Tobit Models with Social Interactions: Complete vs Incomplete Information*

Chao Yang[†1], Lung-fei Lee[‡2], and Xi Qu[§3]

[1]School of Economics, Shanghai University of Finance and Economics; and Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, Shanghai 200433, China
[2]Department of Economics, The Ohio State University, Ohio, 43210, USA
[3]Department of Economics, Antai College of Economics and Management, Shanghai Jiaotong University, Shanghai 200030, China

December 13, 2016

Abstract

In many network data sets, the outcomes of interest are equal to zero for some agents and strictly positive for others. They can be analyzed by Tobit models with social interactions. Under complete information, all the observables and unobservables are publicly known. Under incomplete information, unobservables and some covariates observed by econometricians can be private information for agents. A Cox-type test is proposed for model selection. For property tax rates among adjacent municipalities in North Carolina, significant competing effects are found under both incomplete and complete information. However, the Cox Test is in favor of the complete information model.

JEL-Classification: C31, C35, C57

Keywords: Social Interactions, Tobit Model, Incomplete Information, Complete Information, The Cox Test

*We owe our thanks to Ohio Supercomputer Center for supporting high performance computation.
†420 School of Economics, 777 Guoding Road, Shanghai 200433, China. *Tel.* +86(0)21-6590-2704. *Email*: yang.chao@mail.shufe.edu.cn.

‡475 Arps Hall, 1945 N. High Street, Columbus, OH 43210, USA. *Tel.* 614-292-5508. *Email*: lee.1777@osu.edu.

§1954 Huashan Rd. Shanghai, 200030 China, *Tel.* +86(0)21-5230-1595. *Email*: xiqu@sjtu.edu.cn.

1 Introduction

Models of social interactions have the appealing ability to quantitatively characterize peer's influence on outcomes in a network. In practice, in some data sets observations of individual outcomes are all non-negative with some equal to zero. One example is tax rates set by local governments. The censored outcomes can be explained as the result that agents face binding constraints, such as the non-negative constraint. In a theoretical model on behaviors, zero outcomes can naturally come out (Bramoulle et al. (2014)). A Tobit model with social interactions can be used to account for both the censorship and peer effects in a data set. However, interactions between agents in a network can be modeled under different information structures. Two basic types are the complete information and incomplete information ones. In the former case, agents know the relevant characteristics of all agents in a network. Each agent's behavior is directly affected by those of others in a network. In the second scenario, some relevant features are private information and an agent's outcomes may be affected by her expectations on others' outcomes. These two types of information structures have different implications on interactions in a network or a social group. Xu and Lee (2015) discuss in detail about the identification and estimation of the Tobit model with social interactions under complete information and establish the consistency and asymptotic normality of the maximum likelihood (ML) estimator. A competitive model with incomplete information can be built by applying the general framework of social interactions under incomplete information set up by Yang and Lee (2016a). According to Borkovsky et al. (2015), it is usually debatable on which information structure to use in an empirical game application, when more than one structures can be reasonable. Although in the literature of the game theory, there are extensive discussions on the comparison of equilibria under different information structures in terms of welfare or incentives, e.g., Bergemann and Morris (2016), there is scarce investigation on formal econometric test for selection of models with different information structures.[1] In this paper, we construct a game theoretical framework for Tobit models with social interactions under complete and incomplete information structures and set up a Cox type test for selection between those two models. Our non-nested test complements the literature on game estimation. The theoretical investigation is applied to the property tax competition among adjacent municipalities in North Carolina.

As an effective way to model censored or truncated outcomes, the Tobit model has gained much attention for both empirical and theoretical microeconometrics since Tobin (1958) and Amemiya (1973). Even recently, it is still an active model for studies, e.g., Kumar(2012) proposes an extension of nonparametric estimation methods for nonlinear budget-set models to censored dependent variables. Abreyava and Shen (2014) consider estimation of censored panel-data models with individual-specific slope heterogeneity. Both strings of complete and incomplete information on models of social interactions have been sprung since the last decade. Lee(2007) and Boucher et al.(2014) discuss complete information models where all group and individual features are public information to all agents in a network or a social group. More traditional models of Manski (1993) and Brock and Durlauf (2001) are built on incomplete information, where an agent's actions are affected by her expectations on average behaviors of other agents and group characteristics in a social group. Manski (1993) studies linear models about socially interacted continuous choices; and Brock and Durlauf (2001) investigate binary choices for socially linked agents. In a recent work, Yang and Lee (2016a) extend previous social interaction models to a form of partial information but essentially still under the incomplete information framework. They find that, by allowing an agent to have personal or public information only on other agents' certain individual characteristics but incomplete information not only on their unobserved idiosyncratic shocks but also on remaining exogenous individual characteristics, conditional expectations about other agents' behaviors can be functions of private information. In Yang and Lee (2016a), their discussions have focused on continuous choices and binary choices.

The bases of both our Tobit models with complete or incomplete information are simultaneous-move games when players' actions are bounded from below by no action. In the complete information situation, given others' actions, each individual maximizes her utility function subject to a non-negative constraint.[2] For the case with incomplete information (but may have a partial personal information), her aim is to maximize expected utility function. The expected utility is formed according to available information and is rational in the sense that a profile of conditional expectations is consistent with a Bayesian Nash Equilibrium (BNE). A key condition that ensures unique equilibrium in both models with complete and incomplete information is that the interaction intensity is within a reasonable range. That range corresponds to weak or moderate social interactions, but not strong ones. Strong social interactions might demonstrate multiple (expectation) equilibria which generate both stable and unstable systems. This happens especially for binary choices models as shown by Brock and Durlauf(2001). For the Tobit social interaction model with complete information, interactions occur directly on the censored outcomes. Its Nash equilibrium (NE) can also be derived by a contraction mapping, which is valid in a similar scenario of weak or moderate social interactions. The latter has been established in Xu and Lee (2015). While the Tobit model with complete information has been studied in Qu and Lee (2013a) and Xu and Lee (2015), the Tobit social interaction model with incomplete information is new in the literature. Therefore, in this paper, we discuss its structure and also estimation issues in an appendix.

It is revealed from the Monte Carlo experiments that when we estimate models with both the true and wrong information structures, the estimated log likelihood of a model with a correct information structure tends to be larger than that of a model with wrong information structures. Thus, estimated sample log likelihood provides with an intuitive but informal criterion for model selection. The Cox type tests are based on estimated log likelihoods for non-nested competitive models. Jin and Lee (2013) establish such a test for the selection of linear spatial autoregressive (SAR) models. Here we develop the Cox tests for the Tobit models. Comparing with such tests for linear SAR models, as the Tobit model with complete information is a nonlinear model with spatial (or social) correlation, asymptotic distributions of the Cox test statistics need to be rigorously analyzed. We establish the asymptotic normality of the Cox tests based on the spatial Near-Epoch

Dependence (NED) asymptotics in Jenish and Prucha (2012).

As an empirical application, we study the property tax rates for contiguous municipalities in North Carolina. Tax competition among local governments has been theoretically and empirically studied in the public economics literature (see Brueckner(2003) for a comprehensive review). Most researches consider a tax rate as a continuous variable. However, it is more appropriate to adopt the Tobit model as local governments' choices are subject to the non-negative constraint and some shares of sample observations across regions are zeros. More recently, Porto and Revelli(2013) evaluate three empirical approaches to the analysis of spatially dependent limited tax policies. Their Tobit type models are based on interactions with latent variables and/or spatial time lags, which are different from ours. We model property tax rates as equilibrium outcomes from a static simultaneous-move game with complete information and rational expectations under incomplete information. For the estimation of incomplete information model, we use the nested fixed point ML method to evaluate the expectation functions proposed in the paper. For the complete information model, the likelihood function is derived in Xu and Lee (2015) where computation is relatively simpler. For the sample of municipal property tax rates in North Carolina, we find the existence of strong competition among near-by municipalities for both models with incomplete and complete information. However, our Cox test statistics reject the incomplete information model and are in favor of the complete information one.

This paper proceeds as follows. In Section 2, we build the model with incomplete information and compare it with the complete information one. There we also show that both models can be explained by simultaneous-move games. The Cox test statistics to discriminate the two models are presented in Section 3. Asymptotic distributions of the Cox test statistics are rigorously established. Their computations are also discussed there. Simulation results of the Monte Carlo experiments are presented in Section 4. Section 5 concentrates on the empirical study of the tax competition among municipalities in North Carolina. Section 6 concludes. Equilibrium expectations, parameter identification, and calculations in the relatively new model with incomplete information as well as the details about the Cox tests are in the Appendices. Detailed proofs and results for additional Monte Carlo experiments can be found in the online supplementary files.

2 The Tobit Models

2.1 Model Frameworks: Complete vs. Incomplete Information

Consider n agents, $i = 1, \cdots, n$, who are socially linked. Their social relations are represented by an $n \times n$ weighting matrix, W_n, such that for all $i \neq j$, its (i, j) entry, $W_{n,ij} > 0$, if i connects with j; $W_{n,ij} = 0$ otherwise. The diagonal elements, $W_{n,ii} = 0$ for all $i = 1, \cdots, n$. W_n may be either symmetric or asymmetric. For example, when W_n represents a friendship network, $W_{n,ij} = 1$ if i views j as one of her friends; and $W_{n,ij} = 0$ otherwise. When friendship is mutual, the network is undirected and W_n is symmetric. However, if friendship is not mutual, it is possible that i regards j as one of her friends, but j does not think i is among her friends. Then $W_{n,ij} \neq W_{n,ji}$, the network is direct and W_n is asymmetric. In spatial econometrics, an example is the relative strength of spatial interactions among regions (e.g., counties). In that case, $W_{n,ij}$ for $i \neq j$ may be the reciprocal of the geographic distance between two different counties. Formulated in this way, W_n is symmetric. However, in practice (as in empirical studies of SAR models), once it is row-normalized, W_n would in general be asymmetric. For our models, both direct and indirect networks are allowed.

With the scenario that agents are socially interacted and outcomes of agents, y_i's, are subject to the nonnegative constraint, one may consider the following model:

$$y_i = \max\left\{x_i'\beta + \lambda\widetilde{y}_{-i} + \epsilon_i, 0\right\}, \tag{2.1}$$

for $i = 1, 2, \cdots, n$. x_i's are exogenous explanatory variables for i, which are observable by econometricians. Relevant characteristics of i not observed from a data set are incorporated into the term, ϵ_i, which can be viewed as an idiosyncratic shock to agent i. What distinguishes this framework from the classical Tobit model is the term, $\lambda\widetilde{y}_{-i}$, where the parameter, λ, represents the interaction intensity and the variable, $\widetilde{y}_{-i}$, represents the influences on y_i from the observed or expected outcomes of agents (other than

i) who are linked to i. There are more than one way to specify $\widetilde{y}_{-i}$ based on different information structures. Under the assumption of complete information, every agent in a network knows all relevant features, x_i's and ϵ_i's, of all other members. However, under incomplete information, the unobservable ϵ_i is revealed only to i herself and only some exogenous characteristics in x_i's may be privately known by some other members. Take the case of interactions among students on academic performances as an example. Gender of a student is publicly known to all students and observed in a data set. For the high school GPA, although the econometrician can observe its value, an individual may know some but not all the values of her classmates.

With complete information, Xu and Lee (2015) model interactions as follows:

$$y_i = \max\left\{x_i^{'}\beta + \lambda\sum_{j\neq i} W_{n,ij}y_j + \epsilon_i, 0\right\}, \tag{2.2}$$

which corresponds to assuming $\widetilde{y}_{-i} = \sum_{j\neq i} W_{n,ij}y_j$ in Eq. (2.1). That is to say, under complete information, the outcome of an agent is directly influenced by outcomes of all agents who she is associated with. The model in Eq. (2.2) will be referred to as Model 1 in short.

Under incomplete information, an agent needs to make predictions on others' outcomes based on her available information when making her own decision. In their work on social interactions with a general form of incomplete information, Yang and Lee (2016a) assume that the impact of j on i's outcome is through i's expectation about j's outcome based on i's private information. As private information may be asymmetric across different agents, their model has the feature of expectation heterogeneity. Applying their framework to truncated outcomes is to assume that $\widetilde{y}_{-i} = \sum_{j\neq i} W_{n,ij}E[y_j|\cdot]$, where $\cdot$ refers to the proper information set possessed by i. To specify an information set, one may classify exogenous characteristics in x_i observed by econometricians into three categories, group features, X^g; commonly known individual characteristics, $X^c = (x_1^{c'}, \cdots, x_n^{c'})'$; and some personal traits which might not be publicly known, $X^p = (x_1^{p'}, \cdots, x_n^{p'})'$. That is, $x_i = (x_i^{c'}, x_i^{p'}, x^{g'})'$. The information set of the individual i can be represented by an $n\times 1$ vector J_i to represent her private information about X^p. That is, $J_i(j) = 1$ if x_j^p is known by i; and $J_i(j) = 0$ otherwise. $x^p_{J_i}$ then represents the random vector composed of those x^p_j's that i knows. For simplicity, use Z to collect all public information, including group features, X^g, commonly known individual traits, X^c, the social relations, W_n, as well as the information structure, $J_1, \cdots, J_n$. With such an incomplete information structure, the interaction Tobit model can be modeled as below:

$$y_i = \max\left\{\beta_0 + x_i^{c'}\beta_1 + x_i^{p'}\beta_2 + x^{g'}\beta_3 + \lambda\sum_{j\neq i} W_{n,ij}E[y_j|x^p_{J_i}, Z] + \epsilon_i, 0\right\}. \tag{2.3}$$

The model in Eq.(2.3) will be referred to as Model 2.[3]

An inspection of Eq.(2.2) and Eq.(2.3) reveals that in Model 1, although i's outcome can be affected by the outcomes of her socially associated agents directly, that happens only for those outcomes not censored. From Qu and Lee (2013a) and Xu and Lee (2015), if the agents whose outcomes are not censored are picked out, their outcomes are socially related within a sub-network; but the censored outcomes are affected by the uncensored outcomes with social relations, which distinguish the complete information Tobit model with social interactions from the linear SAR model. In Model 2, nonetheless, i's outcome will be affected by the expectations of every other agents whom she is associated with, no matter whether their actual outcomes are censored or not. Therefore, the two different types of information structures, complete vs. incomplete, will result in different implications on social interactions.

2.2 Game Theoretical Foundation

The above framework for Tobit models with social interactions can be related to a simultaneous-move game where values of actions are bounded below by zero to satisfy the nonnegative constraint. Denote the action taken by agent i by a_i. Assume that $a_i \geq 0$. Her payoff is determined by a quadratic function of her own action:[4]

$$r(a_i, a_{-i}, x_i) = \alpha - \gamma(a_i - x_i'\beta - \lambda\sum_{j\neq i} W_{n,ij}a_j - \epsilon_i)^2. \tag{2.4}$$

Under complete information, the strategy of a player is just to choose an action. A NE is a profile of strategies such that everyone is making her best response to others' strategies. Given actions of others a_{-i}, the best response of i is $a_i^*(a_{-i}) = \max\left\{x_i'\beta_1 + \lambda\sum_{j\neq i} W_{n,ij}a_j + \epsilon_i, 0\right\}$. Hence, an equilibrium outcome is just represented by Model 1 with Eq. (2.2). It is shown by Xu and Lee (2015) that when $|\lambda|\|W_n\|_\infty < 1$, where $\|W_n\|_\infty = \max_{1\le i\le n}\sum_{j=1}^n W_{n,ij}$, there will be a unique NE.

Under incomplete information, ϵ_i is observed only by i herself and also $x_i = (x_i^{c'}, x_i^{p'}, x^{g'})'$; but x_i^p's may not be publicly observable to all players. From Harsanyi (1967a; 1967b), incomplete information can be interpreted by unknown "types". Let $\mathcal{S}_i$ represent the support of x_i^p and $\mathcal{T}$ denote the common support of ϵ_i's. The set of states, $\prod_i^n \mathcal{S}_i \times \mathcal{T}^n$, is defined as the set of possible values of x_i^p's and ϵ_i's for all players. In this case, player i's "type" is her private information about exogenous characteristics, $x_{J_i}^p$, and shock, ϵ_i. Hence, her type set is the corresponding support, $\mathcal{R}_i = \prod_{k:J_i(k)=1}\mathcal{S}_k \times \mathcal{T}$. The signal function is a mapping from the set of states to her type, $\tau_i : \prod_i^n \mathcal{S}_i \times \mathcal{T}^n \to \prod_{k:J_i(k)=1}\mathcal{S}_k \times \mathcal{T}$. Her prior belief on the set of states is the joint distribution of x_i^p's, and the distribution for shocks. The prior belief is the same across all players. A strategy is a plan specifying an action for each possible realization of her type. That is, $s_i : \mathcal{R}_i \to \mathcal{A}_i$, where $\mathcal{A}_i$ is i's set of actions.

Because players move simultaneously, they do not know which actions others will take when they are making their own decisions. They can only form expectations based on their private information. The expected payoff by taking action a_i is as follows:

$$
\begin{aligned}
&E[r(a_i, a_{-i}, x^g, x_i^c, x_i^p)|x_{J_i}^p, Z, \epsilon_i] \\
=&\alpha - \gamma(a_i - \beta_0 - x_i^{c'}\beta_1 - x_i^{p'}\beta_2 - x^{g'}\beta_3 - \lambda\sum_{j\neq i} W_{n,ij}E[a_j|x_{J_i}^p, Z] - \epsilon_i)^2 \\
&+ \gamma\lambda^2\Big((\sum_{j\neq i} W_{n,ij}E[a_j|x_{J_i}^p, Z])^2 - E[(\sum_{j\neq i} W_{n,ij}a_j)^2|x_{J_i}^p, Z]\Big),
\end{aligned}
$$

where the expectation on a_j does not depends on ϵ_i, because ϵ_i's are mutual independent and also independent of X and W_n. Suppose that $\gamma > 0$, if there is no restriction on a_i, the agent's best response is to choose the ideal value,

$$a_i^* = \beta_0 + x^{c'}\beta_1 + x_i^{p'}\beta_2 + x^{g'}\beta_3 + \lambda\sum_{j\neq i} W_{n,ij}E[a_j|x_{J_i}^p, Z] + \epsilon_i.$$

However, with the constraint, $a_i \ge 0$, it is possible to have a corner solution. The optimal action is $a_i = \max\{a_i^*, 0\}$. Thus, $(s_1(x_{J_1}^p, \epsilon_1), \cdots, s_n(x_{J_n}^p, \epsilon_n))$ is a Bayesian Nash

Equilibrium (BNE) characterized by

$$s_i(x^p_{J_i}, \epsilon_i) = \max\left\{\beta_0 + x_i^{c'}\beta_1 + x_i^{p'}\beta_2 + x^{g'}\beta_3 + \lambda\sum_{j\neq i} W_{n,ij}E[s_j(x^p_{J_j}, \epsilon_j)|x^p_{J_i}, Z] + \epsilon_i, 0\right\},$$

for all $i = 1, \cdots, n$. Therefore, y_i's in Model 2 with Eq.(2.3) can be viewed as realizations of actions of a BNE. Applying the general analysis of Yang and Lee (2016a), we can derive sufficient conditions for the existence of a unique BNE. A key condition is also that $|\lambda|\|W_n\|_\infty < 1$, as in the complete information game. Intuitively, this condition means that the intensity of social interactions is moderate.

Xu and Lee (2015) discuss extensively the identification and estimation of Model 1. Because of the direct influence of outcomes among socially connected agents, the likelihood of an individual's action (or outcome) is not independent of those of others. Xu and Lee (2015) show that under generic conditions, the NED property of spatial process in Jenish and Prucha (2012) can be established and the ML estimator is consistent and asymptotically normal. Applying the theoretical analysis by Yang and Lee (2016a) to Model 2, it can also be shown that parameters in that model can be identified under some generic conditions. Different from the complete information model, in the incomplete information case, the individual likelihoods are independent of each other, which makes large sample properties of esimators for Model 2 simpler to be derived than those of Model 1. However, as pointed out by Yang and Lee (2016a), computing the sample likelihood function of Model 2 requires the calculation of equilibrium conditional expectations, which are characterized as a functional fixed point. As a result, they propose the estimation of parameters by nesting a fixed point iteration algorithm inside a ML estimation. This is analogous to the NFXP algorithm proposed by Rust (1987) for dynamic decision making problems. The estimation of Model 2 is more computationally involved than that of Model 1. Details on how to apply the theory for incomplete information models in Yang and Lee (2016a) to censored outcomes can be found in Appendix A. Numerical methods for computing the equilibrium conditional expectations are elaborated in the supplementary file to this paper.

3 Cox Type Tests

As different information structures may imply different interaction effects, it is of significance to provide formal statistical criteria by which to select between them. For example, consider the tax competition among adjacent municipalities. Intuitively, both scenarios are possible. It is reasonable to assume that a local municipal committee knows the financial and residential features of all the municipalities in the same state when setting tax rates. Nonetheless, as municipalities set new tax rates of a new fiscal year before detailed financial and demographical reports for local jurisdictions are published, it is plausible that some features about the residence of a city is privately known by the government of this city or some of its neighbors but not by other local jurisdictions in the same state. In the empirical study of Section 5, the estimated interaction intensities differ between the model with complete information and the one with incomplete information. The estimation results also show that the model predictions change when different information structures are used. Thus, it is a question of theoretical and empirical importance on which information structure to select given a data set. As far as we know, there has been hardly any formal test to select information structures in game estimation with an exception by Gireco (2014). In a model of market entry, Gireco (2014) assumes that an entry cost of a player is affected by two types of shocks (unobserved by econometricians): one is observed by all the players and another is the shocks observed by herself. The relative variance of the shock of the first type compared to the second one plays as an indicator of information structures. The case that it is equal to 0 corresponds to incomplete information; and the case that it is equal to 1 (after normalization) indicates complete information. As the full model nests the models of complete and incomplete information in a same framework, a model with a complete (or incomplete) information will be rejected if its estimates are outside of a confidence region of the full model at a certain significance level. This technique, however, does not work when two types of information structures cannot be nested in a same framework, such as the Tobit models with social interactions with complete vs. incomplete information, where different information structures change the dependence among outcomes in a nonlinear way. In this paper, we propose a Cox type test, which is a non-nested test for competitive models based on estimated sample likelihoods.

3.1 Log Likelihoods and Test Statistics

In the following discussion, to distinguish between Model 1 and Model 2, let m be 1 or 2 respectively, as an index for Model 1 or Model 2. Assume that the error terms $\varepsilon_{m,i}$'s are i.i.d. normally distributed with zero mean and variance of σ_m^2 as in Qu and Lee (2013a) and Xu and Lee (2015). Denote $Y_n = (y_1, ..., y_n)'$ and $X_n = (X_n^c, X_n^p, X^g) = (x_1, ..., x_n)'$. Let $\theta_m = (\delta_m', \theta_x')'$ for $m = 1$ or 2, where $\delta_m = (\lambda_m, \beta_m', \sigma_m)'$ and θ_x is a finite dimensional parameter vector in the conditional density (or probability for a discrete variable) $f(X_n^p|X_n^c, X^g, \theta_x)$ of X_n^p conditional on X_n^c and X^g. θ_x is the same in the two models, but δ_m and θ_m are parameters of Model m. The joint distribution of (Y_n, X_n) under Model m is $f_m(Y_n, X_n|\theta_m) = f_m(Y_n|X_n, \theta_m)\cdot f(X_n^p|X_n^c, X^g, \theta_x)\cdot f(X_n^c, X^g)$. Under Model m, let $L_{mn}^{full}(\theta_m) = \ln f(Y_n, X_n|\theta_m)$ be the joint log likelihood function and $L_{mn}(\theta_m) = \ln f_m(Y_n|X_n, \theta_m)$ be the log likelihood function of Y_n conditional on X_n. Then for Model 1,

$$\begin{aligned} L_{1n}(\theta_1) &= \ln f_1(Y_n|X_n, \delta_1) = \sum_{i=1}^n I(y_i = 0)\ln\Phi(-\gamma_{1i}(\delta_1|Y_n, X_n)) + \ln|I_n - \lambda_1 G_n(Y_n)W_nG_n(Y_n)| \\ &\quad -\frac{1}{2}\sum_{i=1}^n I(y_i > 0)[\ln(2\pi\sigma_1^2) + (\frac{y_i}{\sigma_1} - \gamma_{1i}(\delta_1|Y_n, X_n))^2] \end{aligned}$$

with $\gamma_{1i}(\delta_1|Y_n, X_n) = (\lambda_1 w_{i.}Y_n + x_i'\beta_1)/\sigma_1$, $w_{i.}$ being the ith row of W_n, and $G_n(Y_n) = diag\{I(y_1 > 0), ..., I(y_n > 0)\}$. For this model, we can see that δ_1 appears in the density of $f_1(Y_n|X_n, \delta_1)$; but θ_x in $f(X_n^p|X_n^c, X^g, \theta_x)$ does not enter into the conditional distribution of Y_n on X_n. For Model 2,

$$\begin{aligned} L_{2n}(\theta_2) =& \ln f_2(Y_n|X_n, \delta_2, \theta_x) = \sum_{i=1}^n I(y_i = 0)\ln\Phi(-\gamma_{2i}(\delta_2, \theta_x)) \\ &- \frac{1}{2}\sum_{i=1}^n I(y_{i,n} > 0)[\ln(2\pi\sigma_2^2) + (\frac{y_i}{\sigma_2} - \gamma_{2i}(\delta_2, \theta_x))^2], \end{aligned}$$

with $\gamma_{2i}(\delta_2, \theta_x) = [\lambda_2\sum_{j\neq i} W_{n,ij}E_{\theta_2}(y_j|x_{J_i}^p, Z) + x_i'\beta_2]/\sigma_2$. For this model, we see that both δ_2 and θ_x appear in the density $f_2(Y_n|X_n, \delta_2, \theta_x)$ as θ_x is used for the expectation of y_j's given the information set $x_{J_i}^p$.

For relatively computationally simper estimation of these models, we suggest the use of a two-step estimation. In the first step, we may use data available for X_n^p and estimate θ_x based on its stochastic structure $f(X_n^p|X_n^c, X^g, \theta_x)$, such as the ML estimation when $f(X_n^p|X_n^c, X^g, \theta_x)$ has a known function form. Otherwise, it can be estimated by some other methods, such as GMM. Suppose the estimator $\widehat{\theta}_{xn}$ of θ_x is $\sqrt{n}$ consistent with $\sqrt{n}(\widehat{\theta}_{xn} - \theta_{x0}) \xrightarrow{d} N(0, V_2)$ for some positive definite matrix V_2, which holds regardless the DGP of Y_n is from Model 1 or Model 2. For the second step estimation, δ_m can be estimated as $\widehat{\delta}_{mn} = \arg\max_{\delta_m \in \Delta_m} L_{mn}(\delta_m, \widehat{\theta}_{xn})$[5], with Δ_m being the parameter space of δ_m. Then the two step ML estimator for θ_m is $\widehat{\theta}_{mn} = (\widehat{\delta}'_{mn}, \widehat{\theta}'_{xn})'$. Let $\overline{L}_{mn}(\delta_m, \theta_x : \theta_l) = \int L_{mn}(\delta_m, \theta_x) f_l(Y_n|\theta_l, X_n) dY_n$ be the conditional expectation of $L_{mn}(\delta_m, \theta_x)$ and $\overline{\delta}_{mn}(\theta_l) = \arg\max_{\delta_m \in \Delta_m} \overline{L}_{mn}(\delta_m, \theta_x : \theta_l)$ be the pseudo-true value of δ_m when the model l with parameter $\theta_l = (\delta'_l, \theta'_x)'$ generates the data. Denote the pseudo true value of θ_m when the Model l is the DGP as $\overline{\theta}_{mn}(\theta_l) = (\overline{\delta}_{mn}(\theta_l)', \theta'_x)'$.

The Cox test is based on the recentered log likelihood ratio, i.e., $L_{mn}^{full}(\widehat{\delta}_{mn}, \widehat{\theta}_{xn}) - L_{ln}^{full}(\widehat{\delta}_{ln}, \widehat{\theta}_{xn})$, subtracting its expectation. As $L_{mn}^{full}(\delta_m, \theta_x)$ is the sum of three terms, $L_{mn}(\delta_m, \theta_x)$, $\ln f(X_n^p|X_n^c, X^g, \theta_x)$, and $\ln f(X_n^c, X^g)$, where the second and third terms are the same under the two models, $L_{mn}^{full}(\delta_m, \theta_x) - L_{ln}^{full}(\delta_l, \theta_x) = L_{mn}(\theta_m) - L_{ln}(\theta_l)$. The difference of the full log likelihood comes only from the first term, i.e., the log likelihood function of Y_n conditional on X_n. Therefore, the corresponding Cox test can be constructed using the conditional log likelihoods. To test Model 1 being the true model, the Cox test statistic is based on

$$\frac{1}{\sqrt{n}}\left(L_{2n}(\widehat{\delta}_{2n}, \widehat{\theta}_{xn}) - L_{1n}(\widehat{\delta}_{1n}) - [\overline{L}_{2n}(\overline{\delta}_{2n}(\widehat{\theta}_{1n}), \widehat{\theta}_{xn} : \widehat{\theta}_{1n}) - \overline{L}_{1n}(\widehat{\delta}_{1n} : \widehat{\theta}_{1n})]\right)$$

or

$$\frac{1}{\sqrt{n}}\left(L_{2n}(\widehat{\delta}_{2n}, \widehat{\theta}_{xn}) - L_{1n}(\widehat{\delta}_{1n}) - [\overline{L}_{2n}(\widehat{\delta}_{2n}, \widehat{\theta}_{xn} : \widehat{\theta}_{1n}) - \overline{L}_{1n}(\widehat{\delta}_{1n} : \widehat{\theta}_{1n})]\right);$$

to test Model 2 being the true one, the Cox test statistic is based on

$$\frac{1}{\sqrt{n}}\left(L_{1n}(\widehat{\delta}_{1n}) - L_{2n}(\widehat{\delta}_{2n}, \widehat{\theta}_{xn}) - [\overline{L}_{1n}(\overline{\delta}_{1n}(\widehat{\theta}_{2n}) : \widehat{\theta}_{2n}) - \overline{L}_{2n}(\widehat{\delta}_{2n}, \widehat{\theta}_{xn} : \widehat{\theta}_{2n})]\right)$$

or

$$\frac{1}{\sqrt{n}}\left(L_{1n}(\widehat{\delta}_{1n}) - L_{2n}(\widehat{\delta}_{2n}, \widehat{\theta}_{xn}) - [\overline{L}_{1n}(\widehat{\delta}_{1n} : \widehat{\theta}_{2n}) - \overline{L}_{2n}(\widehat{\delta}_{2n}, \widehat{\theta}_{xn} : \widehat{\theta}_{2n})]\right).$$

The first expression for testing either Model 1 or Model 2 is the original version of the Cox (1961,1962) test, while the second one corresponds to Atkinson (1970) version.

These Cox (1961,1962) and Atkinson (1970) versions, are asymptotically equivalent because

$$\begin{aligned}&\frac{1}{\sqrt{n}}\left(\overline{L}_{mn}(\widehat{\delta}_{mn},\widehat{\theta}_{xn}:\widehat{\theta}_{ln})-\overline{L}_{mn}(\overline{\delta}_{mn}(\widehat{\theta}_{ln}),\widehat{\theta}_{xn}:\widehat{\theta}_{ln})\right)\\ =\;&\frac{1}{2n}\left(\overline{\delta}_{mn}(\widehat{\theta}_{ln})-\widehat{\delta}_{mn}\right)'\frac{\partial^2\overline{L}_{mn}(\widetilde{\delta}_{mn},\widehat{\theta}_{xn}:\widehat{\theta}_{ln})}{\partial\delta_m\partial\delta_m'}\sqrt{n}\left(\overline{\delta}_{mn}(\widehat{\theta}_{ln})-\widehat{\delta}_{mn}\right)=o_p(1),\end{aligned}$$

where $\widetilde{\delta}_{mn}$ is a value between $\widehat{\delta}_{mn}$ and $\overline{\delta}_{mn}(\widehat{\theta}_{ln})$. The last "$=o_p(1)$" holds because $\frac{\partial^2\overline{L}_{mn}(\widetilde{\delta}_{mn},\widehat{\theta}_{xn}:\widehat{\theta}_{ln})}{n\partial\delta_m\partial\delta_m'}=O_p(1)$ and

$\widehat{\delta}_{mn}-\overline{\delta}_{mn}(\widehat{\theta}_{ln})=\widehat{\delta}_{mn}-\overline{\delta}_{mn}(\theta_{l0})+\overline{\delta}_{mn}(\theta_{l0})-\overline{\delta}_{mn}(\widehat{\theta}_{ln})=O_p(\frac{1}{\sqrt{n}})+\frac{\partial\overline{\delta}_{mn}(\widetilde{\theta}_{ln})}{\partial\theta_l'}(\theta_{l0}-\widehat{\theta}_{ln})=o_p(1).$

Here, $\frac{\partial\overline{\delta}_{mn}(\theta_l)}{\partial\theta_l'}=-\left(\frac{\partial^2\overline{L}_{mn}(\delta_m,\theta_x:\theta_l)}{n\partial\delta_m\partial\delta_m'}\right)^{-1}\frac{\partial^2\overline{L}_{mn}(\delta_m,\theta_x:\theta_l)}{n\partial\delta_m\partial\theta_l'}|_{\delta_m=\overline{\delta}_{mn}(\theta_l)}$ is bounded, because $\overline{\delta}_{mn}(\theta_l)$ is the unique solution of $\partial\overline{L}_{mn}(\delta_m,\theta_x:\theta_l)/\partial\delta_m=0$ and we can apply the implicit function theorem.

3.2 Test Statistics Distribution

As the large sample theory on NED spatial processes in Xu and Lee (2015) is general, it can be applied to derive the asymptotic distributions of the Cox test statistics for both Model 1 and Model 2. Under the null hypothesis that the data is generated from Model 1,

$$\begin{aligned}Cox_1(Y_n,X_n)&=\frac{1}{\sqrt{n}}\left(L_{2n}(\widehat{\delta}_{2n},\widehat{\theta}_{xn})-L_{1n}(\widehat{\delta}_{1n})-[\overline{L}_{2n}(\widehat{\delta}_{2n},\widehat{\theta}_{xn}:\widehat{\theta}_{1n})-\overline{L}_{1n}(\widehat{\delta}_{1n}:\widehat{\theta}_{1n})]\right)\\&=\frac{1}{\sqrt{n}}\sum_{i=1}^n[q_{1i}(\theta_{10})-E_{\theta_{10}}(q_{1i}(\theta_{10}))]+o_p(1)\xrightarrow{d}N\left(0,\lim_{n\to\infty}\frac{1}{n}V_{\theta_{10}}(\sum_{i=1}^n q_{1i}(\theta_{10}))\right);\end{aligned}$$

under the null hypothesis that the data is generated from Model 2,

$$\begin{aligned}Cox_2(Y_n,X_n)&=\frac{1}{\sqrt{n}}\left(L_{1n}(\widehat{\delta}_{1n})-L_{2n}(\widehat{\delta}_{2n},\widehat{\theta}_{xn})-[\overline{L}_{1n}(\widehat{\delta}_{1n}:\widehat{\theta}_{2n})-\overline{L}_{2n}(\widehat{\delta}_{2n},\widehat{\theta}_{xn}:\widehat{\theta}_{2n})]\right)\\&=\frac{1}{\sqrt{n}}\sum_{i=1}^n[q_{2i,n}(\theta_{20})-E_{\theta_{20}}(q_{2i}(\theta_{20}))]+o_p(1)\xrightarrow{d}N\left(0,\lim_{n\to\infty}\frac{1}{n}V_{\theta_{20}}(\sum_{i=1}^n q_{2i}(\theta_{20}))\right),\end{aligned}$$

where both the expressions of $q_{1i}(\theta_{10})$ and $q_{2i}(\theta_{20})$ can be found, respectively, in Eq. B.5 and Eq. B.6 of the Appendix. The derivation is discussed in detail in Appendix B.

3.3 Test Statistics Calculation

In general, it is demanding to derive an analytic form of the variances of the Cox test statistics. We employ the bootstrap method to construct them numerically.[6]

Consider the null hypothesis that Model 2 with θ_{20} is the true model. Treat X_n^c, X^g, and X^p as fixed. The Cox test statistic can be constructed step by step as follows:

1. Use the first step estimator $\widehat{\theta}_{xn}$ to generate a bootstrapped sample $X_n^{p,(b)}$ and do estimation to get the first step estimate $\widehat{\theta}_{xn}^{(b)}$, where b represents a bootstrap step. By repeating bootstraps, we have a sequence of consistent bootstrapped estimators of $\widehat{\theta}_{xn}$.

2. Use the second step MLE $\widehat{\delta}_{2n}$ together with $\widehat{\theta}_{xn}$ and the original sample data $X_n = (X_n^c, X_n^p, X^g)$ to generate a bootstrapped sample $Y_{2n}^{(b)}$ from Model 2 as if $\widehat{\theta}_{2n} = (\widehat{\delta}_{2n}', \widehat{\theta}_{xn}')'$ is the true parameter vector. More specifically, we generate $Y_{2n}^{(b)} = (y_{21}^{(b)}, ..., y_{2n}^{(b)})'$ from $\widehat{\theta}_{2n}$: $y_{2i}^{(b)} = \max(0, \widehat{\lambda}_{2n} \sum_{j\neq i} W_{n,ij} E_{\widehat{\theta}_{2n}}(y_{2j}|x_{J_i}^p, Z) + x_i'\widehat{\beta}_{2n} + \varepsilon_{2i}^{(b)})$, where $\varepsilon_{2i}^{(b)} \sim N(0, \widehat{\sigma}_{2n}^2)$, to calculate the conditional log likelihood functions
$$L_{1n}^{(b)}(\delta_1|Y_{2n}^{(b)}, X_n) = \sum_{i=1}^n I(y_{2i}^{(b)} = 0) \ln \Phi(-\gamma_{1i}^{(b)}(\delta_1|Y_{2n}^{(b)}, X_n)) + \ln|I_n - \lambda_1 G(Y_{2n}^{(b)}) W_n G(Y_{2n}^{(b)})| - (1/2)\sum_{i=1}^n I(y_{2i}^{(b)} > 0)\left(\ln(2\pi\sigma_1^2) + (y_{2i}^{(b)}/\sigma_1 - \gamma_{1i}^{(b)}(\delta_1|Y_{2n}^{(b)}, X_n))^2\right);$$
$$L_{2n}^{(b)}(\delta_2, \theta_x|Y_{2n}^{(b)}, X_n) = \sum_{i=1}^n I(y_{2i}^{(b)} = 0) \ln \Phi[-\gamma_{2i}^{(b)}(\theta_2|X_n)] - (1/2)\sum_{i=1}^n I(y_{2i}^{(b)} > 0)\left(\ln(2\pi\sigma_2^2) + [y_{2i}^{(b)}/\sigma_2 - \gamma_{2i}^{(b)}(\theta_2|X_n)]^2\right),$$
where $\gamma_{1i}^{(b)}(\delta_1|Y_{2n}^{(b)}, X_n) = [\lambda_1 w_{i.} Y_{2n}^{(b)} + x_i'\beta_1]/\sigma_1$ and
$$\gamma_{2i}^{(b)}(\theta_2|X_n) = [\lambda_2 \sum_{j\neq i} W_{n,ij} E_{\theta_2}(y_j|x_{J_i}^p, Z) + x_i'\beta_2]/\sigma_2,$$
with $E_{\theta_2}(y_j|x_{J_i}^p, Z)$ being a function of all information available which does not change with bootstrap sample b.[7]

3. Use the generated vector $Y_{2n}^{(b)}$ to do the second step ML estimation for both Models 1 and 2 to get bootstrapped estimates $\widehat{\delta}_{1n}^{(b)}$ and $\widehat{\delta}_{2n}^{(b)}$. To be more specific,
$$\widehat{\delta}_{2n}^{(b)} = \arg\max_{\delta_2} L_{2n}^{(b)}(\delta_2, \widehat{\theta}_{xn}^{(b)}|Y_{2n}^{(b)}, X_n), \text{ and } \widehat{\delta}_{1n}^{(b)} = \arg\max_{\delta_1} L_{1n}^{(b)}(\delta_1|Y_{2n}^{(b)}, X_n).$$

4. For each b, generate a total number of S n-dimensional vectors of $Y_{2n}^{(s_b)}$ from Model 2 by using X_n and treat $\widehat{\theta}_{2n}^{(b)} = (\widehat{\delta}_{2n}^{(b)\prime}, \widehat{\theta}_{xn}^{(b)\prime})'$ as 'true' parameter vector
$$y_{2i}^{(s_b)} = \max(0, \widehat{\lambda}_{2n}^{(b)} \sum_{j \neq i} W_{n,ij} E_{\widehat{\theta}_{2n}^{(b)}}(y_{2j}|x_{J_i}^p, Z) + x_i' \widehat{\beta}_{2n}^{(b)} + \varepsilon_{2i}^{(s_b)}),$$
where $\varepsilon_{2i}^{(s_b)} \sim N(0, \widehat{\sigma}_{2n}^{(b)2})$, $s_b = 1, ..., S$. With these, we can approximate $\overline{L}_{1n}(\widehat{\delta}_{1n}^{(b)} : \widehat{\theta}_{2n}^{(b)})$ by the empirical mean as $\widehat{\overline{L}}_{S,1n}(\widehat{\delta}_{1n}^{(b)} : \widehat{\theta}_{2n}^{(b)}) \equiv (1/S) \sum_{s_b=1}^{S} L_{1n}^{(s_b)}(\widehat{\delta}_{1n}^{(b)}|Y_{2n}^{(s_b)}, X_n)$, and similarly for $\overline{L}_{2n}(\widehat{\delta}_{2n}^{(b)}, \widehat{\theta}_{xn}^{(b)} : \widehat{\theta}_{2n}^{(b)})$.[8]

5. Compute the value of the statistic
$$\begin{aligned} Cox_2(\widehat{\theta}_n^{(b)}|Y_{2n}^{(b)}, X_n) &= \frac{1}{\sqrt{n}} \Big(L_{1n}(\widehat{\delta}_{1n}^{(b)}|Y_{2n}^{(b)}, X_n) - L_{2n}(\widehat{\delta}_{2n}^{(b)}, \widehat{\theta}_{xn}^{(b)}|Y_{2n}^{(b)}, X_n) \Big) \\ &\quad - \frac{1}{\sqrt{n}} \frac{1}{S} \sum_{s_b=1}^{S} [L_{1n}^{(s_b)}(\widehat{\delta}_{1n}^{(b)}|Y_{2n}^{(s_b)}, X_n) - L_{2n}^{(s_b)}(\widehat{\delta}_{2n}^{(b)}, \widehat{\theta}_{xn}^{(b)}|Y_{2n}^{(s_b)}, X_n)] \end{aligned}$$

6. Repeat steps 1) - 5) for many times to have B values of the statistics.

7. Order the statistics to get the upper α quantile critical value $c_{1-\alpha}^B$. Reject the null of Model 2 if $Cox_2(\widehat{\theta}_n|Y_n, X_n) > c_{1-\alpha}^B$. This is equivalent to reject the null if $(1/B) \sum_{b=1}^{B} I(Cox_2(\widehat{\theta}_n^{(b)}|Y_{2n}^{(b)}, X_n) > Cox_2(\widehat{\theta}_n|Y_n, X_n)) < \alpha$.

A bootstrap version of the Cox test statistics under the hypothesis that Model 1 with θ_{10} is the true model can be constructed in an analogous way:

1. Use the second step MLE $\widehat{\delta}_{1n}$ and the original sample data $X_n = (X_n^c, X^g, X_n^p)$ to generate sample $Y_{1n}^{(b)}$ from Model 1 as if $\widehat{\delta}_{1n}$ is the true parameter vector, where b represents a bootstrap step. From this, we generate $Y_{1n}^{(b)} = (y_{11}^{(b)}, ..., y_{1n}^{(b)})'$ from $\widehat{\delta}_{1n}$:
$$y_{1i}^{(b)} = \max(0, \widehat{\lambda}_{1n} \sum_{j \neq i} W_{n,ij} y_{1j}^{(b)} + x_i' \widehat{\beta}_{1n} + \varepsilon_{1i}^{(b)}),$$
where $\varepsilon_{1i}^{(b)} \sim N(0, \widehat{\sigma}_{1n}^2)$, to calculate the conditional log likelihood functions
$$\begin{aligned} L_{1n}^{(b)}(\delta_1|Y_{1n}^{(b)}, X_n) &= \sum_{i=1}^{n} I(y_{1i}^{(b)} = 0) \ln \Phi[-\gamma_{1i}^{(b)}(\delta_1|Y_{1n}^{(b)}, X_n)] + \ln \Big| I_n - \lambda_1 G(Y_{1n}^{(b)}) W_n G(Y_{1n}^{(b)}) \Big| \\ &\quad - \frac{1}{2} \sum_{i=1}^{n} I(y_{1i}^{(b)} > 0)[\ln(2\pi\sigma_1^2) + (\frac{y_{1i}^{(b)}}{\sigma_1} - \gamma_{1i}^{(b)}(\delta_1|Y_{1n}^{(b)}, X_n))^2] \text{ and} \\ L_{2n}^{(b)}(\delta_2, \theta_x|Y_{1n}^{(b)}, X_n) &= \sum_{i=1}^{n} I(y_{1i}^{(b)} = 0) \ln \Phi[-\gamma_{2i}^{(b)}(\theta_2|X_n)] \\ &\quad - \frac{1}{2} \sum_{i=1}^{n} I(y_{1i}^{(b)} > 0) \left(\ln(2\pi\sigma_2^2) + [\frac{y_{1i}^{(b)}}{\sigma_2} - \gamma_{2i}^{(b)}(\theta_2|X_n)]^2 \right), \end{aligned}$$

where $\gamma_{1i}^{(b)}(\delta_1|Y_{1n}^{(b)},X_n) = [\lambda_1 w_{i.}Y_{1n}^{(b)} + x_i'\beta_1]/\sigma_1$ and $\gamma_{2i}^{(b)}(\theta_2|X_n) = [\lambda_2 \sum_{j\neq i} W_{n,ij} E_{\theta_2}(y_j|x_{J_i}^p, Z) + x_i'\beta_2]/\sigma_2$ as before.

2. Use the generated vector $Y_{1n}^{(b)}$ to get the second step ML estimates $\widehat{\delta}_{1n}^{(b)}$ and $\widehat{\delta}_{2n}^{(b)}$, respectively, for both Models 1 and 2. To be more specific, $\widehat{\delta}_{2n}^{(b)} = \arg\max_{\delta_2} L_{2n}^{(b)}(\delta_2, \widehat{\theta}_{xn}^{(b)}|Y_{1n}^{(b)}, X_n)$ and $\widehat{\delta}_{1n}^{(b)} = \arg\max_{\delta_1} L_{1n}^{(b)}(\delta_1|Y_{1n}^{(b)}, X_n)$, where $\widehat{\theta}_{xn}^{(b)}$ is the same as before.

3. For each b, generate a total number of S n-dimensional vectors of $Y_{1n}^{(s_b)}$ from Model 1 by using X_n and treat $\widehat{\delta}_{1n}^{(b)}$ as the 'true' parameter vector
$$y_{1i}^{(s_b)} = \max(0, \widehat{\lambda}_{1n}^{(b)} \sum_{j\neq i} W_{n,ij} y_{1j}^{(s_b)} + x_i'\widehat{\beta}_{1n}^{(b)} + \varepsilon_{1i}^{(s_b)}),$$
where $\varepsilon_{1i}^{(s_b)} \sim N(0, \widehat{\sigma}_{1n}^{(b)2})$, $s_b = 1, ..., S$. From these, we approximate $\overline{L}_{1n}(\widehat{\delta}_{1n}^{(b)} : \widehat{\delta}_{1n}^{(b)})$ by the empirical mean as $\widehat{\overline{L}}_{S,1n}(\widehat{\delta}_{1n}^{(b)} : \widehat{\delta}_{1n}^{(b)}) \equiv \frac{1}{S}\sum_{s_b=1}^{S} L_{1n}^{(s_b)}(\widehat{\delta}_{1n}^{(b)}|Y_{1n}^{(s_b)}, X_n)$, and similarly for $\overline{L}_{2n}(\widehat{\delta}_{2n}^{(b)}, \widehat{\theta}_{xn}^{(b)} : \widehat{\delta}_{1n}^{(b)})$. Or we can use the analytical expression of $\overline{L}_{2n}(\delta_2, \theta_x : \theta_2)$.

4. Compute the value of the statistic
$$\begin{aligned} Cox_1(\widehat{\theta}_n^{(b)}|Y_{1n}^{(b)}, X_n) &= \frac{1}{\sqrt{n}}\Big(L_{2n}(\widehat{\delta}_{2n}^{(b)}, \widehat{\theta}_{xn}^{(b)}|Y_{1n}^{(b)}, X_n) - L_{1n}(\widehat{\delta}_{1n}|Y_{1n}^{(b)}, X_n)\Big) \\ &\quad - \frac{1}{\sqrt{n}}\frac{1}{S}\sum_{s_b=1}^{S}[L_{2n}^{(s_b)}(\widehat{\delta}_{2n}^{(b)}, \widehat{\theta}_{xn}^{(b)}|Y_{1n}^{(s_b)}, X_n) - L_{1n}^{(s_b)}(\widehat{\delta}_{1n}^{(b)}|Y_{1n}^{(s_b)}, X_n)] \end{aligned}$$

5. Repeat steps (1) - (4) for many times to have B values of the statistics.

6. Reject the null if $(1/B)\sum_{b=1}^{B} I(Cox_1(\widehat{\theta}_n^{(b)}|Y_{1n}^{(b)}, X_n) > Cox_1(\widehat{\theta}_n|Y_n, X_n)) < \alpha$.

The above numerical approximations are justified to be valid in Appendix C

4 Monte Carlo Experiments

We investigate finite sample performance of the NFXP ML estimation and Cox tests via Monte Carlo experiments. In the experiments, observed group features in X^g are irrelevant and hence, absent, for simplicity. The idiosyncratic shocks, ϵ_i's, are i.i.d. $N(0, \sigma^2)$. We focus on the estimation of the coefficient of the intercept, β_0, individual commonly known characteristics, β_1, private personal features, β_2, the interaction intensity among socially associated agents, λ, and the standard deviation of the idiosyncratic shocks, σ.

Their true values are $\beta_0 = 0$, $\beta_1 = 1$, $\beta_2 = 1$, and $\sigma = 1$. Both small interaction intensity, $\lambda = 0.3$, and intermediate one, $\lambda = 0.6$, are considered. The generation of the spatial weights matrix and regressors is described next.

For the social networks matrix W_n, where n is the population size of the whole network, $W_{n,ij} = 1$ if agent i links to agent j, where $i \neq j$; and $W_{n,ij} = 0$ otherwise. As usual, $W_{n,ii} = 0$ for all $i = 1, \cdots, n$. Then we row-normalize W_n such that the sum of each row is equal to 1. In the experiments, a sample is composed of a collection of independent groups. Consequently, W_n can be organized as a block-diagonal matrix, with each block representing social relations in one group. We assume that all groups have the same size, 20. Within a group, for every agent, 15 other agents are randomly selected to be linked to her. The number of groups, G, is either 10 or 40.[9]

The commonly known individual characteristics, x_i^c's, are generated as independent standard normal variables for different agents. For x_i^p, we focus on the case that x_i^p's are continuously distributed.[10] Suppose that $x_{i,g}^p = \alpha_g + \epsilon_{i,g}^p$, for $g = 1, \cdots, G$ and $i = 1, \cdots, n$, where α_g's are i.i.d. $N(\mu, \sigma_1^2)$ and $\epsilon_{i,g}^p$'s are i.i.d. $N(0, \sigma_2^2)$ and independent of α_g's. Then within a group g, $(x_{1,g}^p, \cdots, x_{n,g}^p)'$ is jointly normal with a common mean μ, a common variance η^2, and a common correlation coefficient ρ for any $x_{i,g}^p$ and $x_{j,g}^p$ with $i \neq j$, where $\eta^2 = \sigma_1^2 + \sigma_2^2$ and $\rho = \sigma_1^2/(\sigma_1^2 + \sigma_2^2)$. For any $i \neq j$, given $x_{j,g}^p = x$, $x_{i,g}^p$ is $N(\rho x + (1-\rho)\mu, (1-\rho^2)\eta^2)$. For the Monte Carlo study, we choose $\mu = 1$, $\eta = 2$, and $\rho = 0.4$. While values of those parameters are known to agents, econometricians need to estimate them from observed data. In this situation, we adopt a two-step estimation. We first estimate μ, η, and ρ from $X_{i,g}^p$'s. Those estimates are consistent, when the number of independent groups, G, increases to ∞. We then plug those estimates into the likelihood of y and use the NFXP ML method for the second step estimation.[11]

Two information structures are considered. One is complete information. The other is incomplete information where x_i^p is observed by i but not any other group members. In the experiment, we generate data from both information structures. Once a data set is generated, we do estimation under both the true and wrong information structures, computing the estimated likelihoods and the values for the information criteria, and also compute the Cox test statistics using the aforementioned bootstrap method.

For each information structure (complete and incomplete) and interaction intensity ($\lambda = 0.3$ and $\lambda = 0.6$), there are 500 repetitions. Empirical means and standard deviations of estimates are summarized in Tables 1 to 2. From those two tables, it can be seen that the NFXP ML estimation algorithm performs well for both Tobit models. Comparing the estimates of the two models, the complete information model has more accurate estimates. This may be due to more information is available in the complete information model relative to the incomplete one. As the sample size increases, the (empirical) standard deviations for estimates decrease accordingly. In addition, it can also be seen that estimation under a wrong information structure can bring in larger biases. Moreover, estimation under the true information structure can lead to larger estimated sample log likelihoods. That provides a hint on using estimated sample likelihoods to formally build a test statistics for information structures.

The results for the Cox tests computed using the bootstrap method for the samples are tabulated in Table 3. The number of bootstrap samples is $B = 100$. For each bootstrapped sample, the expected log likelihood is approximated by the mean of Monte Carlo simulated log likelihoods with $S = 100$ for each. The test for each value of the interaction intensity and with the true information structure from which the data is generated, the first two columns in Table 3 show the size and the last two columns show the power of the Cox test at three different levels of significance. Empirical sizes are close to the theoretical ones. The power of the test when the data is generated from Model 2 (incomplete information) is a bit small for small interaction intensity but become much larger for the intermediate interaction intensity, $\lambda = 0.6$. That is intuitive, as the difference between Model 1 and Model 2 resides in the way that interaction effects are specified. They would be identical if $\lambda = 0$. As the interaction intensity grows, it will be easier to distinguish between the two models. The power also increases as the sample size increases.

5 Empirical Application

We apply our model to study the relationship of property taxes for adjacent municipalities in North Carolina. By Bordigon et al. (2003), there are three ways through which the local governments' decisions on tax rates and public spendings are correlated through a spatial network. The simplest way is through a common shock. The second way is through information transmission. By comparing local taxes in their region and those in geographically related or similar regions, voters can assess the performance of their current local government officials and vote accordingly. As a result, an incumbent government official may take into account policies made by neighboring jurisdictions when making her own decision, which is called the "yardstick competition". The third way of interactions is a mobile tax base. A total tax revenue depends on both the tax rate per unit property and the tax base (such as residents, firms, and capital). The decision on tax rate in one jurisdiction will not only affect its own tax revenue but also those of others spatially related to it. In the case of property taxes, a higher rate of property tax can increase tax revenue per unit properties owned by its residents. At the same time, residents may have incentives to move to nearby municipalities which offer lower tax rates. Therefore, municipalities will act strategically to compete for taxable properties. According to Wilson and Wildasin (2004), the third one is the real tax competition. It is noted by Bordigon et al. (2003) that it is more suitable to model the "yardstick competition" by the spatial error (SE) model, as it is through the information about unobservables that taxes of different regions are related. For the competition through a mobile tax base, i.e., the third type, it is more suitable to model the competition by a SAR model. In this paper, our empirical application is about the property tax rates set by municipalities in North Carolina. Inspecting the data set, we can see that the tax rates are all non-negative, with some of them equal to 0's. As a result, it is more reasonable to use a Tobit type model framework.

We focus on the case that taxes are determined through a "strategical" competition, i.e., the third type above. Therefore, we use the framework of a simultaneous move game in Section 2. Consider n municipalities in a state. They are related according to geographic vicinity, represented by an $n \times n$ matrix W_n. For any two different cities, $i \neq j$, $W_{n,ij} = W_{n,ji} = 1$ if the distance between them is less than some cutoff value; and $W_{n,ij} = W_{n,ji} = 0$ otherwise. As usual, $W_{n,ii} = 0$, for all i.[12] We use a_i to denote the rate of property tax set by city i. A tax rate must be non-negative, i.e., $a_i \geq 0$ for all $i = 1, \cdots, n$. i's payoffs by choosing a_i when other municipalities choose a_{-i} is given by (2.4). Under complete information, the outcome of a Nash Equilibrium obeys the following rule:

$$a_i = \max\left\{x_i'\beta_1 + \lambda_1 \sum_{j\neq i} W_{n,ij}a_j + \epsilon_i, 0\right\}. \tag{5.1}$$

Under incomplete information, $(a_1, \cdots, a_n)$, is an outcome of BNE and satisfies

$$a_i = \max\left\{\beta_{2,0} + x^{c'}\beta_{2,1} + x_i^{p'}\beta_{2,2} + \lambda_2 \sum_{j\neq i} W_{n,ij}E[a_j|x^p_{J_i}, Z] + \epsilon_i, 0\right\}. \tag{5.2}$$

Therefore, we may use our model to investigate the problem of tax competition. Assume that condition $|\lambda|\|W_n\|_\infty < 1$ is satisfied. Hence, the observed data comes from the unique equilibrium in this model, which can be solved as a fixed point.

For municipalities in North Carolina, we collect data on property tax rates, government finance, and demographics in the 2012 fiscal year as well as geographic statistics (latitudes and longitudes). Data of city property tax rates are from North Carolina Department of Revenue. Information about municipal government finance comes from North Carolina Department of State Treasurer. Data about city median household income is found from "Find theData.org", which is based on the American Community Survey. Latitudes and longitudes are found from "CityLatitudeLongitude.com"[13] We calculate distance between any two cities based on latitudes and longitudes, using the Haversine formula[14]. Sample statistics are summarized in Table 4. From the table, we find a big variety among the 506 municipalities in the sample in terms of demographics and financial status. Among those municipalities, the rate of property tax is strictly positive except for 29 of them where no property taxes are levied.

In defining geographic vicinity, we tried three different cutoff values for distance between two municipalities, 30 kilometers, 50 kilometers, and 100 kilometers, and estimate model parameters under those three associated social weighting matrices. As the population of a city is correlated to the base of property tax, probably it affects the municipal tax rate. Obviously, it is public information. Thus, we include population in X^c. Since it is possible that a municipal government knows more about the financial situation of people living in its own territory than other governments do, it is reasonable to include in X^p residents' financial data in the current period. In their investigation of public safety spendings by local governments in North Carolina under a linear SAR model, Yang and Lee (2016b) take into account the possibility that the municipal median household income may be private information when local governments are making their own policies. Yang and Lee (2016b) try different specifications on the generation of the median household income in their empirical analysis. As the focus in this paper is to test whether the median household income (as well as the idiosyncratic shocks, ϵ_i's) is public information or private information, the city median household income is assumed to depend on two factors: state level median household income and city idiosyncratic income shocks. Let h_t denote the state median household income and $\epsilon^p_{i,t}$ the city-specific shocks at year t. The median household income of city i at t, $MHI_{i,t}$ is then determined by $MHI_{i,t} = h_t + \epsilon^p_{i,t}$. Suppose that for each t, h_t is independent of $\epsilon^p_{i,t}$'s and $\epsilon^p_{i,t}$'s are i.i.d. $N(0, \iota^2)$. h_t is publicly known to all municipalities. In addition to a cross section data for municipal median household incomes in 2012, we collect the time series of the median household income for the state of North Carolina from 1984 to 2012. Let $t = 0, 1, \cdots, 28$. $T = 28$ corresponds to the year 2012. Assume that $\{h_t\}$ is generated from an AR(1) process with a deterministic time trend:[15]

$$h_t = a_h + b_h t + \phi_h h_{t-1} + \epsilon^h_t, \tag{5.3}$$

for $t = 1, \cdots, T$, where a_h and b_h are constants, $|\phi_h| < 1$, ϵ^h_t's are i.i.d. $N(0, \omega^2)$. According to Sims, Stock, and Watson (1990), we can reparametrize this process and transform it into a covariance-stationary time series. Define $a^*_h = a_h/(1-\phi_h) - (b_h\phi_h)/[(1-\phi_h)^2]$, $b^*_h = b_h/(1-\phi_h)$, and $h^*_t = h_t - a^*_h - b^*_h t$. Then Eq. (5.3) can be rewritten as follows: $h^*_t = \phi_h h^*_{t-1} + \epsilon^h_t$. $\{h^*_t\}$ is a zero-mean AR(1) process such that each h^*_t is $N(0, \frac{\omega^2_h}{1-\phi^2_h})$.

Therefore, h_T is normally distributed with mean $a_h^* + b_h^* T$ and variance $\frac{\omega_h^2}{1-\phi_h^2}$. As a result, x_i^p's have an exchangeable joint normal distribution, with mean $a_h^* + b_h^* T$, variance $\eta^2 = \omega^2/(1-\phi_h^2) + \iota^2$, and correlation coefficient, $\rho = \omega^2/[\omega^2 + (1-\phi_h^2)\iota^2]$.

Two models under different information structures are estimated. The first scenario corresponds to the complete information model where municipal populations, median household incomes, and idiosyncratic shocks ϵ_i's are all known by each local government. $\delta = (\beta', \lambda, \sigma)'$ is estimated via MLE. In the second case with incomplete information, both the median household income $MHI_{i,T}$ and the shock ϵ_i are only known by i in the state. $(\theta_x', \delta')'$ is estimated by the 2-step ML algorithm where $\theta_x = (\phi_h, a_h^*, b_h^*, \omega^2, \iota^2)'$ is estimated in the first step by maximizing the log joint likelihood of X^p and $h = (h_1, \cdots, h_T)'$. In the second step, δ is estimated from the conditional log likelihood of Y conditional on X^p using the NFXP ML estimation. The log joint likelihood function of X^p and h is:

$$L_{xn}(\theta_{xn}; X^p, h) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\iota^2) - \frac{1}{2\iota^2}\sum_{i=1}^{n}(x_{i,T}^p - h_T)^2 - \frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\omega^2)$$
$$- \frac{1}{2\omega^2}\sum_{t=1}^{T}(h_t - a_h^*(1-\phi_h) - b_h^*\phi_h - b_h^*(1-\phi_h)t - \phi_h h_{t-1})^2.$$

Define $H_1(n,T) \equiv diag((\sqrt{T}, \sqrt{T}, \sqrt{T^3}, \sqrt{T}, \sqrt{n}))$. Similar to Sims, Stock, and Watson (1990), one can show that the MLE $\widehat{\theta}_{x,n}$ is consistent and $H_1(\widehat{\theta}_{xn} - \theta_0) \xrightarrow{d} N(0, \Omega_{11})$ as both n and T increase to ∞, where

$$\Omega_{11} = (-\lim_{n\to\infty, T\to\infty} E[H_1^{-1}\frac{\partial^2 L_{xn}(\theta_0; X^p, h)}{\partial\theta\partial\theta'}H_1^{-1}])^{-1}$$
$$= (diag((\frac{1}{1-\phi_{h0}^2}, \frac{(1-\phi_{h0})^2}{\omega_0^2}, \frac{(1-\phi_{h0})^2}{3\omega_0^2}, \frac{1}{2\omega_0^2}, \frac{1}{2\iota_0^2})))^{-1}.$$

The Cox test statistics can be adapted here and we can use them to test between the complete vs. incomplete information structures.

Tables 5 and 6 tabulate estimation results of the model under different vicinity cutoffs with complete and incomplete information. Across all the specifications, we see that population and median household income are both significantly related to the municipal property tax rate. The more populated a municipality is, the higher its property tax will be; and the wealthier the residents in a city, the lower the rate of property tax is. With complete information, the intensity of interactions between adjacent local governments is significantly positive for all three cutoff values. With incomplete information, when the municipal median household income is self-known, the estimated λ's are also positive, but significant only when the cutoff value is 30 or 50 kilometers and not for 100 kilometers. A positive λ shows strategic complements, as the decisions of local governments reinforce each other: when a neighboring jurisdiction increases the tax rate, a municipality will levy a higher tax rate in its own territory; and when a neighbor decreases its tax rate, this municipal government will reduce its own tax rate. Therefore, tax competition among local governments is supported by the estimated results when the cutoff values between two neighboring municipalities are not too large.

To understand more implications from magnitudes of estimates, we derive marginal effects and counterfactuals by simulations. Consider changes in covariates or network relations. (X^c, X^p) and W_n are the values before the change and $(\widetilde{X}^c, \widetilde{X}^p)$ and $\widetilde{W}_n$ are the values after the change. Random ζ_i^s for $s = 1, 2, \cdots, S$ are drawn from $N(0, 1)$. In Model m with estimates $\widehat{\theta}_m$ and $\epsilon_{mi}^s = \widehat{\sigma}_m \zeta_i^s$, generate y_{mi}^s with the values (X^c, X^p) and W_n and $\widetilde{y}_{mi}^s$ with $(\widetilde{X}^c, \widetilde{X}^p)$ and $\widetilde{W}_n$, for $m = 0, 1, 2$.[16] We focus on the simulated average tax rate before of a change $(\sum_i^n y_{mi}^s)/n$ and that after the change $(\sum_{i=1}^n \widetilde{y}_{mi}^s)/n$. For a marginal effect of population, consider the case that all the municipal populations are increased by 1%. That is, $\widetilde{x}_i^c = 1.01 x_i^c$ for all i, $\widetilde{X}^p = X^p$ and $\widetilde{W}_n = W_n$. Then the simulated marginal effect is computed as $(1/S) \sum_{s=1}^S [(\sum_{i=1}^n \widetilde{y}_{mi}^s/n) - (\sum_{i=1}^n y_{mi}^s/n)]/[(\sum_{i=1}^n y_{mi}^s/n)]$. From Tables 5 and 6, a 1% increase in municipal populations will induce an increase in tax rate on average across all vicinity cutoffs and information structures; however, the magnitude is small (less than 0.06% in all specifications). The marginal effect of a 1% increase in median household income can similarly be computed by simulations. They are negative but larger in magnitudes compared with marginal effects of populations. For all the cutoffs and information structures, the average simulated tax rate will decrease by more than 0.4%. In addition, the magnitudes are larger in Model 1 relative to those in Model 2. That may be because i's tax rate is affected by j's household income through j's tax rate in Model 1 but through i's expectation on j's tax rate in Model 2. To see the influence of local government interactions on tax rates, fix $(\widetilde{X}^c, \widetilde{X}^p) = (X^c, X^p)$ and let $\widetilde{W}_{n,ij} = 0$, i.e., no interactions. In Tables 5 and 6, we can compare the predicted average tax rates $\overline{\tau}_p = (\sum_{i=1}^n y_{mi}^s)/n$ with the original W_n and $\overline{\tau}_c = (\sum_{i=1}^n \widetilde{y}_{mi}^s)/n$ with $\widetilde{W}_n$. In both Model 1 and Model 2, $\overline{\tau}_c < \overline{\tau}_p$ due to strategic complements. Additionally, when the vicinity cutoff is either 30 or 50 kilometers, as their estimated interactions $\widehat{\lambda}$ in Model 2 are significant and larger in magnitude compared with those in Model 1, the counterfacturals without related neighbors are smaller than those in Model 1.

As different information structures may lead to different implications, it is of practical relevance to select between the two models. From the Monte Carlo study, the estimated log likelihood and information criteria are valuable in model selection. Tables 5 and 6 present the estimated sample log likelihoods and the values for AIC and BIC criteria. According the those criteria, the model without social interaction effects as estimated in Estimation (1) is mostly dominated by estimations for the models with social interactions.[17] For Tobit models with social interactions, the complete information one prevails the incomplete information one. The results for the formal Cox tests are presented in Table 7, which show that the incomplete information model can be rejected but not the complete information one with the empirical sample.

6 Conclusion

Motivated by empirical studies, we build a Tobit model with social interactions under incomplete information and compare it with its counterpart in the complete information case. A formal Cox-type test is proposed as a tool for model selection. As far as we know, our study complements researches on information structure selection in game estimation.

As all the discussions are under the prerequisite that a sufficient condition for the existence of a unique equilibrium is satisfied, it will be an appealing extension to investigate test of models in the presence of multiple equilibria. In his flexible information model, Gireco (2014) solve the problem of multiplicity by assuming a nonparametric selection rule and use the partial identification method for estimation and testing. For studies of models with a possible large interaction intensity where multiplicity is possible, it would be interesting to develop techniques to deal with multiple equilibria in non-nested tests between competing models.

Appendices

A Incomplete Information Models: Equilibrium Expectations, Identification and Calculation

A.1 Equilibrium and Identification

Following the general analysis in Yang and Lee (2016a), the existence of a unique BNE can be established under the following assumptions.

Assumption A.1. *The idiosyncratic shocks ϵ_i's are i.i.d. with the pdf, $f_\epsilon(\cdot)$, and the corresponding CDF, $F_\epsilon(\cdot)$. These idiosyncratic shocks are also independent of all exogenous characteristics and network connections.*

Assumption A.2. $E|\epsilon_i| < \infty$ *for any* $i = 1, \cdots, n$.

Assumption A.3. $\lambda_m \|W_n\|_\infty < 1$, *where* $[-\lambda_m, \lambda_m]$ *is a compact parameter space of* λ *in* $\Re^1$; *elements* $W_{n,ij}$ *of* W_n *are all non-negative and* $\|W_n\|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^n W_{n,ij}$.

As in Yang and Lee (2016a), conditional expectations on agents' behaviors are important component for the model with incomplete information, Model 2. From (2.3), fix $Z = z$, for any $k \neq i$, we have that

$$E[y_i|X^p_{J_k}, z] = E[H(\beta_0 + X^{c'}\beta_1 + X^{p'}_i\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j\neq i} W_{n,ij} E[y_j|X^p_{J_i}, Z])|X^p_{J_k}, z],$$

where $H(\cdot)$ is a real-valued function such that for any $x \in \Re^1$,

$$H(x) = \int_{-\infty}^{+\infty} I(c > -x)(x + c) f_\epsilon(c) dc = x(1 - F_\epsilon(-x)) + \int_{c>-x} c f_\epsilon(c) dc.$$

The features about censored outcomes give the function $H(\cdot)$ the above specific form, which is a monotonic function with uniformly bounded derivatives

For identification, we aim at recovering the model primitives, β, λ, F_x and F_ϵ, from observations on X and Y, given W_n and $\overline{J}_n$.

Definition A.1. *Given social relations $W_n = \overline{W}_n$ and a structure of private information, $J = \overline{J}$, $(\beta, \lambda, F_x, F_\epsilon)$ is observationally equivalent to $(\widetilde{\beta}, \widetilde{\lambda}, \widetilde{F}_x, \widetilde{F}_\epsilon)$, if* $F_{Y,X|\overline{W}_n,\overline{J}}(\cdot,\cdot|\beta,\lambda,F_x,F_\epsilon) =$

$F_{Y,X|\overline{W}_n,\overline{J}}(\cdot,\cdot|\widetilde{\beta},\widetilde{\lambda},\widetilde{F}_x,\widetilde{F}_\epsilon)$. *The true model primitives* $(\beta^*,\lambda^*,F_x^*,F_\epsilon^*)$ *is identified if any* $(\widetilde{\beta},\widetilde{\lambda},\widetilde{F}_x,\widetilde{F}_\epsilon) \neq (\beta^*,\lambda^*,F_x^*,F_\epsilon^*)$ *cannot be observationally equivalent to* $(\beta^*,\lambda^*,F_x^*,F_\epsilon^*)$.

Assumption A.4. *The distribution of exogenous characteristics,* $F_x(\cdot)$*, can be inferred from data about* X.

Assumption A.5. ϵ_i*'s are i.i.d. with the full support,* $\Re^1$*, according to a parametric pdf,* $f_\epsilon(\cdot;\sigma)$*, where the functional form,* $f_\epsilon(\cdot;\cdot)$ *is known but the parameter value of* σ *is unknown. The corresponding CDF is* $F_\epsilon(\cdot;\sigma)$*, which is strictly increasing in its argument.*

Assumption A.4 allows us to focus on β, λ, and F_ϵ. Assumption A.5 parametrizes the distribution of ϵ_i's. It allows the distribution of the idiosyncratic shocks to belong to some scale family, including the normal distribution.

First, due to the feature of the Tobit model, σ can be identified from the relationship between the mean outcome, $E[y_i|X^p,z]$, and the average amount of censoring, $E[I(y_i=0)|X^p,z]$, where X^p refers to the matrix of all privately known characteristics, $(X_1^{p\prime},\cdots,X_n^{p\prime})'$.[18]

Lemma A.1. *Given public information,* $Z=z$*, for all* $i=1,\cdots,n$,

$$\begin{aligned} E[y_i|X^p,z] =& -F_\epsilon^{-1}(E[I(y_i=0)|X^p,z];\sigma)(1-E[I(y_i=0)|X^p,z]) \\ &+\int_{c>F_\epsilon^{-1}(E[I(y_i=0)|X^p,z];\sigma)} cf_\epsilon(c;\sigma)dc. \end{aligned} \tag{A.1}$$

Assumption A.6. $f_\epsilon(c;\sigma)$ *is differentiable with respect to* σ*; and* $\lim_{c\to\infty} c\frac{\partial F_\epsilon(c;\sigma)}{d\sigma}=0$.

Assumption A.7. *The ratio,* $\frac{\partial F_\epsilon(c;\sigma)}{\partial\sigma}/f_\epsilon(c;\sigma)$*, is strictly monotonic with respect to* c.

As an example of Assumptions A.6 and A.7, consider the case that ϵ_i is normally distributed with zero mean and standard deviation σ. Then, $F_\epsilon(c;\sigma)=\Phi(c/\sigma)$ and $f_\epsilon(c;\sigma)=\frac{1}{\sigma}\phi(\frac{c}{\sigma})$, where $\Phi(\cdot)$ and $\phi(\cdot)$ are respectively the CDF and pdf of the standard normal random variable. $\lim_{c\to\infty} c\frac{\partial F_\epsilon(c;\sigma)}{\partial\sigma}=\lim_{c\to\infty}-(\frac{c}{\sigma})^2\phi(\frac{c}{\sigma})=0$ and $\frac{\partial F_\epsilon(c;\sigma)}{\partial\sigma}/f_\epsilon(c;\sigma)=-(\frac{c}{\sigma^2}\phi(\frac{c}{\sigma}))/(\frac{1}{\sigma}\phi(\frac{c}{\sigma}))=-\frac{c}{\sigma}$, which is decreasing in c. Therefore, the sufficient conditions in Assumptions A.6 and A.7 are satisfied.

Proposition A.1. *For any* W_n *and information structure,* J*, if Assumptions A.4 to A.7 are satisfied,* σ *can be identified from the moments,* $E[y|X^p,z]$ *and* $E[I(y=0)|X^p,z]$.

The proof of Proposition A.1 depends on the relationship, (A.1), which is valid with any information structure on X. Actually, it is more transparent to see the identification of the standard deviation in the normal disturbance case through (A.1). By calculation, we can get that

$E[y|X^p, z] = \sigma\Big(\phi(\Phi^{-1}(E[I(y=0)|X^p, z])) - \Phi^{-1}(E[I(y=0)|X^p, z])(1 - E[I(y=0)|X^p, z])\Big)$,

which implies that

$\sigma = E[y|X^p, z] / \Big(\phi(\Phi^{-1}(E[I(y=0)|X^p, z])) - \Phi^{-1}(E[I(y=0)|X^p, z])(1 - E[I(y=0)|X^p, z])\Big)$.

In principle, with appropriate empirical observations, $E[y|X^p, z]$ and $E[I(y=0)|X^p, z]$ might be nonparametrically identified from empirical observations. If so, we can identify σ for any information structure.

Next, we turn to identification of other parameters. Consider the case of a single group. In that case, any group characteristics, X^g, is absorbed by the constant term. Hence, we focus on the identification of β_0, β_1, β_2 and λ.[19] For that purpose, we impose two additional assumptions.

Assumption A.8. *Given social relations $\overline{W}_n$ and an information structure, $\overline{J}$, $E[Y_i|X^p_{J_k}, Z]$ can be identified (nonparametrically), for any $i, k = 1, \cdots, n$.*

Assumption A.9. $\max_{1\leq i\leq n}\Big|det\Big(Var(\Big(X^c_i, \quad X^p_i, \quad \sum_{j\neq i} W_{i,j} E[y_j|X^p_{J_i}, Z]\Big))\Big)\Big| > 0.$

Proposition A.2. *With a given W_n and J, if Assumptions A.4 to A.9 hold, β_0, β_1, β_2, and λ can be identified.*

A.2 Equilibrium Calculation

As it is shown in Yang and Lee (2016a), the function of equilibrium conditional expectations, denoted by ψ^e, can be viewed as a fixed point of an operator in a function space, denoted as T. If $|\lambda|\|W_n\|_\infty < 1$, that operator will be a contraction mapping. Owing to properties of a contraction mapping in a complete metric space $(\Xi, \|\cdot\|)$, $\lim_{l\to\infty} T^l(\xi^0) = \psi^e$, for any $\xi^0 \in \Xi$, according to the norm $\|\cdot\|$. Thus, beginning with any initial guess, by iterating the operator T, we can approximate the equilibrium conditional expectation function, ψ^e, to derive the corresponding BNE. However, in general, since ψ^e is a function of random vectors, to solve it means to derive its realizations for every point on the underlying sample space, which is by no way a trivial task.

Nonetheless, ψ^e changes with elements in the sample space indirectly through the random vector X^p and a predetermined information structure J. Therefore, if X_i^p's are discrete random vectors with a finite support, it suffices to characterize ψ^e by its values on those points, which makes it possible to represent ψ^e by a finite dimensional vector. Although that representation is not applicable when X_i^p's vary continuously on a continuum support, due to the use of quadrature for stochastic integrals as an approximation of integrals generated by expectations, we can approximate every possible realization of ψ^e by a finite dimensional vector. To elucidate the computation, we begin with the simplest case and then move on to more complicated ones.

A.2.1 All Characteristics are Publicly Known

When all exogenous characteristics, X^g, X^c, X^p, are public information, there is no uncertainty other than idiosyncratic shocks, which are i.i.d. and independent of all exogenous characteristics. Given $Z = z$, ψ^e reduces to an $n \times 1$ vector, satisfying

$$\psi_i^e = H(\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j^e), \tag{A.2}$$

for all $i = 1, \cdots, n$. Although we cannot get an analytical solution for the Tobit model, due to the nonlinearity of the function $H(\cdot)$, we can solve ψ^e numerically by contraction mapping iterations at each given vector of parameters.

A.2.2 Self-Known Characteristics

If X_i^p is realized only to i (and econometricians), we call the information structure as "self-known characteristics", in which case $J_i(k) = 0$ for all $k \neq i$. Then any two different agents do not share private information. For $i \neq k$, we have that

$$\psi_i^e(X_k^p) = E[H(\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j^e(X_i^p))|X_k^p, z]. \tag{A.3}$$

Inspecting (A.3), if all X_i^p's are independent of each other conditional on $Z = z$, the realization of X_k^p does not provide new information on X_i^p given $Z = z$. That is to say,

$$\psi_i^e(X_k^p) = E[H(\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j^e(X_i^p))|z],$$

for any $i \neq k$. Since the right-hand side does not depend on the random vector X_k^p, we can view expectations on i's behaviors as a constant. Therefore, with independent self-known characteristics, the equilibrium conditional expectation is a constant vector, such that

$$\psi_i^e = E[H(\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j^e)|z], \tag{A.4}$$

Comparing (A.2) and (A.4), we see that in both cases every two agents $k \neq k^{'}$ have the same expectation on the behavior of a third person, i. However, when all exogenous characteristics are public information, k and $k^{'}$ just integrate over unobserved idiosyncratic shocks in (A.2); but in contrast, when they know just their only realizations for X^p, they have to integrate over X_i^p to predict i's behaviors in (A.4), for X_i^p is not included in Z.

In general, $(X_1^p, \cdots, X_n^p)$ might have a joint distribution such that X_i^p's are correlated. With correlation, conditional expectations would depend on specific private information used to make predictions. Scrutinizing (A.3), if two agents k and $k^{'}$ link to i, i.e., $W_{n,ki} \neq 0$ and $W_{n,k^{'}i} \neq 0$, their private information influences predictions on i's behaviors through the conditional distributions $F_p(X_i^p|X_k^p, z)$ and $F_p(X_i^p|X_{k^{'}}^p, z)$. In general, the two conditional distributions might differ, and k and $k^{'}$ might form different conditional expectations. However, there are circumstances that the two conditional distributions can be the same. In that case, those two agents' predictions on i's behaviors will be identical once they get the same realizations, i.e., $X_k^p = x$ and $X_{k^{'}}^p = x$, where x is a realization. A sufficient condition for such a circumstance is "exchangeability", cited below from Yang and Lee (2016a):

Assumption A.10. *Conditional on public information, $Z = z$, X_i^p's have the same support, $\mathfrak{S}_p$. Their conditional joint distribution, $F^p(X_1^p, \cdots, X_n^p|Z = z)$, is exchangeable, i.e., for any permutation, $s : \{1, \cdots, n\} \to \{1, \cdots, n\}$, $F^p(X_1^p, \cdots, X_n^p|Z = z) = F^p(X_{s(1)}^p, \cdots, X_{s(n)}^p|Z = z)$.*

Under "exchangeability", if $X_k^p = X_{k^{'}}^p = x$,

$$\begin{aligned}
\psi_i^e(X_k^p = x) =& E[H(\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j^e(X_i^p))|X_k^p = x, z] \\
=& E[H(\beta_0 + X_i^{c'}\beta_1 + X_i^{p'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j^e(X_i^p))|X_{k^{'}}^p = x, z] \\
=& \psi_i^e(X_{k^{'}}^p = x).
\end{aligned}$$

for any $k, k^{'}$ with $W_{n,ki} \neq 0$ and $W_{n,k^{'}i} \neq 0$. Therefore, we can directly define ψ_i^e on the common support of X_i^p's $\mathcal{S}_p$, and characterize ψ^e by

$$\psi_i^e(x) = E[H(\beta_0 + X_i^{c'}\beta_1 + y^{'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j^e(y))|x, z],$$

for all $i = q, \cdots, n$ and $x \in \mathcal{S}_p$. In the following discussion, we consider the computation of conditional expectations for two classes of joint distributions that satisfy Assumption A.10.

1. (Discrete X^p) Suppose that X_i^p can only take one of m values in $\{x^l : 1 \leq l \leq m\}$, where each x^l is a vector with specific values. Given public information $Z = z$, the conditional probability function, $f_p(y|x, z)$, is fully captured by an $m \times m$ transition matrix, $P = (P_{ll'})$, where $p_{ll'} = prob(X_i^p = x^{l'}|X_k^p = x^l, Z = z)$, for any i and k such that $k \neq i$. We can represent ψ^e by an $(nm) \times 1$ vector, $(\psi_1^e(x^1), \cdots, \psi_1^e(x^m), \cdots, \psi_n^e(x^1), \cdots, \psi_n^e(x^m))'$, and characterize it by the following system of nonlinear equations:
$$\psi_i^e(x^l) = \sum_{\tilde{l}=1}^{m} p_{l\tilde{l}} H(\beta_0 + X_i^{c'}\beta_1 + x^{\tilde{l}'}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j \neq i} W_{n,ij}\psi_j^e(x^{\tilde{l}})), \qquad \text{(A.5)}$$
for $i = 1, \cdots, n$ and $l, \tilde{l} = 1, \cdots, m$. Beginning with any initial $(nm) \times 1$ vector and iterating the contraction mapping, we can derive ψ^e.

2. (Continuous X^p) Consider the case when $X_1^p, \cdots, X_n^p$ are jointly normal with a common mean μ, a common variance Σ_1, and a common covariance Σ_2 between any X_i^p and X_j^p with $i \neq j$. Then for any $i \neq j$, conditional on $X_j^p = X$, X_i^p is normal with mean $\mu + \Sigma_2\Sigma_1^{-1}(X - \mu)$ and variance $\Sigma_1 - \Sigma_2\Sigma_1^{-1}\Sigma_2'$. In this case, each X_i^p can take any value in $\Re^{k_p}$. Thus, it is impossible to represent ψ^e by a finite dimensional vector. However, for any i and $x \in \Re^{k_p}$, $\psi_i^e(x)$ is determined by an integral, which can be approximated by values of the function on a fixed number of values (quadrature points) using the quadrature method. For illustration, consider the special case that each X_i^p is a single random variable, i.e., its dimension $k_p = 1$. In this case, denote $\Sigma_1 = \eta^2$ and $\Sigma_2 = \rho\eta^2$. We first transform integration in $\Re^1$ into integration over a finite interval, $[-1, 1]$, and then apply the Gauss-Legendre

quadrature.

$$\begin{aligned}
\psi_i^e(x) =& \int_{-\infty}^{+\infty} H(\beta_0 + X_i^{c'}\beta_1 + \widetilde{x}\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j\neq i} W_{n,ij}\psi_j^e(\widetilde{x})) \\
& \cdot \frac{1}{\sqrt{2\pi(1-\rho^2)\eta^2}} \exp(-\frac{(\widetilde{x} - \rho x - (1-\rho)\mu)^2}{2(1-\rho^2)\eta^2}) d\widetilde{x} \\
=& \sqrt{\frac{2}{\pi(1-\rho^2)\eta^2}} \int_{-1}^{1} H(\beta_0 + X_i^{c'}\beta_1 + \log(\frac{\widetilde{z}+1}{1-\widetilde{z}})\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j\neq i} W_{n,ij}\psi_j^e(\log(\frac{\widetilde{z}+1}{1-\widetilde{z}}))) \\
& \cdot \exp(-\frac{(\log(\frac{\widetilde{z}+1}{1-\widetilde{z}}) - \rho x - (1-\rho)\mu)^2}{2(1-\rho^2)\eta^2}) \frac{1}{(\widetilde{z}+1)(1-\widetilde{z})} d\widetilde{z} \\
\approx& \sqrt{\frac{2}{\pi(1-\rho^2)\eta^2}} \sum_{k=1}^{K} \omega_k H(\beta_0 + X_i^{c'}\beta_1 + \log(\frac{\widetilde{z}_k+1}{1-\widetilde{z}_k})\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j\neq i} W_{n,ij}\psi_j^e(\log(\frac{\widetilde{z}_k+1}{1-\widetilde{z}_k}))) \\
& \cdot \exp(-\frac{(\log(\frac{\widetilde{z}_k+1}{1-\widetilde{z}_k}) - \rho x - (1-\rho)\mu)^2}{2(1-\rho^2)\eta^2}) \frac{1}{(\widetilde{z}_k+1)(1-\widetilde{z}_k)},
\end{aligned} \tag{A.6}$$

In Eq A.6, the second equality is derived by a change of integration variable, $\widetilde{x} = \log((\widetilde{z}+1)/(1-\widetilde{z}))$. At last, the approximation is based on standard Gauss-Legendre quadrature, where $\widetilde{z}_k$'s are the abscissae, ω_k's are the corresponding weights, and K is the number of abscissae. Define accordingly, $x_k^p = \log(\frac{\widetilde{z}_k+1}{1-\widetilde{z}_k})$, for $k = 1, \cdots, K$, we get nK equalities,

$$\begin{aligned}
\psi_i^e(x_{k'}^p) =& \sqrt{\frac{2}{\pi(1-\rho^2)\eta^2}} \sum_{k=1}^{K} \omega_k H(\beta_0 + X_i^{c'}\beta_1 + x_k^p\beta_2 + X^{g'}\beta_3 + \lambda \sum_{j\neq i} W_{n,ij}\psi_j^e(x_k^p)) \\
& \cdot \exp(-\frac{(x_k^p - \rho x_{k'}^p - (1-\rho)\mu)^2}{2(1-\rho^2)\eta^2}) \frac{1}{(\widetilde{z}_k+1)(1-\widetilde{z}_k)},
\end{aligned}$$

for all $i = 1, \cdots, n$ and $k' = 1, \cdots, K$. This is very similar to (A.5). Hence, we can solve $\psi_i^e(x_{k'}^p)$'s by contraction mapping iterations. After that, for any $x \in \Re^1$, we can approximate $\psi_i^e(x)$ by (A.6). Owing to the fast convergence of the Gauss-Legendre quadrature, we only need to take a small number of abscissae. In our Monte Carlo experiments, we find that good performance can be achieved in estimation by choosing $K = 8$.

When X_i^p's are of multiple dimensions, multiple-dimension quadrature methods are available but can be computationally intensive. Alternatively, we can use the stochastic integral approximation with importance sampling. Let $h(a_i)$ be a density with its support containing the support of X_i^p so that $f_p(x^p|x)/h(x^p)$ is well defined. Then we can generate K random draws, x_k^p, from $h(\cdot)$. The stochastic approximation will be

$$\psi_i^e(x) \approx \frac{1}{K}\sum_{k=1}^{K} H(\beta_0 + X_i^{c'}\beta_1 + x_k^p\beta_2 + X^{g'}\beta_3 + \lambda\sum_{j\neq i} W_{n,ij}\psi_j^e(x_k^p))\frac{f_p(x_k^p|x)}{h(x_k^p)}.$$

Analogous to previous discussions, we solve $\psi^e(x_k^p)$'s by contraction mapping and then approximate the function $\psi_i^e(x)$ at any point x.

For general information structures without an exchangeable joint distribution, the unique equilibrium can be calculated in a similar way. When all X_i^p's are discrete random vectors with a finite support, we can fully solve ψ^e directly via contraction mapping iterations. When X_i^p's are continuous random variables, we choose a finite number of points for each agent and approximate relevant integration in conditional expectations by a weighted sum. For continuous X_i^p's with stochastic simulation, if the number of simulated points for each agent is the same, say K, the total number of equations for solution will be nK. Since values of ψ^e on those finite number of variables can be solved as a vector by contraction mapping iteration, the values of ψ^e for all realizations can be approximated.

A.3 Group Random Effects

In a sample with groups, there are some possible common factors which can influence all group members but are unknown to econometricians. For example, when making decisions on tax rates, municipal officers know the lobbying power of different parties. Nonetheless, there might not be a measure about that from data sets. Such factors are group unobservables. In this paper, we model group unobservables as random effects.

To be specific, consider G independent groups. For a group g, in addition to X^g, $X^{c,g}$, and $X^{p,g}$, there can be another set of unobserved (for econometricians) group features which can be lumped together into a variable, ω^g. Assume that ω^g is independent of other variables and has a zero mean. The observed censored outcome satisfies

$$y_{i,g} = \max\left\{\beta_0 + X_{i,g}^{c'}\beta_1 + X_{i,g}^{p'}\beta_2 + X^{g'}\beta_3 + \omega^g + \lambda\sum_{j\neq i} W_{n_g,ij}E[y_{j,g}|X_{J_{i,g}}^{p,g}, Z] - \epsilon_{i,g}, 0\right\}.$$

Because ω^g is in public information for agents, its presence will be similar to X^g in the analysis of equilibrium expectation and behaviors. For econometricians, ω^g may be taken as a random variable with zero mean and independent of all X's in order to identify the intercept and β_3 in the presence of unobserved group variables in X^g. The distribution of ω^g can be identified through variation across different groups. However, when unobservable group random effects are taken into account, our estimation method will need to be modified. Since ω^g is unobservable to econometricians, one needs to integrate over its density function $f_\omega(\cdot)$ to construct the sample likelihood function. Denote the size of group g by n_g,

$$
\begin{aligned}
&\log L(Y|X^c, X^p, X^g, W) = \\
&\sum_{g=1}^{G} \log \Big[\int \prod_{i=1}^{n_g} \big(f_\epsilon(y_{i,g} - (\beta_0 + X_{i,g}^{c'}\beta_1 + X_{i,g}^{p'}\beta_2 + X^{g'}\beta_3 + \omega^g + \lambda \sum_{j\neq i} W_{n_g,ij} E[y_{j,g}|X^{p,g}_{J_{i,g}}, z]); \sigma)^{I(y_{i,g}>0)} \\
&\cdot F_\epsilon(\beta_0 + X_{i,g}^{c'}\beta_1 + X_{i,g}^{p'}\beta_2 + X^{g'}\beta_3 + \omega^g + \lambda \sum_{j\neq i} W_{n_g,ij} E[y_{j,g}|X^{p,g}_{J_{i,g}}, z]; \sigma)^{I(y_{i,g})=0}\big) f_\omega(\omega^g) d\omega^g \Big].
\end{aligned}
$$

In estimation, we may use stochastic integration to approximate the integration over the unobserved group random effects, ω^g's. That is, we obtain S independent draws, $\omega^{g,s}$, for $s = 1, \cdots, S$ for each group $g = 1, \cdots, G$ from the density $f_\omega(\cdot;\gamma)$, to construct a simulated sample log likelihood function:

$$
\begin{aligned}
&\log \overline{L}(Y|X^c, X^p, X^g, W) = \\
&\sum_{g=1}^{G} \log \Big[\frac{1}{S} \sum_{s=1}^{S} \prod_{i=1}^{n_g} \big(f_\epsilon(y_{i,g} - (\beta_0 + X_{i,g}^{c'}\beta_1 + X_{i,g}^{p'}\beta_2 + X^{g'}\beta_3 + \omega^{g,s} + \lambda \sum_{j\neq i} W_{n_g,ij} E[y_{j,g}|X^{p,g}_{J_{i,g}}, z]); \sigma)^{I(y_{i,g}>0)} \\
&\cdot F_\epsilon(\beta_0 + X_{i,g}^{c'}\beta_1 + X_{i,g}^{p'}\beta_2 + X^{g'}\beta_3 + \omega^{g,s} + \lambda \sum_{j\neq i} W_{n_g,ij} E[y_{j,g}|X^{p,g}_{J_{i,g}}, z]; \sigma)^{I(y_{i,g})=0}\big) \Big].
\end{aligned}
$$

In estimation, we will use the NFXP ML algorithm, replacing the true likelihood with the above simulated one. Results of Monte Carlo experiments for this estimation method can be found in the supplementary file.

B Asymptotic Distributions of the Test Statistics

B.1 Model Assumptions

To investigate rigorously the asymptotic distribution of the Cox test statistics, we list more concise regularity conditions for the models below.

Assumption B.1. *Individual units in the economy are located in a region* $D_n \subset D \subset \Re^d$ *where the cardinality* $|D_n|$ *of* D_n *goes to infinity as* $n \to \infty$. *The distance between two individuals* $i \neq j$, $d(i,j)$, *is larger than or equal to a specific positive constant, which may be normalized to be 1, without loss of generality.*

Assumption B.2. *The weights* $W_{n,ij}$ *are all non-negative. They satisfy either of the following two conditions, or both. (1) Only individuals whose distances are no larger than some specific constant, say* $\overline{d}_0 > 1$, *can affect each other. That is,* $W_{n,ij} = 0$ *if* $d(i,j) > \overline{d}_0$ *for* $i \neq j$; *(2) For every* n, *the number of columns of* W_n *with* $\lambda_0 \sum_{i=1}^{n} w_{n,ij} > \lambda_m \|W_n\|_\infty$ *is less than or equal to some fixed number, say* N *that is independent of* n; *and there exists an* $\alpha > d$ *and a constant* C_0 *such that* $|W_{n,ij}| < C_0/d^\alpha(i,j)$, *where* λ_0 *is the interaction intensity in the true DGP for Model 1 or 2 and* d *is the dimension of agents' location space in Assumption B.1.*

Assumption B.3. $\sup_{k,i,n} E|x_{ik,n}|^2 < \infty$ *and* $\sup_{i,n} E|\epsilon_{i,n}|^2 < \infty$.

Assumption B.4. *(1)*$\{x_{i,n}\}_{i=1}^n$ *is an* α*-mixing random fields with* α*-mixing coefficient* $\alpha(u,v,r) \leq (u+v)^\tau \widehat{\alpha}(r)$ *for some* $\tau \geq 0$ *with* $\sum_{r=1}^{\infty} r^{d-1}\widehat{\alpha}(r) < \infty$; *(2)* $\sup_{i,k,n} \|x_{ki,n}\|_{4+\Delta} < \infty$ *for some* $\Delta > 0$.

Assumption B.5. *The parameter space* Θ *is a compact subset in a finite dimensional Euclidean space.*

Assumption B.6. θ_0 *is interior in* Θ *and* $\limsup_{n\to\infty}[E\log L_n(\theta) - E\log L_n(\theta_0)] < 0$ *for any* $\theta \neq \theta_0$, *where* $L_n(\cdot)$ *is the likelihood function of the DGP, which can be Model 1 or Model 2, and* θ_0 *is the true parameter vector of the DGP.*

Assumption B.7. *(1)* $\sup_{i,k,n} \|x_{ik,n}\|_{8+\delta} < \infty$ *for some* $\delta > 0$; *(2) For some* $0 < \delta' < 2 + \frac{\delta}{2}$, *the* α*-mixing coefficient of* $\{x_{i,n}\}_{i=1}^n$ *in Assumption B.4 satisfies* $\sum_{r=1}^{\infty} r^{d(\tau_*+1)-1}\widehat{\alpha}(r)^{\frac{\delta'}{4+2\delta'}} < \infty$ *for* $\tau_* = \delta'/(2+\delta')$.

Assumption B.8. $\Sigma_0 = \lim_{n\to\infty}\Sigma_n$ *exists and is nonsingular where* $\Sigma_n = \frac{1}{n}var(\sum_{i=1}^{n} q_{i,n}(\theta_0))$ *with* $q_{i,n}(\theta_0)$ *being* $q_{1i}(\theta_0)$ *of Model 1 in Eq.*(B.5) *or* $q_{2i}(\theta_0)$ *of Model 2 in Eq.*(B.6).

Assumption B.9. *The* α *in Assumption B.2 (II) and* δ *in Assumption B.7 satisfy* $\alpha > d\max\{7 + 24\delta^{-1}, 5 + 32\delta^{-1} + 64\delta^{-2}\}$, *where* d *is the dimension of the space where agents are located as in Assumption B.1.*

Under Assumptions A.1 to A.3 and B.1 to B.9 Xu and Lee (2015) show that the ML estimator of Model 1 is consistent and asymptotically normal. The derivation of the distributions of the Cox tests is based on analysis of the distributions of the ML estimators under Model 1 and 2. In our model with a simple network with exchangeable regressors such as $x_i^p = h + \epsilon_i^p$, h can be treated as a constant for analysis so the mixing condition in Assumption B.4 can be satisfied as long as ϵ_i^p's are mixing across i.

B.2 Pseudo True Values

The first step estimate $\widehat{\theta}_x$ has no effects on the second step estimation of δ_1 in Model 1 because $\ln f_1(Y_n|X_n, \delta_1)$ does not depend on $\widehat{\theta}_x$. Under the hypothesis H_0 that Model 1 is the true model, the consistency and asymptotic normality of the MLE $\widehat{\delta}_{1n}$ is derived in Xu and Lee (2015). It has $\sqrt{n}(\widehat{\delta}_{1n} - \delta_{10}) \xrightarrow{d} N(0, \Sigma_{10}^{-1})$, where Σ_{10} is a positive definite matrix. For $\widehat{\delta}_{2n}$ of Model 2, which is a misspecified model in this case, we have the following property.

Proposition B.1. *Suppose that the true DGP is Model 1 with* θ_{10} *generating the data and Assumptions A.1 to A.9 and B.1 to B.9 are satisfied, then* $\widehat{\delta}_{2n} - \overline{\delta}_{2n}(\theta_{10}) = o_p(1)$ *and* $\sqrt{n}(\widehat{\delta}_{2n} - \overline{\delta}_{2n}(\theta_{10})) \xrightarrow{d} N(0, \Omega_{21})$, *where* Ω_{21} *is a positive definite matrix.*

Next we consider H_0 : Model 2 with true parameter θ_{20}. We have the following proposition:

Proposition B.2. *Suppose that the true DGP is Model 2 with* θ_{20} *generating the data and Assumptions A.1 to A.9 and B.1 to B.9 are satisfied, then* $\sqrt{n}(\widehat{\delta}_{1n} - \overline{\delta}_{1n}(\theta_{20})) \xrightarrow{d} (0, \Omega_{12})$ *for some variance matrix* Ω_{12} *and* $\sqrt{n}(\widehat{\delta}_{2n} - \delta_{20}) \xrightarrow{d} N(0, \Omega_{22})$, *where* Ω_{22} *is a positive definite matrix.*

B.3 Distributions of the Test Statistics

Under H_0 : true model l with parameter θ_{l0},

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\Big(\overline{L}_{mn}(\overline{\delta}_{mn}(\theta_{l0}),\theta_{x0}:\theta_{l0}) - \overline{L}_{mn}(\widehat{\delta}_{mn},\widehat{\theta}_{xn}:\widehat{\theta}_{ln})\Big)\\
=&\frac{1}{\sqrt{n}}[\overline{L}_{mn}(\overline{\delta}_{mn}(\theta_{l0}),\theta_{x0}:\theta_{l0}) - \overline{L}_{mn}(\widehat{\delta}_{mn},\theta_{x0}:\theta_{l0}) + \overline{L}_{mn}(\widehat{\delta}_{mn},\theta_{x0}:\theta_{l0}) - \overline{L}_{mn}(\widehat{\delta}_{mn},\widehat{\theta}_{xn}:\theta_{l0})\\
&+\overline{L}_{mn}(\widehat{\delta}_{mn},\widehat{\theta}_{xn}:\theta_{l0}) - \overline{L}_{mn}(\widehat{\delta}_{mn},\widehat{\theta}_{xn}:\widehat{\theta}_{ln})]\\
=&-[\widehat{\delta}_{mn}-\overline{\delta}_{mn}(\theta_{l0})]'\frac{\partial^2\overline{L}_{mn}(\widetilde{\widetilde{\delta}}_{mn},\theta_{x0}:\theta_{l0})}{2n\partial\delta_m\partial\delta_m'}\sqrt{n}[\widehat{\delta}_{mn}-\overline{\delta}_{mn}(\theta_{l0})]'\\
&-\frac{\partial\overline{L}_{mn}(\widehat{\delta}_{mn},\widetilde{\theta}_{xn}:\theta_{l0})}{n\partial\theta_x'}\sqrt{n}(\widehat{\theta}_{xn}-\theta_{x0}) - \frac{\partial\overline{L}_{mn}(\widehat{\delta}_{mn},\widehat{\theta}_{xn}:\widetilde{\theta}_{ln})}{n\partial\theta_l'}\sqrt{n}(\widehat{\theta}_{ln}-\theta_{l0})\\
=&-\frac{\partial\overline{L}_{mn}(\overline{\delta}_{mn}(\theta_{l0}),\theta_{x0}:\theta_{l0})}{n\partial\theta_x'}\sqrt{n}(\widehat{\theta}_{xn}-\theta_{x0}) - C_{mn,l}(\overline{\delta}_{mn}(\theta_{l0}),\theta_{x0}:\theta_{l0})\sqrt{n}(\widehat{\theta}_{ln}-\theta_{l0}) + o_p(1),
\end{aligned}
\tag{B.1}
$$

where $\widetilde{\widetilde{\delta}}_{mn}$ is a value between $\widehat{\delta}_{mn}$ and $\overline{\delta}_{mn}(\theta_{l0})$, $\widetilde{\theta}_{xn}$ is a value between $\widehat{\theta}_{xn}$ and θ_{x0}, $\widetilde{\theta}_{ln}$ is a value between $\widehat{\theta}_{ln}$ and θ_{l0}, $C_{mn,l}(\delta_m,\theta_x:\theta_l) = \frac{1}{n}\frac{\partial\overline{L}_{mn}(\delta_m,\theta_x:\theta_l)}{\partial\theta_l'}$ is the derivative only with respect to the argument θ_l of $\overline{L}_{mn}(\delta_m,\theta_x:\theta_l)$, where θ_l is from the conditional pdf of $f_l(Y_n|X_n,\theta_l)$, and $\frac{\partial\overline{L}_{mn}(\theta_m,\theta_x:\theta_l)}{\partial\theta_x'}$ is the derivative with respect to θ_x in $\overline{L}_{mn}(\delta_m,\theta_x:\theta_l)$, where θ_x is from $L_{mn}(\delta_m,\theta_{xm})$.

Similarly,

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}[L_{mn}(\widehat{\delta}_{mn},\widehat{\theta}_{xn}) - L_{mn}(\overline{\delta}_{mn}(\theta_{l0}),\theta_{x0})]\\
=&\frac{1}{\sqrt{n}}[L_{mn}(\widehat{\delta}_{mn},\widehat{\theta}_{xn}) - L_{mn}(\overline{\delta}_{mn}(\theta_{l0}),\widehat{\theta}_{x0})] + \frac{1}{\sqrt{n}}[L_{mn}(\overline{\delta}_{mn}(\theta_{l0}),\widehat{\theta}_{x0}) - L_{mn}(\overline{\delta}_{mn}(\theta_{l0}),\theta_{x0})]\\
=&\frac{\partial L_{mn}(\overline{\delta}_{mn}(\theta_{l0}),\widehat{\theta}_{xn})}{n\partial\theta_x'}\sqrt{n}(\widehat{\theta}_{xn}-\theta_{x0}) - [\widehat{\delta}_{mn}-\overline{\delta}_{mn}(\theta_{l0})]'\frac{\partial^2 L_{mn}(\widetilde{\widetilde{\delta}}_{mn},\theta_{x0})}{2n\partial\delta_m\partial\delta_m'}\sqrt{n}[\widehat{\delta}_{mn}-\overline{\delta}_{mn}(\theta_{l0})]\\
=&\frac{\partial L_{mn}(\delta_m,\theta_{x0})}{n\partial\theta_x'}|_{\delta_m=\overline{\delta}_{mn}(\theta_{l0})}\sqrt{n}(\widehat{\theta}_{xn}-\theta_{x0}) + o_p(1).
\end{aligned}
\tag{B.2}
$$

The preceding Eqs.(B.1) and (B.2) are two components of the asymptotic distribution of the Cox test statistic. For the asymptotic distributions of the test statistics under the true model l, we consider

$$
\begin{aligned}
& \frac{1}{\sqrt{n}}\left(L_{mn}(\widehat{\delta}_{mn},\widehat{\theta}_{xn})-\overline{L}_{mn}(\widehat{\delta}_{mn},\widehat{\theta}_{xn}:\widehat{\theta}_{ln})\right) \\
= & \frac{1}{\sqrt{n}}[L_{mn}(\widehat{\delta}_{mn},\widehat{\theta}_{xn})-L_{mn}(\overline{\delta}_{mn}(\theta_{l0}),\theta_{x0})+L_{mn}(\overline{\delta}_{mn}(\theta_{l0}),\theta_{x0})-\overline{L}_{mn}(\overline{\delta}_{mn}(\theta_{l0}),\theta_{x0}:\theta_{l0}) \\
& +\overline{L}_{mn}(\overline{\delta}_{mn}(\theta_{l0}),\theta_{x0}:\theta_{l0})-\overline{L}_{mn}(\widehat{\delta}_{mn},\widehat{\theta}_{xn}:\widehat{\theta}_{ln})] \\
= & \left(\frac{\partial L_{mn}(\delta_m,\theta_{x0})}{n\partial\theta_x'}|_{\delta_m=\overline{\delta}_{mn}(\theta_{l0})}-\frac{\partial\overline{L}_{mn}(\overline{\delta}_{mn}(\theta_{l0}),\theta_{x0}:\theta_{l0})}{n\partial\theta_x'}\right)\sqrt{n}(\widehat{\theta}_{xn}-\theta_{x0})+o_p(1) \\
& +\frac{1}{\sqrt{n}}[L_{mn}(\overline{\delta}_{mn}(\theta_{l0}),\theta_{x0})-\overline{L}_{mn}(\overline{\delta}_{mn}(\theta_{l0}),\theta_{x0}:\theta_{l0})]-C_{mn,l}(\overline{\theta}_{mn}(\theta_{l0}),\theta_{x0}:\theta_{l0})\sqrt{n}(\widehat{\theta}_{ln}-\theta_{l0}) \\
= & \frac{1}{\sqrt{n}}[L_{mn}(\overline{\delta}_{mn}(\theta_{l0}),\theta_{x0})-\overline{L}_{mn}(\overline{\delta}_{mn}(\theta_{l0}),\theta_{x0}:\theta_{l0})]-C_{mn,l}(\overline{\theta}_{mn}(\theta_{l0}),\theta_{x0}:\theta_{l0})\sqrt{n}(\widehat{\theta}_{ln}-\theta_{l0})+o_p(1),
\end{aligned}
$$

where the last "=" holds because

$$
\frac{1}{n}\left(\frac{\partial L_{mn}(\delta_m,\theta_{x0})}{n\partial\theta_x'}|_{\delta_m=\overline{\delta}_{mn}(\theta_{l0})}-\frac{\partial\overline{L}_{mn}(\overline{\delta}_{mn}(\theta_{l0}),\theta_{x0}:\theta_{l0})}{n\partial\theta_x'}\right)=o_p(1)
$$

for both $m=1$ and 2, as $\frac{\partial L_{1n}(\delta_1,\theta_x)}{\partial\theta_x'}=0$ and $\frac{\partial L_{2n}(\delta_2,\theta_x)}{\partial\theta_x'}$ is the sum of n terms with each term satisfying the spatial NED, hence the LLNs can be applied.

Therefore, under the null of Model 1,

$$
\begin{aligned}
& Cox_1(Y_n,X_n)=\frac{1}{\sqrt{n}}\left(L_{2n}(\widehat{\delta}_{2n},\widehat{\theta}_{xn})-L_{1n}(\widehat{\delta}_{1n})-[\overline{L}_{2n}(\widehat{\delta}_{2n},\widehat{\theta}_{xn}:\widehat{\theta}_{1n})-\overline{L}_{1n}(\widehat{\delta}_{1n}:\widehat{\theta}_{1n})]\right) \\
= & \frac{1}{\sqrt{n}}[L_{2n}(\overline{\delta}_{2n}(\theta_{10}),\theta_{x0})-\overline{L}_{2n}(\overline{\delta}_{2n}(\theta_{10}),\theta_{x0}:\theta_{10})]-\frac{1}{\sqrt{n}}[L_{1n}(\delta_{10})-\overline{L}_{1n}(\delta_{10}:\theta_{10})] \quad \text{(B.3)} \\
& +[C_{1n,1}(\delta_{10},\theta_{x0}:\theta_{10})-C_{2n,1}(\overline{\delta}_{2n}(\theta_{10}),\theta_{x0}:\theta_{10})]\sqrt{n}(\widehat{\theta}_{1n}-\theta_{10})+o_p(1)
\end{aligned}
$$

and under the null of Model 2,

$$
\begin{aligned}
& Cox_2(Y_n,X_n)=\frac{1}{\sqrt{n}}\left(L_{1n}(\widehat{\delta}_{1n})-L_{2n}(\widehat{\delta}_{2n},\widehat{\theta}_{xn})-[\overline{L}_{1n}(\widehat{\delta}_{1n}:\widehat{\theta}_{2n})-\overline{L}_{2n}(\widehat{\delta}_{2n},\widehat{\theta}_{xn}:\widehat{\theta}_{2n})]\right) \\
= & \frac{1}{\sqrt{n}}[L_{1n}(\overline{\delta}_{1n}(\theta_{20}))-\overline{L}_{1n}(\overline{\delta}_{1n}(\theta_{20}):\theta_{20})]-\frac{1}{\sqrt{n}}[L_{2n}(\delta_{20},\theta_{x0})-\overline{L}_{2n}(\delta_{20},\theta_{x0}:\theta_{20})] \quad \text{(B.4)} \\
& +[C_{2n,2}(\delta_{20},\theta_{x0}:\theta_{20})-C_{1n,2}(\overline{\delta}_{1n}(\theta_{20}),\theta_{x0}:\theta_{20})]\sqrt{n}(\widehat{\theta}_{2n}-\theta_{20})+o_p(1).
\end{aligned}
$$

Detailed expressions for terms on the right hand side of these Cox test statistics are from the specific likelihood functions of the two models.[20] Let k_x and k_β denote, respectively, the dimension of θ_x and β_m. Then the dimension of $\delta_m=(\beta_m',\lambda,\sigma)'$ is $k_\beta+2$. From Eq.(B.3), Cox_1 can be rewritten as

$$
Cox_1=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[q_{1i}(\theta_{10})-E_{\theta_{10}}(q_{1i}(\theta_{10}))]+o_p(1), \tag{B.5}
$$

with

$$
\begin{aligned}
q_{1i}(\theta_{10}) &= I(y_i=0)\ln\frac{\Phi[-\gamma_{2i}(\overline{\theta}_{2n}(\theta_{10})|X_n)]}{\Phi[-\gamma_{1i}(\delta_{10}|Y_n,X_n)]} \\
&\quad -\frac{1}{2}I(y_i>0)\left(\ln\frac{\overline{\sigma}^2_{2n}(\sigma_{10})}{\sigma^2_{10}}+(\frac{y_i}{\overline{\sigma}_{2n}(\theta_{10})}-\gamma_{2i}(\overline{\theta}_{2n}(\theta_{10})|X_n))^2-\frac{\varepsilon^2_{1i}}{\sigma^2_{10}}]\right) \\
&\quad +D_{11}w_{i.}Y_n\left[I(y_i>0)\frac{\varepsilon_{1i}}{\sigma^2_{10}}-I(y_i=0)\frac{\phi[-\gamma_{1i}(\delta_{10}|Y_n,X_n)]}{\sigma_{10}\Phi[-\gamma_{1i}(\delta_{10}|Y_n,X_n)]}\right] \\
&\quad +D_{13}\left(\frac{1}{\sigma^3_{10}}I(y_i>0)(\varepsilon^2_{1i}-\sigma^2_{10})+I(y_i=0)\frac{(\lambda_{10}w_{i.}Y_n+x_i'\beta_{10})\phi[-\gamma_{1i}(\delta_{10}|Y_n,X_n)]}{\sigma^2_{10}\Phi[-\gamma_{1i}(\delta_{10}|Y_n,X_n)]}\right) \\
&\quad +D_{12}x_i\left[I(y_i>0)\frac{\varepsilon_{1i}}{\sigma^2_{10}}-I(y_i=0)\frac{\phi[-\gamma_{1i}(\delta_{10}|Y_n,X_n)]}{\sigma_{10}\Phi[-\gamma_{1i}(\delta_{10}|Y_n,X_n)]}\right] \\
&\quad -D_{11}[G(Y_n)W_nG(Y_n)(I_n-\lambda_{10}G(Y_n)W_nG(Y_n))^{-1}]_{ii}+\sum_{l=1}^{\infty}\frac{\lambda^l_{10}}{l!}[(G(Y_n)W_nG(Y_n))^l]_{ii},
\end{aligned}
$$

where the last term is from the Taylor expansion that $\ln|I_n-\lambda_{10}A_n|=-\sum_{i=1}^n\sum_{l=1}^\infty\frac{\lambda^l_{10}}{l!}(A^l_n)_{ii}$, and the row vector $(D_{11},D_{12},D_{13})=\lim_{n\to\infty}(C_{1n,1}-C_{2n,1})\binom{\Sigma^{-1}_{1n,1}}{0_{k_x\times k_\beta}}$, with D_{12} being a $1\times k_\beta$ dimensional row vector and D_{11} and D_{13} being scalars. Hence, with functions of exogenous variables factored out, $q_{1i}(\theta_{10})$ involves terms of endogenous variables

$$
\begin{aligned}
I(y_i &= 0),\ I(y_i=0)\ln\Phi(-\gamma_{1i}(\delta_{10}|Y_n,X_n)),\ I(y_i=0)\frac{\phi[-\gamma_{1i}(\delta_{10}|Y_n,X_n)]}{\Phi[-\gamma_{1i}(\delta_{10}|Y_n,X_n)]}, \\
I(y_i &= 0)\frac{\phi[-\gamma_{1i}(\delta_{10}|Y_n,X_n)]}{\Phi[-\gamma_{1i}(\delta_{10}|Y_n,X_n)]}w_{i.}Y_n,\ I(y_i>0)y_i^2,\ I(y_i>0)y_i,\ I(y_i>0)(w_{i.}Y_n)^2, \\
I(y_i &> 0)w_{i.}Y_n,\ [G(Y_n)W_nG(Y_n)(I_n-\lambda_{10}G(Y_n)W_nG(Y_n))^{-1}]_{ii}, \\
I(y_i &> 0)y_iw_{i.}Y_n,\sum_{l=1}^{\infty}\frac{\lambda^l_{10}}{l!}[(G(Y_n)W_nG(Y_n))^l]_{ii}.
\end{aligned}
$$

The NED property of these terms has been established in Xu and Lee (2015).

The Cox test statistics involve the scores evaluated at the true parameter value. Suppose we can express the fist step estimator $\widehat{\theta}_{xn}$ as $\sqrt{n}(\widehat{\theta}_{xn}-\theta_{x0})=[\sum_{i=1}^n(q_{xi}(\theta_{x0})-E_{\theta_{x0}}[q_{xi}(\theta_{x0})])]/sqrtn$, where $q_{xi}(\theta_{x0})$ satisfies the NED property. Since $\gamma_{2i}(\theta_2)$ and $\frac{\partial\gamma_{2i}(\theta_2)}{\partial\theta_2}$ are non-stochastic, we can rewrite

$$
Cox_2=\frac{1}{\sqrt{n}}\sum_{i=1}^n[q_{2i}(\theta_{20})-E_{\theta_{20}}(q_{2i}(\theta_{20}))]+o_p(1), \tag{B.6}
$$

with

$$
\begin{aligned}
&q_{2i}(\theta_{20})\\
=&I(y_{i,n}=0)\ln\frac{\Phi[-\gamma_{1i}(\overline{\delta}_{1n}(\theta_{20})|Y_n,X_n)]}{\Phi[-\gamma_{2i,n}(\theta_{20}|X_n)]}-\sum_{l=1}^{\infty}\frac{\overline{\lambda}_{1n}^l(\theta_{20})}{l!}[(G(Y_n)W_nG(Y_n))^l]_{ii}\\
&-\frac{1}{2}I(y_i>0)\left(\ln\frac{\overline{\sigma}_{1n}^2(\theta_{20})}{\sigma_{20}^2}+[\frac{y_i}{\overline{\sigma}_{1n}(\theta_{20})}-\gamma_{1i}(\overline{\delta}_{1n}(\theta_{20})|Y_n,X_n)]^2-\frac{\varepsilon_{2i}^2}{\sigma_{20}^2}\right)\\
&+D_{21}\frac{\partial\gamma_{2i}(\theta_{20}|X_n)}{\partial\lambda_2}\left(I(y_{i,n}>0)\frac{\varepsilon_{2i}}{\sigma_{20}}-I(y_i=0)\frac{\phi[-\gamma_{2i}(\theta_{20}|X_n)]}{\Phi[-\gamma_{2i}(\theta_{20}|X_n)]}\right)\\
&+D_{23}\left(I(y_{i,n}>0)[\frac{\varepsilon_{2i}}{\sigma_{20}}(\frac{y_i}{\sigma_{20}^2}+\frac{\partial\gamma_{2i}(\theta_{20}|X_n)}{\partial\sigma_2})-\frac{1}{\sigma_{20}}]-I(y_i=0)\frac{\phi[-\gamma_{2i}(\theta_{20}|X_n)]}{\Phi[-\gamma_{2i}(\theta_{20}|X_n)]}\frac{\partial\gamma_{2i}(\theta_{20}|X_n)}{\partial\sigma_2}\right)\\
&+D_{22}\frac{\partial\gamma_{2i}(\theta_{20}|X_n)}{\partial\beta_2}\left(I(y_i>0)\frac{\varepsilon_{2i}}{\sigma_{20}}-I(y_i=0)\frac{\phi[-\gamma_{2i}(\theta_{20}|X_n)]}{\Phi[-\gamma_{2i}(\theta_{20}|X_n)]}\right)\\
&+\lim_{n\to\infty}(C_{2n,2}-C_{1n,2})\begin{pmatrix}\Sigma_{2n,2}^{-1}R_{21}q_{xi}(\theta_{x0})\\ q_{xi}(\theta_{x0})\end{pmatrix},
\end{aligned}
$$

where the row vector $(D_{21},D_{22},D_{23})=\lim_{n\to\infty}(C_{2n,2}-C_{1n,2})\binom{\Sigma_{2n,2}^{-1}}{0_{k_x\times k_\beta}}$ with D_{22} being a $1\times k_\beta$ dimensional row vector and D_{21} and D_{23} being scalars, and $\Sigma_{2n,2}=-\frac{\partial^2 L_{2n}(\delta_{20},\theta_{x0})}{n\partial\delta_2\partial\delta_2'}$. The terms involved endogenous variables in $q_{2i}(\theta_{20})$ are

$$
\begin{aligned}
I(y_i&=0),\ I(y_i=0)\ln\Phi(-\gamma_{1i}(\overline{\delta}_{1n}(\theta_{20})|Y_n,X_n)),\ I(y_i>0)y_i^2,\ I(y_i>0)y_i,\\
I(y_i&>0)(w_{i\cdot}Y_n)^2,\ I(y_i>0)w_{i\cdot}Y_n,\ I(y_i>0)y_iw_{i\cdot}Y_n,\ \sum_{l=1}^{\infty}\frac{\overline{\lambda}_{1n}^l(\theta_{20})}{l!}(\widetilde{W}_n^l)_{ii},\varepsilon_i^x
\end{aligned}
$$

Here, conditional on X_n, y_i is independent and $\sup_{i,n}||y_i||_p<\infty$, so we can consider $\{y_i\}$ as a special case of NED processes. Thus similar to those considered in Xu and Lee (2015), all terms involved in q_{2i} have the spatial NED property that ensures the CLT holds as those in q_{1i}. Hence, the Cox statistics are asymptotically normally distributed.

C Cox Tests: Bootstrapping

C.1 Verification of the Mean Approximation

First, we show the consistency of $\frac{1}{\sqrt{n}}\overline{L}_{mn}(\theta_m:\theta_{ln})$ by the simulated $\frac{1}{\sqrt{n}}\frac{1}{S}\sum_{s_b=1}^{S}L_{mn}^{(s_b)}(\theta_m|Y_{ln}^{(s_b)},X_n)$ with a large number S of random draws, where $Y_{ln}^{(s_b)}$ is generated from θ_{ln} and X_n. Let $S=h(n)$, where $h(\cdot)$ is an increasing function of the sample size. From Chebyshev's inequality,

$$P\left(\left|\frac{1}{S\sqrt{n}}\sum_{s_b=1}^{S}L_{mn}^{(s_b)}(\theta_m|Y_{ln}^{(s_b)},X_n)-E_{\theta_{ln}}[\frac{1}{S\sqrt{n}}\sum_{s_b=1}^{S}L_{mn}^{(s_b)}(\theta_m|Y_{ln}^{(s_b)},X_n)]\right|\geq\varepsilon\right)$$
$$\leq\ \frac{1}{\varepsilon^2}V_{\theta_{ln}}\left(\frac{1}{S\sqrt{n}}\sum_{s_b=1}^{S}L_{mn}^{(s_b)}(\theta_m|Y_{ln}^{(s_b)},X_n)\right).$$

As $V_{\theta_{ln}}\left(\frac{1}{S\sqrt{n}}\sum_{s_b=1}^{S}L_{mn}^{(s_b)}(\theta_m|Y_{ln}^{(s_b)},X_n))\right)=\frac{1}{S^2}\sum_{s_b=1}^{S}V_{\theta_{ln}}\left(\frac{1}{\sqrt{n}}L_{mn}^{(s_b)}(\theta_m|Y_{ln}^{(s_b)},X_n))\right)=O(\frac{1}{S})$, and $E_{\theta_{ln}}[L_{mn}^{(s_b)}(\theta_m|Y_{ln}^{(s_b)},X_n)]=\overline{L}_{mn}(\theta_m:\theta_{ln})$, we have

$$\frac{1}{S\sqrt{n}}\sum_{s_b=1}^{S}L_{mn}^{(s_b)}(\theta_m|Y_{ln}^{(s_b)},X_n)-\frac{1}{\sqrt{n}}\overline{L}_{mn}(\theta_m:\theta_{ln})=O_p(\frac{1}{\sqrt{S}}).\tag{C.1}$$

C.2 Validity of Bootstrapping

Under Model 2 with true DGP of θ_{20}, $Cox_2(\widehat{\theta}_n|Y_n,X_n)$ converges in distribution to a $N\left(0,\lim_{n\to\infty}\frac{1}{n}V_{\theta_{20}}(\sum_{i=1}^n q_{2i}(\theta_{20}|Y_n,X_n))\right)$. Hence, the standardized Cox test is $Cox_2^*(\widehat{\theta}_n|Y_n,X_n)=Cox_2(\widehat{\theta}_n|Y_n,X_n)/\sqrt{\frac{1}{n}V_{\theta_{20}}(\sum_{i=1}^n q_{2i}(\theta_{20}|Y_n,X_n))}$, which converges in distribution to a $N(0,1)$. Under Model 2 with true DGP of $\widehat{\theta}_{2n}$, the bootstrap version has

$$\begin{aligned}
&Cox_2(\widehat{\theta}_n^{(b)}|Y_{2n}^{(b)},X_n)\\
=\ &\frac{1}{\sqrt{n}}\left(L_{1n}(\widehat{\delta}_{1n}^{(b)}|Y_{2n}^{(b)},X_n)-L_{2n}(\widehat{\theta}_{2n}^{(b)}|Y_{2n}^{(b)},X_n)-E_{\widehat{\theta}_{2n}}\left[L_{1n}(\widehat{\delta}_{1n}^{(b)}|Y_{2n}^{(b)},X_n)-L_{2n}(\widehat{\theta}_{2n}^{(b)}|Y_{2n}^{(b)},X_n)\right]\right)\\
&+\frac{1}{\sqrt{n}}\left(E_{\widehat{\theta}_{2n}}(L_{1n}(\widehat{\delta}_{1n}^{(b)}|Y_{2n}^{(b)},X_n)-L_{2n}(\widehat{\theta}_{2n}^{(b)}|Y_{2n}^{(b)},X_n))\right)\\
&-\frac{1}{\sqrt{n}}\frac{1}{S}\sum_{s_b=1}^{S}[L_{1n}^{(s_b)}(\widehat{\delta}_{1n}^{(b)}|Y_{2n}^{(s_b)},X_n)-\overline{L}_{2n}(\widehat{\theta}_{2n}^{(b)}:\widehat{\theta}_{2n}^{(b)})]\\
&\xrightarrow{d}N\left(0,\lim_{n\to\infty}\frac{1}{n}V_{\widehat{\theta}_{2n}}(\sum_{i=1}^{n}q_{2i}(\widehat{\theta}_{2n}|Y_{2n}^{(b)},X_n))\right).
\end{aligned}$$

We claim that $\frac{1}{n}V_{\theta_{20}}(\sum_{i=1}^n q_{2i}(\theta_{20}|Y_n,X_n))-\frac{1}{n}V_{\widehat{\theta}_{2n}}(\sum_{i=1}^n q_{2i}(\widehat{\theta}_{2n}|Y_{2n}^{(b)},X_n))=o_p(1)$, because $\frac{1}{n}V_{\theta_{20}}(\sum_{i=1}^n q_{2i}(\theta_{20}|Y_n,X_n))-\frac{1}{n}V_{\widehat{\theta}_{2n}}(\sum_{i=1}^n q_{2i}(\theta_{20}|Y_{2n}^{(b)},X_n))=o_p(1)$, due to the equicontinuity of $V_{\theta_2}(\frac{1}{\sqrt{n}}\sum_{i=1}^n q_{2i}(\theta_{20}|Y_n,X_n))$ on θ_2 and

$$\frac{1}{n}V_{\widehat{\theta}_{2n}}(\sum_{i=1}^{n}q_{2i}(\theta_{20}|Y_{2n}^{(b)},X_n))-\frac{1}{n}V_{\widehat{\theta}_{2n}}(\sum_{i=1}^{n}q_{2i}(\widehat{\theta}_{2n}|Y_{2n}^{(b)},X_n))=o_p(1)$$

owing to the stochastic equicontinuity of $q_{2i}(\theta_2|Y_n,X_n)$ on θ_2.[21]

Hence,

$$
\begin{aligned}
Cox_2^*(\widehat{\theta}_n^{(b)}|Y_{2n}^{(b)}, X_n) &= \frac{Cox_2(\widehat{\theta}_n^{(b)}|Y_{2n}^{(b)}, X_n)}{\sqrt{\frac{1}{n}V_{\theta_{20}}(\sum_{i=1}^n q_{2i}(\theta_{20}|Y_n, X_n))}} \\
&= \frac{Cox_2(\widehat{\theta}_n^{(b)}|Y_{2n}^{(b)}, X_n)}{\sqrt{\frac{1}{n}V_{\widehat{\theta}_{2n}}(\sum_{i=1}^n q_{2i}(\widehat{\theta}_{2n}|Y_{2n}^{(b)}, X_n))}} \cdot \frac{\sqrt{\frac{1}{n}V_{\widehat{\theta}_{2n}}(\sum_{i=1}^n q_{2i}(\widehat{\theta}_{2n}|Y_{2n}^{(b)}, X_n))}}{\sqrt{\frac{1}{n}V_{\theta_{20}}(\sum_{i=1}^n q_{2i}(\theta_{20}|Y_n, X_n))}}.
\end{aligned}
$$

converges in distribution to a standard normal random variable. Namely, both the asymptotic distributions of $Cox_2^*(\widehat{\theta}_n^{(b)}|Y_{2n}^{(b)}, X_n)$ and $Cox_2^*(\widehat{\theta}_n|Y_n, X_n)$ are asymptotically standard normal. For any constant c, $(1/B)\sum_{b=1}^B I(Cox_2^*(\widehat{\theta}_n^{(b)}|Y_{2n}^{(b)}, X_n) > c) - P(Cox_2^*(\widehat{\theta}_n^{(b)}|Y_{2n}^{(b)}, X_n) > c)$ goes to 0 in probability, so we can find theoretical quantiles from the sample quantiles of $Cox_2^*(\widehat{\theta}_n^{(b)}|Y_{2n}^{(b)}, X_n)$. Therefore, as $p^* = (1/B)\sum_{b=1}^B I(Cox_2(\widehat{\theta}_n^{(b)}|Y_{2n}^{(b)}, X_n) > Cox_2(\widehat{\theta}_n|Y_n, X_n))$ is equal to $(1/B)\sum_{b=1}^B I(Cox_2^*(\widehat{\theta}_n^{(b)}|Y_{2n}^{(b)}, X_n) > Cox_2^*(\widehat{\theta}_n|Y_n, X_n))$, we reject the null if p^* is smaller than a given level of significance α.

References

Abrevaya, Jason. and Shen, Shu. 2014. "Estimation of Censored Panel-Data Models with Slope Heterogeneity". *Journal of Applied Econometrics*, 29:523–548.

Amemiya, Takeshi. 1973. "Regression Analysis When the Dependent Variable is Truncated Normal". *Econometrica*, 41:997–1016.

Atkinson, Anthony C. 1970. "A Method for Discriminating between Models". *Journal of the Royal Statistics Society*, B32:323–353.

Ballester, Coralio., Calvó-Armengol, Antoni., and Zenou, Yves. 2006. "Who is Who in Networks. Wanted: the Key Player". *Econometrica*, 74(5):1403–1417.

Bergemann, Dirk and Morris, Stephen. 2016. "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games". *Theoretical Economics*, 11:487–522.

Blume, Lawrence E., Brock, William A., Durlauf, Stephen N., and Jayaraman, Rajshri. 2015. "Linear Social Interaction Models". *Journal of Political Economy*, 123(2):444–496.

Bordigon, Massimo, Cerniglia, Floriana, and Revelli, Federico. 2003. "In Search of Yardstick Competition: A Spatial Analysis of Italian Municipality Property Tax Setting". *Journal of Urban Economics*, pages 199–217.

Borkovsky, Ron N., Ellickson, Paul B., Gordon, Brett R., Aguirregabiria, Victor, Gardete, Pedro M., Grieco, Paul L. E., Gureckis, Todd M., Ho, Teckhua, Mathevet, Laurent, and Sweeting, Andrew. 2015. "Multiplicity of Equilibria and Information Structures in Empirical Games: Challenges and Prospects, Ssession at the 9th Triennial Choice Symposium". *Marketing Letters*, 26(2):115–125.

Boucher, Vincent, Bramoullé, Yann, Djebbari, Habiba, and Fortin, Bernard. 2014. "Do Peers Affect Student Achievement? Evidence from Canada Using Group Size Variation". *Journal of Applied Econometrics*, 29:91–109.

Bramoullé, Yann, Kranton, Rachel E., and D'Amours, Martin. 2014. "Strategic Interaction and Networks". *The American Economic Review*, 104(3):898–930.

Brock, William A. and Durlauf, Stephen. 2001. "Discrete Choice with Social Interactions". *The Review of Economic Studies*, 42:430–445.

Brock, William A. and Durlauf, Stephen N. 2000. "Interactions-based Models". In HECKMAN, K.J. and LEAMER, E.E., editors, *Handbook of Econometrics: Edition 1*, pages 3297–3380. Elsevier.

Brock, William A. and Durlauf, Stephen N. 2007. "Identification of Binary Choice Models with Social Interactions". *Journal of Econometrics*, 140:52–75.

Brueckner, Jan K. 2003. "Strategic Interactions among Governments: An Overview of Empirical Studies". *International Regional Science Review*, 26:175–188.

Cox, David R. 1961. "Tests of Separate Families of Hypotheses". *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability 1*, pages 105–123.

Cox, David R. 1962. "Further Results on Tests of Separate Families of Hypotheses". *Journal of the Royal Statistical Society, Series B 24*, pages 406–424.

Durlauf, Stephen N. 2000. "A Framework for the Study of Individual Behavior and Social Interactions". *working paper 16,Wisconsin Madison, Social Systems.*

Gireco, Paul L. E. 2014. "Discrete Games with Flexible Information Structures: An Application to Local Grocery Stores". *RAND Journal of Economics*, 45(2):303–340.

Harsanyi, John C. 1967a. "Games with Incomplete Information Played by 'Bayesian' Players, Parts i". *Management Science*, 14(3):159–182.

Harsanyi, John C. 1967b. "Games with Incomplete Information Played by 'Bayesian' Players, Parts ii". *Management Science*, 14(3):320–334.

Jenish, Nazgul and Prucha, Ingmar R. 2012. “On Spatial Process and Asymptotic Inference under Near-Epoch Dependence”. *Journal of Econometrics*, 170:178–190.

Jin, Fei and Lee, Lung-Fei. 2013. “Cox-Type Tests for Competing Spatial Autoregressive Models with Spatial Autoregressive Disturbances”. *Regional Science and Urban Economics*, 43:590–616.

Kumar, Anil. 2012. “Nonparametric Estimation of the Impact of Taxes on Female Labor Supply”. *Journal of Applied Econometrics”*, 27:415–439.

Lee, Lung-Fei. 2007. “Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects”. *Journal of Econometrics*, 140: 333–374.

Lee, Lung-Fei and Li, Ji. 2009. “Binary Choice under Social Interactions: An Empirical Study with and without Subjective Data on Expectations”. *Journal of Applied Econometrics*, 24:257–281.

Lee, Lung-Fei, Li, Ji, and Lin, Xu. 2014. “Binary Choice Models with Social Network under Heterogeneous Rational Expectations”. *The Review of Economics and Statistics*, 96:402–417.

Li, Shanjun, Liu, Yanyan, and Deininger, Klaus. 2013. “How Important are Endogenous Peer Effects in Group Lending? Eestimating A Static Game of Incomplete Information”. *Journal of Applied Econometrics*, 28:864–882.

Manski, Charles F. 1993. “Identification of Endogenous Social Effects: The Reflection Problem”. *The Review of Economic Studies*, 60:531–542.

Manski, Charles F. 2000. “Economic Analysis of Social Interactions”. *Journal of Economic Perspectives*, 14:115–136.

Moffitt, Robert A. 2000. “Policy Interventions, Low-Level Equilibria, and Social Interactions”. *Social Dynamics*, 97:45–82.

Osborne, Martin J. and Rubinstein, Ariel. 1994. *"A Course in Game Theory"*. MIT Press, Cambridge, Massachusetts; London, England.

Porto, Edoardo Di and Revelli, Federico. 2013. "Tax-Limited Reaction Functions". *Journal of Applied Econometrics*, 28:823–839.

Qu, Xi and Lee, Lung-Fei. 2013a. "Locally Most Powerful Tests for Spatial Interactions in the Simultaneous SAR Tobit Model". *Regional Science and Urban Economics*, 43: 307–321.

Qu, Xi. and Lee, Lung-Fei. 2013b. "LM Tests for Spatial Correlation in Spatial Models with Limited Dependent Variables". *Regional Science and Urban Economics*, 42:430–445.

Rust, John. 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher". *Econometrica*, 55(5):999–1033.

Shang, Qingyan. and Lee, Lung-Fei. 2011. "Two-Step Estimation of Endogenous and Exogenous Group Effects". *Econometric Reviews*, 30(2):173–207.

Sims, Christopher A., Stock, James H., and Watson, Mark W. 1990. "Inference in Linear Time Series Models with Some Unit Roots". *Econometrica*, 58:113–144.

Tao, Ji and Lee, Lung-Fei. 2014. "A Social Interaction Model with An Extreme Order Statistics". *The Econometrics Journal*, 17(3):197–240.

Tobin, James. 1958. "Estimation of Relationships for Limited Dependent Variables". *Econometrica*, 26:24–36.

Wilson, John D. and Wildasin, David E. 2004. "Capital Tax Competition: Bane or Boon". *Journal of Public Economics*, 88:1065–1091.

Xu, Xingbai and Lee, Lung-Fei. 2015. "Maximum Likelihood Estimation of A Spatial Autoregressive Tobit Model". *Journal of Econometrics*, 188:264–280.

Yang, Chao and Lee, Lung-Fei. 2016a. "Social Interactions under Incomplete Information with Heterogeneous Expectations". *Journal of Econometrics.* forthcoming.

Yang, Chao and Lee, Lung-Fei. 2016b. "Strategic Interactions on Municipal Public Safety Spending with Correlated Private Information". *working paper.*

Table 1: Tobit Models with Small Interaction Intensity

		Complete Information			
		Complete Information		Incomplete Information	
Model Specification		$n = 20$, $G = 10$	$n = 20$, $G = 40$	$n = 20$, $G = 10$	$n = 20$, $G = 40$
β_0	0	0.0184 (0.1843)	0.0087 (0.0810)	0.1574 (0.8175)	0.0642 (0.4074)
β_1	1	0.9952 (0.0817)	1.0002 (0.0403)	0.9909 (0.0847)	1.0035 (0.0414)
β_2	1	1.0011 (0.0541)	1.0007 (0.0273)	1.0551 (0.1421)	1.0514 (0.0754)
λ	0.3	0.2880 (0.0831)	0.2953 (0.0342)	0.1863 (0.4680)	0.2347 (0.2277)
σ	1	0.9848 (0.0581)	0.9949 (0.0305)	1.0341 (0.0690)	1.0549 (0.0354)
$m \log L$		-224.9726 (18.8686)	-907.6152 (40.8076)	-232.3204 (19.9512)	-942.3631 (43.5445)
r_{true}		0.9640	1.0000	-	-
r_{censor}		0.2755	0.2749	0.2755	0.2749

		Incomplete Information			
		Incomplete Information		complete Information	
Model Specification		$n = 20$, $G = 10$	$n = 20$, $G = 40$	$n = 20$, $G = 10$	$n = 20$, $G = 40$
β_0	0	0.1630 (0.6366)	0.0532 (0.2373)	0.5081 (0.2109)	0.4993 (0.0899)
β_1	1	0.9906 (0.0809)	1.0000 (0.0404)	0.9950 (0.0812)	1.0029 (0.0403)
β_2	1	1.0397 (0.1045)	1.0122 (0.0472)	1.0081 (0.0542)	1.0872 (0.0274)
λ	0.3	0.2025 (0.3579)	0.2693 (0.1320)	0.0081 (0.0972)	0.0134 (0.0406)
σ	1	0.9845 (0.0572)	0.9954 (0.0305)	0.9863 (0.0574)	0.9976 (0.0306)
$m \log L$		-226.5578 (17.0586)	-913.4246 (36.9513)	-226.8745 (17.0564)	-914.9692 (36.9957)
r_{true}		0.5900	0.8460	-	-
r_{censor}		0.2682	0.2685	0.2682	0.2685

		Distribution of Self-Known Characteristics	
	True parameters	$n = 20$, $G = 10$	$n = 20$, $G = 40$
μ	1	0.9872 (0.4032)	0.9899 (0.2148)
η	2	1.9336 (0.1919)	1.9804 (0.1005)
ρ	0.4	0.3463 (0.1184)	0.3851 (0.0591)

Note: G is the number of groups. n is the population for each group. The number of repetitions is 500. $m \log L$ is the estimated sample average log likelihood. r_{true} is the proportion of simulations for which estimated log likelihood is bigger than that under wrong information structure. r_{censor} is the censoring rate. The numbers in parentheses are empirical standard deviation.

Table 2: Tobit Models with Intermediate Interaction Intensity

		Complete Information			
		Complete Information		Incomplete Information	
Model Specification		$n = 20$, $G = 10$	$n = 20$, $G = 40$	$n = 20$, $G = 10$	$n = 20$, $G = 40$
β_0	0	0.0124 (0.1593)	0.0059 (0.0678)	0.2432 (1.5755)	-0.0474 (0.9561)
β_1	1	0.9965 (0.0794)	1.0005 (0.0382)	1.0053 (0.1101)	1.0288 (0.0553)
β_2	1	1.0022 (0.0515)	1.0011 (0.0259)	1.2539 (0.2801)	1.2629 (0.1595)
λ	0.6	0.5935 (0.0449)	0.5975 (0.0183)	0.4160 (0.5579)	0.5004 (0.3290)
σ	1	0.9858 (0.0545)	0.9961 (0.0283)	1.4919 (0.2052)	1.5925 (0.1083)
$m \log L$		-242.4600 (19.2647)	-978.8440 (41.7971)	-308.8009 (30.6033)	-1282.4000 (64.9694)
r_{true}		1.0000	1.0000	-	-
r_{censor}		0.1990	0.1981	0.1990	0.1981

		Incomplete Information			
		Incomplete Information		Complete Information	
Model Specification		$n = 20$, $G = 10$	$n = 20$, $G = 40$	$n = 20$, $G = 10$	$n = 20$, $G = 40$
β_0	0	0.0829 (0.4733)	0.0201 (0.1560)	1.4866 (0.2759)	1.4777 (0.1203)
β_1	1	0.9957 (0.0784)	1.0014 (0.0382)	1.0122 (0.0795)	1.0244 (0.0387)
β_2	1	1.0497 (0.0882)	1.0127 (0.0389)	1.2209 (0.0540)	1.2146 (0.0263)
λ	0.6	0.5625 (0.1622)	0.5908 (0.0498)	0.0589 (0.0859)	0.0632 (0.0379)
σ	1	0.9855 (0.0529)	0.9968 (0.0287)	1.0079 (0.0540)	1.0238 (0.0297)
$m \log L$		-250.1296 (14.3362)	-1008.5000 (31.2863)	-254.2479 (14.5850)	-1028.0000 (31.6930)
r_{true}		0.9280	1.0000	-	-
r_{censor}		0.1579	0.1593		

	True parameters	Distribution of Self-Known Characteristics	
		$n = 20$, $G = 10$	$n = 20$, $G = 40$
μ	1	0.9872 (0.4032)	0.9899 (0.2148)
η	2	1.9336 (0.1919)	1.9804 (0.1005)
ρ	0.4	0.3463 (0.1184)	0.3851 (0.0591)

Note: G is the number of groups. n is the population for each group. The number of repetitions is 500. $m \log L$ is the estimated sample average log likelihood. r_{true} is the proportion of simulations for which estimated log likelihood is bigger than that under wrong information structure. r_{censor} is the censoring rate. The numbers in parentheses are empirical standard deviation.

Table 3: Rejection Rates of Cox Tests for Information structures

$\lambda = 0.3$

		DGP: Complete Information			
		H0: Complete Information H1: Incomplete Information		H0: Incomplete Information H1: Complete Information	
Model Specification		$n = 20$, $G = 10$	$n = 20$, $G = 40$	$n = 20$, $G = 10$	$n = 20$, $G = 40$
		Size		Power	
Significance	0.01	0.0120	0.0080	0.6860	1.0000
	0.05	0.0440	0.0440	0.8820	1.0000
	0.1	0.0940	0.0880	0.9320	1.0000

		DGP: Incomplete Information			
		H0: Incomplete Information H1: Complete Information		H0: Complete Information H1: Incomplete Information	
Model Specification		$n = 20$, $G = 10$	$n = 20$, $G = 40$	$n = 20$, $G = 10$	$n = 20$, $G = 40$
		Size		Power	
Significance	0.01	0.0080	0.0080	0.0240	0.1520
	0.05	0.0520	0.0320	0.1320	0.3900
	0.1	0.1080	0.0940	0.2280	0.5340

$\lambda = 0.6$

		DGP: Complete Information			
		H0: Complete Information H1: Incomplete Information		H0: Incomplete Information H1: Complete Information	
Model Specification		$n = 20$, $G = 10$	$n = 20$, $G = 40$	$n = 20$, $G = 10$	$n = 20$, $G = 40$
		Size		Power	
Significance	0.01	0.0060	0.0100	0.7740	0.9980
	0.05	0.0320	0.0460	0.7820	0.9980
	0.1	0.0760	0.0980	0.7820	0.9980

		DGP: Incomplete Information			
		H0: Incomplete Information H1: Complete Information		H0: Complete Information H1: Incomplete Information	
Model Specification		$n = 20$, $G = 10$	$n = 20$, $G = 40$	$n = 20$, $G = 10$	$n = 20$, $G = 40$
		Size		Power	
Significance	0.01	0.0100	0.0060	0.5340	0.9960
	0.05	0.0600	0.0520	0.7540	1.0000
	0.1	0.1240	0.0960	0.8380	1.0000

Note: G is the number of groups. n is the population for each group.

Table 4: Sample Statistics

Variables		Mean	Standard Deviation	Min	Max
Property Tax Rate	Per $100 Valuation	0.3676	(0.1972)	0	0.8200
Population	$\times 10^3$	10.2842	(45.0819)	0.0250	751.9990
Median Household Income	\$ $\times 10^4$	4.2092	(1.8504)	1.1750	15.7297
Distance	Kilometers	216.7976	(125.4035)	1.2133	767.1423
No. of Observations		506			

Table 5: Tobit Model for Property Tax Competition with Complete Information

		Estimation for City Property Tax Rates		
Estimation	(1)	(2)	(3)	(4)
Constant	0.5384***	0.3535***	0.3398***	0.2363***
	(0.0204)	(0.0323)	(0.0401)	(0.0499)
Population	0.0006***	0.0006***	0.0005***	0.0005***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Household Income	-0.0430***	-0.0371***	-0.0385***	-0.0396***
	(0.0041)	(0.0044)	(0.0045)	(0.0044)
λ		0.4428***	0.4921***	0.7701***
		(0.0613)	(0.0858)	(0.1164)
σ	0.1899***	0.1783***	0.1825***	0.1816***
	(0.0068)	(0.0064)	(0.0066)	(0.0067)
log Likelihood	76.8650	100.5834	92.5753	95.5271
AIC	-145.7301	-191.1667	-175.1506	-181.0541
BIC	-128.8239	-170.0340	-154.0179	-159.9214
Predictions				
$\overline{\tau}_p$	0.3672	0.3695	0.3670	0.3503
$\overline{\tau}_c$	-	0.2182	0.2025	0.1218
%Tax Change per 1% Population Increase	0.0151	0.0282	0.0281	0.0593
%Tax Change per 1% Income Increase	-0.4602	-0.6918	-0.7847	-1.6539
No. of Observations	506	506	506	506
No. of "Neighbors"		10.8656	28.3636	94.0474
		(4.3634)	(9.0854)	(28.1823)
Cutoff Distance		30	50	100

Note: Estimation (1) is the ordinary Tobit regression without social interactions. Estimations (2), (3), and (4) correspond to the Tobit model under the framework of complete information. Two municipalities are viewed as close "neighbors" if the distance between them is less than 30 kilometers for Estimations (2), or less than 50 kilometers for Estimations (3), and for less than 100 kilometers for Estimations (4). Numbers in parentheses are theoretical standard deviations. Estimates that are significant at the %10, %5, and %1 levels are marked by "*", "**", and "***", respectively. $\overline{\tau}_p$ is the average predicted tax rate. $\overline{\tau}_c$ is the counterfactual average tax rate when there are no network relations.

Table 6: Tobit Model for Property Tax Competition with Incomplete Information

		Estimation for City Property Tax Rates		
Estimation	(1)	(5)	(6)	(7)
Constant	0.5384***	0.2838***	0.2549***	0.2878
	(0.0204)	(0.0813)	(0.0846)	(0.1900)
Population	0.0006***	0.0006***	0.0006***	0.0006***
	(0.0002)	(0.0001)	(0.0001)	(0.0001)
Household Income	-0.0430***	-0.0453***	-0.0435***	-0.0422***
	(0.0041)	(0.0046)	(0.0046)	(0.0047)
λ		0.9242***	0.9999***	0.8694
		(0.2904)	(0.2957)	(0.6575)
σ	0.1899***	0.1890***	0.1890***	0.1898***
	(0.0068)	(0.0508)	(0.0510)	(0.0513)
log Likelihood	76.8650	79.0218	79.0007	77.4249
AIC	-145.7301	-148.0435	-148.0013	-144.8498
BIC	-128.8239	-126.9109	-128.8686	-123.7172
log Likelihood				
$\overline{\tau}_p$	0.3672	0.3668	0.3670	0.3671
$\overline{\tau}_c$	-	0.1420	0.1275	0.1535
% Tax Change per 1% Population Increase	0.0151	0.0458	0.0587	0.0416
% Tax Change per 1% Income Increase	-0.4602	-0.4950	-0.4768	-0.4618
No. of Observations	506	506	506	506
No. of "Neighbors"		10.8656	28.3636	94.0474
		(4.3634)	(9.0854)	(28.1823
Cutoff Distance		30	50	100

	Distribution of Municipal and State Median Household Income				
Parameters	ϕ_h	a_h^*	b_h^*	ω^{2*}	ι^{2*}
Estimates	0.6453**	2.2858***	0.0794***	0.0213***	3.4202***
	(0.2628)	(0.3019)	(0.0140)	(0.0066)	(0.1100)
Estimated Moments		$E[X_i^p]$	$Var(X_i^p)$	$Corr(X_i^p, X_j^p)$	
		4.5086	3.4568	0.0106	

Note: Estimation (1) is the ordinary Tobit regression without social interactions. Estimations (5), (6), and (7) correspond to the Tobit model under incomplete information where the municipal median household income is assumed to be self-known when decisions about the new tax rates are made. Two municipalities are viewed as close "neighbors" if the distance between them is less than 30 kilometers for Estimations (5), or less than 50 kilometers for Estimations (6), and for less than 100 kilometers for Estimations (7). Estimates of the (transformed) parameters of the distribution of municipal median household income are present in the lower panel. Numbers in parentheses are theoretical standard deviations. Estimates that are significant at the %10, %5, and %1 levels are marked by "*", "**", and "***", respectively. $\overline{\tau}_p$ is the average predicted tax rate. $\overline{\tau}_c$ is the counterfactual average tax rate when there are no network relations.

Table 7: Testing Information Structures: Property Tax Competition

	Cutoff: 30 kilometers					
	H0: Complete Information H1: Incomplete Information			H0: Incomplete Information H1: Complete Information		
Significance	Sample Statistics	Bootstrapping Cutoff	Rejection	Sample Statistics	Bootstrapping Cutoff	Rejection
1%	0.0513	0.4206	0	1.7579	1.1376	1
5%	0.0513	0.2726	0	1.7579	0.2620	1
10%	0.0513	0.2009	0	1.7579	0.1982	1

	Cutoff: 50 kilometers					
	H0: Complete Information H1: Incomplete Information			H0: Incomplete Information H1: Complete Information		
Significance	Sample Statistics	Bootstrapping Cutoff	Rejection	Sample Statistics	Bootstrapping Cutoff	Rejection
1%	0.0195	0.5661	0	0.9886	0.9021	1
5%	0.0195	0.2350	0	0.9886	0.4465	1
10%	0.0195	0.1984	0	0.9886	0.3260	1

	Cutoff: 100 kilometers					
	H0: Complete Information H1: Incomplete Information			H0: Incomplete Information H1: Complete Information		
Significance	Sample Statistics	Bootstrapping Cutoff	Rejection	Sample Statistics	Bootstrapping Cutoff	Rejection
1%	0.1859	0.9104	0	1.1086	0.4147	1
5%	0.1859	0.4703	0	1.1086	0.1574	1
10%	0.1859	0.3044	0	1.1086	0.1137	1

Note: "Rejection" will be 1 if the result of the test is that the null can be rejected at a certain significance level. The number of bootstrapping simulations is $B = 100$ and the number of simulations for approximating the expected likelihood is $S = 100$.