Article

Strategic Interaction Model with Censored Strategies

Nazgul Jenish

Department of Economics, New York University, 19 West 4th Street, New York, NY 10012, USA;
E-Mail: nazgul.jenish@nyu.edu; Tel.: +1-212-998-3891

Academic Editor: Giuseppe Arbia

Abstract: In this paper, we develop a new model of a static game of incomplete information with a large number of players. The model has two key distinguishing features. First, the strategies are subject to threshold effects, and can be interpreted as dependent censored random variables. Second, in contrast to most of the existing literature, our inferential theory relies on a large number of players, rather than a large number of independent repetitions of the same game. We establish existence and uniqueness of the pure strategy equilibrium, and prove that the censored equilibrium strategies satisfy a near-epoch dependence property. We then show that the normal maximum likelihood and least squares estimators of this censored model are consistent and asymptotically normal. Our model can be useful in a wide variety of settings, including investment, R&D, labor supply, and social interaction applications.

Keywords: tobit model; static incomplete information games; near-epoch dependent spatial processes

JEL classifications: C13, C14, C21

1. Introduction

Identification and estimation of strategic interaction models have recently received a great deal of attention in econometrics, owing to the growing interest and application of stochastic games in various fields including industrial organization, labor, political and international economics. Most of the existing literature has focused on discrete choice games, see [1,2] for a survey of recent results. In this literature, the observed data are assumed to arise from an equilibrium of a game played by a finite number of players, and therefore, to be correlated across players. Typically, the number of players is assumed to

be fixed, and the asymptotic inferential theory relies on a large number of independent repetitions of the same game in different markets or in a single market at different points of time. Two notable exceptions are Menzel [3] and Xu [4], who develop the inferential theory of discrete choice games based on a large number of players.

In this paper, we develop a new model of a static game of incomplete information with a large number of players. The model has two key distinguishing features. First, the strategies are subject to threshold effects, and can be interpreted as dependent censored random variables, e.g., R&D investment and labor supply. Second, the game is played in a single market and is not repeated over time. To develop the asymptotic theory, we instead assume that the number of players grows unboundedly, and the players reside on an exogenously given lattice so that the vector of their choices and characteristics can be viewed as a dependent random field, which can be handled by the limit theorems for near-epoch dependent (NED) random fields established by Jenish and Prucha [5].

We derive this model explicitly in two game-theoretical applications: (i) R&D investment by firms under strategic complementarities; and (ii) labor supply decision by women under peer effects. The set-up is standard for a static game of incomplete information: each player's payoff function depends on her choice and choices of other players, her commonly observed characteristics, and her private characteristic unobserved by other players; players move simultaneously based on their expectations about the choices of other players, and in equilibrium, players have self-consistent expectations, see [6], *i.e.*, their subjective expectations coincide with the expectation based on the equilibrium distribution of strategies conditional on commonly observed variables. We assume that private characteristics are i.i.d. normal across players, and prove existence and uniqueness of the pure strategy equilibrium under some mild conditions. We then show that the censored equilibrium strategies also satisfy the NED property under the same conditions.

Under normality of private shocks, the equilibrium strategies boil down to a Tobit econometric model. However, in contrast to the standard Tobit model, our censored model involves a non-zero threshold parameter that needs to be estimated. Therefore, we use the following two-step semiparametric procedure: we first estimate the threshold by the minimum order statistic of the uncensored subsample, and then estimate the remaining parameters either by the maximum likelihood or least squares method. Unlike the standard Tobit model, the maximum likelihood estimator does not strictly dominate the least squares estimator in our model due to discontinuous dependence of the likelihood on the first-step estimator, which may amplify finite-sample biases stemming from the first-step estimation. This provides a rationale for considering the least squares estimation as an alternative to the maximum likelihood procedure. We establish consistency and asymptotic distributions of these estimators. The minimum order statistic is n-consistent and asymptotically exponentially distributed, while the maximum likelihood and least squares estimators are $\sqrt{n}$-consistent and asymptotically normal. A Monte Carlo study suggests that all these estimators perform well in finite samples.

Finally, we address the computational challenges of our game with a large number of players. The standard estimation of games involves computing the equilibrium for each alternative parameter value and then optimizing the objective function over parameter values, and thus presents a formidable computational burden. To tackle it, we use the constrained optimization algorithm proposed by Su and Judd [7], which treats the equilibrium equations as constraints and optimizes simultaneously over

parameters and equilibrium variables, thereby avoiding calculation of the equilibrium at each iteration on the parameter value. Su and Judd [7] show equivalence of this constrained optimization problem to the original problem. Our simulations confirm viability and a significant computational efficiency of the Su-Judd algorithm in our model.

To our knowledge, the proposed censored model has not been considered in the existing literature. Most of the existing results have dealt with discrete choice games, e.g., [8]. Recently, Xu and Lee [9] analyzed a spatially autoregressive (SAR) Tobit model, which can be viewed as a censored version the Cliff-Ord type linear SAR model with a known spatial weight matrix. Xu and Lee [9] establish the NED property as well as consistency and asymptotic normality of the maximum likelihood estimator using the limit theorems of Jenish and Prucha [5]. Though not explicitly demonstrated, this SAR Tobit model can be interpreted as an equilibrium of a static game of complete information, while our model is a game of incomplete information with a different concept of equilibrium and, consequently, qualitatively different implications. Moreover, the presence of latent endogenous variables and non-zero threshold in our model pose additional statistical and computational difficulties. Thus, the two papers are complementary to each other.

The paper is organized as follows. Section 2 describes and derives the model in two examples. Section 3 establishes existence and uniqueness of the equilibrium, and proves the NED property of the equilibrium strategies. Section 4 discusses identification and estimation of the model. Consistency and asymptotic distributions of the estimators are established in Section 5. Section 6 contains a Monte Carlo study, and Section 7 concludes. All proofs are collected in the appendices.

2. Model

In this paper, we are concerned with estimation of the following econometric model:

$$
Y_{in} = \begin{cases} b_0 + X_i'\beta_0 + \sum_{j \in \mathcal{N}_i} \alpha_{j0} E_i(Y_{jn}) + \varepsilon_i, & b_0 + X_i'\beta_0 + \sum_{j \in \mathcal{N}_i} \alpha_{j0} E_i(Y_{jn}) + \varepsilon_i > \gamma_0 \\ 0, & \text{otherwise} \end{cases} \tag{1}
$$

where Y_{in}, $i = 1, ..., n$, is the choice of agent i, $E_i(Y_{jn}) = E(Y_{jn}|F_i)$ is agent i's expectation, given its information set F_i, about the choices of its neighbors, Y_{jn}, within the neighborhood of radius r, $\mathcal{N}_i = \mathcal{N}_i(r)$, containing a fixed number of neighbors $k = |\mathcal{N}_i|$ that does not depend on i; X_i is the vector of observed characteristics of agent i; ε_i is agent i's private characteristic observed only by agent i, and $(\alpha_0', \beta_0', b_0, \gamma_0)$, $\gamma_0 \geq 0$, is the vector of unknown coefficients. The distribution of private shocks is known to all players. It is assumed that $\varepsilon_i \sim$ i.i.d. $N(0, \sigma_0^2)$, and are independent of $\{X_i\}$. The information set of each player consists of the entire state vector, $\overline{X}_n = (X_1', ..., X_n')'$, and its private information ε_i, *i.e.*, $F_i = (\overline{X}_n, \varepsilon_i)$.

The choice of player i is assumed to be directly affected by its neighbors only in a fixed neighborhood of the known radius, $r > 0$, with respect to some socio-economic metric. However, it will be indirectly affected by all other players. The number of the neighbors within the r-neighborhood of each agent is assumed to be fixed and equal to k. To avoid the incidental parameters problem, the k coefficients $(\alpha_{j0})_{j \in \mathcal{N}_i}$, measuring the effect of these k neighbors, are assumed to depend only on the relative locations of i and j, but not on j or i. We formally specify the metric and the neighborhood structure in the following section.

The above assumptions seem reasonable in many empirical settings. For example, in their R&D decision, firms would take into account R&D of its neighbors within a certain distance in the geographic (or product characteristic) space, rather than all firms in the market. This is due to the fact that technological spillovers, knowledge diffusion and labor mobility—determinants of R&D diffusion—are usually confined to a limited geographical or technological area.

Aside from the unobserved heterogeneity captured by the private shocks ε_i, we do not allow individual heterogeneity in the parameters. The reason is that the model assumes only one repetition of the game and the number of agents growing to infinity to develop the asymptotic theory. Clearly, allowing heterogenous parameters across individuals will result in inconsistency.

Model (1) is fairly general for applications. It can arise as a system of best response functions of a static game of incomplete information among n players. Below, we derive these equations for two strategic interaction models: (i) R&D investment by firms; and (ii) labor supply by women. In these models, decisions of players exhibit strategic complementarities, and are subject to threshold effects.

2.1. Spillovers in R&D Investment

A large body of empirical evidence suggests presence of technology and R&D spillovers among firms, e.g., [10]. Audretsch and Feldman [10] find that knowledge spillovers are more prevalent in the industries that exhibit spatial clustering. Positive R&D spillovers may occur through several channels including knowledge transfers, labor mobility and imitation. Therefore, it is reasonable to expect the magnitude of such a spillover effect to depend on the geographical and technological distances between firms. As a result, firms' R&D expenditures may be spatially correlated, and the magnitude of this correlation often decays with the distance between firms.

The literature distinguishes two major channels through which R&D can raise firms' profits: cost-reducing and demand-creating effects. The former allows firms to carry out process improvements leading to efficiency gains and cost reduction, while the latter enables firms to improve the quality of their product and thereby boost the demand. Levin and Reiss [11] analyze a model of monopolistic competition with both demand-creating and cost-reducing R&D spillovers across n firms. Based on a sample of US manufacturing firms, the authors find statistically significant, sizeable spillovers in the cost-reducing R&D and insignificant, small spillovers in the demand-creating R&D in most industries. Levin and Reiss [11] also find the elasticity of product quality to firm's own R&D to be much higher than that of cost to firm's own R&D. Other theoretical models of R&D spillovers include d'Aspremont and Jacquemin [12], and Motta [13], among others.

Yet all these papers model the R&D investment as a continuous variable thereby neglecting the strong empirical evidence that a sizeable proportion of firms do not undertake R&D activities, see, e.g., [14]. One plausible explanation is that the demand-creating effect of R&D is subject to threshold effects: the quality could be raised only after a certain minimum level R&D investment is attained; R&D has no effect on the quality below this level. Thus, the R&D expenditure could be viewed as a censored decision variable whose optimal values below a certain threshold are unobserved. This type of model in the single-firm setup is analyzed by Gonzalez and Jaumandreu [15].

To study spatial spillovers in R&D investment, we develop a simple model of strategic interaction with a censored decision variable that incorporates the empirical findings discussed above. We consider

a single, monopolistically competitive industry composed of a large number, n, firms, each producing a brand of the same product differentiated by quality. Let p_i, q_i, s_i and y_i denote, respectively, the price, demand, product quality and R&D expenditure of firm i. To derive q_i, we employ a variant of the Dixit-Stiglitz [16] model of monopolistic competition in which the CES utility of a representative consumer is augmented with preference for quality:

$$u(q_1, ..., q_n) = \left[\sum_{i=1}^{n} (q_i s_i^{\eta})^{\rho} \right]^{1/\rho}, \ 0 < \rho < 1$$

where $\eta > 0$ is a quality sensitivity parameter. Utility maximization yields the demand for firm i of the form: $q_i(p_i, s_i) = \overline{K} p_i^{-\nu} s_i^{\epsilon}$, where $\nu = 1/(1-\rho) > 1$ is the elasticity of substitution between the quality-adjusted goods, $\epsilon = \eta(\nu - 1)$, $\overline{K} = I \overline{p}^{-1}$ with I being consumer's income and $\overline{p} = \sum_{i=1}^{n} (p_i/s_i^{\eta})^{1-\nu}$ is a quality-adjusted price index. To obtain non-increasing marginal demand for quality, $\partial^2 q_i / \partial s_i^2 \le 0$, suppose that $\eta \le 1/(\nu - 1)$. If the number of firms is large, it is reasonable to assume the effect of a single firm's decision on the industry index $\overline{p}$ to be negligible, *i.e.*, $\overline{K}$ is constant, and normalize $\overline{K} = 1$.

Following Gonzalez and Jaumandreu [15], we assume that firm's own R&D expenditure affects only its product quality, subject to a technological constraint:

$$s_i = s(y_i) = \begin{cases} (y_i + 1)^{\delta} & \text{if} \ \ y_i > \overline{y} \\ (\overline{y} + 1)^{\delta} & \text{if} \ \ y_i \le \overline{y} \end{cases}$$

where $\overline{y}$ is the minimum investment required for quality improvements, $0 < \delta < 1$ is the R&D sensitivity parameter. Throughout, we use $y_i + 1$ instead of y_i to ensure that the logarithm of the censored investment is defined for zero values, and let $Y_i = \log(y_i + 1)$. It is a convenient normalization, which does not affect the results.

Furthermore, in light of the above empirical findings, we assume that other firms' R&D have only a cost-reducing effect on firm i, and this effect is limited to the fixed r-neighborhood, $\mathcal{N}_i(r)$, of firm i:

$$c_i = c_i(X_i, Y_{-i}, e_i) = \exp\left(X_i'\beta + \alpha \sum_{j \in \mathcal{N}_i} Y_j + e_i \right)$$

where c_i and X_i are, respectively, the marginal cost and vector of observed cost-determinants of firm i, $Y_i = \log(y_i + 1)$ is the log of firm i's investment, $Y_{-i} = (Y_1, .., Y_{i-1}, Y_{i+1}, ..., Y_n)$ is the vector of the log R&D choices of all firms except i, and e_i is firm i's idiosyncratic cost component. The coefficient $\alpha \le 0$ measures the strength of this spillover effect.

Suppose that all firms observe $\overline{X}_n = (X_1, ..., X_n)'$, but e_i is observed only by firm i. Given this uncertainty about the choices of other firms, following Durlauf [17], see also [6], we assume that each firm i decides on its R&D investment based on its beliefs about the choices of the other firms, $E_i(Y_{-i}) = E(Y_{-i}|\overline{X}_n, e_i)$, which are formed as the conditional expectation given all the information available to firm i.

Based on these beliefs, firms choose simultaneously price, p_i, and R&D investment, y_i, to maximize their profits subject to a technological constraint, *i.e.*, solve

$$\max_{p_i \ge 0, y_i \ge 0} \Pi(p_i, y_i, X_i, E_i(Y_{-i}), e_i) = [p_i - c_i(X_i, E_i(Y_{-i}), e_i)] q_i(p_i, s_i(y_i)) - C(y_i) \qquad (2)$$

$$\text{s.t. } q_i\left(p_i, s_i\right) = p_i^{-\nu} s_i^{\epsilon}, \; s_i = \begin{cases} (y_i + 1)^{\delta} & \text{if } y_i > \overline{y} \\ (\overline{y} + 1)^{\delta} & \text{if } y_i \le \overline{y} \end{cases} \quad \text{and } C(y_i) = \begin{cases} y_i + 1 & \text{if } y_i > \overline{y} \\ y_i & \text{if } y_i \le \overline{y} \end{cases} \quad (3)$$

where $C(y_i)$ is the cost of investment. The nonstochastic threshold $\overline{y} \ge 0$ is assumed to be observed and constant across all firms.

Lemma 1. *The solution to optimization problem (2) and (3) is given by (1) with* $\alpha_{j0} = -\tau\alpha$, $j = 1, .., k$, $\beta_0 = -\tau\beta$, $\varepsilon_i = -\tau e_i$, $\tau = (\nu - 1)/(1 - \epsilon\delta)$, $b_0 = \log\left(\left[\epsilon\delta\nu^{-\nu}(1-\nu)^{(1-\nu)}\right]^{1/(1-\epsilon\delta)}\right)$ *and* $\gamma_0 = \log\left((1-\epsilon\delta)^{-1/(\epsilon\delta)}(\overline{y} + 1)\right) > 0$.

If $\alpha < 0$, *i.e.*, R&D of the neighbors has a cost-reducing effect on firm i, then both the probability and intensity of firm i's R&D increases with the expected R&D of its neighbors. In other words, there are strategic complementarities or positive externalities in the R&D decision of firms. Furthermore, the probability of R&D is also increasing in (i) the elasticity of demand with respect to quality, higher ϵ; (ii) the elasticity of quality with respect to R&D, higher δ; and (iii) the market power, lower ν. The latter is consistent with the Schumpeterian argument that economies of scales make R&D more attractive to large firms than to small firms.

2.2. Peer Effects in Female Labor Supply

Our next example involves social interactions in the female labor supply. Suppose the utility of female i is defined over her consumption, c_i, and leisure, l_i, as follows:

$$U(c_i, y_i, h_i) = \frac{c_i^{\delta}}{\delta} + h_i l_i$$

where $0 < \delta < 1$ is the parameter characterizing the relative preference for consumption over leisure. Let y_i be the labor supply of female i, and let the weight on the leisure, h_i, capture the peer effects that depend on the labor supply decisions of female i's peers in her social neighborhood, referred to as friends, as follows:

$$h_i = h_i\left(X_i, Y_{-i}, e_i\right) = \exp(X_i'\beta + \alpha \sum_{j \in \mathcal{N}_i} Y_j + e_i)$$

where X_i is the vector of observed characteristics of woman i, $Y_j = \log(y_j + 1)$ is the log labor supply of woman i's friends, and e_i is her private characteristic unobserved by other women. As in the previous example, we use $y_i + 1$ instead of y_i to ensure that log of the censored labor supply is defined for zero values. In presence of positive peer effects, $\alpha < 0$, which implies mutual reinforcement in the choices within the social group.

As before, all women observe $\overline{X}_n = (X_1, ..., X_n)'$, but e_i is observed only by i. Woman i makes her decision based on her beliefs about the choices of her peer group, $E_i(Y_{-i}) = E(Y_{-i}|\overline{X}_n, e_i)$, which are formed as the conditional expectation given all the information available to woman i. Based on these beliefs, women simultaneously maximize their utility subject to threshold effects:

$$\max_{c_i \ge 0, y_i \ge 0} U(c_i, X_i, E_i(Y_{-i}), e_i) = \frac{c_i^{\delta}}{\delta} + h_i(X_i, E(Y_{-i}), e_i) l_i \quad (4)$$

$$\text{s.t. } c_i = \begin{cases} w\,(y_i + 1) & \text{if } w\,(y_i + 1) > \overline{c} \\ \overline{c} & \text{if } w\,(y_i + 1) \leq \overline{c} \end{cases}, l_i = \begin{cases} T - y_i - 1 & \text{if } w\,(y_i + 1) > \overline{c} \\ T - y_i & \text{if } w\,(y_i + 1) \leq \overline{c} \end{cases} \quad (5)$$

where w is the wage, T is the time endowment, $\overline{c}$ is the reservation labor income, which can be interpreted as welfare or other government transfers. The nonstochastic threshold $\overline{c} \geq 1$ is assumed to be observed and constant across women.

Lemma 2. *The solution to optimization problem (4) and (5) is given by (1) with* $\alpha_{j0} = -\alpha/\,(1 - \delta)$, $j = 1, .., k$, $\beta_0 = -\beta/\,(1 - \delta)$, $\varepsilon_i = -e_i/\,(1 - \delta)$, $b_0 = \log\left(w^{\delta/(1-\delta)}\right)$ *and* $\gamma_0 = \log\left((1 - \delta)^{-1/\delta}\,\overline{c}\right) > 0$.

If $\alpha < 0$, *i.e.*, there are positive peer effects, then both the probability and magnitude of woman i's labor supply increases with the expected labor supply of her peers.

3. Equilibrium: Characterization and Weak Dependence

We assume that in equilibrium, players have self-consistent expectations, *i.e.*, their subjective expectations or beliefs coincide with the expectation based on the equilibrium distribution of strategies conditional on $\overline{X}_n$. That is,

$$E_i\,(Y_j) = E\left(Y_j | \overline{X}_n, \varepsilon_i\right) = E\left(Y_j | \overline{X}_n\right) := \widetilde{Y}_{jn} \quad (6)$$

where the expectation $E\left[\cdot | \overline{X}_n\right]$ is taken with respect to the equilibrium conditional distribution of strategies. The last equality follows from independence of $\{\varepsilon_i\}$, and independence of $\{\varepsilon_i\}$ and $\{X_i\}$.

Suppose that the ε_i are i.i.d. $N\left(0, \sigma_0^2\right)$. Taking conditional expectation of Equation (1) with respect to the equilibrium distribution of strategies, conditional on $\overline{X}_n$, yields:

$$\begin{aligned} \widetilde{Y}_{in} &= \Phi\left(\frac{\sum_{j \in \mathcal{N}_i} \alpha_{j0} \widetilde{Y}_{jn} + X_i' \beta_0 + b_0 - \gamma_0}{\sigma_0}\right)\left(\sum_{j \in \mathcal{N}_i} \alpha_{j0} \widetilde{Y}_{jn} + X_i' \beta_0 + b_0\right) \\ &\quad + \sigma_0 \phi\left(\frac{\sum_{j \in \mathcal{N}_i} \alpha_{j0} \widetilde{Y}_{jn} + X_i' \beta_0 + b_0 - \gamma_0}{\sigma_0}\right) \end{aligned} \quad (7)$$

where Φ and ϕ are, respectively, the c.d.f. and p.d.f. of the standard normal distribution.

Provided that they are well-defined, strategies $\{Y_{in}, i = 1, .., n\}$ are independent across i conditional on $\overline{X}_n$ and have censored normal distributions with the means $\left\{\widetilde{Y}_{in}, i = 1, .., n\right\}$, the common variance σ_0 and the common nonstochastic threshold γ_0. In equilibrium, $\left\{\widetilde{Y}_{in}, i = 1, .., n\right\}$ satisfy system (7). If this system has a unique solution, the corresponding equilibrium strategies, $\{Y_{in}, i = 1, .., n\}$, will be also unique with probability 1, since a censored normal variable is uniquely characterized by its mean, variance and threshold. This leads to the following characterization of equilibrium.

Definition 1. *An equilibrium is a set of policy functions* $\{Y_{in}, i = 1, .., n\}$ *whose conditional mean functions* $\left\{\widetilde{Y}_{in} = E\left(Y_{in} | \overline{X}_n\right)\right\}_{i=1}^{n}$ *satisfy system (7).*

A similar characterization of equilibrium in discrete games of incomplete information is used in [8].

An appealing feature of Equation (1) is that it reduces to the popular Tobit model, which is part of any regression package. However, the difficulty is that $\{Y_{in}\}$ depends on the latent regressors, $\left\{\widetilde{Y}_{in}\right\}$. Thus,

one would need first to obtain consistent estimates of the latent regressors, and then use any consistent estimation procedures for the Tobit model.

Since consistency of any estimation method hinges upon uniqueness of equilibrium, we first prove existence and uniqueness of the pure strategy equilibrium. To this end, we maintain the following assumption.

Assumption 1. *The shocks $\varepsilon_i \sim$ i.i.d. $N\left(0, \sigma_0^2\right)$ and $\lambda = k\left[\phi\left(0\right)\gamma_0/\sigma_0 + 1\right]\overline{\alpha} < 1$, where $\gamma_0 \geq 0$, $\overline{\alpha} = \max_{1 \leq j \leq k} |\alpha_{j0}|$, $\phi\left(\cdot\right)$ is the p.d.f. of the standard normal distribution.*

This assumption restricts the strength of interactions, captured by the coefficients α: interactions must not be too strong for a stable equilibrium to exist. Intuitively, if the interactions are long-ranged and too strong, then the effect of remote neighbors is substantial and may lead to instability and multiple equilibria. Since it involves only the estimated coefficients and the number of neighbors, k, is typically known, the assumption is testable.

Assumption 1 is similar to Assumptions B and C in [4], which restrict the strength of interactions to obtain a unique equilibrium in a discrete choice game of social interactions.

Based on this assumption, we can now show existence and uniqueness of equilibrium.

Theorem 1. *Under Assumption 1, there exists a unique equilibrium of model (1).*

In general, without restrictions on the parameters, multiple equilibria could occur. If one does not want to impose restrictions directly, one can use the Mathematical Program with Equilibrium Constraints (MPEC) routine to deal with multiple equilibria implicitly by choosing the equilibrium that maximizes the empirical likelihood.

In equilibrium, the policy variables will be correlated across players. To characterize their dependence, we assume that the process $\left\{W_{in} = \left(Y_{in}, \widetilde{Y}_{in}, X_i', \varepsilon_i\right)'\right\}$ is indexed by a vector of locations $t(i) = (t_1, ..., t_d) \in \mathbb{Z}^d$ on the lattice $\mathbb{Z}^d$, and hence can be viewed as a random field on $\mathbb{Z}^d$. In other words, $\left\{W_{in} = W_{t(i)n}, t(i) \in \Lambda_n, n \geq 1\right\}$ is a triangular array of vector-valued random fields defined on a probability space $(\Omega, \mathfrak{F}, P)$ and observed on sample regions $\Lambda_n \subset \mathbb{Z}^d$. In the following, to simplify notation, we suppress the index t and write $W_{in} = W_{t(i)n}$. Furthermore, we denote by $\|\cdot\|$ the Euclidian norm in $\mathbb{R}^d$ and by $\|\cdot\|_p = \left[E\|\cdot\|^p\right]^{1/p}$ – the L_p-norm.

Assumption 2. *The data-generating process $\left\{W_{in} = W_{t(i)n}, t(i) \in \Lambda_n, n \geq 1\right\}$ is a triangular array of random fields indexed by $t(i) = (t_1, ..., t_d) \in \mathbb{Z}^d$, where the $\Lambda_n \subset \mathbb{Z}^d$ are the sample regions such that $|\Lambda_n| \to \infty$ as $n \to \infty$. The distance between players i and j is measured by the Euclidian metric: $dist(i, j) = \|t(i) - t(j)\|$.*

This assumption implies that the players' locations are exogenous, *i.e.*, they are known and determined outside the model. Extensions to endogenous locations would require explicit modeling of the location choice, and would therefore considerably complicate the model. This extension is an interesting direction for future research.

Given Assumption 2, it turns out that the equilibrium policy variables satisfy a weak dependence condition known as near-epoch dependence (NED), see [5], under the same condition that ensures uniqueness of equilibrium. For ease of reference, we state definition of NED random variables.

Definition 2. *The triangular array of random fields* $\{W_{tn}, t \in \Lambda_n, n \geq 1\}$, $\|W_{tn}\|_2 < \infty$, *is* L_2-*NED on* $\{V_{tn}, t \in \Lambda_n, n \geq 1\}$ *iff* $\sup_{n, t \in \Lambda_n} \|W_{tn} - E(W_{tn}|\mathfrak{F}_{tn}(m))\|_2 \leq \psi(m)$ *for some sequence* $\psi(m) \to 0$ *as* $m \to \infty$, *where* $\mathfrak{F}_{tn}(m) = \sigma(V_{sn}; s \in \Lambda_n : \|t - s\| \leq m)$.

Theorem 2. *Suppose Assumptions 1 and 2 hold and* $\|X\|_6 = \sup_i \|X_i\|_6 < \infty$, *then (i)* $\{\widetilde{Y}_{in}\}$ *is* L_2-*NED on* $\{X_i\}$ *with the NED numbers* $\psi(m) = 2\|Y\|_2 \lambda^{[m/r]}$, *where* $p = [m/r]$ *is the integer of part of* m/r, *(ii)* $\{Y_{in}\}$ *is* L_2-*NED on* $\{(X_i', \varepsilon_i)\}$ *with the NED numbers* $c\psi^{1/12}(m)$ *for constant* c *that does not depend on* m, *(iii)* $\{\mathbf{1}(Y_{in} > \gamma_0)\}$ *and* $\{\mathbf{1}(Y_{in} = 0)\}$ *are* L_2-*NED on* $\{(X_i', \varepsilon_i)\}$ *with the NED numbers* $(2 + k\overline{\alpha})\psi^{1/3}(m)$.

The value of the constant c is given in the proof, but it is not important for what follows.

4. Identification and Estimation

We now discuss identification and estimation of our model. Let $Z_{in} = \left(\left(\widetilde{Y}_{jn}, j \in \mathcal{N}_i\right), 1, X_i'\right)'$, and let $\delta = ((\alpha_j, j \in \mathcal{N}_i), b, \beta')'$ denote the corresponding vector of the coefficients. Given the specification, it is natural to identify and estimate all unknown parameters based on the likelihood function. The log likelihood function of the model is

$$\log L(\theta, \gamma) = \sum_{i=1}^n \log f\left(Y_i | \overline{X}_n; \theta, \gamma\right) \text{ with} \tag{8}$$

$$\log f\left(Y_i | \overline{X}_n; \theta, \gamma\right) = \mathbf{1}(Y_i = 0)\log \Phi\left(\frac{\gamma - Z_{in}'\delta}{\sigma}\right) + \mathbf{1}(Y_i > \gamma)\log\left[\sigma^{-1}\phi\left(\frac{Y_i - Z_{in}'\delta}{\sigma}\right)\right]$$

where $\theta = (\delta', \sigma)'$. Likelihood function (8) involves an unknown threshold parameter, γ, which is in contrast to the standard Tobit model, where the threshold is assumed to be known and equal to zero. The maximum likelihood (ML) estimator of γ is the minimum order statistics of the uncensored subsample. More specifically, partition the dependent variable and regressor matrix into two parts: $Y = \left(Y_{(0)}', Y_{(1)}'\right)'$ and $Z = \left(Z_{(0)}', Z_{(1)}'\right)'$, where the subscript (0) indicates that observations come from the censored subsample, and the subscript (1) – from the uncensored subsample, and let

$$\widehat{\gamma} = \min\{Y_i : Y_i > 0, i = 1, ..., n\} \tag{9}$$

As shown in Proposition 1 below, $\widehat{\gamma}$ is a consistent estimator of γ. The ML estimators of the other parameters θ can then be obtained by the standard differentiation techniques.

Assumption 3. *Suppose (i)* $\sup_{i,n} \|W_{in}\|_6 < \infty$; *and (ii)* $\{X_i\}$ *is a stationary* α-*mixing process with coefficient satisfying* $\alpha(k, l, m) \leq (k + l)^\varsigma \widehat{\alpha}(m)$, $\varsigma \geq 0$, $\widehat{\alpha}(m)$ *s.t.* $\sum_{m=1}^\infty m^{(d\varsigma/3 - 1)}\widehat{\alpha}^{1/6}(m) < \infty$.[1]

Proposition 1. *Under Assumptions 1–3,* $\lim_{n \to \infty} P(n(\widehat{\gamma} - \gamma_0) < z) = 1 - \exp(-az)$ *for* $z > 0$, *where*

$$a = E\left[\Phi\left(\frac{Z_{in}'\delta - \gamma}{\sigma}\right)\right] E\left[\frac{\phi\left(\left(Z_{in,(1)}'\delta - \gamma\right)/\sigma\right)}{\sigma\Phi\left(\left(Z_{in,(1)}'\delta - \gamma\right)/\sigma\right)}\right]$$

and $Z_{in,(1)}$ *is the regressor vector from the uncensored subsample.*

---

[1] For the definition of mixing coefficients, see [18].

Thus, $\widehat{\gamma}$ is n-consistent and asymptotically exponentially distributed. For i.i.d. sample, this result has been established by Carson and Sun [19]. So, Proposition 1 extends [19] to a spatially dependent case. The superconsistency of $\widehat{\gamma}$ is a well-known consequence of the dependence of the support of Y_i on γ.

Proposition 1 implies that γ_0 is identified. The remaining parameters, $\theta_0 = (\delta_0', \sigma_0)'$ can now be identified from the likelihood function. Alternatively, one can identify θ_0 from the conditional mean function and estimate it by the least squares procedure:

$$\varphi(Z_{in}, \theta, \gamma) = E\left(Y_i | \overline{X}_n\right) = \Phi\left(\frac{Z_{in}'\delta - \gamma}{\sigma}\right) Z_{in}'\delta + \sigma\phi\left(\frac{Z_{in}'\delta - \gamma}{\sigma}\right) \tag{10}$$

In contrast to the standard Tobit model with zero-threshold, the ML estimator does not strictly dominate the least squares (LS) estimator in our model due to the presence of the first-step estimator. The thing is that the LS objective function is continuous in γ, while the likelihood function is not. The latter implies that small finite-sample biases in $\widehat{\gamma}$ may cause sizeable finite-sample biases in the ML estimates of θ. This prediction is confirmed by the simulation results of Section 6, which suggest larger finite-sample biases in the ML than in the LS estimator. This is the main rationale for considering the LS procedure as an alternative to the ML procedure in our model.

Thus, estimation of model (1) could be carried out in two steps. First, estimate the threshold parameter γ by the minimum order statistic of the uncensored subsample $\widehat{\gamma}$, and substitute it for the true γ_0 in (8) and (10). Then, estimate the remaining parameters θ in (8) and (10) by the ML or LS procedures, respectively. Note that the least squares estimator of γ in (1) will be imprecise due to near multicollinearity of the intercept and threshold. Therefore, we use the first-step estimator $\widehat{\gamma}$ in both procedures.

We now present sufficient conditions for identification of θ.

Assumption 4. *Suppose (i) at least one of components of X_i has the full support, $\mathbb{R}$; and (ii) $E\left(Z_{in}Z_{in}'\right)$ is positive definite.*

Theorem 3. *Under Assumptions 1–4, $Q\left(\theta, \gamma_0\right) = E\left[m\left(W_{in}, \theta, \gamma_0\right)\right]$ is uniquely maximized at θ_0 for*

A. $m\left(W_{in}, \theta, \gamma_0\right) = \log f\left(Y_i | \overline{X}_n; \theta, \gamma_0\right)$, where $\log f\left(Y_i | \overline{X}_n; \theta, \gamma\right)$ is as defined in (8), and

B. $m\left(W_{in}, \theta, \gamma_0\right) = -\left(Y_i - \varphi(Z_{in}, \theta, \gamma_0)\right)^2$, where $\varphi(Z_{in}, \theta, \gamma)$ is as defined in (10).

Practically, the second-step estimation of θ could be implemented through the following nested fixed-point (NFXP) algorithm: (i) in an inner loop, for a given θ, find the unique solution of the equilibrium equations (7) by the fixed-point algorithm; and (ii) in an outer loop, search over $\theta \in \Theta$ that maximizes the objective function. Let $\widetilde{Y}\left(\theta, \gamma\right) = \left(\widetilde{Y}_1, ..., \widetilde{Y}_n\right)$ be the solution of the equilibrium equations (7). Then, the resulting estimator can be represented as

$$\widehat{\theta} = \arg\max_{\theta \in \Theta}\left\{Q_n\left(\theta, \widetilde{Y}\left(\theta, \widehat{\gamma}\right), \widehat{\gamma}\right) = \frac{1}{n}\sum_{i=1}^{n} m\left(W_{in}, \theta, \widehat{\gamma}\right)\right\} \tag{11}$$

where $m\left(\cdot, \cdot, \cdot\right)$ is either the log likelihood function defined in (8) or minus the squared deviation of Y_i from the conditional mean defined in (10). This formulation makes explicit the dependence of the equilibrium variables on the estimated parameters. Given superconsistency of the first-step estimator, the resulting second-step maximum likelihood or least squares estimators of θ will be root-n consistent,

asymptotically normal and independent of $\widehat{\gamma}$, as shown in Theorem 4 below. However, the NFXP algorithm will be computationally costly for large cross-sectional datasets, e.g., $n \geq 200$.

To overcome this problem, we instead use the constrained optimization algorithm proposed by Su and Judd [7]. The idea is to solve the following constrained optimization problem:

$$\max_{(\theta, \widetilde{Y})} Q_n\left(\theta, \widetilde{Y}, \widehat{\gamma}\right) \text{ subject to } h(\widetilde{Y}, \theta, \widehat{\gamma}) = 0 \tag{12}$$

where $h(\widetilde{Y}, \theta, \gamma) = 0$ is the vector representation of the equilibrium system (7). Note that $\widetilde{Y}$ in this formulation does not depend on θ, and is chosen simultaneously with θ to maximize the objective function subject to the equilibrium constraints. This obviates the need to solve the multi-dimensional fixed point problem for $\widetilde{Y}$ at each iteration on θ.

Su and Judd [7] prove equivalence of problems (11) and (12) provided that the model is identified. They also demonstrate the computational advantage of this constrained optimization algorithm over the NFXP algorithm in the context of a single-agent dynamic discrete choice model. In particular, they show that the proposed algorithm leads, on average, to a ten-fold reduction of the computational time relative to the NFXP algorithm.

Since our model is identified by Theorem 3, the maximizer $\widehat{\theta}$ of problem (11) equals the maximizer $\bar{\theta}$ of problem (12) by Proposition 1 of Su and Judd [7], and one can thus replace the computationally intensive problem (11) by the simpler problem (12). We investigate performance of the constrained optimization algorithm for our model in the Monte Carlo study of Section 6.

5. Consistency and Asymptotic Normality

We next show consistency and asymptotic normality of the maximum likelihood and least squares estimators. To this end, we need the following assumption.

Assumption 5. *Suppose*

(i) $\Theta = \Delta \times \Sigma$ *and* Γ *are compact,* $\Sigma = [\sigma_1; \sigma_2]$, $\sigma_1 > 0$, $\theta_0 \in int(\Theta)$, $\gamma_0 \in int(\Gamma)$.

(ii) $\sup_{i,n} E\left[\sup_\Theta \log^2 \Phi\left((\gamma_0 - Z'_{in}\delta)/\sigma\right)\right] < \infty$, $\sup_{i,n} E\left[\sup_\Theta (Y_{in} - Z'_{in}\delta)^2\right] < \infty$,
$\sup_{i,n} E\left\|\sup_\Theta \frac{\partial m(W_{in}, \theta, \gamma_0)}{\partial \theta}\right\| < \infty$.

Part (i) Assumption 5 is the standard condition on the parameter space and the true parameter value; Part (ii) is used to verify uniform convergence of various sample functions. Generally, the above assumptions are slightly stronger than those in the fully parametric Tobit model with zero-threshold, since our Tobit estimator of θ relies on a nonparametric first-step estimator of γ.

Theorem 4. *Under Assumptions 1–5, the maximum likelihood and least squares estimators are both consistent and asymptotically normal,* i.e., $\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow{d} N(0, \Omega)$, *where* $\Omega = H_0^{-1} S_0 H_0^{-1}$ *with* $H_0 = E\left[\frac{\partial^2}{\partial\theta\partial\theta'} m(W_{in}, \theta_0, \gamma_0)\right]$ *and* $S_0 = Var\left[\frac{\partial}{\partial\theta} m(W_{in}, \theta_0, \gamma_0)\right]$.

Thus, both the maximum likelihood and least squares estimators of θ are $\sqrt{n}$-consistent and asymptotically normal. To conduct inference, it remains to obtain a consistent estimate of the covariance matrix S_0. For this purpose, one can employ the following spatial HAC estimator:

$$\widehat{S}\left(\widehat{\theta}, \widehat{\gamma}\right) = \frac{1}{n} \sum_{i \in \Lambda_n} \sum_{j \in \Lambda_n : |i-j| \leq h_n} K((i-j)/h_n) \frac{\partial}{\partial \theta} m(W_{in}, \widehat{\theta}, \widehat{\gamma}) \frac{\partial}{\partial \theta'} m(W_{jn}, \widehat{\theta}, \widehat{\gamma})$$

where $K((i-j)/h_n) = K((i_1-j_1)/h_n, ..., (i_d-j_d)/h_n)$ is a d-dimensional symmetric kernel, and h_n is a bandwidth parameter. Jenish [20] proves consistency of this estimator for more general nonparametric estimators of γ. In our model, consistency is achieved by bandwidth parameters satisfying $h_n = O(n^{1/(3d)})$.

6. Numerical Results

In this section, we examine the finite sample properties of the maximum likelihood (ML) and least squares (LS) estimators of our censored model, as well the performance of the Su-Judd [7] algorithm.

Throughout, data $\{W_{i_1,i_2}\}$ reside on the two-dimensional lattice $\mathbb{Z}^2$, where $(i_1, i_2) \in \mathbb{Z}^2$ denotes, for simplicity, the vector of coordinates. The data are simulated on a rectangular grid of $(m_1 + 300) \times (m_2 + 300)$ locations. To control for boundary effects, we discard the 300 outer boundary points along each of the axes and use the sample of size $n = m_1 m_2$ for estimation.

Our experiment consists of two stages: (i) simulation; and (ii) estimation. In the first stage, we first simulate two i.i.d. $N(0,1)$ processes $\{\varepsilon_{i_1,i_2}\}$ and $\{\eta_{i_1,i_2}\}$, which are independent of each other. Next, using the fixed-point algorithm, we generate the process $\{X_{i_1,i_2}\}$ according to:

$$X_{i_1,i_2} = 0.2 \left(X_{i_1-1,i_2} + X_{i_1,i_2-1} + X_{i_1+1,i_2} + X_{i_1,i_2+1}\right) + \eta_{i_1,i_2}$$

and then the process $\left\{\widetilde{Y}_{i_1,i_2}\right\}$ according to:

$$\widetilde{Y}_{i_1,i_2} = \Phi\left(\frac{\alpha\left(\widetilde{Y}_{i_1-1,i_2} + \widetilde{Y}_{i_1,i_2-1} + \widetilde{Y}_{i_1+1,i_2} + \widetilde{Y}_{i_1,i_2+1}\right) + X_{i_1,i_2}\beta + b - \gamma}{\sigma}\right) \cdot$$
$$\cdot \left[\alpha\left(\widetilde{Y}_{i_1-1,i_2} + \widetilde{Y}_{i_1,i_2-1} + \widetilde{Y}_{i_1+1,i_2} + \widetilde{Y}_{i_1,i_2+1}\right) + X_{i_1,i_2}\beta + b\right] - \quad (13)$$
$$-\sigma\phi\left(\frac{\alpha\left(\widetilde{Y}_{i_1-1,i_2} + \widetilde{Y}_{i_1,i_2-1} + \widetilde{Y}_{i_1+1,i_2} + \widetilde{Y}_{i_1,i_2+1}\right) + X_{i_1,i_2}\beta + b - \gamma}{\sigma}\right)$$

with $\alpha = 0.2$, $\beta = 1$, $b = 1.25$, $\gamma = 2$ and $\sigma = 1$. Last, we form the process $\{Y_{i_1,i_2}\}$ according to:

$$Y_{i_1,i_2} = Y_{i_1,i_2}^* \mathbf{1}\left(Y_{i_1,i_2}^* > \gamma\right), \text{ where}$$
$$Y_{i_1,i_2}^* = b + X_{i_1,i_2}\beta + \alpha\left(\widetilde{Y}_{i_1-1,i_2} + \widetilde{Y}_{i_1,i_2-1} + \widetilde{Y}_{i_1+1,i_2} + \widetilde{Y}_{i_1,i_2+1}\right) + \varepsilon_{i_1,i_2}$$

In the second stage, we first construct the minimum order statistic estimator of γ – $\widehat{\gamma}$ – and then use the Su-Judd [7] constrained optimization algorithm to estimate the remaining parameters $\theta = (\alpha, \beta', b, \sigma)'$. As discussed in Section 4, we estimate the n-dimensional vector of endogenous

variables $\widetilde{Y} = \left\{\widetilde{Y}_{i_1,i_2}\right\}$ jointly with θ, instead of computing it at each iteration on θ, *i.e.*, we solve the constrained optimization problem:

$$\max_{(\theta,\widetilde{Y})} Q_n\left(\theta,\widetilde{Y},\widehat{\gamma}\right) \text{ subject to } h(\widetilde{Y},\theta,\widehat{\gamma}) = 0$$

where $h(\widetilde{Y},\theta,\gamma) = 0$ is the vector representation of the equilibrium system (13).

The estimation results based on 1000 Monte-Carlo repetitions are presented in Tables 1–5. Table 1 reports the estimates of the autoregressive parameter $\alpha = 0.2$.

Table 1. Estimation of auto-regressive parameter: $\alpha = 0.2$.

Sample Size	Mean	Bias (%)	SD	RMSE	25th Pct	50th Pct	75th Pct
Maximum Likelihood							
$N = 200$	0.1978	−1.1212	0.0056	0.0060	0.1946	0.1979	0.2013
$N = 400$	0.1989	−0.5674	0.0030	0.0032	0.1970	0.1991	0.2008
$N = 600$	0.1991	−0.4255	0.0022	0.0024	0.1977	0.1992	0.2007
$N = 800$	0.1994	−0.3052	0.0018	0.0019	0.1983	0.1994	0.2006
$N = 1000$	0.1995	−0.2691	0.0015	0.0016	0.1984	0.1995	0.2005
Least Squares							
$N = 200$	0.1989	−0.5452	0.0059	0.0060	0.1953	0.1990	0.2027
$N = 400$	0.1995	−0.2578	0.0033	0.0034	0.1974	0.1996	0.2018
$N = 600$	0.1996	−0.1981	0.0023	0.0024	0.1982	0.1997	0.2013
$N = 800$	0.1998	−0.1162	0.0019	0.0019	0.1985	0.1999	0.2011
$N = 1000$	0.1998	−0.0787	0.0017	0.0017	0.1986	0.1998	0.2011

Both the ML and LS estimators of α behave well for all sample sizes. The finite-sample bias declines rapidly from about 1.1% ($n = 200$) to 0.3% (for $n = 1000$) in the case of the ML estimator, and from 0.55% ($n = 200$) to 0.08% ($n = 1000$) in the case of the LS estimator. These results suggest that a five-fold increase in the sample size leads to a more than three-fold reduction in the ML bias, and a more than six-fold decrease in the LS bias, which is consistent with our asymptotic theory. The standard errors also fall off rapidly with the sample size. A similar pattern is observed for the estimates of the slope β, shown in Table 2.

Table 3 contains the minimum order statistic estimates of γ. The finite-sample bias diminishes from 4.5% ($n = 200$) to 0.8% (for $n = 1000$), which means that a five-fold increase in the sample size is associated with more than a five-fold reduction in the bias. This is in line with the theoretical prediction of n-consistency of the minimum order statistics.

Table 2. Estimation of slope: $\beta = 1.0$.

Sample Size	Mean	Bias (%)	SD	RMSE	25th Pct	50th Pct	75th Pct
Maximum Likelihood							
$N = 200$	1.0053	0.5286	0.0666	0.0668	0.9612	1.0072	1.0507
$N = 400$	1.0037	0.3665	0.0419	0.0420	0.9756	1.0022	1.0295
$N = 600$	1.0032	0.3248	0.0324	0.0326	0.9789	1.0038	1.0251
$N = 800$	1.0022	0.2228	0.0254	0.0255	0.9844	1.0014	1.0197
$N = 1000$	1.0019	0.1886	0.0232	0.0233	0.9870	1.0019	1.0169
Least Squares							
$N = 200$	1.0050	0.5027	0.0712	0.0714	0.9538	1.0038	1.0522
$N = 400$	1.0018	0.1758	0.0448	0.0449	0.9717	0.9988	1.0320
$N = 600$	1.0015	0.1534	0.0340	0.0340	0.9775	1.0009	1.0241
$N = 800$	1.0009	0.0909	0.0271	0.0271	0.9820	0.9999	1.0180
$N = 1000$	1.0002	0.0241	0.0250	0.0250	0.9840	1.0007	1.0161

Table 3. Estimation of threshold: $\gamma = 2.0$.

Sample Size	Mean	Bias (%)	SD	RMSE	25th Pct	50th Pct	75th Pct
Min. Order Statistics							
$N = 200$	2.0899	4.4969	0.1128	0.1443	2.0219	2.0528	2.1111
$N = 400$	2.0438	2.1888	0.0462	0.0637	2.0123	2.0288	2.0610
$N = 600$	2.0294	1.4715	0.0310	0.0427	2.0082	2.0198	2.0408
$N = 800$	2.0189	0.9471	0.0193	0.0270	2.0050	2.0130	2.0260
$N = 1000$	2.0161	0.8047	0.0166	0.0231	2.0045	2.0109	2.0219

Next, Tables 4 and 5 present the estimates of the intercept b and the standard deviation σ, respectively.

The maximum likelihood estimates of b and σ exhibit larger biases than those of α and β. However, they decrease as the sample size increases: from 5.5% ($n = 200$) to 1.4% ($n = 1000$) in the case of b, and from 15.9% ($n = 200$) to 6.9% ($n = 1000$) in the case of σ. Thus, the biases still halve when the sample size increases four-fold, consistent with the asymptotic theory. Large small-sample biases could be explained by weak identification or near multicollinearity introduced by the inverse Mills ratio, which is approximately linear over a wide range of its argument.

Table 4. Estimation of intercept: $b_0 = 1.25$.

Sample Size	Mean	Bias (%)	SD	RMSE	25th Pct	50th Pct	75th Pct
Maximum Likelihood							
$N = 200$	1.3185	5.4812	0.1352	0.1516	1.2353	1.3156	1.3950
$N = 400$	1.2862	2.8955	0.0749	0.0832	1.2372	1.2804	1.3321
$N = 600$	1.2778	2.2216	0.0539	0.0607	1.2415	1.2744	1.3130
$N = 800$	1.2702	1.6191	0.0438	0.0482	1.2403	1.2698	1.2990
$N = 1000$	1.2679	1.4353	0.0374	0.0415	1.2428	1.2668	1.2924
Least Squares							
$N = 200$	1.2834	2.6748	0.1449	0.1487	1.1917	1.2808	1.3699
$N = 400$	1.2671	1.3653	0.0843	0.0860	1.2121	1.2630	1.3197
$N = 600$	1.2635	1.0803	0.0582	0.0598	1.2236	1.2625	1.3011
$N = 800$	1.2582	0.6562	0.0483	0.0489	1.2252	1.2572	1.2896
$N = 1000$	1.2560	0.4775	0.0427	0.0431	1.2253	1.2557	1.2852

Table 5. Estimation of standard deviation: $\sigma = 1.0$.

Sample Size	Mean	Bias (%)	SD	RMSE	25th Pct	50th Pct	75th Pct
Maximum Likelihood							
$N = 200$	0.8405	-15.9486	0.0575	0.1695	0.8010	0.8397	0.8798
$N = 400$	0.8966	-10.3429	0.0397	0.1108	0.8700	0.8972	0.9243
$N = 600$	0.9179	-8.2084	0.0316	0.0879	0.8962	0.9185	0.9392
$N = 800$	0.9262	-7.3842	0.0273	0.0787	0.9075	0.9267	0.9448
$N = 1000$	0.9303	-6.9696	0.0246	0.0739	0.9134	0.9305	0.9468
Least Squares							
$N = 200$	0.9860	-1.3956	0.4799	0.4801	0.7597	0.9956	1.2585
$N = 400$	0.9598	-4.0236	0.2901	0.2929	0.8227	0.9858	1.1342
$N = 600$	0.9784	-2.1637	0.2000	0.2012	0.8833	0.9881	1.1034
$N = 800$	0.9828	-1.7215	0.1594	0.1604	0.8930	0.9963	1.0809
$N = 1000$	0.9303	-0.8312	0.1241	0.1243	0.9114	0.9941	1.0732

Interestingly, the LS estimates of all parameters, including b and σ, have smaller finite-sample biases than the respective ML estimates. The reason is that the LS objective function is continuous in the first-step nonparametric estimator of γ, while the likelihood function is not, and small first-step biases in γ may get disproportionately amplified and translate into sizeable second-step biases in θ. Consequently, the biases in the ML estimates of b and σ due to weak identification are further exacerbated by discontinuity of the likelihood function. Nevertheless, as expected, the LS estimator has larger

standard errors than the ML estimator across all parameters. Thus, in contrast to the standard Tobit model with zero-threshold, the ML estimator does not strictly dominate the LS estimator in our model.

Finally, Table 6 reports the computational time and the number of converged iterations for the Su-Judd [7] algorithm. The algorithm performs well for all sample sizes: converges in almost 99% of iterations and the time costs are less than two hours even in large sample sizes such as $n = 1000$. For comparison, the NFXP algorithm will take about 130–150 hours to estimate the model for same sample sizes. Thus,the Su-Judd [7] algorithm offers a considerable time savings over standard nested fixed-point algorithms.

Table 6. Algorithm Performance

Maximum Likelihood					
Sample Size	$N = 200$	$N = 400$	$N = 600$	$N = 800$	$N = 1000$
Number of converged iterations	1000	999	991	992	961
Run Time	2 min	6 min	27 min	61 min	119 min
Least Squares					
Sample Size	$N = 200$	$N = 400$	$N = 600$	$N = 800$	$N = 1000$
Number of converged iterations	942	996	997	999	998
Run Time	2 min	5 min	24 min	57 min	112 min

Overall, the simulations results are consistent with our asymptotic theory: the finite-sample biases and standard errors of the ML and LS estimators decay rapidly with the sample size. Moreover, the Su-Judd [7] constrained optimization algorithm seems to be a viable and effective numerical procedure for estimating games with large number of players, including our model.

7. Conclusions

In this paper, we study identification and estimation of a static game of incomplete information with censored strategies. Specifically, we show existence and uniqueness of an equilibrium as well as its weak dependence property under a condition that restricts the strength of interactions among the players. We then show identification of the parameters and estimate them by maximum likelihood and least squares procedures. The resulting estimators are shown to be consistent and asymptotically normal. We also demonstrate application of our results to modeling spillovers in firms' R&D investment and peer effects in female labor supply.

One direction for future research is to relax the normality assumption on the errors and obtain identification under more general error distributions whose conditional mean functions satisfy contraction mapping conditions, similar to the one used in the paper. Another extension could be to allow for random threshold effects in the outcome variable, using some parametric family of distributions. One can also allow for truncated strategies by slight modifications in the likelihood function and Assumption 1. Finally, instead of the regular lattice, one can consider players located at the nodes of some graph, which describes the network structure, as in the social interactions literature.

Acknowledgments

Conflicts of Interest

The author declares no conflict of interest.

Appendix

A. Proofs for Section 2

Throughout appendices, C denotes a generic constant that does not depend on n and may vary from line to line. Also, $|A| = trace^{1/2}(A'A)$ denotes the norm of a nonrandom matrix A.

Proof of Lemma 1: The first-order conditions with respect to price are the same for both $y_i > \overline{y}$ and $y_i \leq \overline{y}$. They imply the following optimal price: $\frac{\partial \Pi}{\partial p_i} = p_i^{-\nu} s_i^\epsilon - \nu [p_i - c_i] p_i^{-\nu-1} s_i^\epsilon = 0 \Rightarrow p_i^* = \frac{\nu}{\nu-1} c_i$.

First, if $y_i > \overline{y}$, then the first-order conditions with respect to investment imply the following optimal investment: $\frac{\partial \Pi}{\partial y_i} = \epsilon\delta [p_i - c_i] p_i^{-\nu} (y_i + 1)^{\epsilon\delta-1} - 1 = 0 \Rightarrow y_i^* + 1 = Bc_i^{-\tau}$, where $B = \left(\epsilon\delta\nu^{-\nu} (\nu-1)^{(\nu-1)} \right)^{1/(1-\epsilon\delta)}$ and $\tau = (\nu-1)/(1-\epsilon\delta) > 0$ since, by assumption, $0 < \epsilon\delta < 1$ and $\nu > 1$. The value of the profit at the optimal values of price and investment is $\Pi_1 = (1 - \epsilon\delta) (p_i^*)^{1-\nu} (y_i^* + 1)^{\epsilon\delta} /\nu$.

Next, if $y_i \leq \overline{y}$, then $\Pi = [p_i - c_i] p_i^{-\nu} (\overline{y} + 1)^{\epsilon\delta} - y_i$ and, hence, it is optimal to set $y_i^* = 0$. The value of the profit in the second case is $\Pi_2 = \Pi(p_i^*, 0) = (p_i^*)^{1-\nu} (\overline{y} + 1)^{\epsilon\delta} /\nu$. Thus, firm i engages in R&D iff $\Pi_1 > \Pi_2 \Leftrightarrow (y_i^* + 1) > (1 - \epsilon\delta)^{-1/(\epsilon\delta)} (\overline{y} + 1)$, and hence the optimal investment is given by:

$$Y_i = \begin{cases} b_0 - X_i'(\tau\beta) - (\tau\alpha) \sum_{j \in \mathcal{N}_i} E_i(Y_j) - \tau e_i, & \text{if } b_0 - X_i'(\tau\beta) - (\tau\alpha) \sum_{j \in \mathcal{N}_i} E_i(Y_j) - \tau e_i > \gamma_0 \\ 0, & \text{otherwise} \end{cases}$$

where $b_0 = \log B$ and $\gamma_0 = \log \left((1 - \epsilon\delta)^{-1/(\epsilon\delta)} (\overline{y} + 1) \right) > 0$ since $0 < \epsilon\delta < 1$ and $\overline{y} > 0$. ∎

Proof of Lemma 2: First, if $w(y_i + 1) > \overline{c}$, then $U = \frac{w^\delta (y_i+1)^\delta}{\delta} + h_i(T - y_i - 1)$ and the first-order conditions imply the following labor supply: $\frac{\partial U}{\partial y_i} = w^\delta (y_i + 1)^{\delta-1} - h_i = 0 \Rightarrow y_i^* + 1 = w^{\delta/(1-\delta)} h_i^{1/(\delta-1)}$. The optimal utility is $U_1 = \frac{w^\delta (y_i^*+1)^\delta}{\delta} + h_i T - w^\delta (y_i^* + 1)^\delta$.

Next, if $w(y_i + 1) \leq \overline{c}$, then $U = \frac{w^\delta \overline{c}^\delta}{\delta} + h_i(T - y_i)$ and, hence, it is optimal to set $y_i^* = 0$. The optimal utility in the second case is $U_2 = \frac{w^\delta \overline{c}^\delta}{\delta} + h_i T$. Thus, firm i engages in R&D iff $U_1 > U_2 \Leftrightarrow (y_i^* + 1) > (1 - \delta)^{-1/\delta} \overline{c}$, and hence the optimal investment is given by:

$$Y_i = \begin{cases} b_0 - X_i' \frac{\beta}{1-\delta} - \frac{\alpha}{1-\delta} \sum_{j \in \mathcal{N}_i} E_i(Y_j) - \frac{e_i}{1-\delta}, & \text{if } b_0 - X_i' \frac{\beta}{1-\delta} - \frac{\alpha}{1-\delta} \sum_{j \in \mathcal{N}_i} E_i(Y_j) - \frac{e_i}{1-\delta} > \gamma_0 \\ 0, & \text{otherwise} \end{cases}$$

where $b_0 = \log \left(w^{\delta/(1-\delta)} \right)$ and $\gamma_0 = \log \left((1 - \delta)^{-1/\delta} \overline{c} \right) > 0$ since $\overline{c} \geq 1$ and $0 < \delta < 1$. ∎

B. Proofs for Section 3

Proof of Theorem 1: In the following, we suppress dependence of variables on n and write $\widetilde{Y}_i = \widetilde{Y}_{in}$.

To prove the theorem, it suffices to show that the mapping $\mathbf{G} = (G_1, ..., G_n) : \mathbb{R}^n \to \mathbb{R}^n$ is a contraction mapping w.r.t. $\left\{ \widetilde{Y}_i \right\}_{i=1}^n$, with the components given by:

$$
\begin{aligned}
G_i &= G\left(\left(\widetilde{Y}_j \right)_{j \in \mathcal{N}_i}, X_i \right) = \Phi \left(\frac{\sum_{j \in \mathcal{N}_i} \alpha_{j0} \widetilde{Y}_j + X_i' \beta_0 + b_0 - \gamma_0}{\sigma_0} \right) \left(\sum_{j \in \mathcal{N}_i} \alpha_{j0} \widetilde{Y}_j + X_i' \beta_0 + b_0 \right) \\
&\quad + \sigma_0 \phi \left(\frac{\sum_{j \in \mathcal{N}_i} \alpha_{j0} \widetilde{Y}_j + X_i' \beta_0 + b_0 - \gamma_0}{\sigma_0} \right)
\end{aligned}
\tag{B1}
$$

The result would then follow by the Banach Fixed Point Theorem.

To simplify notation, let $z = \left(\sum_{j \in \mathcal{N}_i} \alpha_{j0} \widetilde{Y}_j + X_i' \beta_0 + b_0 - \gamma_0 \right) \sigma_0^{-1}$. Differentiating G_i w.r.t. $\widetilde{Y}_j$ gives

$$
\frac{\partial G_i}{\partial \widetilde{Y}_j} = \frac{\alpha_{j0}}{\sigma_0} \phi(z) (\sigma_0 z + \gamma_0) + \alpha_{j0} \Phi(z) - \alpha_{j0} z \phi(z) = \frac{\alpha_{j0} \gamma_0}{\sigma_0} \phi(z) + \alpha_{j0} \Phi(z)
$$

where we used $\phi'(z) = -z\phi(z)$. It then follows that $\sup \left| \frac{\partial G_i}{\partial \widetilde{Y}_j} \right| \le |\alpha_{j0}| \left[\phi(0) \gamma_0 \sigma_0^{-1} + 1 \right]$. Hence, by the Mean Value Theorem, for any X_i and any $Y, Y' \in \mathbb{R}^n$:

$$
|G(Y', X_i) - G(Y, X_i)| < \sum_{j \in \mathcal{N}_i} \sup \left| \frac{\partial G}{\partial \widetilde{Y}_j} \right| \cdot |Y_j - Y_j'| \le \lambda_0 \sum_{j \in \mathcal{N}_i} |Y_j - Y_j'|, \; \lambda_0 < 1
\tag{B2}
$$

by Assumption 1, where $\lambda_0 = \overline{\alpha} \left[\phi(0) \gamma_0 \sigma_0^{-1} + 1 \right]$ and $\overline{\alpha} = \max_{1 \le j \le k} |\alpha_{j0}|$.

Consequently, the vector mapping $\mathbf{G}$ satisfies the following Lipschitz condition:

$$
\|\mathbf{G}(Y) - \mathbf{G}(Y')\| \le \sum_{i=1}^n |G(Y', X_i) - G(Y, X_i)| \le \lambda \sum_{i=1}^n |Y_i - Y_i'|
$$

with the Lipschitz coefficient $\lambda = k\lambda_0 < 1$, by Assumption 1. ∎

Proof of Theorem 2: The proof is similar to that of Proposition 1 of Jenish [18]. Let $\mathcal{N}_i(m) = \{ j, j \ne i, 1 \le j \le n : \|t(i) - t(j)\| \le m \}$ be the m-neighborhood of agent i that excludes i, and $\mathcal{N}_i^o(m) = \mathcal{N}_i(m) \cup \{i\}$ be the neighborhood of agent i that includes i, where $t(i) \in \mathbb{Z}^d$ is i's location.

Let $\widetilde{Y}_{in} = F_i(X_1, ..., X_n)$, $i = 1, ..., n$, be the unique solution of the system $\widetilde{Y}_{in} = G(\left(\widetilde{Y}_{jn} \right)_{j \in \mathcal{N}_i(r)}, X_i)$, $i = 1, ..., n$, defined in (B1). To simplify, we suppress dependence of variables on n, and write $\widetilde{Y}_i = \widetilde{Y}_{in}$.

Fix i, and define $\eta_i^{(m)} = (X_j)_{j \in \mathcal{N}_i^o(m)}$ and $\zeta_i^{(m)} = (X_j)_{j \in \Lambda_n \setminus \mathcal{N}_i^o(m)}$ for any $m \in \mathbb{N}$, so that we can partition $X = \{X_1, ..., X_n\} = \left(\eta_i^{(m)}; \zeta_i^{(m)} \right)$. Suppose that the underlying probability space is rich enough that there exists a random variable U uniformly distributed on $[0, 1]$ that is independent of $\{X_j\}$. Then, by Lemma A1 in Jenish [18], there exists a function $h(U, \eta_i^{(m)})$ such that the process $X^{(m)} = \left\{ X_j^{(m)} \right\}_{j=1}^n = \left(\eta_i^{(m)}; h(U, \eta_i^{(m)}) \right)$ has the same distribution as X.

We now construct an approximation to $\widetilde{Y}_i$. Define

$$\widetilde{Y}_j^{(m)} = F_j\left(X_1^{(m)}, ..., X_n^{(m)}\right) = F_j\left(\eta_i^{(m)}; h(U, \eta_i^{(m)})\right) \tag{B3}$$

Since $X^{(m)}$ has the same distribution as X, we have $\widetilde{Y}_j^{(m)} = G\left(\left(\left(\widetilde{Y}_l^{(m)}\right)_{l\in\mathcal{N}_j(r)}, X_j^{(m)}\right)\right.$. If $j \in \mathcal{N}_i(m)$, then $X_j^{(m)} = X_j$ and by the Lipschitz condition (B2), we have

$$\left\|\widetilde{Y}_j - \widetilde{Y}_j^{(m)}\right\|_2 \leq \lambda_0 \sum_{l\in\mathcal{N}_j(r)} \left\|\widetilde{Y}_l - \widetilde{Y}_l^{(m)}\right\|_2 \tag{B4}$$

Consider two cases: $m \geq r$ and $m < r$. If $m \geq r$, recursive use of (B4) gives

$$\left\|\widetilde{Y}_i - \widetilde{Y}_i^{(m)}\right\|_2 \leq \lambda_0 \sum_{j_1\in\mathcal{N}_i(r)} \left\|\widetilde{Y}_{j_1} - \widetilde{Y}_{j_1}^{(m)}\right\|_2 \leq \lambda_0^2 \sum_{j_1\in\mathcal{N}_i(r)} \sum_{j_2\in\mathcal{N}_{j_1}(r)} \left\|\widetilde{Y}_{j_2} - \widetilde{Y}_{j_2}^{(m)}\right\|_2$$

$$\leq \quad ... \leq \lambda_0^p \sum_{j_1\in\mathcal{N}_i(r)} ... \sum_{j_p\in\mathcal{N}_{j_{p-1}}(r)} \left\|\widetilde{Y}_{j_p} - \widetilde{Y}_{j_p}^{(m)}\right\|_2 \leq 2\left\|Y\right\|_2 \lambda^p$$

where $\lambda = \lambda_0 k < 1$. If $m < r$, then $p = [m/r] < 1$ and again we will have the same bound on the error:

$$\left\|\widetilde{Y}_i - \widetilde{Y}_i^{(m)}\right\|_2 \leq \lambda_0 \sum_{j_1\in\mathcal{N}_i(r)} \left\|\widetilde{Y}_{j_1} - \widetilde{Y}_{j_1}^{(m)}\right\|_2 \leq 2\left\|Y\right\|_2 \lambda \leq 2\left\|Y\right\|_2 \lambda^p$$

Since i was arbitrary, using the same arguments, we can approximate the entire process $\left\{\widetilde{Y}_i\right\}$ by $\left\{\widetilde{Y}_i^{(m)}\right\}$, which has a similar functional form but a more tractable dependence structure. The approximation error is given by $\sup_{n,i\in\Lambda_n} \left\|\widetilde{Y}_i - \widetilde{Y}_i^{(m)}\right\|_2 \leq 2\left\|Y\right\|_2 \lambda^p$.

We now verify the NED condition $\sup_{n,i\in\Lambda_n} \left\|\widetilde{Y}_i - E(\widetilde{Y}_i|\mathfrak{F}_{in}^X(m))\right\|_2 \to 0$ as $m \to \infty$, where $\mathfrak{F}_{in}^X(m) = \sigma(X_j, j \in \mathcal{N}_i^o(m))$. If $2m >$ diameter(Λ_n), then $\left\|\widetilde{Y}_i - E(\widetilde{Y}_i|\mathfrak{F}_{in}^X(m))\right\|_2 = 0$.

If $2m \leq$ diameter(Λ_n), approximate $\widetilde{Y}_i$ by $\widetilde{Y}_i^{(m)}$ as defined in (B3). Note that by construction, $E[\widetilde{Y}_i|\mathfrak{F}_{in}^X(m)] = E[\widetilde{Y}_i^{(m)}|\mathfrak{F}_{in}^X(m)] = \int_0^1 F_i\left(\eta_i^{(m)}; h(u, \eta_i^{(m)})\right) du$. Then, by the Jensen inequality and independence of U and $\{X_i\}$, we have

$$
\begin{aligned}
\left\|\widetilde{Y}_i - E(\widetilde{Y}_i|\mathfrak{F}_{in}^X(m)))\right\|_2 &= \left\|F_i\left(\eta_i^{(m)}; \zeta_i^{(m)}\right) - \int_0^1 F_i\left(\eta_i^{(m)}; h(u, \eta_i^{(m)})\right) du\right\|_2 \\
&\leq \left\{E\int_0^1 \left|F_i\left(\eta_i^{(m)}; \zeta_i^{(m)}\right) - F_i\left(\eta_i^{(m)}; h(u, \eta_i^{(m)})\right)\right|^2 du\right\}^{1/2} \\
&= \left\|F_i\left(\eta_i^{(m)}; \zeta_i^{(m)}\right) - F_i\left(\eta_i^{(m)}; h(U, \eta_i^{(m)})\right)\right\|_2 \\
&= \left\|\widetilde{Y}_i - \widetilde{Y}_i^{(m)}\right\|_2 \leq \psi(m) = 2\left\|Y\right\|_2 \lambda^{[m/r]} \to 0 \text{ as } m \to \infty
\end{aligned}
$$

We now show that $\{Y_i\}$ is L_2-NED on $\{(X_i, \varepsilon_i)\}$. Let $\mathfrak{F}_{in}(m) = \sigma((X_j, \varepsilon_j), j \in \mathcal{N}_i^o(m))$ and define Z_i as

$$Z_i = g\left(\left(\widetilde{Y}_j\right)_{j\in\mathcal{N}_i(r)}\right) = \begin{cases} b_0 + X_i'\beta_0 + \sum_{j\in\mathcal{N}_i(r)} \alpha_{j0}\widetilde{Y}_j + \varepsilon_i, & b_0 + X_i'\beta_0 + \sum_{j\in\mathcal{N}_i(r)} \alpha_{j0}\widetilde{Y}_j + \varepsilon_i > \gamma_0 \\ \gamma_0, & \text{otherwise} \end{cases}$$

Then, $Y_i = Z_i \mathbf{1}\{Z_i > \gamma_0\}$. The proof will proceed in three steps: (i) verify the NED property of Z_i; (ii) show the NED property of $\mathbf{1}(Y_i > 0) = \mathbf{1}\{Z_i > \gamma_0\}$; and (iii) apply Proposition 3 of Jenish and Prucha [5] to the product of the two processes to show its NED property.

It can be easily verified that $g(y_1, ..., y_k)$ satisfies a Lipschitz inequality: $|g(y_1, ..., y_k) - g(y'_1, ..., y'_k)| \leq \overline{\alpha} \sum_{j=1}^{k} |y_j - y'_j|$, where $\overline{\alpha} = \max_{1 \leq j \leq k} |\alpha_{j0}|$. By the least mean squared property of the conditional mean,

$$
\begin{aligned}
\left\| Z_i - E(Z_i | \mathfrak{F}_{in}(m)) \right\|_2 &\leq \left\| g\left(\widetilde{Y}_{-i}\right) - g\left(E\left(\widetilde{Y}_{-i} | \mathfrak{F}_{in}(m)\right)\right) \right\|_2 \\
&\leq \overline{\alpha} \sum_{j \in \mathcal{N}_i(r)} \left\| \widetilde{Y}_j - E\left[\widetilde{Y}_j | \mathfrak{F}_{in}^X(m)\right] \right\|_2 \leq k\overline{\alpha}\psi(m) \to 0 \text{ as } m \to \infty
\end{aligned}
$$

since $E\left[\widetilde{Y}_j | \mathfrak{F}_{in}(m)\right] = E\left[\widetilde{Y}_j | \mathfrak{F}_{in}^X(m)\right]$ by independence of $\{\varepsilon_i\}$ and $\left\{X_i, \widetilde{Y}_j\right\}$.

Next, let $\varphi(z) = \mathbf{1}\{z > \gamma_0\}$, $a > 0$ be some positive scalar to be chosen later. Define the function:

$$
\varphi_m(z) = \begin{cases}
1, & z > \gamma_0 + \psi^a(m) \\
\psi^{-a}(m)(z - \gamma_0), & \gamma_0 \leq z \leq \gamma_0 + \psi^a(m) \\
0, & z < \gamma_0
\end{cases}
$$

This piecewise linear function converges pointwise to $\mathbf{1}\{z > \gamma_0\}$ as $m \to \infty$; but in contrast to $\mathbf{1}\{z > \gamma_0\}$, it is Lipschitz-continuous with the Lipschitz coefficient $\psi^{-a}(m)$, *i.e.*, for all z_1, z_2 $|\varphi_m(z_1) - \varphi_m(z_2)| \leq \psi^{-a}(m)|z_1 - z_2|$. Moreover, observe that $\sup_z |\varphi(z) - \varphi_m(z)| \leq 1$, and consequently,

$$
\begin{aligned}
\|\varphi(z) - \varphi_m(z)\|_2 &\leq \left[\int \mathbf{1}\{\gamma_0 \leq z \leq \gamma_0 + \psi^a(m)\} \, dF(z) \right]^{1/2} \\
&= \left[F(\gamma_0 + \psi^a(m)) - F(\gamma_0) \right]^{1/2} \leq \left[f(\widetilde{z})\psi^a(m) \right]^{1/2} \leq C\psi^{a/2}(m)
\end{aligned}
$$

where $f(\cdot)$ is the p.d.f. of z and $\widetilde{z} \in (\gamma_0, \gamma_0 + \psi^a(m))$. Using the above inequalities gives

$$
\begin{aligned}
&\|\varphi(Z_i) - E(\varphi(Z_i)|\mathfrak{F}_{in}(m))\|_2 \\
&\leq \|\varphi(Z_i) - \varphi_m(Z_i)\|_2 + \|\varphi_m(Z_i) - E(\varphi_m(Z_i)|\mathfrak{F}_{in}(m))\|_2 + \|E(\varphi(Z_i)|\mathfrak{F}_{in}(m)) - E(\varphi_m(Z_i)|\mathfrak{F}_{in}(m))\|_2 \\
&\leq 2\|\varphi(Z_i) - \varphi_m(Z_i)\|_2 + \|\varphi_m(Z_i) - E(\varphi_m(Z_i)|\mathfrak{F}_{in}(m))\|_2 \leq 2\psi^{a/2}(m) + k\overline{\alpha}\psi^{1-a}(m).
\end{aligned}
$$

Now, minimize the order of magnitude of the variable in the last line by setting $a = 2/3$. Thus, $\mathbf{1}\{Z_i > \gamma_0\} = \mathbf{1}(Y_i > 0)$ is L_2-NED on $\{(X_i, \varepsilon_i)\}$ with the NED numbers $(2 + k\overline{\alpha})\psi^{1/3}(m)$.

Next, observe that $\mathbf{1}(Y_i = 0) = 1 - \mathbf{1}(Y_i > 0)$. Hence, by Proposition 2 of Jenish and Prucha [5], $\{\mathbf{1}(Y_i = 0)\}$ is L_2-NED on $\{(X_i, \varepsilon_i)\}$ with the same NED numbers. Finally, noting that

$$
|z\mathbf{1}\{z > \gamma_0\} - z'\mathbf{1}\{z' > \gamma_0\}| \leq (|z| + \varphi(z'))(|z - z'| + |\varphi(z) - \varphi(z')|)
$$

and using Proposition 3 of Jenish and Prucha [5] with $B(z, z') = |z| + \varphi(z')$ and $r = 3$, we have that $\{Y_i\}$ is L_2-NED on $\{(X_i, \varepsilon_i)\}$ with the NED numbers $c\psi^{1/12}(m)$, where $c = 2\|Y\|_6^{3/2}\|Y\|_2^{1/4}(2 + k\overline{\alpha})^{1/4}$. ∎

C. Proofs for Section 4

Proof of Proposition 1:

The proof of this proposition follows Carson and Sun [19]. The latter paper is not readily applicable to our model since it relies on LLN for independent processes. Let n_0 and n_1 denote, respectively, the sizes of the censored and uncensored subsamples, and let $Y_{i,(0)}$ and $Y_{i,(1)}$ denote observations from the censored and uncensored subsample, respectively. Conditional on the state variables $\overline{X}_n$, the uncensored subsample $\left\{Y_{i,(1)}\right\}$ is independent and follows a truncated normal distribution. Consequently,

$$
\begin{aligned}
P\left(\min\left\{Y_{(1)}\right\} > \gamma + z/n | \overline{X}_n, n_1\right) &= P\left(Y_{1,(1)} > \gamma + z/n, ..., Y_{n_1,(1)} > \gamma + z/n | \overline{X}_n, n_1\right) \\
&= \prod_{i=1}^{n_1}\left(1 - \frac{\Phi\left[\left(\gamma + z/n - Z'_{in,(1)}\delta\right)/\sigma\right] - \Phi\left[\left(\gamma - Z'_{in,(1)}\delta\right)/\sigma\right]}{1 - \Phi\left[\left(\gamma - Z'_{in,(1)}\delta\right)/\sigma\right]}\right) \\
&= \prod_{i=1}^{n_1}\left(1 - \frac{z}{n\sigma}\frac{\phi\left[\left(\gamma - Z'_{in,(1)}\delta\right)/\sigma\right] + o_p(1)}{1 - \Phi\left[\left(\gamma - Z'_{in,(1)}\delta\right)/\sigma\right]}\right) \quad \text{(C1)}
\end{aligned}
$$

where $o_p(1)$ holds uniformly over i since $\phi(\cdot)$ is uniformly continuous on $\mathbb{R}$. Now, let

$$
\mu_{in} = \frac{1}{\sigma}\frac{\phi\left[\left(\gamma - Z'_{in,(1)}\delta\right)/\sigma\right]}{1 - \Phi\left[\left(\gamma - Z'_{in,(1)}\delta\right)/\sigma\right]} \text{ and } \mu = E(\mu_{in})
$$

Next, we establish some inequalities for the r.v. μ_{in}. By Lemma D1, Z_{in} is L_2-NED on $\{X_i\}$ with geometrically decaying coefficients. Then, by Proposition 3 of Jenish and Prucha [5], $\{\mu_{in}\}$ is also L_2-NED on $\{X_i\}$, which is mixing satisfying Assumption 3. Consequently, $\{\mu_{in}\}$ satisfies the LLN of Jenish and Prucha [5], *i.e.*, $\frac{1}{n_1}\sum_{i=1}^{n_1}\mu_{in} = \mu + o_p(1)$. By the Mill's Ratio inequality, $|\phi(z)/(1-\Phi(z))| \leq K|z|$ for some $K < \infty$, and hence, $\sup_{i,n} E\left(|\mu_{in}|^6\right) \leq \overline{M} < \infty$, since $\sup_{i,n} E\left(|Z_{in}|^6\right) < \infty$ by assumption. Next, let $M_k = \max_{1 \leq i \leq k}\mu_{in}$, for $k \leq n$ and define sets $A_1(c) = \{M_1 > c\}$, $A_k(c) = \{M_{k-1} \leq c, \mu_{kn} > c\}$ for $k = 2, ..., n$. Clearly, $\{M_n > c\} \subseteq \cup_{k=1}^{n}A_k(c)$ and $A_k(c) \subseteq \{\mu_{kn} > c\}$. Then, by the Markov inequality,

$$
P\left(\max_{1 \leq i \leq n}\mu_{in} > n^{1/6}C\right) \leq \sum_{k=1}^{n}P\left(A_k(n^{1/6}C)\right) \leq \sum_{k=1}^{n}P\left(\mu_{kn} > n^{1/6}C\right) \leq \overline{M}C^{-6}
$$

which implies $\max_{1 \leq i \leq n}\mu_{in} = O_p\left(n^{1/6}\right)$. Now, using (C1) and the last inequality gives

$$
\begin{aligned}
\frac{1}{n_1}\log P\left(\min\left\{Y_{(1)}\right\} > \gamma + z/n | \overline{X}_n, n_1\right) &= \frac{1}{n_1}\sum_{i=1}^{n_1}\log\left(1 - \frac{z}{n}\mu_{in} + o_p\left(\frac{1}{n}\right)\right) \\
&= -\frac{1}{n_1}\sum_{i=1}^{n_1}\left[\frac{z}{n}\mu_{in} + o_p\left(\frac{1}{n}\right) + O\left(\frac{\mu_{in}^2}{n^2}\right)\right] = \left[-\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{z}{n}\mu_{in}\right] + o_p\left(\frac{1}{n}\right) + O\left(\frac{\max_{1 \leq i \leq n}\mu_{in}^2}{n^2}\right) \\
&= \left[-\frac{z}{n}\frac{1}{n_1}\sum_{i=1}^{n_1}\mu_{in}\right] + o_p\left(\frac{1}{n}\right) + O_p\left(n^{-5/3}\right) = -\frac{z}{n}\left[\mu + o_p(1)\right] + o_p\left(\frac{1}{n}\right) = -\frac{z\mu}{n}\left(1 + o_p(1)\right)
\end{aligned}
$$

Thus, $P\left(\min\left\{Y_{(1)}\right\} > \gamma + z/n|n_1\right) = \exp\left[-\frac{n_1}{n}z\mu\left(1 + o_p\left(1\right)\right)\right]$. Moreover, by Theorem 2, $\mathbf{1}\left(Y_i > \gamma\right)$ is L_2-NED on $\left\{\left(X_i', \varepsilon_i\right)'\right\}$, and hence, satisfies the LLN of Jenish and Prucha [5]:

$$\frac{n_1}{n} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left(Y_i > \gamma\right) \xrightarrow{p} E\left[1 - \Phi\left(\left(\gamma - Z_{in}'\delta\right)/\sigma\right)\right] := \kappa$$

Consequently, $P\left(\min\left\{Y_{(1)}\right\} > \gamma + z/n|n_1\right) \to \exp\left(-az\right)$, where $a = \kappa\mu$. As the right-hand side of the last expression does not depend on n_1, it is also the limit of the unconditional probability. Thus,

$$\lim_{n\to\infty} P\left(n\left(\widehat{\gamma} - \gamma\right) \le z\right) = 1 - \exp\left(-az\right) \text{ for } z > 0 \quad \blacksquare$$

Proof of Theorem 3:

By Proposition 1, γ_0 is identified, so it remains to prove identification of θ by showing that the population objective function $Q\left(\theta, \gamma_0\right) = E\left[m\left(W_{in}, \theta, \gamma_0\right)\right]$ is uniquely maximized at θ_0.

A. Identification in ML

To prove identification in the ML case, it suffices to verify the Kullback-Leibler information inequality, *i.e.*, for all $\theta \in \Theta$ s.t. $\theta \ne \theta_0$

$$\log f\left(Y_i|\overline{X}_n; \theta, \gamma_0\right) \ne \log f_i\left(Y_i|\overline{X}_n; \theta_0, \gamma_0\right) \text{ with positive probability.} \tag{C2}$$

Observe that $\log f\left(Y_i|\overline{X}_n; \theta, \gamma_0\right) - \log f_i\left(Y_i|\overline{X}_n; \theta_0, \gamma_0\right) = \mathbf{1}\left(Y_i = 0\right)\left[\log\Phi\left(\frac{\gamma_0 - Z_{in}'\delta}{\sigma}\right) - \log\Phi\left(\frac{\gamma_0 - Z_{in}'\delta_0}{\sigma_0}\right)\right] - \frac{1}{2}\mathbf{1}\left(Y_i > \gamma_0\right)\left[\log\sigma^2 - \log\sigma_0^2\right] - \mathbf{1}\left(Y_i > \gamma_0\right)\left[\frac{\left(Y_i - Z_{in}'\delta\right)^2}{2\sigma^2} - \frac{\left(Y_i - Z_{in}'\delta_0\right)^2}{2\sigma_0^2}\right]$.

Clearly, $\log f\left(Y_i|\overline{X}_n; \theta, \gamma_0\right) \ne \log f\left(Y_i|\overline{X}_n; \theta_0, \gamma_0\right)$ with positive probability if $\sigma^2 \ne \sigma_0^2$. Suppose the opposite were true, *i.e.*, $\log f\left(Y_i|\overline{X}_n; \theta, \gamma_0\right) = \log f\left(Y_i|\overline{X}_n; \theta_0, \gamma_0\right)$ w.p.1. Then, we would have $\log\frac{\sigma^2}{\sigma_0^2} - \frac{\left(Y_i - Z_{in}'\delta_0\right)^2}{\sigma_0^2} + \frac{\left(Y_i - Z_{in}'\delta\right)^2}{\sigma^2} = 0$ for $Y_i > \gamma_0$, which implies that the r.v. $\left(Y_i - Z_{in}'\delta_0\right)^2$ and $\left(Y_i - Z_{in}'\delta\right)^2$ would have a point mass, which is impossible since at least one of the regressors in X_i has a full support by Assumption 4. So, σ_0^2 is identified, and we can focus on the case: $\sigma^2 = \sigma_0^2$ but $\delta \ne \delta_0$.

If $\delta \ne \delta_0$, then $E\left(Z_{in}'\delta - Z_{in}'\delta_0\right)^2 = E\left[Z_{in}'\left(\delta - \delta_0\right)\right]^2 = \left(\delta - \delta_0\right)'E\left(Z_{in}Z_{in}'\right)\left(\delta - \delta_0\right) > 0$, since $E\left(Z_{in}Z_{in}'\right)$ is positive definite by Assumption 4. Hence, $Z_{in}'\delta \ne Z_{in}'\delta_0$ with positive probability. Consequently, $Y_i - Z_{in}'\delta \ne Y_i - Z_{in}'\delta_0$ with positive probability. Since $Z_{in}'\delta \ne Z_{in}'\delta_0$ with positive probability, and $\Phi\left(\cdot\right)$ and $\log$ are strictly increasing functions, we also have $\log\Phi\left(\frac{\gamma_0 - Z_{in}'\delta}{\sigma_0}\right) \ne \log\Phi\left(\frac{\gamma_0 - Z_{in}'\delta_0}{\sigma_0}\right)$ with positive probability, which proves (C2).

B. Identification in LS

To prove identification in the LS case, it suffices to verify that for all $\theta \in \Theta$ s.t. $\theta \ne \theta_0$

$$\varphi\left(Z_{in}, \theta, \gamma_0\right) \ne \varphi\left(Z_{in}, \theta_0, \gamma_0\right) \text{ with positive probability.} \tag{C3}$$

Recall $\varphi(Z_{in}, \theta, \gamma_0) = \Phi\left(\frac{Z_{in}'\delta - \gamma_0}{\sigma}\right)Z_{in}'\delta + \sigma\phi\left(\frac{Z_{in}'\delta - \gamma_0}{\sigma}\right)$. Re-parametrize: $\upsilon = \delta/\sigma$ and $\sigma = \sigma$. This mapping is one-to-one, and hence its suffices to prove identification of the parameters $\upsilon_0 = \delta_0/\sigma_0$ and σ_0.

By similar arguments as in part A, if $\upsilon \ne \upsilon_0$, then $Z_{in}'\upsilon \ne Z_{in}'\upsilon_0$ with positive probability. Denote $u = Z_{in}'\upsilon$ and $u_0 = Z_{in}'\upsilon_0$, and consider the function $\widetilde{\varphi}(u, \sigma) = \sigma\left[u\Phi\left(u - \gamma_0/\sigma\right) + \phi\left(u - \gamma_0/\sigma\right)\right]$.

It is strictly increasing in σ. To see this, let $\widetilde{\gamma} = \gamma_0/\sigma$ and note that $\frac{\partial \widetilde{\varphi}(u,\sigma)}{\partial \sigma} = u\Phi(u - \widetilde{\gamma}) + (1 + \widetilde{\gamma}^2)\phi(u - \widetilde{\gamma}) > 0$, because if $u < 0$, then $-u \leq (\widetilde{\gamma} - u) < \frac{\phi(\widetilde{\gamma}-u)}{1-\Phi(\widetilde{\gamma}-u)} = \frac{\phi(u-\widetilde{\gamma})}{\Phi(u-\widetilde{\gamma})}$ by the Mill's Ratio inequality since $\widetilde{\gamma} = \gamma_0/\sigma \geq 0$. Moreover, $\widetilde{\varphi}(u, \sigma)$ is also strictly increasing in u:

$$\frac{\partial \widetilde{\varphi}(u, \sigma)}{\partial u} = \sigma\left[\Phi(u - \widetilde{\gamma}) + u\phi(u - \widetilde{\gamma}) - (u - \widetilde{\gamma})\phi(u - \widetilde{\gamma})\right] = \sigma\left[\Phi(u - \widetilde{\gamma}) + \widetilde{\gamma}\phi(u - \widetilde{\gamma})\right] > 0,$$

which proves identification of υ_0 and σ_0, and thus completes the proof of the theorem. ∎

D. Proofs for Section 5

The following lemma collects formulas and some properties of the score and Hessian functions for the ML and LS estimators, which are used throughout the proofs. Let $m_{in} = m(W_{in}, \theta, \gamma)$, $s(W_{in}, \theta, \gamma) = \frac{\partial m_{in}}{\partial \theta} = \left(\frac{\partial m_{in}}{\partial \delta'}; \frac{\partial m_{in}}{\partial \sigma^2}\right)'$ denote the score function, and let the Hessian matrix be denoted by:

$$H(W_{in}, \theta, \gamma) = \frac{\partial^2 m_{in}}{\partial \theta \partial \theta'} = \begin{bmatrix} \frac{\partial^2 m_{in}}{\partial \delta \partial \delta'} & \frac{\partial^2 m_{in}}{\partial \delta \partial \sigma^2} \\ \frac{\partial^2 m_{in}}{\partial \sigma^2 \partial \delta'} & \frac{\partial^2 m_{in}}{\partial (\sigma^2)^2} \end{bmatrix}$$

Lemma D1. *Let* $\phi_{in} = \phi\left(\frac{\gamma - Z'_{in}\delta}{\sigma}\right)$, $\Phi_{in} = \Phi\left(\frac{\gamma - Z'_{in}\delta}{\sigma}\right)$ *and* $\varphi_{in} = \varphi(Z_{in}, \theta, \gamma)$.

A. For the ML estimator, the components of the score and Hessian are given by:

$$\frac{\partial m_{in}}{\partial \delta} = -\mathbf{1}(Y_i = 0)\frac{\phi_{in}}{\sigma\Phi_{in}}Z_{in} + \mathbf{1}(Y_i > \gamma)\frac{1}{\sigma^2}(Y_i - Z'_{in}\delta)Z_{in} \tag{D1}$$

$$\frac{\partial m_{in}}{\partial \sigma^2} = -\mathbf{1}(Y_i = 0)\frac{\phi_{in}}{2\sigma^3\Phi_{in}}(\gamma - Z'_{in}\delta) - \frac{1}{2\sigma^2}\mathbf{1}(Y_i > \gamma) + \frac{1}{2\sigma^4}\mathbf{1}(Y_i > \gamma)(Y_i - Z'_{in}\delta)^2$$

$$\frac{\partial^2 m_{in}}{\partial \delta \partial \delta'} = -\mathbf{1}(Y_i = 0)\frac{\phi_{in}}{\sigma\Phi_{in}^2}\left[\frac{1}{\sigma}\phi_{in} - \frac{1}{\sigma^2}(\gamma - Z'_{in}\delta)\Phi_{in}\right]Z_{in}Z'_{in} - \mathbf{1}(Y_i > \gamma)\frac{1}{\sigma^2}Z_{in}Z'_{in}$$

$$\frac{\partial^2 m_{in}}{\partial \sigma^2 \partial \delta'} = -\mathbf{1}(Y_i = 0)\frac{\phi_{in}}{2\sigma^3\Phi_{in}^2}\left[\frac{1}{\sigma}(\gamma - Z'_{in}\delta)\phi_{in} - \Phi_{in} + \frac{1}{\sigma^2}(\gamma - Z'_{in}\delta)^2\Phi_{in}\right]Z_{in}$$
$$\qquad - \frac{1}{\sigma^4}\mathbf{1}(Y_i > \gamma)(Y_i - Z'_{in}\delta)Z_{in}$$

$$\frac{\partial^2 m_{in}}{\partial(\sigma^2)^2} = \mathbf{1}(Y_i = 0)\frac{\sigma^{-5}\phi_{in}}{4\Phi_{in}^2}\left[\frac{(Z'_{in}\delta - \gamma)^3}{\sigma^2}\Phi_{in} - 3(\gamma - Z'_{in}\delta)\Phi_{in} - \frac{(Z'_{in}\delta - \gamma)^2}{\sigma}\phi_{in}\right]$$
$$\qquad + \frac{1}{2\sigma^4}\mathbf{1}(Y_i > \gamma) - \frac{1}{\sigma^6}\mathbf{1}(Y_i > \gamma)(Y_i - Z'_{in}\delta)^2$$

B. For the LS estimator, the components of the score and Hessian are given by:

$$\frac{\partial m_{in}}{\partial \delta} = 2(Y_i - \varphi_{in})\left[\Phi_{in} + \frac{\gamma}{\sigma}\phi_{in}\right]Z_{in}, \tag{D2}$$

$$\frac{\partial m_{in}}{\partial \sigma^2} = (Y_i - \varphi_{in})\phi_{in}\left[\frac{\gamma(\gamma - Z'_{in}\delta\gamma)}{\sigma^3} + \frac{1}{\sigma}\right]$$

$$\frac{\partial^2 m_{in}}{\partial \delta \partial \delta'} = -2\left[\Phi_{in} + \frac{\gamma}{\sigma}\phi_{in}\right]^2 Z_{in}Z'_{in} + 2(Y_i - \varphi_{in})\frac{\phi_{in}}{\sigma}\left[1 - \frac{(Z'_{in}\delta - \gamma)\gamma}{\sigma^2}\right]Z_{in}Z'_{in}$$

$$\frac{\partial^2 m_{in}}{\partial \sigma^2 \partial \delta'} = -\phi_{in}\left[\left(\Phi_{in} + \frac{\gamma}{\sigma}\phi_{in}\right)\frac{\gamma(\gamma - Z'_{in}\delta\gamma) + \sigma^2}{\sigma^3} - (Y_i - \varphi_{in})\frac{\gamma(\gamma - Z'_{in}\delta\gamma)^2 + \sigma^2(\gamma - Z'_{in}\delta) - \gamma\sigma}{\sigma^4}\right]Z_{in}$$

$$\frac{\partial^2 m_{in}}{\partial(\sigma^2)^2} = -\phi_{in}^2\left[\frac{\gamma^2 - Z'_{in}\delta}{\sigma^3} + \frac{1}{\sigma}\right]^2 + \frac{1}{2}(Y_i - \varphi_{in})\phi_{in}\left[-\frac{(Z'_{in}\delta - \gamma)^3\gamma}{\sigma^7} + \frac{(Z'_{in}\delta - \gamma)(Z'_{in}\delta + 2\gamma)}{\sigma^5} - \frac{1}{\sigma^3}\right]$$

C. Under Assumptions 1–3(i), $\{m(W_{in}, \theta, \gamma)\}$ and $\{H(W_{in}, \theta_0, \gamma_0)\}$ are L_1-NED on $\{(X_i', \varepsilon_i)'\}$ with coefficients decaying at geometric rates, and the score $\{s(W_{in}, \theta_0, \gamma_0)\}$ is L_2-NED on $\{(X_i', \varepsilon_i)'\}$ with coefficients decaying at geometric rates.

The rates of the NED coefficients are required only for verification of the assumptions of the CLT and LLN for NED processes of Jenish and Prucha [5], which rely on polynomial decay rates. Clearly, the NED coefficients declining at any geometric rate will automatically satisfy these theorems, and their exact orders of magnitude are unimportant for the proofs our results.

Proof of Lemma D1: Parts A and B follow by straightforward differentiation. To show part C, observe that by Theorem 2 and Proposition 3 of Jenish and Prucha [5], $\{\mathbf{1}(Y_{in} > \gamma_0)\}$, $\{\mathbf{1}(Y_{in} = 0)\}$, $\left\{\Phi\left(\frac{\gamma - Z_{in}'\delta}{\sigma}\right)\right\}$, $\left\{\phi\left(\frac{\gamma - Z_{in}'\delta}{\sigma}\right)\right\}$ and $\{(Y_i - Z_{in}'\delta)^2\}$ are L_2-NED on $\{(X_i', \varepsilon_i)'\}$ with the NED coefficients decaying at geometric rates. Then, by Theorem 17.9 of Davidson [21], all products and sums of L_2-NED terms are L_1-NED variables with the NED coefficients decaying at the slowest rate of the multiples or summands. Thus, $\{m_{in}\}$ is L_1-NED on $\{(X_i', \varepsilon_i)'\}$ for both the ML and LS estimators. B analogous arguments, the Hessians $H(W_{in}, \theta_0, \gamma_0)$ are L_1-NED on $\{(X_i', \varepsilon_i)'\}$ for both the ML and LS estimators. Finally, the score of the ML estimator is L_2-NED on $\{(X_i', \varepsilon_i)'\}$ with geometric decay rates by Example 17.17 of Davidson [21] as the sum of products of $\mathbf{1}(Y_i = 0)$ or $\mathbf{1}(Y_i > \gamma)$, which are bounded and L_2-NED, and some smooth functions, each of which is L_2-NED on $\{(X_i', \varepsilon_i)'\}$ with geometric decay rates. By analogous arguments, the score of the LS estimator is also L_2-NED on $\{(X_i', \varepsilon_i)'\}$ with geometric decay rates. ∎

Proof of Theorem 4: We first show consistency. Let $Q(\theta, \gamma) = E[m(W_{in}, \theta, \gamma)]$. To prove consistency of $\widehat{\theta}_n$, it suffices to verify assumptions of Lemma A-1 of Andrews [22], namely: (a) $\sup_{\theta \in \Theta} |Q_n(\theta, \widehat{\gamma}) - Q(\theta, \gamma_0)| \xrightarrow{p} 0$, and (b) for every neighborhood Θ_0 of θ, $\sup_{\theta \in \Theta/\Theta_0} Q(\theta, \gamma_0) < Q(\theta_0, \gamma_0)$. Condition (b) is satisfied both for the ML and LS estimators by the identification theorem, Theorem 3. So, it only remains to verify condition (a).

A. We first verify condition (a) for the ML estimator. Note that

$$\sup_{\theta \in \Theta} |Q_n(\theta, \widehat{\gamma}) - Q(\theta, \gamma_0)| \le \sup_{\theta \in \Theta} |Q_n(\theta, \gamma_0) - Q(\theta, \gamma_0)| + \sup_{\theta \in \Theta} |Q_n(\theta, \widehat{\gamma}) - Q_n(\theta, \gamma_0)| \quad \text{(D3)}$$

We now show that both terms on the r.h.s. of this inequality go to zero in probability. Consider the first term. By Lemma D1, $\{m(W_{in}, \theta, \gamma_0)\}$ is L_1-NED on $(X_i', \varepsilon_i)'$ with the NED coefficients decaying at a geometric rate, and $(X_i', \varepsilon_i)'$ is α-mixing satisfying Assumption 3(ii). Moreover, $E|m(W_{in}, \theta, \gamma)| < \infty$. Then, by Theorem 1 of Jenish and Prucha [5], $\{m(W_{in}, \theta, \gamma_0)\}$ satisfies a pointwise LLN on Θ. Next, since Θ is compact, and $m(W_{in}, \theta, \gamma_0)$ is continuously differentiable in θ, $m(W_{in}, \theta, \gamma_0)$ satisfies a Lipschitz condition in θ with a coefficient satisfying: $\sup_{i,n} \sup_{\Theta} E\left|\frac{\partial m(W_{in}, \theta, \gamma_0)}{\partial \theta}\right| < \infty$, by Assumption 5. Thus, $\{m(W_{in}, \theta, \gamma_0)\}$ is L_1-stochastically equicontinuous on Θ, and hence, by the ULLN of Jenish and Prucha [23], the first term on the r.h.s. of (D3) converges to zero.

Next, write out the second term in (D3) as

$$
\begin{aligned}
Q_n\left(\theta,\widehat{\gamma}\right) - Q_n\left(\theta,\gamma_0\right) &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left(Y_i=0\right)\left[\log\Phi\left(\frac{\widehat{\gamma}-Z_{in}'\delta}{\sigma}\right) - \log\Phi\left(\frac{\gamma_0-Z_{in}'\delta}{\sigma}\right)\right] \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\log\left[\sigma^{-1}\phi\left(\frac{Y_i-Z_{in}'\delta}{\sigma}\right)\right]\left[\mathbf{1}\left(Y_i>\widehat{\gamma}\right) - \mathbf{1}\left(Y_i>\gamma_0\right)\right]
\end{aligned}
$$

By construction, $\widehat{\gamma} = \min\left\{Y_i : Y_i > 0\right\} \geq \gamma_0$, and hence,

$$
\mathbf{1}\left(Y_i>\widehat{\gamma}\right) - \mathbf{1}\left(Y_i>\gamma_0\right) = \begin{cases} 0, & \text{if } Y_i \neq Y_{(1)} = \widehat{\gamma} \\ -\mathbf{1}\left(Y_i>\gamma_0\right), & \text{if } Y_i = \min\left\{Y_i : Y_i > 0\right\} = \widehat{\gamma} \end{cases} \tag{D4}
$$

Note that the minimal value in the uncensored subsample, $\{Y_i > 0\}$, is attained only by a single observation in the subsample. This is because the variables $\{Y_i\}$ are i.i.d. continuously distributed on $(\gamma_0, +\infty)$ so that the probability of the event $Y_i = Y_j$, $i \neq j$, is zero. Then,

$$
\begin{aligned}
Q_n\left(\theta,\widehat{\gamma}\right) - Q_n\left(\theta,\gamma_0\right) &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left(Y_i=0\right)\left[\log\Phi\left(\frac{\widehat{\gamma}-Z_{in}'\delta}{\sigma}\right) - \log\Phi\left(\frac{\gamma_0-Z_{in}'\delta}{\sigma}\right)\right] \\
&\quad - \frac{1}{n}\mathbf{1}\left(\widehat{\gamma}>\gamma_0\right)\log\left[\sigma^{-1}\phi\left(\frac{\widehat{\gamma}-Z_{in,(1)}'\delta}{\sigma}\right)\right]
\end{aligned}
$$

Since $\log\Phi\left(z\right)$ is continuously differentiable on $\mathbb{R}$, we have by the Mean Value Theorem:

$$
\log\Phi\left(\frac{\widehat{\gamma}-Z_{in}'\delta}{\sigma}\right) - \log\Phi\left(\frac{\gamma_0-Z_{in}'\delta}{\sigma}\right) = \frac{\phi\left(\frac{\widetilde{\gamma}-Z_{in}'\delta}{\sigma}\right)}{\sigma\Phi\left(\frac{\widetilde{\gamma}-Z_{in}'\delta}{\sigma}\right)}\left(\widehat{\gamma}-\gamma_0\right)
$$

where $\widetilde{\gamma}$ is between $\widehat{\gamma}$ and γ_0, and $\widetilde{\gamma} \xrightarrow{p} \gamma_0$. Then,

$$
\begin{aligned}
\sup_{\theta\in\Theta}\left|Q_n\left(\theta,\widehat{\gamma}\right) - Q_n\left(\theta,\gamma_0\right)\right| &\leq \left|\widehat{\gamma}-\gamma_0\right|\frac{1}{n}\left\{\sum_{i=1}^{n}\sup_{\theta\in\Theta}\left|\frac{\phi\left(\frac{\widetilde{\gamma}-Z_{in}'\delta}{\sigma}\right)}{\sigma\Phi\left(\frac{\widetilde{\gamma}-Z_{in}'\delta}{\sigma}\right)}\right|\right\} \\
&\quad + \frac{1}{n}\mathbf{1}\left(\widehat{\gamma}>\gamma_0\right)\sup_{\theta\in\Theta}\left|\log\left[\sigma^{-1}\phi\left(\frac{\widehat{\gamma}-Z_{in,(1)}'\delta}{\sigma}\right)\right]\right| \xrightarrow{p} 0
\end{aligned}
$$

since $\widehat{\gamma} \xrightarrow{p} \gamma_0$, $\sup_{x\in\mathbb{R}}\frac{\phi(x)}{\Phi(x)} \leq C$, $\sigma \in [\sigma_1,\sigma_2]$, $\sigma_1 > 0$, $E\left[\sup_\delta\left(Y_{in}-Z_{in}'\delta\right)^2\right] < \infty$, by Assumption 5, and by Proposition 1: $E\left(\mathbf{1}\left(\widehat{\gamma}>\gamma_0\right)\right) = P\left(\widehat{\gamma}>\gamma_0\right) = P\left(n\left(\widehat{\gamma}-\gamma_0\right)>0\right) \to 1$. Thus, $\sup_{\theta\in\Theta}\left|Q_n\left(\theta,\widehat{\gamma}\right) - Q\left(\theta,\gamma_0\right)\right| \xrightarrow{p} 0$, and hence, the ML estimator $\widehat{\theta}_n \xrightarrow{p} \theta_0$.

B. We now verify condition (a) for the LS estimator. Using similar arguments as in part A, $\{m\left(W_{in},\theta,\gamma_0\right)\}$ is L_1-NED and L_1-stochastically equicontinuous on Θ, and hence, by the ULLN of Jenish and Prucha [23], the first term on the r.h.s. of (D3) converges to zero.

As for the second term, observe that by the Mean Value Theorem:

$$Q_n(\theta, \widehat{\gamma}) - Q_n(\theta, \gamma_0) = \frac{1}{n} \sum_{i=1}^{n} [m(W_{in}, \theta, \widehat{\gamma}) - m(W_{in}, \theta, \gamma_0)]$$

$$= \frac{1}{n} \sum_{i=1}^{n} [\varphi(W_{in}, \theta, \gamma_0) - \varphi(W_{in}, \theta, \widehat{\gamma})] [2Y_i + \varphi(\theta, \gamma_0) + \varphi(\theta, \widehat{\gamma})]$$

$$= (\widehat{\gamma} - \gamma_0) \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{\gamma_0}{\sigma} \phi\left(\frac{Z'_{in}\delta - \widetilde{\gamma}}{\sigma}\right) [2Y_i + \varphi(\theta, \gamma_0) + \varphi(\theta, \widehat{\gamma})] \right\} \overset{p}{\to} 0$$

since $\widehat{\gamma} \overset{p}{\to} \gamma_0$ and the r.v. in braces is $O_p(1)$, where $\varphi(W_{in}, \theta, \gamma) = \Phi\left(\frac{Z'_{in}\delta - \gamma}{\sigma}\right) Z'_{in}\delta + \sigma\phi\left(\frac{Z'_{in}\delta - \gamma}{\sigma}\right)$ and $\frac{\partial \varphi}{\partial \gamma} = \gamma\phi\left((Z'_{in}\delta - \gamma)/\sigma\right)/\sigma$. This completes the proof of consistency

Proof of asymptotic normality: Taking the mean value expansion of the first-order conditions around θ_0 gives:

$$0 = \frac{\partial Q_n(\widehat{\theta}, \widehat{\gamma})}{\partial \theta} = \frac{1}{n} \sum_{i=1}^{n} s(W_{in}, \theta_0, \widehat{\gamma}) + \left[\frac{1}{n} \sum_{i=1}^{n} H\left(W_{in}, \widetilde{\theta}, \widehat{\gamma}\right)\right] \left(\widehat{\theta}_n - \theta_0\right)$$

where $\widetilde{\theta}$ lies between $\widehat{\theta}$ and θ_0, and $\widehat{\theta}_n \overset{p}{\to} \theta_0$. For both ML and LS estimators, we verify below that

$$\frac{1}{n} \sum_{i=1}^{n} H\left(W_{in}, \widetilde{\theta}, \widehat{\gamma}\right) \overset{p}{\to} H_0 = E\left[H(W_{in}, \theta_0, \gamma_0)\right] \tag{D5}$$

where H_0 is nonsingular by assumption. Then,

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) = -H_0^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [s(W_{in}, \theta_0, \widehat{\gamma}) - Es(W_{in}, \theta_0, \widehat{\gamma})] - H_0^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Es(W_{in}, \theta_0, \widehat{\gamma}) + o_p(1)$$

Define: $\nu_n(\gamma) = n^{-1/2} \sum_{i=1}^{n} [s(W_{in}, \theta_0, \gamma) - Es(W_{in}, \theta_0, \gamma)]$, and re-write

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) = -H_0^{-1}\nu_n(\gamma_0) - H_0^{-1}[\nu_n(\widehat{\gamma}) - \nu_n(\gamma_0)] - H_0^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Es(W_{in}, \theta_0, \widehat{\gamma}) + o_p(1)$$

By Lemma D1, $\{s(W_{in}, \theta_0, \gamma_0)\}$ is L_2-NED on $\{(X'_i, \varepsilon_i)'\}$ with the NED coefficients decaying at a geometric rate both in the case of the ML and LS estimators. In turn, $\{(X'_i, \varepsilon_i)'\}$ is mixing with α-mixing coefficients satisfying Assumption 3. Then, by the CLT of Jenish and Prucha [5], $\nu_n(\gamma_0) \overset{d}{\to} N(0, S_0)$, where $S_0 = Var[s(W_{in}, \theta_0, \gamma_0)]$. So to prove the theorem, it remains to verify (D5) and show that

$$\nu_n(\widehat{\gamma}) - \nu_n(\gamma_0) \overset{p}{\to} 0, \tag{D6}$$

$$n^{-1/2} \sum_{i=1}^{n} Es(W_{in}, \theta_0, \widehat{\gamma}) \overset{p}{\to} 0 \tag{D7}$$

A. *ML estimator*

We first verify (D7). Note that by the population log likelihood equality

$$Es(W_{in}, \theta_0, \gamma_0) = E\left[\frac{\partial}{\partial \theta} \log f(W_{in}, \theta_0, \gamma_0)\right] = 0 \tag{D8}$$

Next partition $Es\left(W_{in}, \theta_0, \gamma\right) = \left[E \frac{\partial}{\partial \delta'} m(W_{in}, \theta_0, \gamma); E \frac{\partial}{\partial \sigma^2} m(W_{in}, \theta_0, \gamma)\right]'$ and use formulas (D1) to find the conditional expectations for $\gamma \geq \gamma_0$:

$$
\begin{aligned}
h_\delta(W_{in}, \theta_0, \gamma) &\equiv E\left[\frac{\partial}{\partial \delta} m(W_{in}, \theta_0, \gamma)|\overline{X}_n\right] \\
&= -\Phi\left(\frac{\gamma_0 - Z_{in}'\delta_0}{\sigma_0}\right) \frac{\phi\left(\frac{\gamma - Z_{in}'\delta_0}{\sigma_0}\right)}{\sigma_0 \Phi\left(\frac{\gamma - Z_{in}'\delta_0}{\sigma_0}\right)} Z_{in} + \frac{Z_{in}}{\sigma_0^2} E\left[\varepsilon_i \mathbf{1}\left(\varepsilon_i > \gamma - Z_{in}'\delta_0\right)|\overline{X}_n\right] \\
&= -\Phi\left(\frac{\gamma_0 - Z_{in}'\delta_0}{\sigma_0}\right) \frac{\phi\left(\frac{\gamma - Z_{in}'\delta_0}{\sigma_0}\right)}{\sigma_0 \Phi\left(\frac{\gamma - Z_{in}'\delta_0}{\sigma_0}\right)} Z_{in} + \frac{\phi\left(\frac{Z_{in}'\delta_0 - \gamma}{\sigma_0}\right)}{\Phi\left(\frac{Z_{in}'\delta_0 - \gamma}{\sigma_0}\right)} \frac{Z_{in}}{\sigma_0}
\end{aligned}
$$

$$
\begin{aligned}
h_\sigma(W_{in}, \theta_0, \gamma) &\equiv E\left[\frac{\partial}{\partial \sigma^2} m(W_{in}, \theta_0, \gamma)|\overline{X}_n\right] = -E\left[\mathbf{1}\left(Y_i = 0\right)|\overline{X}_n\right] \frac{\phi\left(\frac{\gamma - Z_{in}'\delta_0}{\sigma_0}\right)}{2\sigma_0^3 \Phi\left(\frac{\gamma - Z_{in}'\delta_0}{\sigma_0}\right)} (\gamma - Z_{in}'\delta_0) \\
&\quad - \frac{1}{2\sigma_0^2} E\left[\mathbf{1}\left(Y_i > \gamma\right)|\overline{X}_n\right] + \frac{1}{2\sigma_0^4} E\left[\mathbf{1}\left(Y_i > \gamma\right)\left(Y_i - Z_{in}'\delta_0\right)^2|\overline{X}_n\right] \\
&= -\Phi\left(\frac{\gamma_0 - Z_{in}'\delta_0}{\sigma_0}\right) \frac{\phi\left(\frac{\gamma - Z_{in}'\delta_0}{\sigma_0}\right)}{2\sigma_0^3 \Phi\left(\frac{\gamma - Z_{in}'\delta_0}{\sigma_0}\right)} (\gamma - Z_{in}'\delta_0) - \\
&\quad - \frac{1}{2\sigma_0^2} \Phi\left(\frac{Z_{in}'\delta_0 - \gamma}{\sigma_0}\right) + \frac{1}{2\sigma_0^2} Z_{in}\left[1 + (\gamma - Z_{in}'\delta_0) \frac{\phi\left(\frac{Z_{in}'\delta_0 - \gamma}{\sigma_0}\right)}{\Phi\left(\frac{Z_{in}'\delta_0 - \gamma}{\sigma_0}\right)}\right]
\end{aligned}
$$

since by Amemiya [24], $E\left[\varepsilon_i^2 \mathbf{1}\left(\varepsilon_i > \gamma - Z_{in}'\delta_0\right)|\overline{X}_n\right] = \sigma_0^2\left[1 - (Z_{in}'\delta_0 - \gamma)\phi\left(\frac{Z_{in}'\delta_0 - \gamma}{\sigma_0}\right)/\Phi\left(\frac{Z_{in}'\delta_0 - \gamma}{\sigma_0}\right)\right]$.

By the Law of Iterated Expectations, $Es\left(W_{in}, \theta_0, \gamma\right) = \left(E\left[h_\delta(W_{in}, \theta_0, \gamma)\right]', E\left[h_\sigma(W_{in}, \theta_0, \gamma)\right]\right)'$.

Next, observe that $h_\delta(W_{in}, \theta_0, \gamma)$ and $h_\sigma(W_{in}, \theta_0, \gamma)$ are continuous in γ uniformly on Γ, since Γ is compact. Moreover, $h_\delta(W_{in}, \theta_0, \gamma)$ and $h_\sigma(W_{in}, \theta_0, \gamma)$ both satisfy domination conditions in γ in the ϵ-neighborhood of γ_0. Then, since $\widehat{\gamma} \geq \gamma_0$ and $\widehat{\gamma} \downarrow \gamma_0$ in probability,

$$
\begin{aligned}
p \lim_{n\to\infty} E\left[h_\delta(W_{in}, \theta_0, \widehat{\gamma})\right] &= E\left[p \lim_{n\to\infty} h_\delta(W_{in}, \theta_0, \widehat{\gamma})\right] = E\left[h_\delta(W_{in}, \theta_0, \gamma_0)\right] = 0 \\
p \lim_{n\to\infty} E\left[h_\sigma(W_{in}, \theta_0, \widehat{\gamma})\right] &= E\left[p \lim_{n\to\infty} h_\sigma(W_{in}, \theta_0, \widehat{\gamma})\right] = E\left[h_\sigma(W_{in}, \theta_0, \gamma_0)\right] = 0
\end{aligned}
$$

by (D8), which verifies (D7). We next verify (D6). Write

$$
\nu_n(\widehat{\gamma}) - \nu_n(\gamma_0) = n^{-1/2} \sum_{i=1}^n \left[s\left(W_{in}, \theta_0, \widehat{\gamma}\right) - s\left(W_{in}, \theta_0, \gamma_0\right)\right] - n^{-1/2} \sum_{i=1}^n Es\left(W_{in}, \theta_0, \widehat{\gamma}\right)
$$

since $Es\left(W_{in}, \theta_0, \gamma_0\right) = 0$. It was already shown that $n^{-1/2} \sum_{i=1}^{n} Es\left(W_{in}, \theta_0, \widehat{\gamma}\right) \xrightarrow{p} 0$. It remains to show that $n^{-1/2} \sum_{i=1}^{n} \left[s\left(W_{in}, \theta_0, \widehat{\gamma}\right) - s\left(W_{in}, \theta_0, \gamma_0\right)\right] \xrightarrow{p} 0$. Partition $s\left(W_{in}, \theta_0, \widehat{\gamma}\right) - s\left(W_{in}, \theta_0, \gamma_0\right) = \left(s_{1i}\left(\widehat{\gamma}, \gamma_0\right)', s_{2i}\left(\widehat{\gamma}, \gamma_0\right)\right)'$, where

$$
\begin{aligned}
s_{1i}\left(\widehat{\gamma}, \gamma_0\right) &= \sigma_0^{-1} \mathbf{1}\left(Y_i = 0\right) Z_{in} \left[\frac{\phi\left(\frac{\gamma_0 - Z_{in}'\delta_0}{\sigma_0}\right)}{\Phi\left(\frac{\gamma_0 - Z_{in}'\delta_0}{\sigma_0}\right)} - \frac{\phi\left(\frac{\widehat{\gamma} - Z_{in}'\delta_0}{\sigma_0}\right)}{\Phi\left(\frac{\widehat{\gamma} - Z_{in}'\delta_0}{\sigma_0}\right)}\right] \\
&\quad + \frac{1}{\sigma_0^2}\left(Y_i - Z_{in}'\delta_0\right) Z_{in} \left[\mathbf{1}\left(Y_i > \widehat{\gamma}\right) - \mathbf{1}\left(Y_i > \gamma_0\right)\right];
\end{aligned}
$$

$$
\begin{aligned}
s_{2i}\left(\widehat{\gamma}, \gamma_0\right) &= -\frac{1}{2\sigma_0^3} \mathbf{1}\left(Y_i = 0\right) \left[\frac{\phi\left(\frac{\widehat{\gamma} - Z_{in}'\delta_0}{\sigma_0}\right)}{\Phi\left(\frac{\widehat{\gamma} - Z_{in}'\delta_0}{\sigma_0}\right)}\left(\widehat{\gamma} - Z_{in}'\delta_0\right) - \frac{\phi\left(\frac{\gamma_0 - Z_{in}'\delta_0}{\sigma_0}\right)}{\Phi\left(\frac{\gamma_0 - Z_{in}'\delta_0}{\sigma_0}\right)}\left(\gamma_0 - Z_{in}'\delta_0\right)\right] \\
&\quad - \frac{1}{2\sigma_0^2}\left[\mathbf{1}\left(Y_i > \widehat{\gamma}\right) - \mathbf{1}\left(Y_i > \gamma_0\right)\right] + \frac{1}{2\sigma_0^4}\left(Y_i - Z_{in}'\delta_0\right)^2 \left[\mathbf{1}\left(Y_i > \widehat{\gamma}\right) - \mathbf{1}\left(Y_i > \gamma_0\right)\right]
\end{aligned}
$$

Using (D4) and similar arguments as in the proof of consistency, we have

$$
\begin{aligned}
n^{-1/2} \sum_{i=1}^{n} s_{1i}\left(\widehat{\gamma}, \gamma_0\right) &= n^{-1/2} \sum_{i=1}^{n} \sigma_0^{-1} \mathbf{1}\left(Y_i = 0\right) Z_{in} \left[\frac{\phi\left(\frac{\gamma_0 - Z_{in}'\delta_0}{\sigma_0}\right)}{\Phi\left(\frac{\gamma_0 - Z_{in}'\delta_0}{\sigma_0}\right)} - \frac{\phi\left(\frac{\widehat{\gamma} - Z_{in}'\delta_0}{\sigma_0}\right)}{\Phi\left(\frac{\widehat{\gamma} - Z_{in}'\delta_0}{\sigma_0}\right)}\right] \\
&\quad - n^{-1/2} \frac{1}{\sigma_0^2}\left(\widehat{\gamma} - Z_{in,(1)}'\delta_0\right) Z_{in,(1)} \mathbf{1}\left(\widehat{\gamma} > \gamma_0\right) := A_{1n} + A_{2n},
\end{aligned}
$$

$$
\begin{aligned}
n^{-1/2} \sum_{i=1}^{n} s_{2i}\left(\widehat{\gamma}, \gamma_0\right) &= -\frac{1}{2\sigma_0^3} n^{-1/2} \sum_{i=1}^{n} \mathbf{1}\left(Y_i = 0\right) \left[\frac{\phi\left(\frac{\widehat{\gamma} - Z_{in}'\delta_0}{\sigma_0}\right)}{\Phi\left(\frac{\widehat{\gamma} - Z_{in}'\delta_0}{\sigma_0}\right)}\left(\widehat{\gamma} - Z_{in}'\delta_0\right) - \frac{\phi\left(\frac{\gamma_0 - Z_{in}'\delta_0}{\sigma_0}\right)}{\Phi\left(\frac{\gamma_0 - Z_{in}'\delta_0}{\sigma_0}\right)}\left(\gamma_0 - Z_{in}'\delta_0\right)\right] \\
&\quad + \frac{1}{2\sigma_0^2} n^{-1/2} \mathbf{1}\left(\widehat{\gamma} > \gamma_0\right) - \frac{1}{2\sigma_0^4} n^{-1/2}\left(\widehat{\gamma} - Z_{in,(1)}'\delta_0\right)^2 \mathbf{1}\left(\widehat{\gamma} > \gamma_0\right)
\end{aligned}
$$

Since the function $g(z) = \phi(z) / \Phi(z)$ is continuously differentiable on $\mathbb{R}$, we have by the Mean Value Theorem: $g\left(\frac{\widehat{\gamma} - Z_{in}'\delta_0}{\sigma_0}\right) - g\left(\frac{\gamma_0 - Z_{in}'\delta_0}{\sigma_0}\right) = \sigma_0^{-1} g'\left(\frac{\widetilde{\gamma} - Z_{in}'\delta_0}{\sigma_0}\right)\left(\widehat{\gamma} - \gamma_0\right)$, where $\widetilde{\gamma}$ is between $\widehat{\gamma}$ and γ_0, and $\widetilde{\gamma} \xrightarrow{p} \gamma_0$. Hence, the first term in $n^{-1/2} \sum_{i=1}^{n} s_{1i}\left(\widehat{\gamma}, \gamma_0\right)$ satisfies

$$
A_{1n} = n^{-1/2}\left[n\left(\widehat{\gamma} - \gamma_0\right)\right] n^{-1} \sum_{i=1}^{n} \mathbf{1}\left(Y_i = 0\right) Z_{in} g'\left(\frac{\widetilde{\gamma} - Z_{in}'\delta_0}{\sigma_0}\right) = n^{-1/2} O_p(1) O_p(1) = o_p(1)
$$

since $n\left(\widehat{\gamma} - \gamma_0\right) = O_p(1)$ by Proposition 1. The second term satisfies

$$
E\left|A_{2n}\right| \leq n^{-1/2} \left[E\left|\left(\widehat{\gamma} - Z_{in,(1)}'\delta_0\right) Z_{in,(1)}\right|^2\right]^{1/2} P^{1/2}\left(n\left(\widehat{\gamma} - \gamma_0\right) > 0\right) \to 0
$$

since $P\left(n\left(\widehat{\gamma} - \gamma_0\right) > 0\right) \to 1$ by Proposition 1. Hence, $n^{-1/2} \sum_{i=1}^{n} s_{1i}\left(\widehat{\gamma}, \gamma_0\right) = o_p(1)$. Similarly, $n^{-1/2} \sum_{i=1}^{n} s_{2i}\left(\widehat{\gamma}, \gamma_0\right) = o_p(1)$. Thus, $n^{-1/2} \sum_{i=1}^{n} \left[s\left(W_{in}, \theta_0, \widehat{\gamma}\right) - s\left(W_{in}, \theta_0, \gamma_0\right)\right] = o_p(1)$, proving (D6).

Finally, we verify (D5). Write

$$
\left|\frac{1}{n} \sum_{i=1}^{n} H_{in}\left(\widetilde{\theta}, \widehat{\gamma}\right) - H_0\right| \leq \left|\frac{1}{n} \sum_{i=1}^{n} \left[H_{in}\left(\widetilde{\theta}, \widehat{\gamma}\right) - H_{in}\left(\theta_0, \gamma_0\right)\right]\right| + \left|\frac{1}{n} \sum_{i=1}^{n} H_{in}\left(\theta_0, \gamma_0\right) - H_0\right|
$$

We need to show that each of the terms on the r.h.s. of the last inequality is $o_p(1)$. The second term is $o_p(1)$ by the LLN of Jenish and Prucha [5] since $H(W_{in}, \theta_0, \gamma_0)$ is L_1-NED on $(X'_i, \varepsilon_i)'$.

Now consider the first term. Using formulas (D1), we have

$$
\frac{\partial^2 m_{in}\left(\widetilde{\theta}, \widehat{\gamma}\right)}{\partial\delta\partial\delta'} - \frac{\partial^2 m_{in}(\theta_0, \gamma_0)}{\partial\delta\partial\delta'} = -\mathbf{1}(Y_i = 0)\left[g_{1in}\left(\widetilde{\theta}, \widehat{\gamma}\right) - g_{1in}(\theta_0, \gamma_0)\right] Z_{in}Z'_{in}
$$
$$
- \left[\mathbf{1}(Y_i > \widehat{\gamma}) - \mathbf{1}(Y_i > \gamma_0)\right]\frac{1}{\sigma^2} Z_{in}Z'_{in}
$$

where $g_{1in}(\theta, \gamma) = \dfrac{\phi\left(\frac{\gamma - Z'_{in}\delta}{\sigma}\right)}{\sigma\Phi^2\left(\frac{\gamma - Z'_{in}\delta}{\sigma}\right)}\left[\frac{1}{\sigma}\phi\left(\frac{\gamma - Z'_{in}\delta}{\sigma}\right) - \frac{1}{\sigma^2}\Phi\left(\frac{\gamma - Z'_{in}\delta}{\sigma}\right)(\gamma - Z'_{in}\delta)\right]$ is continuously

differentiable in (θ', γ), and hence by the Mean Value Theorem

$$
g_{1in}\left(\widetilde{\theta}, \widehat{\gamma}\right) - g_{1in}(\theta_0, \gamma_0) = \frac{\partial g_{1in}(\theta^*, \gamma^*)}{\partial\theta'}\left(\widetilde{\theta} - \theta_0\right) + \frac{\partial g_{1in}(\theta^*, \gamma^*)}{\partial\gamma}\left(\widehat{\gamma} - \gamma_0\right)
$$

where (θ^*, γ^*) lie between $\left(\widetilde{\theta}', \widehat{\gamma}\right)'$ and $(\theta'_0, \gamma_0)'$. Then, using (D4) and similar arguments as in the proof of consistency, we have

$$
\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 m_{in}\left(\widetilde{\theta}, \widehat{\gamma}\right)}{\partial\delta\partial\delta'} - \frac{\partial^2 m_{in}(\theta_0, \gamma_0)}{\partial\delta\partial\delta'}\right| \leq \left|\widetilde{\theta} - \theta_0\right|\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\partial g_{1in}(\theta^*, \gamma^*)}{\partial\theta'}\right||Z_{in}Z'_{in}|
$$
$$
+ |\widehat{\gamma} - \gamma_0|\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\partial g_{1in}(\theta^*, \gamma^*)}{\partial\gamma}\right||Z_{in}Z'_{in}| + \frac{1}{n}\mathbf{1}(\widehat{\gamma} > \gamma_0)\sigma^{-2}\left|Z_{in,(1)}Z'_{in,(1)}\right| \xrightarrow{p} 0
$$

since $\left|\widetilde{\theta} - \theta_0\right| \xrightarrow{p} 0$, $|\widehat{\gamma} - \gamma_0| \xrightarrow{p} 0$, and $E[\mathbf{1}(\widehat{\gamma} > \gamma_0)] = P(n(\widehat{\gamma} - \gamma_0) > 0) \to 1$. By similar arguments,

$$
\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 m_{in}\left(\widetilde{\theta}, \widehat{\gamma}\right)}{\partial(\sigma^2)^2} - \frac{\partial^2 m_{in}(\theta_0, \gamma_0)}{\partial(\sigma^2)^2}\right| \xrightarrow{p} 0 \text{ and } \left|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 m_{in}\left(\widetilde{\theta}, \widehat{\gamma}\right)}{\partial\sigma^2\partial\delta'} - \frac{\partial^2 m_{in}(\theta_0, \gamma_0)}{\partial\sigma^2\partial\delta'}\right| \xrightarrow{p} 0
$$

which verifies (D5), and hence, asymptotic normality of the MLE.

B. *LS estimator*

The proof of (D5) for the LSE is analogous to that in the case of the MLE. Next, by (D2), the score $s_{in}(\theta_0, \gamma)$ is continuously differentiable in γ, and hence, by the Mean Value Theorem,

$$
n^{-1/2}\sum_{i=1}^{n}\left[s_{in}(\theta_0, \widehat{\gamma}) - s_{in}(\theta_0, \gamma_0)\right] \leq n^{-1/2}n|\widehat{\gamma} - \gamma_0|\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\partial s_{in}(\theta_0, \widetilde{\gamma})}{\partial\gamma}\right| \xrightarrow{p} 0
$$

where $\widetilde{\gamma}$ is between $\widehat{\gamma}$ and γ, since $n(\widehat{\gamma} - \gamma_0) = O_p(1)$ and $n^{-1}\sum_{i=1}^{n}|\partial s_{in}(\theta_0, \widetilde{\gamma})/\partial\gamma| = O_p(1)$. Now, observe that $\nu_n(\widehat{\gamma}) - \nu_n(\gamma_0) = n^{-1/2}\sum_{i=1}^{n}[s_{in}(\theta_0, \widehat{\gamma}) - s_{in}(\theta_0, \gamma_0)] - n^{-1/2}\sum_{i=1}^{n}Es_{in}(\theta_0, \widehat{\gamma})$. By the population moment condition, $Es_{in}(\theta_0, \gamma_0) = E\left[\frac{\partial}{\partial\theta}m(W_{in}, \theta_0, \gamma_0)\right] = 0$, and hence

$$
\left|n^{-1/2}\sum_{i=1}^{n}Es_{in}(\theta_0, \widehat{\gamma}) - Es_{in}(\theta_0, \gamma_0)\right| \leq n^{-1/2}n|\widehat{\gamma} - \gamma_0|\frac{1}{n}\sum_{i=1}^{n}E\left|\frac{\partial s_{in}(\theta_0, \widetilde{\gamma})}{\partial\gamma}\right| \xrightarrow{p} 0
$$

which verifies both (D7) and (D6). The proof of the theorem is thus complete. ∎

References

1. Bajari, P.; Hong, H.; Nekipelov, D. Game theory and econometrics: A survey of some recent research. Working Paper. University of Minnesota, Minneapolis, MN, USA, 2010.
2. De Paula, A. Econometric analysis of games with multiple equilibria. *Annu. Rev. Econ.* **2013**, *5*, 107–131.
3. Menzel, K. Inference for games with many players. Working Paper. NYU, New York, NY, USA, 2012.
4. Xu, H. Social interactions: A game theoretical approach. Working Paper, University of Texas at Austin, Texas, Austin, 2011.
5. Jenish, N.; Prucha, I.R. On spatial processes and asymptotic inference under near-epoch dependence. *J. Econom.* **2012**, *170*, 178–190.
6. Brock, W.; Durlauf, S. Discrete choice with social interaction. *Rev. Econ. Stud.* **2001**, *68*, 235–260.
7. Su, C.-L.; Judd, K.L. Constrained optimization approaches to estimation of structural models. *Econometrica* **2012**, *80*, 2213–2230.
8. Bajari, P.; Hong, H.; Krainer, J.; Nekipelov, D. Estimating static models of strategic interactions. *J. Bus. Econ. Stat.* **2010**, *28*, 469–482.
9. Xu, X.; Lee, L.-F. Maximum likelihood estimation of a spatial autoregressive Tobit model. Working Paper. The Ohio State University, Columbus, OH, USA, 2014.
10. Audretsch, D.B.; Feldman, M.P. R&D spillovers and the geography of innovation and production. *Am. Econ. Rev.* **1996**, *86*, 253–273.
11. Levin, R.; Reiss, P. Cost-reducing and demand-creating R&D with spillovers. *Rand J. Econ.* **1988**, *19*, 538–556.
12. d'Aspremont, C.; Jacquemin, A. Cooperative and noncooperative R&D in duopoly with spillovers. *Am. Econ. Rev.* **1988**, *78*, 1133–1137.
13. Motta, M. Cooperative R&D and vertical product differentiation. *Int. J. Ind. Org.* **1992**, *10*, 643–661.
14. Cohen, W.M.; Klepper, S. The anatomy of industry R&D intensity distributions. *Am. Econ. Rev.* **1992**, *82*, 773–788.
15. Gonzalez, X.; Jaumandreu, J. Threshold effects in product R&D decisions: Theoretical Framework and Empirical Analysis. Studies on the Spanish Economy, FEDEA. Available online: http://econpapers.repec.org/paper/fdafdaeee/45.htm (accessed on 10 October 2013).
16. Dixit, A.; Stiglitz, J. Monopolistic competition and optimum product diversity. *Am. Econ. Rev.* **1977**, *67*, 297–308.
17. Durlauf, S. A framework for the study of individual behavior and social interactions. Unpublished Manuscript, 2001.
18. Jenish, N. Nonparametric spatial regression under near-epoch dependence. *J. Econom.* **2012**, *167*, 224–239.
19. Carson, R.; Sun, Y. The Tobit model with a non-zero threshold. *Econom. J.* **2007**, *10*, 488–502.

20. Jenish, N. Spatial semiparametric model with endogenous regressors. *Econometric Theory* **2014**, forthcoming.
21. Davidson, J. *Stochastic Limit Theory*; Oxford University Press: Oxford, UK, 1994,
22. Andrews, D.W.K. Asymptotics for semi-parametric econometric models via stochastic equicontinuity. *Econometrica* **1994**, *62*, 43–72.
23. Jenish, N.; Prucha, I.R. Central limit theorems and uniform laws of large numbers for arrays of random fields. *J. Econom.* **2009**, *150*, 86–98.
24. Amemiya, T. Regression analysis when the dependent variable is truncated normal. *Econometrica* **1973**, *41*, 997–1016.