix. *Journal of Econometrics 160*(2), 349–371.

Lee, L.-F. (2004). Asymptotic distribution of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica 76*(6), 1899–1925.

LeSage, J. and R. Pace (2009). *Introduction to Spatial Econometrics*. CRC Press.

Lin, X. (2010). Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics 28*(4), 825–860.

MacKinnon, J. G. and M. D. Webb (2017). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics 32*, 233–254.

McMillen, D. (1995). Selection bias in spatial econometric models. *Journal of Regional Science 35*(3), 417–436.

Mukherjee, B. and D. A. Singer (2010). International institutions and domestic compensation: The IMF and the politics of capital account liberalization. *American Journal of Political Science 54*(1), 45–60.

Newey, W. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics, Vol. 4*, pp. 2111–2245. Elsevier.

Newey, W. and K. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Pinkse, J. and M. Slade (1998). Contracting in space: An application of spatial statistics to discrete-choice models. *Journal of Econometrics 85*(1), 125–154.

Poirier, D. and P. Ruud (1988). Probit with dependent observations. *Review of Economic Studies 55*(4), 593–614.

Pötscher, B. M. and I. Prucha (1997). *Dynamic Nonlinear Econometric Models: Asymptotic Theory.* Springer Heidelberg.

Qu, X. and L.-F. Lee (2012). LM tests for spatial correlation in spatial models with limited dependent variables. *Regional Science and Urban Economics 42*(3), 430–445.

Qu, X. and L.-F. Lee (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics 184*(2), 209–232.

Robinson, P. (1982). On the asymptotic properties of estimators of models containing limited dependent variables. *Econometrica 50*(1), 27–41.

Sharma, A., L. Siciliani, and A. Harris (2013). Waiting times and socioeconomic status: Does sample selection matter? *Economic Modelling 33*(0), 659–667.

Su, L. and Z. Yang (2015). QML estimation of dynamic panel data models with spatial errors. *Journal of Econometrics 185*(1), 230–258.

Vella, F. (1998). Estimating models with sample selection bias: A survey. *The Journal of Human Resources 33*(1), 127–169.

Wang, H., E. M. Iglesias, and J. M. Wooldridge (2013). Partial maximum likelihood estimation of spatial probit models. *Journal of Econometrics 172*(1), 77–89.

Wang, W. and L.-F. Lee (2013). Estimation of spatial autoregressive models with randomly missing data in the dependent variable. *The Econometrics Journal 16*(1), 73–102.

Ward, P. S., R. J. G. M. Florax, and A. Flores-Lagunes (2014). Climate change and agricultural productivity in Sub-Saharan Africa: a spatial sample selection model. *European Review of Agricultural Economics 41(2)*, 199–226.

Wooldridge, J. M. (1994). Estimation and inference for dependent processes. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics, Vol. 4*, pp. 2639 – 2738. Elsevier.

Xu, X. and L.-F. Lee (2015a). Maximum likelihood estimation of a spatial autoregressive Tobit model. *Journal of Econometrics 188*(1), 264–280.

Xu, X. and L.-F. Lee (2015b). A spatial autoregressive model with a nonlinear transformation of the dependent variable. *Journal of Econometrics 186*(1), 1–18.

Xu, X. and L.-F. Lee (2018). Sieve maximum likelihood estimation of the spatial autoregressive Tobit model. *Journal of Econometrics 203*(1), 96–112.

# Supplementary Material for "Estimation of Spatial Sample Selection Models: A Partial Maximum Likelihood Approach"

Renata Rabovič* Pavel Čížek†

February 27, 2020

This supplementary material provides the derivations of the likelihood function and marginal effects, the verification of the first-order conditions, an example of sufficient conditions for Assumptions 1($ii$), 5, and 10, two additional graphs, two additional technical lemmas and their proofs as well as proofs of Lemmas C.4–C.9 and D.1–D.6 in Appendices C and D, respectively.

## Appendix G The Likelihood Function, First-order Conditions, and Marginal Effects

### G.1 The likelihood function

There are four scenarios: $y^s_{g1,n} = 1$ and $y^s_{g2,n} = 0$, $y^s_{g1,n} = 0$ and $y^s_{g2,n} = 1$, $y^s_{g1,n} = y^s_{g2,n} = 1$, and $y^s_{g1,n} = y^s_{g2,n} = 0$. We derive the log-likelihood contribution based on the third scenario, while for the other scenarios it can be done in a similar way. Let $f(\cdot)$ without any index denote a generic density function. Then the Bayes rule and Assumption 2($ii$) imply that $d^{11}_{g,n} f(y^s_{g,n} = \iota_2, y^o_{g,n}|X^s_n, X^o_n) = d^{11}_{g,n} f(y^s_{g,n} = \iota_2, y^{*o}_{g,n}|X^s_n, X^o_n) = P[y^s_{g,n} = \iota_2|y^{*o}_{g,n}, X^s_n, X^o_n] \cdot d^{11}_{g,n} f(y^{*o}_{g,n}|X^o_n)$. By Assumption 2, $y^{*o}_{g,n}|X^o_n \sim \mathcal{N}(S^o_{g\cdot,n}(\lambda^o)X^o_n\beta^o, \Omega^{oo}_{g,n}(\theta))$, thus $d^{11}_{g,n} f(y^{*o}_{g,n}|X^o_n) =$

*Corresponding author. Faculty of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD, United Kingdom. Email: rr574@cam.ac.uk, Phone: +44 122 333 5283.

†CentER, Dept. of Econometrics & Operations Research, Tilburg University, P.O. Box 90153, 5000LE Tilburg, The Netherlands. Email: P.Cizek@tilburguniversity.edu, Phone: +31 13 466 8723.

$d^{11}_{g,n}\phi_2(y^{*o}_{g,n} - S^o_{g\cdot,n}(\lambda^o)X^o_n\beta^o, \Omega^{oo}_{g,n}(\theta)) = d^{11}_{g,n}\phi_2(y^o_{g,n} - S^o_{g\cdot,n}(\lambda^o)X^o_n\beta^o, \Omega^{oo}_{g,n}(\theta))$. Next,

$$\begin{aligned}
P[y^s_{g,n} = \iota_2 | y^{*o}_{g,n}, X^s_n, X^o_n] &= P[y^{*s}_{g,n} > 0 | y^{*o}_{g,n}, X^s_n, X^o_n] \\
&= P[S^s_{g\cdot,n}(\lambda^s)X^s_n\beta^s + \varepsilon^s_{g,n}(\lambda^s) > 0 | y^{*o}_{g,n}, X^s_n, X^o_n] \\
&= P[-\varepsilon^s_{g,n}(\lambda^s) < S^s_{g\cdot,n}(\lambda^s)X^s_n\beta^s | \varepsilon^o_{g,n}(\lambda^o), X^s_n, X^o_n] \\
&= P[-\varepsilon^s_{g,n}(\lambda^s) < \tilde{S}^s_{g\cdot,n}(\lambda^s)X^s_n\beta^s | \varepsilon^o_{g,n}(\lambda^o), X^s_n, X^o_n],
\end{aligned}$$

where $\tilde{S}^s_{g\cdot,n}(\lambda)$ is defined in Section 3 with $\zeta_{g,n} = \iota_2$ in this case. Given the definitions of $\tilde{\Omega}^{ss}_{g,n}(\theta)$, $\tilde{\Omega}^{so}_{g,n}(\theta)$, and $\Omega^{oo}_{g,n}(\theta)$ in Section 3 with $\zeta_{g,n} = \iota_2$, note that

$$\left( \begin{array}{c} -\varepsilon^s_{g,n}(\lambda^s) \\ \varepsilon^o_{g,n}(\lambda^o) \end{array} \middle| X^s_n, X^o_n \right) \sim \mathcal{N}\left( 0, \begin{pmatrix} \tilde{\Omega}^{ss}_{g,n}(\theta) & \tilde{\Omega}^{so}_{g,n}(\theta) \\ \tilde{\Omega}^{so\prime}_{g,n}(\theta) & \Omega^{oo}_{g,n}(\theta) \end{pmatrix} \right).$$

Thus, $-\varepsilon^s_{g,n}(\lambda^s)|\varepsilon^o_{g,n}(\lambda^o), X^s_n, X^o_n \sim \mathcal{N}(\tilde{\Omega}^{so}_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)\varepsilon^o_{g,n}(\lambda^o), \Sigma^{11}_{g,n}(\theta))$, where $\Sigma^{11}_{g,n}(\theta) = \tilde{\Omega}^{ss}_{g,n}(\theta) - \tilde{\Omega}^{so}_{g,n}(\theta)\,\Omega^{oo-1}_{g,n}(\theta)\tilde{\Omega}^{so\prime}_{g,n}(\theta)$. Substituting for $\varepsilon^o_{g,n}(\lambda^o)$ from model (3) and interchanging $y^{*o}_{g,n}$ and $y^o_{g,n}$ as before, the likelihood contribution equals $d^{11}_{g,n}P[y^s_{g,n} = \iota_2|y^{*o}_{g,n}, X^s_n, X^o_n]\cdot f(y^{*o}_{g,n}|X^o_n) = d^{11}_{g,n}\Phi_2(\tilde{S}^s_{g\cdot,n}(\lambda^s)X^s_n\beta^s - \mu^{11}_{g,n}(\theta), \Sigma^{11}_{g,n}(\theta))\ \phi_2\left(y^o_{g,n} - S^o_{g\cdot,n}(\lambda^o)X^o_n\beta^o, \Omega^{oo}_{g,n}(\theta)\right)$, where $\mu^{11}_{g,n}(\theta) = \tilde{\Omega}^{so}_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)(y^o_{g,n} - S^o_{g\cdot,n}(\lambda^o)X^o_n\beta^o)$. The result in (4) follows by noting that $z_{g,n}(\theta) = y^o_{g,n} - S^o_{g\cdot,n}(\lambda^o)X^o_n\beta^o$ and $v^{11}_{g,n}(\theta) = \mathrm{Diag}(\Sigma^{11}_{g,n}(\theta))^{-1/2}(\tilde{S}^s_{g\cdot,n}(\lambda^s)X^s_n\beta^s - \mu^{11}_{g,n}(\theta))$. The log-likelihood contributions based on the other scenarios can be obtained similarly, see (4).

It is interesting to note that this likelihood function becomes identical to the HMLE likelihood function if the pairs $g$ of observations are selected in such a way that the observations in the pair are independent of each other, that is, $\Omega^{ss}_{g12,n} = \Omega^{so}_{g12,n} = \Omega^{so}_{g21,n} = \Omega^{oo}_{g12,n} = 0$. Then the bivariate density and distribution functions $\phi_2$ and $\Phi_2$ are just the products of the corresponding marginal density and distribution functions such as $\Phi_2\left(v^{00}_{g,n}(\theta), R^{00}_{g,n}(\theta)\right) = \Phi(v^{00}_{g1,n}(\theta)) \cdot \Phi(v^{00}_{g2,n}(\theta))$, and the logarithms of these products are just sums of the logarithms of the univariate density and distribution functions.

## G.2 First-order conditions

Although (4) does not represent the full MLE, the likelihood function is correctly specified for each grouped pair of observations, and therefore, is maximized at the true value of the parameters $\theta_0$ by the information inequality. Therefore, we can show that the population first-order conditions hold at $\theta_0$ for any pair $g \in \mathcal{G}_n$. For a given $g$, consider first the conditional expectation of the first-order conditions

$$
\begin{aligned}
E\Bigg\{ \frac{\partial}{\partial\theta}\Bigg( & \mathbb{1}(y^s_{g1,n}=1, y^s_{g2,n}=0)\ln\left[\frac{1}{\sqrt{\Omega^{oo}_{g11,n}(\theta)}}\phi\left(\frac{z_{g1,n}(\theta)}{\sqrt{\Omega^{oo}_{g11,n}(\theta)}}\right)\Phi_2\left(v^{10}_{g,n}(\theta), R^{10}_{g,n}(\theta)\right)\right] \\
& + \mathbb{1}(y^s_{g1,n}=0, y^s_{g2,n}=1)\ln\left[\frac{1}{\sqrt{\Omega^{oo}_{g22,n}(\theta)}}\phi\left(\frac{z_{g2,n}(\theta)}{\sqrt{\Omega^{oo}_{g22,n}(\theta)}}\right)\Phi_2\left(v^{01}_{g,n}(\theta), R^{01}_{g,n}(\theta)\right)\right] \\
& + \mathbb{1}(y^s_{g1,n}=1, y^s_{g2,n}=1)\ln\left[\phi_2\left(z_{g,n}(\theta), \Omega^{oo}_{g,n}(\theta)\right)\Phi_2\left(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta)\right)\right] \\
& + \mathbb{1}(y^s_{g1,n}=0, y^s_{g2,n}=0)\ln\left[\Phi_2\left(v^{00}_{g,n}(\theta), R^{00}_{g,n}(\theta)\right)\right]\Bigg)\Bigg| X^s_n, X^o_n\Bigg\},
\end{aligned}
\tag{G.1}
$$

which should equal to zero at the true parameter values. Since $E[\mathbb{1}(y^s_{g,n}=a)\partial\ln(\cdot)/\partial\theta|X^s_n, X^o_n] = E[\partial\ln(\cdot)/\partial\theta|y^s_{g,n} = a, X^s_n, X^o_n]\ P[y^s_{g,n} = a|X^s_n, X^o_n]$, where $a = (a_1, a_2)'$, we can rewrite this first-order derivative using notation $E^{10a_1+a_2}[\cdot] = E[\cdot|y^s_{g,n} = a, X^s_n, X^o_n]$ as

$$
\begin{aligned}
& E^{10}\left\{\left[\frac{\phi\left(\frac{z_{g1,n}(\theta)}{\sqrt{\Omega^{oo}_{g11,n}(\theta)}}\right)}{\sqrt{\Omega^{oo}_{g11,n}(\theta)}}\right]^{-1}\frac{\partial}{\partial\theta}\left[\frac{1}{\sqrt{\Omega^{oo}_{g11,n}(\theta)}}\phi\left(\frac{z_{g1,n}(\theta)}{\sqrt{\Omega^{oo}_{g11,n}(\theta)}}\right)\Phi_2\left(v^{10}_{g,n}(\theta), R^{10}_{g,n}(\theta)\right)\right]\frac{P[y^s_{g,n}=(1,0)'|X^s_n, X^o_n]}{\Phi_2\left(v^{10}_{g,n}(\theta), R^{10}_{g,n}(\theta)\right)}\right\} \\
& +E^{01}\left\{\left[\frac{\phi\left(\frac{z_{g2,n}(\theta)}{\sqrt{\Omega^{oo}_{g22,n}(\theta)}}\right)}{\sqrt{\Omega^{oo}_{g22,n}(\theta)}}\right]^{-1}\frac{\partial}{\partial\theta}\left[\frac{1}{\sqrt{\Omega^{oo}_{g22,n}(\theta)}}\phi\left(\frac{z_{g2,n}(\theta)}{\sqrt{\Omega^{oo}_{g22,n}(\theta)}}\right)\Phi_2\left(v^{01}_{g,n}(\theta), R^{01}_{g,n}(\theta)\right)\right]\frac{P[y^s_{g,n}=(0,1)'|X^s_n, X^o_n]}{\Phi_2\left(v^{01}_{g,n}(\theta), R^{01}_{g,n}(\theta)\right)}\right\} \\
& +E^{11}\left\{\left[\phi_2\left(z_{g,n}(\theta), \Omega^{oo}_{g,n}(\theta)\right)\right]^{-1}\frac{\partial}{\partial\theta}\left[\phi_2\left(z_{g,n}(\theta), \Omega^{oo}_{g,n}(\theta)\right)\Phi_2\left(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta)\right)\right]\frac{P[y^s_{g,n}=(1,1)'|X^s_n, X^o_n]}{\Phi_2\left(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta)\right)}\right\} \\
& +E^{00}\left\{\frac{\partial}{\partial\theta}\left[\Phi_2\left(v^{00}_{g,n}(\theta), R^{00}_{g,n}(\theta)\right)\right]\frac{P[y^s_{g,n}=(0,0)'|X^s_n, X^o_n]}{\Phi_2\left(v^{00}_{g,n}(\theta), R^{00}_{g,n}(\theta)\right)}\right\}.
\end{aligned}
$$

Next, the product rule applied to the partial derivatives in the above expression results in a sum of two parts. The first part contains the derivatives of the logarithm of the bivariate normal distribution functions:

$$
E^{10}\left\{\frac{\partial\Phi_2\left(v^{10}_{g,n}(\theta), R^{10}_{g,n}(\theta)\right)}{\partial\theta}\frac{P[y^s_{g,n}=(1,0)'|X^s_n, X^o_n]}{\Phi_2\left(v^{10}_{g,n}(\theta), R^{10}_{g,n}(\theta)\right)}\right\} \tag{G.2}
$$

$$
+E^{01}\left\{\frac{\partial\Phi_2\left(v^{01}_{g,n}(\theta), R^{01}_{g,n}(\theta)\right)}{\partial\theta}\frac{P[y^s_{g,n}=(0,1)'|X^s_n, X^o_n]}{\Phi_2\left(v^{01}_{g,n}(\theta), R^{01}_{g,n}(\theta)\right)}\right\} \tag{G.3}
$$

$$
+E^{11}\left\{\frac{\partial\Phi_2\left(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta)\right)}{\partial\theta}\frac{P[y^s_{g,n}=(1,1)'|X^s_n, X^o_n]}{\Phi_2\left(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta)\right)}\right\} \tag{G.4}
$$

$$+E^{00}\left\{\frac{\partial\Phi_2\left(v^{00}_{g,n}(\theta),R^{00}_{g,n}(\theta)\right)}{\partial\theta}\frac{P[y^s_{g,n}=(0,0)'|X^s_n,X^o_n]}{\Phi_2\left(v^{00}_{g,n}(\theta),R^{00}_{g,n}(\theta)\right)}\right\}, \tag{G.5}$$

see Lemma C.4 for the form of the derivatives in the above expression. The second part contains the partial derivatives of the univariate and bivariate normal densities:

$$E^{10}\left\{\left[\frac{\phi\left(\frac{z_{g1,n}(\theta)}{\sqrt{\Omega^{oo}_{g11,n}(\theta)}}\right)}{\sqrt{\Omega^{oo}_{g11,n}(\theta)}}\right]^{-1}\frac{\partial}{\partial\theta}\left[\frac{1}{\sqrt{\Omega^{oo}_{g11,n}(\theta)}}\phi\left(\frac{z_{g1,n}(\theta)}{\sqrt{\Omega^{oo}_{g11,n}(\theta)}}\right)\right]P[y^s_{g,n}=(1,0)'|X^s_n,X^o_n]\right\} \tag{G.6}$$

$$+E^{01}\left\{\left[\frac{\phi\left(\frac{z_{g2,n}(\theta)}{\sqrt{\Omega^{oo}_{g22,n}(\theta)}}\right)}{\sqrt{\Omega^{oo}_{g22,n}(\theta)}}\right]^{-1}\frac{\partial}{\partial\theta}\left[\frac{1}{\sqrt{\Omega^{oo}_{g22,n}(\theta)}}\phi\left(\frac{z_{g2,n}(\theta)}{\sqrt{\Omega^{oo}_{g22,n}(\theta)}}\right)\right]P[y^s_{g,n}=(0,1)'|X^s_n,X^o_n]\right\} \tag{G.7}$$

$$+E^{11}\left\{\left[\phi_2\left(z_{g,n}(\theta),\Omega^{oo}_{g,n}(\theta)\right)\right]^{-1}\frac{\partial}{\partial\theta}\left[\phi_2\left(z_{g,n}(\theta),\Omega^{oo}_{g,n}(\theta)\right)\right]P[y^s_{g,n}=(1,1)'|X^s_n,X^o_n]\right\}. \tag{G.8}$$

We have to show now that the sum of the two parts, that is the sum of (G.2)–(G.8), equals zero. Given that the verification can be done in a similar though not identical way for each parameter in $\theta$ and that the proof is analogous in the univariate and bivariate cases, we verify the conditions first in a simple case of parameters $\beta^o$ in (G.2) and (G.6) and then provide a more general example of $\sigma$ in (G.4) and (G.8); the remaining terms and parameters can be handled analogously to show the validity of the first-order conditions.

First, consider the sum of (G.2) and (G.6) and its derivatives with respect to $\beta^o$ evaluated at $\theta_0$:

$$E^{10}\left\{\frac{\frac{\partial\Phi_2\left(v^{10}_{g,n}(\theta),R^{10}_{g,n}(\theta)\right)}{\partial\beta^o}\big|_{\theta=\theta_0}}{\Phi_2\left(v^{10}_{g,n}(\theta_0),R^{10}_{g,n}(\theta_0)\right)}+\left[\frac{\phi\left(\frac{z_{g1,n}(\theta_0)}{\sqrt{\Omega^{oo}_{g11,n}(\theta_0)}}\right)}{\sqrt{\Omega^{oo}_{g11,n}(\theta_0)}}\right]^{-1}\frac{\frac{\partial z_{g1,n}(\theta)}{\partial\beta^o}\big|_{\theta=\theta_0}}{\Omega^{oo}_{g11,n}(\theta_0)}\phi^{(1)}\left(\frac{z_{g1,n}(\theta_0)}{\sqrt{\Omega^{oo}_{g11,n}(\theta_0)}}\right)\right\}P[y^s_{g,n}=(1,0)'|X^s_n,X^o_n],$$

where $\phi^{(1)}$ is the first-order derivative of the standard normal density function. This conditional expectation can be expressed as an integral with respect to the dependent variable $y^o_{g1,n}$ using its conditional density $f(y^s_{g,n}=(1,0)',y^o_{g1,n}|X^s_n,X^o_n)/P[y^s_{g,n}=(1,0)'|X^s_n,X^o_n]$, where the denominator was derived earlier in this appendix (see also Section 3):

$$\int_{-\infty}^{+\infty}\left\{\frac{\partial\Phi_2\left(v^{10}_{g,n}(\theta),R^{10}_{g,n}(\theta)\right)}{\partial\beta^o}\bigg|_{\theta=\theta_0}\frac{\phi\left(\frac{z_{g1,n}(\theta_0)}{\sqrt{\Omega^{oo}_{g11,n}(\theta_0)}}\right)}{\sqrt{\Omega^{oo}_{g11,n}(\theta_0)}}+\frac{\frac{\partial z_{g1,n}(\theta)}{\partial\beta^o}\big|_{\theta=\theta_0}}{\Omega^{oo}_{g11,n}(\theta_0)}\phi^{(1)}\left(\frac{z_{g1,n}(\theta_0)}{\sqrt{\Omega^{oo}_{g11,n}(\theta_0)}}\right)\Phi_2\left(v^{10}_{g,n}(\theta_0),R^{10}_{g,n}(\theta_0)\right)\right\}dy.$$

Let $u(\theta)=z_{g1,n}(\theta)/\sqrt{\Omega^{oo}_{g11,n}(\theta)}$ with $u=u(\theta_0)$. Then $v^{10}_{g,n}(\theta_0)=(\mathrm{Diag}(\Sigma^{10}_{g,n}(\theta_0)))^{-1/2}[q_{g,n}(\theta_0)-\tilde{\Omega}^{so}_{g\cdot 1,n}(\theta_0)u/\sqrt{\Omega^{oo}_{g11,n}(\theta_0)}]$

and

$$\left.\frac{\partial \Phi_2\left(v_{g,n}^{10}(\theta), R_{g,n}^{10}(\theta)\right)}{\partial \beta^o}\right|_{\theta=\theta_0} = \frac{1}{\sqrt{\Omega_{g11,n}^{oo}(\theta_0)}} \frac{\partial \Phi_2\left(v_{g,n}^{10}(\theta_0), R_{g,n}^{10}(\theta_0)\right)}{\partial u} \left.\frac{\partial z_{g1,n}(\theta)}{\partial \beta^o}\right|_{\theta=\theta_0}.$$

Substituting $u = z_{g1,n}(\theta_0)/\sqrt{\Omega_{g11,n}^{oo}(\theta_0)}$ , the integral above can be rewritten as

$$\int_{-\infty}^{+\infty} \left\{ \frac{\partial}{\partial u}\Phi_2\left(v_{g,n}^{10}(\theta_0), R_{g,n}^{10}(\theta_0)\right) \frac{\frac{\partial z_{g1,n}(\theta)}{\partial \beta^o}\big|_{\theta=\theta_0}}{\sqrt{\Omega_{g11,n}^{oo}(\theta_0)}} \phi\left(u\right) + \frac{\frac{\partial z_{g1,n}(\theta)}{\partial \beta^o}\big|_{\theta=\theta_0}}{\sqrt{\Omega_{g11,n}^{oo}(\theta_0)}} \phi^{(1)}\left(u\right) \Phi_2\left(v_{g,n}^{10}(\theta_0), R_{g,n}^{10}(\theta_0)\right) \right\} du = 0,$$

where the result is obtained by applying integration by parts.

Next, we can apply a similar argument to the parameter $\sigma$ in the sum of (G.4) and (G.8):

$$E^{11}\left\{ \frac{\frac{\partial \Phi_2\left(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta)\right)}{\partial \sigma}\big|_{\theta=\theta_0}}{\Phi_2\left(v_{g,n}^{11}(\theta_0), R_{g,n}^{11}(\theta_0)\right)} + \frac{\frac{\partial \phi_2\left(z_{g,n}(\theta), \Omega_{g,n}^{oo}(\theta)\right)}{\partial \sigma}\big|_{\theta=\theta_0}}{\phi_2\left(z_{g,n}(\theta_0), \Omega_{g,n}^{oo}(\theta_0)\right)} \right\} P[y_{g,n}^s = (1,1)' | X_n^s, X_n^o].$$

We can again write this conditional expectation as an integral with respect to the dependent variable $y_{g,n}^o$ using its conditional density $f(y_{g,n}^s = (1,1)', y_{g,n}^o | X_n^s, X_n^o)/P[y_{g,n}^s = (1,1)' | X_n^s, X_n^o]$, where the denominator was derived in the beginning of this appendix:

$$\iint_{-\infty}^{+\infty} \left\{ \left.\frac{\partial \Phi_2\left(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta)\right)}{\partial \sigma}\right|_{\theta=\theta_0} \phi_2\left(z_{g,n}(\theta_0), \Omega_{g,n}^{oo}(\theta_0)\right) + \left.\frac{\partial \phi_2\left(z_{g,n}(\theta), \Omega_{g,n}^{oo}(\theta)\right)}{\partial \sigma}\right|_{\theta=\theta_0} \Phi_2\left(v_{g,n}^{11}(\theta_0), R_{g,n}^{11}(\theta_0)\right) \right\} dy_1 dy_2. \tag{G.9}$$

Let us first note that $\tilde{\Omega}_{g,n}^{so}(\theta)$ and $\Omega_{g,n}^{oo}(\theta)$ are linearly and quadratically proportional to $\sigma$, respectively, and $z_{g,n}(\theta)$ does not depend on $\sigma$. Using $w(\theta) = [\Omega_{g,n}^{oo}(\theta)]^{-1/2} z_{g,n}(\theta)$ with $w = w(\theta_0)$, $\partial w(\theta)/\partial \sigma = -w(\theta)/\sigma$. Then the derivative

$$\frac{\partial}{\partial \sigma}\left[\phi_2\left(z_{g,n}(\theta), \Omega_{g,n}^{oo}(\theta)\right)\right] = \frac{\partial}{\partial \sigma}[|\Omega_{g,n}^{oo}(\theta)|^{-1/2}\phi_2\left([\Omega_{g,n}^{oo}(\theta)]^{-1/2} z_{g,n}(\theta)\right)] = \frac{\partial}{\partial \sigma}\left[|\Omega_{g,n}^{oo}(\theta)|^{-1/2}\phi_2\left(w(\theta)\right)\right]$$

evaluated at $\theta_0$ can be written as

$$\left[\frac{-2 \cdot \phi_2\left(w\right)}{\sigma_0 |\Omega_{g,n}^{oo}(\theta_0)|^{1/2}}\right] + |\Omega_{g,n}^{oo}(\theta_0)|^{-1/2} \left.\frac{\partial \phi_2(w)}{w'} \frac{\partial w(\theta)}{\partial \sigma}\right|_{\theta=\theta_0} = \left[\frac{-2 \cdot \phi_2\left(w\right)}{\sigma_0 |\Omega_{g,n}^{oo}(\theta_0)|^{1/2}}\right] - |\Omega_{g,n}^{oo}(\theta_0)|^{-1/2} \frac{\partial \phi_2(w)}{\partial w'} \frac{w}{\sigma_0}.$$

Similarly, the fact that $q_{g,n}(\theta)$, $\Sigma_{g,n}^{11}(\theta)$, and $\tilde{\Omega}_{g\cdot 1,n}^{so}(\theta)[\Omega_{g,n}^{oo}(\theta)]^{-1/2}$ do not depend on $\sigma$ and $v_{g,n}^{11}(\theta) = (\mathrm{Diag}(\Sigma_{g,n}^{11}(\theta)))^{-1/2} \times$

$[q_{g,n}(\theta) - \tilde{\Omega}^{so}_{g,n}(\theta)[\Omega^{oo}_{g,n}(\theta)]^{-1/2}w(\theta)]$ imply that

$$\left.\frac{\partial \Phi_2\left(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta)\right)}{\partial \sigma}\right|_{\theta=\theta_0} = \left.\left[\frac{\partial \Phi_2\left(v^{11}_{g,n}(\theta), R^{11}_{g,n}(\theta)\right)}{\partial w(\theta)'}\frac{\partial w(\theta)}{\partial \sigma}\right]\right|_{\theta=\theta_0} = -\frac{\partial \Phi_2\left(v^{11}_{g,n}(\theta_0), R^{11}_{g,n}(\theta_0)\right)}{\partial w'}\frac{w}{\sigma_0}.$$

The integral (G.9) can be thus rewritten using $\phi_2\left(z_{g,n}(\theta), \Omega^{oo}_{g,n}(\theta)\right) = |\Omega^{oo}_{g,n}(\theta)|^{-1/2}\phi_2\left([\Omega^{oo}_{g,n}(\theta)]^{-1/2}z_{g,n}(\theta)\right)$ and substitution $w = [\Omega^{oo}_{g,n}(\theta_0)]^{-1/2}z_{g,n}(\theta_0)$ as the sum of two integrals, one for each partial derivative with respect to $w_k$, $k = 1, 2$:

$$\frac{-1}{\sigma_0|\Omega^{oo}_{g,n}(\theta_0)|}\iint\limits_{-\infty}^{+\infty}\left\{\frac{\partial \Phi_2\left(v^{11}_{g,n}(\theta_0), R^{11}_{g,n}(\theta_0)\right)}{\partial w_k}w_k\phi_2(w) + \frac{\partial[\phi_2(w)w_k]}{\partial w_k}\Phi_2\left(v^{11}_{g,n}(\theta_0), R^{11}_{g,n}(\theta_0)\right)\right\}dw_1 dw_2.$$

Using Fubini's theorem and integration by parts, this integral is again equal to zero for $k = 1, 2$ and thus the first-order condition is satisfied.

The above expressions demonstrate that the first-order conditions can be verified for the parameters of the sample selection model. It is also interesting to note that the above strategy would be also applicable under other elliptically contoured distributions.

## G.3 Marginal effects

**Spatial lag model.** Let $P[y^s = 1|X^s_n] = \left(P[y^s_{1,n} = 1|X^s_n], \ldots, P[y^s_{2n,n} = 1|X^s_n]\right)'$. Then $\partial P[y^s = 1|X^s_n]/\partial X^{s'}_{\cdot l,n} = \text{Diag}(\chi_n(\theta_0))S^s_n(\lambda^s_0)\beta^s_{0l}$ is a matrix of marginal effects associated with a regressor $l$, where

$$\chi_n(\theta_0) = \left(\frac{\phi(b_{1,n}(\theta_0))}{\sqrt{\Omega^{ss}_{11,n}(\theta_0)}}, \ldots, \frac{\phi(b_{2n,n}(\theta_0))}{\sqrt{\Omega^{ss}_{2n2n,n}(\theta_0)}}\right)'$$

with

$$b_n(\theta_0) = \left(\frac{S^s_{1\cdot,n}(\lambda^s_0)X^s_n\beta^s_0}{\sqrt{\Omega^{ss}_{11,n}(\theta_0)}}, \ldots, \frac{S^s_{2n\cdot,n}(\lambda^s_0)X^s_n\beta^s_0}{\sqrt{\Omega^{ss}_{2n2n,n}(\theta_0)}}\right)'.$$

LeSage and Pace (2009) propose to use three types of marginal effects for a spatial lag model: total $((\partial P[y^s = 1|X^s_n]/\partial X^{s'}_{\cdot l,n})\iota_{2n})$, direct $(\text{diag}(\partial P[y^s = 1|X^s_n]/\partial X^{s'}_{\cdot l,n}))$, and indirect that is equal to the difference of the first two marginal effects; $\iota_{2n}$ denotes here the $2n$-dimensional vector of ones. The average total, average direct, and average indirect effects are obtained by calculating the averages of these vectors.

Next, note that

$$
\begin{aligned}
E[y_{i,n}^{*o}|y_{i,n}^{s}=1,X_n^s,X_n^o] &= E[S_{i\cdot,n}^{o}(\lambda_0^o)X_n^o\beta_0^o+\varepsilon_{i,n}^{o}(\lambda_0^o)|y_{i,n}^{s}=1,X_n^s,X_n^o] \\
&= S_{i\cdot,n}^{o}(\lambda_0^o)X_n^o\beta_0^o+E[\varepsilon_{i,n}^{o}(\lambda_0^o)|\varepsilon_{i,n}^{s}(\lambda_0^s)>-S_{i\cdot,n}^{s}(\lambda_0^s)X_n^s\beta_0^s,X_n^s] \\
&= S_{i\cdot,n}^{o}(\lambda_0^o)X_n^o\beta_0^o+\frac{\Omega_{ii,n}^{so}(\theta_0)}{\sqrt{\Omega_{ii,n}^{ss}(\theta_0)\Omega_{ii,n}^{oo}(\theta_0)}}\sqrt{\Omega_{ii,n}^{oo}(\theta_0)}\frac{\phi\left(-\frac{S_{i\cdot,n}^{s}(\lambda_0^s)X_n^s\beta_0^s}{\sqrt{\Omega_{ii,n}^{ss}(\theta_0)}}\right)}{1-\Phi\left(-\frac{S_{i\cdot,n}^{s}(\lambda_0^s)X_n^s\beta_0^s}{\sqrt{\Omega_{ii,n}^{ss}(\theta_0)}}\right)} \\
&= S_{i\cdot,n}^{o}(\lambda_0^o)X_n^o\beta_0^o+\frac{\Omega_{ii,n}^{so}(\theta_0)}{\sqrt{\Omega_{ii,n}^{ss}(\theta_0)}}\frac{\phi\left(\frac{S_{i\cdot,n}^{s}(\lambda_0^s)X_n^s\beta_0^s}{\sqrt{\Omega_{ii,n}^{ss}(\theta_0)}}\right)}{\Phi\left(\frac{S_{i\cdot,n}^{s}(\lambda_0^s)X_n^s\beta_0^s}{\sqrt{\Omega_{ii,n}^{ss}(\theta_0)}}\right)},
\end{aligned}
$$

where the third equality follows by Theorem 24.5 of Greene (2008), which states that if $y$ and $z$ have a bivariate normal distribution with means $\mu_y$ and $\mu_z$, standard deviations $\sigma_y$ and $\sigma_z$, and correlation coefficient $\rho$, then $E[y|z>a]=\mu_y+\rho\sigma_y\phi(\alpha_z)/(1-\Phi(\alpha_z))$ with $\alpha_z=(a-\mu_z)/\sigma_z$. Thus, the marginal effect of $E[y_{i,n}^{*o}|y_{i,n}^{s}=1,X_n^s,X_n^o]$ with respect to $X_{\cdot l,n}^{o}$ depends on whether the explanatory variable is present in both the selection and outcome equations or only in one. Without loss of generality, let the first $L_1$ explanatory variables be the same in both the selection and outcome equations and ordered in the same way, while the remaining $L-L_1$ variables be different. Moreover, denote $E[y^{*o}|y^s=1,X_n^s,X_n^o]=\left(E[y_{1,n}^{*o}|y_{1,n}^{s}=1,X_n^s,X_n^o],\ldots,E[y_{2n,n}^{*o}|y_{2n,n}^{s}=1,X_n^s,X_n^o]\right)'$.

*Case 1. $l\leq L_1$:*

Let $\pi_n(\theta_0)=(\phi(b_{1,n}(\theta_0))/\Phi(b_{1,n}(\theta_0)),\ldots,\phi(b_{2n,n}(\theta_0))/\Phi(b_{2n,n}(\theta_0)))'$ and $\gamma_n(\theta_0)=\left(\Omega_{11,n}^{so}(\theta_0)/\Omega_{11,n}^{ss}(\theta_0),\ldots,\right.$ $\left.\Omega_{2n2n,n}^{so}(\theta_0)/\ \Omega_{2n2n,n}^{ss}(\theta_0)\right)'$. Then

$$
\frac{\partial E[y^{*o}|y^s=1,X_n^s,X_n^o]}{\partial X_{\cdot l,n}^{o'}}=S_n^o(\lambda_0^o)\beta_{0l}^o-\mathrm{Diag}(\gamma_n(\theta_0))\left(\mathrm{Diag}(b_n(\theta_0))\mathrm{Diag}(\pi_n(\theta_0))+\mathrm{Diag}(\pi_n(\theta_0))^2\right)S_n^s(\lambda_0^s)\beta_{0l}^s.
$$

*Case 2. $l>L_1$:*

Now the exogenous variable is present only in the outcome equation, thus the formula simplifies:

$$
\frac{\partial E[y^{*o}|y^s=1,X_n^s,X_n^o]}{\partial X_{\cdot l,n}^{o'}}=S_n^o(\lambda_0^o)\beta_{0l}^o.
$$

The total, direct, and indirect marginal effects for both cases are obtained analogously to $\partial P[y^s=1|X_n^s]/\partial X_{\cdot l,n}^{s'}$.

**Spatial error model.** In the spatial error case, the indirect marginal effects are equal to zero. It is thus enough to consider the marginal effects with respect to "own" exogenous variables:

$$\frac{\partial P[y_{i,n}^s = 1|X_{i\cdot,n}^s]}{\partial X_{il,n}^s} = \phi(b_{i,n}^e(\theta_0))\frac{\beta_{0l}^s}{\sqrt{\Omega_{ii,n}^{ss}(\theta_0)}}$$

for $i = 1, \ldots, 2n$, where $b_{i,n}^e(\theta_0) = X_{i\cdot,n}^s \beta_0^s / \sqrt{\Omega_{ii,n}^{ss}(\theta_0)}$. As before, the marginal effect of $E[y_{i,n}^{*o}|y_{i,n}^s = 1, X_{i\cdot,n}^s, X_{i\cdot,n}^o]$ with respect to $X_{il,n}^o$ depends on whether the explanatory variable is in both equations or only in one.

*Case 1.* $l \leq L_1$:

$$\frac{\partial E[y_{i,n}^{*o}|y_{i,n}^s = 1, X_{i\cdot,n}^s, X_{i\cdot,n}^o]}{\partial X_{il,n}^o} = \beta_{0l}^o - \gamma_{i,n}(\theta_0)\left(b_{i,n}^e(\theta_0)\pi_{i,n}^e(\theta_0) + \pi_{i,n}^{e^2}(\theta_0)\right)\beta_{0l}^s$$

with $\pi_{i,n}^e(\theta_0) = \phi(b_{i,n}^e(\theta_0))/\Phi(b_{i,n}^e(\theta_0))$.

*Case 2.* $l > L_1$:

$$\frac{\partial E[y_{i,n}^{*o}|y_{i,n}^s = 1, X_{i\cdot,n}^s, X_{i\cdot,n}^o]}{\partial X_{il,n}^o} = \beta_{0l}^o.$$

# Appendix H Example of sufficient conditions for Assumptions 1($ii$), 5, and 10

Let us provide an example of sufficient conditions for Assumptions 1($ii$), 5, and 10 based on the distances between the locations $l_i$ and $l_j$ of observations $i$ and $j$: $\mathtt{d}(i,j) = \|l_i - l_j\|$. Note that this distance between two observations is closely connected to the distance between two pairs of observations $g = \{g_1, g_2\}$ and $\dot{g} = \{\dot{g}_1, \dot{g}_2\}$ since $d(g,\dot{g}) = \min\{\|(l_{g_1}', l_{g_2}')' - (l_{\dot{g}_1}', l_{\dot{g}_2}')'\|, \|(l_{g_1}', l_{g_2}')' - (l_{\dot{g}_2}', l_{\dot{g}_1}')'\|\}$ and thus $d^2(g,\dot{g}) = \min\{\mathtt{d}^2(g_1,\dot{g}_1) + \mathtt{d}^2(g_2,\dot{g}_2), \mathtt{d}^2(g_1,\dot{g}_2) + \mathtt{d}^2(g_2,\dot{g}_1)\}$.

As in Qu and Lee (2015), let us now assume that the spatial weight matrix $W_n^b$, $b \in \{s, o\}$, satisfies Assumption 3 and the following constraint: $0 \leq W_{ij,n}^b \leq C_c \mathtt{d}(i,j)^{-C_p \tilde{d}}$ for some $C_c \geq 0$, $C_p > 2$, and $i \neq j$. Moreover, the number of columns with their column sums exceeding $C_W^b = \sup_n \|W_n^b\|_\infty < \infty$ is assumed to be bounded. Furthermore, let $\sup_\lambda |\lambda| C_W^b < 1$. This assumption guarantees that $I_{2n} - \lambda W_n^b$ is invertible for all values of $\lambda$ satisfying $|\lambda| < 1/C_W^b$. This is not a restrictive assumption because in empirical applications the weight matrices are typically normalized in such a way that the sum of each row is equal to 1 – implying that $C_W^b = 1$ – and the parameter space is chosen to be $(-1, 1)$. Finally, it has to hold that $d(g,\dot{g}) > s \Rightarrow \min\{\mathtt{d}(g_1,\dot{g}_1), \mathtt{d}(g_1,\dot{g}_2), \mathtt{d}(g_2,\dot{g}_1), \mathtt{d}(g_2,\dot{g}_2)\} > C_d s$ for any

$s > s_0$, any two pairs $g \in \mathcal{G}_n$ and $\dot{g} \in \mathcal{G}_n$, and some $C_d > 0$ and $s_0 > 0$. For example, this condition holds if $\mathtt{d}(g_1, g_2) < \bar{g} < \infty$ for each pair $g = \{g_1, g_2\} \in \mathcal{G}_n$. Given that the pairs are chosen by the user, this is not a restrictive assumption.

We will show that these assumptions imply Assumptions 1($ii$), 5, and 10. We have already assumed above that the row sum matrix norm of $W_n^b$ is bounded. Claim C.1.1 of Qu and Lee (2015) implies that the column sum matrix norm is also bounded. Under the above stated assumptions, Qu and Lee (2015) proved in their Lemma C.1.6, equation (C.1), that for any $m, n, s \in \mathbb{N}$ and $i = 1, \ldots, 2n$,

$$\sum_{j:\mathtt{d}(i,j)>s} ([W_n^b]^m)_{ij} \leq C_1 (C_W^b)^m m^{C_p \tilde{d}+2} s^{(1-C_p)\tilde{d}}, \tag{H.1}$$

where $C_1 > 0$.

Using this result, let us now consider all $|\lambda| \leq C_\lambda^b$ for some $0 < C_\lambda^b < 1/C_W^b$ and analyze $S_n^b(\lambda) = (I_{2n} - \lambda W_n^b)^{-1}$. Since $S_n^b(\lambda) = \sum_{l=0}^{\infty} \lambda^l [W_n^b]^l$, it follows that $\|S_n^b(\lambda)\|_\infty \leq \sum_{l=0}^{\infty} (C_\lambda^b C_W^b)^l < \infty$ uniformly in $n$ and $\lambda$ and thus $\sup_{n,i} \sup_{|\lambda| \leq C_\lambda^b} |S_{ij,n}^b(\lambda)| < \infty$. Hence, Assumption 1($ii$) holds. Similarly for any $i = 1, \ldots, 2n$ and $n, s \in \mathbb{N}$, it follows from (H.1) that

$$\sum_{j:\mathtt{d}(i,j)>s} |S_{ij,n}^b(\lambda)| = \sum_{l=0}^{\infty} \sum_{j:\mathtt{d}(i,j)>s} |\lambda|^l ([W_n^b]^l)_{ij} \leq \sum_{l=0}^{\infty} |\lambda|^l C_1 (C_W^b)^l l^{C_p \tilde{d}+2} s^{(1-C_p)\tilde{d}} \leq \sum_{l=0}^{\infty} C_1 (C_\lambda^b C_W^b)^l l^{C_p \tilde{d}+2} s^{(1-C_p)\tilde{d}},$$

and given the independence of the right-hand side on $n$ and $\lambda$, that

$$\sup_{n,i} \sup_{|\lambda| \leq C_\lambda^b} \sum_{j:\mathtt{d}(i,j)>s} |S_{ij,n}^b(\lambda)| \leq C_0 s^{(1-C_p)\tilde{d}}$$

for some $C_0 > 0$.

Proceeding now to Assumption 5, the equivalence of the matrix norms – $\|S_{g\dot{g},n}\| \leq 2\|S_{g\dot{g},n}\|_\infty$ – implies that

$$1/\sqrt{2} < \sup_{n,g} \sup_{|\lambda| \leq C_\lambda^b} \sum_{\dot{g} \in \mathcal{G}_n} \|S_{g\dot{g},n}^b(\lambda)\| < \sup_{n,g} \sup_{|\lambda| < C_\lambda^b} \sum_{\dot{g} \in \mathcal{G}_n} 2\|S_{g\dot{g},n}^b(\lambda)\|_\infty < \infty,$$

where the lower bound is obtained for $\lambda = 0$, and for all $|\lambda| \leq C_\lambda^b$ and $s > s_0$, that

$$\sup_{n,g} \sum_{\dot{g}:d(g,\dot{g})>s} \|S_{g\dot{g},n}^b(\lambda)\| \leq 2 \left\{ \sup_{n,g} \sum_{j:\mathtt{d}(g_1,j)>C_d s} |S_{g_1 j,n}^b(\lambda)| + \sup_{n,g} \sum_{j:\mathtt{d}(g_2,j)>C_d s} |S_{g_2 j,n}^b(\lambda)| \right\} \leq 4C_0 (C_d s)^{(1-C_p)\tilde{d}}.$$

Hence, Assumption 5 is satisfied since $C_p > 2$.

Finally, let us note that the above result implies that $\psi(s) = \max\{\psi^s(s), \psi^o(s)\} \leq C_\psi (C_d s)^{(1-C_p)\tilde{d}}$. Assumption 10 requires that $\sum_{s=1}^{\infty} s^{2\tilde{d}-1}[\psi(s)]^{C_r} < \infty$, where $C_r = (r-2)/(12r-12)$ for some $r > 2$. This sum can be bounded by

$$\sum_{s=1}^{\infty} s^{2\tilde{d}-1}[\psi(s)]^{C_r} \leq \sum_{s=1}^{\infty} s^{2\tilde{d}-1}[C_\psi (C_d s)^{(1-C_p)\tilde{d}}]^{C_r},$$

which is finite if $C_p > 1 + 2/C_r$.

## Appendix I Some Additional Graphs

In Tables 1 and 2 (Appendix A), the bias, standard deviation, and RMSE of $\hat{\beta}_n^s$ obtained by both HMLE and PMLE when $\lambda^s = \lambda^o = 0.85$ and $2n = 158$ are very high. Figures 1 and 2 show that these results are mainly driven by one Monte Carlo iteration: Figure 1 reports the estimates of $\beta^s$ obtained in all the iterations, whereas Figure 2 shows exactly the same estimates but after excluding the Monte Carlo iteration with the most prominent estimates in Figure 1.

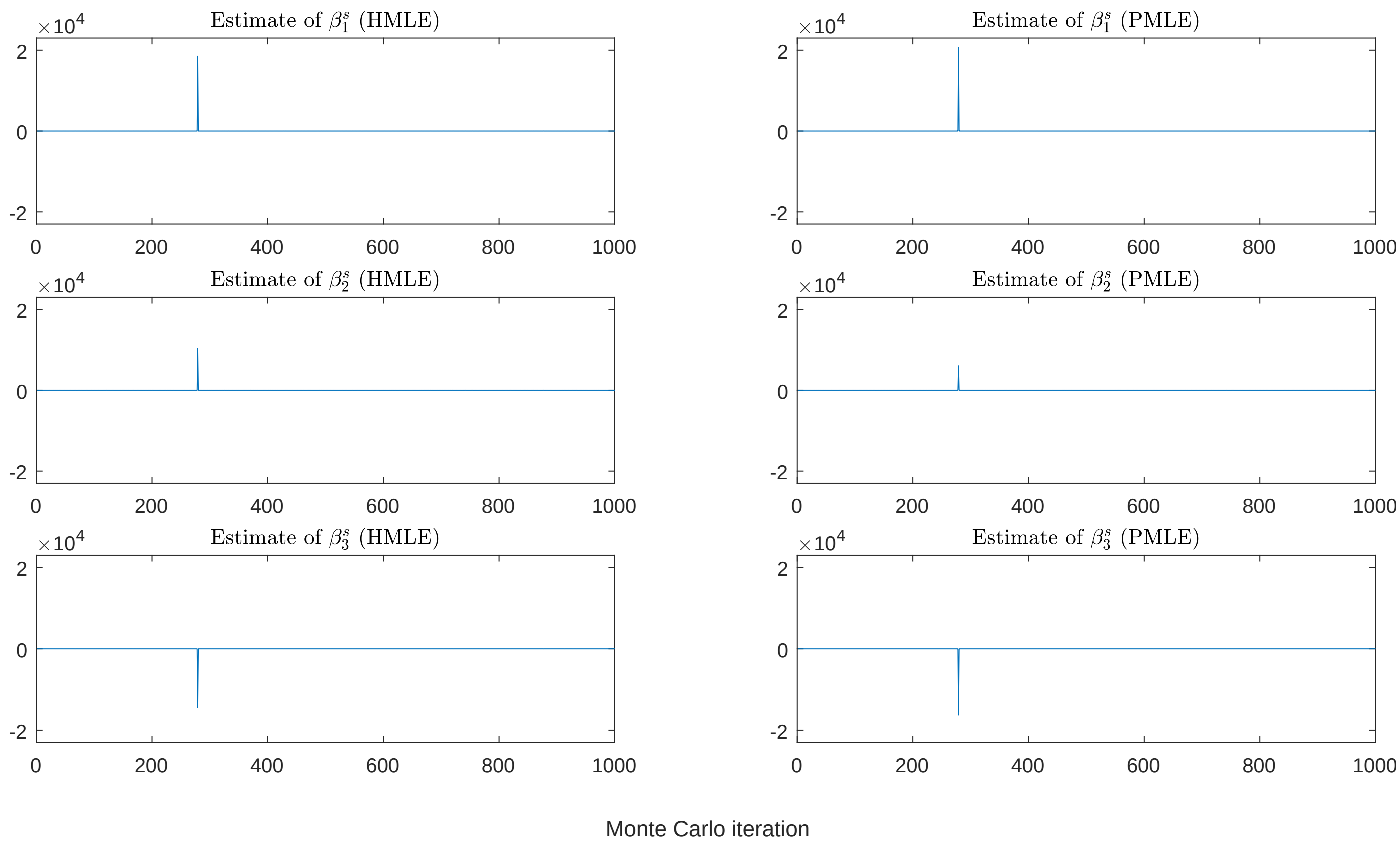


**Figure 1:** The estimates of $\beta^s$ obtained in all the Monte Carlo iterations used to construct the bias, standard deviation, and RMSE in Tables 1 and 2 in Appendix A when $\lambda^s = \lambda^o = 0.85$ and $2n = 158$.

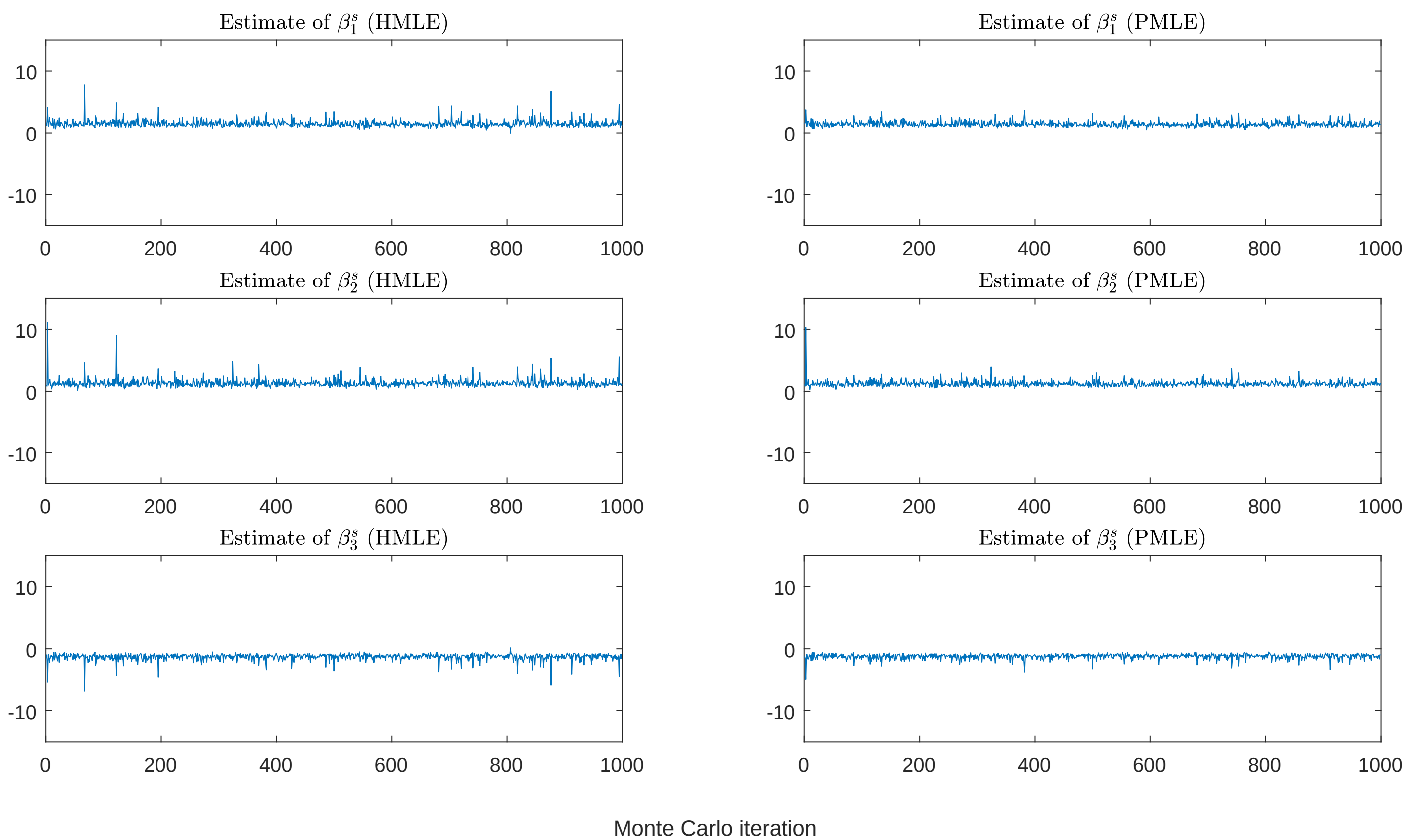


**Figure 2:** The estimates of $\beta^s$ obtained in all the Monte Carlo iterations used to construct the bias, standard deviation, and RMSE in Tables 1 and 2 in Appendix A when $\lambda^s = \lambda^o = 0.85$ and $2n = 158$ excluding the most prominent iteration in Figure 1.

# Appendix J Some Additional Technical Lemmas

**Lemma J.1.** *Let $X \sim \mathcal{N}(0,1)$. Then for any given $r \in \mathbb{N}$, there is some constant $C_1 > 0$ such that for any $c \in \mathbb{R}$, $E[|X|^r | X \leq c] \leq |c|^{r-1}\phi(c)/\Phi(c) + (r-1)E[|X|^{r-2}|X \leq c]$ for $r \geq 2$ with $E[|X||X \leq c] \leq \phi(c)/\Phi(c) + C_1$ and $E[|X|^0 | X \leq c] = 1$.*

*Proof.*

**Case 1.** $c \leq 0$:

Consider $r = 0$. Then $E[|X|^0|X \leq c] = \int_{-\infty}^{c} \phi(x)/\Phi(c)\mathrm{d}x = \Phi(c)/\Phi(c) = 1$. If $r = 1$, then $E[|X||X \leq c] = \int_{-\infty}^{c}(-x)\phi(x)/\Phi(c)\mathrm{d}x = \int_{-\infty}^{c} \phi'(x)/\Phi(c)\mathrm{d}x = \phi(c)/\Phi(c)$, where the second equality follows by observing that $\phi'(x) = -x\phi(x)$. If $r \geq 2$, then by integration by parts,

$$
\begin{aligned}
E[|X|^r|X \leq c] &= \int_{-\infty}^{c} |x|^r \frac{\phi(x)}{\Phi(c)}\mathrm{d}x = \int_{-\infty}^{c} (-x)^r \frac{\phi(x)}{\Phi(c)}\mathrm{d}x = (-1)^{r-1} \int_{-\infty}^{c} x^{r-1} \frac{\phi'(x)}{\Phi(c)}\mathrm{d}x \\
&= (-1)^{r-1} \left( c^{r-1} \frac{\phi(c)}{\Phi(c)} - (r-1) \int_{-\infty}^{c} x^{r-2} \frac{\phi(x)}{\Phi(c)}\mathrm{d}x \right) \\
&= |c|^{r-1} \frac{\phi(c)}{\Phi(c)} + (r-1) \int_{-\infty}^{c} (-x)^{r-2} \frac{\phi(x)}{\Phi(c)}\mathrm{d}x = |c|^{r-1} \frac{\phi(c)}{\Phi(c)} + (r-1)E[|X|^{r-2}|X \leq c].
\end{aligned}
$$

**Case 2.** $c > 0$:

As in Case 1, $E[|X|^0|X \leq c] = 1$. If $r = 1$, then

$$
\begin{aligned}
E[|X||X \leq c] &= \int_{-\infty}^{c} |x| \frac{\phi(x)}{\Phi(c)}\mathrm{d}x = \int_{-\infty}^{0} (-x) \frac{\phi(x)}{\Phi(c)}\mathrm{d}x + \int_{0}^{c} x \frac{\phi(x)}{\Phi(c)}\mathrm{d}x = \int_{-\infty}^{0} \frac{\phi'(x)}{\Phi(c)}\mathrm{d}x - \int_{0}^{c} \frac{\phi'(x)}{\Phi(c)}\mathrm{d}x \\
&= -\frac{\phi(c)}{\Phi(c)} + \frac{2\phi(0)}{\Phi(c)} \leq \frac{\phi(c)}{\Phi(c)} + \frac{2\phi(0)}{\Phi(0)} \leq \frac{\phi(c)}{\Phi(c)} + C_1.
\end{aligned}
$$

Consider $r \geq 2$. Then by integration by parts, it holds that

$$
\begin{aligned}
E[|X|^r|X \leq c] &= \int_{-\infty}^{c} |x|^r \frac{\phi(x)}{\Phi(c)}\mathrm{d}x = \int_{-\infty}^{0} (-x)^r \frac{\phi(x)}{\Phi(c)}\mathrm{d}x + \int_{0}^{c} x^r \frac{\phi(x)}{\Phi(c)}\mathrm{d}x \\
&= (-1)^{r-1} \int_{-\infty}^{0} x^{r-1} \frac{\phi'(x)}{\Phi(c)}\mathrm{d}x - \int_{0}^{c} x^{r-1} \frac{\phi'(x)}{\Phi(c)}\mathrm{d}x \\
&= (-1)^{r-2}(r-1) \int_{-\infty}^{0} x^{r-2} \frac{\phi(x)}{\Phi(c)}\mathrm{d}x - c^{r-1} \frac{\phi(c)}{\Phi(c)} + (r-1) \int_{0}^{c} x^{r-2} \frac{\phi(x)}{\Phi(c)}\mathrm{d}x \\
&= -c^{r-1} \frac{\phi(c)}{\Phi(c)} + (r-1) \int_{-\infty}^{c} |x|^{r-2} \frac{\phi(x)}{\Phi(c)}\mathrm{d}x \leq |c|^{r-1} \frac{\phi(c)}{\Phi(c)} + (r-1)E[|X|^{r-2}|X \leq c].
\end{aligned}
$$

□

**Lemma J.2.** *Let $Y_{g,n} = (Y_{gk,n})_{k=1}^K$ be a $K$-dimensional random vector. Then for some $p \geq 1$, $\{Y_{g,n}\}_{g\in\mathcal{G}_n}$ is a uniform $L_p$-NED random field with NED coefficients $\psi(s)$ if and only if for each $k = 1,\ldots,K$, $\{Y_{gk,n}\}_{g\in\mathcal{G}_n}$ is a uniform $L_p$-NED random field with NED coefficients $\psi(s)$.*

*Proof.* We start with the 'if' part. By Loève's $c_r$-inequality, it follows that

$$\begin{aligned}
\|Y_{g,n} - E[Y_{g,n}|\mathcal{F}_{g,n}(s)]\|_p &= (E\|Y_{g,n} - E[Y_{g,n}|\mathcal{F}_{g,n}(s)]\|^p)^{1/p} \\
&= \left(E\left[\sum_{k=1}^K |Y_{gk,n} - E[Y_{gk,n}|\mathcal{F}_{g,n}(s)]|^2\right]^{p/2}\right)^{1/p} \\
&\leq C_1\left(\sum_{k=1}^K E|Y_{gk,n} - E[Y_{gk,n}|\mathcal{F}_{g,n}(s)]|^p\right)^{1/p} \\
&\leq C_1\sum_{k=1}^K \|Y_{gk,n} - E[Y_{gk,n}|\mathcal{F}_{g,n}(s)]\|_p \\
&\leq \psi(s)C_1\sum_{k=1}^K t_{g,n}^k,
\end{aligned}$$

where $\{t_{g,n}^k\}_{g\in\mathcal{G}_n}$ is the NED scaling factor for an element $k$ with $\sup_{n,g}\sum_{k=1}^K t_{g,n}^k < \infty$ because for each $k = 1,\ldots,K$, $\{Y_{gk,n}\}_{g\in\mathcal{G}_n}$ is a uniform random field.

We continue with the 'only if' part:

$$\|Y_{gk,n} - E[Y_{gk,n}|\mathcal{F}_{g,n}(s)]\|_p \leq \|Y_{g,n} - E[Y_{g,n}|\mathcal{F}_{g,n}(s)]\|_p \leq t_{g,n}\psi(s),$$

where $\{t_{g,n}\}_{g\in\mathcal{G}_n}$ is the NED scaling factor for random field $\{Y_{g,n}\}_{g\in\mathcal{G}_n}$ with $\sup_{n,g} t_{g,n} < \infty$ because $\{Y_{g,n}\}_{g\in\mathcal{G}_n}$ is a uniform random field. $\square$

# Appendix K Proof of Technical Lemmas in Appendix C

***Proof of Lemma C.4.*** Noting that $\partial \ln \Phi_2(v,R)/\partial v = \Phi_2(v,R)^{-1}\partial\Phi_2(v,R)/\partial v$ and $\partial \ln \Phi_2(v,R)/\partial\theta = \Phi_2(v,R)^{-1}$ $\partial\Phi_2(v,R)/\partial\theta$, we apply differentiation under the integral sign twice to compute $\partial\Phi_2(v,R)/\partial v$ and $\partial\Phi_2(v,R)/\partial\theta$:

$$\begin{aligned}
\frac{\partial\Phi_2(v,R)}{\partial v} &= \int_{-\infty}^{v_2}\phi_2((v_1,z_2)',R)\mathrm{d}z_2\frac{\partial v_1}{\partial v} + \int_{-\infty}^{v_1}\phi_2((z_1,v_2)',R)\mathrm{d}z_1\frac{\partial v_2}{\partial v} \\
&= \left(\int_{-\infty}^{v_2}\phi_2((v_1,z_2)',R)\mathrm{d}z_2, \int_{-\infty}^{v_1}\phi_2((z_1,v_2)',R)\mathrm{d}z_1\right)'
\end{aligned} \tag{K.1}$$

and

$$\frac{\partial \Phi_2(v,R)}{\partial \theta} = \int_{-\infty}^{v_2} \phi_2((v_1,z_2)',R)\mathrm{d}z_2 \frac{\partial v_1}{\partial \theta} + \int_{-\infty}^{v_1} \phi_2((z_1,v_2)',R)\mathrm{d}z_1 \frac{\partial v_2}{\partial \theta} + \int_{-\infty}^{v_1}\int_{-\infty}^{v_2} \frac{\partial \phi_2(z,R)}{\partial \theta}\mathrm{d}z_2\mathrm{d}z_1. \tag{K.2}$$

Note that

$$\phi_2((v_1,z_2)',R) = \frac{1}{\sqrt{1-\rho^2}}\phi(v_1)\phi\left(\frac{z_2-\rho v_1}{\sqrt{1-\rho^2}}\right).$$

Thus,

$$\int_{-\infty}^{v_2} \phi_2((v_1,z_2)',R)\mathrm{d}z_2 = \phi(v_1)\Phi\left(\frac{v_2-\rho v_1}{\sqrt{1-\rho^2}}\right). \tag{K.3}$$

Similarly,

$$\int_{-\infty}^{v_1} \phi_2((z_1,v_2)',R)\mathrm{d}z_1 = \phi(v_2)\Phi\left(\frac{v_1-\rho v_2}{\sqrt{1-\rho^2}}\right). \tag{K.4}$$

The first claim now follows from (K.1), (K.3), and (K.4). Further, it is easy to show that

$$\frac{\partial \phi_2(z,R)}{\partial \theta} = -\frac{1}{2}\phi_2(z,R)\left(\frac{\partial \ln|R|}{\partial \theta} + \frac{\partial P(z,R)}{\partial \theta}\right). \tag{K.5}$$

Hence,

$$\int_{-\infty}^{v_1}\int_{-\infty}^{v_2} \frac{\partial \phi_2(z,R)}{\partial \theta}\mathrm{d}z_2\mathrm{d}z_1 = -\frac{1}{2}\Phi_2(v,R)\left(\frac{\partial \ln|R|}{\partial \theta} + E_V\left[\frac{\partial P(V,R)}{\partial \theta}\Big| V \le v\right]\right), \tag{K.6}$$

where $V \sim \mathcal{N}(0,R)$. The second conclusion follows by combining (K.2) with (K.3), (K.4), and (K.6). For the last claim of the lemma, note that

$$\frac{\partial^2 \ln \Phi_2(v,R)}{\partial \theta \partial \theta'} = \frac{1}{\Phi_2(v,R)}\frac{\partial^2 \Phi_2(v,R)}{\partial \theta \partial \theta'} - \frac{1}{\Phi_2^2(v,R)}\frac{\partial \Phi_2(v,R)}{\partial \theta}\frac{\partial \Phi_2(v,R)}{\partial \theta'}. \tag{K.7}$$

Based on the results (K.2)–(K.4) and the definition of $\xi(v,R)$, it follows

$$\begin{aligned}\frac{1}{\Phi_2(v,R)}\frac{\partial^2 \Phi_2(v,R)}{\partial \theta \partial \theta'} &= \frac{1}{\Phi_2(v,R)}\frac{\partial v_1}{\partial \theta}\frac{\partial \int_{-\infty}^{v_2}\phi_2((v_1,z_2)',R)\mathrm{d}z_2}{\partial \theta'} + \xi_1(v,R)\frac{\partial^2 v_1}{\partial \theta \partial \theta'} \\ &+ \frac{1}{\Phi_2(v,R)}\frac{\partial v_2}{\partial \theta}\frac{\partial \int_{-\infty}^{v_1}\phi_2((z_1,v_2)',R)\mathrm{d}z_1}{\partial \theta'} + \xi_2(v,R)\frac{\partial^2 v_2}{\partial \theta \partial \theta'} \\ &+ \frac{1}{\Phi_2(v,R)}\frac{\partial \int_{-\infty}^{v_1}\int_{-\infty}^{v_2}\frac{\partial \phi_2(z,R)}{\partial \theta}\mathrm{d}z_2\mathrm{d}z_1}{\partial \theta'}.\end{aligned} \tag{K.8}$$

By applying differentiation under the integral sign, it follows as in (K.5)–(K.6) that

$$\begin{aligned}\frac{\partial \int_{-\infty}^{v_2} \phi_2((v_1, z_2)', R)\mathrm{d}z_2}{\partial \theta'} &= \phi_2(v, R)\frac{\partial v_2}{\partial \theta'} + \int_{-\infty}^{v_2} \frac{\partial \phi_2((v_1, z_2)', R)}{\partial \theta'}\mathrm{d}z_2 \\ &= \phi_2(v, R)\frac{\partial v_2}{\partial \theta'} - \frac{1}{2}\phi(v_1)\Phi\left(\frac{v_2 - \rho v_1}{\sqrt{1-\rho^2}}\right)\left(\frac{\partial \ln|R|}{\partial \theta'} + E_{\widetilde{V}_2}\left[\frac{\partial P((v_1, \widetilde{V}_2)', R)}{\partial \theta'}\Big|\widetilde{V}_2 \leq v_2\right]\right).\end{aligned} \tag{K.9}$$

Then by definition of $\kappa(v, R)$, $\xi(v, R)$, and $A(v, R)$,

$$\frac{1}{\Phi_2(v, R)}\frac{\partial v_1}{\partial \theta}\frac{\partial \int_{-\infty}^{v_2} \phi_2((v_1, z_2)', R)\mathrm{d}z_2}{\partial \theta'} = \kappa(v, R)\frac{\partial v_1}{\partial \theta}\frac{\partial v_2}{\partial \theta'} - \frac{1}{2}A(v, R). \tag{K.10}$$

Symmetrically,

$$\frac{1}{\Phi_2(v, R)}\frac{\partial v_2}{\partial \theta}\frac{\partial \int_{-\infty}^{v_1} \phi_2((z_1, v_2)', R)\mathrm{d}z_1}{\partial \theta'} = \kappa(v, R)\frac{\partial v_2}{\partial \theta}\frac{\partial v_1}{\partial \theta'} - \frac{1}{2}B(v, R). \tag{K.11}$$

We proceed with the last term in (K.8). By applying differentiation under the integral sign twice,

$$\begin{aligned}&\frac{1}{\Phi_2(v, R)}\frac{\partial \int_{-\infty}^{v_1}\int_{-\infty}^{v_2} \frac{\partial \phi_2(z,R)}{\partial \theta}\mathrm{d}z_2\mathrm{d}z_1}{\partial \theta'} \\ &= \frac{1}{\Phi_2(v, R)}\left(\int_{-\infty}^{v_2} \frac{\partial \phi_2((v_1, z_2)', R)}{\partial \theta}\mathrm{d}z_2\frac{\partial v_1}{\partial \theta'} + \int_{-\infty}^{v_1} \frac{\partial \phi_2((z_1, v_2)', R)}{\partial \theta}\mathrm{d}z_1\frac{\partial v_2}{\partial \theta'} + \int_{-\infty}^{v_1}\int_{-\infty}^{v_2} \frac{\partial^2 \phi_2(z, R)}{\partial \theta \partial \theta'}\mathrm{d}z_2\mathrm{d}z_1\right) \\ &= -\frac{1}{2}\xi_1(v, R)\left(\frac{\partial \ln|R|}{\partial \theta} + E_{\widetilde{V}_2}\left[\frac{\partial P((v_1, \widetilde{V}_2)', R)}{\partial \theta}\Big|\widetilde{V}_2 \leq v_2\right]\right)\frac{\partial v_1}{\partial \theta'} \\ &\quad - \frac{1}{2}\xi_2(v, R)\left(\frac{\partial \ln|R|}{\partial \theta} + E_{\widetilde{V}_1}\left[\frac{\partial P((\widetilde{V}_1, v_2)', R)}{\partial \theta}\Big|\widetilde{V}_1 \leq v_1\right]\right)\frac{\partial v_2}{\partial \theta'} + \frac{1}{\Phi_2(v, R)}\int_{-\infty}^{v_1}\int_{-\infty}^{v_2} \frac{\partial^2 \phi_2(z, R)}{\partial \theta \partial \theta'}\mathrm{d}z_2\mathrm{d}z_1 \\ &= -\frac{1}{2}A'(v, R) - \frac{1}{2}B'(v, R) + \frac{1}{\Phi_2(v, R)}\int_{-\infty}^{v_1}\int_{-\infty}^{v_2} \frac{\partial^2 \phi_2(z, R)}{\partial \theta \partial \theta'}\mathrm{d}z_2\mathrm{d}z_1,\end{aligned} \tag{K.12}$$

where the second equality follows in the same way as in (K.9). It can easily be shown that

$$\frac{\partial^2 \phi_2(z, R)}{\partial \theta \partial \theta'} = -\frac{1}{2}\phi_2(z, R)G(z, R),$$

where $G$ is defined (C.3). Thus,

$$\frac{1}{\Phi_2(v, R)}\int_{-\infty}^{v_1}\int_{-\infty}^{v_2} \frac{\partial^2 \phi_2(z, R)}{\partial \theta \partial \theta'}\mathrm{d}z_2\mathrm{d}z_1 = -\frac{1}{2}E_V[G(V, R)|V \leq v]. \tag{K.13}$$

The conclusion follows by combining (K.7) with (K.8), (K.10), (K.11), (K.12), and (K.13). □

***Proof of Lemma C.5***. We will start with the first claim by deriving the bounds when $(v_1, v_2) \in (-1, +\infty) \times$

$(-1,+\infty)$ and $(v_1,v_2)\notin(-1,+\infty)\times(-1,+\infty)$ and afterwards we will combine the results.

**Case 1.** $(v_1,v_2)\in(-1,+\infty)\times(-1,+\infty)$:

$$\frac{\phi(v_1)\Phi\left(\frac{v_2-\rho v_1}{\sqrt{1-\rho^2}}\right)}{\Phi_2(v,R)}\le\frac{1}{\sqrt{2\pi}\Phi_2(-\iota_2,R)},$$

where $\iota_2$ is a 2-dimensional vector of ones. Thus, we need to derive the lower bound for $\Phi_2(-\iota_2,R)$. Since $R$ is a symmetric matrix, there exists an orthogonal matrix $O$ such that $R=O\mathrm{Diag}\{\tau_1,\tau_2\}O'$, where $\tau_1\le\tau_2$ are the eigenvalues of $R$. Thus, $R^{-1}=O\mathrm{Diag}\{\tau_1^{-1},\tau_2^{-1}\}O'$. From Exercise 12.39 of Abadir and Magnus (2005), it holds for any symmetric matrix $A$ that $z'Az\le\mathrm{maxeig}(A)z'z$. Hence,

$$z'R^{-1}z=z'O\mathrm{Diag}\{\tau_1^{-1},\tau_2^{-1}\}O'z\le\frac{1}{\tau_1}z'OO'z=\frac{1}{\tau_1}z'z=\left(\frac{z_1}{\sqrt{\tau_1}}\right)^2+\left(\frac{z_2}{\sqrt{\tau_1}}\right)^2 \tag{K.14}$$

and

$$\begin{aligned}
\Phi_2(-\iota_2,R)&=\int_{-\infty}^{-1}\int_{-\infty}^{-1}\frac{1}{2\pi|R|^{1/2}}\exp\left(-\frac{1}{2}z'R^{-1}z\right)\mathrm{d}z_2\mathrm{d}z_1\\
&=\int_{-\infty}^{-1}\int_{-\infty}^{-1}\frac{1}{2\pi\sqrt{\tau_1\tau_2}}\exp\left(-\frac{1}{2}z'R^{-1}z\right)\mathrm{d}z_2\mathrm{d}z_1\\
&\ge\int_{-\infty}^{-1}\int_{-\infty}^{-1}\frac{1}{2\pi\sqrt{\tau_1\tau_2}}\exp\left(-\frac{1}{2}\left(\left(\frac{z_1}{\sqrt{\tau_1}}\right)^2+\left(\frac{z_2}{\sqrt{\tau_1}}\right)^2\right)\right)\mathrm{d}z_2\mathrm{d}z_1\\
&=\sqrt{\frac{\tau_1}{\tau_2}}\int_{-\infty}^{-1}\int_{-\infty}^{-1}\frac{1}{\tau_1}\phi\left(\frac{z_1}{\sqrt{\tau_1}}\right)\phi\left(\frac{z_2}{\sqrt{\tau_1}}\right)\mathrm{d}z_2\mathrm{d}z_1 \qquad\text{(K.15)}\\
&=\sqrt{\frac{\tau_1}{\tau_2}}\Phi^2\left(\frac{-1}{\sqrt{\tau_1}}\right)\\
&\ge\sqrt{\frac{1-|\rho|}{2}}\Phi^2\left(\frac{-1}{\sqrt{1-|\rho|}}\right)\\
&=\left(\frac{1-|\rho|}{2}\right)^{1/2}\left(1-\Phi\left((1-|\rho|)^{-1/2}\right)\right)^2,
\end{aligned}$$

where the last inequality follows by noticing that $\tau_1=\min\{1-\rho,1+\rho\}=1-|\rho|$ and $\tau_2=\max\{1-\rho,1+\rho\}<2$. Hence,

$$\frac{\phi(v_1)\Phi\left(\frac{v_2-\rho v_1}{\sqrt{1-\rho^2}}\right)}{\Phi_2(v,R)}\le C_1(1-|\rho|)^{-1/2}\left(1-\Phi\left((1-|\rho|)^{-1/2}\right)\right)^{-2}\le C_1(1-|\rho|)^{-7}\left(1-\Phi\left((1-|\rho|)^{-1/2}\right)\right)^{-2}. \tag{K.16}$$

**Case 2.** $(v_1,v_2)\notin(-1,+\infty)\times(-1,+\infty)$:

First of all, we will derive the bound for $\phi(v_1)\Phi\left(\frac{v_2-\rho v_1}{\sqrt{1-\rho^2}}\right)$; afterwards we will derive the bound for the entire expression. Let $\alpha = z^{\star\prime} R^{-1} z^\star$ with $z^* = (z_1^*, z_2^*)' = \arg\min_z z' R^{-1} z$, such that $z \le v$. Then

$$
\begin{aligned}
\phi(v_1)\Phi\left(\frac{v_2 - \rho v_1}{\sqrt{1-\rho^2}}\right) &= \int\limits_{-\infty}^{v_2} \phi_2((v_1, z_2)', R)\mathrm{d}z_2 \\
&= \int_{-\infty}^{v_2}\int_{-\infty}^{v_1} \phi_2(z, R) \cdot \frac{\rho z_2 - z_1}{1-\rho^2}\mathrm{d}z_1\mathrm{d}z_2 \\
&= \int_{-\infty}^{v_2}\int_{-\infty}^{v_1} \frac{\exp(-z' R^{-1} z/2)}{2\pi(1-\rho^2)^{1/2}} \cdot \frac{\rho z_2 - z_1}{1-\rho^2}\mathrm{d}z_1\mathrm{d}z_2 \\
&\le \frac{1}{2\pi(1-\rho^2)^{3/2}} \int_{-\infty}^{v_2}\int_{-\infty}^{v_1} \exp\left(-\frac{1}{2} z' R^{-1} z\right)(|z_1| + |z_2|)\mathrm{d}z_1\mathrm{d}z_2 \\
&\le \frac{1}{2\pi(1-\rho^2)^{3/2}} \int_{-\infty}^{v_2}\int_{-\infty}^{v_1} \exp\left(-\frac{1}{2}\alpha\right)\frac{\max\{6,\alpha\}^3}{(z' R^{-1} z)^3}(|z_1| + |z_2|)\mathrm{d}z_1\mathrm{d}z_2 \\
&\le \frac{\exp(-\alpha/2)\max\{6,\alpha\}^3}{2\pi(1-|\rho|)^{3/2}} \int_{-\infty}^{v_2}\int_{-\infty}^{v_1} \frac{|z_1| + |z_2|}{(z' R^{-1} z)^3}\mathrm{d}z_1\mathrm{d}z_2,
\end{aligned}
$$

where the second inequality follows from the following observation: the derivative of $\exp(-\alpha/2)(\max\{6,\alpha\}/t)^3/\exp(-t/2)$ indicates that the minimum of this function for $t \ge \alpha$ is attained at $t = \max\{6, \alpha\}$; the minimum of this function is at least 1. The double integral will be now proved to be bounded by a constant.

*Case (i).* $v_1 \le -1$ and $v_2 \le -1$:

If $z_1 \le -1$ and $z_2 \le -1$, then $z' R^{-1} z = ((z_1 - z_2)^2 + 2(1-\rho)z_1 z_2)/(1-\rho^2) \ge 2(1-\rho)z_1 z_2/(1-\rho^2) = 2z_1 z_2/(1+\rho) > z_1 z_2 > 0$. Hence,

$$
\begin{aligned}
\int_{-\infty}^{v_2}\int_{-\infty}^{v_1} \frac{|z_1| + |z_2|}{(z' R^{-1} z)^3}\mathrm{d}z_1\mathrm{d}z_2 \le \int_{-\infty}^{v_2}\int_{-\infty}^{v_1} \frac{|z_1| + |z_2|}{(z_1 z_2)^3}\mathrm{d}z_1\mathrm{d}z_2 &= \int_{-\infty}^{v_2}\int_{-\infty}^{v_1} \left(\frac{-1}{z_1^2 z_2^3} + \frac{-1}{z_1^3 z_2^2}\right)\mathrm{d}z_1\mathrm{d}z_2 \\
&= \int_{-\infty}^{v_2} \left(\frac{1}{v_1 z_2^3} + \frac{1}{2v_1^2 z_2^2}\right)\mathrm{d}z_2 = \frac{1}{2}\left(\frac{-1}{v_1 v_2^2} + \frac{-1}{v_1^2 v_2}\right) \le 1.
\end{aligned}
$$

*Case (ii).* $-1 < v_1 \le 1$ and $v_2 \le -1$:

$$
\int_{-\infty}^{v_2}\int_{-\infty}^{v_1} \frac{|z_1| + |z_2|}{(z' R^{-1} z)^3}\mathrm{d}z_1\mathrm{d}z_2 = \int_{-\infty}^{v_2}\int_{-\infty}^{-1} \frac{|z_1| + |z_2|}{(z' R^{-1} z)^3}\mathrm{d}z_1\mathrm{d}z_2 + \int_{-\infty}^{v_2}\int_{-1}^{v_1} \frac{|z_1| + |z_2|}{(z' R^{-1} z)^3}\mathrm{d}z_1\mathrm{d}z_2 \tag{K.17}
$$

The first double integral is bounded by a constant as it is shown in Case (i). Note that if $-1 < z_1 \leq 1$ and $z_2 \leq -1$, then $z'R^{-1}z = ((z_1 - \rho z_2)^2 + (1-\rho^2)z_2^2)/(1-\rho^2) \geq z_2^2 > 0$. Thus,

$$\begin{aligned}\int_{-\infty}^{v_2}\int_{-1}^{v_1} \frac{|z_1|+|z_2|}{(z'R^{-1}z)^3}\mathrm{d}z_1\mathrm{d}z_2 &\leq \int_{-\infty}^{v_2}\int_{-1}^{v_1} \frac{1+|z_2|}{(z'R^{-1}z)^3}\mathrm{d}z_1\mathrm{d}z_2 \leq \int_{-\infty}^{v_2}\int_{-1}^{v_1} \frac{1+|z_2|}{z_2^6}\mathrm{d}z_1\mathrm{d}z_2 \\ &= (v_1+1)\int_{-\infty}^{v_2}\left(\frac{1}{z_2^6}+\frac{-1}{z_2^5}\right)\mathrm{d}z_2 \leq 2\int_{-\infty}^{v_2}\left(\frac{1}{z_2^6}+\frac{-1}{z_2^5}\right)\mathrm{d}z_2 \\ &= 2\left(\frac{-1}{5v_2^5}+\frac{1}{4v_2^4}\right) \leq 2\left(\frac{1}{5}+\frac{1}{4}\right) < 1.\end{aligned}$$

It concludes the proof that the integral in (K.17) is bounded by a constant.

*Case (iii).* $v_1 > 1$ and $v_2 \leq -1$:

$$\int_{-\infty}^{v_2}\int_{-\infty}^{v_1} \frac{|z_1|+|z_2|}{(z'R^{-1}z)^3}\mathrm{d}z_1\mathrm{d}z_2 = \int_{-\infty}^{v_2}\int_{-\infty}^{1} \frac{|z_1|+|z_2|}{(z'R^{-1}z)^3}\mathrm{d}z_1\mathrm{d}z_2 + \int_{-\infty}^{v_2}\int_{1}^{v_1} \frac{|z_1|+|z_2|}{(z'R^{-1}z)^3}\mathrm{d}z_1\mathrm{d}z_2 \tag{K.18}$$

We have already shown in Case (ii) that the first double integral is bounded by a constant. For the second integral, note that, if $z_1 > 1$ and $z_2 \leq -1$, then $z'R^{-1}z = ((z_1+z_2)^2 - 2(1+\rho)z_1z_2)/(1-\rho^2) \geq -2(1+\rho)z_1z_2/(1-\rho^2) = -2z_1z_2/(1-\rho) > -z_1z_2 > 0$. Hence,

$$\begin{aligned}\int_{-\infty}^{v_2}\int_{1}^{v_1} \frac{|z_1|+|z_2|}{(z'R^{-1}z)^3}\mathrm{d}z_1\mathrm{d}z_2 &\leq \int_{-\infty}^{v_2}\int_{1}^{v_1} \frac{|z_1|+|z_2|}{(-z_1z_2)^3}\mathrm{d}z_1\mathrm{d}z_2 = \int_{-\infty}^{v_2}\int_{1}^{v_1}\left(\frac{-1}{z_1^2z_2^3}+\frac{1}{z_1^3z_2^2}\right)\mathrm{d}z_1\mathrm{d}z_2 \\ &= \int_{-\infty}^{v_2}\frac{1}{z_2^3}\left(\frac{1}{v_1}-1\right)+\frac{-1}{2z_2^2}\left(\frac{1}{v_1^2}-1\right)\mathrm{d}z_2 = \frac{1}{2}\left(\frac{-1}{v_2^2}\left(\frac{1}{v_1}-1\right)+\frac{1}{v_2}\left(\frac{1}{v_1^2}-1\right)\right) \\ &\leq 1.\end{aligned}$$

It concludes the proof that the integral in (K.18) is bounded by a constant. Cases when $v_1 \leq -1$ and $-1 < v_2 \leq 1$ or $v_2 > 1$ can be proven analogously. Thus for some constant $C_4 > 0$,

$$\frac{\phi(v_1)\Phi\left(\frac{v_2-\rho v_1}{\sqrt{1-\rho^2}}\right)}{\Phi_2(v,R)} \leq \frac{C_4(1-|\rho|)^{-3/2}\exp(-\alpha/2)\max\{6,\alpha\}^3}{\Phi_2(v,R)}, \tag{K.19}$$

if $(v_1, v_2) \notin (-1, +\infty) \times (-1, +\infty)$. First, we will establish the bound for $\exp(-\alpha/2)/\Phi_2(v, R)$. The proof is similar to the proof of Proposition 3.2 of Hashorva and Hüsler (2003). Let $t = (t_1, t_2)' = R^{-1} z^\star$. Then

$$
\begin{aligned}
(z + z^\star)' R^{-1} (z + z^\star) &= z' R^{-1} z + 2 z' R^{-1} z^\star + {z^\star}' R^{-1} z^\star \\
&\leq \frac{1}{\tau_1} z' O O' z + 2 z' R^{-1} z^\star + {z^\star}' R^{-1} z^\star \\
&= \frac{1}{\tau_1} z' z + 2 z' R^{-1} z^\star + \alpha \\
&= \frac{1}{\tau_1} z' z + 2 z' t + \alpha \\
&= \left( \frac{z_1}{\sqrt{\tau_1}} + \sqrt{\tau_1} t_1 \right)^2 + \left( \frac{z_2}{\sqrt{\tau_1}} + \sqrt{\tau_1} t_2 \right)^2 - \left( \sqrt{\tau_1} t_1 \right)^2 - \left( \sqrt{\tau_1} t_2 \right)^2 + \alpha,
\end{aligned}
$$

where the inequality follows in the same way as in (K.14). Thus,

$$
\begin{aligned}
&\frac{\Phi_2(v, R)}{\exp(-\alpha/2)} = \int_{-\infty}^{v_1} \int_{-\infty}^{v_2} \frac{1}{2\pi |R|^{1/2}} \exp\left( -\frac{1}{2} (z' R^{-1} z - \alpha) \right) \mathrm{d}z_2 \mathrm{d}z_1 \\
&= \int_{-\infty}^{v_1 - z_1^\star} \int_{-\infty}^{v_2 - z_2^\star} \frac{1}{2\pi \sqrt{\tau_1 \tau_2}} \exp\left( -\frac{1}{2} \left( (z + z^\star)' R^{-1} (z + z^\star) - \alpha \right) \right) \mathrm{d}z_2 \mathrm{d}z_1 \\
&\geq \int_{-\infty}^{v_1 - z_1^\star} \int_{-\infty}^{v_2 - z_2^\star} \frac{1}{2\pi \sqrt{\tau_1 \tau_2}} \exp\left( -\frac{1}{2} \left( \left( \frac{z_1}{\sqrt{\tau_1}} + \sqrt{\tau_1} t_1 \right)^2 + \left( \frac{z_2}{\sqrt{\tau_1}} + \sqrt{\tau_1} t_2 \right)^2 - \left( \sqrt{\tau_1} t_1 \right)^2 - \left( \sqrt{\tau_1} t_2 \right)^2 \right) \right) \mathrm{d}z_2 \mathrm{d}z_1 \\
&\geq \int_{-\infty}^{0} \int_{-\infty}^{0} \frac{1}{2\pi \sqrt{\tau_1 \tau_2}} \exp\left( -\frac{1}{2} \left( \left( \frac{z_1}{\sqrt{\tau_1}} + \sqrt{\tau_1} t_1 \right)^2 + \left( \frac{z_2}{\sqrt{\tau_1}} + \sqrt{\tau_1} t_2 \right)^2 - \left( \sqrt{\tau_1} t_1 \right)^2 - \left( \sqrt{\tau_1} t_2 \right)^2 \right) \right) \mathrm{d}z_2 \mathrm{d}z_1 \\
&= \frac{1}{2\pi} \sqrt{\frac{\tau_1}{\tau_2}} \frac{\Phi\left( \sqrt{\tau_1} t_1 \right)}{\phi\left( \sqrt{\tau_1} t_1 \right)} \frac{\Phi\left( \sqrt{\tau_1} t_2 \right)}{\phi\left( \sqrt{\tau_1} t_2 \right)} \\
&\geq \frac{1}{2\pi} \sqrt{\frac{1 - |\rho|}{2}} \frac{\Phi\left( \sqrt{\tau_1} t_1 \right)}{\phi\left( \sqrt{\tau_1} t_1 \right)} \frac{\Phi\left( \sqrt{\tau_1} t_2 \right)}{\phi\left( \sqrt{\tau_1} t_2 \right)}.
\end{aligned}
$$

It follows from the proof of Lemma A.9 by Xu and Lee (2015) that $\phi(x)/\Phi(x) \leq 2(|x| + C_2)$. Thus,

$$
\frac{\exp(-\alpha/2)}{\Phi_2(v, R)} \leq 8\sqrt{2}\pi (1 - |\rho|)^{-1/2} \left( |\sqrt{\tau_1} t_1| + C_2 \right) \left( |\sqrt{\tau_1} t_2| + C_2 \right) \leq C_5 (1 - |\rho|)^{-1/2} \left( |t_1| + C_2 \right) \left( |t_2| + C_2 \right),
$$

for some constant $C_5 > 0$, since $|\tau_1| \leq 1$.

It is not difficult to see that the solution to $\min_z z' R^{-1} z$ s.t. $z \leq v$ with $(v_1, v_2) \notin (-1, +\infty) \times (-1, +\infty)$ is unique and takes one of the three values $(v_1, v_2)'$, $(v_1, \rho v_1)'$, or $(\rho v_2, v_2)'$ depending on the values of $v_1$, $v_2$, and $\rho$ (similarly to Example 1 in Hashorva and Hüsler, 2003). If $z^\star = (v_1, v_2)'$, then $t = (v_1 - \rho v_2,\ v_2 - \rho v_1)'/(1 - \rho^2)$ and

$$
\frac{\exp(-\alpha/2)}{\Phi_2(v, R)} \leq C_5 (1 - |\rho|)^{-1/2} \left( \left| \frac{v_1 - \rho v_2}{1 - \rho^2} \right| + C_2 \right) \left( \left| \frac{v_2 - \rho v_1}{1 - \rho^2} \right| + C_2 \right) \leq C_5 (1 - |\rho|)^{-5/2} (|v_1| + |v_2| + C_2)^2. \quad \text{(K.20)}
$$

If $z^\star = (v_1, \rho v_1)'$, then $t = (v_1,\ 0)'$ and

$$\frac{\exp(-\alpha/2)}{\Phi_2(v,R)} \leq C_5(1-|\rho|)^{-1/2}(|v_1|+C_2)C_2 \leq C_5(1-|\rho|)^{-5/2}(|v_1|+|v_2|+C_2)^2. \tag{K.21}$$

The bound when $z^\star = (\rho v_2, v_2)'$ can be derived analogously. Next, we calculate the bound for $\alpha^3$:

$$\begin{aligned}\alpha^3 &\leq (1-\rho^2)^{-3}(v_1^2 - 2\rho v_1 v_2 + v_2^2)^3 \leq (1-\rho^2)^{-3}(|v_1|^2 + 2|v_1||v_2| + |v_2|^2)^3 \\ &= (1-\rho^2)^{-3}(|v_1|+|v_2|)^6 \leq (1-|\rho|)^{-3}(|v_1|+|v_2|)^6.\end{aligned}$$

Hence,

$$\max\{6,\alpha\}^3 \leq (1-|\rho|)^{-3}(|v_1|+|v_2|+C_2)^6. \tag{K.22}$$

It follows from combining (K.19), (K.20), (K.21), and (K.22) that

$$\frac{\phi(v_1)\Phi\left(\frac{v_2-\rho v_1}{\sqrt{1-\rho^2}}\right)}{\Phi_2(v,R)} \leq C_1(1-|\rho|)^{-7}(|v_1|+|v_2|+C_2)^8, \tag{K.23}$$

if $(v_1, v_2) \notin (-1,+\infty)\times(-1,+\infty)$. The conclusion is obtained by combining (K.16) with (K.23).

We continue with the second claim of the lemma. If $(v_1, v_2) \in (-1,+\infty)\times(-1,+\infty)$, then clearly

$$\begin{aligned}\frac{\phi_2(v,R)}{\Phi_2(v,R)} &\leq \frac{1}{2\pi(1-\rho^2)^{1/2}\Phi_2(-\iota_2,R)} \leq \frac{1}{2\pi(1-|\rho|)^{1/2}\Phi_2(-\iota_2,R)} \leq C_1(1-|\rho|)^{-1}\left(1-\Phi\left((1-|\rho|)^{-1/2}\right)\right)^{-2} \\ &\leq C_1(1-|\rho|)^{-3}\left(1-\Phi\left((1-|\rho|)^{-1/2}\right)\right)^{-2},\end{aligned} \tag{K.24}$$

where the second inequality follows from (K.15). If $(v_1, v_2) \notin (-1,+\infty)\times(-1,+\infty)$,

$$\frac{\phi_2(v,R)}{\Phi_2(v,R)} = \frac{\exp(-v'R^{-1}v/2)}{2\pi(1-\rho^2)^{1/2}\Phi_2(v,R)} \leq \frac{\exp(-\alpha/2)}{2\pi(1-|\rho|)^{1/2}\Phi_2(v,R)} \leq C_1(1-|\rho|)^{-3}(|v_1|+|v_2|+C_2)^2, \tag{K.25}$$

where the result follows from (K.20) and (K.21). The conclusion is obtained by combining (K.24) and (K.25).

The third claim follows in the same way as in (K.15) with $-\iota_2$ replaced with the 2-dimensional vector of zeros. Thus,

$$\Phi_2(0,R) \geq \frac{(1-|\rho|)^{1/2}}{2^{5/2}} \geq C_3(1-|\rho|)^{1/2}.$$

□

***Proof of Lemma C.6***. Given a matrix function $F$ and a matrix $X$, we proceed as follows: (i) compute the differential of $F(X)$, (ii) vectorize to obtain $\mathrm{d}\operatorname{vec} F(X) = A(X)\mathrm{d}\operatorname{vec} X$, and (iii) conclude that $\partial \operatorname{vec} F(X)/\partial(\operatorname{vec} X)' = A(X)$ (see Magnus and Neudecker, 1999, for more details). The differential of the first function is given by $\mathrm{d}|F(\theta)| = |F(\theta)|\mathrm{Tr}\left(F^{-1}(\theta)\mathrm{d}F(\theta)\right) = |F(\theta)|\left(\operatorname{vec} F^{-1}(\theta)\right)'\mathrm{d}\operatorname{vec} F(\theta) = |F(\theta)|(\operatorname{vec} F^{-1}(\theta))'(\partial \operatorname{vec} F(\theta)/\partial\theta')\mathrm{d}\theta$, where we used that $\mathrm{Tr}(A'B) = (\operatorname{vec} A)' \operatorname{vec} B$ and $F(\theta)$ is symmetric implying that $F^{-1}(\theta)$ is symmetric as well. Hence,

$$\frac{\partial|F(\theta)|}{\partial\theta'} = |F(\theta)|\left(\operatorname{vec} F^{-1}(\theta)\right)' \frac{\partial \operatorname{vec} F(\theta)}{\partial\theta'}.$$

Thus given the definition of $K(\theta)$,

$$\frac{\partial|F(\theta)|}{\partial\theta} = |F(\theta)|K(\theta) \operatorname{vec} F^{-1}(\theta).$$

In order to obtain the second order derivative, we calculate the differential once more:

$$\begin{aligned}
\mathrm{d}\left(\frac{\partial|F(\theta)|}{\partial\theta}\right) &= \mathrm{d}|F(\theta)|K(\theta)\operatorname{vec} F^{-1}(\theta) + |F(\theta)|\mathrm{d}K(\theta)\operatorname{vec} F^{-1}(\theta) + |F(\theta)|K(\theta)\mathrm{d}\operatorname{vec} F^{-1}(\theta) \\
&= K(\theta)\operatorname{vec} F^{-1}(\theta)\mathrm{d}|F(\theta)| + |F(\theta)|\left((\operatorname{vec} F^{-1}(\theta))' \otimes I_p\right)\mathrm{d}\operatorname{vec} K(\theta) + |F(\theta)|K(\theta)\mathrm{d}\operatorname{vec} F^{-1}(\theta) \\
&= \left(K(\theta)\operatorname{vec} F^{-1}(\theta)\frac{\partial|F(\theta)|}{\partial\theta'} + |F(\theta)|\left((\operatorname{vec} F^{-1}(\theta))' \otimes I_p\right)\frac{\partial \operatorname{vec} K(\theta)}{\partial\theta'} + |F(\theta)|K(\theta)\frac{\partial \operatorname{vec} F^{-1}(\theta)}{\partial\theta'}\right)\mathrm{d}\theta,
\end{aligned}$$

where the second equality follows from $\operatorname{vec}(ABC) = (C' \otimes A)\operatorname{vec} B$. The result follows.

Next,

$$\begin{aligned}
\mathrm{d}(f'(\theta)F^{-1}(\theta)f(\theta)) &= 2f'(\theta)F^{-1}(\theta)\mathrm{d}f(\theta) + f'(\theta)\mathrm{d}F^{-1}(\theta)f(\theta) \\
&= 2f'(\theta)F^{-1}(\theta)\mathrm{d}f(\theta) + (f'(\theta) \otimes f'(\theta))\,\mathrm{d}\operatorname{vec} F^{-1}(\theta) \\
&= \left(2f'(\theta)F^{-1}(\theta)\partial f(\theta)/\partial\theta' + (f'(\theta) \otimes f'(\theta))\,\partial \operatorname{vec} F^{-1}(\theta)/\partial\theta'\right)\mathrm{d}\theta,
\end{aligned}$$

where the second equality follows from $\operatorname{vec}(ABC) = (C' \otimes A)\operatorname{vec} B$. Thus,

$$\frac{\partial(f'(\theta)F^{-1}(\theta)f(\theta))}{\partial\theta'} = 2f'(\theta)F^{-1}(\theta)\frac{\partial f(\theta)}{\partial\theta'} + (f'(\theta) \otimes f'(\theta))\frac{\partial \operatorname{vec} F^{-1}(\theta)}{\partial\theta'}.$$

The result is obtained by taking a transpose of this expression. We continue with the second differential:

$$
\begin{aligned}
\mathrm{d}\left(\frac{\partial(f'(\theta)F^{-1}(\theta)f(\theta))}{\partial\theta}\right) &= 2\mathrm{d}L(\theta)F^{-1}(\theta)f(\theta) + 2L(\theta)\mathrm{d}F^{-1}(\theta)f(\theta) + 2L(\theta)F^{-1}(\theta)\mathrm{d}f(\theta) \\
&\quad + \mathrm{d}M(\theta)\left(f(\theta)\otimes f(\theta)\right) + M(\theta)\mathrm{d}\left(f(\theta)\otimes f(\theta)\right) \\
&= 2\left(f'(\theta)F^{-1}(\theta)\otimes I_p\right)\mathrm{d}\operatorname{vec} L(\theta) + 2\left(f'(\theta)\otimes L(\theta)\right)\mathrm{d}\operatorname{vec} F^{-1}(\theta) + 2L(\theta)F^{-1}(\theta)\mathrm{d}f(\theta) \\
&\quad + \left(f'(\theta)\otimes f'(\theta)\otimes I_p\right)\mathrm{d}\operatorname{vec} M(\theta) + M(\theta)(\mathrm{K}_{1n}\otimes I_n)\left[(I_n\otimes f(\theta))\mathrm{d}f(\theta) + (f(\theta)\otimes I_n)\mathrm{d}f(\theta)\right] \\
&= 2\left(f'(\theta)F^{-1}(\theta)\otimes I_p\right)\mathrm{d}\operatorname{vec} L(\theta) + 2\left(f'(\theta)\otimes L(\theta)\right)\mathrm{d}\operatorname{vec} F^{-1}(\theta) + \left(f'(\theta)\otimes f'(\theta)\otimes I_p\right)\mathrm{d}\operatorname{vec} M(\theta) \\
&\quad + \left(2L(\theta)F^{-1}(\theta) + M(\theta)(\mathrm{K}_{1n}\otimes I_n)\left[(I_n\otimes f(\theta)) + (f(\theta)\otimes I_n)\right]\right)\mathrm{d}f(\theta) \\
&= \Bigg(2\left(f'(\theta)F^{-1}(\theta)\otimes I_p\right)\frac{\partial\operatorname{vec} L(\theta)}{\partial\theta'} + 2\left(f'(\theta)\otimes L(\theta)\right)\frac{\partial\operatorname{vec} F^{-1}(\theta)}{\partial\theta'} + \left(f'(\theta)\otimes f'(\theta)\otimes I_p\right)\frac{\partial\operatorname{vec} M(\theta)}{\partial\theta'} \\
&\quad + \left(2L(\theta)F^{-1}(\theta) + M(\theta)(\mathrm{K}_{1n}\otimes I_n)\left[(I_n\otimes f(\theta)) + (f(\theta)\otimes I_n)\right]\right)\frac{\partial f(\theta)}{\partial\theta'}\Bigg)\mathrm{d}\theta,
\end{aligned}
$$

where the second equality follows from $\operatorname{vec}(ABC) = (C'\otimes A)\operatorname{vec} B$ and for $X$ and $Y$ being $n\times q$ and $p\times r$ matrices, $\mathrm{d}\operatorname{vec}(X\otimes Y) = (I_q\otimes \mathrm{K}_{rn}\otimes I_p)\left[(I_{nq}\otimes\operatorname{vec} Y)\mathrm{d}\operatorname{vec} X + (\operatorname{vec} X\otimes I_{pr})\mathrm{d}\operatorname{vec} Y\right]$ as derived in Magnus and Neudecker (1999, p. 185). The conclusion follows. $\square$

***Proof of Lemma C.7***.

$$
\begin{aligned}
\|A\otimes B\| &= \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=1}^{p}\sum_{k=1}^{q}(A_{ij}B_{lk})^2} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}A_{ij}^2\sum_{l=1}^{p}\sum_{k=1}^{q}B_{lk}^2} \\
&= \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}A_{ij}^2\|B\|^2} = \sqrt{\|A\|^2\|B\|^2} = \|A\|\|B\|.
\end{aligned}
$$

$\square$

***Proof of Lemma C.8***. Based on equation (9) in Muthén (1990),

$$
\begin{aligned}
E[X_1^2|X\le v] &= 1 - v_1\frac{\phi(v_1)\Phi\left(\frac{v_2-\rho v_1}{\sqrt{1-\rho^2}}\right)}{\Phi_2(v,R)} - \rho^2 v_2\frac{\phi(v_2)\Phi\left(\frac{v_1-\rho v_2}{\sqrt{1-\rho^2}}\right)}{\Phi_2(v,R)} + \rho\sqrt{1-\rho^2}\frac{\phi(v_2)\phi\left(\frac{v_1-\rho v_2}{\sqrt{1-\rho^2}}\right)}{\Phi_2(v,R)} \\
&= 1 - v_1\frac{\phi(v_1)\Phi\left(\frac{v_2-\rho v_1}{\sqrt{1-\rho^2}}\right)}{\Phi_2(v,R)} - \rho^2 v_2\frac{\phi(v_2)\Phi\left(\frac{v_1-\rho v_2}{\sqrt{1-\rho^2}}\right)}{\Phi_2(v,R)} + \rho(1-\rho^2)\frac{\phi_2(v,R)}{\Phi_2(v,R)} \\
&= 1 - v_1\xi_1(v,R) - \rho^2 v_2\xi_2(v,R) + \rho(1-\rho^2)\kappa(v,R),
\end{aligned}
$$

where the second equality follows by observing that $\phi(v_2)\phi((v_1-\rho v_2)/\sqrt{1-\rho^2})/\sqrt{1-\rho^2} = \phi_2(v,R)$, whereas the last equality follows by the definitions of $\xi(v,R)$ and $\kappa(v,R)$ in (C.1) and (C.2), respectively. Symmetrically,

$$E[X_2^2|X \leq v] = 1 - v_2\xi_2(v,R) - \rho^2 v_1\xi_1(v,R) + \rho(1-\rho^2)\kappa(v,R).$$

From equation (11) in Muthén (1990),

$$\begin{aligned}
E[X_1X_2|X \leq v] &= \rho - \rho v_1 \frac{\phi(v_1)\Phi\left(\frac{v_2-\rho v_1}{\sqrt{1-\rho^2}}\right)}{\Phi_2(v,R)} - \rho v_2 \frac{\phi(v_2)\Phi\left(\frac{v_1-\rho v_2}{\sqrt{1-\rho^2}}\right)}{\Phi_2(v,R)} + \sqrt{1-\rho^2}\frac{\phi(v_2)\phi\left(\frac{v_1-\rho v_2}{\sqrt{1-\rho^2}}\right)}{\Phi_2(v,R)} \\
&= \rho - \rho v_1\xi_1(v,R) - \rho v_2\xi_2(v,R) + (1-\rho^2)\kappa(v,R).
\end{aligned}$$

The conclusion follows by noticing that $E[XX'|X \leq v] = (E[X_1^2|X \leq v]\ E[X_1X_2|X \leq v];\ E[X_1X_2|X \leq v]\ E[X_2^2|X \leq v])$. □

***Proof of Lemma C.9***. The proof closely follows the proof of Theorem 17.9 of Davidson (1994). Let $X_{i,n}^{(s)} = E[X_{i,n}|\mathcal{F}_{i,n}(s)]$ and $Y_{i,n}^{(s)} = E[Y_{i,n}|\mathcal{F}_{i,n}(s)]$. Then

$$\begin{aligned}
&\|X_{i,n}Y_{i,n} - E[X_{i,n}Y_{i,n}|\mathcal{F}_{i,n}(s)]\|_p \\
&= \|(X_{i,n}Y_{i,n} - X_{i,n}Y_{i,n}^{(s)}) + (X_{i,n}Y_{i,n}^{(s)} - X_{i,n}^{(s)}Y_{i,n}^{(s)}) - E[(X_{i,n} - X_{i,n}^{(s)})(Y_{i,n} - Y_{i,n}^{(s)})|\mathcal{F}_{i,n}(s)]\|_p \\
&\leq \|X_{i,n}(Y_{i,n} - Y_{i,n}^{(s)})\|_p + \|Y_{i,n}^{(s)}(X_{i,n} - X_{i,n}^{(s)})\|_p + \|E[(X_{i,n} - X_{i,n}^{(s)})(Y_{i,n} - Y_{i,n}^{(s)})|\mathcal{F}_{i,n}(s)]\|_p \\
&\leq \|X_{i,n}\|_{2p}\|Y_{i,n} - Y_{i,n}^{(s)}\|_{2p} + \|Y_{i,n}^{(s)}\|_{2p}\|X_{i,n} - X_{i,n}^{(s)}\|_{2p} + \|E[(X_{i,n} - X_{i,n}^{(s)})(Y_{i,n} - Y_{i,n}^{(s)})|\mathcal{F}_{i,n}(s)]\|_p \\
&\leq \|X_{i,n}\|_{2p}\|Y_{i,n} - Y_{i,n}^{(s)}\|_{2p} + \|Y_{i,n}\|_{2p}\|X_{i,n} - X_{i,n}^{(s)}\|_{2p} + \|(X_{i,n} - X_{i,n}^{(s)})(Y_{i,n} - Y_{i,n}^{(s)})\|_p \\
&\leq \|X_{i,n}\|_{2p}\|Y_{i,n} - Y_{i,n}^{(s)}\|_{2p} + \|Y_{i,n}\|_{2p}\|X_{i,n} - X_{i,n}^{(s)}\|_{2p} + \|X_{i,n} - X_{i,n}^{(s)}\|_{2p}\|Y_{i,n} - Y_{i,n}^{(s)}\|_{2p} \\
&\leq \|X_{i,n}\|_{2p}t_{i,n}^Y\psi^Y(s) + \|Y_{i,n}\|_{2p}t_{i,n}^X\psi^X(s) + t_{i,n}^X\psi^X(s)t_{i,n}^Y\psi^Y(s) \\
&\leq t_{i,n}\psi(s),
\end{aligned}$$

where the first and second inequalities are implied by the Minkowski's and Cauchy-Schwartz inequalities, respectively, whereas the third inequality follows by the conditional Jensen's inequality and law of iterated expectations; the fourth inequality again follows by the Cauchy-Schwartz inequality. The final claim of the lemma follows from Definition 2. □

# Appendix L Proof of Lemmas in Appendix D

***Proof of Lemma D.1***. *(i)* Let $\tau^b_{1g,n}(\theta) \leq \tau^b_{2g,n}(\theta)$ be the eigenvalues of $\Omega^b_{g,n}(\theta)$. Then

$$\inf_{n,g}\inf_{\theta\in\Theta}|\Omega^b_{g,n}(\theta)| = \inf_{n,g}\inf_{\theta\in\Theta}(\tau^b_{1g,n}(\theta)\tau^b_{2g,n}(\theta)) \geq \inf_{n,g}\inf_{\theta\in\Theta}\tau^{b^2}_{1g,n}(\theta) > 0$$

by Assumption 6.

In the same way as in the proof of Lemma 2 by Xu and Lee (2015), it follows that $\inf_{n,i}\inf_{\theta\in\Theta}\Omega^b_{ii,n}(\theta) \geq \inf_n(\|I_{2n} - \lambda^b W^b_n\|_\infty \|I_{2n} - \lambda^b W^b_n\|_1 \cdot \min\{1,\sigma^2\})^{-1} > 0$ by Assumptions 1$(ii)$, 2$(i)$, and 7.

*(ii)* Next, let $d, e \in \{s, o\}$. Then uniformly in $n \in \mathbb{N}$, $g \in \mathcal{G}_n$, and $\theta \in \Theta$, $\|\Omega^c_{g,n}(\theta)\| \leq C_1\|\Omega^c_{g,n}(\theta)\|_\infty \leq C_1\|\Omega^c_n(\theta)\|_\infty \leq C_1C_2\|(I_{2n}-\lambda^d W^d_n)^{-1}(I_{2n}-\lambda^e W^{e'}_n)^{-1}\|_\infty \leq C_1C_2\|(I_{2n}-\lambda^d W^d_n)^{-1}\|_\infty\|(I_{2n}-\lambda^e W^e_n)^{-1}\|_1 < \infty$ for some constants $C_1, C_2 > 0$. The first inequality is implied by the equivalence of matrix norms on finite dimensional matrix spaces, whereas the third inequality follows by compactness of the parameter space. The conclusion is implied by Assumption 1$(ii)$.

Next, note that

$$\left\|\frac{\partial \operatorname{vec}\Omega^c_{g,n}(\theta)}{\partial\theta'}\right\| = \sqrt{\left\|\frac{\partial\Omega^c_{g,n}(\theta)}{\partial\lambda^s}\right\|^2 + \left\|\frac{\partial\Omega^c_{g,n}(\theta)}{\partial\lambda^o}\right\|^2 + \left\|\frac{\partial\Omega^c_{g,n}(\theta)}{\partial\rho}\right\|^2 + \left\|\frac{\partial\Omega^c_{g,n}(\theta)}{\partial\sigma^2}\right\|^2}.$$

We will show that $\|\partial\Omega^c_{g,n}(\theta)/\partial\lambda^s\|$ is uniformly bounded, while the boundedness of the other terms can be established in a similar way. For some constant $C_3 > 0$, uniformly in $n \in \mathbb{N}$, $g \in \mathcal{G}_n$, and $\theta \in \Theta$

$$\begin{aligned}
\left\|\frac{\partial\Omega^c_{g,n}(\theta)}{\partial\lambda^s}\right\| &\leq C_3\left\|\frac{\partial\Omega^c_{g,n}(\theta)}{\partial\lambda^s}\right\|_\infty \\
&\leq C_3\left\|\frac{\partial\Omega^c_{n}(\theta)}{\partial\lambda^s}\right\|_\infty \\
&\leq C_2C_3\left\|\frac{\partial((I_{2n}-\lambda^d W^d_n)^{-1}(I_{2n}-\lambda^e W^{e'}_n)^{-1})}{\partial\lambda^s}\right\|_\infty \\
&\leq C_2C_3\Big(\|(I_{2n}-\lambda^d W^d_n)^{-1}W^d_n(I_{2n}-\lambda^d W^d_n)^{-1}(I_{2n}-\lambda^e W^{e'}_n)^{-1}\|_\infty \\
&\quad+\|(I_{2n}-\lambda^d W^d_n)^{-1}(I_{2n}-\lambda^e W^{e'}_n)^{-1}W^{e'}_n(I_{2n}-\lambda^e W^{e'}_n)^{-1}\|_\infty\Big) < \infty.
\end{aligned}$$

The first inequality follows by the equivalence of matrix norms on finite dimensional matrix spaces. The result is implied by sub-multiplicity of matrix norms and Assumption 1$(ii)$.

*(iii)* Note that $\|\Omega^{b-1}_{g,n}(\theta)\| = |\Omega^b_{g,n}(\theta)|^{-1}\|\Omega^b_{g,n}(\theta)\| < \infty$ uniformly in $n \in \mathbb{N}$, $g \in \mathcal{G}_n$, and $\theta \in \Theta$ by parts *(i)* and

*(ii)*.

Consequently, by Lemma C.7,

$$\begin{aligned}\left\|\frac{\partial \operatorname{vec} \Omega_{g,n}^{b-1}(\theta)}{\partial \theta'}\right\| &= \left\|(\Omega_{g,n}^{b-1}(\theta) \otimes \Omega_{g,n}^{b-1}(\theta)) \frac{\partial \operatorname{vec} \Omega_{g,n}^{b}(\theta)}{\partial \theta'}\right\| \\ &\leq \|\Omega_{g,n}^{b-1}(\theta)\|^2 \left\|\frac{\partial \operatorname{vec} \Omega_{g,n}^{b}(\theta)}{\partial \theta'}\right\| < \infty.\end{aligned}$$

Lemma C.6 implies that

$$\begin{aligned}\left\|\frac{\partial |\Omega_{g,n}^{b}(\theta)|}{\partial \theta}\right\| &= \left\||\Omega_{g,n}^{b}(\theta)| \left(\frac{\partial \operatorname{vec} \Omega_{g,n}^{b}(\theta)}{\partial \theta'}\right)' \operatorname{vec} \Omega_{g,n}^{b-1}(\theta)\right\| \\ &\leq |\Omega_{g,n}^{b}(\theta)| \left\|\frac{\partial \operatorname{vec} \Omega_{g,n}^{b}(\theta)}{\partial \theta'}\right\| \|\Omega_{g,n}^{b-1}(\theta)\| < \infty\end{aligned}$$

uniformly in $n \in \mathbb{N}$, $g \in \mathcal{G}_n$, and $\theta \in \Theta$, where the boundedness of $|\Omega_{g,n}^{b}(\theta)|$ is implied by the boundedness of $\|\Omega_{g,n}^{b}(\theta)\|$. □

***Proof of Lemma D.2***. *(i)* From Exercise 12.39 in Abadir and Magnus (2005), for any symmetric matrix $A$ and compatible vector $x$, $x'Ax \geq \operatorname{mineig}(A)x'x$, where $\operatorname{mineig}(A)$ is the minimum eigenvalue of $A$. Let $\tau_{1g,n}^{11}(\theta) \leq \tau_{2g,n}^{11}(\theta)$ be the eigenvalues of $\Sigma_{g,n}^{11}(\theta)$ and consider vectors $x = (1, 0)'$ and $x = (0, 1)'$ for $j = 1$ and $j = 2$, respectively. Then $\inf\limits_{n,g} \inf\limits_{\theta \in \Theta} \Sigma_{gjj,n}^{11}(\theta) \geq \inf\limits_{n,g} \inf\limits_{\theta \in \Theta} \tau_{1g,n}^{11}(\theta) > 0$ by Assumption 6.

Next, note that

$$|\Sigma_{g,n}^{11}(\theta)| = \Sigma_{g11,n}^{11}(\theta)\Sigma_{g22,n}^{11}(\theta) - \Sigma_{g12,n}^{11^2}(\theta) = \Sigma_{g11,n}^{11}(\theta)\Sigma_{g22,n}^{11}(\theta)(1 - \rho_{g,n}^{11^2}(\theta)).$$

Thus,

$$|\rho_{g,n}^{11}(\theta)| = \sqrt{1 - \frac{|\Sigma_{g,n}^{11}(\theta)|}{\Sigma_{g11,n}^{11}(\theta)\Sigma_{g22,n}^{11}(\theta)}} = \sqrt{1 - \frac{\tau_{1g,n}^{11}(\theta)\tau_{2g,n}^{11}(\theta)}{\Sigma_{g11,n}^{11}(\theta)\Sigma_{g22,n}^{11}(\theta)}} \leq \sqrt{1 - \frac{\tau_{1g,n}^{11^2}(\theta)}{\Sigma_{g11,n}^{11}(\theta)\Sigma_{g22,n}^{11}(\theta)}} < 1$$

uniformly in $g \in \mathcal{G}_n$, $n \in \mathbb{N}$, and $\theta \in \Theta$ by Assumption 6 and the fact that $\inf\limits_{n,g} \inf\limits_{\theta \in \Theta} \min\limits_{j=1,2} \Sigma_{gjj,n}^{11}(\theta) > 0$.

*(ii)* Let $M^{11}_{g,n}(\theta) = \mathrm{Diag}(\Sigma^{11}_{g,n}(\theta))^{-1/2}$. Then

$$\begin{aligned}
\|R^{11}_{g,n}(\theta)\| &= \|M^{11}_{g,n}(\theta)\Sigma^{11}_{g,n}(\theta)M^{11}_{g,n}(\theta)\| \\
&\leq \|M^{11}_{g,n}(\theta)\|^2\|\Sigma^{11}_{g,n}(\theta)\| \\
&= \|M^{11}_{g,n}(\theta)\|^2\|\tilde{\Omega}^{ss}_{g,n}(\theta) - \tilde{\Omega}^{so}_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)\tilde{\Omega}^{so'}_{g,n}(\theta)\| \\
&\leq \|M^{11}_{g,n}(\theta)\|^2\left(\|\Omega^{ss}_{g,n}(\theta)\| + \|\Omega^{so}_{g,n}(\theta)\|^2\|\Omega^{oo-1}_{g,n}(\theta)\|\right) \\
&< \infty
\end{aligned} \tag{L.1}$$

uniformly in $n \in \mathbb{N}$, $g \in \mathcal{G}_n$, and $\theta \in \Theta$ because of Lemma D.1 and the first part of this lemma, which guarantees that $\|M^{11}_{g,n}(\theta)\|$ is uniformly bounded. By Lemma C.6,

$$\begin{aligned}
\left\|\frac{\partial|R^{11}_{g,n}(\theta)|}{\partial\theta}\right\| &= \left\||R^{11}_{g,n}(\theta)|\left(\frac{\partial\,\mathrm{vec}\,R^{11}_{g,n}(\theta)}{\partial\theta'}\right)'\mathrm{vec}\,R^{11^{-1}}_{g,n}(\theta)\right\| \\
&\leq |R^{11}_{g,n}(\theta)|\left\|\frac{\partial\,\mathrm{vec}\,R^{11}_{g,n}(\theta)}{\partial\theta'}\right\|\|R^{11^{-1}}_{g,n}(\theta)\|.
\end{aligned} \tag{L.2}$$

The uniform boundedness of $\|R^{11}_{g,n}(\theta)\|$ implies that $|R^{11}_{g,n}(\theta)|$ is uniformly bounded as well. By the first part of the proof, $\inf_{n,g}\inf_{\theta\in\Theta}|R^{11}_{g,n}(\theta)| = \inf_{n,g}\inf_{\theta\in\Theta}(1-\rho^{11^2}_{g,n}(\theta)) > 0$. Thus, the last term in (L.2) is uniformly bounded by noticing that $\|R^{11^{-1}}_{g,n}(\theta)\| = |R^{11}_{g,n}(\theta)|^{-1}\|R^{11}_{g,n}(\theta)\|$.

It remains to show that the second term in (L.2) is uniformly bounded:

$$\begin{aligned}
&\left\|\frac{\partial\,\mathrm{vec}\,R^{11}_{g,n}(\theta)}{\partial\theta'}\right\| = \left\|\frac{\partial\,\mathrm{vec}(M^{11}_{g,n}(\theta)\Sigma^{11}_{g,n}(\theta)M^{11}_{g,n}(\theta))}{\partial\theta'}\right\| \\
&= \left\|\left[\left(M^{11}_{g,n}(\theta)\Sigma^{11}_{g,n}(\theta)\otimes I_2\right) + \left(I_2\otimes M^{11}_{g,n}(\theta)\Sigma^{11}_{g,n}(\theta)\right)\right]\frac{\partial\,\mathrm{vec}\,M^{11}_{g,n}(\theta)}{\partial\theta'} + \left(M^{11}_{g,n}(\theta)\otimes M^{11}_{g,n}(\theta)\right)\frac{\partial\,\mathrm{vec}\,\Sigma^{11}_{g,n}(\theta)}{\partial\theta'}\right\| \\
&\leq \left(\|M^{11}_{g,n}(\theta)\Sigma^{11}_{g,n}(\theta)\otimes I_2\| + \|I_2\otimes M^{11}_{g,n}(\theta)\Sigma^{11}_{g,n}(\theta)\|\right)\left\|\frac{\partial\,\mathrm{vec}\,M^{11}_{g,n}(\theta)}{\partial\theta'}\right\| \\
&\qquad + \|M^{11}_{g,n}(\theta)\otimes M^{11}_{g,n}(\theta)\|\left\|\frac{\partial\,\mathrm{vec}\,\Sigma^{11}_{g,n}(\theta)}{\partial\theta'}\right\| \\
&\leq 2\sqrt{2}\|M^{11}_{g,n}(\theta)\|\|\Sigma^{11}_{g,n}(\theta)\|\left\|\frac{\partial\,\mathrm{vec}\,M^{11}_{g,n}(\theta)}{\partial\theta'}\right\| + \|M^{11}_{g,n}(\theta)\|^2\left\|\frac{\partial\,\mathrm{vec}\,\Sigma^{11}_{g,n}(\theta)}{\partial\theta'}\right\|.
\end{aligned}$$

The norm of $M^{11}_{g,n}(\theta)$ is uniformly bounded due to the first part of this lemma, whereas the boundedness of $\|\Sigma^{11}_{g,n}(\theta)\|$

is implied by the proof of the boundedness of $\|R_{g,n}^{11}(\theta)\|$ in (L.1). Note that

$$\left\|\frac{\partial \operatorname{vec} M_{g,n}^{11}(\theta)}{\partial \theta'}\right\| = \sqrt{\left\|\frac{\partial M_{g,n}^{11}(\theta)}{\partial \lambda^s}\right\|^2 + \left\|\frac{\partial M_{g,n}^{11}(\theta)}{\partial \lambda^o}\right\|^2 + \left\|\frac{\partial M_{g,n}^{11}(\theta)}{\partial \rho}\right\|^2 + \left\|\frac{\partial M_{g,n}^{11}(\theta)}{\partial \sigma^2}\right\|^2}.$$

We will show that the first term is uniformly bounded, while the uniform boundedness of the other terms can be established in a similar way:

$$\left\|\frac{\partial M_{g,n}^{11}(\theta)}{\partial \lambda^s}\right\| = \left\|-\frac{1}{2}\mathrm{Diag}(\Sigma_{g,n}^{11}(\theta))^{-3/2}\mathrm{Diag}\left(\frac{\partial \Sigma_{g,n}^{11}(\theta)}{\partial \lambda^s}\right)\right\| \leq \|\mathrm{Diag}(\Sigma_{g,n}^{11}(\theta))^{-3/2}\| \left\|\frac{\partial \Sigma_{g,n}^{11}(\theta)}{\partial \lambda^s}\right\|.$$

The first term on the right hand side is uniformly bounded due to the first part of the lemma. Regarding the second term, note that $\|\partial \Sigma_{g,n}^{11}(\theta)/\partial \lambda^s\| = \|\partial(\tilde{\Omega}_{g,n}^{ss}(\theta) - \tilde{\Omega}_{g,n}^{so}(\theta)\Omega_{g,n}^{oo-1}(\theta)\tilde{\Omega}_{g,n}^{so'}(\theta))/\partial \lambda^s\|$; it is uniformly bounded by the triangle inequality, the product rule, sub-multiplicity of matrix norms, and Lemma D.1. The boundedness of $\|\partial \operatorname{vec} \Sigma_{g,n}^{11}(\theta)/\partial \theta'\|$ can be established similarly. It concludes the proof that the second term in (L.2) is uniformly bounded.

Finally,

$$\left\|\frac{\partial \operatorname{vec} R_{g,n}^{11^{-1}}(\theta)}{\partial \theta'}\right\| = \left\|\left(R_{g,n}^{11^{-1}}(\theta) \otimes R_{g,n}^{11^{-1}}(\theta)\right)\frac{\partial \operatorname{vec} R_{g,n}^{11}(\theta)}{\partial \theta'}\right\| \leq \|R_{g,n}^{11^{-1}}(\theta)\|^2 \left\|\frac{\partial \operatorname{vec} R_{g,n}^{11}(\theta)}{\partial \theta'}\right\|$$

is uniformly bounded in $n \in \mathbb{N}$, $g \in \mathcal{G}_n$, and $\theta \in \Theta$ by the previous results of this proof. $\square$

***Proof of Lemma D.3***. The proof holds under Assumption 4($ii$) or Assumption 9($ii$). For the sake of simplicity, we will refer to Assumption 4($ii$) only.

Employing the equivalence of vector norms on finite dimensional vector spaces and Loève's $c_r$-inequality, it follows for some constant $C_1 > 0$ that uniformly in $n \in \mathbb{N}$ and $g \in \mathcal{G}_n$

$$\begin{aligned} E\left[\sup_{\theta\in\Theta}\|S_{g\cdot,n}^b(\lambda^b)X_n^b\beta^b\|\right]^p &\leq C_1 E\left[\sup_{\theta\in\Theta}|S_{g1\cdot,n}^b(\lambda^b)X_n^b\beta^b| + \sup_{\theta\in\Theta}|S_{g2\cdot,n}^b(\lambda^b)X_n^b\beta^b|\right]^p \\ &\leq 2^{p-1}C_1\left(E\left[\sup_{\theta\in\Theta}|S_{g1\cdot,n}^b(\lambda^b)X_n^b\beta^b|\right]^p + E\left[\sup_{\theta\in\Theta}|S_{g2\cdot,n}^b(\lambda^b)X_n^b\beta^b|\right]^p\right) \qquad \text{(L.3)} \\ &\leq 2^p C_1 \sup_{n,i} E\left[\sup_{\theta\in\Theta}|S_{i\cdot,n}^b(\lambda^b)X_n^b\beta^b|\right]^p. \end{aligned}$$

We proceed with the last term in (L.3): uniformly in $n \in \mathbb{N}$ and $i \in \{1, 2, \ldots, 2n\}$,

$$\begin{aligned}
E\left[\sup_{\theta\in\Theta} |S^b_{i\cdot,n}(\lambda^b) X^b_n \beta^b|\right]^p &= E\left[\sup_{\theta\in\Theta} |\sum_{j=1}^{2n} S^b_{ij,n}(\lambda^b) X^b_{j\cdot,n}\beta^b|\right]^p \\
&\leq E\left[\sup_{\theta\in\Theta} \sum_{j=1}^{2n} |S^b_{ij,n}(\lambda^b)| \|X^b_{j\cdot,n}\| \|\beta^b\|\right]^p \\
&\leq \sup_{\theta\in\Theta} \|\beta^b\|^p \sup_{\theta\in\Theta}\left(\sum_{j=1}^{2n} |S^b_{ij,n}(\lambda^b)|\right)^p E\left[\sup_{\theta\in\Theta} \frac{\sum_{j=1}^{2n} |S^b_{ij,n}(\lambda^b)| \|X^b_{j\cdot,n}\|}{\sum_{j=1}^{2n} |S^b_{ij,n}(\lambda^b)|}\right]^p \qquad (L.4)\\
&\leq \sup_{\theta\in\Theta} \|\beta^b\|^p \sup_{\theta\in\Theta}\left(\sum_{j=1}^{2n} |S^b_{ij,n}(\lambda^b)|\right)^p \sup_{\theta\in\Theta} \frac{\sum_{j=1}^{2n} |S^b_{ij,n}(\lambda^b)| E\|X^b_{j\cdot,n}\|^p}{\sum_{j=1}^{2n} |S^b_{ij,n}(\lambda^b)|} \\
&\leq \sup_{\theta\in\Theta} \|\beta^b\|^p \sup_n \sup_{\theta\in\Theta} \|S^b_n(\lambda^b)\|^p_\infty \sup_{n,i} E\|X^b_{i\cdot,n}\|^p \\
&< \infty,
\end{aligned}$$

where the third inequality follows by Jensen's inequality for convex functions. The conclusion is implied by Assumptions 1($ii$), 4($ii$), and 7.

Next,

$$\begin{aligned}
E\left[\sup_{\theta\in\Theta} \|z_{g,n}(\theta)\|\right]^p &= E\left[\sup_{\theta\in\Theta} \|y^o_{g,n} - S^o_{g\cdot,n}(\lambda^o) X^o_n \beta^o\|\right]^p \\
&\leq 2^{p-1}\left(E\|y^o_{g,n}\|^p + E\left[\sup_{\theta\in\Theta} \|S^o_{g\cdot,n}(\lambda^o) X^o_n \beta^o\|\right]^p\right),
\end{aligned} \qquad (L.5)$$

where the inequality follows by the triangle and Loève's $c_r$-inequalities. Given that we have already shown that the second term in (L.5) is uniformly bounded, it is enough to establish that $\sup_{n,i} E|y^o_{i,n}|^p < \infty$:

$$\begin{aligned}
E|y^o_{i,n}|^p &\leq E\left[|y^{*o}_{i,n}|^p \middle| y^s_{i,n} = 1\right] = E\left[|S^o_{i\cdot,n}(\lambda^o_0) X^o_n \beta^o_0 + \varepsilon^o_{i,n}(\lambda^o_0)|^p \middle| y^s_{i,n} = 1\right] \\
&\leq 2^{p-1}\left(E\left[|S^o_{i\cdot,n}(\lambda^o_0) X^o_n \beta^o_0|^p \middle| y^s_{i,n} = 1\right] + E\left[|\varepsilon^o_{i,n}(\lambda^o_0)|^p \middle| y^s_{i,n} = 1\right]\right).
\end{aligned}$$

Now we show that each term is bounded. In a similar way as in (L.4),

$$E\left[|S^o_{i\cdot,n}(\lambda^o_0) X^o_n \beta^o_0|^p \middle| y^s_{i,n} = 1\right] \leq \|\beta^o_0\|^p \sup_n \|S^o_n(\lambda^o_0)\|^p_\infty \sup_{n,i} E\left[\|X^o_{i\cdot,n}\|^p \middle| y^s_{i,n} = 1\right] < \infty$$

by Assumptions 1($ii$), 4($ii$), and 7. By the law of iterated expectations,

$$E\left[|\varepsilon^o_{i,n}(\lambda^o_0)|^p \middle| y^s_{i,n} = 1\right] = E\left[E\left[|\varepsilon^o_{i,n}(\lambda^o_0)|^p \middle| y^s_{i,n} = 1, X^s_n\right] \middle| y^s_{i,n} = 1\right]. \qquad (L.6)$$

Firstly, we will find the inner expectation by applying the law of iterated expectations once more:

$$E\left[|\varepsilon^o_{i,n}(\lambda^o_0)|^p \middle| y^s_{i,n}=1, X^s_n\right] = E[E[|\varepsilon^o_{i,n}(\lambda^o_0)|^p|\varepsilon^s_{i,n}(\lambda^s_0), X^s_n]|y^s_{i,n}=1, X^s_n] = E\left[E\left[|\varepsilon^o_{i,n}(\lambda^o_0)|^p \middle|\varepsilon^s_{i,n}(\lambda^s_0)\right] \middle| y^s_{i,n}=1, X^s_n\right], \tag{L.7}$$

where the last equality follows by Assumption 2($ii$).

Note that $(\varepsilon^o_{i,n}(\lambda^o_0), \varepsilon^s_{i,n}(\lambda^s_0))' \sim \mathcal{N}(0, [\Omega^{oo}_{ii,n}(\theta_0)\ \Omega^{so}_{ii,n}(\theta_0); \Omega^{so}_{ii,n}(\theta_0)\ \Omega^{ss}_{ii,n}(\theta_0)])$. Thus, $\varepsilon^o_{i,n}(\lambda^o_0)|\varepsilon^s_{i,n}(\lambda^s_0) \sim \mathcal{N}(\tilde{\mu}_{i,n}(\theta_0), \tilde{\sigma}^2_{i,n}(\theta_0))$ with $\tilde{\mu}_{i,n}(\theta_0) = \sqrt{\Omega^{oo}_{ii,n}(\theta_0)/\Omega^{ss}_{ii,n}(\theta_0)}\rho^{so}_{i,n}(\theta_0)\varepsilon^s_{i,n}(\lambda^s_0)$ and $\tilde{\sigma}^2_{i,n}(\theta_0) = (1-\rho^{so^2}_{i,n}(\theta_0))\Omega^{oo}_{ii,n}(\theta_0)$, where $\rho^{so}_{i,n}(\theta_0)$ is the correlation coefficient of $\varepsilon^s_{i,n}(\lambda^s_0)$ and $\varepsilon^o_{i,n}(\lambda^o_0)$. Hence, the inner expectation in (L.7) can be bounded by

$$\begin{aligned}
E\left[|\varepsilon^o_{i,n}(\lambda^o_0)|^p \middle|\varepsilon^s_{i,n}(\lambda^s_0)\right] &= E\left[|\varepsilon^o_{i,n}(\lambda^o_0) - \tilde{\mu}_{i,n}(\theta_0) + \tilde{\mu}_{i,n}(\theta_0)|^p \middle|\varepsilon^s_{i,n}(\lambda^s_0)\right] \\
&\le 2^{p-1}\left(E\left[|\varepsilon^o_{i,n}(\lambda^o_0) - \tilde{\mu}_{i,n}(\theta_0)|^p \middle|\varepsilon^s_{i,n}(\lambda^s_0)\right] + E\left[|\tilde{\mu}_{i,n}(\theta_0)|^p \middle|\varepsilon^s_{i,n}(\lambda^s_0)\right]\right) \\
&= 2^{p-1}\left(\tilde{\sigma}^p_{i,n}(\theta_0) 2^{p/2}\Gamma\left(\frac{p+1}{2}\right)/\sqrt{\pi} + |\tilde{\mu}_{i,n}(\theta_0)|^p\right) \\
&\le C_2 + C_3|\varepsilon^s_{i,n}(\lambda^s_0)|^p
\end{aligned}$$

for some constants $C_2, C_3 > 0$, where $\Gamma(\cdot)$ is the Gamma function. The second equality is implied by the following fact: if $X \sim \mathcal{N}(0,\sigma^2)$, then for any $p \in (-1, +\infty)$, $E[|X|^p] = \sigma^p 2^{p/2}\Gamma((p+1)/2)/\sqrt{\pi}$ (Kamat, 1953). The conclusion follows by noticing that Lemma D.1 implies the uniform boundedness from zero of $\Omega^{ss}_{ii,n}(\theta_0)$ and the uniform boundedness of $\Omega^{oo}_{ii,n}(\theta_0)$ which is implied by the uniform boundedness of $\|\Omega^{oo}_{g,n}(\theta_0)\|$. Thus, the expectation in (L.7) becomes

$$\begin{aligned}
E[|\varepsilon^o_{i,n}(\lambda^o_0)|^p|y^s_{i,n}=1, X^s_n] &\le C_2 + C_3 E[|\varepsilon^s_{i,n}(\lambda^s_0)|^p| - \varepsilon^s_{i,n}(\lambda^s_0) < S^s_{i\cdot,n}(\lambda^s_0)X^s_n\beta^s_0, X^s_n] \\
&= C_2 + C_3(\Omega^{ss}_{ii,n}(\theta_0))^{p/2} E\left[\left|\frac{\varepsilon^s_{i,n}(\lambda^s_0)}{\sqrt{\Omega^{ss}_{ii,n}(\theta_0)}}\right|^p \middle| -\frac{\varepsilon^s_{i,n}(\lambda^s_0)}{\sqrt{\Omega^{ss}_{ii,n}(\theta_0)}} < \frac{S^s_{i\cdot,n}(\lambda^s_0)X^s_n\beta^s_0}{\sqrt{\Omega^{ss}_{ii,n}(\theta_0)}}, X^s_n\right] \\
&\le C_2 + C_3(\Omega^{ss}_{ii,n}(\theta_0))^{p/2}\left(E\left[\left|\frac{\varepsilon^s_{i,n}(\lambda^s_0)}{\sqrt{\Omega^{ss}_{ii,n}(\theta_0)}}\right|^r \middle| -\frac{\varepsilon^s_{i,n}(\lambda^s_0)}{\sqrt{\Omega^{ss}_{ii,n}(\theta_0)}} < \frac{S^s_{i\cdot,n}(\lambda^s_0)X^s_n\beta^s_0}{\sqrt{\Omega^{ss}_{ii,n}(\theta_0)}}, X^s_n\right]\right)^{p/r} \\
&= C_2 + C_3(\Omega^{ss}_{ii,n}(\theta_0))^{p/2}\vartheta^{p/r}_r(m_{i,n}(\theta_0)) \\
&\le C_2 + C_4\vartheta^{p/r}_r(m_{i,n}(\theta_0))
\end{aligned}$$

for some constant $C_4 > 0$, where $r$ is the smallest integer at least as large as $p$, $m_{i,n}(\theta_0) = S^s_{i\cdot,n}(\lambda^s_0)X^s_n\beta^s_0/\sqrt{\Omega^{ss}_{ii,n}(\theta_0)}$, and $\vartheta_r(m_{i,n}(\theta_0)) \le |m_{i,n}(\theta_0)|^{r-1}\phi(m_{i,n}(\theta_0))/\Phi(m_{i,n}(\theta_0)) + (r-1)\vartheta_{r-2}(m_{i,n}(\theta_0))$ for $r \ge 2$ with $\vartheta_1(m_{i,n}(\theta_0)) \le \phi(m_{i,n}(\theta_0))/\Phi(m_{i,n}(\theta_0)) + C_5$ and $\vartheta_0 = 1$, for some constant $C_5 > 0$. The second inequality is implied by Hölder's inequality, the second equality follows by Lemma J.1, whereas the last inequality follows from the uniform boundedness of $\Omega^{ss}_{ii,n}(\theta_0)$, which is implied by the uniform boundedness of $\|\Omega^{ss}_{g,n}(\theta)\|$ established in Lemma D.1. Consider $r \geq 2$. It follows from (L.6) that

$$\begin{aligned}
E[|\varepsilon^{o}_{i,n}(\lambda^{o}_0)|^p | y^s_{i,n}=1] &\leq C_2 + C_4 E[\vartheta^{p/r}_r(m_{i,n}(\theta_0))|y^s_{i,n}=1] \\
&\leq C_2 + C_4 E\left[|m_{i,n}(\theta_0)|^{r-1}\frac{\phi(m_{i,n}(\theta_0))}{\Phi(m_{i,n}(\theta_0))} + (r-1)\vartheta_{r-2}(m_{i,n}(\theta_0))\Big|y^s_{i,n}=1\right]^{p/r} \\
&\leq C_2 + C_4\left(E\left[|m_{i,n}(\theta_0)|^{r-1}\frac{\phi(m_{i,n}(\theta_0))}{\Phi(m_{i,n}(\theta_0))}\Big|y^s_{i,n}=1\right]^{p/r} + (r-1)^{p/r}E[\vartheta_{r-2}(m_{i,n}(\theta_0))|y^s_{i,n}=1]^{p/r}\right).
\end{aligned} \tag{L.8}$$

From the proof of Lemma A.9 by Xu and Lee (2015), it follows that $\phi(x)/\Phi(x) \leq 2(|x| + C_6)$, for some constant $C_6 > 0$. Hence,

$$\begin{aligned}
E\left[|m_{i,n}(\theta_0)|^{r-1}\frac{\phi(m_{i,n}(\theta_0))}{\Phi(m_{i,n}(\theta_0))}\Big|y^s_{i,n}=1\right]^{p/r} &\leq 2^{p/r}E\left[|m_{i,n}(\theta_0)|^{r-1}(|m_{i,n}(\theta_0)|+C_6)|y^s_{i,n}=1\right]^{p/r} \\
&\leq 2^{p/r}\left(E[|m_{i,n}(\theta_0)|^p|y^s_{i,n}=1] + C_6^{p/r}E[|m_{i,n}(\theta_0)|^{p(r-1)/r}|y^s_{i,n}=1]\right),
\end{aligned}$$

where the last inequality follows by Loève's $c_r$-inequality. In order to show that the first term in (L.8) is uniformly bounded, by Hölder's inequality it is enough to establish that $E[|m_{i,n}(\theta_0)|^p|y^s_{i,n}=1]$ is uniformly bounded. In the same way as in (L.4),

$$\begin{aligned}
E\left[|m_{i,n}(\theta_0)|^p\big|y^s_{i,n}=1\right] &= E\left[\left|\frac{S^s_{i\cdot,n}(\lambda^s_0)X^s_n\beta^s_0}{\sqrt{\Omega^{ss}_{ii,n}(\theta_0)}}\right|^p \big|y^s_{i,n}=1\right] \\
&\leq \sup_{n,i}\Omega^{ss^{-p/2}}_{ii,n}(\theta_0)\|\beta^s_0\|^p \sup_n \|S^s_n(\lambda^s_0)\|^p_\infty \sup_{n,i} E\left[\|X^s_{i\cdot,n}\|^p\big|y^s_{i,n}=1\right] < \infty,
\end{aligned}$$

where the conclusion is implied by Assumptions 1($ii$), 4($ii$), and 7 and Lemma D.1.

It is easy to show using recursion and Hölder's inequality that the second term in (L.8) is uniformly bounded if $E[|m_{i,n}(\theta_0)|^p|y^s_{i,n}=1]$ is uniformly bounded. This condition is sufficient for the case when $r=1$ as well. It completes the proof that $\sup_{\theta\in\Theta}\|z_{g,n}(\theta)\|$ is uniformly $L_p$-bounded.

We continue by showing that $\sup_{\theta\in\Theta}\|\mu^{11}_{g,n}(\theta)\|$ is uniformly $L_p$-bounded:

$$\begin{aligned}
E\left[\sup_{\theta\in\Theta}\|\mu^{11}_{g,n}(\theta)\|\right]^p &= E\left[\sup_{\theta\in\Theta}\|\tilde{\Omega}^{so}_{g,n}(\theta)\Omega^{oo^{-1}}_{g,n}(\theta)z_{g,n}(\theta)\|\right]^p \\
&\leq \sup_{\theta\in\Theta}(\|\Omega^{so}_{g,n}(\theta)\|\|\Omega^{oo^{-1}}_{g,n}(\theta)\|)^p E\left[\sup_{\theta\in\Theta}\|z_{g,n}(\theta)\|\right]^p < \infty
\end{aligned}$$

uniformly in $n \in \mathbb{N}$ and $g \in \mathcal{G}_n$, where the conclusion is implied by Lemma D.1 and the previous results of this proof.

In the same way as in (L.3), $\sup_{\theta\in\Theta} \|v^{11}_{g,n}(\theta)\|$ is uniformly $L_p$-bounded if $\sup_{\theta\in\Theta} |v^{11}_{gj,n}(\theta)|$ is uniformly bounded for $j = 1, 2$:

$$
\begin{aligned}
E\left[\sup_{\theta\in\Theta} |v^{11}_{gj,n}(\theta)|\right]^p &= E\left[\sup_{\theta\in\Theta} \left|\frac{\tilde{S}^s_{gj\cdot,n}(\lambda^s)X^s_n\beta^s - \mu^{11}_{gj,n}(\theta)}{\sqrt{\Sigma^{11}_{gjj,n}(\theta)}}\right|\right]^p \\
&\leq 2^{p-1} \sup_{\theta\in\Theta} \Sigma^{11^{-p/2}}_{gjj,n}(\theta) \left(E\left[\sup_{\theta\in\Theta} |S^s_{gj\cdot,n}(\lambda^s)X^s_n\beta^s|\right]^p + E\left[\sup_{\theta\in\Theta} |\mu^{11}_{gj,n}(\theta)|\right]^p\right) < \infty
\end{aligned}
$$

uniformly in $n \in \mathbb{N}$ and $g \in \mathcal{G}_n$, where the conclusion follows by the previous results of this proof and Lemma D.2. □

***Proof of Lemma D.4***. The proof holds under Assumption 4($ii$) or Assumption 9($ii$). For the sake of simplicity, we will refer to Assumption 4($ii$) only.

Applying Loève's $c_r$-inequality twice leads to the following bound:

$$
\begin{aligned}
E\left[\sup_{\theta\in\Theta} \left\|\frac{\partial z_{g,n}(\theta)}{\partial\theta'}\right\|\right]^p &= E\left[\sup_{\theta\in\Theta} \left\|\frac{\partial(y^o_{g,n} - S^o_{g\cdot,n}(\lambda^o)X^o_n\beta^o)}{\partial\theta'}\right\|\right]^p \\
&= E\left[\sup_{\theta\in\Theta} \left\|\frac{\partial(S^o_{g\cdot,n}(\lambda^o)X^o_n\beta^o)}{\partial\theta'}\right\|\right]^p \\
&= E\left[\sup_{\theta\in\Theta} \sqrt{\left\|\frac{\partial(S^o_{g\cdot,n}(\lambda^o)X^o_n\beta^o)}{\partial\lambda^o}\right\|^2 + \left\|\frac{\partial(S^o_{g\cdot,n}(\lambda^o)X^o_n\beta^o)}{\partial\beta^{o\prime}}\right\|^2}\right]^p \\
&\leq 2^{p-1}\left(E\left[\sup_{\theta\in\Theta} \left\|\frac{\partial(S^o_{g\cdot,n}(\lambda^o)X^o_n\beta^o)}{\partial\lambda^o}\right\|\right]^p + E\left[\sup_{\theta\in\Theta} \left\|\frac{\partial(S^o_{g\cdot,n}(\lambda^o)X^o_n\beta^o)}{\partial\beta^{o\prime}}\right\|\right]^p\right) \\
&= 2^{p-1}\left(E\left[\sup_{\theta\in\Theta} \left\|\frac{\partial S^o_{g\cdot,n}(\lambda^o)}{\partial\lambda^o}X^o_n\beta^o\right\|\right]^p + E\left[\sup_{\theta\in\Theta} \|S^o_{g\cdot,n}(\lambda^o)X^o_n\|\right]^p\right).
\end{aligned} \tag{L.9}
$$

The uniform boundedness of the second term can be proven in the same way as the uniform boundedness of $E[\sup_{\theta\in\Theta} \|S^b_{g\cdot,n}(\lambda^b)X^b_n\beta^b\|]^p$ in Lemma D.3. In the same way as in (L.3) and (L.4), we can show that the first term is uniformly bounded if

$$
\sup_{\theta\in\Theta} \|\beta^o\|^p \sup_n \sup_{\theta\in\Theta} \left\|\frac{\partial S^o_n(\lambda^o)}{\partial\lambda^o}\right\|^p_\infty \sup_{n,i} E\|X^o_{i\cdot,n}\|^p < \infty. \tag{L.10}
$$

The first and the last terms in (L.10) are bounded by Assumptions 7 and 4($ii$), respectively. The second term in (L.10) is uniformly bounded because

$$
\left\|\frac{\partial S^o_n(\lambda^o)}{\partial\lambda^o}\right\|_\infty = \left\|\frac{\partial(I_{2n} - \lambda^o W^o_n)^{-1}}{\partial\lambda^o}\right\|_\infty = \|(I_{2n} - \lambda^o W^o_n)^{-1}W^o_n(I_{2n} - \lambda^o W^o_n)^{-1}\|_\infty < \infty, \tag{L.11}
$$

where the result follows from the sub-multiplicativity of matrix norms and Assumption 1($ii$).

Next, let $M^{11}_{g,n}(\theta) = \mathrm{Diag}(\Sigma^{11}_{g,n}(\theta))^{-1/2}$. Then

$$\begin{aligned}
E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial v^{11}_{g,n}(\theta)}{\partial\theta'}\right\|\right]^p &= E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial(M^{11}_{g,n}(\theta)(q_{g,n}(\theta)-\mu^{11}_{g,n}(\theta)))}{\partial\theta'}\right\|\right]^p \\
&= E\left[\sup_{\theta\in\Theta}\left\|((q_{g,n}(\theta)-\mu^{11}_{g,n}(\theta))'\otimes I_2)\frac{\partial\operatorname{vec}M^{11}_{g,n}(\theta)}{\partial\theta'} + M^{11}_{g,n}(\theta)\left(\frac{\partial q_{g,n}(\theta)}{\partial\theta'} - \frac{\partial\mu^{11}_{g,n}(\theta)}{\partial\theta'}\right)\right\|\right]^p \\
&\le 2^{p-1}\left(\sup_{\theta\in\Theta}\left\|\frac{\partial\operatorname{vec}M^{11}_{g,n}(\theta)}{\partial\theta'}\right\|^p E\left[\sup_{\theta\in\Theta}\|(q_{g,n}(\theta)-\mu^{11}_{g,n}(\theta))'\otimes I_2)\|\right]^p\right. \qquad \text{(L.12)}\\
&\qquad\left. +\sup_{\theta\in\Theta}\|M^{11}_{g,n}(\theta)\|^p E\left[\sup_{\theta\in\Theta}\left(\left\|\frac{\partial q_{g,n}(\theta)}{\partial\theta'}\right\| + \left\|\frac{\partial\mu^{11}_{g,n}(\theta)}{\partial\theta'}\right\|\right)\right]^p\right)
\end{aligned}$$

uniformly in $n\in\mathbb{N}$ and $g\in\mathcal{G}_n$. We have already shown in the proof of Lemma D.2 that the norms of $\partial\operatorname{vec}M^{11}_{g,n}(\theta)/\partial\theta'$ and $M^{11}_{g,n}(\theta)$ are uniformly bounded. By Lemma C.7,

$$\begin{aligned}
E\left[\sup_{\theta\in\Theta}\|(q_{g,n}(\theta)-\mu^{11}_{g,n}(\theta))'\otimes I_2\|\right]^p &= E\left[\sup_{\theta\in\Theta}\|q_{g,n}(\theta)-\mu^{11}_{g,n}(\theta)\|\|I_2\|\right]^p \\
&\le 2^{3p/2-1}\left(E\left[\sup_{\theta\in\Theta}\|q_{g,n}(\theta)\|\right]^p + E\left[\sup_{\theta\in\Theta}\|\mu^{11}_{g,n}(\theta)\|\right]^p\right) \\
&< \infty
\end{aligned}$$

uniformly in $n\in\mathbb{N}$ and $g\in\mathcal{G}_n$ by Lemma D.3 because $\|q_{g,n}(\theta)\| = \|\tilde{S}^s_{g\cdot,n}(\lambda^s)X^s_n\beta^s\| = \|S^s_{g\cdot,n}(\lambda^s)X^s_n\beta^s\|$. It remains to show that the last term in (L.12) is uniformly bounded. By Loève's $c_r$-inequality,

$$E\left[\sup_{\theta\in\Theta}\left(\left\|\frac{\partial q_{g,n}(\theta)}{\partial\theta'}\right\| + \left\|\frac{\partial\mu^{11}_{g,n}(\theta)}{\partial\theta'}\right\|\right)\right]^p \le 2^{p-1}\left(E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial q_{g,n}(\theta)}{\partial\theta'}\right\|\right]^p + E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial\mu^{11}_{g,n}(\theta)}{\partial\theta'}\right\|\right]^p\right).$$

We can show that the first term is uniformly bounded in the same way as we proved earlier in this proof that $\sup_{n,g} E[\sup_{\theta\in\Theta}\|\partial(S^o_{g\cdot,n}(\lambda^o)X^o_n\beta^o)/\partial\theta'\|]^p < \infty$. For the second term, note that

$$\begin{aligned}
E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial\mu^{11}_{g,n}(\theta)}{\partial\theta'}\right\|\right]^p &= E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial(\tilde{\Omega}^{so}_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)z_{g,n}(\theta))}{\partial\theta'}\right\|\right]^p \\
&= E\left[\sup_{\theta\in\Theta}\left\|(z'_{g,n}(\theta)\otimes I_2)\frac{\partial\operatorname{vec}(\tilde{\Omega}^{so}_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta))}{\partial\theta'} + \tilde{\Omega}^{so}_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)\frac{\partial z_{g,n}(\theta)}{\partial\theta'}\right\|\right]^p \\
&\le 2^{p-1}\left(\sup_{\theta\in\Theta}\left\|\frac{\partial\operatorname{vec}(\tilde{\Omega}^{so}_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta))}{\partial\theta'}\right\|^p 2^{p/2}E\left[\sup_{\theta\in\Theta}\|z_{g,n}(\theta)\|\right]^p\right. \\
&\qquad\left. +\sup_{\theta\in\Theta}(\|\Omega^{so}_{g,n}(\theta)\|\|\Omega^{oo-1}_{g,n}(\theta)\|)^p E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial z_{g,n}(\theta)}{\partial\theta'}\right\|\right]^p\right)
\end{aligned}$$

is uniformly bounded. The conclusion then follows by Lemmas D.1 and D.3, the previous results of this lemma, and by noticing that

$$\begin{aligned}\left\|\frac{\partial \operatorname{vec}(\tilde{\Omega}^{so}_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta))}{\partial\theta'}\right\| &= \left\|(\Omega^{oo-1}_{g,n}(\theta)\otimes I_2)\frac{\partial \operatorname{vec}\tilde{\Omega}^{so}_{g,n}(\theta)}{\partial\theta'} + (I_2\otimes\tilde{\Omega}^{so}_{g,n}(\theta))\frac{\partial \operatorname{vec}\Omega^{oo-1}_{g,n}(\theta)}{\partial\theta'}\right\| \\ &\le \sqrt{2}\left(\|\Omega^{oo-1}_{g,n}(\theta)\|\left\|\frac{\partial \operatorname{vec}\Omega^{so}_{g,n}(\theta)}{\partial\theta'}\right\| + \|\Omega^{so}_{g,n}(\theta)\|\left\|\frac{\partial \operatorname{vec}\Omega^{oo-1}_{g,n}(\theta)}{\partial\theta'}\right\|\right) \\ &< \infty\end{aligned}$$

uniformly in $n \in \mathbb{N}$ and $g \in \mathcal{G}_n$ by Lemma D.1. □

***Proof of Lemma D.5***. We start with establishing the uniform $L_p$-NED, $p \in \{2,4\}$, property for $\{y^{*b}_{g,n}\}_{g\in\mathcal{G}_n}$, $b \in \{s,o\}$, which will be needed later in the proof. Since the bounds derived in this proof are uniform on $\Theta$, we write for the sake of simplicity $\theta$ instead of $\theta_n$. Using now the definition of NED and the conditional Jensen's inequality, it follows

$$\begin{aligned}\|y^{*b}_{g,n} - E[y^{*b}_{g,n}|\mathcal{F}_{g,n}(s)]\|_p &= \|S^b_{g\cdot,n}(\lambda^b_0)\left(X^b_n\beta^b_0 + u^b_n - E[X^b_n\beta^b_0 + u^b_n|\mathcal{F}_{g,n}(s)]\right)\|_p \\ &= \|\sum_{\dot{g}\in\mathcal{G}_n} S^b_{g\dot{g},n}(\lambda^b_0)\left(X^b_{\dot{g}\cdot,n}\beta^b_0 + u^b_{\dot{g},n} - E[X^b_{\dot{g}\cdot,n}\beta^b_0 + u^b_{\dot{g},n}|\mathcal{F}_{g,n}(s)]\right)\|_p \\ &= \|\sum_{\dot{g}:d(g,\dot{g})>s} S^b_{g\dot{g},n}(\lambda^b_0)\left(X^b_{\dot{g}\cdot,n}\beta^b_0 + u^b_{\dot{g},n} - E[X^b_{\dot{g}\cdot,n}\beta^b_0 + u^b_{\dot{g},n}|\mathcal{F}_{g,n}(s)]\right)\|_p \\ &\le \sum_{\dot{g}:d(g,\dot{g})>s} \|S^b_{g\dot{g},n}(\lambda^b_0)\|\left(\|X^b_{\dot{g}\cdot,n}\beta^b_0 + u^b_{\dot{g},n}\|_p + \|E[X^b_{\dot{g}\cdot,n}\beta^b_0 + u^b_{\dot{g},n}|\mathcal{F}_{g,n}(s)]\|_p\right) \\ &\le 2\sum_{\dot{g}:d(g,\dot{g})>s} \|S^b_{g\dot{g},n}(\lambda^b_0)\|\left(\|X^b_{\dot{g}\cdot,n}\|_p\|\beta^b_0\| + \|u^b_{\dot{g},n}\|_p\right) \\ &\le 2\sup_{n,g}\sup_{\theta\in\Theta}\sum_{\dot{g}\in\mathcal{G}_n}\|S^b_{g\dot{g},n}(\lambda^b)\|\cdot\left(\sup_{n,g}\|X^b_{g\cdot,n}\|_p\|\beta^b_0\| + \sup_{n,g}\|u^b_{g,n}\|_p\right) \\ &\quad \times \frac{\sup_{n,g}\sup_{\theta\in\Theta}\sum_{\dot{g}:d(g,\dot{g})>s}\|S^b_{g\dot{g},n}(\lambda^b)\|}{\sup_{n,g}\sup_{\theta\in\Theta}\sum_{\dot{g}\in\mathcal{G}_n}\|S^b_{g\dot{g},n}(\lambda^b)\|} \\ &\le t^{y^{*b}}\psi^b(s),\end{aligned}$$

where $t^{y^{*b}} = 2\sup_{n,g}\sup_{\theta\in\Theta}\sum_{\dot{g}\in\mathcal{G}_n}\|S^b_{g\dot{g},n}(\lambda^b)\|\left(\sup_{n,g}\|X^b_{g\cdot,n}\|_p\|\beta^b_0\| + \sup_{n,g}\|u^b_{g,n}\|_p\right)$ and $\psi^b(s) \le 1$ for $b \in \{s,o\}$ with $\psi^b(s) = \sup_{n,g}\sup_{\theta\in\Theta}\sum_{\dot{g}:d(g,\dot{g})>s}\|S^b_{g\dot{g},n}(\lambda^b)\|/\sup_{n,g}\sup_{\theta\in\Theta}\sum_{\dot{g}\in\mathcal{G}_n}\|S^b_{g\dot{g},n}(\lambda^b)\|$, $b \in \{s,o\}$. The first and second inequalities follow by Minkowski's and the conditional Jensen's inequalities, respectively. Given Assumption 7, $t^{y^{*b}}$ is bounded provided that $E\|X^b_{g\cdot,n}\|^p$, $E\|u^b_{g,n}\|^p$, and $\sum_{\dot{g}\in\mathcal{G}_n}\|S^b_{g\dot{g},n}(\lambda^b)\|$ are uniformly bounded. Since $p \in \{2,4\}$, by Liapunov's inequality it is enough to establish the results for $p = 4$. Notice that $E\|X^b_{g\cdot,n}\|^4 = E\left[\|X^b_{g1\cdot,n}\|^2 + \|X^b_{g2\cdot,n}\|^2\right]^2 \leq \sup_{n,i} 4E\|X^b_{i\cdot,n}\|^4 < \infty$ by Assumption 4($ii$). Because a normal distribution has infinitely many moments and $\sup_{n,g} E\|u^b_{g,n}\|^4 \leq \sup_{n,i} 4E|u^b_{i,n}|^4$, Assumption 2($i$) implies that $\sup_{n,g} E\|u^b_{g,n}\|^4 < \infty$, whereas equivalence of matrix norms on finite dimensional matrix spaces implies that uniformly in $n \in \mathbb{N}$, $g \in \mathcal{G}_n$, and $\theta \in \Theta$

$$\begin{aligned}\sum_{\dot{g}\in\mathcal{G}_n} \|S^b_{g\dot{g},n}(\lambda^b)\| &\leq C_1 \sum_{\dot{g}\in\mathcal{G}_n} \|S^b_{g\dot{g},n}(\lambda^b)\|_\infty \leq C_1 \sum_{\dot{g}\in\mathcal{G}_n} \left(\|S^b_{g1\dot{g},n}(\lambda^b)\|_\infty + \|S^b_{g2\dot{g},n}(\lambda^b)\|_\infty\right) \\ &\leq 2C_1\|S^b_{g\cdot,n}(\lambda^b)\|_\infty \leq 2C_1\|S^b_n(\lambda^b)\|_\infty < \infty\end{aligned} \tag{L.13}$$

for some constant $C_1 > 0$ by Assumption 1($ii$). Note that by Assumption 5, $\lim_{s\to\infty} \psi^b(s) = 0$. Thus, $\{y^{*b}_{g,n}\}_{g\in\mathcal{G}_n}$ is a uniform $L_4$- and $L_2$-NED random field with NED coefficients $\psi^b(s)$.

Recall that $d^{11}_{g,n} = \mathbb{1}(y^s_{g1,n} = 1, y^s_{g2,n} = 1) = \mathbb{1}(y^{*s}_{g1,n} > 0) \cdot \mathbb{1}(y^{*s}_{g2,n} > 0)$. From the proof of Proposition 2 by Xu and Lee (2015), it follows that for some constants $C_2, C_3 > 0$ and $j = 1, 2$,

$$\|\mathbb{1}(y^{*s}_{gj,n} > 0) - E[\mathbb{1}(y^{*s}_{gj,n} > 0)|\mathcal{F}_{g,n}(s)]\|_2 \leq (1 + C_2)\|y^{*s}_{gj,n} - E[y^{*s}_{gj,n}|\mathcal{F}_{g,n}(s)]\|_2^{1/3} \leq (1 + C_2)C_3\psi^{1/3}(s),$$

where the last inequality follows by Lemma J.2 and the fact that $\{y^{*s}_{g,n}\}_{g\in\mathcal{G}_n}$ is uniformly $L_2$-NED with NED coefficients $\psi^s(s)$. Since $|\mathbb{1}(y^{*s}_{gj,n} > 0) - E[\mathbb{1}(y^{*s}_{gj,n} > 0)|\mathcal{F}_{g,n}(s)]|^4 \leq |\mathbb{1}(y^{*s}_{gj,n} > 0) - E[\mathbb{1}(y^{*s}_{gj,n} > 0)|\mathcal{F}_{g,n}(s)]|^2$ implies that $\|\mathbb{1}(y^{*s}_{gj,n} > 0) - E[\mathbb{1}(y^{*s}_{gj,n} > 0)|\mathcal{F}_{g,n}(s)]\|_4 \leq \|\mathbb{1}(y^{*s}_{gj,n} > 0) - E[\mathbb{1}(y^{*s}_{gj,n} > 0)|\mathcal{F}_{g,n}(s)]\|_2^{1/2}$, $\{\mathbb{1}(y^{*s}_{gj,n} > 0)\}_{g\in\mathcal{G}_n}$ is a uniform $L_4$-NED random field with NED coefficients $[\psi^s(s)]^{1/6}$. Given that $\{\mathbb{1}(y^{*s}_{gj,n} > 0)\}_{g\in\mathcal{G}_n}$ is uniformly $L_4$-bounded, Lemma C.9 implies that $\{d^{11}_{g,n}\}_{g\in\mathcal{G}_n}$ is a uniform $L_2$-NED random field with NED coefficients $\psi^s(s)^{1/6} + \psi^s(s)^{1/6} + \psi^s(s)^{1/3} \leq 3[\psi^s(s)]^{1/6}$.[1]

From the definition, $z_{g,n}(\theta) = y^o_{g,n} - S^o_{g\cdot,n}(\lambda^o)X^o_n\beta^o$, and we will now establish the uniform NED property for each term of this summation and find their NED coefficients. Note that $y^o_{g,n} = \left(\mathbb{1}(y^{*s}_{g1,n} > 0)y^{*o}_{g1,n}\ ,\ \mathbb{1}(y^{*s}_{g2,n} > 0)y^{*o}_{g2,n}\right)'$. Since by Lemma J.2 and the previous results of this proof, $\{y^{*o}_{gj,n}\}_{g\in\mathcal{G}_n}$ and $\mathbb{1}(y^{*s}_{gj,n} > 0)$, $j = 1, 2$, are uniformly $L_4$-NED with NED coefficients $\psi^o(s)$ and $[\psi^s(s)]^{1/6}$, respectively, Lemma C.9 implies that $\{\mathbb{1}(y^{*s}_{gj,n} > 0)y^{*o}_{gj,n}\}_{g\in\mathcal{G}_n}$ is a uniform $L_2$-NED random field with NED coefficients $[\psi^s(s)]^{1/6} + \psi^o(s) + \psi^o(s)[\psi^s(s)]^{1/6} \leq 3\psi^{1/6}(s)$. By Lemma J.2, the same property is transfered to $\{y^o_{g,n}\}_{g\in\mathcal{G}_n}$. It is easy to see from the proof of $\{y^{*o}_{g,n}\}_{g\in\mathcal{G}_n}$ being an $L_2$-NED random field that $\{S^o_{g\cdot,n}(\lambda^o)X^o_n\beta^o\}_{g\in\mathcal{G}_n}$ is a uniform $L_2$-NED random field with NED coefficients $\psi^o(s)$. Hence, $\{z_{g,n}(\theta)\}_{g\in\mathcal{G}_n}$ is a uniform $L_2$-NED random field with NED coefficients $[\psi^o(s)]^{1/6} + \psi^o(s) \leq 2[\psi^o(s)]^{1/6}$.

Finally, since $v^{11}_{g,n}(\theta) = \mathrm{Diag}(\Sigma^{11}_{g,n}(\theta))^{-1/2}(q_{g,n}(\theta) - \mu^{11}_{g,n}(\theta))$ and $\|\mathrm{Diag}(\Sigma^{11}_{g,n}(\theta))^{-1/2}\|$ is uniformly bounded by Lemma D.2, it suffices to establish the uniform NED property for $q_{g,n}(\theta)$ and $\mu^{11}_{g,n}(\theta)$ and find their NED

[1] Note that in this case we can treat $[\psi^s(s)]^{1/6}$ as the NED coefficient because 3 can be treated as a part of the NED scaling factor.

coefficients. Recall that $q_{g,n}(\theta) = \tilde{S}^s_{g\cdot,n}(\lambda^s)X^s_n\beta^s$, thus it is a uniform $L_2$-NED random field with NED coefficients $\psi^s(s)$. Moreover, $\mu^{11}_{g,n}(\theta) = \tilde{\Omega}^{so}_{g,n}(\theta)\Omega^{oo-1}_{g,n}(\theta)z_{g,n}(\theta)$. Given that the norms of $\tilde{\Omega}^{so}_{g,n}(\theta)$ and $\Omega^{oo-1}_{g,n}(\theta)$ are uniformly bounded by Lemma D.1, $\{\mu^{11}_{g,n}(\theta)\}_{g\in\mathcal{G}_n}$ is a uniform $L_2$-NED random field with NED coefficients $[\psi^o(s)]^{1/6}$. The conclusion follows by setting $\psi(s) = \max\{\psi^s(s), \psi^o(s)\}$. $\square$

***Proof of Lemma D.6***. Let us denote $\theta_0 = (\beta^{s'}_0, \beta^{o'}_0, \lambda^s_0, \lambda^o_0, \rho_0, \sigma^2_0)'$ and $\theta_n = (\beta^{s'}_n, \beta^{o'}_n, \lambda^s_n, \lambda^o_n, \rho_n, \sigma^2_n)'$. Then by the elementwise mean value theorem, there exists $\xi_n$ with elements between $\lambda^s_n$ and $\lambda^s_0$ such that[2]

$$\begin{aligned}
\hat{y}^{*s}_{g,n} - y^{*s}_{g,n} &= S^s_{g\cdot,n}(\lambda^s_n)\left(X^s_n\beta^s_n + u^s_n\right) - S^s_{g\cdot,n}(\lambda^s_0)\left(X^s_n\beta^s_0 + u^s_n\right) \\
&= S^s_{g\cdot,n}(\lambda^s_n)X^s_n(\beta^s_n - \beta^s_0) + \left(S^s_{g\cdot,n}(\lambda^s_n) - S^s_{g\cdot,n}(\lambda^s_0)\right)\left(X^s_n\beta^s_0 + u^s_n\right) \\
&= S^s_{g\cdot,n}(\lambda^s_n)X^s_n(\beta^s_n - \beta^s_0) + \frac{\partial S^s_{g\cdot,n}(\xi_n)}{\partial\lambda}(\lambda^s_n - \lambda^s_0)\left(X^s_n\beta^s_0 + u^s_n\right) \\
&= \sum_{\dot{g}\in\mathcal{G}_n} S^s_{g\dot{g},n}(\lambda^s_n)X^s_{\dot{g}\cdot,n}(\beta^s_n - \beta^s_0) + \sum_{\dot{g}\in\mathcal{G}_n}\frac{\partial S^s_{g\dot{g},n}(\xi_n)}{\partial\lambda}(\lambda^s_n - \lambda^s_0)\left(X^s_{\dot{g}\cdot,n}\beta^s_0 + u^s_{\dot{g},n}\right).
\end{aligned} \tag{L.14}$$

Recall that $\frac{\partial S^s_n(\xi_n)}{\partial\lambda} = S^s_n(\xi_n)W^s_nS^s_n(\xi_n)$, see (L.11), which together with $S^s_n(\lambda^s_n)$ has row and column sums bounded uniformly in $n$, $\xi_n$, and $\lambda^s$ by Assumption 1$(ii)$. As in (L.13),

$$\begin{aligned}
\sum_{\dot{g}\in\mathcal{G}_n}\left\|\frac{\partial S^s_{g\dot{g},n}(\xi_n)}{\partial\lambda}\right\| &\le C_1\sum_{\dot{g}\in\mathcal{G}_n}\left\|\frac{\partial S^s_{g\dot{g},n}(\xi_n)}{\partial\lambda}\right\|_\infty \le C_1\sum_{\dot{g}\in\mathcal{G}_n}\left(\left\|\frac{\partial S^s_{g1\dot{g},n}(\xi_n)}{\partial\lambda}\right\|_\infty + \left\|\frac{\partial S^s_{g2\dot{g},n}(\xi_n)}{\partial\lambda}\right\|_\infty\right) \\
&\le 2C_1\left\|\frac{\partial S^s_{g\cdot,n}(\xi_n)}{\partial\lambda}\right\|_\infty \le 2C_1\left\|\frac{\partial S^s_n(\xi_n)}{\partial\lambda}\right\|_\infty < \infty
\end{aligned} \tag{L.15}$$

for some constant $C_1 > 0$ uniformly in $\xi_n, g$, and $n$. We have also verified in the proof of Lemma D.5 that $\sup_{n,g} E\|X^s_{g\cdot,n}\|^4 < \infty$ and $\sup_{n,g} E\|u^s_{g,n}\|^4 < \infty$. Hence as $\xi_n \in [\min\{\lambda^s_0, \lambda^s_n\}, \max\{\lambda^s_0, \lambda^s_n\}]$,

$$\begin{aligned}
\|\hat{y}^{*s}_{g,n} - y^{*s}_{g,n}\|_2 &\le \|\beta^s_n - \beta^s_0\|\sum_{\dot{g}\in\mathcal{G}_n}\|S^s_{g\dot{g},n}(\lambda^s_n)\|\|X^s_{\dot{g}\cdot,n}\|_2 + |\lambda^s_n - \lambda^s_0|\sum_{\dot{g}\in\mathcal{G}_n}\left\|\frac{\partial S^s_{g\dot{g},n}(\xi_n)}{\partial\lambda}\right\|\left(\|X^s_{\dot{g}\cdot,n}\|_2\|\beta^s_0\| + \|u^s_{\dot{g},n}\|_2\right) \\
&\le \|\beta^s_n - \beta^s_0\|\sup_{n,g}\|X^s_{g\cdot,n}\|\sup_{n,g}\sup_{\theta\in\Theta}\sum_{\dot{g}\in\mathcal{G}_n}\|S^s_{g\dot{g},n}(\lambda^s)\| \\
&\quad + |\lambda^s_n - \lambda^s_0|\left(\sup_{n,g}\|X^s_{g\cdot,n}\|_2\|\beta^s_0\| + \sup_{n,g}\|u^s_{g,n}\|_2\right)\sup_{n,g}\sup_{\theta\in\Theta}\sum_{\dot{g}\in\mathcal{G}_n}\left\|\frac{\partial S^s_{g\dot{g},n}(\lambda^s)}{\partial\lambda}\right\| = O(n^{-1/2})
\end{aligned}$$

as $n \to \infty$.

Next, we establish the $L_2$-NED property for $\{\hat{y}^{*s}_{g,n} - y^{*s}_{g,n}\}_{g\in\mathcal{G}_n}$. Using the definition of NED and the conditional

[2] The mean value theorem is applied for each element of $S^s_{g\cdot,n}(\lambda^s_n) - S^s_{g\cdot,n}(\lambda^s_0)$ separately, and therefore, $\xi_n$ might differ for each element of $\partial S^s_{g\cdot,n}/\partial\lambda$.

Jensen's and Minkowski's inequalities, it follows from (L.14) that

$$
\begin{aligned}
&\|\hat{y}^{*s}_{g,n} - y^{*s}_{g,n} - E[\hat{y}^{*s}_{g,n} - y^{*s}_{g,n}|\mathcal{F}_{g,n}(s)]\|_2 \\
&\leq \left\| \sum_{\dot{g}\in\mathcal{G}_n} S^s_{g\dot{g},n}(\lambda^s_n)\left(X^s_{\dot{g}\cdot,n}(\beta^s_n - \beta^s_0) - E[X^s_{\dot{g}\cdot,n}(\beta^s_n - \beta^s_0)|\mathcal{F}_{g,n}(s)]\right)\right\|_2 \\
&+ \left\| \sum_{\dot{g}\in\mathcal{G}_n} \frac{\partial S^s_{g\dot{g},n}(\xi_n)}{\partial\lambda}(\lambda^s_n - \lambda^s_0)\left(X^s_{\dot{g}\cdot,n}\beta^s_0 + u^s_{\dot{g},n} - E[X^s_{\dot{g}\cdot,n}\beta^s_0 + u^s_{\dot{g},n}|\mathcal{F}_{g,n}(s)]\right)\right\|_2 \\
&= \left\| \sum_{\dot{g}:d(g,\dot{g})>s} S^s_{g\dot{g},n}(\lambda^s_n)\left(X^s_{\dot{g}\cdot,n}(\beta^s_n - \beta^s_0) - E[X^s_{\dot{g}\cdot,n}(\beta^s_n - \beta^s_0)|\mathcal{F}_{g,n}(s)]\right)\right\|_2 \\
&+ \left\| \sum_{\dot{g}:d(g,\dot{g})>s} \frac{\partial S^s_{g\dot{g},n}(\xi_n)}{\partial\lambda}(\lambda^s_n - \lambda^s_0)\left(X^s_{\dot{g}\cdot,n}\beta^s_0 + u^s_{\dot{g},n} - E[X^s_{\dot{g}\cdot,n}\beta^s_0 + u^s_{\dot{g},n}|\mathcal{F}_{g,n}(s)]\right)\right\|_2 \\
&\leq \|\beta^s_n - \beta^s_0\| \sum_{\dot{g}:d(g,\dot{g})>s} \|S^s_{g\dot{g},n}(\lambda^s_n)\| \left(\|X^s_{\dot{g}\cdot,n}\|_2 + \|E[X^s_{\dot{g}\cdot,n}|\mathcal{F}_{g,n}(s)]\|_2\right) \\
&+ |\lambda^s_n - \lambda^s_0| \sum_{\dot{g}:d(g,\dot{g})>s} \left\|\frac{\partial S^s_{g\dot{g},n}(\xi_n)}{\partial\lambda}\right\| \left(\|X^s_{\dot{g}\cdot,n}\beta^s_0 + u^s_{\dot{g},n}\|_2 + \|E[X^s_{\dot{g}\cdot,n}\beta^s_0 + u^s_{\dot{g},n}|\mathcal{F}_{g,n}(s)]\|_2\right) \\
&\leq 2\|\beta^s_n - \beta^s_0\| \sum_{\dot{g}:d(g,\dot{g})>s} \|S^s_{g\dot{g},n}(\lambda^s_n)\|\|X^s_{\dot{g}\cdot,n}\|_2 + 2|\lambda^s_n - \lambda^s_0| \sum_{\dot{g}:d(g,\dot{g})>s} \left\|\frac{\partial S^s_{g\dot{g},n}(\xi_n)}{\partial\lambda}\right\| \left(\|X^s_{\dot{g}\cdot,n}\|_2\|\beta^s_0\| + \|u^s_{\dot{g},n}\|_2\right) \\
&\leq 2\|\beta^s_n - \beta^s_0\| \sup_{n,g}\sup_{\theta\in\Theta} \sum_{\dot{g}\in\mathcal{G}_n} \|S^s_{g\dot{g},n}(\lambda^s)\| \cdot \sup_{n,g}\|X^s_{g\cdot,n}\|_2 \times \frac{\sup_{n,g}\sup_{\theta\in\Theta}\sum_{\dot{g}:d(g,\dot{g})>s}\|S^s_{g\dot{g},n}(\lambda^s)\|}{\sup_{n,g}\sup_{\theta\in\Theta}\sum_{\dot{g}\in\mathcal{G}_n}\|S^s_{g\dot{g},n}(\lambda^s)\|} \\
&+ 2|\lambda^s_n - \lambda^s_0| \sup_{n,g}\sup_{\theta\in\Theta} \sum_{\dot{g}\in\mathcal{G}_n} \left\|\frac{\partial S^s_{g\dot{g},n}(\lambda^s)}{\partial\lambda}\right\| \cdot \left(\sup_{n,g}\|X^s_{g\cdot,n}\|_2\|\beta^s_0\| + \sup_{n,g}\|u^s_{g,n}\|_2\right) \times \frac{\sup_{n,g}\sup_{\theta\in\Theta}\sum_{\dot{g}:d(g,\dot{g})>s}\left\|\frac{\partial S^s_{g\dot{g},n}(\lambda^s)}{\partial\lambda}\right\|}{\sup_{n,g}\sup_{\theta\in\Theta}\sum_{\dot{g}\in\mathcal{G}_n}\left\|\frac{\partial S^s_{g\dot{g},n}(\lambda^s)}{\partial\lambda}\right\|} \\
&\leq O(n^{-1/2})t^{\hat{y}^{*s}-y^{*s}} \max\{\psi^s(s), \tilde{\psi}^s(s)\},
\end{aligned}
$$

where $t^{\hat{y}^{*s}-y^{*s}} = 2\sup_{n,g}\sup_{\theta\in\Theta}\sum_{\dot{g}\in\mathcal{G}_n}\left[\left\|\frac{\partial S^s_{g\dot{g},n}(\lambda^s)}{\partial\lambda}\right\| + \|S^s_{g\dot{g},n}(\lambda^s)\|\right]\left(\sup_{n,g}\|X^s_{g\cdot,n}\|_2(1+\|\beta^s_0\|) + \sup_{n,g}\|u^s_{g,n}\|_2\right)$. Since we have shown that $t^{\hat{y}^{*s}-y^{*s}}$ is finite in the first part of the proof and $\frac{\partial S^s_n(\lambda^s)}{\partial\lambda} = S^s_n(\lambda^s)W^s_nS^s_n(\lambda^s)$, see (L.11), the $L_2$-NED property of $\sqrt{n}(\hat{y}^{*s}_{g,n} - y^{*s}_{g,n})$ follows from Assumptions 5 and 13. □

# References


Abadir, K. and J. Magnus (2005). *Matrix Algebra.* Cambridge University Press.

Davidson, J. (1994). *Stochastic Limit Theory.* Oxford University Press.

Greene, W. (2008). *Econometric Analysis*. Prentice-Hall.

Hashorva, E. and J. Hüsler (2003). On multivariate Gaussian tails. *Annals of the Institute of Statistical Mathematics 55*(3), 507–522.

Kamat, A. R. (1953). Incomplete and absolute moments of the multivariate normal distribution with some applications. *Biometrika 40*(1), 20–34.

LeSage, J. and R. Pace (2009). *Introduction to Spatial Econometrics*. CRC Press.

Magnus, J. and H. Neudecker (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons.

Muthén, B. (1990). Moments of the censored and truncated bivariate normal distribution. *British Journal of Mathematical and Statistical Psychology 43*(1), 131–143.

Qu, X. and L.-F. Lee (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics 184*(2), 209–232.

Xu, X. and L.-F. Lee (2015). Maximum likelihood estimation of a spatial autoregressive Tobit model. *Journal of Econometrics 188*(1), 264–280.