Semiparametric Dynamic Logit Model with Endogenous Networks

Brice Romuald Gueyap Kounga*

This version: October 21, 2026

Click here for the latest version

Abstract

This paper studies the identification and estimation of a semiparametric dynamic logit model with nonparametric link formation, in which a time-varying index of unobserved characteristics affects outcomes and links. The outcome depends on covariates, the past outcome, a fixed effect, and an unknown function of that index. Conditioning on a switch by the same agent across waves removes the fixed effect but not the change in the unknown function. When this change is common within network types and followed by a stable period, comparing two agents of the same network type, meaning they have the same probability of linking with each of the other agents, removes it and identifies covariate effects and state dependence. The pairwise conditional likelihood estimator is consistent, including when the index is continuous, and asymptotically normal when the index takes finitely many values and matches are exact. In simulations, bias falls with sample size, and coverage is near nominal among successful fits. The fixed-effects dynamic logit and pooled logits with peer-average controls remain biased when the index moves. A second estimator, conditioning on recovered network classes, is more precise. A feasibility exercise on the Glasgow Teenage Friends and Lifestyle Study examines how past use and pocket money affect adolescents' recorded tobacco or cannabis use and documents the data requirements.

Keywords: Dynamic logit, semiparametric estimation, social networks, network formation, conditional maximum likelihood, state dependence, unobserved heterogeneity.

JEL Classification: C14, C23, C25, D85.

*Department of Economics, Western University. Email: bgueyapk@uwo.ca. Special thanks to Nail Kashaev, Salvador Navarro, Roy Allen, and Tim Conley for their comments and suggestions in the preparation of this paper. The views expressed herein and any remaining errors are mine.

1 Introduction

Teenagers who smoke this year are likely to smoke next year, and their friends often smoke too (Fletcher, 2010; Nakajima, 2007). Farmers deciding whether to adopt a new crop technology rely on their own past experience, and farmers who exchange information with each other often make the same decision (Conley and Udry, 2010). Two mechanisms can produce such patterns. The first is state dependence: a past choice directly raises the probability of the current one, as when smoking becomes a habit. The second is unobserved characteristics (e.g., a taste for risk, impulsivity, patience, or ability) that affect both the choice and the way people form ties, so that choices persist and cluster among friends or neighbors without any habit. To study such choices, the econometrician should therefore use a model that allows for state dependence and for an endogenous network. The network is endogenous when people form their ties partly because of the same unobserved characteristics that drive their choices.

This paper studies such a model: a dynamic logit with an endogenous network, in which the unobserved characteristics that drive both the choice and the ties may change over time. Its main result is that matching students on their network type, the probabilities with which they befriend each other student, identifies the effects of past choices and of observed covariates when the change in the unobserved term between two waves is common to students of the same network type and is followed by a stable wave. The levels of the unobserved term and the fixed effects may differ across matched students: conditioning on each student's own switch removes the levels, and the comparison across matched students removes the common change. I propose two estimators and examine the method in simulations and in data on adolescent smoking.

The economic question is which mechanism drives persistence, and the distinction matters for policy. If persistence comes from state dependence, an intervention that stops a student from smoking for one year also lowers her probability of smoking in later years. If persistence comes from stable characteristics, the student returns to the same probability of smoking once the intervention ends (Heckman, 1981a). Separating the two mechanisms is hard, because the characteristics that drive smoking, friendships, and persistence are not observed. A risk-tolerant student may be more likely to smoke, to keep smoking, and to befriend other risk-takers, and the researcher has no data on risk tolerance. Friendships and smoking are then related for reasons that have nothing to do with habit or with observed covariates such as pocket money.

Existing methods do not resolve this problem, for different reasons. A pooled logit omits the heterogeneity, which is correlated with past smoking and with the covariates, so its estimates are inconsistent in general. Adding friends' characteristics as controls does not

repair this, because the controls capture only the part of the heterogeneity reflected in friends' observables. Methods that condition on fixed effects, such as the dynamic logit of Honoré and Kyriazidou (2000), or that otherwise do not use the network (Ouyang and Yang, 2025), remove the heterogeneity only if it does not change over time. During adolescence, attitudes to risk and to social norms do change. Methods for endogenous networks specify a model of link formation (Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2016; Arduini et al., 2015) or rely on control functions or exclusion restrictions (Johnsson and Moon, 2021; Kashaev and Lazzati, 2025). Each rests on restrictions on the formation process that differ from the ones used here. This paper restricts the arguments of the link rule but leaves its functional form unknown rather than assuming a parametric form for it.

This paper provides identification and estimation results for a dynamic logit model with a fixed effect and a time-varying unobserved characteristic that also drives link formation, without a parametric model of the links. It combines the path conditioning of the dynamic logit with the network-type matching of Auerbach (2022) and Gueyap Kounga (2023). Conditioning on the order in which a student switches between two waves, with the covariates held stable over the following wave, removes any characteristic that does not change over the window. A student's network type is the list of her probabilities of befriending each other student, given her past smoking. Comparing two students of the same network type and with the same past smoking then removes the change in the unobserved term, provided the change is common to students of the same type. Auerbach (2022) and Gueyap Kounga (2023) use a single cross-section and therefore require equal levels of the unobserved term within a network type. This paper requires equal changes, which allows a fixed effect that matching is not assumed to equalize. The price is a set of further restrictions: stability of the unobserved term over the wave after the transition, strict exogeneity of the covariates and the unobserved characteristic, a dyadic link rule on a scalar index, and enough support among the matched comparisons. In the smoking example, the common change and the stability that follows it describe a cohort observed around a scheduled school transition: students of the same network type, who befriend the same kinds of peers, meet the same new environment, so their unobserved terms move by the same amount over the transition, whatever their levels and their fixed effects, and stay put over the following wave. Section 4 and Appendix D make these conditions precise. Under these restrictions the covariate and state dependence coefficients are identified.

The network is useful because a student's friendships carry indirect information about the characteristic that drives them. The observed network does not reveal a student's own value of the characteristic, but it does reveal whether two students have the same network type. The codegree of two students is the number of friends they have in common. The

codegree distance between two students aggregates, over all third students, the differences between their codegrees with each third student. As the network grows, it vanishes exactly when the two students share a network type, under the support and separation conditions of Proposition 1.

The outcome is whether student i smokes at wave t. The model is a dynamic logit in which the outcome depends on observed covariates X_{it}, such as pocket money; on the student's own smoking at the previous wave, y_{it-1}; on a fixed effect μ_i; and on an unknown function λ of an index of unobserved characteristics w_{it} that changes over time. The model applies to any agents, and the example refers to students. The parameters of interest are β, the coefficients on the covariates, and α, the coefficient on past smoking, which measures state dependence on the log-odds scale. Habit is one interpretation of a positive α, not an identified mechanism. Two unobserved terms enter: the fixed effect μ_i is the part of the unobserved characteristics that does not change over time and does not enter the friendship rule, such as a physiological susceptibility to nicotine, and the function λ is an unknown function of the student's index of unobserved social characteristics $w_{it} \in [0, 1]$, such as risk attitude, attachment to norms, and impulsivity, which change during adolescence. Following Auerbach (2022), I call λ the social influence function. The name refers to the student's own pull toward smoking. The function is not an effect of friends' smoking, since the outcome equation contains neither friends' outcomes nor the realized network, and the model identifies no peer effect. The same w_{it} governs whom the student befriends, which is why the network is informative about w_{it} and why the network is endogenous. With $\lambda \equiv 0$, the model is the fixed-effects dynamic logit of Honoré and Kyriazidou (2000). Reading β and α as causal effects requires the assumptions of Section 4. The function λ, the fixed effects, and the link rule are nuisance components, which the method removes without estimating, and the level of λ is not recovered.

The identification argument uses at least four dates: the initial outcome and three waves. Take two students with the same initial outcome and the same network type at wave 1, who switch in opposite directions between waves 1 and 2 and whose covariates and social influence term do not change over wave 3. Conditioning on the switch cancels each student's fixed effect and level of λ but not the change in λ between waves 1 and 2. Under the identifying assumption, the change is the same for both students and cancels when they are compared. The probability that one rather than the other started smoking then depends only on (β, α) and on observed variables. Since the codegree distance identifies which students share a network type (Proposition 1), β and α are identified under overlap and rank conditions (Theorem 1), together with the common change in λ within each network class (Corollary 1).

The identifying assumption, Assumption 2, has two parts. The change in the social influence function between waves 1 and 2 is the same for all students of the same network

type at wave 1 with the same initial outcome, and the function does not change between waves 2 and 3. The assumption allows a shift common to every student and excludes changes that differ across students of the same network type, such as a characteristic with its own idiosyncratic innovations. It cannot be tested from the data alone, because the social influence function is never observed. The maintained assumptions (Assumption 1) are logistic errors, strict exogeneity of the covariates and of the unobserved characteristic, and a link rule on the two students' unobserved characteristics and past outcomes, with independent link shocks and no observed covariates.

The estimator is a conditional maximum likelihood estimator over pairs of switching students. Each pair is weighted by its estimated codegree distance and by the stability of each student's covariates over the following wave. The weights are kernels, and their bandwidths set how similar two students must be to count as a match. The estimator is consistent, including with continuous network types and covariates, with bandwidths that shrink at the rates of Assumption 4 (Theorem 2). Suppose that the characteristic takes finitely many values and that covariate changes have an atom at zero. Exact matches then occur with positive probability, and the estimated weights select them with probability approaching one. The estimator is root-n consistent and asymptotically normal, with a variance clustered by student (Theorem 3). With three waves, the same holds for a bandwidth computed from the first-wave network (Corollary 2). With continuous types, the paper establishes consistency only, for the reasons given in Section 6.6. When the characteristic takes finitely many values, switchers whose estimated distances fall below the bandwidth, with the same initial outcome, form classes whose members share the change in λ. Conditioning on the number of switches in each direction within a class gives a second, class-conditional estimator. This estimator is root-n consistent and asymptotically normal with one transition period (Proposition 2) and more precise in the simulations. In practice, a gap in the estimated distances and classes that do not change with the bandwidth are what the class estimator needs, although neither certifies that the classes are correct. The pairwise estimator is the general one, the only one the theory covers when types are continuous, and the safer choice when the classes are uncertain.

Simulations in designs that satisfy the assumptions agree with the theory. The bias of the pairwise estimator falls with the sample size and its confidence intervals have close to nominal coverage among successful fits. The class-conditional estimator is more precise than the pairwise estimator. Pooled logits, with or without peer-average controls, and the within-agent estimator of Honoré and Kyriazidou (2000) remain biased when the characteristic moves, with coverage that falls as the sample grows.

I apply the estimators to the Glasgow Teenage Friends and Lifestyle Study, a panel

of 96 students observed over three waves, to study how past smoking and pocket money affect smoking, with the outcome defined as any tobacco or cannabis use. The fits treat friendship links as symmetric, set the unobserved initial outcome to zero, drop the covariate-stability condition, which no switching student satisfies, and compare all pairs of switchers rather than network-matched pairs, because at every localized bandwidth the pairwise and class-conditional likelihoods are separated. The conditional fits are stabilized by the Firth penalization. I find that the point estimate of state dependence is positive in each of the five unlocalized fits, falling from 2.22 in a pooled logit and 2.10 with peer-average controls to 1.83 when each student is compared with herself over time, 1.41 when students are paired with each other, and 0.89 when all switchers are compared within a single class. The pocket money coefficient, per pound of monthly pocket money, is imprecisely estimated in every specification, and alcohol use is the covariate most strongly associated with smoking. These fits are descriptive: they implement the conditioning but not the network matching, their ordering is sensitive to the penalization, and the baseline unpenalized maxima exist while several robustness specifications are separated. A dataset suited to the method needs an observed initial outcome, enough stable switchers within recoverable network classes, and a sufficiently informative measurement of friendships.

Section 2 reviews the literature and Section 3 presents the model. Sections 4 and 5 give the identification results and the two estimators, and Section 6 establishes consistency and the finite-type distribution theory. Section 7 reports the simulations, Section 8 examines feasibility in the Glasgow data, and Section 9 concludes. Appendix A contains the proofs, Appendix B explains what is identified about the social influence function, Appendix C notes that the binary estimator applies to an ordered outcome dichotomized at one known cutoff, and Appendices D–F describe the observation window, report the descriptive fits for the Glasgow data, and collect supplementary simulation tables.

2 Literature Review

This paper lies at the intersection of three strands of the econometrics literature: dynamic logit models, semiparametric panel data methods, and the econometrics of endogenous social networks. Each literature provides tools for part of the problem considered here: conditional likelihood removes time-invariant heterogeneity in dynamic logit models, semiparametric panel methods handle unknown nuisance functions for continuous outcomes, and the network-based methods reviewed below use linking behavior to control for latent characteristics in a single cross-section.

An important starting point for panel logit methods is Chamberlain (1980), who showed

that the logistic specification admits sufficient statistics for individual fixed effects, allowing consistent estimation without specifying the distribution of unobserved heterogeneity. Honoré and Kyriazidou (2000) extended this approach to dynamic settings by exploiting conditional likelihood arguments based on pairs of observations with identical sufficient statistics, requiring covariates to remain locally constant across adjacent periods. More recent contributions have expanded identification and estimation methods for dynamic logit models. Honoré and Weidner (2025) derive moment conditions that eliminate nuisance parameters without relying on conditional likelihood, while Honoré et al. (2025) and Muris et al. (2025) extend these ideas to ordered outcomes. Shin (2025) derives sufficient statistics for a first-order Markov feedback process in a dynamic panel logit and shows that conditional likelihood identification then requires restrictions on that process and on the initial condition. The present paper maintains strict exogeneity of the covariates and of the latent characteristic. Ouyang and Yang (2025) develop a two-step maximum score approach for dynamic binary choice that does not rely on the logistic distribution. It matches observations on an estimated index and relies on additional restrictions on serial dependence and on the tails of the error distribution. Dano et al. (2025) study logit models with general fixed-effect structures for panel and network data, including individual effects, time trends, and two-way and dyadic effects. Outside the logistic family, Honoré and Tamer (2006) show that the dynamic parameter is generally only partially identified, and Chamberlain (2010) characterizes the special informational role of the logistic error. These results delimit what any approach in this class, including the present one, can deliver without the logistic assumption. Much of this literature treats unobserved heterogeneity as a time-invariant individual effect. The richer structures of Dano et al. (2025) allow time variation, but through fixed effects whose structure is specified in terms of the observed indices of units and periods, such as an individual effect combined with a time trend or with time effects. The heterogeneity considered here keeps the individual fixed effect of that literature and adds an unknown function of a latent characteristic whose movement is not restricted to a parametric trend: at the transitions used for identification, its increment is shared within network classes, and the latent term is stable over the following wave. The added term can be removed only because agents with the same network type share that increment, so that the network supplies the information needed to compare them. The present paper addresses this case by combining conditional likelihood arguments, which remove the fixed effect, with network-type matching, which removes the movement, and local temporal smoothing, which selects agents whose covariates are stable after the transition. Grouped fixed-effects methods (Bonhomme and Manresa, 2015; Bonhomme et al., 2022) also accommodate heterogeneity that varies over time, by classifying units into groups and estimating group-specific effects. The class-conditional estimator of Section 5.3 also works

with groups, but its groups are network classes recovered from the codegree distance, and it removes their effects by conditioning on each group's number of switches, as conditional logistic regression does for matched sets (Breslow and Day, 1980), rather than estimating them.

A second related literature studies semiparametric estimation in panel data models with unknown nuisance functions. The partially linear framework of Robinson (1988) established that finite-dimensional parameters can be estimated consistently without specifying the functional form of a nonparametric component. Extensions to panel data include the kernel-based estimators of Henderson et al. (2008) and the marginal integration approach of Qian and Wang (2012), both of which accommodate unknown functions while controlling for individual heterogeneity. These methods, however, are developed for continuous outcomes with linear or partially linear conditional mean models. They do not address nonlinear dynamic logit models in which the nuisance function depends on an unobserved characteristic that is itself recovered only indirectly through network information.

The identification of peer effects and endogenous social interactions has been central since Manski (1993), who showed that endogenous interactions, contextual effects, and correlated unobservables are generically not separately identifiable without additional restrictions. A large empirical literature, surveyed by Bramoullé et al. (2020) and Jackson et al. (2017), documents the importance of social interactions in education (Sacerdote, 2001; Calvó-Armengol et al., 2009), consumption (De Giorgi et al., 2020), labor markets (Beugnot et al., 2013), and health behaviors (Christakis and Fowler, 2008). Existing approaches to endogenous network formation generally rely on one of three strategies: a parametric or semiparametric model for link formation (Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2016; Arduini et al., 2015), identification from equilibrium restrictions in strategic network formation models (Badev, 2021), or control-function or exclusion-restriction arguments under alternative identifying assumptions (Johnsson and Moon, 2021; Kashaev and Lazzati, 2025). These approaches differ substantially, but all require assumptions on the network formation process that may be difficult to justify empirically. The network-type approach instead combines a nonparametric link function with restrictions connecting network profiles to outcome heterogeneity and uses the similarity of observed linking behavior to recover equivalence classes of latent characteristics.

A related line of work exploits network data to control for unobserved heterogeneity through the structure of the graph itself: Graham (2017) models degree heterogeneity in network formation with agent-specific effects, and Zeleneev (2020) develops matching on unobserved network types under nonparametric heterogeneity, in a spirit close to the codegree approach used here but for outcomes defined on the network rather than for a dynamic panel

of individual choices. Gao and Niu (2026) study identification in dynamic dyadic network formation models with dyad fixed effects, where the object of interest is link formation itself. Starck (2025) and Xu (2025) use network data to control for unobservables in a cross-section, the former by forming comparison groups of similar agents to estimate treatment effects, the latter through a network control function for nonparametric partial effects that generalizes Auerbach (2022).

The papers most closely related to the present work are Auerbach (2022) and Gueyap Kounga (2023). Auerbach (2022) introduced the idea of matching agents according to their network formation behavior in partially linear regression models, showing that network similarity can control for latent heterogeneity under an equivalence restriction, without parametrically specifying the link function. His framework is static and applies to continuous outcomes with partially linear conditional means. Gueyap Kounga (2023) extended this idea to semiparametric logit models in a static cross-sectional setting, establishing point identification through codegree-based matching and proposing a kernel-weighted maximum likelihood estimator. Both papers require agents with the same network type to have the same unobserved term in the outcome equation. The present paper extends this line of research to dynamic panel logit models and relaxes that requirement to equal changes of the unobserved term, which allows an individual fixed effect that is not a separate argument of the link function and that network matching is not assumed to equalize. The dynamic setting introduces identification challenges that do not arise in static environments. State dependence interacts with time-varying latent heterogeneity, requiring conditional likelihood arguments together with local temporal smoothing, while the evolution of the network requires matching to hold across time rather than at a single cross-section. The resulting framework combines dynamic panel logit methods with network-type matching to identify and estimate a semiparametric logit model under endogenous network formation without imposing a parametric specification on the link formation process.

3 Model

Consider a panel of n agents observed over $T+1$ periods $t = 0, 1, \ldots, T$, where $T \geq 3$. Each agent $i \in \{1, \ldots, n\}$ is characterized by an observed vector of time-varying covariates $X_{it} \in \mathbb{R}^k$, a binary outcome $y_{it} \in \{0, 1\}$, an index of unobserved social characteristics $w_{it} \in [0, 1]$, and an unobserved agent-specific term $\mu_i \in \mathbb{R}$ that is constant over time. The econometrician observes the outcomes $(y_{i0}, y_{i1}, \ldots, y_{iT})$, including the initial outcome y_{i0}, the covariates $(X_{i1}, \ldots, X_{iT})$, and the network D_t defined below at each period $t \geq 1$. Neither the index w_{it} nor the term μ_i is ever observed.

The binary outcome is generated by the dynamic semiparametric logit model:

$$y_{it} = \mathbb{1}\left\{X_{it}'\beta + \alpha y_{it-1} + \lambda(w_{it}) + \mu_i - \varepsilon_{it} \geq 0\right\}, \quad t = 1, \ldots, T, \tag{1}$$

where $\beta \in \mathbb{R}^k$ is a vector of slope parameters, $\alpha \in \mathbb{R}$ captures state dependence, $\lambda : [0,1] \to \mathbb{R}$ is an unknown measurable function, which I call the social influence function, μ_i is an agent effect, and ε_{it} is an idiosyncratic error following a standard logistic distribution $F(x) = e^x/(1+e^x)$. The explanatory variable X_{it} does not contain an intercept, as the terms $\lambda(w_{it})$ and μ_i absorb it. The index is partially linear because X_{it} enters linearly while λ is unrestricted in functional form.

The unobserved heterogeneity has two components, distinguished by whether they are arguments of the link function. The index w_{it} collects unobserved social characteristics, such as motivation, trust, risk tolerance, or social capital, that influence both the agent's economic behavior and the formation of social ties. Unlike the heterogeneity in standard fixed-effects models, w_{it} may vary across both agents and time periods, accommodating settings where the relevant social characteristics evolve. The agent effect μ_i collects the persistent unobserved characteristics that affect the outcome but are not arguments of the link function. When a persistent characteristic affects the links only in part, the part that the links do not reveal belongs to μ_i. Agents of the same network type need not have the same μ_i, and the analysis does not require it. In the smoking application, μ_i may stand for a physiological susceptibility to nicotine or for the smoking environment at home. It is treated as a fixed effect: it is constant over time, it may depend on (X_i, w_i) and on the initial outcome in an arbitrary way, and nothing is assumed about its distribution. The two components are eliminated by different arguments. The agent effect cancels within each agent over time, as in the conditional likelihood approach of Chamberlain (1980) and Honoré and Kyriazidou (2000). The change of $\lambda(w_{it})$ over time, which a within-agent comparison cannot remove, is eliminated by comparing agents whose network types coincide.

In addition to outcomes and covariates, the researcher observes a binary adjacency matrix D_t of dimension $n \times n$ at each period t. The element $D_{ijt} = 1$ if there is a direct link between agents i and j at time t, and $D_{ijt} = 0$ otherwise. Self-links are excluded ($D_{iit} = 0$) and all links are undirected ($D_{ijt} = D_{jit}$). The network evolves over time according to a nonparametric link formation model:

$$D_{ijt} = \mathbb{1}\left\{f(w_{it}, w_{jt}; y_{it-1}, y_{jt-1}) \geq \eta_{ijt}\right\} \mathbb{1}(i \neq j), \quad i, j = 1, \ldots, n, \quad t = 1, \ldots, T, \tag{2}$$

where $f : [0,1]^2 \times \{0,1\}^2 \to [0,1]$ is an unknown jointly symmetric measurable function and η_{ijt} is an idiosyncratic time-varying disturbance. The link formation function f depends on

the agents' current indices of unobserved social characteristics (w_{it}, w_{jt}) and their lagged outcomes (y_{it-1}, y_{jt-1}), allowing network structure to be influenced by prior behavior. This creates a dynamic feedback between outcomes and network formation.

Joint symmetry means $f(w, w'; y, y') = f(w', w; y', y)$: exchanging the two agents, together with their lagged outcomes, leaves the linking probability unchanged, as undirected links require. For each sample size n the link shocks form an $n \times n$ symmetric array $\{\eta_{ijt}\}_{i,j=1}^{n}$, and the dependence on n is left implicit.

The link formation function f in (2) excludes the observed covariates X_{it}, a restriction of the model. Allowing links to depend directly on observed characteristics would require conditional network profiles and additional support and estimation arguments, an extension not studied here.

This specification has four features. First, the model nests two existing models. With $\lambda \equiv 0$ it is the dynamic logit model with individual fixed effects of Honoré and Kyriazidou (2000). With $\mu_i \equiv 0$, $\alpha = 0$, $w_{it} = w_i$, and a link function that does not depend on the lagged outcomes, it is the static semiparametric logit model with an endogenous network of Gueyap Kounga (2023). The contribution of the paper is the combination: a time-varying unobserved term that the network reveals, on top of a fixed effect that it does not. Second, the inclusion of lagged outcomes in the link formation function captures empirically relevant patterns such as homophily on past behavior, for example the tendency of program adopters to form ties with other adopters. Third, the time-varying nature of w_{it} accommodates evolving social characteristics, though the identification strategy also covers the simpler case of time-invariant w_i. Finally, the estimands are the slope and state dependence parameters (β, α), and the functions λ and f and the effects μ_i are nuisance components.

Two clarifications concern the scope of the model. First, the name "social influence function" refers to the agent's own unobserved heterogeneity, the pull toward the outcome exerted by the characteristics that also shape the agent's ties. It does not refer to an effect of other agents' outcomes, and the paper identifies no peer effect. The outcome equation contains neither friends' outcomes nor the realized adjacency matrix, so the network serves only as information about a confounder. In particular, identifying state dependence while allowing for latent heterogeneity that is related to the network does not identify the effect of changing a particular friend's smoking, which would require friends' outcomes in the outcome equation and restrictions that this paper does not impose. The level of λ is not separately identified from μ_i, which can absorb any time-invariant component of it. What can be recovered is its change between waves (Corollary 1). Second, although the functional form of f is left unspecified, the network model is not unrestricted: links are conditionally independent across dyads given the agents' indices and lagged outcomes, the index is scalar,

and links are symmetric.

Figure 1 illustrates the joint evolution of outcomes and network structure over three periods using the adolescent smoking example.

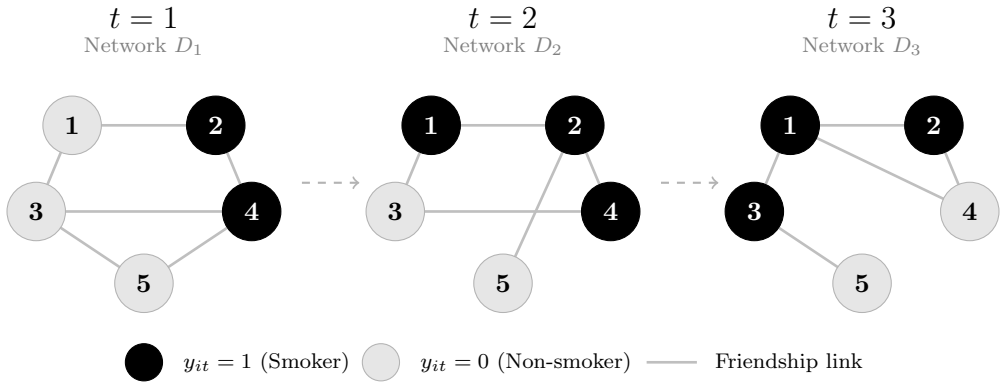


Figure 1: Evolution of outcomes and network structure over three periods. Black nodes represent smokers ($y_{it} = 1$) and gray nodes represent non-smokers ($y_{it} = 0$). Agent 1 transitions from non-smoking to smoking between $t = 1$ and $t = 2$ and persists at $t = 3$, illustrating state dependence (α). Agent 3 begins smoking at $t = 3$ as evolving unobserved characteristics increase risk-taking. Agent 4 quits smoking at $t = 3$ following a shift in unobserved characteristics and simultaneously forms a new link with agent 1. The friendship network changes at each period: the econometrician observes both outcomes and links but cannot directly observe the latent characteristics w_{it} that drive both processes.

The model accommodates a range of empirical settings. In the adolescent smoking application of Fletcher (2010), who uses the longitudinal Add Health data, y_{it} indicates whether student i smokes in survey wave t, the friendship network D_t is observed at each wave and evolves as students form and dissolve friendships, and y_{it-1} captures state dependence through habit formation and addiction. The unobserved characteristic w_{it} can be interpreted as a latent index of sensation-seeking or risk attitude that evolves during adolescence, affecting both the propensity to smoke and the tendency to befriend other risk-taking students. Since w_{it} enters both the outcome equation and the link formation process, standard logit estimates of the slope and state dependence parameters are biased.

A similar structure arises in the agricultural technology adoption setting of Conley and Udry (2010), who study pineapple farmers in Ghana making repeated planting decisions across growing seasons. Here y_{it} indicates whether farmer i adopts the new technology in season t, the information network D_t records which farmers exchange agricultural advice and evolves as farmers update their contacts, and y_{it-1} captures learning-by-doing and switching costs. The unobserved characteristic w_{it} represents latent farming ability and openness to innovation, which may evolve over the seasons for reasons other than the farmer's own past adoption. More able farmers are both more likely to adopt and more sought after as advisors,

generating the same confounding pathway between network structure and outcomes.

The framework also applies to the study of juvenile delinquency by Patacchini and Zenou (2012), who use longitudinal friendship network data from Add Health. In this setting y_{it} indicates whether adolescent i engages in delinquent behavior in wave t, the friendship network D_t evolves across survey waves, and y_{it-1} captures persistence due to reputation effects, reduced stigma, or accumulated criminal capital. The unobserved characteristic w_{it} represents an evolving index of impulsivity or attachment to conventional norms, which influences both the propensity toward delinquency and the formation of friendships with other delinquent youth through conformist motives.

4 Identification

This section develops the main identification results. I first define the concept of network-type equivalence for the dynamic setting, then establish point identification of the parameters (β, α).

The functions λ and f cannot be identified separately. The latent index w_{it} is not observed, and the network data alone cannot pin it down. To see this, consider the homophily specification $f(w_{it}, w_{jt}; y_{it-1}, y_{jt-1}) = 1 - (w_{it} - w_{jt})^2$. If $\{w_{it}\}_{i,t}$ generates a sequence of adjacency matrices $\{D_t\}_t$, then the transformed profile $\{\tilde{w}_{it}\}_{i,t} = \{1 - w_{it}\}_{i,t}$ generates exactly the same matrices, since $(w_{it} - w_{jt})^2 = ((1 - w_{it}) - (1 - w_{jt}))^2 = (\tilde{w}_{it} - \tilde{w}_{jt})^2$. The econometrician cannot distinguish between w_{it} and $\tilde{w}_{it}$ from the observed links. If the outcome function is also relabeled as $\tilde{\lambda}(u) = \lambda(1 - u)$, then $\tilde{\lambda}(\tilde{w}_{it}) = \lambda(w_{it})$ and outcome probabilities are unchanged. The numerical labels of the latent index and the functions indexed by those labels are therefore not separately identified. The agent effect adds a second layer: since μ_i may depend on w_i, the composite value $\lambda(w_{it})$ is not identified either. What is identified is its common movement between the transition waves, on the network classes that the codegree distance recovers (Corollary 1). The identification strategy targets the structural parameters (β, α).

4.1 Network-Type Equivalence

The parameters of interest are β and α. The functions λ and f are treated as nuisance components. Following the approach developed in Gueyap Kounga (2023) for the static case, identification relies on comparing agents with equivalent linking profiles. In the static case the comparison requires their unobserved contributions to outcomes to be equal, whereas in the dynamic case with an agent effect it requires only that the changes of those contributions

between the transition waves be equal.

In the dynamic setting, each agent i's type at period t consists of two components: the index of unobserved social characteristics w_{it} and the observed lagged outcome y_{it-1}. The link formation function $f(w_{it}, w_{jt};\, y_{it-1}, y_{jt-1})$ in (2) depends on both components for each agent. Since the lagged outcomes are observed and discrete, they can be conditioned on directly from the data, reducing the identification problem to controlling for the unobserved component w_{it} through network matching.

For each fixed pair of lagged outcomes $(y, y') \in \{0,1\}^2$, define the *conditional individual link function* of agent i at period t as:

$$f_{w_{it};\, y,y'}(s) = f(w_{it}, s;\, y, y'), \quad s \in [0,1].$$

This is the conditional probability that an agent whose index is w_{it} and whose lagged outcome is y links with agents of every other social characteristic in $[0,1]$ whose lagged outcome is y'. For each fixed (y, y'), this is a function of a single argument $s \in [0,1]$, exactly as in the static individual link function $f_{w_i}(s) = f(w_i, s)$ of Auerbach (2022) and Gueyap Kounga (2023). The key difference is that the dynamic setting introduces four such functions, one for each configuration $(y, y') \in \{0,1\}^2$, whereas the static setting has only one.

To quantify the similarity between two agents' linking behavior, define the *network-type distance* at period t for agents sharing the same lagged outcome $y_{it-1} = y_{jt-1} = y$ as:

$$\begin{aligned}\rho_{ijt} = \|f_{w_{it};\, y,\cdot} - f_{w_{jt};\, y,\cdot}\|_2 &= \left[\sum_{y' \in \{0,1\}} \int_0^1 \left(f_{w_{it};\, y,y'}(s) - f_{w_{jt};\, y,y'}(s)\right)^2 ds\right]^{1/2} \\ &= \left[\sum_{y' \in \{0,1\}} \int_0^1 \left(f(w_{it}, s;\, y, y') - f(w_{jt}, s;\, y, y')\right)^2 ds\right]^{1/2}. \qquad (3)\end{aligned}$$

This aggregates the L^2 distance between the conditional individual link functions over both possible partner lagged outcomes. When $\rho_{ijt} = 0$, agents i and j produce the same conditional linking probabilities for almost every partner characteristic $s \in [0,1]$ and each lagged outcome $y' \in \{0,1\}$, and they are then called *conditionally network-type equivalent* at period t.

Conditional network-type equivalence does not imply that the two agents share the same value of the index of unobserved social characteristics. To see this, consider the link formation function $f(w_{it}, w_{jt}; y_{it-1}, y_{jt-1}) = 1 - (w_{it} - \frac{1}{2})^2 - (w_{jt} - \frac{1}{2})^2$. The individual link function is $f_{w;\, y,y'}(s) = 1 - (w - \frac{1}{2})^2 - (s - \frac{1}{2})^2$, which depends on w only through $(w - \frac{1}{2})^2$. Agents with $w_{it} = 0.3$ and $w_{jt} = 0.7$ produce identical linking probabilities with every other agent, so $\rho_{ijt} = 0$ even though $w_{it} \neq w_{jt}$. In the adolescent smoking application, a low-effort student

and a high-effort student may form friendships with the same types of peers yet differ in their underlying characteristic. The identification strategy does not require distinguishing between them, only that their social influence functions change by the same amount between the transition waves. Figure 2 illustrates such a pair.

When the link function does not depend on lagged outcomes, so that

$$f(w_{it}, w_{jt};\, y_{it-1}, y_{jt-1}) = g(w_{it}, w_{jt})$$

for some function g, the conditional individual link function $f_{w_{it};\,y,y'}(s) = g(w_{it}, s)$ is the same for all (y, y'), and the distance reduces to $\rho_{ijt} = \sqrt{2}\,\|g_{w_{it}} - g_{w_{jt}}\|_2$. In this case, every period is informative for type matching, regardless of lagged outcomes.

I now state the formal assumptions. Throughout, $X_i = (X_{i1}, \ldots, X_{iT})$, $w_i = (w_{i1}, \ldots, w_{iT})$, and $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})$ denote agent i's covariate path, latent-index path, and outcome shocks, and $\chi_i = (y_{i0}, X_i, w_i, \mu_i)$ collects the agent's primitives other than the shocks.

Assumption 1 (Sampling and distributional structure)**.**

(i) *The vectors $(\chi_i, \varepsilon_i) = (y_{i0}, X_i, w_i, \mu_i, \varepsilon_i)$, $i = 1, \ldots, n$, are i.i.d.*

(ii) *For each t, the array $\{\eta_{ijt}\}_{i,j=1}^{n}$ is symmetric, with i.i.d. standard uniform entries above the diagonal. The arrays are independent across t and independent of $\{(\chi_i, \varepsilon_i)\}_{i=1}^{n}$.*

(iii) *For each t, w_{it} has a distribution $G_{w,t}$ on $[0, 1]$, which may differ across periods.*

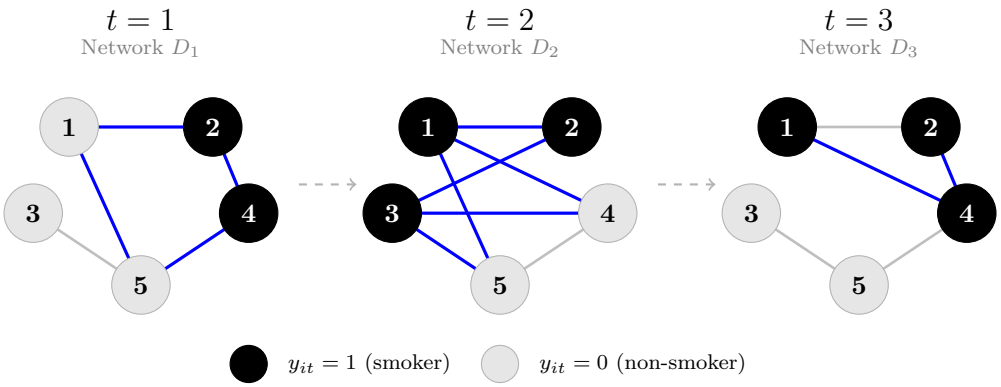
(iv) *The shocks ε_{it} are standard logistic and i.i.d. across i and t, and the vector ε_i is independent of $\chi_i = (y_{i0}, X_i, w_i, \mu_i)$.*

(v) *The outcomes y_{it} and the adjacency matrices D_t, $t = 1, \ldots, T$, are generated by equations* (1) *and* (2)*.*

(vi) *λ and f are measurable, and f is jointly symmetric: $f(w_1, w_2;\, y, y') = f(w_2, w_1;\, y', y)$.*

Assumption 1(i) includes the initial outcome and the agent effect in the vector that is sampled independently across agents. The initial outcome may depend on the covariate and latent-index paths and on the agent effect in an arbitrary way, and the agent effect may depend on the covariate and latent-index paths and on the initial outcome in an arbitrary way. Both are compatible with conditional likelihood, but part (iv) requires the outcome shocks to be independent of them. This excludes, for example, $y_{i0} = \mathbb{1}\{\varepsilon_{i1} > 0\}$, under which ε_{i1} would remain unconditionally logistic and independent of (X_i, w_i) but would be truncated given the initial outcome, so that the first transition probability used in Lemma 1 would be

(a) Conditionally network-type equivalent agents



(b) Movement of the social influence term under Assumption 2

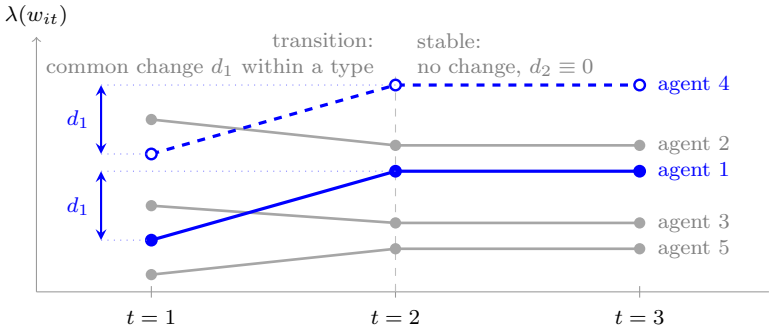


Figure 2: Network-type equivalence and the movement of the social influence term. Panel (a): conditionally network-type equivalent agents. At $t = 1$, agents 1 and 4 are connected to the same set of third agents (blue links): both link to agents 2 and 5, and neither links to agent 3. Here the realized neighborhoods are consistent with $\rho_{14,1} = 0$ provided $y_{1,0} = y_{4,0}$. By contrast, the realized neighborhoods of agents 2 and 5 differ, since agent 5 is linked with agent 3 while agent 2 is not. The illustration takes this to reflect different link functions, although a single realized graph cannot by itself rule out equal link probabilities. At $t = 2$, agents 1 and 3 share the same lagged outcome ($y_{1,1} = y_{3,1} = 0$) and are connected to the same set of third agents: both link to agents 2, 4, and 5. Their realized neighborhoods coincide in this illustration, which is consistent with $\rho_{13,2} = 0$. At $t = 3$, agents 1 and 2 share the same lagged outcome ($y_{1,2} = y_{2,2} = 1$) and, excluding their mutual link, are connected to the same set of third agents: among agents 3, 4, and 5, both link only to agent 4, consistent with $\rho_{12,3} = 0$. Panel (b): the social influence term $\lambda(w_{it})$ of the same five agents under Assumption 2. Agents 1 and 4, which are equivalent at $t = 1$, have different levels but the same change d_1 between $t = 1$ and $t = 2$ (parallel segments, equal arrows), while agents 2, 3, and 5, whose network types differ from theirs at $t = 1$, change by different amounts. Between $t = 2$ and $t = 3$, the term does not change for any agent. The levels of λ, like the agent effects μ_i, are unrestricted. Agents 1 and 4 also switch in opposite directions between $t = 1$ and $t = 2$, which is the comparison that the identification argument uses.

wrong. Part (i) allows temporal dependence within an agent's covariate and latent paths but requires the primitives to be drawn independently across agents, which rules out aggregate shocks common to all agents but does not restrict the dependence that arises endogenously through the network. Part (ii) specifies the link shocks for all waves jointly: within each wave they are symmetric ($\eta_{ijt} = \eta_{jit}$, consistent with undirected links) and i.i.d. above the diagonal, and across waves they are independent, so that the links of different dyads are independent given the agents' latent indices and realized lagged outcomes (dyads that share an agent share her realized lagged outcome, so they are not independent given the primitives alone). The independence of the link shocks from the agent primitives means that unobserved factors affecting both linking and outcomes operate through w_{it}. In particular, the agent effect μ_i does not enter the link function as a separate argument, which is what distinguishes it from $\lambda(w_{it})$, although it may be associated with the links through its dependence on w_i and through the lagged outcomes that enter f. This is a restriction: any time-varying part of the unobserved heterogeneity must be a function of the index that drives links, and whatever affects the outcome without shaping ties must be constant over time. Within that restriction the split is a definition: any persistent characteristic that shapes ties belongs to w_{it}, and μ_i is what remains. The arrays have dimension $n \times n$ at sample size n, a dependence left implicit in the notation.

Assumption 1(iii) lets the distribution of the latent index differ across periods, since the index moves over time. Nothing in the identification argument uses a particular marginal distribution. The population codegree distance of Section 5 integrates against the period-specific joint distribution of $(w_{lt}, y_{l,t-1})$. The uniform distribution of η_{ijt} makes $f(w_{it}, w_{jt};\, y_{it-1}, y_{jt-1})$ a conditional linking probability.

Assumption 1(iv) contains two substantive restrictions. The logistic distribution is what makes the conditional likelihood work: the cancellation of the social influence terms relies on the log-linear structure of the logistic odds ratio, a property first exploited by Andersen (1970) and Chamberlain (1980) to eliminate fixed effects through sufficient statistics and shown by Magnac (2004) to be essentially unique to the logistic family. See also Chamberlain (2010) and, for what survives without it, the partial identification results of Honoré and Tamer (2006). The independence of ε_i from $(y_{i0}, X_i, w_i, \mu_i)$ is used in the factorization of path probabilities in Lemma 1, which conditions on the entire covariate and latent-index paths, including their future values, and it excludes feedback from realized outcomes into future covariates or into the future latent index. This is a real restriction in the motivating applications: it rules out, for instance, a risk-attitude index w_{it} that responds to the agent's own past smoking, and covariates such as alcohol use that respond to past smoking shocks. Serial dependence of any form in (X_i, w_i) remains allowed. Assumption 1(v) ties the observed data to the structural

equations, and part (vi) is the joint symmetry discussed after equation (2).

A remark on conditioning is in order before the next assumption. When network types vary continuously, the events $\{\rho_{ijt} = 0\}$ and $\{f_{w_{it}} = \varphi\}$ have probability zero, and conditional expectations given these events are not defined in the elementary sense. For statements about population objects, such conditional expectations are defined as limits over shrinking neighborhoods, in the manner of Section 2.2.2 of Auerbach (2022). Write $f_w := (f_{w;\,y,y'})_{(y,y')\in\{0,1\}^2}$ for the full profile of conditional individual link functions of a type w, and $\|f_u - \varphi\|_2^2 := \sum_{y,y'\in\{0,1\}} \int_0^1 \big(f(u,s;\,y,y') - \varphi_{y,y'}(s)\big)^2 ds$ for the distance between such a profile and a candidate profile φ. For any integrable random vector Π_{ij} indexed at the pair level,

$$\mathbb{E}\left[\Pi_{ij} \,|\, \rho_{ijt} = 0\right] := \lim_{h\downarrow 0} \mathbb{E}\left[\Pi_{ij} \,|\, \rho_{ijt} \le h\right], \tag{4}$$

and, for objects Π_j attached to a single agent j and a network type φ in the support of $f_{w_{jt}}$,

$$\mathbb{E}\big[\Pi_j \,\big|\, f_{w_{jt}} = \varphi\big] := \lim_{h\downarrow 0} \mathbb{E}\Big[\Pi_j \,\Big|\, w_{jt} \in \big\{u \in [0,1] : \|f_u - \varphi\|_2 \le h\big\}\Big],$$

evaluated at $\varphi = f_{w_{it}}$ when the conditioning is on agent i's network type. Analogously, for conditioning on covariate stability,

$$\mathbb{E}\left[\,\cdot \,|\, X_{i,s+2} = X_{i,s+1}\right] := \lim_{h\downarrow 0} \mathbb{E}\left[\,\cdot \,\big|\, \|X_{i,s+2} - X_{i,s+1}\|^2 \le h\right].$$

Whenever a conditional expectation of this form is used, the relevant limit is assumed to exist. Separate limits of this kind need not coincide with the limit of a criterion that localizes in several directions at once with particular relative bandwidths. For that reason, the identification results in the continuous case (Theorem 1(b) and Corollary 1(b)) localize jointly in the codegree distance and in covariate stability, with explicit bounds, and the estimation theory of Section 6 works directly with the localized criterion that the estimator computes. The convention only gives a meaning to informal statements such as "conditional on the network type," and no formal result relies on it. When types or covariates have discrete support, the conditioning events have positive probability, the limits coincide with elementary conditional expectations, and no limiting argument is needed.

The identification argument does not require agents with the same network type to have the same value of the social influence function. It requires a restriction on how $\lambda(w_{it})$ moves between periods. Let $\mathcal{S} \subseteq \{1, \ldots, T-2\}$ denote the set of transition periods used in estimation.

Assumption 2 (Latent index dynamics)**.**

(i) (Movement common to network types.) For each $t \le T-1$ there is a measurable

function $d_t : [0,1] \to \mathbb{R}$ *such that* $\lambda(w_{i,t+1}) - \lambda(w_{it}) = d_t(w_{it})$ *almost surely, and* d_t *is uniformly continuous with respect to the network-type distance: for every* $\kappa > 0$ *there exists an* $r > 0$ *such that*

$$\sup_{u,v \in [0,1]:\, \|f_{u;\,y,\cdot} - f_{v;\,y,\cdot}\|_2 \le r \text{ for some } y \in \{0,1\}} \big(d_t(u) - d_t(v)\big)^2 \;\le\; \kappa,$$

where $\|f_{u;\,y,\cdot} - f_{v;\,y,\cdot}\|_2$ *is the distance between conditional individual link functions in equation* (3)*. The magnitude of* d_t *is unrestricted.*

(ii) (Stability after the transition.) For each $s \in \mathcal{S}$, $d_{s+1} \equiv 0$.

Assumption 2 is the restriction that the time-varying generalization requires, and it is what the identification result from observables uses. Part (i) says that the change in the social influence term between consecutive periods is common to agents who form links in the same way and similar for agents who form links in similar ways. In particular, d_t is constant on network-type equivalence classes, and it may be large and need not shrink with the sample. The uniform continuity is stated in the metric of equation (3), in the manner of Assumption 3 in Auerbach (2022), and the quantifier "for some y" matters: matched pairs share a single realized lagged outcome at each period, so the data certify proximity of the link functions only in the stratum the pair realizes. Part (ii) says that each transition period used in estimation is followed by a stable period.

The assumption restricts the changes of the social influence function, not its level. Two agents with the same network type may have different values of $\lambda(w_{it})$ at every date and different agent effects μ_i. What they share is the movement of λ between consecutive dates, which Figure 2(b) illustrates for the agents of panel (a). This is weaker than requiring equal values of λ within a network type, as in Auerbach (2022) and Gueyap Kounga (2023), and it is what allows the model to carry an agent effect that network matching does not equalize. The assumption is plausible when the same characteristics that drive the outcome also drive the formation of ties, so that a change in those characteristics shows up in both. It fails when the movement of the social influence function differs across agents of the same network type, as when the characteristic that moves affects the outcome but plays no role in the formation of ties and moves idiosyncratically. A movement common to every agent is allowed, whether or not it works through the formation of ties, because it is constant within every class. Since λ is never observed, the assumption cannot be tested from the data alone. Three observations order its content. First, when $w_{it} = w_i$ is time-invariant, $d_t \equiv 0$ and both parts hold trivially. Second, when the latent index evolves by a common map, $w_{i,t+1} = g_t(w_{it})$, the change is $d_t = \lambda \circ g_t - \lambda$, and part (i) holds if this induced change is

constant on each network-type equivalence class and uniformly continuous in the metric of equation (3). Movements of any size are then allowed. A common map that preserves the classes does not by itself ensure this when λ varies within a class. For instance, with the lag-independent link function $f(u,v) = \frac{1}{4} + \frac{1}{4}\big[(u-\frac{1}{2})^2 + (v-\frac{1}{2})^2\big]$, the indices u and $1-u$ have identical link functions, the map $g(u) = 1-u$ sends each index to a member of its own class, and yet with $\lambda(u) = u$ the change $d(u) = 1-2u$ equals 0.6 at $u = 0.2$ and -0.6 at $u = 0.8$, so two equivalent indices move by different amounts. Examples in which the condition holds are a cohort-wide change in risk attitude between two survey waves whose size depends only on the student's position in the friendship structure, and the finite-type designs of Section 7 that satisfy the assumption, in which the supported types have distinct link profiles, so that every equivalence class is a single type, and types move between network blocks by a common permutation. Third, part (ii) interacts with the choice of $\mathcal{S}$: if every interior period is used, it forces $d_2 = \cdots = d_{T-1} = 0$, so that the only unrestricted common movement is between periods 1 and 2. With longer panels, one may instead use alternating windows of movement and stability and restrict $\mathcal{S}$ accordingly. With $T = 3$, $\mathcal{S} = \{1\}$ and the restriction is $d_2 \equiv 0$. What the assumption excludes is movement that differs by a non-vanishing amount across agents with the same network type, for example a stationary autoregression with nondegenerate innovations. In that case, neither within-agent nor cross-agent comparisons remove the movement. Section 6 relaxes parts (i) and (ii) to allow an idiosyncratic component whose maximum over the sample shrinks as the sample grows (Assumption 3), which is the intermediate case of slowly evolving characteristics in short panels.

Appendix D describes the observation window that the assumption requires in the smoking example, a scheduled school transition followed by a stable period, and how the waves of the Glasgow data map onto it.

Two consequences are used below. First, if agents i and j have the same network type at period s in the stratum of their common lagged outcome, then $d_s(w_{is}) = d_s(w_{js})$, because the uniform-continuity condition applied at every $\kappa > 0$ forces d_s to be constant on equivalence classes, and by part (i)

$$\lambda(w_{i,s+1}) - \lambda(w_{is}) = d_s(w_{is}) = d_s(w_{js}) = \lambda(w_{j,s+1}) - \lambda(w_{js}),$$

so the two agents' social influence terms change by the same amount between s and $s+1$, whether or not their levels, or their agent effects, are equal. For agents whose network types are close but not equal, write $\varpi(r) = \sup\{|d_t(u) - d_t(v)| : \|f_{u;\,y,\cdot} - f_{v;\,y,\cdot}\|_2 \le r \text{ for some } y\}$ for the modulus of continuity in part (i), which tends to zero as $r \to 0$ and is what the estimation theory uses. Second, by part (ii), $\lambda(w_{m,s+2}) = \lambda(w_{m,s+1})$ for every agent m and

every $s \in \mathcal{S}$. The first consequence is what makes equivalence at period s, which the data can certify for pairs sharing the lagged outcome at $s-1$, sufficient for the cancellation in Lemma 2. Equivalence at $s+1$ cannot be certified for the pairs the estimator uses, since their lagged outcomes differ at $s+1$ (Section 5).

4.2 Identification of Slope and State Dependence Parameters

The identification strategy combines three elements: (i) the Chamberlain conditional likelihood approach for logit models, which eliminates the agent effect μ_i and the level of λ within each agent, (ii) cross-agent matching on network types to eliminate the change in the social influence function, and (iii) stability of covariates and of the social influence term in the period that follows the transition.

For any interior period $s \in \{1, \ldots, T-2\}$ (so that period $s+2$ is observed), define two outcome paths for agent i that differ only in the ordering of outcomes at periods s and $s+1$:

$$A_s = \{y_{i0}, \ldots, y_{i,s-1}, \underbrace{0}_{y_{is}}, \underbrace{1}_{y_{i,s+1}}, y_{i,s+2}, \ldots, y_{iT}\},$$
$$B_s = \{y_{i0}, \ldots, y_{i,s-1}, \underbrace{1}_{y_{is}}, \underbrace{0}_{y_{i,s+1}}, y_{i,s+2}, \ldots, y_{iT}\},$$

where the outcomes outside periods s and $s+1$ are fixed at the same values in both paths. The total $y_{is} + y_{i,s+1} = 1$ is the same for both paths. However, the lagged values entering the model differ at periods $s+1$ and $s+2$: in path A_s the lag at $s+1$ is 0 and the lag at $s+2$ is 1, whereas in path B_s the lag at $s+1$ is 1 and the lag at $s+2$ is 0.

Lemma 1 (Within-agent likelihood ratio)**.** *Under Assumption 1, conditional on $\chi_i = (y_{i0}, X_i, w_i, \mu_i)$ with $X_{i,s+2} = X_{i,s+1}$ and $\lambda(w_{i,s+2}) = \lambda(w_{i,s+1})$, the likelihood ratio for agent i between paths B_s and A_s satisfies:*

$$\frac{\mathbb{P}(y_i \in B_s \mid \chi_i)}{\mathbb{P}(y_i \in A_s \mid \chi_i)} = \exp\left[(X_{is} - X_{i,s+1})'\beta + \alpha(y_{i,s-1} - y_{i,s+2}) + \lambda(w_{is}) - \lambda(w_{i,s+1})\right]. \tag{5}$$

The ratio does not depend on μ_i.

The proof is in Appendix A.1. For $T = 3$ (so $s = 1$) the content of the lemma, and of the comparison with within-agent conditioning, can be seen in one display. Write $c_{it} = X_{it}'\beta + \lambda(w_{it}) + \mu_i$ for the index without the lag and $\Gamma(c) = \log(1 + e^{c+\alpha}) - \log(1 + e^c)$. Without imposing any stability condition, the exact log odds of the two switched paths are

$$\log \frac{\mathbb{P}(y_i \in B_1 \mid \chi_i)}{\mathbb{P}(y_i \in A_1 \mid \chi_i)} = (X_{i1} - X_{i2})'\beta + \alpha(y_{i0} - y_{i3}) + \big[\lambda(w_{i1}) - \lambda(w_{i2})\big] + \big[\Gamma(c_{i3}) - \Gamma(c_{i2})\big]. \tag{6}$$

The agent effect cancels in $c_{i1} - c_{i2}$, which is the covariate term plus the first bracket, because it is constant over time. What survives of λ there is its change between periods 1 and 2, not its level. This is the within-agent elimination of a fixed effect of Chamberlain (1980) and Honoré and Kyriazidou (2000), applied to $\lambda(w_{it}) + \mu_i$. Equation (5) imposes the stability condition of the lemma, under which $c_{i3} = c_{i2}$ and the last bracket vanishes. Equation (6) imposes no stability condition, which is why the Γ terms appear. The last bracket is the period-3 term, and μ_i does enter it through the level c_{i2} unless $c_{i3} = c_{i2}$. Since $\Gamma'(c) = F(c+\alpha) - F(c)$ lies in $(-1, 1)$, the bracket is bounded by $|c_{i3} - c_{i2}| \leq \|\beta\| \, \|X_{i3} - X_{i2}\| + |\lambda(w_{i3}) - \lambda(w_{i2})|$, which is free of μ_i, and it vanishes when covariates and social influence are stable between periods 2 and 3, which is what the covariate kernel K_2 of Section 5 and Assumption 2(ii) target. Within-agent conditioning can remove the first bracket only by assuming it away, as the fixed-effect specification of Honoré and Kyriazidou (2000) does. Cross-agent matching removes it without that assumption: for two agents whose social influence terms change by the same amount between periods 1 and 2, the bracket is the same for both, so it cancels in the ratio of their two ratios.

The likelihood ratio in (5) still contains the nuisance term $\lambda(w_{is}) - \lambda(w_{i,s+1})$. To eliminate it, I compare two agents, whose nuisance terms cancel if and only if $\lambda(w_{is}) - \lambda(w_{i,s+1}) = \lambda(w_{js}) - \lambda(w_{j,s+1})$. Only equal changes are needed. The levels of λ and the agent effects may differ across the two agents, because each cancels within its own agent.

To state the comparison with its complete conditioning information, fix an interior period s and two agents $i \neq j$. Write $y_{a,-s}$ for agent a's outcomes at periods other than s and $s+1$ (including y_{a0}), and let

$$\mathcal{F}_{ijs} = \sigma\big(X_i, X_j, y_{i,-s}, y_{j,-s}\big).$$

The pair is *discordant* at s if $y_{is} + y_{i,s+1} = y_{js} + y_{j,s+1} = 1$ and $y_{is} \neq y_{js}$, an event denoted $\mathcal{D}_{ijs}$. Given $\mathcal{F}_{ijs}$, the only outcome configurations compatible with it are $(y_i, y_j) \in A_s \times B_s$ and $(y_i, y_j) \in B_s \times A_s$, with the outcomes outside the window fixed at their realized values. Define the latent event and the covariate-stability event

$$\mathcal{C}_{ijs} = \big\{\lambda(w_{is}) - \lambda(w_{i,s+1}) = \lambda(w_{js}) - \lambda(w_{j,s+1});\ \lambda(w_{a,s+2}) = \lambda(w_{a,s+1}) \text{ for } a \in \{i, j\}\big\},$$
$$\mathcal{X}_{ijs} = \big\{X_{a,s+2} = X_{a,s+1} \text{ for } a \in \{i, j\}\big\},$$

and the index

$$V_{ijs}(b, a) = (\Delta_s X_j - \Delta_s X_i)'b + a\,(y_{j,s-1} - y_{j,s+2} - y_{i,s-1} + y_{i,s+2}), \qquad \Delta_s X_i = X_{is} - X_{i,s+1}.$$

Lemma 2 (Conditional likelihood identity)**.** *Under Assumption 1, on the event* $\mathcal{C}_{ijs} \cap \mathcal{X}_{ijs}$*,*

$$\mathbb{P}\big(y_i \in A_s,\, y_j \in B_s \,\big|\, \mathcal{F}_{ijs},\, w_i,\, w_j,\, \mathcal{D}_{ijs}\big) = F\big(V_{ijs}(\beta,\alpha)\big). \tag{7}$$

The same identity holds with (μ_i, μ_j) *added to the conditioning information. Consequently, for every event* E *in* $\sigma(\mathcal{F}_{ijs}, w_i, w_j, \mu_i, \mu_j)$ *with* $E \subseteq \mathcal{C}_{ijs} \cap \mathcal{X}_{ijs}$ *and positive probability,* $\mathbb{P}(y_i \in A_s, y_j \in B_s \mid \mathcal{F}_{ijs}, E, \mathcal{D}_{ijs}) = F(V_{ijs}(\beta,\alpha))$*.*

Proof. By Assumption 1(i) and (iv), the outcome paths of agents i and j are conditionally independent given (χ_i, χ_j), and each agent's conditional path distribution depends only on its own primitives. Hence, given $\mathcal{F}_{ijs}$ and (w_i, w_j, μ_i, μ_j),

$$\frac{\mathbb{P}(y_i \in A_s, y_j \in B_s \mid \cdot)}{\mathbb{P}(y_i \in B_s, y_j \in A_s \mid \cdot)} = \frac{\mathbb{P}(y_i \in A_s \mid \chi_i)}{\mathbb{P}(y_i \in B_s \mid \chi_i)} \cdot \frac{\mathbb{P}(y_j \in B_s \mid \chi_j)}{\mathbb{P}(y_j \in A_s \mid \chi_j)}.$$

On $\mathcal{C}_{ijs} \cap \mathcal{X}_{ijs}$, Lemma 1 applies to both agents, the agent effects drop out of both ratios, and the equality of the changes in λ gives

$$\begin{aligned}\frac{\mathbb{P}(y_i \in A_s, y_j \in B_s \mid \cdot)}{\mathbb{P}(y_i \in B_s, y_j \in A_s \mid \cdot)} &= \exp\left[(\Delta_s X_j - \Delta_s X_i)'\beta + \alpha(y_{j,s-1} - y_{j,s+2} - y_{i,s-1} + y_{i,s+2})\right] \\ &= \exp\big(V_{ijs}(\beta,\alpha)\big).\end{aligned}$$

Given $\mathcal{D}_{ijs}$ and the outcomes outside the window, exactly two configurations are possible, with conditional probabilities proportional to Π and Π', where $\Pi/\Pi' = \exp(V_{ijs})$. Hence the conditional probability of $A_s \times B_s$ is $\Pi/(\Pi + \Pi') = F(V_{ijs})$ given $(\mathcal{F}_{ijs}, w_i, w_j, \mu_i, \mu_j, \mathcal{D}_{ijs})$. Since the right side is $\mathcal{F}_{ijs}$-measurable, integrating out (μ_i, μ_j) gives (7), and the last statement follows by iterated expectations, because the right side equals $F(V_{ijs})$ for every (w_i, w_j, μ_i, μ_j) in E. □

Lemma 2 is a conditional likelihood identity on a latent event. It is not yet an identification result, because $\mathcal{C}_{ijs}$ involves the unobserved index. The network supplies an observable event contained in $\mathcal{C}_{ijs}$, and Proposition 1 makes precise in what sense it is observable. Let $\delta_{ijs} = \delta_s(w_{is}, w_{js};\, y_{i,s-1})$ denote the population codegree distance between agents i and j at period s, defined in Section 5 for pairs with $y_{i,s-1} = y_{j,s-1}$, and let $\hat{\delta}_{ijs}$ be its sample counterpart, a statistic of the observed networks and lagged outcomes. Following Auerbach (2022), identification is understood relative to the growing-network experiment: the observed data are the outcomes, covariates, and networks $\{(y_i, X_i)\}_{i \le n}$, $\{D_t\}_{t \le T}$ of samples of size $n \to \infty$ drawn under Assumption 1, and a pair-level quantity is identified in this experiment when it is the probability limit, uniformly over pairs, of a statistic of the observed data. A

parameter is identified when it is determined by the joint law of the observed data together with such limits.

Proposition 1 (The codegree distance as an identified object)**.** *Suppose Assumption 1 holds.*

(a) For every s, $\max_{i\neq j} |\hat{\delta}^2_{ijs} - \delta^2_{ijs}|\, \mathbb{1}\{y_{i,s-1} = y_{j,s-1}\} = O_p((\log n/n)^{1/2})$, so δ_{ijs} is identified in the growing-network experiment.

(b) Under Assumption 6 (continuity of f and full support), for pairs sharing the lagged outcome y at $s-1$, $\delta_{ijs} = 0$ if and only if $\|f_{w_{is};\,y,\cdot} - f_{w_{js};\,y,\cdot}\|_2 = 0$, and the modulus

$$\tau(h) := \sup\big\{\|f_{u;\,y,\cdot} - f_{v;\,y,\cdot}\|_2 : y \in \{0,1\},\ \delta_s(u,v;\,y)^2 \le h\big\}$$

tends to zero as $h \downarrow 0$. Consequently, under Assumption 2(i), $\delta_{ijs} = 0$ implies $d_s(w_{is}) = d_s(w_{js})$, and $\delta^2_{ijs} \le h$ implies $|d_s(w_{is}) - d_s(w_{js})| \le \varpi(\tau(h))$, with ϖ the modulus of continuity of d_s defined after Assumption 2.

(c) In finite-type designs, under Assumption 7(i)–(ii) of Section 6.4, $\delta_{ijs} = 0$ if and only if the two agents are network-type equivalent on the support, in which case $d_s(w_{is}) = d_s(w_{js})$, and every positive value of δ^2_{ijs} is at least $\bar{\epsilon} > 0$, with $\bar{\epsilon}$ as in Assumption 7(v).

Part (a) is Lemma 4(a) and (d) in Appendix A, and parts (b) and (c) are proved in Appendix A.4. Define the observable selection event and its neighborhood version,

$$\begin{aligned}\mathcal{O}_{ijs} &= \big\{y_{i,s-1} = y_{j,s-1},\ \delta_{ijs} = 0\big\} \cap \mathcal{X}_{ijs},\\ \mathcal{O}_{ijs}(h) &= \big\{y_{i,s-1} = y_{j,s-1},\ \delta^2_{ijs} \le h\big\} \cap \big\{\|X_{a,s+2} - X_{a,s+1}\|^2 \le h \text{ for } a \in \{i,j\}\big\},\end{aligned}$$

and the regressor $Z_{ijs} = \big(\Delta_s X_i - \Delta_s X_j,\ y_{i,s-1} - y_{i,s+2} - y_{j,s-1} + y_{j,s+2}\big)'$, so that $V_{ijs}(b,a) = -Z'_{ijs}(b',a)'$. Write $\theta = (\beta',\alpha)'$. The event $\mathcal{O}_{ijs}$ has positive probability when types and covariates have discrete support, and probability zero when they are continuous. The theorem treats the two cases separately.

Theorem 1 (Identification of β and α)**.** *Suppose Assumptions 1 and 2 hold and fix $s \in \mathcal{S}$.*

(a) (Exact matching.) Suppose that $\mathbb{P}(\mathcal{O}_{ijs} \cap \mathcal{D}_{ijs}) > 0$ and that $\delta_{ijs} = 0$ implies $d_s(w_{is}) = d_s(w_{js})$ for pairs sharing the lagged outcome at $s-1$. By Proposition 1, the implication holds in finite-type designs under Assumption 7(i)–(ii), without the continuity requirement in Assumption 2(i), and in general under Assumption 6. Then $\mathcal{O}_{ijs} \subseteq \mathcal{C}_{ijs} \cap \mathcal{X}_{ijs}$ and

$$\mathbb{P}\big(y_i \in A_s,\, y_j \in B_s \,\big|\, \mathcal{F}_{ijs},\, \mathcal{O}_{ijs},\, \mathcal{D}_{ijs}\big) = F\big(V_{ijs}(\beta,\alpha)\big). \tag{8}$$

If, in addition, $\mathbb{E}[Z_{ijs}Z_{ijs}' \mid \mathcal{O}_{ijs}, \mathcal{D}_{ijs}]$ *is positive definite, then* (β, α) *is the unique maximizer of the conditional expected log-likelihood* $\mathbb{E}[m_{ijs}(b,a) \mid \mathcal{O}_{ijs}, \mathcal{D}_{ijs}]$*, with* m_{ijs} *defined in Section 5, and is therefore identified.*

(b) (Continuous types.) Suppose Assumption 6 holds and $\mathbb{P}(\mathcal{O}_{ijs}(h) \cap \mathcal{D}_{ijs}) > 0$ *for every* $h > 0$*. Then, on* $\mathcal{O}_{ijs}(h)$*,*

$$\mathbb{P}\big(y_i \in A_s,\, y_j \in B_s \,\big|\, \mathcal{F}_{ijs}, w_i, w_j, \mu_i, \mu_j, \mathcal{D}_{ijs}\big) = F\big(V_{ijs}(\beta,\alpha) + \Delta_{ijs}\big), \qquad |\Delta_{ijs}| \le \varpi_1(h) := \varpi(\tau(h)) + 2\|\beta\|\, h^{1/2}, \tag{9}$$

where the remainder Δ_{ijs} *may depend on the agent effects but its bound does not. Hence* $\big|\mathbb{P}(y_i \in A_s, y_j \in B_s \mid \mathcal{F}_{ijs}, \mathcal{O}_{ijs}(h), \mathcal{D}_{ijs}) - F(V_{ijs}(\beta,\alpha))\big| \le \varpi_1(h)/4 \to 0$ *as* $h \downarrow 0$*. If, in addition,* $\sup_{0<h\le\bar{h}} \mathbb{E}[\|Z_{ijs}\|^4 \mid \mathcal{O}_{ijs}(h), \mathcal{D}_{ijs}] < \infty$ *for some* $\bar{h} > 0$ *and* $\liminf_{h\downarrow 0} \lambda_{\min}\big(\mathbb{E}[F(Z_{ijs}'\theta)(1 - F(Z_{ijs}'\theta))Z_{ijs}Z_{ijs}' \mid \mathcal{O}_{ijs}(h), \mathcal{D}_{ijs}]\big) > 0$*, then for every sufficiently small* h *the localized criterion* $M_h(b,a) = \mathbb{E}[m_{ijs}(b,a) \mid \mathcal{O}_{ijs}(h), \mathcal{D}_{ijs}]$ *has a unique maximizer* θ_h *on the parameter space* Θ*, a compact convex set with* (β,α) *in its interior as in Assumption 5(ii), and* $\theta_h \to (\beta,\alpha)$ *as* $h \downarrow 0$*. In this sense* (β,α) *is identified as the limit of the localized comparison, which is the object that the estimator of Section 5 computes.*

Proof. (a) On $\mathcal{O}_{ijs}$, the implication assumed in part (a) gives $d_s(w_{is}) = d_s(w_{js})$, hence $\lambda(w_{i,s+1}) - \lambda(w_{is}) = \lambda(w_{j,s+1}) - \lambda(w_{js})$. By Assumption 2(ii), $\lambda(w_{a,s+2}) = \lambda(w_{a,s+1})$ for $a \in \{i,j\}$. Thus $\mathcal{O}_{ijs} \subseteq \mathcal{C}_{ijs} \cap \mathcal{X}_{ijs}$. Since $\mathcal{O}_{ijs}$ is determined by $(X_i, X_j, y_{i,s-1}, y_{j,s-1}, w_{is}, w_{js})$, it belongs to $\sigma(\mathcal{F}_{ijs}, w_i, w_j)$, and (8) follows from Lemma 2. The agent effects never enter. The expected conditional log-likelihood is then that of a correctly specified logit in the regressor Z_{ijs}, which is strictly concave with a unique maximizer at (β,α) under the rank condition. Part (b) is proved in Appendix A.4. □

Four remarks clarify the theorem. First, the conditioning information is explicit: the probability in (8) conditions on the covariate paths and on the outcomes outside the window, which is why its right side varies with them. Second, the rank condition is imposed on the population actually selected, including the shared lagged outcome at $s-1$. On that population the second component of Z_{ijs} reduces to $y_{j,s+2} - y_{i,s+2}$, and a rank condition imposed without the shared lag would not suffice: if $X_{i,s} - X_{i,s+1}$ were a function of $y_{i,s-1}$ alone, the first component would vanish on every selected pair while having positive variance on the larger population. Third, $\mathcal{O}_{ijs}$ can be refined by any further event in $\sigma(\mathcal{F}_{ijs}, w_i, w_j)$ without affecting (8). This covers the aggregated distance used by the estimator, which averages period-specific distances over periods at which the pair shares the lagged outcome.

For a discordant pair the set of such periods is the same under both orderings of the swapped outcomes: periods $s+1$ and $s+2$ are excluded under both, because the lagged outcomes differ there, and every other period's lag lies outside the window. In part (b), the same refinement replaces $\delta_{ijs}^2 \leq h$ by a bound on the aggregated distance, since $\delta_{ijs}^2 \leq T\bar{\delta}_{ij}^2$ whenever period s contributes to the aggregate. Finally, the theorem places no restriction on the agent effects and does not require agents with the same network type to have the same value of λ: the comparison uses only the changes of the social influence term, which are common within network types by Assumption 2(i).

Lemma 3 (Minimum number of time periods)**.** *The identification strategy requires $T \geq 3$ observed periods (plus the initial condition y_{i0}), since it uses three consecutive periods $(s, s+1, s+2)$. The condition on period $s+2$ is needed because swapping outcomes at $(s, s+1)$ changes the lag entering the model at $s+2$. With additional time periods, several transition periods can be used, providing additional identifying variation. When $T = 2$ and $\alpha \neq 0$, the within-agent conditional likelihood based on the transition event $y_{i1} + y_{i2} = 1$ depends on the unknown function λ through a non-log-linear term that cannot be eliminated by the conditioning used in this paper. Consequently, the parameters (β, α) are not identified* by this strategy *at $T = 2$. The lemma does not rule out identification through other functionals of the data. (When $\alpha = 0$ is known a priori, the offending term equals one, and β is identified at $T = 2$ by the cross-agent comparison of Lemma 2 applied to the two-period window, which needs neither period $s+2$ nor Assumption 2(ii). The impossibility is specific to the dynamic parameter.)*

Proof. With $T = 2$, the only paths satisfying $y_{i1} + y_{i2} = 1$ are $A_1 = (y_{i0}, 0, 1)$ and $B_1 = (y_{i0}, 1, 0)$. Define $c_{i1} = X_{i1}'\beta + \alpha y_{i0} + \lambda(w_{i1}) + \mu_i$ and $c_{i2} = X_{i2}'\beta + \lambda(w_{i2}) + \mu_i$. Conditional on χ_i, the likelihoods are:

$$\mathbb{P}(y_i \in A_1 \mid \chi_i) = \big[1 - F(c_{i1})\big] \cdot F(c_{i2}) = \frac{e^{c_{i2}}}{(1+e^{c_{i1}})(1+e^{c_{i2}})},$$
$$\mathbb{P}(y_i \in B_1 \mid \chi_i) = F(c_{i1}) \cdot \big[1 - F(c_{i2}+\alpha)\big] = \frac{e^{c_{i1}}}{(1+e^{c_{i1}})(1+e^{c_{i2}+\alpha})},$$

where the index at period 2 in path B_1 is $c_{i2} + \alpha$ because the lag $y_{i1} = 1$ adds α to the linear index. The conditional probability is:

$$\mathbb{P}(y_i \in A_1 \mid y_{i1} + y_{i2} = 1, \chi_i) = \frac{e^{c_{i2}}(1+e^{c_{i2}+\alpha})}{e^{c_{i2}}(1+e^{c_{i2}+\alpha}) + e^{c_{i1}}(1+e^{c_{i2}})}.$$

When $\alpha \neq 0$, this does not reduce to the logistic form $F(V)$ for any index V that is free of λ and μ_i: the term $(1+e^{c_{i2}+\alpha})/(1+e^{c_{i2}})$, which depends on $\lambda(w_{i2})$ and μ_i through c_{i2},

enters the conditional probability in a non-log-linear way. This is the same obstacle that Honoré and Kyriazidou (2000) face with $T = 2$. (At $\alpha = 0$ the term equals one and the expression reduces to $F(c_{i2} - c_{i1})$, in which μ_i has already canceled and the remaining term $\lambda(w_{i2}) - \lambda(w_{i1}) = d_1(w_{i1})$ is removed by matching on network type as in Lemma 2, consistent with the caveat in the statement.) When $T \geq 3$, period $s+2$ contributes the reciprocal factor $(1 + e^c)/(1 + e^{c+\alpha})$ under the stability condition, canceling this term exactly (cf. the proof of Lemma 1). With $T = 2$, no such period exists, and the nuisance function λ remains in the conditional likelihood. □

The identification strategy is closely related to Honoré and Kyriazidou (2000), whose path-swapping argument eliminates the individual fixed effect μ_i and, with it, the level of λ. The difference is the use of network-type matching to handle the change in $\lambda(w_{it})$, a time-varying component of the unobserved heterogeneity that within-agent techniques alone cannot eliminate.

Given identification of (β, α), what can be learned about the social influence function is its movement, not its level, and only on network classes that the observed network can recover. For an agent j and a transition period s, write $\mathcal{T}_{js} = \{y_{js} + y_{j,s+1} = 1\}$ for the event that the agent switches between s and $s+1$, and $Z_{js} = \big(X_{js} - X_{j,s+1},\ y_{j,s-1} - y_{j,s+2}\big)'$ for the within-agent regressor, so that the pair regressor of Theorem 1 is $Z_{ijs} = Z_{is} - Z_{js}$. Fix a lag stratum y and a reference agent i with $y_{i,s-1} = y$. The agents that link like i at period s are identified through the codegree distance, not through the full link profile: by Proposition 1(a), δ_{ijs} is identified in the growing-network experiment for every j with $y_{j,s-1} = y$, whereas the profile $f_{w_{js};\,y,\cdot}$ can differ from that of i at partner types that are absent from the population, where the network carries no information. Define the observable class of i and its neighborhoods,

$$\mathcal{M}_{is} = \big\{j \neq i : y_{j,s-1} = y,\ \delta_{ijs} = 0\big\}, \qquad \mathcal{M}_{is}(h) = \big\{j \neq i : y_{j,s-1} = y,\ \delta^2_{ijs} \leq h\big\},$$

and the stable-switch events $\mathcal{E}_{js} = \mathcal{T}_{js} \cap \{X_{j,s+2} = X_{j,s+1}\}$ and $\mathcal{E}_{js}(h) = \mathcal{T}_{js} \cap \{\|X_{j,s+2} - X_{j,s+1}\|^2 \leq h\}$. Probabilities and expectations below are taken over a second agent j, which is independent of i by Assumption 1(i), given the primitives χ_i of the reference agent. They depend on i only through (w_{is}, y), and, because membership in $\mathcal{M}_{is}$ or $\mathcal{M}_{is}(h)$ is decided by the identified distance δ_{ijs}, they are functionals of the joint law of the observed data and of that distance.

Corollary 1 (Identification of the common movement)**.** *Suppose Assumptions 1 and 2 hold and (β, α) is known. Fix $s \in \mathcal{S}$, a lag stratum y, and a reference agent i with $y_{i,s-1} = y$.*

(a) (Finite support.) Suppose Assumption 7(i)–(ii) holds, in which case the continuity requirement in Assumption 2(i) is not needed, and $\mathbb{P}(j \in \mathcal{M}_{is},\ \mathcal{E}_{js} \mid \chi_i) > 0$. *Then* $d_s(w_{is})$ *is the unique solution* d *of*

$$\mathbb{E}\big[\mathbb{1}\{y_j \in B_s\} - F(Z'_{js}\theta - d) \,\big|\, j \in \mathcal{M}_{is},\ \mathcal{E}_{js},\ \chi_i\big] = 0, \tag{10}$$

and, for every value $(x, \bar{y})$ *of* $(X_j, y_{j,-s})$ *that has positive probability given* $\{j \in \mathcal{M}_{is}\} \cap \mathcal{E}_{js}$, $d_s(w_{is}) = z'\theta - F^{-1}\big(\mathbb{P}(y_j \in B_s \mid j \in \mathcal{M}_{is},\ \mathcal{E}_{js},\ X_j = x,\ y_{j,-s} = \bar{y},\ \chi_i)\big)$, *where* z *is the value of* Z_{js} *at* $(x, \bar{y})$.

(b) (Continuous types.) Suppose Assumption 6 holds and $\mathbb{P}(j \in \mathcal{M}_{is}(h),\ \mathcal{E}_{js}(h) \mid \chi_i) > 0$ *for every* $h > 0$. *Then, for every* $h > 0$, *the equation*

$$\mathbb{E}\big[\mathbb{1}\{y_j \in B_s\} - F(Z'_{js}\theta - d) \,\big|\, j \in \mathcal{M}_{is}(h),\ \mathcal{E}_{js}(h),\ \chi_i\big] = 0 \tag{11}$$

has a unique solution d_h, *and* $|d_h - d_s(w_{is})| \le \varpi_1(h)$, *with* ϖ_1 *as in Theorem 1(b). Hence* $d_s(w_{is}) = \lim_{h \downarrow 0} d_h$ *is identified.*

No recovery claim is made for reference agents whose class does not satisfy the stated overlap condition, for which equations (10) *and* (11) *have no observable counterpart. A supported type whose covariates never remain stable between* $s+1$ *and* $s+2$ *is an example. The level of* λ *is not identified: the outcome distribution is unchanged if* λ *is replaced by* $\lambda + c$ *and* μ_i *by* $\mu_i - c$ *for any constant* c, *and, when* $w_{it} = w_i$ *is time-invariant, only the sum* $\lambda(w_i) + \mu_i$ *enters the outcomes and no decomposition of it is identified.*

Proof. (a) The proof has two steps. The first justifies conditioning on the observable class. By Proposition 1(a), the distance δ_{ijs} is identified, so the conditional law of (y_j, X_j) given $\{j \in \mathcal{M}_{is}\} \cap \mathcal{E}_{js}$ is a functional of the data and of that distance. By Proposition 1(c), $j \in \mathcal{M}_{is}$ if and only if w_{js} is equivalent on the support to w_{is} in stratum y, and Assumption 7(ii) then gives $d_s(w_{js}) = d_s(w_{is})$: the movement is common on the observable class. The second step is the within-agent inversion. Condition on χ_i, χ_j, $y_{j,-s}$, and $\mathcal{E}_{js}$. Since (χ_j, ε_j) is independent of (χ_i, ε_i), the conditional probability of B_s depends only on $(\chi_j, y_{j,-s})$. On $\mathcal{T}_{js}$ the path of j is A_s or B_s, and the stability conditions of Lemma 1 hold by $X_{j,s+2} = X_{j,s+1}$ and Assumption 2(ii), so Lemma 1 and Assumption 2(i) give $\mathbb{P}(y_j \in B_s \mid \chi_i, \chi_j, y_{j,-s}, \mathcal{E}_{js}) = F\big(Z'_{js}\theta - d_s(w_{js})\big)$, which does not depend on μ_j. On $\{j \in \mathcal{M}_{is}\}$ this equals $F(Z'_{js}\theta - d_s(w_{is}))$. The event $\{j \in \mathcal{M}_{is}\}$ is determined by $(w_{is}, w_{js}, y_{j,s-1})$ and Z_{js} is a function of $(X_j, y_{j,-s})$, so iterated expectations over the coarser conditioning sets give (10) at $d = d_s(w_{is})$ and the pointwise formula, whose conditioning event has positive probability. The left side of (10) is strictly

increasing in d, with derivative $\mathbb{E}[F'(Z'_{js}\theta - d) \mid \cdot] > 0$, so the solution is unique. (b) is proved in Appendix A.4. The non-identification statements follow because $(\lambda + c, \mu_i - c)$ generates the same outcomes as (λ, μ_i) and because μ_i may be an arbitrary function of w_i under Assumption 1(i). □

The observable-class premise cannot be dropped. Suppose that the latent index takes only the values 0 and 1, each with positive probability, that each agent's index is reflected between waves 1 and 2, $w_{i2} = 1 - w_{i1}$ with $w_{i0} = w_{i1}$, and is constant afterwards, so that $\mathcal{S} = \{1\}$ with $T = 3$, that $\lambda(u) = u$, and that links are lag-independent with $f(u, v) = \frac{1}{2} + \frac{1}{4}\big[uv(1 - v) + vu(1-u)\big]$. The movements are $d_1(0) = 1$ and $d_1(1) = -1$, and $d_2 \equiv 0$. The full link profiles of the two indices differ, since $f(0, \cdot) \equiv \frac{1}{2}$ while $f(1, v) = \frac{1}{2} + \frac{1}{4}v(1 - v)$, and because $u \mapsto f(u, \cdot)$ is continuous and injective on the compact set $[0, 1]$, its inverse is uniformly continuous, so the movement $d_1(u) = 1 - 2u$ satisfies the continuity condition of Assumption 2(i). Explicitly, $f(u, v) - f(u', v) = \frac{1}{4}v(u - u')(2 - u - u' - v)$ gives $\|f_{u;\,y,\cdot} - f_{u';\,y,\cdot}\|_2 \geq |u - u'|/\sqrt{640}$. But every partner in the population has index 0 or 1, and between such partners the link probability is always $\frac{1}{2}$, so the codegree distance between the two indices is zero. Conditioning on the full profile would separate agents that move by $+1$ from agents that move by -1, but the network alone cannot distinguish these full-profile classes, so codegree-based conditioning cannot separate agents whose latent terms move by different amounts in this example. The example does not show that no other observed information could: the stated setup does not exclude, for instance, an observed initial covariate that reveals the index. Part (a) excludes the example through Assumption 7(ii), which requires the movement to be constant on classes of indices that are equivalent on the support, and part (b) excludes it through the full-support condition of Assumption 6.

The corollary is the within-agent counterpart of Theorem 1: among switchers with stable covariates in an observable network class, the odds of switching down rather than up reveal the common movement of the social influence term, which is what the within-agent estimator of Honoré and Kyriazidou (2000) omits. The corollary also explains why inverting the conditional choice probability given the covariates, the lag, and the network type does not recover λ here: that probability averages $F(X'_{it}\beta + \alpha y_{i,t-1} + \lambda(w_{it}) + \mu_i)$ over the conditional distribution of μ_i, and its inversion does not return λ. Appendix B records these points and describes an estimator of d_s, for which no asymptotic theory is claimed.

5 Estimation

This section develops a feasible estimator that implements the identification strategy of Section 4. The main challenge is that the conditioning events, network-type equivalence ($\rho_{ijt} = 0$) and covariate stability ($X_{i,s+2} = X_{i,s+1}$), hold with probability zero in continuous distributions, and the network distance ρ_{ijt} is not directly observable.

5.1 Estimating Network Proximity

Following Auerbach (2022) and Gueyap Kounga (2023), I use codegree information to construct a feasible measure of network similarity. In the static setting, the codegree function directly proxies the type distance. In the dynamic setting, agent i's type at period t consists of the index of unobserved social characteristics w_{it} and the observed lagged outcome y_{it-1}. Since the lagged outcome is observed, the codegree can be conditioned on it to isolate the unobserved component.

Define the conditional codegree function for two agents with social characteristics (w_{it}, w_{jt}) and lagged outcomes (y_{it-1}, y_{jt-1}) as:

$$p(w_{it}, w_{jt}; y_{it-1}, y_{jt-1}) = \int f(w_{it}, u;\, y_{it-1}, y)\, f(w_{jt}, u;\, y_{jt-1}, y)\, dG_t(u, y),$$

where G_t denotes the joint population distribution of (w_{lt}, y_{lt-1}) at period t. This distribution may change over time. The period dependence of the codegree function p is left implicit in the notation. This function measures the expected fraction of common neighbors between an agent of type (w_{it}, y_{it-1}) and an agent of type (w_{jt}, y_{jt-1}): the integration runs over all possible third agents with type $u \in [0, 1]$ and lagged outcome $y \in \{0, 1\}$.

For each agent i at period t, define the *conditional individual codegree function* as:

$$p_{w_{it};\, y_{it-1}, y_{jt-1}}(\cdot) \equiv p(w_{it}, \cdot; y_{it-1}, y_{jt-1}) : [0, 1] \to [0, 1].$$

This is the codegree profile of agent i: for each potential partner type w_{jt} and partner lagged outcome y_{jt-1}, it gives the expected fraction of common neighbors between agents i and j.

The *conditional codegree distance* for agents i and j with $y_{it-1} = y_{jt-1}$ at period t is the $L^2(G_t)$ distance between their codegree profiles, with the partner's type and lagged outcome (s, y') integrated against their joint population distribution G_t:

$$\delta_{ijt} = \left\| p_{w_{it};\, y_{it-1}, \cdot} - p_{w_{jt};\, y_{jt-1}, \cdot} \right\|_{L^2(G_t)} = \left[\int \left(p_{w_{it};\, y_{it-1}, y'}(s) - p_{w_{jt};\, y_{jt-1}, y'}(s) \right)^2 dG_t(s, y') \right]^{1/2}.$$

The G_t-weighting is deliberate: it is exactly the metric in which the empirical codegree statistic of equation (12) below converges (Lemma 4(a) establishes $\hat{\delta}^2_{ijt} \to_p \delta^2_{ijt} = \mathbb{E}_k[(p_{ikt} - p_{jkt})^2]$, an average over third agents drawn from the population, which is the display above), so a single object is both the population target and the probability limit. An equal-stratum-weight variant, summing over $y' \in \{0, 1\}$ with Lebesgue ds, has the same zero set and defines an equivalent topology when G_t has full support and f is continuous (Assumption 6), but not in general: when the population has atoms or lacks support at some partner types, the two metrics can differ in what they distinguish. The variant is in any case not the limit of the sample statistic, and every assumption stated in terms of the codegree distance uses the G_t-weighted metric. Since $y_{it-1} = y_{jt-1}$, the distance depends only on the type difference w_{it} versus w_{jt}. Equivalently, substituting the definition of p:

$$\delta_{ijt} = \left[\int \left(\int \left[f(w_{it}, u;\, y_{it-1}, y) - f(w_{jt}, u;\, y_{it-1}, y) \right] f(s, u;\, y', y)\, dG_t(u, y) \right)^2 dG_t(s, y') \right]^{1/2}.$$

Since the codegree function is obtained by integrating the individual link function against $f(s, u;\, y', y)\, dG_t(u, y)$, conditionally network-type equivalent agents ($\rho_{ijt} = 0$) share the same codegree profile when f is continuous, so $\rho_{ijt} = 0$ then implies $\delta_{ijt} = 0$ (Lemma 4(b)). Without continuity, the implication can fail at atoms of the type distribution, since ρ integrates against Lebesgue measure and δ against G_t. The codegree distance therefore provides a feasible proxy for network-type equivalence: under the continuity and full-support conditions of Assumption 6, Lemma 4(c) establishes the converse separation, so that pairs selected by $\delta_{ijt} \approx 0$ have $\rho_{ijt} \approx 0$ and hence, by Assumption 2(i), approximately equal changes of the social influence term between t and $t + 1$, which is what the identification argument requires.

The codegree distance can be consistently estimated from the observed adjacency matrices. For each period t, define the empirical codegree distance:

$$\hat{\delta}_{ijt} = \left[\frac{1}{n} \sum_{k=1}^{n} \left(\frac{1}{n} \sum_{l=1}^{n} D_{klt}(D_{ilt} - D_{jlt}) \right)^2 \right]^{1/2}. \tag{12}$$

When $y_{it-1} = y_{jt-1} = y$, the difference $D_{ilt} - D_{jlt}$ is driven exclusively by the type difference w_{it} versus w_{jt}: for any third agent l with $y_{lt-1} = y'$, $\mathbb{E}[D_{ilt} - D_{jlt} \mid w, y] = f(w_{it}, w_{lt};\, y, y') - f(w_{jt}, w_{lt};\, y, y') = f_{w_{it};\, y, y'}(w_{lt}) - f_{w_{jt};\, y, y'}(w_{lt})$, which is zero when $\rho_{ijt} = 0$. When $y_{it-1} \neq y_{jt-1}$, the conditional individual link functions of the two agents are evaluated at different own lagged outcomes, so the codegree difference conflates type differences with the lagged-outcome difference and cannot reliably proxy the type distance. Accordingly, the estimator restricts

attention to pairs sharing the same lagged outcome at the relevant period.

When f does not depend on lagged outcomes, the codegree difference between two agents is driven by their type difference whatever their lagged outcomes, so the shared-lag restriction is unnecessary for interpreting the distance in that special case, and the codegree reduces to the codegree of the static case. The estimator defined below nevertheless retains the restriction, and equation (13) still excludes the periods at which the lagged outcomes differ.

For a discordant pair, the stratification discards the distance at $s+1$: $y_{is} \neq y_{js}$ by construction, so at $s+1$ (and again at $s+2$) the codegree difference conflates the type difference with the lagged-outcome difference and cannot proxy the type distance. This is why Theorem 1 certifies equivalence only at period s, where the pair shares the lagged outcome, and lets Assumption 2(i) deliver the equal change of the social influence term between s and $s+1$. Other periods at which the pair shares the lagged outcome can be added to the certification: they do not change the identity of Theorem 1 (third remark after the theorem), and they reduce the noise in the estimated distance.

Accordingly, I define the *aggregated lag-stratified codegree distance*

$$\hat{\bar{\delta}}^2_{ij} \;=\; \frac{\sum_{t=1}^{T} \mathbb{1}\{y_{i,t-1} = y_{j,t-1}\}\, \hat{\delta}^2_{ijt}}{\max\big(1, \sum_{t=1}^{T} \mathbb{1}\{y_{i,t-1} = y_{j,t-1}\}\big)}, \tag{13}$$

with the convention $\hat{\bar{\delta}}^2_{ij} = +\infty$ when the two agents never share a lagged outcome. Each period contributes only when the two agents share the same lagged outcome, so that every term isolates the latent type difference, and the average over periods requires the pair to display similar linking behavior throughout the panel rather than at a single wave. A small value of $\hat{\bar{\delta}}^2_{ij}$ is therefore evidence that $\rho_{ijt} \approx 0$ at every informative period, including period s, which is the observable event of Theorem 1. With $T = 3$ and $s = 1$, the only informative period for a discordant pair is $t = 1$, and the aggregated distance reduces to $\hat{\delta}^2_{ij1}$.

5.2 Feasible Estimator

The feasible estimator is a kernel-weighted conditional maximum likelihood estimator. For each transition period $s \in \mathcal{S}$, define the set of "discordant" agent pairs:

$$\mathcal{P}_s = \{(i,j) : i < j,\ y_{i,s-1} = y_{j,s-1},\ y_{is} + y_{i,s+1} = 1,\ y_{js} + y_{j,s+1} = 1,\ y_{is} \neq y_{js}\}\,.$$

These are pairs where both agents share the same lagged outcome at period $s-1$ and transition between 0 and 1 across periods $(s, s+1)$, but in different orders. The condition $y_{i,s-1} = y_{j,s-1}$ ensures that the empirical codegree distance $\hat{\delta}_{ijs}$ is a valid proxy for network-type equivalence

at period s.

For each such pair and period, define the conditioning variable:

$$v_{ijs}(b,a) = (\Delta_s X_i - \Delta_s X_j)'b + a(y_{i,s-1} - y_{i,s+2} - y_{j,s-1} + y_{j,s+2}),$$

where $\Delta_s X_i = X_{is} - X_{i,s+1}$.

The sample objective function is:

$$\Omega_n(b,a) = \sum_{s\in\mathcal{S}} \sum_{(i,j)\in\mathcal{P}_s} K_1\left(\frac{\hat{\bar{\delta}}_{ij}^2}{h_1}\right) K_2\left(\frac{\|X_{i,s+2} - X_{i,s+1}\|^2}{h_2}\right) K_2\left(\frac{\|X_{j,s+2} - X_{j,s+1}\|^2}{h_2}\right) \cdot m_{ijs}(b,a), \tag{14}$$

where:

$$m_{ijs}(b,a) = \mathbb{1}(y_{is}=0)\log F\left[v_{jis}(b,a)\right] + \mathbb{1}(y_{is}=1)\log F\left[v_{ijs}(b,a)\right].$$

The kernel K_1 weights agent pairs by their aggregated codegree similarity. By Assumption 2(i) and Proposition 1(b)–(c), pairs with small aggregated codegree distance have social influence terms that change by approximately the same amount between s and $s+1$, which is the cross-agent condition of Lemma 2. The kernel K_2 weights observations by the local stability of covariates between periods $s+1$ and $s+2$. It approximates the condition $X_{i,s+2} = X_{i,s+1}$ of Lemma 1, while Assumption 2(ii) supplies the within-agent stability condition $\lambda(w_{i,s+2}) = \lambda(w_{i,s+1})$. The bandwidths h_1 and h_2 are sequences converging to zero as $n \to \infty$, except in finite-type designs with discrete covariates, where fixed bandwidths small enough relative to the separation gaps select exact matches (Assumption 7(v) in Section 6).

The estimator is:

$$(\hat{\beta}, \hat{\alpha}) = \arg\max_{(b,a)} \Omega_n(b,a). \tag{15}$$

Throughout the rest of the paper I write $\theta = (\beta', \alpha)'$ for the parameter vector and $\hat{\theta} = (\hat{\beta}', \hat{\alpha})'$ for the estimator in (15), so that Θ is the parameter space for θ and (b,a) denotes a generic candidate value.

The asymptotic results maximize over a compact parameter space Θ (Assumption 5), where a maximizer always exists. The simulations and the application instead seek an unconstrained maximum and record a separated likelihood, which has none, as a failed fit. Separation in a finite sample does not conflict with the theory, which concerns the constrained maximizer as n grows. The Firth-penalized fits of Section 8 maximize a different criterion and are not covered by Theorems 2 and 3.

The kernel K_2 is necessitated by the dynamic structure, and it serves only the covariate condition of Lemma 1: the condition $\lambda(w_{i,s+2}) = \lambda(w_{i,s+1})$ comes from Assumption 2(ii), and

it holds automatically when $w_{it} = w_i$ is time-invariant. Because K_2 localizes the squared change $\|X_{i,s+2} - X_{i,s+1}\|^2$, the radius of the stability window in covariate units is $\sqrt{h_2}$, and with several covariates the squared norm adds changes measured in different units, so the covariates should be put on a common scale before h_2 is chosen. With integer-valued covariates, as in the simulations, $h_2 < 1$ selects exact stability.

The empirical codegree distances concentrate around their population values, whose scale is design dependent and can be numerically small (the smallest positive population distance in the baseline design of Section 7 is 3.3×10^{-3}), so a bandwidth h_1 chosen as a fixed function of n alone can exceed the largest distance in the sample, in which case every pair receives weight and no matching occurs. In sparse networks the same formula can leave too few pairs with positive weight. A scale-adaptive choice is therefore useful as a sensitivity diagnostic: set h_1 to a low quantile, say the fifth or tenth percentile, of the empirical distribution of $\{\hat{\delta}^2_{ij}\}_{i<j}$. The quantile targets a share of the candidate distances, subject to ties. It does not guarantee a share of pairs that survive the discordance, shared-lag, and covariate-stability restrictions, which is why the diagnostics below should be reported with it. It is a diagnostic device rather than part of the theory: a fixed quantile of the estimated distances converges to a positive threshold in continuous designs, so it does not shrink, and in finite-type designs a quantile that falls inside the atom of exact matches is driven by estimation noise. Theorems 2 and 3 are stated for deterministic bandwidth sequences, and the main simulations of Section 7 use such sequences, calibrated with knowledge of the design that is not available in practice. In the sparse design of that section, the quantile rule selects mostly wrong matches.

With $T = 3$, the following rule uses only the data, and in finite-type designs Theorem 3 and Proposition 2 continue to hold with it (Corollary 2). Let S_n be the largest $\hat{\delta}^2_{ij1}$ over the pairs $i < j \leq \lceil \sqrt{n} \rceil$ with $y_{i0} = y_{j0}$, with $S_n = 0$ if there is none, and set

$$\hat{h}_1 = 0.1 \max\{0.001,\ S_n\} (n/500)^{-1/4}. \tag{16}$$

In a finite-type design, S_n converges in probability to the largest population distance between two types with the same initial outcome, so $\hat{h}_1$ scales with the distances of the design and shrinks at the rate $n^{-1/4}$. The floor keeps it positive. The rule uses neither the types nor the separation gap, and its constants are not optimized. It depends only on the first-wave network and the initial outcomes, which preserves the exactness of the class likelihood in Proposition 2(b), whereas a bandwidth chosen with later networks or with the orderings need not. Nothing ensures that $\hat{h}_1$ is below half the separation gap in a given sample, and Section 7.4 shows when it is.

Three diagnostics should always be reported alongside the estimates: the total number of

pairs receiving positive weight; the kernel-weighted effective sample size $(\sum_p w_p)^2 / \sum_p w_p^2$ over the weighted pairs p, which is a measure of the concentration of the weights (it equals the number of weighted pairs when the weights are equal and is smaller otherwise) and not a count of independent agents, since pairs sharing an agent are dependent; and, over the set of eligible agents, those that switch between s and $s+1$ and have covariates within the stability window, the number of agents that contribute at least one weighted pair, the number left with no partner, and the minimum number of weighted partners among the contributing agents, $\min_{i \in \mathcal{C}} \sum_{j \neq i} \mathbb{1}\{\text{pair } (i,j) \text{ receives positive weight}\}$, where $\mathcal{C}$ is the set of contributing agents. The unmatched count says whether any eligible agent is excluded, and the minimum says how thinly the agents who do enter are matched. Agents who do not switch have no partners by construction and belong to neither count. Table 3 reports all of these.

5.3 Conditioning on Network Classes

The pairwise estimator uses the network one pair at a time: each weighted pair contributes the conditional probability of Lemma 2, whether or not its two agents also belong to a larger group of agents with the same network type. When types take finitely many values, the network certifies more than pairs. Among the agents who switch between s and $s+1$ and have stable covariates, those who share the lagged outcome at $s-1$ and are at zero codegree distance from one another form a class whose members share the movement d_s, and, by the within-agent identity in the proof of Corollary 1, the ordering of each member's switch is a logit with the same intercept for the whole class. A logit with a group-specific intercept admits a sufficient statistic for that intercept, the number of members with a given outcome, and conditioning on it removes the intercept, as conditioning on the number of periods with a positive outcome removes the individual effect in the conditional logit of Andersen (1970) and Chamberlain (1980). Here the class plays the role of the individual, and its members the role of the periods. The resulting likelihood is the conditional logistic regression for matched sets with several cases of Breslow and Day (1980), with network classes as the matched sets.

Formally, fix $s \in \mathcal{S}$ and let

$$\mathcal{E}_s = \left\{ j : y_{js} + y_{j,s+1} = 1,\ X_{j,s+2} = X_{j,s+1} \right\}$$

be the set of eligible agents, who switch between s and $s+1$ and have stable covariates between $s+1$ and $s+2$. With discrete covariates, this is the set that the covariate kernel selects at any bandwidth below $\bar{\epsilon}_X$ (Assumption 7(iv)). For $j \in \mathcal{E}_s$, let $b_{js} = \mathbb{1}\{y_{js} = 1\}$ indicate the ordering B_s, and let Z_{js} be the within-agent regressor defined before Corollary 1.

The feasible classes are the connected components of the graph on $\mathcal{E}_s$ that joins i and j when

$$y_{i,s-1} = y_{j,s-1} \quad \text{and} \quad \hat{\delta}^2_{ijs} < h_1, \tag{17}$$

with the network bandwidth of the pairwise estimator. The distance is that of period s: two members of a class may share the lagged outcome at $s+1$, and the networks of periods $s+1$ and $s+2$ depend on the outcomes whose ordering is modeled, so those periods cannot enter the construction. Other periods at which the two agents share the lagged outcome, before s or after $s+2$, could be averaged in, as in (13), and with $T = 3$ there are none. For a class C with $N_C = \sum_{j \in C} b_{js}$ members of ordering B_s, the class-conditional log likelihood is

$$\ell_C(\theta) = \sum_{j \in C} b_{js}\, Z'_{js}\theta - \log \sum_{S \subseteq C:\ |S| = N_C} \exp\Big(\sum_{j \in S} Z'_{js}\theta \Big), \tag{18}$$

the logarithm of the conditional probability of the observed orderings given N_C when they are independent logits with a common intercept. A class in which every member has the same ordering contributes zero. The class-conditional estimator is

$$\hat{\theta}^{\mathrm{C}} = \arg\max_{\theta \in \Theta} \sum_{s \in \mathcal{S}} \sum_{C} \ell_C(\theta), \tag{19}$$

where the inner sum runs over the feasible classes at s. The sum inside the logarithm in (18) is the elementary symmetric polynomial of order N_C in the values $\exp(Z'_{js}\theta)$, $j \in C$. This polynomial and its first two derivatives are computed by a recursion over the members of the class (Gail et al., 1981), as in standard software for conditional logistic regression, at a cost proportional to $|C|\, N_C$ rather than to the number of subsets. The objective is concave, since each ℓ_C is the log likelihood of an exponential family in θ.

Three features distinguish the two estimators. First, for a class with one member of each ordering, (18) equals $\log F\big((Z_{js} - Z_{is})'\theta\big)$, with j the member of ordering B_s, which is the pair's contribution to the pairwise objective (14) at unit weight. For larger classes, the pairwise estimator replaces the conditional probability of the configuration of the whole class by the product of the conditional probabilities of its discordant pairs, which makes it a pairwise, or composite, version of the class likelihood, in the way that a conditional logit built from pairs of periods is a pairwise version of the conditional logit of Chamberlain (1980). Pairs that share an agent carry overlapping information, whereas the class likelihood uses each class's sufficient statistic once, and Remark 2 argues that, in finite-type designs with $T = 3$, the class estimator is at least as efficient asymptotically. Second, the class construction is transitive and therefore more sensitive to an erroneous edge: a single pair of

agents of different types whose estimated distance falls below the bandwidth merges their two classes, whereas in the pairwise estimator it affects one pair. A feasible class that splits a population class, by contrast, leaves the likelihood correctly specified and costs only efficiency (Proposition 2(b)). Third, the class estimator needs classes: exact covariate stability and a network whose types take finitely many values, separated by a gap that the bandwidth respects. In continuous designs the connected components of the thresholded graph chain together agents whose types, and movements, differ by much more than the bandwidth, so the construction does not deliver groups with a common movement, whereas the pairwise weights bound the distance of every compared pair. The pairwise estimator is therefore the one that the consistency theory of Section 6 covers in general designs, and the class estimator is an alternative for finite-type designs, where it is more precise in the simulations of Section 7. In practice the diagnostics guide the choice: a visible gap in the estimated distances and a partition that is stable across bandwidths favor the class estimator, without certifying that its classes are correct, since a merge of two classes can survive a change of bandwidth. Otherwise, the pairwise estimator is the safer choice. Estimating the class intercepts together with θ, by a logit with one dummy per class, is a third procedure. It is first-order equivalent to the class-conditional estimator in finite-type designs, as the proof of Proposition 2 shows, but it estimates one nuisance parameter per class from that class's members, and its small-sample bias is larger (Table 18).

The diagnostics recommended for the pairwise estimator should be reported for the class estimator as well, together with the number of feasible classes, their sizes, the number of eligible agents in classes that contain both orderings, and the sensitivity of the partition to the bandwidth, since the merging of two classes changes the likelihood.

6 Asymptotic Properties

This section establishes the large-sample properties of the estimator $(\hat{\beta}, \hat{\alpha})$ defined in (15). Throughout, T is treated as fixed and asymptotics are studied as $n \to \infty$. The results are of two kinds. Consistency holds in general designs, including continuous network types and covariates, with shrinking bandwidths. Asymptotic normality at the root-n rate, centered at the true value, is established for finite-type designs, in which network types take finitely many values and exact matches occur with positive probability. For continuous designs, the section explains what a distribution theory would require and why it is not established here. Section 6.5 gives the corresponding results for the class-conditional estimator of Section 5.3 in finite-type designs, and Corollary 2 extends Theorem 3 and Proposition 2 to the data-driven bandwidth (16). Proofs are collected in Appendix A.

6.1 Sampling as the Sample Grows

The estimation theory allows the idiosyncratic part of the movement of the latent index to be small but nonzero. Because the maximum over a growing sample of a variable with a fixed law cannot shrink to zero unless the variable is zero, this requires the law of the primitives to depend on n.

Assumption 3 (Drifting idiosyncratic movement)**.** *For each n, the primitives are drawn under a law $\mathbb{P}_n$ (a triangular array) that satisfies Assumption 1, with the functions λ and f, the distribution of the shocks, and the parameters (β, α) not depending on n. Under $\mathbb{P}_n$,*

$$\lambda(w_{i,t+1}) - \lambda(w_{it}) = d_t(w_{it}) + e_{it,n}, \qquad |e_{it,n}| \le \zeta_n \text{ almost surely}, \quad t \le T-1,$$

for a deterministic sequence $\zeta_n \to 0$, where the functions d_t do not depend on n, satisfy the uniform-continuity requirement of Assumption 2(i), and satisfy $d_{s+1} \equiv 0$ for $s \in \mathcal{S}$. For each t, the distribution $G_{t,n}$ of $(w_{it}, y_{i,t-1})$ under $\mathbb{P}_n$ converges weakly to a limit G_t.

Assumption 3 is an asymptotic device of the drifting-parameter kind. It describes a latent index whose idiosyncratic movement between adjacent waves is small relative to the dispersion of the index across agents, and it reduces to Assumption 2 under a fixed law when $\zeta_n \equiv 0$. Like the identification results, the estimation theory places no restriction on the agent effects, whose law under $\mathbb{P}_n$ may be arbitrary. The common component d_t is unrestricted in size. Because the idiosyncratic component changes the law of the latent index, the population codegree distances also depend on n. The weak convergence of $G_{t,n}$ makes them converge uniformly, since f is continuous and bounded (Assumption 6 below). In the finite-type theorem of Section 6.4, the law does not depend on n and $e_{it,n} \equiv 0$.

6.2 Kernels, Bandwidths, and Regularity

Throughout this section, $\bar{\delta}^2_{ij}$ denotes the population aggregated codegree distance, obtained from (13) by replacing each estimated per-period distance $\hat{\delta}^2_{ijt}$ with its population counterpart δ^2_{ijt} from Lemma 4. The hatted symbol $\hat{\bar{\delta}}^2_{ij}$ is the sample statistic. Under Assumption 3, the population distances are computed under $\mathbb{P}_n$. Let $\mathbb{1}_{ijs}$ be the indicator that the unordered pair $\{i, j\}$ belongs to $\mathcal{P}_s$ (it collects the shared lagged outcome at $s-1$, both switch conditions, and discordance), write $K_2^{(a)} = K_2(\|X_{a,s+2} - X_{a,s+1}\|^2/h_2)$ for $a \in \{i, j\}$, and define the kernel weight at the population distances, its enlarged-window counterpart, and the expected

weight mass,

$$W_{ijs} = K_1\left(\frac{\bar{\delta}_{ij}^2}{h_1}\right) K_2^{(i)} K_2^{(j)}, \qquad \bar{W}_{ijs} = \mathbb{1}\{\bar{\delta}_{ij}^2 < 2h_1\}\, K_2^{(i)} K_2^{(j)}, \qquad \omega_n = \sum_{s\in\mathcal{S}} \mathbb{E}_n\big[\mathbb{1}_{ijs} W_{ijs}\big],$$

where $\mathbb{E}_n$ is the expectation under $\mathbb{P}_n$ and (i, j) is a generic pair of distinct agents. The quantity $n\,\omega_n$ is, up to the normalization of the kernels, the expected number of effective partners of an agent.

Assumption 4 (Kernel and bandwidth conditions). *(i) The kernel functions K_1 and K_2 are non-negative, have support $[0, 1)$, satisfy $K_\ell(0) > 0$, and are Lipschitz on $[0, \infty)$.*

(ii) $h_1 \to 0$, $h_2 \to 0$, $nh_1^2/\log n \to \infty$, and $n\,\omega_n \to \infty$.

The Lipschitz condition on $[0, \infty)$ requires the kernels to vanish continuously at the boundary of their support. The Epanechnikov-type kernel $K(x) = \frac{3}{4}(1 - x^2)\mathbb{1}(0 \le x < 1)$ used in Sections 7 and 8 satisfies it. No differentiability is required, because the theory never expands the kernel weights beyond first order. The condition $nh_1^2/\log n \to \infty$ makes the uniform estimation error of the codegree distances, of order $(\log n/n)^{1/2}$ by Lemma 4(d), small relative to the network bandwidth. The condition $n\,\omega_n \to \infty$ requires the expected number of effective partners to grow. It combines the network match probability and the covariate-stability probability, and it reflects the cost of covariate-stability smoothing familiar from Honoré and Kyriazidou (2000): when covariates are continuously distributed, the stability probability vanishes with h_2. Because the codegree profile is an unknown function, the network match probability cannot be approximated by a polynomial in h_1 of known order, so the condition is stated directly. In practice it is monitored through the diagnostics recommended in Section 5, in particular the number of pairs with positive weight and the per-agent minimum match count.

Define the localized criterion and the localized information matrix of the actual selected population,

$$\begin{aligned} M_n(b, a) &= \frac{1}{\omega_n} \sum_{s\in\mathcal{S}} \mathbb{E}_n\big[\mathbb{1}_{ijs} W_{ijs}\, m_{ijs}(b, a)\big], \\ \Sigma_n &= \frac{1}{\omega_n} \sum_{s\in\mathcal{S}} \mathbb{E}_n\big[\mathbb{1}_{ijs} W_{ijs}\, F(Z_{ijs}'\theta)\big(1 - F(Z_{ijs}'\theta)\big)\, Z_{ijs} Z_{ijs}'\big]. \end{aligned} \tag{20}$$

Assumption 5 (Regularity conditions). *(i) $\mathbb{E}_n[\|X_{it}\|^4]$ is bounded uniformly in n for all t.*

(ii) The parameter space Θ is compact and convex and contains $\theta = (\beta', \alpha)'$ in its interior.

(iii) (Moments in the matching windows.) $\sup_n \omega_n^{-1} \sum_{s \in \mathcal{S}} \mathbb{E}_n\big[\mathbb{1}_{ijs} \bar{W}_{ijs} (1 + \|Z_{ijs}\|)^4\big] < \infty$.

(iv) (Rank on the selected population.) $\liminf_n \lambda_{\min}(\Sigma_n) > 0$, *where* $\lambda_{\min}$ *denotes the smallest eigenvalue.*

Part (iii) bounds the moments of the regressor within the matching windows, and, because $\bar{W}_{ijs}$ uses the window of twice the network bandwidth, it also requires the mass of pairs just outside the window to be comparable to the mass inside it. An unconditional moment condition does not deliver these requirements, because the windows shrink. Part (iv) is stated for the population that the estimator actually selects: the shared lagged outcome at $s-1$, discordance, the covariate localization, and the network localization are all built into $\mathbb{1}_{ijs} W_{ijs}$. A rank condition imposed on a larger population need not survive the additional restrictions: if $X_{i,s} - X_{i,s+1}$ were a function of $y_{i,s-1}$ alone, the covariate component of Z_{ijs} would vanish on every selected pair, since selected pairs share $y_{i,s-1}$, and the objective would be flat in β at every sample size, even though the second moment of Z_{ijs} could be positive definite on the larger population. The assumption does not require the localized criterion M_n to converge, nor does it equate its behavior with separately defined conditional expectations, which could differ across relative bandwidths at a probability-zero event. The proof shows that the correctly specified comparison criterion M_n^*, which is M_n with the index perturbation of the inexact matches removed, is maximized at the truth with a curvature bounded away from zero uniformly in n and that M_n is uniformly close to M_n^* as the bandwidths and the drift vanish. The argument works along subsequences. At fixed bandwidths M_n itself is generally maximized at a pseudo-true value that differs from θ (Remark 1).

Assumption 6 (Separation of network types)**.** *The link formation function f is continuous on $[0,1]^2 \times \{0,1\}^2$ and not constant, and for each t the distribution G_t of $(w_{it}, y_{i,t-1})$ (under Assumption 3, its limit) has full support on $[0,1] \times \{0,1\}$.*

Continuity makes the population codegree distance a continuous function of the latent indices, which the kernel localization argument uses, and non-constancy rules out a link function that is the same for every agent. Non-constancy does not by itself guarantee that the network is informative about the latent index (a link function that depends only on the observed lags is not constant but carries no information about w_{it}), which is the role of the full-support condition and of Lemma 4(c), while identifying variation in the outcomes is the role of the rank condition. The full-support condition is what makes the codegree distance separate agents that are not network-type equivalent. The codegree compares two agents through their links to third agents, and it can only see differences in link functions at partner types that are present in the population. For a continuous link function, the

separation is then automatic on the support: writing $g(\tau, y') = f(u, \tau;\, y, y') - f(v, \tau;\, y, y')$ and $(\mathcal{K}g)(\cdot) = \int f(\cdot, \tau;\, \cdot, y')\, g(\tau, y')\, dG_t(\tau, y')$ for the codegree transform, one has the identity

$$\int g^2\, dG_t = (\mathcal{K}g)(u, y) - (\mathcal{K}g)(v, y),$$

so if the codegree distance vanishes, which means that $\mathcal{K}g = 0$ almost everywhere and hence, by continuity, everywhere on the support of G_t, then $g = 0$ on that support (Lemma 4(c)). Full support converts agreement on the support into agreement of the link functions in the metric (3) used in Assumption 2(i), which integrates over all partner characteristics and both lag strata. Without full support, two agents whose link functions differ only at partner types or lag strata that are absent from the population cannot be distinguished by the codegree, and the relevant equivalence is agreement on the support. Finite-type designs replace this assumption by the directly verifiable separation condition of Assumption 7, which does not require continuity of f or full support.

6.3 Consistency

Theorem 2 (Consistency)**.** *Under Assumptions 3–6, the estimator* $(\hat{\beta}, \hat{\alpha})$ *defined in* (15) *is consistent:* $(\hat{\beta}, \hat{\alpha}) \xrightarrow{p} (\beta, \alpha)$.

Assumption 3 contains Assumption 1 for each n and Assumption 2 up to the vanishing idiosyncratic component. The proof, given in Appendix A, exploits the concavity of the logistic log-likelihood. Each contribution $m_{ijs}(b, a)$ is concave in (b, a) and the kernel weights are non-negative, so the sample objective is a random concave function on the convex set Θ, and the convexity lemma for extremum estimators (Newey and McFadden, 1994, Theorem 2.7), applied along subsequences, reduces the argument to pointwise convergence of the normalized objective to correctly specified criteria M_n^* that are maximized at (β, α) with non-vanishing curvature. The localized criterion M_n that the estimator targets is uniformly close to M_n^* as the bandwidths shrink. Three steps establish this. First, the estimated distances can be replaced by the population distances, because their uniform estimation error is small relative to the bandwidth. Second, the normalized objective at the population distances concentrates around $M_n(b, a)$, by a variance bound for a second-order U-statistic in the i.i.d. agent primitives, controlled by $n\omega_n \to \infty$. Third, on the pairs with positive weight, the true conditional probability of the discordant configuration is a logit in $Z_{ijs}'\theta$ up to an index perturbation bounded by three terms: the difference between the two agents' common movements d_s, which vanishes with h_1 by Assumptions 2(i) and 6; the covariate movement inside the stability window, of order $h_2^{1/2}$; and the idiosyncratic movement, of order ζ_n. The

localized criterion is therefore asymptotically that of a correctly specified logit on the selected population, whose information matrix is bounded away from singularity, so it is maximized at the truth with a curvature that does not vanish as n grows.

Remark 1 (Fixed bandwidths and the pseudo-true value)**.** *If (h_1, h_2) are held fixed as $n \to \infty$ under a fixed law, the same argument shows that $\hat{\theta}$ converges in probability to $\theta_h = \arg\max_{(b,a)\in\Theta} M_h(b, a)$, the maximizer of the localized criterion* (20) *at those bandwidths, provided that maximizer is unique. The pseudo-true value θ_h generally differs from θ in continuous designs, because matches at a positive bandwidth are inexact, and it converges to θ as the bandwidths shrink under the conditions above. Inexact matches do not always cause bias: when $w_{it} = w_i$ is time-invariant and the selected agents' covariates are exactly stable, the remainder in the proof of Lemma 1 and the common movement vanish for every pair, whatever their types, so $\theta_h = \theta$ at any fixed network bandwidth for which the selected rank condition holds. Fixed-quantile implementations in continuous designs estimate θ_h at the corresponding limiting threshold.*

6.4 Asymptotic Normality in Finite-Type Designs

The distribution theory covers designs in which exact matches occur with positive probability. Network types take finitely many values, as in stochastic block specifications with one latent value per block, and the covariates have mass points at exact stability. Survey data with covariates recorded on discrete scales fit this case naturally.

Write $\delta_t(u, v; y)$ for the population codegree distance at period t between agents with latent indices u and v and common lagged outcome y, and say that two supported indices u, v are *equivalent on the support* at period t in stratum y if $f(u, \tau; y, y') = f(v, \tau; y, y')$ for every (τ, y') in the support of G_t.

Assumption 7 (Finite-type design)**.** *(i) For each t, w_{it} takes values in a finite set $\mathcal{W}_t \subset [0, 1]$, each value with positive probability, so that the latent path w_i has finite support.*

(ii) (Equivalence on the support.) For each t and y, if two supported indices $u, v \in \mathcal{W}_t$ are equivalent on the support at period t in stratum y, then $d_t(u) = d_t(v)$.

(iii) (Exact dynamics.) The law of the primitives does not depend on n, so that Assumption 2 holds exactly and there is no idiosyncratic movement.

(iv) For each $s \in \mathcal{S}$, $\mathbb{P}(X_{i,s+2} = X_{i,s+1}) > 0$, and there is $\bar{\epsilon}_X > 0$ with $\mathbb{P}(0 < \|X_{i,s+2} - X_{i,s+1}\|^2 < \bar{\epsilon}_X) = 0$.

(v) (Bandwidths.) $h_2 \le \bar{\epsilon}_X$ *is fixed;* $h_1 \le \bar{\epsilon}/(2(T-2))$ *for all* n *large, where* $\bar{\epsilon}$ *is the smallest positive value of* $\delta_t(u,v;\,y)^2$ *over* t, y, *and supported* u, v; *and* $\sqrt{n}\,h_1/\log n \to \infty$.

(vi) (Rank.) With $\mathbb{1}^{\mathrm{ex}}_{ijs} = \mathbb{1}_{ijs}\mathbb{1}\{\bar{\delta}_{ij} = 0,\ X_{i,s+2} = X_{i,s+1},\ X_{j,s+2} = X_{j,s+1}\}$, *the matrix*

$$\Sigma_{\mathrm{ex}} = \sum_{s\in\mathcal{S}} \mathbb{E}\big[\mathbb{1}^{\mathrm{ex}}_{ijs}\, F(Z'_{ijs}\theta)\big(1 - F(Z'_{ijs}\theta)\big)\, Z_{ijs}Z'_{ijs}\big]$$

is positive definite.

Because the bandwidths in part (v) select exactly the pairs with $\mathbb{1}^{\mathrm{ex}}_{ijs} = 1$, the matrix Σ_n of Assumption 5 equals Σ_{ex} divided by $\sum_s \mathbb{P}(\mathbb{1}^{\mathrm{ex}}_{ijs} = 1)$ in this design, so part (vi) is the finite-type form of the rank condition in Assumption 5(iv). Part (ii) replaces the continuity requirement in Assumption 2(i) and Assumption 6. It is stated on the support because the codegree can only compare agents through partners who are present in the population. By the identity displayed after Assumption 6, which requires no continuity when the support is finite, the population codegree distance between two supported indices vanishes exactly when they are equivalent on the support (Lemma 4(c)). Since there are finitely many supported indices, the positive distances are bounded below by $\bar{\epsilon} > 0$, and this separation gap can be computed from the block matrix of link probabilities and the type probabilities. No continuity of f across block boundaries is needed, and only the common movement d_t, not λ itself, needs to be constant on classes of supported types. Because periods $s+1$ and $s+2$ never enter the aggregated distance of a discordant pair, that distance averages at most $T-2$ per-period distances, so a pair that is not equivalent at some contributing period has $\bar{\delta}^2_{ij} \ge \bar{\epsilon}/(T-2)$, which is why part (v) bounds h_1 by $\bar{\epsilon}/(2(T-2))$. With $T = 3$, the bound is $\bar{\epsilon}/2$. Parts (ii) and (iii) together make the exact-match event reproduce the conditions of Lemma 2: for a pair that shares the lagged outcome at $s-1$ and has $\bar{\delta}_{ij} = 0$, the two agents are equivalent on the support at s, hence have equal changes of the social influence term between s and $s+1$ by part (ii), and both are stable between $s+1$ and $s+2$ by Assumption 2(ii). Part (v) allows a fixed network bandwidth below the stated fraction of the separation gap, or one that shrinks slowly. Unlike an empirical quantile, such a bandwidth does not fall inside the atom of exact matches.

Let $\mathcal{Z}_i = (y_{i0}, \ldots, y_{iT}, X_i, w_i, \mu_i)$ denote agent i's complete node data, including the agent effect, and let

$$S_{ijs}(\theta) = \big\{\mathbb{1}(y_{is} = 1) - F(Z'_{ijs}\theta)\big\} Z_{ijs}$$

denote the score of m_{ijs} on the pairs in $\mathcal{P}_s$. The score is symmetric in (i,j).

Theorem 3 (Asymptotic normality, finite-type designs)**.** *Under Assumption 1, Assumption 2 with the continuity requirement in part (i) replaced by Assumption 7(ii), and Assumptions 4(i),*

5(i)–(ii), and 7,

$$\sqrt{n}\begin{pmatrix}\hat{\beta}-\beta\\ \hat{\alpha}-\alpha\end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0,\ 4\Sigma_{\text{ex}}^{-1}V_{\text{ex}}\Sigma_{\text{ex}}^{-1}\right), \qquad V_{\text{ex}}=\text{Var}(\psi_i), \qquad \psi_i=\mathbb{E}\Big[\sum_{s\in\mathcal{S}}\mathbb{1}_{ijs}^{\text{ex}}\,S_{ijs}(\theta)\,\Big|\,\mathcal{Z}_i\Big],$$

where the expectation defining ψ_i is over an independent agent j.

Three features of the finite-type regime deliver the result, and all three are established in the proof. First, with probability approaching one, the estimated weights are positive exactly on the pairs of the exact-match event, and on those pairs they differ from $K_1(0)K_2(0)^2$ by at most $O_p(\log n/(nh_1))$ uniformly, which is $o_p(n^{-1/2})$ under Assumption 7(v). The estimation of the codegree distances therefore does not affect the limiting distribution. Second, the estimator is thereby asymptotically equivalent to an oracle conditional logit whose objective is a second-order U-process in the i.i.d. node data $\mathcal{Z}_i$. The link shocks enter only through the estimated weights and drop out. Third, by Lemma 2 and Assumption 7(ii)–(iii), the exact-match score has mean zero given the conditioning information of that lemma, so the expected oracle criterion is maximized at the truth at every sample size and no bias term appears. The factor 4 in the variance is the familiar one for second-order U-statistics, and ψ_i sums the contributions of all transition periods before the variance is taken, so V_{ex} retains cross-period covariances.

The asymptotic variance is estimated by plug-in. Let $\hat{W}_{ijs}$ denote the estimated weight of pair (i,j) at s, $\bar{K}=K_1(0)K_2(0)^2$, and

$$\hat{\Sigma}=\binom{n}{2}^{-1}\sum_{s\in\mathcal{S}}\sum_{i<j}\frac{\hat{W}_{ijs}}{\bar{K}}\,\mathbb{1}_{ijs}\,F(Z_{ijs}'\hat{\theta})\big(1-F(Z_{ijs}'\hat{\theta})\big)Z_{ijs}Z_{ijs}',$$

$$\hat{\psi}_i=\frac{1}{n-1}\sum_{j\neq i}\sum_{s\in\mathcal{S}}\frac{\hat{W}_{ijs}}{\bar{K}}\,\mathbb{1}_{ijs}\,S_{ijs}(\hat{\theta}),$$

with $\hat{V}$ the sample variance of $\hat{\psi}_i$. Then $4\hat{\Sigma}^{-1}\hat{V}\hat{\Sigma}^{-1}/n$ estimates the variance of $\hat{\theta}$, and consistency of $\hat{\Sigma}$ and $\hat{V}$ for Σ_{ex} and V_{ex} follows from the oracle equivalence established in the proof of Theorem 3 and the law of large numbers for U-statistics. This estimator coincides, up to finite-sample normalization, with the sandwich that clusters the pair-level scores by agent. When observations are grouped, the agent-level terms $\hat{\psi}_i$ can be summed within groups before forming the variance, yielding a cluster-robust version. That version is not covered by Theorem 3, which assumes independent agents.

6.5 The Class-Conditional Estimator in Finite-Type Designs

The class-conditional estimator is analyzed in the finite-type regime of Assumption 7, with a single transition period, $\mathcal{S} = \{s\}$, as in the designs of Section 7. The case of several transition periods is discussed after Remark 2. Equivalence on the support at period s in stratum y is an equivalence relation on $\mathcal{W}_s$. The population class of an eligible agent j is $c_j = (y_{j,s-1}, [w_{js}])$, where $[w_{js}]$ is the set of supported indices equivalent to w_{js} on the support in stratum $y_{j,s-1}$. There are finitely many classes, and, by Assumption 7(ii), d_s takes a single value d_c on each class c. Let $e_j = \mathbb{1}\{j \in \mathcal{E}_s\}$, $p_j = F(Z'_{js}\theta - d_{c_j})$, and $q_j = p_j(1-p_j)$, and define, over a single agent,

$$\bar{Z}_c = \frac{\mathbb{E}\big[e_j \mathbb{1}\{c_j = c\}\, q_j Z_{js}\big]}{\mathbb{E}\big[e_j \mathbb{1}\{c_j = c\}\, q_j\big]}, \qquad \tilde{S}_j = e_j\,(b_{js} - p_j)\,(Z_{js} - \bar{Z}_{c_j}),$$
$$\mathcal{I}_{\mathrm{C}} = \mathbb{E}\big[e_j\, q_j\, (Z_{js} - \bar{Z}_{c_j})(Z_{js} - \bar{Z}_{c_j})'\big],$$

with $\bar{Z}_c$ defined for the classes that eligible agents occupy with positive probability. The matrix $\mathcal{I}_{\mathrm{C}}$ is the information about θ in a logit for the eligible agents' orderings with one free intercept per class, after the intercepts are profiled out, and it is positive definite when the regressor varies within classes.

Proposition 2 (Class-conditional estimator)**.** *Suppose that Assumption 1, Assumption 2 with the continuity requirement in part (i) replaced by Assumption 7(ii), and Assumptions 5(i)–(ii) and 7(i)–(v) hold, with $\mathcal{S} = \{s\}$, and that $\mathcal{I}_{\mathrm{C}}$ is positive definite.*

(a) (Class recovery.) With probability approaching one, the feasible classes at s are exactly the sets $\{j \in \mathcal{E}_s : c_j = c\}$.

(b) (Exact conditional likelihood.) Let $\mathcal{G}_s$ be the σ-field generated by the primitives $\{\chi_j\}_{j\le n}$, the outcomes $\{y_{j,-s}\}_{j\le n}$ outside periods s and $s{+}1$, the networks D_t with $t \notin \{s{+}1, s{+}2\}$, and the switch events $\{\mathcal{T}_{js}\}_{j\le n}$. Given $\mathcal{G}_s$, the orderings b_{js}, $j \in \mathcal{E}_s$, are independent, with $\mathbb{P}(b_{js} = 1 \mid \mathcal{G}_s) = F(Z'_{js}\theta - d_s(w_{js}))$. Consequently, for every $\mathcal{G}_s$-measurable set $C \subseteq \mathcal{E}_s$ whose members have the same movement $d_s(w_{js})$, $\exp\{\ell_C(\theta)\}$ is the conditional probability of the observed orderings in C given $\mathcal{G}_s$ and N_C. This holds, at every sample size, for every feasible class that does not merge agents with different movements, and in particular for every feasible class on the event of part (a).

(c) (Asymptotic distribution.) $\sqrt{n}\,(\hat{\theta}^{\mathrm{C}} - \theta) = \mathcal{I}_{\mathrm{C}}^{-1}\, n^{-1/2} \sum_{j=1}^{n} \tilde{S}_j + o_p(1) \xrightarrow{d} \mathcal{N}\big(0, \mathcal{I}_{\mathrm{C}}^{-1}\big)$, and n^{-1} times the negative Hessian of the objective in (19) at $\hat{\theta}^{\mathrm{C}}$ converges in probability to $\mathcal{I}_{\mathrm{C}}$.

Part (a) is the class-recovery step: the per-period codegree distance is estimated uniformly at the rate $(\log n/n)^{1/2}$ (Proposition 1(a)), and its positive population values are at least $\bar{\epsilon} \geq 2h_1$, so the thresholded graph eventually joins exactly the pairs of eligible agents in the same population class, whose connected components are the classes because equivalence is transitive. Part (b) makes precise that the class likelihood is exact rather than asymptotic: it requires only that no feasible class merges agents whose movements differ, and splitting a population class is harmless. Part (c) combines the two. Its proof in Appendix A.8 compares the class-conditional log likelihood with the profile log likelihood of the logit with one intercept per class. The two differ by the logarithm of the probability that a class has its observed number of switches of ordering B_s, which lies between $-\log(|C|+1)$ and zero, because that number is a mode of its distribution when the intercept is profiled out (Darroch, 1964), and whose gradient is bounded in probability, by the approximation of Hájek (1964) for the moments of independent Bernoulli variables conditioned on their sum (Lemma 5). The class estimator is therefore asymptotically equivalent to maximum likelihood with free class intercepts, which is why the profiled information $\mathcal{I}_{\mathrm{C}}$ appears. With a single transition period each eligible agent enters one class, so the inverse Hessian is a valid variance estimate, and no clustering is needed.

Remark 2 (Efficiency relative to the pairwise estimator when $T = 3$)**.** *Suppose $T = 3$, so that $\mathcal{S} = \{1\}$, and the conditions of Proposition 2 and Theorem 3 hold. Then $4\Sigma_{\mathrm{ex}}^{-1} V_{\mathrm{ex}} \Sigma_{\mathrm{ex}}^{-1} - \mathcal{I}_{\mathrm{C}}^{-1}$ is positive semidefinite, so the class estimator is at least as efficient asymptotically as the pairwise estimator. Both estimators use only the eligible agents' orderings, regressors, and population classes, which they recover with probability approaching one, and both are consistent in the model in which, given the regressors and classes, the orderings are independent logits with one intercept per class, with the intercepts and the law of the regressors unrestricted. In that model $\mathcal{I}_{\mathrm{C}}$ is the efficient information for θ, and $\mathcal{I}_{\mathrm{C}}^{-1}\tilde{S}_j$ is the efficient influence function, which the class estimator attains by Proposition 2(c). The pairwise estimator is asymptotically linear with influence function $2\Sigma_{\mathrm{ex}}^{-1}\psi_j$, and its estimating equation has mean zero at every law in that model, so its influence function is a gradient and the estimator is regular. By the convolution theorem (van der Vaart, 1998, Chapter 25), its asymptotic variance is at least the efficient one. Equality holds only when the two influence functions coincide, and the simulations of Section 7 measure the difference in finite samples. The comparison concerns this reduced model of the orderings, not the full model of outcomes and links, and no efficiency claim is made for $T > 3$, where the pairwise weights also use pair-specific distances at other periods. The argument is given in outline rather than as a formal proof.*

With several transition periods, the class log likelihoods of all periods can be summed in (19), and an agent who switches in several windows enters a class in each. Proposition 2

does not cover this case, which would require recovering the classes in all windows jointly and an expansion that accounts for the dependence of an agent's orderings across windows. A natural variance estimator is $\mathcal{I}_{\mathrm{C}}^{-1} V_{\mathrm{C}} \mathcal{I}_{\mathrm{C}}^{-1}$, with $\mathcal{I}_{\mathrm{C}}$ summed over periods and V_{C} estimated by clustering by agent the terms $(b_{js} - \hat{p}_{js})(Z_{js} - \hat{\bar{Z}}_C)$, where $\hat{p}_{js} = F(Z'_{js}\hat{\theta}^{\mathrm{C}} - \hat{\gamma}_C)$, $\hat{\gamma}_C$ solves $\sum_{j \in C} F(Z'_{js}\hat{\theta}^{\mathrm{C}} - \gamma) = N_C$, and $\hat{\bar{Z}}_C$ is the $\hat{p}(1-\hat{p})$-weighted mean of Z_{js} in the class, but its validity is not established here. The inverse Hessian alone would not account for that dependence.

Corollary 2 (Data-driven network bandwidth)**.** *Let $T = 3$, so that $\mathcal{S} = \{1\}$, and suppose that the conditions of Theorem 3 hold, except that in Assumption 7(v) the network bandwidth is a random variable $\hat{h}_1$ with $c_- n^{-1/4} \leq \hat{h}_1 \leq c_+ n^{-1/4}$ almost surely, for constants $0 < c_- \leq c_+ < \infty$. Then the conclusion of Theorem 3 and the consistency of its variance estimator hold with $h_1 = \hat{h}_1$, and so, if $\mathcal{I}_{\mathrm{C}}$ is positive definite, do parts (a) and (c) of Proposition 2. If $\hat{h}_1$ is also a function of D_1 and $(y_{10}, \ldots, y_{n0})$, part (b) holds as well. The rule* (16) *satisfies these conditions, with $c_- = 10^{-4} \cdot 500^{1/4}$ and $c_+ = 0.1 \cdot 500^{1/4}$.*

The proof (Appendix A.9) uses only the envelope: the bounds under which the estimated weights select the exact matches hold uniformly over bandwidths between $c_- n^{-1/4}$ and $c_+ n^{-1/4}$, so they hold at $\hat{h}_1$ whatever its dependence on the data. Measurability matters only for the exactness of the class likelihood at every sample size: D_1 and the initial outcomes are $\mathcal{G}_1$-measurable, so the feasible classes remain so. The envelope is loose for the rule (16), whose S_n is about 0.007 in the baseline design of Section 7, far below its bound of one.

6.6 Continuous Designs

When network types or covariates vary continuously, matches are inexact at every positive bandwidth. Theorem 2 establishes consistency with shrinking bandwidths, and Remark 1 describes the limit at fixed bandwidths. The paper does not establish a limiting distribution in these designs, and I explain why, because the reasons also determine how the simulation results for continuous designs should be read.

First, the estimated weights of different pairs share network rows: the link D_{ij} enters both agents' neighborhoods, and each neighborhood depends on overlapping sets of third agents. Per-agent influence terms built from an agent's realized links are therefore not independent across agents, and a central limit theorem requires an expansion in the independent node primitives and dyadic shocks, with the dependence of each component controlled.

Second, the estimation error of the codegree distance has two sources: link noise given the latent types, and sampling variation in the distribution of the auxiliary agents' latent types. The second does not vanish when link noise is averaged out. For example, with $f(u, v) = uv$

and w uniform, the population distance is $\delta_{ij}^2 = (w_i - w_j)^2 m_2^3$ with $m_2 = \mathbb{E}[w^2]$, whereas replacing links by their probabilities in the statistic yields, up to repeated-index terms, $(w_i - w_j)^2 M_{2n}^3$ with $M_{2n} = n^{-1} \sum_l w_l^2$, whose fluctuation includes $3(w_i - w_j)^2 m_2^2 (M_{2n} - m_2)$, of order $n^{-1/2}$. A first-stage correction must account for this term.

Third, the variance normalization depends on how the probability of being matched varies across agents. A localized estimator's leading variance contains a factor of the form $\mathbb{E}[r_h(w_i)^2]/(\mathbb{E}[r_h(w_i)])^2$, where $r_h(w)$ is an agent's conditional match mass, and this factor can diverge as $h_1 \to 0$ (Auerbach, 2021, discussion of Proposition C2). For example, with $T = 3$, $w_{it} = w_i$ uniform, $\lambda = 0$, $\mu_i = 0$, $\beta = \alpha = 0$, exactly stable covariates, and the smooth, separating link function $f(u, v) = \frac{1}{4} + \frac{1}{4}(u^4 + v^4)$, the squared codegree distance is proportional to $(u^4 - v^4)^2$, the match mass behaves like $b^{1/4}$ for $w \lesssim b^{1/4}$ and like b/w^3 otherwise with $b = h_1^{1/2}$, and the factor is of order $h_1^{-1/8}$. The leading variance of the oracle estimator is then of order $1/(n h_1^{1/8})$, not $1/n$. There is therefore no universal convergence rate in continuous designs, and a valid theory must use a variance normalization that accommodates heterogeneous match masses, possibly after equalizing agents' match counts as Auerbach (2021) suggests.

Finally, consider bias correction. If the pseudo-true value admitted an expansion

$$\theta_h = \theta + \sum_{l=1}^{L_1} C_l^{(1)} h_1^{l/\vartheta_1} + \sum_{l=1}^{L_2} C_l^{(2)} h_2^{l/\vartheta_2} + R(h_1, h_2)$$

with known exponents, the combination

$$\bar{\theta} = \sum_{l=1}^{M+1} a_l \, \hat{\theta}_{c_l h}, \qquad \sum_l a_l = 1, \qquad \sum_l a_l \, c_l^{e_m} = 0 \ \text{ for } m = 1, \ldots, M, \tag{21}$$

of estimators computed at bandwidths $(c_l h_1, c_l h_2)$, where $\{e_1, \ldots, e_M\}$ is the merged set of distinct exponents, would remove every listed term: the weights solve a generalized Vandermonde system and are unique (Honoré and Powell, 2005). The exponents are not derived here for any continuous design, symmetry of the kernel window can remove first-order terms, and inference based on $\bar{\theta}$ would in any case require the joint central limit theorem that the three preceding points show to be missing. The generalized jackknife is therefore recorded as an algebraic device, not as an inference result.

For these reasons, in continuous designs the paper reports point estimates together with the finite-type standard errors for comparison only. Confidence intervals based on them are not justified by the theory, and the simulations of Section 7 report their coverage as a diagnostic.

7 Monte Carlo Simulations

This section studies the finite-sample behavior of the proposed estimator in designs that satisfy the assumptions of the theorems and in designs that deliberately violate them. The simulations address five questions. First, in the finite-type regime of Theorem 3, with deterministic bandwidths, is the estimator centered at the truth, do the estimated standard errors match the sampling variability, and do the intervals have their nominal coverage? Second, what does estimating the matches cost relative to an oracle that uses the true network types? Third, how does the estimator compare with within-agent conditioning (Honoré and Kyriazidou, 2000)? Fourth, where does the method lose ground or fail: when the heterogeneity is time invariant, so that within-agent conditioning is valid and more precise, with movement that is not common to network types, with sparse or weakly informative networks, and with covariates in link formation? Fifth, in the finite-type designs, how much does the class-conditional estimator of Section 5.3, which uses the same comparisons class by class rather than pair by pair, gain over the pairwise estimator, and does the gain come from the network classes or only from allowing a common movement? A continuous design, in which only consistency is established, and the data-driven bandwidth (16) complete the section.

7.1 Data Generating Processes

In every design, $T = 3$, the true parameters are $\beta_0 = 1$ and $\alpha_0 = 0.5$, the social influence function is $\lambda(w) = (4w^3 - 3)/2$, the agent effect is $\mu_i \sim \mathcal{N}(0, 1)$, independent of the latent index and absent from the link function, the outcome shocks are standard logistic, the link shocks are uniform and independent across dyads and waves, and the initial outcome is generated from the model without a lag, $y_{i0} = \mathbb{1}\{X_{i0}\beta_0 + \lambda(w_{i0}) + \mu_i - \varepsilon_{i0} \geq 0\}$. The covariate depends on both unobserved terms, so that the naive estimator is inconsistent for two reasons, and the agent effect makes any comparison that requires equal levels of the unobserved term invalid.

Finite-type designs. The latent index takes four values, $w \in \{0.15, 0.40, 0.65, 0.90\}$, labeled types $k = 0, 1, 2, 3$ and drawn with equal probability. Types are constant at $t = 0, 1$. At $t = 2$, every agent moves by the common cyclic permutation $k \mapsto k + 1 \pmod 4$, and types are constant again at $t = 3$. The type distribution is therefore the same in every period, the movement of λ between the transition waves 1 and 2 is common to agents of the same type ($d_1 \neq 0$), and there is no movement after the transition ($d_2 = 0$), so Assumption 2 holds with $\mathcal{S} = \{1\}$. Since the four types have distinct link profiles, every network-type equivalence

class is a single type, and the induced change $d_1 = \lambda \circ g_1 - \lambda$, with g_1 the common map defined by the permutation, is trivially constant on each class. The covariate is integer valued, $X_{it} = \text{round}(2w_{it} + 0.5\mu_i + \upsilon_{it})$ with $\upsilon_{it} \sim \mathcal{N}(0,1)$, so it is correlated with the latent index and with the agent effect, and exact covariate stability has positive probability. Links form with probability $f(w_{it}, w_{jt}) = P_{k_{it}k_{jt}}$, where

$$P = \begin{pmatrix} 0.06 & 0.66 & 0.65 & 0.33 \\ 0.66 & 0.59 & 0.11 & 0.15 \\ 0.65 & 0.11 & 0.49 & 0.61 \\ 0.33 & 0.15 & 0.61 & 0.05 \end{pmatrix}.$$

The four types have distinct codegree profiles: the smallest population squared codegree distance between different types, computed exactly from P and the equal type shares, is $\bar{\epsilon} = 3.336 \times 10^{-3}$, so the separation gap of Assumption 7 is positive. The matrix is chosen so that the average latent index of an agent's neighbors carries little information about the agent's own index, which is the case in which peer-average controls cannot proxy for the latent characteristic. This baseline design, which I call *sbm4*, satisfies Assumptions 1, 2, and 7. Three variants keep these assumptions and change one feature of the design.

- *sbm4dyn*: link probabilities are multiplied by $1 + 0.4\, y_{i,t-1} y_{j,t-1}$, so that links depend on lagged outcomes and the lag stratification of the codegree distance binds. The separation gap, computed over (type, y_{i0}) cells with the population cell shares, is 3.852×10^{-3}.
- *fixed*: types do not move, so the heterogeneity is time invariant and within-agent conditioning is valid.
- *weak*: the block matrix is shrunk halfway toward its mean, which divides the separation gap by ten ($\bar{\epsilon} = 3.281 \times 10^{-4}$). The design satisfies the assumptions, but the gap is close to the sampling noise of the estimated distances at the sample sizes considered.

Three further variants leave the assumptions, each in one respect.

- *idio*: at $t = 2$, a randomly chosen 30% of agents draw a new type uniformly instead of following the permutation, so the movement is not common to network types and Assumption 2(i) fails.
- *xlink*: link probabilities are multiplied by 1.5 when $X_{it} = X_{jt}$ and by 0.75 otherwise, so that observed covariates enter the link function, which the network model of equation (2) excludes. Agents of the same type with different covariates then have different codegree profiles. The separation gap between different types, computed over $(\text{type}, X_{i1}, y_{i0})$ cells, is 8.120×10^{-4}.

- *sparse*: link probabilities are rescaled so that the expected degree is 8 at every sample size, outside the dense-graphon asymptotics of Section 6. This design also changes the bandwidth rule, as explained in Section 7.2.

Continuous design. The latent index is $w_{i0} = w_{i1} \sim U[0,1]$ and $w_{i2} = w_{i3} = w_{i1}^{1/2}$, so the movement between the transition waves is common to agents with the same index and there is no movement afterward. The distribution of the index differs across periods, which Assumption 1(iii) allows. Links follow the steep beta model $f(u,v) = F(8(u+v-1))$, with F the logistic distribution function. The link function is continuous and strictly increasing in each argument and therefore separates types. The covariate is continuous, $X_{it} = (3w_{it}^3 + 0.5\mu_i + \upsilon_{it} + 1)/3$. This design, called *cont*, is in the regime of Theorem 2 only.

7.2 Estimators and Bandwidths

Six estimators are compared (five in the continuous design, which omits the controls estimator), and four class-based estimators are added. The *naive* estimator is a pooled dynamic logit with an intercept, ignoring λ. The *controls* estimator adds the averages of neighbors' covariates and lagged outcomes. The *HK* estimator is the within-agent conditional logit of Honoré and Kyriazidou (2000) on the transition event $y_{i1} + y_{i2} = 1$, with regressors $(X_{i1} - X_{i2}, y_{i0} - y_{i3})$ and the covariate-stability weight. The *proposed* estimator is (15) with $\mathcal{S} = \{1\}$. The *oracle* estimator is the same pairwise estimator computed with population matches: in the finite-type designs, the true type match; in the continuous design, the population codegree distances $\delta_1(w_{i1}, w_{j1})^2$, evaluated numerically by Gauss–Legendre quadrature from the link function and the uniform distribution of the index, so that neither the link noise nor the sampling variation of the other agents' latent indices enters the oracle's first stage. It isolates the cost of estimating the matches. The *infeasible* estimator is a pooled dynamic logit that includes the true unobserved term $\lambda(w_{it}) + \mu_i$ as a regressor. The *class* estimator is (19) with $\mathcal{S} = \{1\}$ and the feasible classes of (17), computed from the same estimated distances, the same bandwidth h_1, and the same covariate-stability selection as the proposed estimator. The *class oracle* uses the population classes, the true type at $t = 1$ crossed with y_{i0}. The *joint* estimator estimates one intercept per feasible class together with (β, α) instead of conditioning the intercepts out. The *one-class* estimator, a within-agent conditional logit with a common time effect conditioned out, uses a single class per lag stratum, so it allows one common movement of λ for all eligible agents that share y_{i0} and uses no network information. In the continuous design the class estimator uses the connected components of the thresholded graph, which are not classes of equal movement there, with the covariate-stability window used as an indicator. The class-based estimators are computed by a separate program that regenerates each saved

sample from its seed. Before computing them, it reproduces the proposed estimator from the regenerated sample, which it does in all 13,500 replications, to a largest absolute difference of 7×10^{-13} in the estimates and standard errors.

The bandwidths are deterministic sequences with the rates that the theory requires. In the finite-type designs, $h_1 = 0.5\,\bar{\epsilon}\,(n/500)^{-1/4}$, where $\bar{\epsilon}$ is the population separation gap of the design, a fixed constant computed once from the block matrix and the population shares of the (type, y_{i0}) or $(\text{type}, X_{i1}, y_{i0})$ cells (3.336×10^{-3} in the baseline, fixed, and idiosyncratic designs, and the values given above for the lag-dependent, weak, and covariate-link designs). The cell shares are obtained by numerical integration of the model for the initial outcome and the covariate, and the code that computes them accompanies the paper. With $T = 3$, only period 1 enters the aggregated distance of a discordant pair, so this sequence satisfies Assumption 7(v). The calibration uses the population link probabilities and is therefore infeasible in practice. It is the sequence for which Theorem 3 is stated, and Section 7.4 replaces it by the data-driven rule (16). A variant of the rule that recomputes $\bar{\epsilon}$ in each replication from the realized type shares, which makes the bandwidth random, gives practically the same results in the baseline design: the set of pairs with positive weight is identical in 91 to 100 percent of the replications, the median absolute change in $\hat{\beta}$ is at most 0.0015, the bias for β changes by at most 0.001, and its coverage is unchanged to three decimals. Since the covariate is integer valued, $h_2 = 0.5$ selects exact stability, for HK as well as for the pairwise estimators. The exception for h_1 is the sparse design, in which the separation gap is far below the sampling noise of the estimated distances, so no bandwidth below the gap can select matches. There, h_1 is set to the 0.10 quantile of the estimated distances among pairs sharing the initial outcome, a practical rule that the theory does not cover. In the continuous design, $h_1 = c\,(n/500)^{-1/5}$, where $c = 5.13 \times 10^{-4}$ is the population 0.10 quantile of the squared codegree distance between two independent uniform indices at $t = 1$, computed by numerical integration (quadrature for the codegree integrals and an 8000×8000 midpoint grid with Richardson extrapolation for the quantile, accurate to about 0.2 percent), and $h_2 = n^{-1/5}/5$, for HK as well. All pairwise estimators use the kernel $K(x) = \frac{3}{4}(1 - x^2)\mathbb{1}(0 \le x < 1)$.

Standard errors are the MLE standard errors for the pooled logits, a sandwich for HK, and, for the proposed and oracle estimators, the sandwich that clusters the pair-level scores by agent, which is the plug-in estimator of Theorem 3 in the finite-type designs. For the class-conditional estimators, they are the inverse of the negative Hessian of the class-conditional log likelihood, as in Proposition 2(c), and for the joint estimator the inverse information of the joint logit. In the continuous design no distribution theory is established, and the same standard errors are reported only to show how far they are from the sampling variability. Each design is run with $R = 500$ replications. At that number, the Monte Carlo standard

error of a coverage rate near 95% is about one percentage point, and Table 4 reports it cell by cell. A fit counts as failed when fewer than three pairs (for the within-agent estimator, fewer than three agents) receive positive weight, when the Hessian is singular, when the Newton iterations do not converge within 100 steps (with a tolerance of 10^{-9} on the step), or when a coefficient exceeds 50 in absolute value. The last two are symptoms of a likelihood without a maximizer, but they do not establish it. For every failed fit of the pairwise and within-agent estimators whose selected sample is usable, that is, every failure due to a singular Hessian, nonconvergence, or divergence, a linear program therefore checks whether the weighted likelihood is completely separated, in which case a direction orders every positive-weight pair correctly, or quasi-completely separated, in which case a direction orders every pair weakly and some strictly (Albert and Anderson, 1984). In either case, the maximum likelihood estimate does not exist. Fits that fail because fewer than three pairs or agents are selected are classified separately and are not checked. For the class-based estimators, the same classification applies with fewer than three eligible agents in classes that contain both orderings, or a design without full rank, in the role of too few pairs, and the separation check is applied to the logit with one dummy per class. The tables report the number of failed fits of each kind and, for the optimizer failures, the number certified as separated in one of the two senses. All statistics are computed over the replications with an estimate, and Table 4 reports both the coverage conditional on a returned interval and the share of all attempted replications that return a covering interval.

7.3 Results

Tables 1 and 2 report the baseline design, Tables 3 and 4 the matching, inference, and success diagnostics of the proposed estimator, Table 5 the class-conditional estimator, and Table 6 the benchmark designs at $n = 2000$. Appendix F reports the other designs and sample sizes.

The finite-type regime. The baseline design is the regime of Theorem 3. From $n = 1000$ onward, the estimated weights select exactly the same-type pairs (precision and recall equal one, against 0.996 at $n = 500$), and the proposed estimator agrees with the oracle to within 0.001 in bias, standard deviation, and RMSE, as the oracle equivalence in the proof of the theorem implies. Its bias for β falls from 0.240 at $n = 500$ to 0.026 at $n = 3000$, and its bias for α is below 0.03 in absolute value from $n = 1000$ onward. The oracle has the same small-sample bias, so the bias comes from a conditional logit that rests on few agents, not from the estimated matches. At $n = 500$, about 50 agents are eligible and about 36 have a partner (Table 3). The information grows with the number of contributing agents, not with the number of pairs, because pairs that share an agent are dependent. At that sample size 69

Table 1: Monte Carlo results for β: baseline finite-type design

n	Naive	Controls	HK	Proposed	Oracle	Class	Class orc.	Infeas.
Bias								
500	0.247	0.242	0.347	0.240	0.235	0.149	0.128	0.009
1000	0.245	0.239	0.296	0.151	0.151	0.085	0.084	0.008
2000	0.240	0.233	0.258	0.070	0.069	0.042	0.042	0.001
3000	0.240	0.233	0.228	0.026	0.026	0.009	0.009	0.001
Standard deviation								
500	0.069	0.069	0.450	0.633	0.628	0.544	0.533	0.076
1000	0.053	0.053	0.316	0.433	0.434	0.372	0.372	0.056
2000	0.034	0.034	0.191	0.253	0.253	0.219	0.219	0.036
3000	0.030	0.030	0.154	0.199	0.199	0.174	0.174	0.033
Root mean squared error								
500	0.256	0.252	0.568	0.677	0.670	0.563	0.548	0.076
1000	0.251	0.245	0.433	0.458	0.459	0.381	0.381	0.056
2000	0.242	0.236	0.321	0.262	0.262	0.222	0.222	0.036
3000	0.242	0.235	0.275	0.201	0.201	0.174	0.174	0.033
Coverage of nominal 95% interval								
500	0.050	0.058	0.957	0.968	0.968	0.963	0.963	0.934
1000	0.000	0.004	0.876	0.950	0.950	0.974	0.974	0.944
2000	0.000	0.000	0.772	0.944	0.944	0.960	0.960	0.970
3000	0.000	0.000	0.754	0.942	0.944	0.950	0.950	0.936
Replications without an estimate								
500	0	0	12	69	69	65	69	0
1000	0	0	0	0	0	0	0	0
2000	0	0	0	0	0	0	0	0
3000	0	0	0	0	0	0	0	0

Notes: $R = 500$ replications; $\beta_0 = 1$, $\alpha_0 = 0.5$. Naive: pooled dynamic logit. Controls: pooled dynamic logit with neighbors' average covariate and lagged outcome. HK: within-agent conditional logit of Honoré and Kyriazidou (2000). Proposed: estimator (15). Oracle: the same pairwise estimator with population matches. Class: the class-conditional estimator (19), with classes formed from the same estimated distances and bandwidth. Class orc.: the same likelihood with the population classes. Infeas.: pooled dynamic logit including the true unobserved term $\lambda(w_{it}) + \mu_i$. Bandwidths: $h_1 = 0.5\,\bar{\epsilon}\,(n/500)^{-1/4}$, with $\bar{\epsilon}$ the population separation gap of the design, a fixed constant computed from the block matrix and the population type shares, and exact covariate stability ($h_2 = 0.5$ on the integer covariate). Coverage: share of replications in which the nominal 95% Wald interval contains the true value. Standard errors are MLE for the pooled logits, a sandwich for HK, the sandwich clustering pair-level scores by agent for the pairwise estimators, and the inverse of the negative Hessian of the class-conditional log likelihood for the class estimators. Statistics are computed over the replications with an estimate, and failures are defined in Section 7.2.

fits fail, all on separated likelihoods. Conditional on a returned interval, coverage is between 0.94 and 0.97 for both parameters at every sample size, but at $n = 500$ only 83 percent of all attempts return an interval that covers β_0 (Table 4). The clustered standard errors are slightly too small: the ratio of the Monte Carlo standard deviation to the mean standard error is between 1.02 and 1.16.

Table 2: Monte Carlo results for α: baseline finite-type design

n	Naive	Controls	HK	Proposed	Oracle	Class	Class orc.	Infeas.
Bias								
500	0.701	0.723	0.373	−0.076	−0.075	−0.044	−0.050	−0.016
1000	0.713	0.737	0.377	0.000	0.000	0.003	0.002	0.004
2000	0.708	0.732	0.374	0.016	0.016	−0.005	−0.005	−0.004
3000	0.709	0.734	0.371	0.027	0.027	0.024	0.024	−0.005
Standard deviation								
500	0.143	0.145	0.736	1.299	1.296	1.164	1.131	0.151
1000	0.100	0.103	0.485	0.873	0.872	0.758	0.758	0.108
2000	0.070	0.073	0.330	0.543	0.543	0.478	0.478	0.073
3000	0.057	0.060	0.267	0.418	0.418	0.380	0.380	0.060
Root mean squared error								
500	0.715	0.737	0.825	1.300	1.297	1.163	1.131	0.152
1000	0.720	0.744	0.614	0.872	0.872	0.757	0.757	0.108
2000	0.712	0.736	0.499	0.543	0.542	0.477	0.477	0.073
3000	0.712	0.736	0.457	0.419	0.419	0.380	0.380	0.061
Coverage of nominal 95% interval								
500	0.002	0.000	0.932	0.961	0.961	0.991	0.991	0.938
1000	0.000	0.000	0.898	0.938	0.940	0.964	0.964	0.932
2000	0.000	0.000	0.802	0.942	0.942	0.952	0.952	0.946
3000	0.000	0.000	0.690	0.944	0.944	0.954	0.954	0.956
Replications without an estimate								
500	0	0	12	69	69	65	69	0
1000	0	0	0	0	0	0	0	0
2000	0	0	0	0	0	0	0	0
3000	0	0	0	0	0	0	0	0

Notes: $R = 500$ replications; $\beta_0 = 1$, $\alpha_0 = 0.5$. Naive: pooled dynamic logit. Controls: pooled dynamic logit with neighbors' average covariate and lagged outcome. HK: within-agent conditional logit of Honoré and Kyriazidou (2000). Proposed: estimator (15). Oracle: the same pairwise estimator with population matches. Class: the class-conditional estimator (19), with classes formed from the same estimated distances and bandwidth. Class orc.: the same likelihood with the population classes. Infeas.: pooled dynamic logit including the true unobserved term $\lambda(w_{it}) + \mu_i$. Bandwidths: $h_1 = 0.5\,\bar{\epsilon}\,(n/500)^{-1/4}$, with $\bar{\epsilon}$ the population separation gap of the design, a fixed constant computed from the block matrix and the population type shares, and exact covariate stability ($h_2 = 0.5$ on the integer covariate). Coverage: share of replications in which the nominal 95% Wald interval contains the true value. Standard errors are MLE for the pooled logits, a sandwich for HK, the sandwich clustering pair-level scores by agent for the pairwise estimators, and the inverse of the negative Hessian of the class-conditional log likelihood for the class estimators. Statistics are computed over the replications with an estimate, and failures are defined in Section 7.2.

Estimators that do not use the path comparison. The naive and controls estimators are biased by 0.23 to 0.25 for β and 0.70 to 0.74 for α at every sample size: a pooled logit attributes the omitted persistent effect to state dependence, and neighbors' averages barely help, because friends' characteristics carry little information about an agent's own type and none about the agent effect. The within-agent estimator removes the agent effect and the level of λ but not the common movement of λ, so its bias, 0.23 to 0.35 for β and about 0.37

Table 3: Baseline and lag-dependent designs: matching, contributing-agent, and inference diagnostics for the proposed estimator

				Agents				Matching		Proposed estimator					
Design	n	Pairs	ESS	Elig.	Contrib.	Unm.	Min.	Prec.	Recall	Par.	Med. bias	SD	Mean SE	SD/SE	Coverage
sbm4	500	48	48	50	36	14	1.0	1.00	1.00	β	0.150	0.633	0.614	1.03	0.968
										α	−0.008	1.299	1.247	1.04	0.961
sbm4	1000	190	190	99	86	14	1.1	1.00	1.00	β	0.086	0.433	0.375	1.16	0.950
										α	0.034	0.873	0.804	1.09	0.938
sbm4	2000	772	772	200	191	9	1.6	1.00	1.00	β	0.047	0.253	0.236	1.07	0.944
										α	0.027	0.543	0.524	1.04	0.942
sbm4	3000	1733	1733	299	294	5	2.1	1.00	1.00	β	0.017	0.199	0.187	1.07	0.942
										α	0.041	0.418	0.411	1.02	0.944
sbm4dyn	500	48	48	50	36	14	1.0	1.00	1.00	β	0.151	0.632	0.608	1.04	0.965
										α	−0.019	1.427	1.279	1.12	0.941
sbm4dyn	1000	192	192	100	86	14	1.2	1.00	1.00	β	0.069	0.411	0.373	1.10	0.950
										α	−0.030	0.917	0.809	1.13	0.942
sbm4dyn	2000	757	757	200	190	10	1.6	1.00	1.00	β	0.052	0.262	0.240	1.09	0.948
										α	−0.005	0.551	0.521	1.06	0.938
sbm4dyn	3000	1734	1733	300	295	4	2.0	1.00	1.00	β	0.042	0.187	0.190	0.98	0.950
										α	−0.008	0.415	0.411	1.01	0.946

Notes: Pairs: average number of pairs with positive weight. ESS: kernel-weighted effective sample size $(\sum_p w_p)^2/\sum_p w_p^2$. Elig.: average number of eligible agents, who switch between waves 1 and 2 and have a positive covariate-stability weight (the agents that can enter a discordant pair and the agents in the within-agent likelihood). Contrib.: average number of eligible agents with at least one positive-weight partner. Unm.: average number of eligible agents with no partner. Min.: average over replications of the smallest number of positive-weight partners among contributing agents. Pair counts and ESS measure the concentration of the weights, not the number of independent agents, because pairs that share an agent are dependent. Prec.: share of weighted pairs whose members have the same type. Recall: share of same-type discordant pairs with stable covariates that receive positive weight. Med. bias: median of $\hat{\theta} - \theta_0$. SD: Monte Carlo standard deviation. Mean SE: average of the clustered standard errors. The mean length of the 95% interval is 3.92 times this number. Coverage of the nominal 95% interval. Failed fits: sbm4 at $n = 500$: Proposed 69 (69 optimizer failures, all certified as separated); sbm4dyn at $n = 500$: Proposed 75 (75 optimizer failures, all certified as separated); and sbm4dyn at $n = 1000$: Proposed 1 (1 optimizer failure, certified as separated).

for α, does not vanish, and its coverage for β falls from 0.96 at $n = 500$ to 0.75 at $n = 3000$. It still has the smaller RMSE for both parameters at $n \leq 1000$ (for β, 0.568 against 0.677 at $n = 500$), whereas the proposed estimator has the smaller RMSE for β from $n = 2000$ and for both parameters at $n = 3000$: removing a bias that does not shrink costs variance, which pays off only as n grows. Restricting each comparison to the replications in which both estimators succeed does not reverse any ranking (Table 16).

Conditioning on network classes. In the baseline and lag-dependent designs, the class-conditional estimator has the smaller RMSE for both parameters at every sample size, by 6 to 20 percent, with coverage between 0.95 and 0.99 (Table 5). The gain comes from the likelihood, not from the classification: from $n = 1000$ the feasible classes are the population

Table 4: Baseline and lag-dependent designs: success-qualified diagnostics for the proposed estimator

Design	n	Fits	Par.	Bias	5th pct.	95th pct.	SD	Mean SE	Length	Coverage	MC SE	Covering share
Baseline	500	431/500	β	0.240	−0.610	1.529	0.633	0.614	2.41	0.968	0.009	0.834
			α	−0.076	−2.231	1.947	1.299	1.247	4.89	0.961	0.009	0.828
Baseline	1000	500/500	β	0.151	−0.398	0.861	0.433	0.375	1.47	0.950	0.010	0.950
			α	0.000	−1.476	1.366	0.873	0.804	3.15	0.938	0.011	0.938
Baseline	2000	500/500	β	0.070	−0.277	0.538	0.253	0.236	0.92	0.944	0.010	0.944
			α	0.016	−0.844	0.948	0.543	0.524	2.05	0.942	0.010	0.942
Baseline	3000	500/500	β	0.026	−0.282	0.366	0.199	0.187	0.73	0.942	0.010	0.942
			α	0.027	−0.647	0.734	0.418	0.411	1.61	0.944	0.010	0.944
Lag-dependent links	500	425/500	β	0.234	−0.650	1.412	0.632	0.608	2.38	0.965	0.009	0.820
			α	−0.055	−2.436	2.127	1.427	1.279	5.02	0.941	0.011	0.800
Lag-dependent links	1000	499/500	β	0.143	−0.399	0.857	0.411	0.373	1.46	0.950	0.010	0.948
			α	0.008	−1.317	1.568	0.917	0.809	3.17	0.942	0.010	0.940
Lag-dependent links	2000	500/500	β	0.068	−0.311	0.482	0.262	0.240	0.94	0.948	0.010	0.948
			α	0.000	−0.893	0.881	0.551	0.521	2.04	0.938	0.011	0.938
Lag-dependent links	3000	500/500	β	0.043	−0.258	0.359	0.187	0.190	0.74	0.950	0.010	0.950
			α	0.004	−0.699	0.724	0.415	0.411	1.61	0.946	0.010	0.946

Notes: Fits: replications returning an estimate, out of 500 attempted. Bias, 5th and 95th percentiles of $\hat{\theta} - \theta_0$, Monte Carlo standard deviation (SD), and mean clustered standard error, computed over the successful fits. Length: mean length of the nominal 95% Wald interval. Coverage: share of successful fits whose interval contains the true value, with its Monte Carlo standard error (MC SE). Covering share: share of all 500 attempts that return an interval containing the true value, which is not a coverage probability but an operational success rate. Failed fits: Baseline at $n = 500$: Proposed 69 (69 optimizer failures, all certified as separated); Lag-dependent links at $n = 500$: Proposed 75 (75 optimizer failures, all certified as separated); and Lag-dependent links at $n = 1000$: Proposed 1 (1 optimizer failure, certified as separated).

classes in at least 99.6 percent of the replications, and with the population classes the gain over the pairwise oracle is 6 to 19 percent, because the class likelihood uses each class's sufficient statistic once, whereas the pairwise likelihood counts overlapping pairs. Conditioning matters: estimating the class intercepts instead raises the bias for β at $n = 500$ from 0.149 to 0.485 (Table 18). The network classes matter: a single class per lag stratum, which uses no network information, is biased by 0.229 at $n = 3000$, as is the within-agent estimator (0.228), because the movements differ across types. In the weak-network design, the transitive class graph joins same-type agents whose direct distance exceeds the bandwidth. Its partitions split classes but never merge them, which keeps the class likelihood exact, and the RMSE for β at $n = 2000$ falls from 0.362 to 0.246. Outside the assumptions, the class estimator is biased where the pairwise estimator is (Tables 6, 14, and 17).

Designs in which the assumptions hold. With lag-dependent links, the lag-stratified distance still separates the types, and the results are those of the baseline design (Table 6). When the heterogeneity does not move, $\lambda(w_i) + \mu_i$ is a fixed effect, the within-agent estimator is valid and more precise (RMSE for β of 0.195, against 0.223 for the pairwise and 0.207 for the class estimator), and the network adds nothing. In the weak-network design, the

Table 5: Pairwise and class-conditional estimators: baseline and lag-dependent designs

		Successful fits			RMSE for β			RMSE for α			Class: coverage		Classes		Time (ms)	
Design	n	Prop.	Class	Common	Prop.	Class	Class orc.	Prop.	Class	Class orc.	β	α	Exact	Used	Prop.	Class
Baseline	500	431	435	431	0.677	0.551	0.548	1.300	1.136	1.131	0.963	0.991	88.2	36	0.5	12.6
	1000	500	500	500	0.458	0.381	0.381	0.872	0.757	0.757	0.974	0.964	99.6	86	0.6	14.0
	2000	500	500	500	0.262	0.222	0.222	0.543	0.477	0.477	0.960	0.952	100.0	191	1.0	22.4
	3000	500	500	500	0.201	0.174	0.174	0.419	0.380	0.380	0.950	0.954	100.0	294	1.4	34.3
Lag-dependent links	500	425	426	424	0.674	0.541	0.540	1.426	1.206	1.225	0.984	0.984	93.4	36	0.5	6.2
	1000	499	499	499	0.435	0.354	0.354	0.917	0.791	0.791	0.968	0.952	99.6	86	0.5	20.9
	2000	500	500	500	0.270	0.232	0.232	0.550	0.472	0.472	0.960	0.950	100.0	190	0.9	21.8
	3000	500	500	500	0.191	0.172	0.172	0.415	0.391	0.391	0.962	0.950	100.0	295	1.4	35.0

Notes: $R = 500$ replications. Prop.: the pairwise estimator (15). Class: the class-conditional estimator (19), with feasible classes formed from the pairwise estimator's estimated distances and bandwidth h_1. Class orc.: the same likelihood with the population classes (type at $t = 1$ crossed with y_{i0}). Successful fits: replications in which the estimator returns finite estimates and standard errors. Common: replications in which the pairwise, class, and class-oracle estimators all succeed (in the continuous design, which has no class oracle, the first two). RMSE on the common replications. Class coverage: coverage of the nominal 95% interval based on the inverse of the negative Hessian of the class-conditional log likelihood, over its own successful fits. Exact: percentage of replications in which the feasible classes coincide with the population classes among the eligible agents. Used: average number of eligible agents in classes that contain both orderings. Time: median computing time per replication, in milliseconds on one thread (two single-threaded jobs ran concurrently), of each estimator's own step (pairs and fit for the pairwise estimator, connected components and fit for the class estimator). The codegree distances among eligible agents, which both use, take a further 0.7 to 100 milliseconds between $n = 500$ and $n = 3000$. Failed fits are optimizer failures, all certified as separated, except one class-oracle fit with a rank-deficient design: Baseline at $n = 500$: Proposed 69, Class 65, and Class orc. 69 (68 optimizer failures and the rank-deficient fit); Lag-dependent links at $n = 500$: Proposed 75, Class 74, and Class orc. 76; and Lag-dependent links at $n = 1000$: 1 for each estimator.

separation gap is ten times smaller. At $n = 1000$, it is below the noise of the estimated distances, so no pair is selected (Table 14), and recall rises to 0.45 at $n = 2000$ and 0.79 at $n = 3000$ (Table 15). The theory needs a gap that is large relative to $(\log n/n)^{1/2}$.

Designs outside the assumptions. When 30% of the agents move idiosyncratically, Assumption 2(i) fails, and the bias of the proposed estimator rises to 0.131 for β and 0.122 for α, against 0.070 and 0.016 in the baseline design. Its coverage stays near nominal only because its standard errors are large. When covariates enter link formation, same-type agents with different covariates have different codegree profiles, so recall falls to about 0.32 with precision one (Table 15), and the RMSE for β is 0.394, against 0.249 for the oracle. In the sparse design, with expected degree 8, fewer than one in five weighted pairs have the same type, and the bias for β is 0.555 with coverage 0.78, while the oracle is centered: with bounded degrees, the codegree distance carries little information about types, as in the application of Section 8.

The continuous design. Only consistency is established here (Theorem 2). With the shrinking bandwidths, the bias of the proposed estimator for β falls from 0.431 at $n = 500$

Table 6: Benchmark designs at $n = 2000$

	HK			Proposed			Oracle			Class		
Design	Bias	RMSE	Cov.	Bias	RMSE	Cov.	Bias	RMSE	Cov.	Bias	RMSE	Cov.
β ($\beta_0 = 1$)												
Designs that satisfy the assumptions												
Baseline	0.258	0.321	0.77	0.070	0.262	0.94	0.069	0.262	0.94	0.042	0.222	0.96
Lag-dependent links	0.264	0.331	0.81	0.068	0.270	0.95	0.068	0.271	0.95	0.047	0.232	0.96
Fixed types	0.030	0.195	0.94	0.046	0.223	0.96	0.046	0.223	0.96	0.032	0.207	0.96
Weak network	0.252	0.322	0.81	0.110	0.362	0.94	0.064	0.268	0.95	0.041	0.246	0.94
Designs outside the assumptions												
Idiosyncratic movement	0.245	0.317	0.79	0.131	0.287	0.93	0.131	0.287	0.93	0.107	0.253	0.94
Covariates in links	0.258	0.327	0.78	0.104	0.394	0.95	0.064	0.249	0.94	0.051	0.288	0.95
Sparse network	0.223	0.292	0.85	0.555	0.833	0.78	0.036	0.239	0.96	0.312	0.431	0.90
α ($\alpha_0 = 0.5$)												
Designs that satisfy the assumptions												
Baseline	0.374	0.499	0.80	0.016	0.543	0.94	0.016	0.542	0.94	−0.005	0.477	0.95
Lag-dependent links	0.400	0.508	0.79	0.000	0.550	0.94	0.000	0.550	0.94	0.001	0.472	0.95
Fixed types	0.011	0.321	0.93	0.023	0.454	0.95	0.023	0.454	0.95	0.023	0.433	0.95
Weak network	0.359	0.489	0.81	−0.058	0.721	0.93	−0.026	0.568	0.93	−0.003	0.522	0.93
Designs outside the assumptions												
Idiosyncratic movement	0.352	0.482	0.79	0.122	0.534	0.95	0.122	0.533	0.95	0.144	0.493	0.94
Covariates in links	0.364	0.490	0.80	−0.022	0.641	0.96	−0.012	0.541	0.94	−0.015	0.516	0.95
Sparse network	0.395	0.494	0.79	0.487	1.195	0.87	0.029	0.501	0.96	0.404	0.735	0.91

Notes: $R = 500$ replications. Bias, root mean squared error, and coverage of the nominal 95% interval, computed over the replications with an estimate. HK: within-agent conditional logit of Honoré and Kyriazidou (2000). Proposed: estimator (15). Oracle: the same pairwise estimator with population matches. Class: the class-conditional estimator (19) with classes formed from the same estimated distances and bandwidth. The sparse design uses the 0.10 quantile of the estimated distances as h_1, and all other finite-type designs use $h_1 = 0.5\,\bar{\epsilon}\,(n/500)^{-1/4}$ with $\bar{\epsilon}$ the population separation gap of the design (Section 7.2). The covariate-link and sparse designs violate the maintained network model (covariates in the link function and bounded degrees, respectively), and the idiosyncratic design violates Assumption 2(i). Failed fits: Sparse network at $n = 2000$: Proposed 2 (2 optimizer failures, all certified as separated).

to between 0.03 and 0.06 at $n \geq 2000$, within 0.05 of the oracle with population distances at every sample size, while the within-agent estimator's bias stays between 0.30 and 0.44 (Tables 19 and 20). Coverage, conditional on a returned interval, is between 0.95 and 0.98, but without a distribution theory it describes this design only. The class estimator also has a smaller RMSE than the pairwise estimator, but the gain comes from pooling: its connected components are not classes of equal movement, and a single class without network information does as well or better (RMSE for β of 0.867 against 1.108 at $n = 500$ and 0.334 against 0.330 at $n = 3000$, with the class values from Table 17 and the single-class values not tabulated).

Summary. When its assumptions hold, the proposed estimator agrees with the oracle, its bias shrinks with n, and its intervals have close to nominal coverage among successful

fits. When these assumptions hold and the heterogeneity moves, it and the class-conditional estimator are the only feasible estimators compared whose biases vanish, at the cost of precision in small samples. In the baseline and lag-dependent designs, conditioning on network classes reduces the RMSE by 6 to 20 percent. The method loses precision when the network is weakly informative or covariates enter link formation, and it fails when the movement is not common to network types or the network is sparse.

7.4 A Data-Driven Bandwidth

Table 7 replaces the calibrated bandwidth by the rule (16), with the same seeds, so the oracle rows reproduce Tables 1, 2, 6, and 14. In the baseline design, $\hat{h}_1$ lies between 0.11 and 0.33 times the separation gap in every replication, so the selected pairs are all of the same type, and from $n = 1000$ almost all same-type pairs are selected (recall at least 0.99): the pairwise estimator is within 0.008 of the oracle in bias and RMSE, and the class partition is exact in every replication. At $n = 500$ it returns 426 estimates, against 431 with the calibration, with an RMSE for β of 0.715, against 0.677. In the weak-network design, whose gap is ten times smaller, the noise of the estimated distances is of the order of the gap: at $n = 500$ and 1000 the rule selects only 28 and 39 percent of the exact matches, and the pairwise estimator returns 33 and 322 estimates (none with the calibration at $n = 1000$). At $n = 2000$, every replication returns one, with an RMSE for β of 0.304, against 0.362 for the calibration. Across the three sample sizes, the class partitions merge two classes in 3.6 to 7.8 percent of the weak-design replications. The constant matters in small samples: halving it leaves 352 estimates in the baseline design at $n = 500$ and none in the weak design at $n \leq 1000$, and doubling it admits false matches in the weak design (precision 0.66 at $n = 500$). The rule works where the gap is large relative to the noise of the estimated distances, which is what the theory requires.

8 Empirical Application

This section takes the method to longitudinal data on adolescent substance use in a Scottish secondary school, where friendship networks and behavior are observed at three survey waves. The exercise is a feasibility study rather than an estimation of the identified model. The data fall short of the model's requirements in ways documented below: the initial outcome is not observed, the covariate-stability condition cannot be imposed, the network is sparse and its observation process is not the dyadic one of Assumption 1, and at every localized bandwidth examined, the conditional likelihood is completely separated. The conditional

Table 7: Data-driven network bandwidth: baseline and weak-network designs

			β			α			
n	Estimator	Fits	Bias	RMSE	Cov.	Bias	RMSE	Cov.	Merge
Baseline									
500	Pairwise	426	0.256	0.715	0.965	−0.060	1.396	0.955	–
	Class	430	0.127	0.549	0.963	−0.049	1.132	0.991	0.0
	Oracle	431	0.235	0.670	0.968	−0.075	1.297	0.961	–
1000	Pairwise	500	0.152	0.457	0.950	0.003	0.880	0.934	–
	Class	500	0.084	0.381	0.974	0.002	0.757	0.964	0.0
	Oracle	500	0.151	0.459	0.950	0.000	0.872	0.940	–
2000	Pairwise	500	0.070	0.263	0.942	0.016	0.546	0.938	–
	Class	500	0.042	0.222	0.960	−0.005	0.477	0.952	0.0
	Oracle	500	0.069	0.262	0.944	0.016	0.542	0.942	–
Weak network									
500	Pairwise	33	−0.087	0.850	0.970	−0.328	1.373	1.000	–
	Class	45	0.012	0.498	1.000	0.015	1.185	0.978	5.4
	Oracle	425	0.238	0.665	0.967	−0.020	1.393	0.951	–
1000	Pairwise	322	0.278	0.787	0.941	0.117	1.657	0.922	–
	Class	341	0.100	0.423	0.947	0.094	0.881	0.962	7.8
	Oracle	500	0.133	0.451	0.950	0.018	0.849	0.948	–
2000	Pairwise	500	0.083	0.304	0.934	−0.051	0.629	0.938	–
	Class	500	0.043	0.239	0.942	−0.001	0.516	0.944	3.6
	Oracle	500	0.064	0.268	0.946	−0.026	0.568	0.930	–

Notes: $R = 500$ replications per cell. Pairwise and Class: estimators (15) and (19) with $h_1 = \hat{h}_1$ of (16). Oracle: the pairwise estimator with population matches. $h_2 = 0.5$ (exact covariate stability). Fits: replications with an estimate. Bias, RMSE, and coverage of the nominal 95% interval are computed over them, with the standard errors of Section 7.2. Merge: percentage of replications in which a feasible class contains agents of different population classes.

estimates reported below are therefore descriptive alternative fits, and no inferential claims are attached to them.

8.1 Data

The data come from the Teenage Friends and Lifestyle Study (TFLS), a longitudinal survey of adolescent health behavior and social networks conducted at a secondary school in Glasgow, Scotland (Michell and West, 1996; Pearson and Michell, 2000).[1] The study tracked 160 students from an older cohort across three waves of data collection: wave 1 in February 1995, wave 2 in January 1996, and wave 3 in January 1997. Of these, 129 were present at all three waves. After restricting to students with complete outcome and covariate data across all waves, the estimation sample contains $n = 96$ students and $n \times T = 288$ student-wave

[1]The data are publicly available at `https://www.stats.ox.ac.uk/~snijders/siena/Glasgow_data.htm`.

observations. The networks used below are the subgraphs induced by these 96 students: nominations to and from the other 64 students of the cohort are discarded. This complete-case construction changes the observed neighborhoods and the distribution of partners against which codegrees are computed, and the paper makes no correction for it. A missing-node or nonresponse model would be needed to assess its consequences.

These data have been used extensively to study the coevolution of friendship networks and substance use with stochastic actor-oriented models, which separate the selection of friends on behavior from the influence of friends' behavior (Steglich et al., 2006, 2010). Those models specify network and behavior dynamics jointly and parametrically, with actors making decisions in continuous time between waves, and their object of interest is peer influence. The model of this paper asks a different question: it contains no influence of friends' outcomes, and it uses the network only as information about a latent own characteristic that confounds the effects of the covariates and of past behavior. The two approaches rest on different assumptions and address different parameters. Studies using these models find that friends' smoking may influence a student's smoking (Steglich et al., 2010; Schaefer et al., 2012). If it does and links are homophilous on lagged smoking, the omitted influence is correlated with the lagged outcome and is not common within a network class, so the estimate of α absorbs part of it. The state dependence estimated here is state dependence in a model without contagion.

The setting motivates the framework, although, as this section documents, the data fall short of several of its requirements. Adolescent smoking is thought to be state dependent, through habit formation, nicotine dependence, and reduced social stigma (Fletcher, 2010; Nakajima, 2007). Unobserved characteristics such as risk attitude, impulsivity, or attachment to conventional norms also evolve during adolescence and affect both substance use and friendship formation: more risk-tolerant students may be more likely both to smoke and to befriend other risk-taking peers. The proposed estimator targets this confounding pathway between network structure and smoking.

8.2 Variables

Table 8 reports summary statistics for the estimation sample across the three survey waves.

The binary outcome y_{it} equals one if student i reports any tobacco use (occasional or regular) or any cannabis use (tried once, occasional, or regular) at wave t and zero otherwise. The composite is hard to interpret in a dynamic model: the cannabis category "tried once" can record past initiation rather than current use, so a value of one can persist mechanically. The estimand is therefore persistence in the recorded composite status, and the exercise

Table 8: Descriptive statistics: Glasgow TFLS sample.

	Wave 1 (1995)		Wave 2 (1996)		Wave 3 (1997)		Range
	Mean	SD	Mean	SD	Mean	SD	all waves
Panel A: Outcome and covariates							
Smoking	0.417	0.496	0.521	0.502	0.708	0.457	0–1
Pocket money	9.67	7.51	11.70	8.36	16.21	10.40	0–50
Romantic relationship	0.302	0.462	0.229	0.423	0.323	0.470	0–1
Female	0.458	0.501	0.458	0.501	0.458	0.501	0–1
Age	13.34	0.30	14.34	0.30	15.34	0.30	12.8–16.3
Alcohol	2.781	1.038	3.062	1.064	3.479	0.894	1–5
Leisure	2.616	0.308	2.748	0.260	2.736	0.279	2–4
Smoking at home	0.594	0.494	0.594	0.494	0.594	0.494	0–1
Panel B: Friend network							
Density	0.0408		0.0390		0.0373		
Degree	3.88	2.17	3.71	1.90	3.54	1.80	0–10
Students				96			
Waves				3			
Observations ($n \times T$)				288			

The range column reports the minimum and maximum over all three waves. Smoking equals one if the student reports any tobacco use (occasional or regular) or any cannabis use (tried once, occasional, or regular). Pocket money is the monthly amount in British pounds. Romantic relationship equals one if the student reports being in a relationship. Female equals one for girls. Age is recorded in years with decimal precision (computed from month of birth), measured at wave 1, and incremented by one year per wave. Alcohol is an ordinal scale: 1 (non-drinker), 2 (once or twice a year), 3 (once a month), 4 (once a week), and 5 (more than once a week). Leisure is the average across fifteen leisure activity items, each coded from 1 (most days) to 4 (less often or never), so a higher value means less frequent activity. Smoking at home equals one if the student reports that someone smokes at home. It is coded as the first non-missing report across the three waves and is therefore time-invariant by construction. Networks are symmetrized using the OR rule. Degree is the number of undirected links per student.

makes no claim to isolate addiction or habit formation. The tobacco-only outcome is reported as a robustness check in Table 11. Prevalence in the estimation sample rises from 41.7% at wave 1 to 52.1% at wave 2 and 70.8% at wave 3, consistent with rapid uptake during early adolescence.

The identification strategy requires $T \geq 3$ modeled periods plus an initial condition y_{i0}, which I set to zero for all students, using the three waves as $t = 1, 2, 3$. Students are 13 at wave 1, an age at which regular smoking is uncommon, but the assumption is substantive: 41.7% already report some use at wave 1, so imposing universal non-use one year earlier attributes all wave-1 use to initiation within that year. Because y_{i0} enters the state dependence

regressor directly, misclassification of the initial condition can affect $\hat{\alpha}$. This is the classical initial conditions problem of Heckman (1981b). Assumption 1 treats the initial outcome as observed and independent of the later outcome shocks, and neither property can be checked here.

Four time-varying covariates enter the proposed estimator. *Pocket money* is self-reported monthly disposable income in pounds, rising on average from 9.67 at wave 1 to 16.21 at wave 3, and captures the capacity to buy tobacco or cannabis. The survey's "don't know" code (-1) is treated as missing, while zero is a valid response. *Romantic relationship* indicates being in a relationship, between 23% and 32% across waves, and captures social involvement and maturity. *Alcohol* is an ordinal five-point scale from 1 (non-drinker) to 5 (more than once a week), rising from 2.78 to 3.48. Drinking and smoking co-occur strongly in adolescent populations and may reflect common risk factors. *Leisure* is the mean reported frequency across the fifteen out-of-school leisure items in the questionnaire (sports, reading, cinema, and so on), not a validated measure of structured activity. Each item is coded from 1 (most days) to 4 (less often or never), so that a higher value means less frequent participation, and the mean is stable between 2.62 and 2.75 across waves. Pocket money and leisure take 47 and 27 distinct values in the estimation sample, with 40 of the 288 pocket-money observations non-integer, so both are effectively continuous rather than recorded on a coarse discrete scale.

Assumption 1(iv) requires the entire covariate path to be independent of the outcome shocks. This is implausible for several of these covariates during adolescence: alcohol use, romantic involvement, leisure activity, and pocket money can all respond to past substance use, and the latent characteristic may respond to it as well. Such feedback violates the strict exogeneity on which the path factorization of Lemma 1 rests. It is a substantive identification concern, and a larger sample would not resolve it.

Time variation is required by the conditional likelihood approach: the pairwise differencing that eliminates $\lambda(w_{it})$ also differences the covariates, so any time-invariant characteristic drops out. The naive and controls specifications of Appendix E, which do not rely on that differencing, add three covariates the proposed estimator cannot use: *female* (45.8% of the sample), *age* (mean 13.34 at wave 1, incremented by one year per wave, so it cancels exactly under differencing), and *smoking at home*, an indicator for whether someone smokes at the student's home, coded from the first non-missing report.

8.3 Network Construction

At each wave, students nominated up to six friends within the cohort, with nominations coded as 1 (best friend) or 2 (friend). The analysis uses the "any friend" network, which

counts a link whenever at least one of the two students nominated the other at level 1 or 2. The directed adjacency matrix is symmetrized using the OR rule: $D_{ijt} = \max(D^{\text{dir}}_{ijt}, D^{\text{dir}}_{jit})$. The resulting undirected networks have densities of approximately 0.04 at each wave, with an average degree around 3.5 to 3.9 links per student.

The observation process departs from the network model of equation (2) in three ways. First, the model takes links to be undirected and formed from symmetric dyadic shocks that are conditionally independent across dyads (Assumption 1(ii)). The six-nomination cap violates this independence: one outgoing nomination constrains another, so the links of a student are dependent even given the latent characteristics, and the problem is not only that high-degree students' neighborhoods are truncated. Restricting attention to reciprocated nominations would not remove this dependence and would make the network sparser. Symmetrizing directed nominations by the OR rule does not by itself break dyadic independence, but it changes what a link records: an unreciprocated nomination, whether or not it reflects a friendship, becomes a link. Second, the asymptotic theory of Section 6 is that of a fixed dense graphon, in which degrees grow in proportion to the number of students, and the bounded-degree and capped-nomination features of these data are not covered by that asymptotic experiment. Average degree here is below four in the any-friend network and below one in the best-friend network (0.81, 0.67, and 0.81 links per student across the three waves, so its codegree distances rest on very little linking information), so the approximation that underlies the codegree theory is poor, and a larger school with degrees that remain bounded would not by itself deliver more informative codegrees. Third, the networks are induced subgraphs of the complete-case sample, as noted above. None of these problems enters the unlocalized conditional columns of Table 10, which do not use the network. All of them enter the localized estimates of Table 9, where the sparsity is part of why so few pairs fall inside the kernel window.

8.4 Feasibility of the Estimator

The outcome equation takes the form

$$y_{it} = \mathbb{1}\{X'_{it}\beta + \alpha\, y_{it-1} + \lambda(w_{it}) + \mu_i - \varepsilon_{it} \geq 0\}, \quad t = 1, 2, 3,$$

where X_{it} contains pocket money, romantic relationship, alcohol, and leisure; y_{it-1} captures state dependence in smoking; μ_i is a student fixed effect, which may stand for a physiological susceptibility to nicotine or for persistent features of the home environment; and $\lambda(w_{it})$ is an unknown function of the latent social characteristic w_{it} that also governs friendship formation. The unobserved characteristic w_{it} can be interpreted as an evolving index of risk attitude

or social orientation, which affects both substance use and the tendency to befriend other risk-taking students.

Appendix E reports five descriptive fits on these data: a naive dynamic logit, a dynamic logit with peer-average controls, a within-agent conditional fit in the spirit of Honoré and Kyriazidou (2000), an unlocalized cross-agent pairwise fit, and an unlocalized class-conditional fit. Two departures from the estimator of Section 5 are forced by these data, and together they mean that the conditional fits do not estimate the identified model. First, the covariate-stability kernel K_2 is infeasible: none of the 22 switching students has stable covariates between waves 2 and 3, and any meaningful bandwidth leaves at most one weighted pair, so the conditional fits omit it. Without it, the log odds of equation (6) retain the term $\Gamma(c_{i3}) - \Gamma(c_{i2})$, which is bounded in absolute value by $|c_{i3} - c_{i2}|$ but is not shown to be negligible in these data and which depends on the level of the individual effect and of λ through c_{i2}. Second, the unlocalized pairwise fit sets $h_1 = \infty$, so that all 96 discordant lag-matched pairs receive the same weight. The largest aggregated codegree distance among these pairs is 1.8×10^{-4}, so a bandwidth that is large relative to it, such as $h_1 = n^{-1/5} \approx 0.40$, gives weights that differ from equal weights by less than one part in a million and the same fit, whereas bandwidths of the order of the distances localize. The localized fits, with quantile bandwidths, are those of Table 9. The unlocalized fit therefore does not use the network, is numerically identical across network definitions, and does not remove the difference between the two agents' common movements, which need not vanish without matching. At the same unlocalized bandwidth, the feasible classes of the class-conditional estimator of Section 5.3 are the lag strata, and since $y_{i0} = 0$ for every student there is a single class of 22 switchers: the class-conditional fit is then the within-agent conditional logit with one common movement of λ for all students, conditioned out, and it and the unlocalized pairwise fit are the full and the pairwise conditional likelihoods of the same class. With the stability kernel omitted, the Γ remainder generally still depends on the level of the individual effect and of λ, so none of the conditional fits delivers the cancellation of Lemma 1, and differences between their coefficients are differences between objectives.

Table 9 reports what happens when the codegree kernel actually binds: the bandwidth is set to the q-th quantile of the finite aggregated codegree distances, so that roughly a fraction q of the discordant lag-matched pairs receives positive weight. This is the network-localized version of the pairwise fit, still omitting the K_2 condition of Section 5, and the table shows what it can and cannot do at $n = 96$. The central diagnostic finding is that at every localized bandwidth reported in the table (the 0.10, 0.20, and 0.30 quantiles with the any-friend network and the 0.10 quantile with the best-friend network, but not the 0.05 quantile, which leaves only five pairs), the weighted conditional likelihood is completely separated (Albert

and Anderson, 1984): a unit-norm direction orders every weighted pair correctly (Panel A), and at the attempted maximum the share of pair-level fitted probabilities outside $[0.01, 0.99]$ is 1.000 in every column, with maximal fitted indices between 38 and 67. Under complete separation the likelihood has no maximizer, its supremum is approached only as $\|\theta\| \to \infty$, the point at which a quasi-Newton optimizer stops is arbitrary, and the associated standard errors invert a numerically singular information matrix. The table therefore does not report unpenalized point estimates, which do not exist as maximizers. It reports the separation diagnostics together with the Firth-penalized estimates (Firth, 1993), which provide finite estimates for the penalized specifications reported here. In the localized fits the penalized criterion is the kernel-weighted log likelihood plus one half of the log determinant of its weighted information, with the raw kernel weights $K(\hat{\tilde{\delta}}_{ij}^2/h_1)$, not weights normalized to sum to one or to the number of pairs. Because the weight of the penalty relative to the fit depends on this scale, penalized coefficients from objectives with different weight scales, such as the localized and unlocalized fits, are not directly comparable.

The table is reported in full because its message is part of the paper's argument. At the localized bandwidths of the table, the effective sample size drops to between 7 and 23, and the likelihoods are completely separated: with so few weighted pairs, some linear combination of the regressors perfectly predicts the discordant ordering, the MLE does not exist, and only penalized estimates are finite. The Firth-penalized estimates are correspondingly modest and unstable across bandwidths (only alcohol and leisure keep a consistent sign across all four columns, and the state dependence coefficient changes sign across them), with standard errors that make none of the localized estimates informative. The conclusion, backed by the separation diagnostics rather than by erratic point estimates, is that $n = 96$ with a sparse, capped network does not support the network-localized estimator at any of the bandwidths examined. The class-conditional estimator of Section 5.3 does no better at the same bandwidths: at each of them only one feasible class contains both orderings, with between 9 and 15 of the 22 switchers, and its likelihood with the class intercept is completely separated, with unit-norm margins between 0.20 and 0.34. The Glasgow data can be used to compute the unlocalized conditional fits of Appendix E, which omit both the covariate-stability condition and the network matching, but not to implement the codegree localization that the method requires. The application therefore documents the data requirements of the method rather than its performance, and nothing in it validates the procedure. Two of the four covariates (romantic relationship and alcohol) are recorded on coarse discrete scales, but pocket money and leisure are effectively continuous and, more to the point, no switching student keeps them unchanged between waves 2 and 3, so the atom of exact covariate stability that the finite-type regime of Theorem 3 requires is absent in these data, even setting the

Table 9: Network-localized estimates: separation diagnostics and penalized fits.

	Any friend			Best friend
	$q = 0.10$	$q = 0.20$	$q = 0.30$	$q = 0.10$
Panel A: Separation diagnostics (unpenalized likelihood)				
Share of fitted prob. outside $[0.01, 0.99]$	1.000	1.000	1.000	1.000
Max |fitted index| at optimizer stop	38.2	67.2	51.7	53.9
Max. min. signed margin ($\|b\| = 1$)	1.540	0.566	0.483	0.729
MLE exists	No	No	No	No
Panel B: Firth-penalized estimates				
Pocket money	−0.005 (0.047)	−0.009 (0.040)	−0.004 (0.038)	0.009 (0.193)
Romantic relationship	−0.658 (2.481)	−0.388 (1.765)	−0.786 (1.628)	0.126 (1.061)
Alcohol	0.317 (0.513)	0.625 (0.413)	0.824 (0.397)	0.847 (0.600)
Leisure	0.867 (2.797)	0.820 (1.975)	1.502 (1.820)	0.179 (3.188)
Lagged smoking ($\hat{\alpha}$)	0.176 (1.514)	−0.217 (0.996)	−0.403 (0.980)	0.876 (1.184)
Pairs with positive weight	9	21	26	14
Pair-weight ESS	7.0	17.1	23.0	14.0

h_1: q-th quantile of the finite aggregated lag-stratified codegree distances. K_2 is omitted (see text). ESS $= (\sum_p w_p)^2 / \sum_p w_p^2$ over the weighted pairs, not a count of independent students. Panel A: evaluated where the BFGS optimizer stops on the unpenalized likelihood. The signed-margin row is the largest minimum margin of a unit-norm direction over the pair rows, oriented so that the realized ordering is positive, and a positive value certifies complete separation (see text). Panel B: Firth-penalized estimates (Firth, 1993), with the unpenalized sandwich at the penalized estimates in parentheses, for scale only. No stars: the effective samples are too small for asymptotic approximations.

network aside. And a suitable dataset would combine at least four waves, so that the initial outcome is observed, covariates recorded on discrete scales or with enough stable observations to impose the covariate-stability condition, and a network observed without a nomination cap whose degrees are large enough for codegrees to be informative. Such data, collected at close and evenly spaced dates around a scheduled transition as described in Appendix D, could address the obstacles documented here, but they would not by themselves validate the method: the overlap of network types among switchers, the common movement of Assumption 2, the strict exogeneity of the covariates, and the rank condition on the selected pairs would still have to be assessed, and none of them is a matter of sample size alone.

9 Conclusion

This paper studies a dynamic logit model in which a binary outcome depends on covariates, the past outcome, a fixed effect, and the social influence function, an unknown function of a time-varying unobserved characteristic that also drives link formation. Covariate effects and state dependence are identified when the change of that function between two waves is common to agents of the same network type and is followed by a stable wave: conditioning on an agent's own switch removes the fixed effect, and comparing agents of the same network type removes the change.

The pairwise conditional likelihood estimator weights pairs of switchers by their estimated codegree distance and the stability of their covariates. It is consistent, including with continuous types, and root-n asymptotically normal when the characteristic takes finitely many values and covariate changes have an atom at zero. With three waves, the same holds for a bandwidth computed from the first-wave network. In that case, a second estimator conditions on network classes of switchers. It is root-n asymptotically normal with one transition period and, when $T = 3$, at least as efficient asymptotically as the pairwise estimator, by an argument sketched in the paper.

In simulations that satisfy the assumptions, the pairwise estimator agrees with an oracle that knows the true matches, its bias falls with the sample size, and its intervals have close to nominal coverage among successful fits, while pooled logits and the within-agent estimator remain biased when the characteristic moves. The class estimator reduces the root mean squared error of the pairwise estimator by 6 to 20 percent in the two main designs. The method fails when the movement is not common to network types or when the network is sparse. In the Glasgow Teenage Friends and Lifestyle Study, the network-localized likelihood is completely separated at every bandwidth examined, so the application documents the data the method requires rather than its performance.

Three extensions remain open: a distribution theory for continuous types, which must handle the dependence created by shared links, the sampling of latent types, and heterogeneous match probabilities; feedback from outcome shocks into later covariates and into the latent characteristic; and directed, mismeasured, or multiple networks.

A Appendix

This appendix collects the proofs of the main results. I begin with the proof of the within-agent likelihood ratio and a detailed derivation of the identification argument for $T = 3$, then establish a lemma on the empirical codegree distance and use it to complete the proofs

of Proposition 1, Theorem 1, and Corollary 1. I then prove consistency (Theorem 2) and asymptotic normality in finite-type designs (Theorem 3), establish a lemma on conditional sums of Bernoulli variables, and prove Proposition 2 and Corollary 2. What is and is not identified about λ is discussed in Appendix B, the extension to ordered outcomes in Appendix C, the observation window of the smoking example in Appendix D, and descriptive fits for the Glasgow data in Appendix E. The supplementary tables are in Appendix F.

A.1 Proof of Lemma 1

Proof. By Assumption 1(iv), the shocks $(\varepsilon_{i1}, \dots, \varepsilon_{iT})$ are i.i.d. logistic and independent of $\chi_i = (y_{i0}, X_i, w_i, \mu_i)$. Since $y_{it} = \mathbb{1}\{X_{it}'\beta + \alpha y_{i,t-1} + \lambda(w_{it}) + \mu_i - \varepsilon_{it} \geq 0\}$, conditional on χ_i the outcomes form a Markov chain with transition probabilities

$$\mathbb{P}(y_{it} = 1 \mid y_{i0}, \dots, y_{i,t-1}, X_i, w_i, \mu_i) = F\big(X_{it}'\beta + \alpha y_{i,t-1} + \lambda(w_{it}) + \mu_i\big),$$

and conditioning on the entire covariate and latent paths, including their future values, does not alter these probabilities. Hence the probability of a path factorizes into per-period transition probabilities, with the initial outcome held fixed. Write $\pi_t(y \mid y')$ for the probability of outcome y at t given lag y'. For path A_s,

$$\begin{aligned}\mathbb{P}(y_i \in A_s \mid \chi_i) = &\left[\prod_{\substack{t=1 \\ t\neq s,s+1,s+2}}^{T} \pi_t(y_{it} \mid y_{i,t-1})\right] \big(1 - F(X_{is}'\beta + \alpha y_{i,s-1} + \lambda(w_{is}) + \mu_i)\big) \\ &\times F\big(X_{i,s+1}'\beta + \lambda(w_{i,s+1}) + \mu_i\big)\, \pi_{s+2}(y_{i,s+2} \mid 1),\end{aligned}$$

where the contribution at period s reflects $y_{is} = 0$ with lag $y_{i,s-1}$, the contribution at $s+1$ reflects $y_{i,s+1} = 1$ with lag 0, and the contribution at $s+2$ reflects the outcome $y_{i,s+2}$ with lag 1. For path B_s,

$$\begin{aligned}\mathbb{P}(y_i \in B_s \mid \chi_i) = &\left[\prod_{\substack{t=1 \\ t\neq s,s+1,s+2}}^{T} \pi_t(y_{it} \mid y_{i,t-1})\right] F\big(X_{is}'\beta + \alpha y_{i,s-1} + \lambda(w_{is}) + \mu_i\big) \\ &\times \big(1 - F(X_{i,s+1}'\beta + \alpha + \lambda(w_{i,s+1}) + \mu_i)\big)\, \pi_{s+2}(y_{i,s+2} \mid 0).\end{aligned}$$

The products over $t \notin \{s, s+1, s+2\}$ coincide, because the two paths agree outside periods s and $s+1$ and the lags entering those periods are the same. Let $\gamma_s = X_{is}'\beta + \alpha y_{i,s-1} + \lambda(w_{is}) + \mu_i$, $c = X_{i,s+1}'\beta + \lambda(w_{i,s+1}) + \mu_i$, and $c' = X_{i,s+2}'\beta + \lambda(w_{i,s+2}) + \mu_i$. Using $F(x)/(1 - F(x)) = e^x$,

the ratio of the period-s terms is e^{γ_s}, and the ratio of the terms at $s+1$ and $s+2$ is

$$\frac{1+e^{c}}{e^{c}(1+e^{c+\alpha})}\cdot\frac{e^{y_{i,s+2}c'}\,(1+e^{c'+\alpha})}{(1+e^{c'})\,e^{y_{i,s+2}(c'+\alpha)}}=\exp\big[-c-\alpha y_{i,s+2}+\Gamma(c')-\Gamma(c)\big],$$

with $\Gamma(x)=\log(1+e^{x+\alpha})-\log(1+e^{x})$. Under $X_{i,s+2}=X_{i,s+1}$ and $\lambda(w_{i,s+2})=\lambda(w_{i,s+1})$, $c'=c$ and the Γ terms cancel. Combining and noting that γ_s-c does not involve μ_i,

$$\frac{\mathbb{P}(y_i\in B_s\mid\chi_i)}{\mathbb{P}(y_i\in A_s\mid\chi_i)}=\exp\big[(X_{is}-X_{i,s+1})'\beta+\alpha(y_{i,s-1}-y_{i,s+2})+\lambda(w_{is})-\lambda(w_{i,s+1})\big].\qquad\square$$

The display before the last step also gives the exact log odds without any stability condition, which is equation (6) for $T=3$. Because $\Gamma'(x)=F(x+\alpha)-F(x)\in(-1,1)$, the remainder satisfies $|\Gamma(c')-\Gamma(c)|\le|c'-c|$. This bound is used in the proof of Theorem 2.

A.2 Derivation for $T=3$

For concreteness, I write out the argument of Lemma 2 and Theorem 1 for $T=3$, so that $\mathcal{S}=\{1\}$. The paths are

$$A=\{y_{i0},0,1,y_{i3}\},\qquad B=\{y_{i0},1,0,y_{i3}\},$$

with y_{i0} and y_{i3} held at their realized values. Write $c_{it}=X_{it}'\beta+\lambda(w_{it})+\mu_i$. Conditional on χ_i,

$$\begin{aligned}\mathbb{P}(y_i\in A\mid\chi_i)&=\frac{1}{1+e^{c_{i1}+\alpha y_{i0}}}\cdot\frac{e^{c_{i2}}}{1+e^{c_{i2}}}\cdot\frac{e^{y_{i3}(c_{i3}+\alpha)}}{1+e^{c_{i3}+\alpha}},\\ \mathbb{P}(y_i\in B\mid\chi_i)&=\frac{e^{c_{i1}+\alpha y_{i0}}}{1+e^{c_{i1}+\alpha y_{i0}}}\cdot\frac{1}{1+e^{c_{i2}+\alpha}}\cdot\frac{e^{y_{i3}c_{i3}}}{1+e^{c_{i3}}}.\end{aligned}$$

Taking the ratio and using $F(x)/(1-F(x))=e^{x}$ gives equation (6): the agent effect cancels in $c_{i1}-c_{i2}=(X_{i1}-X_{i2})'\beta+\lambda(w_{i1})-\lambda(w_{i2})$ and enters only the bracket $\Gamma(c_{i3})-\Gamma(c_{i2})$, through the levels. If $X_{i3}=X_{i2}$ and $\lambda(w_{i3})=\lambda(w_{i2})$, then $c_{i3}=c_{i2}$ and Lemma 1 gives

$$\frac{\mathbb{P}(y_i\in B\mid\chi_i)}{\mathbb{P}(y_i\in A\mid\chi_i)}=\exp\big[(X_{i1}-X_{i2})'\beta+\alpha(y_{i0}-y_{i3})+\lambda(w_{i1})-\lambda(w_{i2})\big].$$

Take a second agent j with the same stability conditions and with $\lambda(w_{i1})-\lambda(w_{i2})=\lambda(w_{j1})-\lambda(w_{j2})$, which is the event $\mathcal{C}_{ij1}\cap\mathcal{X}_{ij1}$. The levels $\lambda(w_{i1})$ and $\lambda(w_{j1})$ need not be equal, and neither need the agent effects μ_i and μ_j. Since the two agents' paths are independent

given their primitives, with $\Delta X_a = X_{a1} - X_{a2}$,

$$\frac{\mathbb{P}(y_i \in B \mid \cdot)\,\mathbb{P}(y_j \in A \mid \cdot)}{\mathbb{P}(y_i \in A \mid \cdot)\,\mathbb{P}(y_j \in B \mid \cdot)} = \exp\big[(\Delta X_i - \Delta X_j)'\beta + \alpha(y_{i0} - y_{i3} - y_{j0} + y_{j3})\big],$$

because the two changes of λ cancel. Given the discordance event $\mathcal{D}_{ij1}$ and the outcomes at periods 0 and 3, only the configurations $A \times B$ and $B \times A$ are possible, so

$$\begin{aligned}\mathbb{P}\big(y_i \in A, y_j \in B \,\big|\, \mathcal{F}_{ij1}, w_i, w_j, \mathcal{D}_{ij1}\big) &= F\big[(\Delta X_j - \Delta X_i)'\beta + \alpha(y_{j0} - y_{j3} - y_{i0} + y_{i3})\big] \\ &= F\big(V_{ij1}(\beta, \alpha)\big).\end{aligned}$$

Under Assumption 2, the event $\mathcal{C}_{ij1}$ is implied by network-type equivalence at period 1 in the stratum of the shared initial outcome together with $d_2 \equiv 0$, which is the content of Theorem 1 for $T = 3$.

A.3 The Empirical Codegree Distance

Throughout the remainder of the appendix, fix a period t and write $\xi_l = (w_{lt}, y_{l,t-1})$ for agent l's type at t, with distribution G_t. By Assumption 1, the ξ_l are i.i.d. across agents, and, conditional on $\xi = (\xi_1, \ldots, \xi_n)$, the links D_{klt} with $k < l$ are independent with means $f_{klt} = f(w_{kt}, w_{lt};\, y_{k,t-1}, y_{l,t-1})$, because the link shocks are independent of the outcome shocks and of the primitives. The diagonal of the adjacency matrix is zero. For a pair (i, j) with $y_{i,t-1} = y_{j,t-1} = y$, write $g_l = f_{ilt} - f_{jlt}$ and let $p_{ikt} = \int f(w_{it}, \tau;\, y, y')\, f(w_{kt}, \tau;\, y_{k,t-1}, y')\, dG_t(\tau, y')$ denote the population codegree, so that $\delta^2_{ijt} = \mathbb{E}_k[(p_{ikt} - p_{jkt})^2]$, where $\mathbb{E}_k$ integrates over the type of an independent agent k.

Lemma 4 (Empirical codegree distance)**.** *Suppose Assumption 1 holds. For each period t:*

(a) For any pair (i, j) with $y_{i,t-1} = y_{j,t-1}$, $\hat{\delta}^2_{ijt} \xrightarrow{p} \delta^2_{ijt}$.

(b) If f is continuous, $\rho_{ijt} = 0$ implies $\delta_{ijt} = 0$.

(c) Suppose that f is continuous or that G_t has finite support, and let (w_{it}, y) and (w_{jt}, y) lie in the support of G_t. Then $\delta_{ijt} = 0$ if and only if $f(w_{it}, \tau;\, y, y') = f(w_{jt}, \tau;\, y, y')$ for every (τ, y') in the support of G_t. Under Assumption 6, $\delta_{ijt} = 0$ if and only if $\rho_{ijt} = 0$.

(d) Uniformly over pairs,

$$\max_{i \neq j} \big|\hat{\delta}^2_{ijt} - \delta^2_{ijt}\big|\, \mathbb{1}\{y_{i,t-1} = y_{j,t-1}\} = O_p\big((\log n / n)^{1/2}\big),$$

and, for pairs that are equivalent on the support, $\max_{i\neq j} \hat{\delta}^2_{ijt}\,\mathbb{1}\{g_l = 0 \text{ for all } l\} = O_p\big(\log n/n\big)$. *The constants in these bounds do not depend on the law of the primitives, so the bounds hold uniformly along the triangular array of Assumption 3.*

Proof. Decomposition. Fix a pair (i,j) with a shared lag and write $\hat{C}_k = \hat{c}_{ikt} - \hat{c}_{jkt} = n^{-1}\sum_l D_{klt}(D_{ilt} - D_{jlt})$, so that $\hat{\delta}^2_{ijt} = n^{-1}\sum_k \hat{C}^2_k$. The two terms $k \in \{i,j\}$ share a dyad with the inner products and are bounded by one, so they contribute at most $2/n$. For $k \notin \{i,j\}$, the inner terms with $l \in \{i,j,k\}$ are bounded by one and contribute at most $2/n$ to $\hat{C}_k$. The remaining inner terms, $l \notin \{i,j,k\}$, involve the distinct dyads (k,l), (i,l), and (j,l), so, conditional on ξ, they are independent across l with means $f_{klt}\,g_l$. Hence

$$\hat{C}_k = a_k + \nu_k + r_k, \qquad a_k = \frac{1}{n}\sum_{l\notin\{i,j,k\}} f_{klt}\,g_l, \qquad \nu_k = \frac{1}{n}\sum_{l\notin\{i,j,k\}} \big[D_{klt}(D_{ilt} - D_{jlt}) - f_{klt}\,g_l\big],$$

with $|r_k| \le 2/n$, $|a_k| \le 1$, and $\mathbb{E}[\nu_k \mid \xi] = 0$, and

$$\hat{\delta}^2_{ijt} = \underbrace{\frac{1}{n}\sum_{k\notin\{i,j\}} a_k^2}_{\text{I}} + \underbrace{\frac{2}{n}\sum_{k\notin\{i,j\}} a_k\nu_k}_{\text{II}} + \underbrace{\frac{1}{n}\sum_{k\notin\{i,j\}} \nu_k^2}_{\text{III}} + R_{ij}, \qquad |R_{ij}| \le \frac{C}{n}.$$

The inner averages ν_k share observations across k, since each involves agent i's and agent j's links with every l. The argument below never uses independence across k, only bounds that hold simultaneously for all k. I carries the sampling variation of the latent types, and II and III carry the link noise.

Link noise. Conditional on ξ, ν_k is an average of at most n independent, centered variables with range at most 2, so Hoeffding's inequality gives $\mathbb{P}(|\nu_k| > x \mid \xi) \le 2\exp(-nx^2/8)$ for n large. A union bound over the at most n^3 triples (i,j,k) with $x = (32\log n/n)^{1/2}$ gives $\max_{i,j,k}|\nu_k| = O_p((\log n/n)^{1/2})$, with constants that do not depend on ξ or on the law of the primitives. Therefore, uniformly over pairs, $|\text{II}| \le 2\max_k|\nu_k| = O_p((\log n/n)^{1/2})$ and $0 \le \text{III} \le \max_k \nu_k^2 = O_p(\log n/n)$.

Sampling of latent types. Expanding the square,

$$\text{I} = \frac{1}{n^3}\sum_{k\notin\{i,j\}}\ \sum_{l,l'\notin\{i,j,k\}} f_{klt}\,f_{kl't}\,g_l\,g_{l'}.$$

Conditional on (ξ_i,ξ_j), this is, up to terms with a repeated index that contribute $O(1/n)$, a U-statistic of order three in the i.i.d. types of the other agents with a kernel bounded by one.

Its mean is

$$\mathbb{E}\big[f_{klt}f_{kl't}g_lg_{l'}\,\big|\,\xi_i,\xi_j\big]=\mathbb{E}_k\Big[\big(\mathbb{E}_l[f_{klt}\,g_l]\big)^2\Big]=\mathbb{E}_k\big[(p_{ikt}-p_{jkt})^2\big]=\delta_{ijt}^2,$$

where the middle equality uses $\mathbb{E}_l[f_{klt}\,f_{ilt}]=p_{ikt}$, which is the definition of p_{ikt}. The joint symmetry of f (Assumption 1(vi)) is used in the decomposition above, where it gives $D_{klt}=D_{lkt}$ the mean f_{klt} also when $l<k$. Hoeffding's inequality for U-statistics (Hoeffding, 1963) gives, conditional on (ξ_i,ξ_j), $\mathbb{P}(|\mathrm{I}-\delta_{ijt}^2|>x+C/n)\le 2\exp(-c\,nx^2)$ with universal constants, and a union bound over the at most n^2 pairs gives $\max_{i,j}|\mathrm{I}-\delta_{ijt}^2|=O_p((\log n/n)^{1/2})$. Combining the three terms proves the first bound in part (d), which implies part (a).

Exact matches. If the two agents are equivalent on the support, then $g_l=0$ for every agent l, because every agent's type lies in the support of G_t almost surely. Then $a_k=0$ for every k, so $\mathrm{I}=\mathrm{II}=0$ exactly, and $\hat\delta_{ijt}^2=\mathrm{III}+R_{ij}=O_p(\log n/n)$ uniformly over such pairs. The term R_{ij} is non-negative up to $O(1/n)$ and generates an upward bias of order $1/n$, which is within the stated rate.

Part (b). If $\rho_{ijt}=0$, the link functions of the two agents coincide Lebesgue-almost everywhere, hence everywhere by continuity, so $f_{ilt}=f_{jlt}$ for every partner, $p_{ikt}=p_{jkt}$ for every k, and $\delta_{ijt}=0$. Without continuity the implication can fail when G_t has atoms, because the metric (3) ignores Lebesgue-null sets while the codegree does not.

Part (c). Write $g(\tau,y')=f(w_{it},\tau;\,y,y')-f(w_{jt},\tau;\,y,y')$ and $(\mathcal{K}g)(\sigma,y'')=\int f(\sigma,\tau;\,y'',y')\,g(\tau,y')\,dG_t(\tau,y')$, so that $p_{ikt}-p_{jkt}=(\mathcal{K}g)(\xi_k)$ and $\delta_{ijt}^2=\int(\mathcal{K}g)^2\,dG_t$. By definition of g,

$$\begin{aligned}\int g^2\,dG_t&=\int\big[f(w_{it},\tau;\,y,y')-f(w_{jt},\tau;\,y,y')\big]\,g(\tau,y')\,dG_t(\tau,y')\\&=(\mathcal{K}g)(w_{it},y)-(\mathcal{K}g)(w_{jt},y).\end{aligned}$$

If $\delta_{ijt}=0$, then $\mathcal{K}g=0$ G_t-almost everywhere. When G_t has finite support, this means $\mathcal{K}g=0$ at every support point. When f is continuous, $\mathcal{K}g$ is continuous and vanishes on the support. In either case $\mathcal{K}g$ vanishes at the two support points (w_{it},y) and (w_{jt},y), so $\int g^2\,dG_t=0$, and $g=0$ G_t-almost everywhere, hence at every support point under either condition. The converse is immediate. No definiteness property of the kernel f is involved, but some regularity is: if f is discontinuous and G_t is continuous, $\mathcal{K}g$ can vanish almost everywhere without vanishing at the two evaluation points. What the codegree cannot see is a difference between link functions at partner types or lag strata outside the support of G_t: the metric (3) integrates over all $\tau\in[0,1]$ and both strata, whereas the codegree integrates against G_t. Under Assumption 6, G_t has full support and f is continuous, so agreement on

the support is agreement everywhere, and $\delta_{ijt} = 0$ if and only if $\rho_{ijt} = 0$. $\square$

The uniform bounds in part (d) extend to the aggregated distance $\hat{\bar{\delta}}^2_{ij}$ of equation (13), which averages at most T period-specific distances over a set of periods that is observed and is the same for the sample statistic and its population counterpart.

The following consequence of part (c) is used in the proof of Theorem 2. Under Assumptions 3 and 6, write $\delta_{t,n}(u, v;\, y)$ for the population codegree distance under $\mathbb{P}_n$ and $\delta_t(u, v;\, y)$ for its limit under G_t. Because f is uniformly continuous on its compact domain and $G_{t,n}$ converges weakly to G_t, $\delta^2_{t,n} \to \delta^2_t$ uniformly on $[0,1]^2 \times \{0,1\}$. The functions δ^2_t and $\rho_t(u, v;\, y) = \|f_{u;\, y,\cdot} - f_{v;\, y,\cdot}\|_2$ are continuous on the compact set $[0,1]^2$ and, by part (c) and full support, have the same zero set. Hence

$$\tau_n(h) := \sup\big\{\rho_t(u, v;\, y) : t \le T,\ y \in \{0,1\},\ \delta_{t,n}(u, v;\, y)^2 \le h\big\} \longrightarrow 0 \tag{22}$$

as $h \to 0$ and $n \to \infty$.

A.4 Proofs of Proposition 1(b)–(c), Theorem 1(b), and Corollary 1(b)

Proposition 1(b). Under Assumption 6 with a fixed law, Lemma 4(c) gives that $\delta_s(u, v;\, y) = 0$ if and only if $\|f_{u;\, y,\cdot} - f_{v;\, y,\cdot}\|_2 = 0$, and the argument leading to (22), with $G_{t,n} = G_t$ for every n, shows that $\tau(h) \to 0$ as $h \downarrow 0$: the functions δ^2_s and ρ_s are continuous on the compact set $[0,1]^2$ for each y and have the same zero set, so if $\tau(h)$ did not vanish there would be a sequence (u_k, v_k, y_k) with $\delta_s(u_k, v_k;\, y_k)^2 \to 0$ and $\rho_s(u_k, v_k;\, y_k)$ bounded away from zero, and a limit point along a subsequence would have $\delta_s = 0$ and $\rho_s > 0$, a contradiction. The two consequences follow from the definition of ϖ after Assumption 2, since $\|f_{u;\, y,\cdot} - f_{v;\, y,\cdot}\|_2 \le \tau(h)$ whenever $\delta_s(u, v;\, y)^2 \le h$.

Proposition 1(c). Under Assumption 7(i), G_s has finite support, so Lemma 4(c) applies without continuity: $\delta_{ijs} = 0$ if and only if the two link functions agree at every supported partner type and lag stratum, which is equivalence on the support, and Assumption 7(ii) then gives $d_s(w_{is}) = d_s(w_{js})$. Since the support is finite, the set of positive values of δ^2_s over supported pairs is finite, so its minimum is positive, and it is at least the separation gap $\bar{\epsilon}$ of Assumption 7(v), which is the minimum over all periods.

Theorem 1(b). Condition on $(\mathcal{F}_{ijs}, w_i, w_j, \mu_i, \mu_j)$ and on $\mathcal{D}_{ijs}$. By the exact log-odds formula in the proof of Lemma 1, applied to both agents without any stability condition,

$$\mathbb{P}\big(y_i \in A_s,\, y_j \in B_s \,\big|\, \cdot\,\big) = F\big(V_{ijs}(\beta, \alpha) + \Delta_{ijs}\big), \qquad \Delta_{ijs} = \big[d_s(w_{is}) - d_s(w_{js})\big] + R_j - R_i,$$

where $R_a = \Gamma(c_{a,s+2}) - \Gamma(c_{a,s+1})$ with $c_{at} = X'_{at}\beta + \lambda(w_{at}) + \mu_a$ and $\Gamma(c) = \log(1+e^{c+\alpha}) - \log(1+e^c)$. Since $|\Gamma'| < 1$, $|R_a| \leq |c_{a,s+2} - c_{a,s+1}|$, and by Assumption 2(ii), $c_{a,s+2} - c_{a,s+1} = (X_{a,s+2} - X_{a,s+1})'\beta$, in which the agent effect cancels. The value of R_a depends on the level $c_{a,s+1}$, hence on μ_a, but its bound does not. On $\mathcal{O}_{ijs}(h)$, $\|X_{a,s+2} - X_{a,s+1}\| \leq h^{1/2}$ and, by Proposition 1(b), $|d_s(w_{is}) - d_s(w_{js})| \leq \varpi(\tau(h))$, which gives (9). Because $|F(v+\Delta) - F(v)| \leq |\Delta|/4$ and $F(V_{ijs}(\beta,\alpha))$ is $\mathcal{F}_{ijs}$-measurable, integrating over (w_i, w_j, μ_i, μ_j) on $\mathcal{O}_{ijs}(h)$ gives the bound $\varpi_1(h)/4$ on the coarser conditional probability.

For the last statement, note that m_{ijs} records the configuration $(y_i, y_j) \in B_s \times A_s$, the complement of the one in (9) given the discordance event, and that $V_{ijs}(\beta,\alpha) = -Z'_{ijs}\theta$. Hence its conditional probability is $q_{ijs} = 1 - F(V_{ijs}(\beta,\alpha) + \Delta_{ijs}) = F(Z'_{ijs}\theta - \Delta_{ijs})$, with the same remainder Δ_{ijs} as in (9). Only $|\Delta_{ijs}| \leq \varpi_1(h)$ is used below. Write $\ell(p;x) = p\log F(x) + (1-p)\log(1-F(x))$. By iterated expectations, $M_h(b,a) = \mathbb{E}[\ell(q_{ijs};\, Z'_{ijs}(b,a)') \mid \mathcal{O}_{ijs}(h), \mathcal{D}_{ijs}]$, and M_h is concave on Θ. It is strictly concave for small h, because its Hessian is $-\mathbb{E}[F(Z'_{ijs}(b,a)')(1-F(Z'_{ijs}(b,a)'))Z_{ijs}Z'_{ijs} \mid \cdot]$, which is negative definite whenever $\mathbb{E}[Z_{ijs}Z'_{ijs} \mid \mathcal{O}_{ijs}(h), \mathcal{D}_{ijs}]$ is positive definite, and the rank condition of the theorem implies the latter for all small h. Hence θ_h exists and is unique. Let M_h^* be the same expression with q_{ijs} replaced by $F(Z'_{ijs}\theta)$. The argument of Step 5 of the proof of Theorem 2, with $-\Delta_{ijs}$ in place of the remainder there (which is defined for the same configuration with the opposite sign convention, but only its absolute value is used), the indicator of $\mathcal{O}_{ijs}(h) \cap \mathcal{D}_{ijs}$, normalized by its probability, in place of $\mathbb{1}_{ijs}W_{ijs}/\omega_n$, and $\varpi_1(h)$ in place of κ_n, gives $\sup_\Theta |M_h - M_h^*| \leq C\varpi_1(h)$ by the moment condition and $M_h^*(\theta) - M_h^*(\theta+v) \geq \frac{1}{4}c_0 e^{-R\|v\|}\|v\|^2$ for all small h by the rank condition. Therefore $\frac{1}{4}c_0 e^{-R\|\theta_h - \theta\|}\|\theta_h - \theta\|^2 \leq M_h^*(\theta) - M_h^*(\theta_h) \leq M_h(\theta) - M_h(\theta_h) + 2C\varpi_1(h) \leq 2C\varpi_1(h)$, since θ_h maximizes M_h, and Θ is compact, so $\theta_h \to \theta$ as $h \downarrow 0$. □

Corollary 1(b). Write $d^* = d_s(w_{is})$ and $\Psi_h(d)$ for the left side of (11). Condition first on χ_i, χ_j, $y_{j,-s}$, and $\mathcal{T}_{js}$. By the independence of agents i and j, the conditional probability below depends only on $(\chi_j, y_{j,-s})$. By the exact log-odds formula in the proof of Lemma 1, applied without any stability condition, $\mathbb{P}(y_j \in B_s \mid \chi_i, \chi_j, y_{j,-s}, \mathcal{T}_{js}) = F\big(Z'_{js}\theta - d_s(w_{js}) + R_j\big)$ with $R_j = \Gamma(c_{j,s+2}) - \Gamma(c_{j,s+1})$, and, as in the proof of Theorem 1(b), $|R_j| \leq \|\beta\|\,\|X_{j,s+2} - X_{j,s+1}\| \leq \|\beta\|h^{1/2}$ on $\mathcal{E}_{js}(h)$ by Assumption 2(ii). The remainder R_j may depend on μ_j, but its bound does not. On $\{j \in \mathcal{M}_{is}(h)\}$, Proposition 1(b) gives $|d_s(w_{js}) - d^*| \leq \varpi(\tau(h))$. Hence the conditional probability equals $F(Z'_{js}\theta - d^* + e_j)$ with $|e_j| \leq \varpi(\tau(h)) + \|\beta\|h^{1/2} \leq \varpi_1(h)$, uniformly in the agent effect. Iterated expectations, which are valid because the event $\{j \in \mathcal{M}_{is}(h)\} \cap \mathcal{E}_{js}(h)$ is determined by $(w_{is}, w_{js}, X_j, y_{j,-s})$ and $\mathcal{T}_{js}$, give $\Psi_h(d) = \mathbb{E}\big[F(Z'_{js}\theta - d^* + e_j) - F(Z'_{js}\theta - d) \,\big|\, \cdot\,\big]$, where the conditioning is on $\{j \in \mathcal{M}_{is}(h)\} \cap \mathcal{E}_{js}(h)$ and χ_i. The function Ψ_h is continuous and strictly increasing, and by dominated convergence it tends to

$\mathbb{E}[F(Z'_{js}\theta - d^* + e_j) \mid \cdot] > 0$ as $d \to \infty$ and to $\mathbb{E}[F(Z'_{js}\theta - d^* + e_j) \mid \cdot] - 1 < 0$ as $d \to -\infty$, so it has a unique zero d_h. Write $\epsilon_h = \varpi(\tau(h)) + \|\beta\| h^{1/2}$, so that $|e_j| \leq \epsilon_h \leq \varpi_1(h)$. Because F is increasing, $F(Z'_{js}\theta - d^* - \epsilon_h) \leq F(Z'_{js}\theta - d^* + e_j) \leq F(Z'_{js}\theta - d^* + \epsilon_h)$ pointwise, and taking the same conditional expectations gives

$$\Psi_h(d^* - \epsilon_h) \leq 0 \leq \Psi_h(d^* + \epsilon_h).$$

Since Ψ_h is continuous and strictly increasing, its unique zero lies between these endpoints, so $|d_h - d^*| \leq \epsilon_h \leq \varpi_1(h)$. The argument uses no moment of Z_{js}: because F is bounded, Ψ_h is well defined, continuous, and strictly increasing whatever the conditional law of Z_{js}. The conditions used above to obtain the probability representation and the overlap of the neighborhoods are still needed, and the estimation results of Section 6 keep their own moment assumptions. Since $\varpi_1(h) \to 0$ by Proposition 1(b), $d_h \to d^*$. □

A.5 Proof of Theorem 2

Let $N = n(n-1)/2$ and $Q_n(b,a) = \Omega_n(b,a)/(N\omega_n)$, which has the same maximizer as Ω_n. Fix $(b,a) \in \Theta$. Since $-\log F(v) = \log(1 + e^{-v}) \leq \log 2 + |v|$ and Θ is compact, $|m_{ijs}(b,a)| \leq C(1 + \|Z_{ijs}\|)$ for a constant C that does not depend on (b,a) in Θ. Note also that $W_{ijs} \leq \|K_1\|_\infty \bar{W}_{ijs}$.

Step 1: Reduction by concavity. Each $m_{ijs}(b,a)$ is a logistic log-likelihood in a linear index and is concave in (b,a), and the kernel weights are non-negative, so Q_n is a random concave function on the convex set Θ, with θ in its interior. Let M_n^* be the deterministic concave functions defined in Step 5. It suffices to show (a) that $Q_n(b,a) - M_n^*(b,a) \to_p 0$ for each (b,a) and (b) that the M_n^* are uniformly bounded on Θ and satisfy, for every $\eta > 0$, $\inf_{\|v\| \geq \eta,\, \theta+v \in \Theta}[M_n^*(\theta) - M_n^*(\theta+v)] \geq c(\eta) > 0$ for all n large. Given any subsequence, uniform boundedness and a diagonal argument yield a further subsequence along which M_n^* converges pointwise on a countable dense subset of Θ. Since uniformly bounded concave functions are equicontinuous on compact subsets of the interior, the convergence extends to the interior, with a concave limit that inherits the curvature bound and is therefore uniquely maximized at θ. Theorem 2.7 of Newey and McFadden (1994), whose argument uses only a compact convex neighborhood of θ in the interior, then gives $\hat{\theta} \to_p \theta$ along the further subsequence. Since every subsequence has such a further subsequence, $\hat{\theta} \to_p \theta$.

Step 2: Estimated distances. Let $\tilde{Q}_n$ denote Q_n with $\hat{\bar{\delta}}_{ij}^2$ replaced by $\bar{\delta}_{ij}^2$ in K_1, and let $u_n = \max_{i \neq j} |\hat{\bar{\delta}}_{ij}^2 - \bar{\delta}_{ij}^2|$ over pairs with a finite aggregated distance. By Lemma 4(d), $u_n = O_p((\log n/n)^{1/2}) = o_p(h_1)$ by Assumption 4(ii). On the event $\{u_n \leq h_1\}$, the two kernel values differ only if $\bar{\delta}_{ij}^2 < 2h_1$, and by the Lipschitz property they differ by at most Lu_n/h_1.

Hence

$$|Q_n(b,a) - \tilde{Q}_n(b,a)| \le \frac{Lu_n}{h_1} \cdot \frac{C}{N\omega_n} \sum_{s\in\mathcal{S}} \sum_{i<j} \mathbb{1}_{ijs} \bar{W}_{ijs}(1 + \|Z_{ijs}\|).$$

The second factor has expectation bounded by Assumption 5(iii), so it is $O_p(1)$, and the right side is $o_p(1)$.

Step 3: Concentration. The population weight W_{ijs}, the selection indicator, and m_{ijs} are functions of the node data $\mathcal{Z}_i = (y_{i0}, \ldots, y_{iT}, X_i, w_i, \mu_i)$ and $\mathcal{Z}_j$ only, and the kernel $\phi(\mathcal{Z}_i, \mathcal{Z}_j) = \sum_{s\in\mathcal{S}} \mathbb{1}_{ijs} W_{ijs} m_{ijs}(b,a)$ is symmetric, because m_{ijs} takes the same value when the roles of i and j are exchanged on a discordant pair. Thus $\tilde{Q}_n(b,a)$ is a U-statistic of order two in i.i.d. data with mean $M_n(b,a)$. By the variance formula for U-statistics,

$$\operatorname{Var}\big(\tilde{Q}_n(b,a)\big) \le \frac{1}{\omega_n^2} \left[\frac{4}{n} \mathbb{E}\big[\mathbb{E}(|\phi| \mid \mathcal{Z}_i)^2\big] + \frac{2}{N} \mathbb{E}[\phi^2] \right].$$

By the bound on $|m_{ijs}|$ and Assumption 5(iii), $\mathbb{E}[\phi^2] \le C \sum_s \mathbb{E}[\mathbb{1}_{ijs} \bar{W}_{ijs}(1 + \|Z_{ijs}\|)^2] \le C' \omega_n$. By the Cauchy–Schwarz inequality, $\mathbb{E}(|\phi| \mid \mathcal{Z}_i)^2 \le C\, \mathbb{E}\big[\sum_s \mathbb{1}_{ijs} \bar{W}_{ijs}(1 + \|Z_{ijs}\|)^2 \,\big|\, \mathcal{Z}_i\big] \cdot \mathbb{E}\big[\sum_s \mathbb{1}_{ijs} \bar{W}_{ijs} \,\big|\, \mathcal{Z}_i\big]$, and the second factor is bounded, so $\mathbb{E}[\mathbb{E}(|\phi| \mid \mathcal{Z}_i)^2] \le C'' \omega_n$. Hence $\operatorname{Var}(\tilde{Q}_n(b,a)) \le C/(n\omega_n) \to 0$ by Assumption 4(ii), and $\tilde{Q}_n(b,a) - M_n(b,a) \to_p 0$. Together with Step 2, $Q_n(b,a) - M_n(b,a) \to_p 0$.

Step 4: The discordant configuration on the selected pairs. Condition on $(\mathcal{F}_{ijs}, w_i, w_j, \mu_i, \mu_j)$ and on the discordance event. The weight W_{ijs} and the selection indicator are measurable with respect to this information: the set of periods entering the aggregated distance excludes $s+1$ and $s+2$ under both orderings of the swapped outcomes (third remark after Theorem 1), and the remaining periods' distances depend only on the latent indices and on lags outside the window. By the exact log-odds formula in the proof of Lemma 1, applied to both agents,

$$\mathbb{P}\big(y_{is} = 1 \,\big|\, \mathcal{F}_{ijs}, w_i, w_j, \mu_i, \mu_j, \mathcal{D}_{ijs}\big) = F\big(Z'_{ijs}\theta + \Delta_{ijs}\big),$$
$$\Delta_{ijs} = \big[\lambda(w_{is}) - \lambda(w_{i,s+1})\big] - \big[\lambda(w_{js}) - \lambda(w_{j,s+1})\big] + R_i - R_j,$$

where $R_a = \Gamma(c_{a,s+2}) - \Gamma(c_{a,s+1})$ satisfies $|R_a| \le |c_{a,s+2} - c_{a,s+1}|$ with $c_{at} = X'_{at}\beta + \lambda(w_{at}) + \mu_a$. The remainders R_a, and hence Δ_{ijs}, may depend on the agent effects through the levels $c_{a,s+1}$, but their bounds do not, because the agent effects cancel in the differences $c_{a,s+2} - c_{a,s+1}$. The bound below therefore holds uniformly in (μ_i, μ_j), which is all that Step 5 uses. By Assumption 3, $\lambda(w_{a,s+1}) - \lambda(w_{as}) = d_s(w_{as}) + e_{as,n}$ and, since $d_{s+1} \equiv 0$, $|\lambda(w_{a,s+2}) - \lambda(w_{a,s+1})| \le \zeta_n$. On a pair with $W_{ijs} > 0$, $\|X_{a,s+2} - X_{a,s+1}\| < h_2^{1/2}$ for $a \in \{i,j\}$, and $\delta_{s,n}(w_{is}, w_{js};\, y_{i,s-1})^2 \le T\bar{\delta}_{ij}^2 < Th_1$, because period s contributes to the aggregated distance. Hence, with ϖ the modulus of continuity of d_s defined after Assumption 2 and τ_n

from (22),

$$|\Delta_{ijs}| \le \varpi\big(\tau_n(Th_1)\big) + 2\|\beta\|\, h_2^{1/2} + 4\zeta_n =: \kappa_n \longrightarrow 0$$

on every pair with positive weight.

Step 5: Curvature of the localized criterion. Write $\ell(p;x) = p\log F(x) + (1-p)\log(1-F(x))$, which is linear in p. By Step 4 and iterated expectations over (w_i, w_j, μ_i, μ_j), $M_n(b,a) = \omega_n^{-1}\sum_s \mathbb{E}_n[\mathbb{1}_{ijs}W_{ijs}\,\ell(F(Z'_{ijs}\theta + \Delta_{ijs});\, Z'_{ijs}(b,a)')]$, where the expectation integrates over the agent effects on which Δ_{ijs} may depend. Let M_n^* denote the same expression with $\Delta_{ijs} = 0$. It is a deterministic concave function whose comparison logit index $Z'_{ijs}\theta$ contains no individual-effect term, although its expectation still uses the law of the selected population, which can depend on the individual effects. Since $|\partial\ell/\partial p| = |x|$ and $|F(v+\Delta) - F(v)| \le |\Delta|/4$, with the bound on $|\Delta_{ijs}|$ holding uniformly in the agent effects,

$$|M_n(b,a) - M_n^*(b,a)| \le \frac{\kappa_n}{4}\,\frac{1}{\omega_n}\sum_s \mathbb{E}_n\big[\mathbb{1}_{ijs}W_{ijs}\,|Z'_{ijs}(b,a)'|\big] \longrightarrow 0,$$

so $Q_n(b,a) - M_n^*(b,a) \to_p 0$ by Step 3, which is condition (a) of Step 1. The same bound and Assumption 5(iii) show that the M_n^* are uniformly bounded on Θ. M_n^* is the expected log-likelihood of a correctly specified logit, so for $v \neq 0$,

$$M_n^*(\theta) - M_n^*(\theta+v) = \frac{1}{\omega_n}\sum_s \mathbb{E}_n\big[\mathbb{1}_{ijs}W_{ijs}\,\mathrm{KL}\big(Z'_{ijs}\theta,\ Z'_{ijs}(\theta+v)\big)\big],$$

where $\mathrm{KL}(a, a+x) = \int_0^x (x-u)F'(a+u)\,du$ is the Kullback–Leibler divergence between Bernoulli distributions with logits a and $a+x$. For the logistic distribution, $F'(a+u) \ge F'(a)e^{-|u|}$, so $\mathrm{KL}(a,a+x) \ge \frac{1}{2}F'(a)x^2e^{-|x|}$. Restricting to $\|Z_{ijs}\| \le R$ and using $F' \le 1/4$,

$$M_n^*(\theta) - M_n^*(\theta+v) \ge \frac{e^{-R\|v\|}}{2}\left[v'\Sigma_n v - \frac{\|v\|^2}{4R^2}\,\frac{1}{\omega_n}\sum_s \mathbb{E}_n\big[\mathbb{1}_{ijs}W_{ijs}\|Z_{ijs}\|^4\big]\right].$$

By Assumption 5(iii) the last average is bounded, and by part (iv) there is $c_0 > 0$ with $v'\Sigma_n v \ge c_0\|v\|^2$ for n large. Choosing R large enough, the right side is at least $\frac{1}{4}c_0e^{-R\|v\|}\|v\|^2$ for all n large. For $\|v\| \ge \eta$, concavity of M_n^* along the ray from θ through $\theta+v$ gives $M_n^*(\theta) - M_n^*(\theta+v) \ge M_n^*(\theta) - M_n^*(\theta+\eta v/\|v\|) \ge \frac{1}{4}c_0e^{-R\eta}\eta^2$, which is condition (b) of Step 1 and completes the proof. □

A.6 Proof of Theorem 3

Let $\bar{K} = K_1(0)K_2(0)^2$, $Q_n(b,a) = \Omega_n(b,a)/(N\bar{K})$, and let

$$Q_n^o(b,a) = \frac{1}{N}\sum_{s\in\mathcal{S}}\sum_{i<j}\mathbb{1}_{ijs}^{\text{ex}}\, m_{ijs}(b,a)$$

be the oracle criterion that uses the exact-match indicator in place of the kernel weights. The law of the primitives is fixed, and all moments below are finite by Assumption 5(i), since Z_{ijs} is a linear function of the two agents' covariates and of bounded outcomes.

Step 1: The estimated weights select the exact matches. Let $\hat{W}_{ijs}$ be the estimated weight of pair (i,j) at s. Because $h_2 \le \bar{\epsilon}_X$, $K_2(\|X_{a,s+2} - X_{a,s+1}\|^2/h_2) = K_2(0)\mathbb{1}\{X_{a,s+2} = X_{a,s+1}\}$ almost surely. By Assumption 7(v) and the definition of $\bar{\epsilon}$, a pair in $\mathcal{P}_s$ with $\bar{\delta}_{ij}^2 > 0$ has $\bar{\delta}_{ij}^2 \ge \bar{\epsilon}/(T-2) \ge 2h_1$. By Lemma 4(d), with probability approaching one all such pairs have $\hat{\bar{\delta}}_{ij}^2 > h_1$ and therefore receive zero weight. For pairs with $\bar{\delta}_{ij} = 0$, the two agents are equivalent on the support at every contributing period, so Lemma 4(d) gives $\max \hat{\bar{\delta}}_{ij}^2 = O_p(\log n/n)$ and, by the Lipschitz property,

$$\max_{i<j,\ s}\ \mathbb{1}_{ijs}\left|\hat{W}_{ijs}/\bar{K} - \mathbb{1}_{ijs}^{\text{ex}}\right| = O_p\left(\frac{\log n}{nh_1}\right) = o_p(n^{-1/2})$$

with probability approaching one, by Assumption 7(v).

Step 2: The oracle score is exactly centered. On the event $\mathbb{1}_{ijs}^{\text{ex}} = 1$, the agents share the lag at $s-1$ and have $\bar{\delta}_{ij} = 0$, so by Lemma 4(c) they are equivalent on the support at s. By Assumption 7(ii), $d_s(w_{is}) = d_s(w_{js})$, so their social influence terms change by the same amount between s and $s+1$. By Assumption 2(ii), both are stable between $s+1$ and $s+2$, and their covariates are exactly stable. The event is therefore contained in $\mathcal{C}_{ijs} \cap \mathcal{X}_{ijs}$, and it is measurable with respect to $(\mathcal{F}_{ijs}, w_i, w_j)$ and the discordance event, as in Step 4 of the proof of Theorem 2. Lemma 2 gives $\mathbb{P}(y_{is} = 1 \mid \mathcal{F}_{ijs}, w_i, w_j, \mathcal{D}_{ijs}) = F(Z_{ijs}'\theta)$ on this event, hence $\mathbb{E}[\mathbb{1}_{ijs}^{\text{ex}} S_{ijs}(\theta) \mid \mathcal{F}_{ijs}, w_i, w_j] = 0$, and in particular $\mathbb{E}[\mathbb{1}_{ijs}^{\text{ex}} S_{ijs}(\theta)] = 0$.

Step 3: Consistency. By Step 1,

$$|Q_n(b,a) - Q_n^o(b,a)| \le o_p(n^{-1/2})\, N^{-1}\sum_s\sum_{i<j}\mathbb{1}_{ijs}^{\text{ex}}|m_{ijs}(b,a)| = o_p(1).$$

The oracle criterion is a U-statistic with integrable symmetric kernel, so $Q_n^o(b,a) \to_p Q^o(b,a) = \sum_s \mathbb{E}[\mathbb{1}_{ijs}^{\text{ex}} m_{ijs}(b,a)]$. The function Q^o is concave with gradient $\sum_s \mathbb{E}[\mathbb{1}_{ijs}^{\text{ex}} S_{ijs}(\theta)] = 0$ at θ by Step 2, and its Hessian $-\sum_s \mathbb{E}[\mathbb{1}_{ijs}^{\text{ex}} F'(Z_{ijs}'b) Z_{ijs} Z_{ijs}']$, where from here on $b \in \Theta$ denotes a value of the whole parameter vector, is negative definite at every b, because $F' > 0$ and

$\sum_s \mathbb{E}[\mathbb{1}^{\text{ex}}_{ijs} Z_{ijs} Z'_{ijs}]$ is positive definite when Σ_{ex} is. Hence θ is the unique maximizer of Q^o, and, since Q_n is concave, Theorem 2.7 of Newey and McFadden (1994) gives $\hat{\theta} \to_p \theta$.

Step 4: Score. Let $U_n = N^{-1} \sum_s \sum_{i<j} \mathbb{1}^{\text{ex}}_{ijs} S_{ijs}(\theta)$. By Step 1,

$$\|\nabla Q_n(\theta) - U_n\| \le o_p(n^{-1/2})\, N^{-1} \sum_s \sum_{i<j} \mathbb{1}^{\text{ex}}_{ijs} \|S_{ijs}(\theta)\| = o_p(n^{-1/2}).$$

The statistic U_n is a U-statistic of order two in the i.i.d. node data with the kernel $k(\mathcal{Z}_i, \mathcal{Z}_j) = \sum_s \mathbb{1}^{\text{ex}}_{ijs} S_{ijs}(\theta)$, which is symmetric because $S_{ijs} = S_{jis}$ on discordant pairs. By Step 2 the kernel has mean zero, and it has finite second moment. Hoeffding's decomposition gives $\sqrt{n}\, U_n = 2n^{-1/2} \sum_i \psi_i + O_p(n^{-1/2})$ with $\psi_i = \mathbb{E}[k(\mathcal{Z}_i, \mathcal{Z}_j) \mid \mathcal{Z}_i]$, and the central limit theorem gives $\sqrt{n}\, U_n \xrightarrow{d} \mathcal{N}(0, 4V_{\text{ex}})$. Note that ψ_i sums over the transition periods before the conditional expectation is taken, so the variance retains covariances across periods.

Step 5: Hessian and conclusion. The Hessian of Q_n is

$$H_n(b) = -N^{-1} \sum_s \sum_{i<j} (\hat{W}_{ijs}/\bar{K}) \mathbb{1}_{ijs} F'(Z'_{ijs} b) Z_{ijs} Z'_{ijs}.$$

By Step 1, $H_n(b)$ differs from its oracle counterpart by $o_p(1)$ uniformly in b. Since $|F'(x) - F'(x')| \le |x - x'|$, the oracle Hessian is Lipschitz in b with an integrable envelope proportional to $\|Z_{ijs}\|^3$, so the uniform law of large numbers for U-statistics gives $\sup_{\|b-\theta\| \le \eta} \|H_n(b) + \Sigma_{\text{ex}}(b)\| \to_p 0$ for small $\eta > 0$, with $\Sigma_{\text{ex}}(b) = \sum_s \mathbb{E}[\mathbb{1}^{\text{ex}}_{ijs} F'(Z'_{ijs} b) Z_{ijs} Z'_{ijs}]$ continuous and $\Sigma_{\text{ex}}(\theta) = \Sigma_{\text{ex}}$. With probability approaching one, $\hat{\theta}$ is interior and satisfies the first-order condition, and

$$0 = \nabla Q_n(\hat{\theta}) = \nabla Q_n(\theta) + \left[\int_0^1 H_n\big(\theta + u(\hat{\theta} - \theta)\big)\, du \right] (\hat{\theta} - \theta).$$

The integrated Hessian converges in probability to $-\Sigma_{\text{ex}}$ because $\hat{\theta} \to_p \theta$, so $\sqrt{n}(\hat{\theta} - \theta) = \Sigma_{\text{ex}}^{-1} \sqrt{n}\, U_n + o_p(1) \xrightarrow{d} \mathcal{N}(0, 4\Sigma_{\text{ex}}^{-1} V_{\text{ex}} \Sigma_{\text{ex}}^{-1})$. □

Variance estimation. Write $\tilde{\Sigma}(b)$ and $\tilde{\psi}_i(b)$ for the oracle versions of $\hat{\Sigma}$ and $\hat{\psi}_i$ evaluated at b, in which $\hat{W}_{ijs}/\bar{K}$ is replaced by $\mathbb{1}^{\text{ex}}_{ijs}$, and let $r_n = \max_{i<j,\, s} \mathbb{1}_{ijs} |\hat{W}_{ijs}/\bar{K} - \mathbb{1}^{\text{ex}}_{ijs}|$. By Step 1, $r_n = O_p(\log n/(nh_1)) = o_p(1)$, and with probability approaching one the two sets of weights

differ only on exact-match pairs. On that event, since $\|S_{ijs}(b)\| \le \|Z_{ijs}\|$ and $F(1-F) \le \frac{1}{4}$,

$$\|\hat{\psi}_i - \tilde{\psi}_i(\hat{\theta})\| \le r_n A_i, \qquad A_i = \frac{1}{n-1} \sum_{j \neq i} \sum_{s \in \mathcal{S}} \mathbb{1}^{\mathrm{ex}}_{ijs} \|Z_{ijs}\|,$$

$$\|\hat{\Sigma} - \tilde{\Sigma}(\hat{\theta})\| \le \frac{r_n}{4} \binom{n}{2}^{-1} \sum_{s \in \mathcal{S}} \sum_{i<j} \mathbb{1}^{\mathrm{ex}}_{ijs} \|Z_{ijs}\|^2,$$

so that $n^{-1} \sum_i \|\hat{\psi}_i - \tilde{\psi}_i(\hat{\theta})\|^2 \le r_n^2\, n^{-1} \sum_i A_i^2$. By Jensen's inequality, $\mathbb{E}[A_i^2] \le \mathbb{E}\big[(\sum_s \mathbb{1}^{\mathrm{ex}}_{ijs} \|Z_{ijs}\|)^2\big] \le |\mathcal{S}| \sum_s \mathbb{E}\big[\mathbb{1}^{\mathrm{ex}}_{ijs} \|Z_{ijs}\|^2\big] < \infty$ for fixed T by Assumption 5(i), so $n^{-1} \sum_i A_i^2 = O_p(1)$ and the average squared difference is $o_p(1)$. The same moment bound makes the average in the bound for $\hat{\Sigma}$ equal to $O_p(1)$, so $\hat{\Sigma} - \tilde{\Sigma}(\hat{\theta}) = o_p(1)$. No bound on the maximum over agents is needed, only these averages. The oracle $\tilde{\Sigma}(\hat{\theta})$ is a U-statistic evaluated at $\hat{\theta}$ and converges to Σ_{ex} by the argument of Step 5. For the influence terms, three replacements lead from $\hat{\psi}_i$ to ψ_i. First, the one just shown. Second, since $|F(Z'_{ijs}\hat{\theta}) - F(Z'_{ijs}\theta)| \le \frac{1}{4} \|Z_{ijs}\|\, \|\hat{\theta} - \theta\|$, $\|\tilde{\psi}_i(\hat{\theta}) - \tilde{\psi}_i(\theta)\| \le \frac{1}{4} \|\hat{\theta} - \theta\|\, B_i$ with $B_i = (n-1)^{-1} \sum_{j \neq i} \sum_s \mathbb{1}^{\mathrm{ex}}_{ijs} \|Z_{ijs}\|^2$, and $n^{-1} \sum_i B_i^2 = O_p(1)$ by the fourth moments of Assumption 5(i), so this average squared difference is $O_p(\|\hat{\theta} - \theta\|^2) = o_p(1)$. Third, given $\mathcal{Z}_i$ the summands of $\tilde{\psi}_i(\theta) = (n-1)^{-1} \sum_{j \neq i} k(\mathcal{Z}_i, \mathcal{Z}_j)$ are i.i.d. with mean ψ_i, and $\mathbb{E}\|k(\mathcal{Z}_i, \mathcal{Z}_j)\|^2 < \infty$ by Assumption 5(i), so $\mathbb{E}\|\tilde{\psi}_i(\theta) - \psi_i\|^2 \le \mathbb{E}\|k(\mathcal{Z}_i, \mathcal{Z}_j)\|^2/(n-1)$. For each of the three replacements the average squared difference $n^{-1} \sum_i \|\cdot\|^2$ is $o_p(1)$, so by Minkowski's inequality $n^{-1} \sum_i \|\hat{\psi}_i - \psi_i\|^2 = o_p(1)$. Since $n^{-1} \sum_i \|\psi_i\|^2 = O_p(1)$, the Cauchy–Schwarz inequality shows that the sample mean and second moment of the $\hat{\psi}_i$ differ from those of the ψ_i by $o_p(1)$. By the law of large numbers, $n^{-1} \sum_i \psi_i \to_p \mathbb{E}[\psi_i] = 0$ and $n^{-1} \sum_i \psi_i \psi_i' \to_p V_{\mathrm{ex}}$, so $\hat{V} \to_p V_{\mathrm{ex}}$ and the plug-in variance is consistent (cf. Serfling, 1980, Chapter 5).

A.7 Conditional Sums of Bernoulli Variables

Lemma 5 (Conditional Bernoulli sums)**.** *Let $B_1, \dots, B_m$ be independent Bernoulli variables with success probabilities $p_j \in (0,1)$ such that $\sum_j p_j = N$ is an integer. Write $q_j = p_j(1-p_j)$, $d = \sum_j q_j$, $\pi_j = \mathbb{P}(B_j = 1 \mid \sum_l B_l = N)$, and, for vectors $z_1, \dots, z_m$ of fixed dimension, $\bar{z} = d^{-1} \sum_j q_j z_j$. There are constants d_0 and C, which do not depend on m or on the p_j, such that whenever $d \ge d_0$:*

(a) $\big\| \sum_j (\pi_j - p_j) z_j \big\| \le C d^{-1} \sum_j \|z_j\|$;

(b) $\big\| \mathrm{Var}\big(\sum_j B_j z_j \,\big|\, \sum_l B_l = N \big) - \sum_j q_j (z_j - \bar{z})(z_j - \bar{z})' \big\| \le C d^{-1} \sum_j \|z_j\|^2 + C d^{-2} \big(\sum_j \|z_j\| \big)^2$.

Proof. By Hájek (1964) and Theorem 1 of Boistard et al. (2012), $\pi_j = p_j + O(d^{-1})$ and $\mathrm{Cov}(B_j, B_k \mid \sum_l B_l = N) = -q_j q_k/d + O(d^{-2})$ for $j \neq k$, uniformly in the indices, and hence $\pi_j(1-\pi_j) = q_j + O(d^{-1})$. Part (a) follows by multiplying the first expansion by z_j and summing. For part (b), summing the conditional variances and covariances against $z_j z_k'$ gives the leading term $\sum_j q_j z_j z_j' - d^{-1}(\sum_j q_j z_j)(\sum_j q_j z_j)' = \sum_j q_j (z_j - \bar{z})(z_j - \bar{z})'$ and the diagonal correction $d^{-1}\sum_j q_j^2 z_j z_j'$, whose norm is at most $(4d)^{-1}\sum_j \|z_j\|^2$ because $q_j \leq 1/4$. The $O(d^{-1})$ error in the variances contributes at most $Cd^{-1}\sum_j \|z_j\|^2$, and the $O(d^{-2})$ error in the covariances at most $Cd^{-2}(\sum_j \|z_j\|)^2$. □

A.8 Proof of Proposition 2

Part (a). By Proposition 1(a), $\max_{i \neq j} |\hat{\delta}^2_{ijs} - \delta^2_{ijs}| \mathbb{1}\{y_{i,s-1} = y_{j,s-1}\} = O_p((\log n/n)^{1/2})$, which is $o_p(h_1)$ because $\sqrt{n}\, h_1 / \log n \to \infty$ by Assumption 7(v). By Proposition 1(c), a pair that shares the lagged outcome at $s-1$ has $\delta^2_{ijs} = 0$ if its members are in the same population class and $\delta^2_{ijs} \geq \bar{\epsilon}$ otherwise, and $\bar{\epsilon} \geq 2(T-2)h_1 \geq 2h_1$ by Assumption 7(v). Hence, with probability approaching one, for every pair of eligible agents that share the lagged outcome, $\hat{\delta}^2_{ijs} < h_1$ exactly when the two agents are in the same population class, and the edges of the graph (17) are exactly the pairs of eligible agents in the same population class. Equivalence on the support is reflexive, symmetric, and transitive, so these edges form a complete graph on each set $\{j \in \mathcal{E}_s : c_j = c\}$ and join no two different sets. The connected components are these sets.

Part (b). Let $\mathcal{H}_s$ be the σ-field generated by the primitives, all link shocks, the outcomes $\{y_{j,-s}\}_{j \leq n}$, and the switch events. By Assumption 1(i), (ii), and (iv), given the primitives and the link shocks, the shock vectors ε_j are independent across agents, so the within-window outcomes $(y_{js}, y_{j,s+1})$ are independent across agents given $\mathcal{H}_s$, and the conditional law of agent j's pair depends only on $(\chi_j, y_{j,-s})$ and $\mathcal{T}_{js}$. For an eligible agent the pair is $(1,0)$ or $(0,1)$, and the ratio of their probabilities is given by Lemma 1, whose stability conditions hold because $X_{j,s+2} = X_{j,s+1}$ and by Assumption 2(ii). Since $\lambda(w_{js}) - \lambda(w_{j,s+1}) = -d_s(w_{js})$ by Assumption 2(i), $\mathbb{P}(b_{js} = 1 \mid \mathcal{H}_s) = F(Z_{js}'\theta - d_s(w_{js}))$. The network D_t is a function of the primitives, the link shocks at t, and the outcomes at $t-1$. For $t \notin \{s+1, s+2\}$ these outcomes are among the $y_{j,-s}$, so $\mathcal{G}_s \subseteq \mathcal{H}_s$, and since the conditional law given $\mathcal{H}_s$ is $\mathcal{G}_s$-measurable, the tower property gives the same law given $\mathcal{G}_s$. For a $\mathcal{G}_s$-measurable set C on which $d_s(w_{js}) = d_C$, the probability of orderings $a \in \{0,1\}^C$ given $\mathcal{G}_s$ is therefore

$$\prod_{j \in C} \frac{\exp\{a_j(Z_{js}'\theta - d_C)\}}{1 + \exp(Z_{js}'\theta - d_C)} = \exp\Big(\sum_{j \in C} a_j Z_{js}'\theta\Big) \frac{e^{-d_C \sum_j a_j}}{\prod_{j \in C}\big(1 + e^{Z_{js}'\theta - d_C}\big)},$$

and summing over the a with $\sum_j a_j = N_C$ shows that the conditional probability of the observed orderings given N_C is $\exp\{\ell_C(\theta)\}$, free of d_C. The feasible classes are $\mathcal{G}_s$-measurable, because they are functions of D_s, of the outcomes at $s-1$, of the covariates, and of the switch events.

Part (c), Step 1: reduction to the population classes. By part (a), with probability approaching one the objective in (19) equals $L_n(\theta) = \sum_c \ell_{C_c}(\theta)$ with $C_c = \{j \in \mathcal{E}_s : c_j = c\}$, so it suffices to prove the result for the maximizer of L_n. Let K be the number of classes with $\mathbb{P}(e_j = 1, c_j = c) > 0$. The other sets C_c are empty almost surely. Since $|C_c|/n$ converges to a positive limit and the orderings are independent logits by part (b), each of the K sets contains both orderings with probability approaching one, which is assumed below.

Step 2: comparison with the profile likelihood. For a class C with $0 < N_C < |C|$ and $\gamma \in \mathbb{R}$, let $L_C(\theta, \gamma) = \sum_{j \in C} \big[b_{js}(Z_{js}'\theta - \gamma) - \log(1 + e^{Z_{js}'\theta - \gamma})\big]$, the log likelihood of a logit with intercept $-\gamma$, and let $P_{\theta,\gamma}$ be the law of independent Bernoulli variables with success probabilities $F(Z_{js}'\theta - \gamma)$, $j \in C$. The factorization in part (b) gives, for every γ,

$$\ell_C(\theta) = L_C(\theta, \gamma) - \log P_{\theta,\gamma}\Big(\sum_{j \in C} b_{js} = N_C\Big).$$

Let $\gamma_C(\theta)$ solve $\sum_{j \in C} F(Z_{js}'\theta - \gamma) = N_C$, which has a unique solution because the left side decreases continuously from $|C|$ to 0. It is the first-order condition of the strictly concave function $\gamma \mapsto L_C(\theta, \gamma)$, so $L_C(\theta, \gamma_C(\theta)) = \max_\gamma L_C(\theta, \gamma) =: L_C^{\mathrm{P}}(\theta)$, the profile log likelihood. Under $P_{\theta,\gamma_C(\theta)}$ the number of successes has mean N_C, an integer, so N_C is a mode of its distribution (Darroch, 1964), and its probability is at least $1/(|C|+1)$. Hence

$$0 \le \ell_C(\theta) - L_C^{\mathrm{P}}(\theta) \le \log(|C| + 1) \qquad \text{for every } \theta. \tag{23}$$

Step 3: consistency. Let $L_n^{\mathrm{J}}(\theta, \gamma) = \sum_c L_{C_c}(\theta, \gamma_c)$ for $\gamma \in \mathbb{R}^K$, the log likelihood of a logit for the eligible agents' orderings with one intercept per class, which is a sum of i.i.d. terms, and $L_n^{\mathrm{P}}(\theta) = \max_\gamma L_n^{\mathrm{J}}(\theta, \gamma) = \sum_c L_{C_c}^{\mathrm{P}}(\theta)$. By (23), $\sup_\theta n^{-1}|L_n(\theta) - L_n^{\mathrm{P}}(\theta)| \le K \log(n+1)/n \to 0$. The function $n^{-1}L_n^{\mathrm{J}}$ is concave and converges pointwise in probability to $Q(\theta, \gamma) = \mathbb{E}\big[e_j\{b_{js}(Z_{js}'\theta - \gamma_{c_j}) - \log(1 + e^{Z_{js}'\theta - \gamma_{c_j}})\}\big]$, which is strictly concave, because its Hessian is $-\mathbb{E}[e_j F'(Z_{js}'\theta - \gamma_{c_j}) g_j g_j']$ with $g_j = (Z_{js}', -\iota_{c_j}')'$ and ι_c the c-th unit vector, and this matrix is negative definite when the regressor varies within classes, which the positive definiteness of $\mathcal{I}_{\mathrm{C}}$ implies. By part (b) and the information inequality, Q is uniquely maximized at (θ, d), where d collects the class movements. For each θ, pointwise convergence of the concave functions $n^{-1}L_n^{\mathrm{J}}(\theta, \cdot)$ implies uniform convergence on compact sets (the convexity lemma used in the proof of Theorem 2.7 of Newey and McFadden, 1994). Since

the limit $Q(\theta, \cdot)$ is strictly concave with a unique maximizer, and the maximizers of concave functions that converge uniformly on a neighborhood of it lie in that neighborhood with probability approaching one, $n^{-1}L_n^{\mathrm{P}}(\theta) = \max_\gamma n^{-1}L_n^{\mathrm{J}}(\theta, \gamma)$ converges in probability to $Q^{\mathrm{P}}(\theta) = \max_\gamma Q(\theta, \gamma)$, which is concave with unique maximizer θ. Hence $n^{-1}L_n(\theta) \to_p Q^{\mathrm{P}}(\theta)$ for every θ. Since L_n is concave, Theorem 2.7 of Newey and McFadden (1994) gives $\hat{\theta}^{\mathrm{C}} \to_p \theta$.

Step 4: asymptotic normality. Fix u, let $\theta_u = \theta + u/\sqrt{n}$, and write $L_n = L_n^{\mathrm{P}} + R_n$. The profile part admits the expansion

$$L_n^{\mathrm{P}}(\theta_u) - L_n^{\mathrm{P}}(\theta) = u'U_n - \tfrac{1}{2}u'\mathcal{I}_{\mathrm{C}}u + o_p(1), \qquad U_n = n^{-1/2}\sum_{j=1}^{n}\tilde{S}_j,$$

which is the local expansion of the profile log likelihood of a regular parametric model with $\dim\theta + K$ parameters and i.i.d. observations: expanding $L_n^{\mathrm{J}}(\theta_u, d + v/\sqrt{n})$ to second order in (u, v), with the Hessian replaced by its expectation by the law of large numbers and the third-order term of order $n^{-1/2}$ by the moments of Assumption 5(i), and maximizing the quadratic over v, whose maximizer is bounded in probability, gives the score for θ minus its projection on the intercept scores, which is U_n, and the Schur complement of the intercept block of the information, which is $\mathcal{I}_{\mathrm{C}}$. For the remainder $R_n(\theta) = \sum_c[\ell_{C_c}(\theta) - L_{C_c}^{\mathrm{P}}(\theta)]$, the gradient of ℓ_C is $\sum_{j\in C}(b_{js} - \pi_j(\theta))Z_{js}$, where $\pi_j(\theta)$ is the probability that $b_{js} = 1$ under the conditional law of the orderings given N_C, and, by the envelope theorem, the gradient of L_C^{P} is $\sum_{j\in C}(b_{js} - p_j(\theta))Z_{js}$ with $p_j(\theta) = F(Z_{js}'\theta - \gamma_C(\theta))$. Hence $\nabla R_n(\theta) = \sum_c\sum_{j\in C_c}(p_j(\theta) - \pi_j(\theta))Z_{js}$. The conditional law of the orderings given N_C is the rejective, or conditional Poisson, sampling design with working probabilities $p_j(\theta)$, whose sum is N_C. Let $\sigma_C^2(\theta') = \sum_{j\in C}p_j(\theta')(1 - p_j(\theta'))$. For $\|\theta' - \theta\| \le \|u\|/\sqrt{n}$, $\sigma_{C_c}^2(\theta')/n$ converges in probability, uniformly, to $\mathbb{E}[e_j\mathbb{1}\{c_j = c\}q_j] > 0$, so Lemma 5(a), applied conditionally on the regressors with $z_j = Z_{js}$ and $d = \sigma_C^2(\theta')$, gives $\sup_{\theta'}\|\nabla R_n(\theta')\| \le O_p(1)\, n^{-1}\sum_j e_j\|Z_{js}\| = O_p(1)$ over that ball, and $R_n(\theta_u) - R_n(\theta) = O_p(n^{-1/2}) = o_p(1)$ by the mean value theorem. Therefore $L_n(\theta_u) - L_n(\theta) = u'U_n - \frac{1}{2}u'\mathcal{I}_{\mathrm{C}}u + o_p(1)$ for every u, and since L_n is concave, the basic corollary of Hjort and Pollard (1993) gives $\sqrt{n}(\hat{\theta}^{\mathrm{C}} - \theta) = \mathcal{I}_{\mathrm{C}}^{-1}U_n + o_p(1)$. The $\tilde{S}_j$ are i.i.d. with mean zero, because $\mathbb{E}[b_{js} - p_j \mid \mathcal{G}_s] = 0$ on $\{e_j = 1\}$ by part (b), and with variance $\mathbb{E}[e_jq_j(Z_{js} - \bar{Z}_{c_j})(Z_{js} - \bar{Z}_{c_j})'] = \mathcal{I}_{\mathrm{C}}$, because the conditional variance of b_{js} is q_j. The central limit theorem gives the limit $\mathcal{N}(0, \mathcal{I}_{\mathrm{C}}^{-1})$.

Step 5: variance estimation. The negative Hessian of ℓ_C at θ' is the conditional covariance matrix of $\sum_{j\in C}b_{js}Z_{js}$ given N_C under θ'. By Lemma 5(b) with $z_j = Z_{js}$, $-\nabla^2\ell_C(\theta') = \sum_{j\in C}p_j(\theta')(1 - p_j(\theta'))(Z_{js} - \bar{Z}_C(\theta'))(Z_{js} - \bar{Z}_C(\theta'))' + O_p(1)$ uniformly for θ' in a neighborhood of θ, with $\bar{Z}_C(\theta')$ the $p_j(1-p_j)$-weighted mean of Z_{js} over the class, because $\sigma_C^2(\theta') \ge \varrho n$ with probability approaching one for some $\varrho > 0$ and $n^{-1}\sum_j\|Z_{js}\|^2 = O_p(1)$ by Assumption 5(i).

Evaluated at $\hat{\theta}^{\mathrm{C}}$, whose class intercepts $\gamma_C(\hat{\theta}^{\mathrm{C}})$ are consistent for the movements d_c by Steps 3 and 4, the leading term divided by n converges in probability to $\mathcal{I}_{\mathrm{C}}$ by the law of large numbers, which proves the consistency of the variance estimator. □

A.9 Proof of Corollary 2

Let $H_n = [c_- n^{-1/4}, c_+ n^{-1/4}]$. With $T = 3$, the aggregated distance of a discordant pair is $\hat{\delta}^2_{ij1}$, and $2c_+ n^{-1/4} \leq \bar{\epsilon}$ for n large. By Lemma 4(d), with probability approaching one, every pair that shares the initial outcome and is not equivalent on the support at $t = 1$ has $\hat{\delta}^2_{ij1} \geq \bar{\epsilon} - O_p((\log n/n)^{1/2}) > c_+ n^{-1/4}$, and every equivalent pair has $\hat{\delta}^2_{ij1} = O_p(\log n/n)$. Hence, with probability approaching one, the estimated weights $\hat{W}_{ij1}(h)$ at every $h \in H_n$ are positive exactly on the exact matches, and, by the Lipschitz property of K_1,

$$\sup_{h \in H_n} \max_{i<j} \mathbb{1}_{ij1} \left| \hat{W}_{ij1}(h)/\bar{K} - \mathbb{1}^{\mathrm{ex}}_{ij1} \right| = O_p\left(\frac{\log n}{c_- n^{3/4}} \right) = o_p(n^{-1/2}).$$

This is the conclusion of Step 1 of the proof of Theorem 3, uniformly over H_n. Steps 2–5 and the proof for the variance estimator use the weights only through it, so they hold at $\hat{h}_1 \in H_n$. Likewise, the argument for Proposition 2(a) holds uniformly over H_n, because $O_p((\log n/n)^{1/2}) = o_p(c_- n^{-1/4})$, and part (c) uses the bandwidth only through part (a). The proof of part (b) uses only the $\mathcal{G}_1$-measurability of the feasible classes, which holds when $\hat{h}_1$ is a function of D_1 and the initial outcomes, both among the generators of $\mathcal{G}_1$. Finally, each inner average in (12) lies in $[-1, 1]$, so $0 \leq S_n \leq 1$, and the rule (16) lies in H_n with the stated constants. □

B The Social Influence Function: What Can and Cannot Be Recovered

This appendix records what the model identifies about the social influence function once the agent effect is present, and describes an estimator of its common movement. The material is secondary to the main results, which treat λ, f, and the agent effects as nuisance components, and no asymptotic theory is claimed for the estimator described here.

B.1 The Level of λ Is Not Identified

In the model without an agent effect, the social influence function can be recovered on the network classes that the codegree distance identifies, under the conditions of Proposition 1

and with conditioning on the network type understood through those classes as in Corollary 1, by inverting the conditional choice probability: for an agent of network type φ at period t, $\mathbb{P}(y_{it} = 1 \mid y_{i,t-1}, X_{it}, f_{w_{it}} = \varphi) = F(X_{it}'\beta + \alpha y_{i,t-1} + \lambda(w_{it}))$ when λ is constant on the class of φ, and F^{-1} of this probability minus the observed part of the index returns $\lambda(w_{it})$. With the agent effect, the same probability is

$$\mathbb{P}(y_{it} = 1 \mid y_{i,t-1}, X_{it}, f_{w_{it}} = \varphi) = \mathbb{E}\Big[F\big(X_{it}'\beta + \alpha y_{i,t-1} + \lambda(w_{it}) + \mu_i\big) \,\Big|\, y_{i,t-1}, X_{it}, f_{w_{it}} = \varphi\Big],$$

an average of F over the conditional distribution of μ_i given the lag, the covariates, and the network type. Because μ_i may depend on (w_i, X_i, y_{i0}) arbitrarily (Assumption 1(i)) and the lagged outcome depends on μ_i, this distribution is unrestricted, and inverting F does not return $\lambda(w_{it})$. Two observations make the non-identification precise. First, replacing (λ, μ_i) by $(\lambda + c, \mu_i - c)$ for a constant c leaves every outcome probability unchanged. Second, when the latent index is time-invariant, $w_{it} = w_i$, only the sum $\lambda(w_i) + \mu_i$ enters the outcome equation, and any function of w_i can be moved between the two terms. The paper therefore makes no claim about the level of the social influence function, and the name refers to the component of the unobserved heterogeneity whose movement the network reveals, not to a function that the data recover.

B.2 An Estimator of the Common Movement

Corollary 1 identifies the common movement $d_s(w_{is})$ of a reference agent i at a transition period s as the zero of a moment over the agents in i's observable network class, or in its codegree neighborhoods: among agents who switch between s and $s+1$, have stable covariates between $s+1$ and $s+2$, and are at zero (or small) codegree distance from i, the log odds of switching down rather than up equal $Z_{js}'\theta - d_s(w_{is})$, up to a remainder that vanishes with the neighborhood in continuous designs. Given the first-step estimator $\hat{\theta} = (\hat{\beta}', \hat{\alpha})'$, the sample counterpart of (10) and (11) is a kernel-weighted logit with a single free parameter, in which the codegree kernel replaces membership in the class and the covariate kernel replaces exact stability. Let $\mathcal{T}_s$ be the set of switchers at s, and for $j \in \mathcal{T}_s$ with $y_{j,s-1} = y_{i,s-1}$ let Z_{js} be the within-agent regressor defined before Corollary 1 and $b_{js} = \mathbb{1}\{y_j \in B_s\}$. Then $\hat{d}_s(w_{is})$ solves

$$\sum_{j \in \mathcal{T}_s:\ y_{j,s-1} = y_{i,s-1}} K_1\left(\frac{\hat{\delta}_{ijs}^2}{h_1}\right) K_2\left(\frac{\|X_{j,s+2} - X_{j,s+1}\|^2}{h_2}\right)\Big[b_{js} - F\big(Z_{js}'\hat{\theta} - d\big)\Big] = 0 \qquad (24)$$

in d, which exists and is unique whenever the weighted switchers include both orderings, because the left side is strictly increasing in d. A cell in which no switcher receives positive

weight, or in which all weighted switchers have the same ordering, returns no finite estimate, which is the sample counterpart of the support qualification in Corollary 1. In finite-type designs with discrete covariates and bandwidths below the separation gaps, the weights select the switchers of agent i's own type, and (24) is the score equation for the intercept of a conditional logit within that cell with the slopes fixed at $\hat{\theta}$. The slopes can be taken from the class-conditional estimator (19), which removes the class intercepts by conditioning on each class's number of switches of each kind and never estimates them, and (24) then recovers each intercept given those slopes. Estimating θ and the class intercepts jointly, by a logit with one dummy per class on the eligible agents, is a different procedure: it is first-order equivalent in finite-type designs, as the proof of Proposition 2 shows, but it estimates one nuisance parameter per class from that class's members, and in the simulations of Section 7 its small-sample bias is larger than that of exact conditioning (Table 18). Section 6.6 explains why no distribution theory is available in continuous designs. In finite-type designs, a theory for (24) would have to account for the first-step estimation of θ, whose error is of the same order as the cell-average error, and it is not developed here.

C Extension to Ordered Outcomes

The binary logit framework of the main text can be extended to accommodate ordered discrete outcomes. Many outcomes of interest in applied work are naturally ordinal: self-reported health takes values from "poor" to "excellent," educational attainment is coded in ordered categories, and substance use, as in the Glasgow TFLS data where tobacco consumption is classified as non-smoker, occasional, or regular, is measured on ordinal scales. This section shows how the identification and estimation strategy extends to the dynamic ordered logit model with endogenous networks, using the threshold-dichotomization insight of Muris et al. (2025) and Honoré et al. (2025). The extension presented here is deliberately minimal: it dichotomizes the outcome at the single, known state-dependence threshold c, which reduces the ordered model exactly to the binary framework of the preceding sections. A composite version, pooling the conditional likelihoods across all thresholds (whose dichotomized errors are perfectly dependent, since they share the same latent ε_{it}), is left for future work, and no formal results are claimed beyond those inherited from the binary case.

C.1 Model

Consider an ordered outcome $Y_{it} \in \{1, \ldots, J\}$ generated by a latent variable model:

$$Y_{it}^* = X_{it}'\beta + \alpha^{\text{ord}}\, \mathbb{1}\{Y_{it-1} \geq c\} + \lambda(w_{it}) + \mu_i - \varepsilon_{it}, \tag{25}$$

$$Y_{it} = \begin{cases} 1 & \text{if } Y_{it}^* < \varsigma_2, \\ 2 & \text{if } \varsigma_2 \leq Y_{it}^* < \varsigma_3, \\ \vdots & \\ J & \text{if } Y_{it}^* \geq \varsigma_J, \end{cases}$$

where $\varsigma_2 < \varsigma_3 < \cdots < \varsigma_J$ are unknown threshold parameters, $c \in \{2, \ldots, J\}$ is a known cutoff for the lagged dependent variable, α^{ord} is the autoregressive parameter, and all other quantities are as in the binary model. The error term ε_{it} follows a standard logistic distribution, and $\lambda(w_{it})$ is the unknown social influence function of the latent index w_{it}, which also governs network formation through (2). The state dependence specification $\alpha^{\text{ord}}\, \mathbb{1}\{Y_{it-1} \geq c\}$ uses a binary indicator of whether the previous outcome exceeded a threshold, following Muris et al. (2025), and it nests the binary model when $J = 2$ and $c = 2$.

C.2 Dichotomization and Identification

The key insight, due to Muris et al. (2025), is that the ordered model can be analyzed through a collection of binary logit models obtained by dichotomizing the outcome at each threshold. I normalize $\varsigma_c = 0$ without loss of generality, where c is as in equation (25). Define the binary indicator:

$$B_{it}(c) = \mathbb{1}\{Y_{it} \geq c\}.$$

Under the logistic distributional assumption, $B_{it}(c)$ follows a binary logit model:

$$B_{it}(c) = \mathbb{1}\{X_{it}'\beta + \alpha^{\text{ord}}\, B_{it-1}(c) + \lambda(w_{it}) + \mu_i - \varepsilon_{it} \geq 0\}. \tag{26}$$

This has the same structure as the binary model in (1), with $B_{it-1}(c)$ playing the role of the lagged outcome. For the network part of the argument to carry over, the link function must depend on the lagged outcome only through $B_{i,t-1}(c)$. If it depends on the full lagged outcome $Y_{i,t-1}$, the codegree distances would have to be stratified on $Y_{i,t-1}$, and the conditioning and swap-invariance arguments of Section 4 would have to be redone for that finer stratification, in which a discordant pair need not share the lagged outcome at the periods used for matching. That case is not covered by the binary proof and is not claimed here. The identification strategy of Section 4 therefore applies to the binary model at this specified cutoff, subject to the same assumptions: matching agents with the same network type eliminates the change of $\lambda(w_{it})$ between the transition periods.

The estimator replaces y_{it} with $B_{it}(c)$ throughout the objective function (14), that is, in the discordant-pair sets $\mathcal{P}_s$, the conditioning variables v_{ijs}, and the contributions m_{ijs}. Let $\Omega_n^{\text{ord}}(b, a)$ denote the resulting objective. Information at the other thresholds does not enter it, so the estimator is the binary estimator applied to $B_{it}(c)$, and applying the results of Sections 4–6 requires their assumptions to hold for this dichotomized model. The consistency result and the finite-type distribution theory then carry over, and the limitations for continuous designs discussed in Section 6.6 carry over unchanged. The estimator is:

$$(\hat{\beta}^{\text{ord}}, \hat{\alpha}^{\text{ord}}) = \arg\max_{(b,a)}\ \Omega_n^{\text{ord}}(b, a).$$

Recovery of the other thresholds ς_j, $j \neq c$, would require additional identification and estimation arguments for probabilities across cutoffs, which are not developed here.

The ordered logit extension is directly relevant to the empirical application in Section 8. The TFLS records tobacco use on a three-point scale: non-smoker ($Y_{it} = 1$), occasional smoker ($Y_{it} = 2$), and regular smoker ($Y_{it} = 3$). The latent ordered model distinguishes these categories, but the proposed single-cutoff estimator uses only whether the student is

above or below the chosen cutoff. With $J = 3$ and $c = 2$ (state dependence then operates through whether the student smoked at all in the previous wave), the binary reduction distinguishes non-use from any tobacco use and does not separately estimate the transition from occasional to regular use. This tobacco-only illustration is distinct from the tobacco-or-cannabis composite used in Section 8.

D The Observation Window in the Smoking Example

In the smoking example, Assumption 2 describes a particular observation window. Consider a cohort observed at four dates, $t = 0, 1, 2, 3$, at comparable intervals, around a scheduled school transition between $t = 1$ and $t = 2$, for instance the move into a new stage of secondary school with new subjects, classes, and timetables. The outcome at $t = 0$ precedes the window and fixes the lag stratum. The outcome and the network at $t = 1$ precede the transition, and that network carries the information about each student's network class. The outcomes at $t = 2$ and $t = 3$ follow the transition, the second later within the same stage. A special case of Assumption 2 is

$$\lambda(w_{i1}) = \ell_{c(i)}, \qquad \lambda(w_{i2}) = \lambda(w_{i3}) = \ell_{c(i)} + \Delta_{c(i)},$$

where $c(i)$ is student i's network class before the transition, the baseline ℓ_c and the increment Δ_c may differ in sign and size across classes, and persistent differences between students of the same class are carried by the agent effect μ_i. The display is a special case of the model when a latent index and a link function that represent it exist, for instance a finite latent state that records the class and whether the transition has occurred. Together with the model, the assumption then has three substantive parts. First, students who formed friendships in the same way before the transition share the change in their pull toward smoking that the transition brings, for example because the change operates through their position in the friendship structure, such as the salience of status or of a peer crowd, while their persistent predispositions can differ. Second, the adjustment is complete by $t = 2$, and the latent term does not move between $t = 2$ and $t = 3$, which is more plausible when the last two dates are close together within the same stage than when they are a year apart. Third, the state-dependence parameter is the same across the intervals, which requires comparable spacing of the dates. The second and third parts pull in different directions if only the last interval is shortened, so the design calls for dates that are close together and evenly spaced throughout, which in turn makes switches rarer. The realized network can still change within the window, because links depend on lagged smoking and link shocks change, so stability

concerns the latent contribution to the outcome and not every friendship. A scheduled transition does not by itself imply any of these conditions: continued individual adjustment after $t = 2$, a latent propensity that responds to recent smoking, and a direct effect of friends' behavior all fall outside the model, and the example illustrates the design that the assumption requires rather than showing that school transitions satisfy it. If every increment Δ_c is zero, within-agent conditioning is valid and the network is not needed. If the increments differ across classes, a single common time effect does not remove them. In the Glasgow data of Section 8, the three waves are a year apart and were collected in the second, third, and fourth years of secondary school. In the Scottish system of the time, the third and fourth years were those of the Standard Grade courses (McVittie, 2008), so the first two waves straddle a scheduled change of stage and the last two lie within one stage. The data nevertheless fall short of this design: the outcome before the window is not observed, and stability of the latent term over a full year is a strong requirement.

E Descriptive Fits for the Glasgow Data

This appendix reports the five descriptive fits referred to in Section 8.4, together with the robustness exercises on the outcome definition and the initial condition. None of these fits is the estimator of Section 5, and no inferential claim is attached to any of them.

Five specifications are reported in Table 10: the naive dynamic logit, with $y_{i,t-1}$ and all seven covariates (pocket money, romantic relationship, alcohol, leisure, female, age, and smoking at home); the dynamic logit with network controls, which adds peer averages of every covariate and of lagged smoking, computed on the any-friend network at each wave; a within-agent conditional logit in the spirit of Honoré and Kyriazidou (2000); a cross-agent pairwise conditional logit using the four time-varying covariates ($k = 4$); and the class-conditional logit (19) at the same unlocalized bandwidth, whose single feasible class contains the 22 switching students, so that it is the within-agent conditional logit with one common movement of the social influence term, conditioned out. The last two columns are not the proposed estimators, for the reasons given in Section 8.4: the covariate-stability kernel is infeasible and the codegree kernel does not bind, so they omit both. The network-localized version is reported in Section 8.4. No significance levels are attached to any column: the pooled standard errors ignore the dependence within students, and the conditional standard errors are not justified by any result in this paper.

The three conditional columns are reported with Firth penalization (Firth, 1993): the criterion is the log-likelihood plus one half of the log determinant of its information matrix, computed as if the agents (column 3) or the pairs (column 4) were independent observations,

and, for column 5, the information of the class-conditional likelihood, as in the penalized conditional logistic regression of Heinze and Puhr (2010). Penalization is used because the unpenalized likelihoods are near-separated: a linear-programming check shows that none of them is separated in the sense of Albert and Anderson (1984), so the three unpenalized maxima exist, but many fitted probabilities at the within-agent and pairwise maxima are extreme. For a pairwise criterion in which pairs sharing a student are dependent, the bias-reduction interpretation of Firth's penalty for independent observations does not carry over, and the standard errors reported for the penalized fits are not justified by any result in this paper and are reported for scale only.

The state dependence estimates fall as more conditioning is imposed. The naive and controls estimates of α are 2.222 and 2.101. The within-agent fit returns 1.827, the cross-agent pairwise fit 1.413, with a standard error for scale of 1.368, and the class-conditional fit 0.893, with a standard error for scale of 1.536. The unpenalized class-conditional value, 2.134, is close to the unpenalized pairwise value, 2.281. The two maximize the full and the pairwise conditional likelihoods of the same single class, but with 22 agents and the Γ remainder of equation (6) present in both objectives their closeness is descriptive only, and the gap between their penalized values reflects the different penalties. The last two steps are not the mechanism this paper proposes: columns (4) and (5) compare all switchers with one another without using the network, which allows a common movement of λ for all students but not a movement that differs across network classes, so the move from 1.827 to 1.413 and 0.893 reflects a change from within-agent to cross-agent conditioning, not the removal of the class-specific movement that codegree matching targets, and it carries no theoretical claim to remove confounding that HK leaves behind. Coefficient differences do not measure how much confounding each method removes, since the estimators also differ in conditioning and penalization, and the last step should not be read as evidence about the network channel, whose feasibility in these data is the subject of Section 8.4 and Table 9.

Three qualifications apply, because the data alone cannot support the ordering. The estimates are imprecise, and their pairwise differences are untested. The ordering depends on the penalization: without it, the within-agent estimate rises to 3.884 and the monotonicity breaks, so the declining sequence is an artifact of the penalization. And $n = 96$ is far below the sample sizes of Section 7, so nothing there implies that column (4) or column (5) is more reliable than column (3) *in this sample*. The ordering of the point estimates is descriptive, and these data cannot adjudicate between the conditional fits.

Where the estimators differ most is in the covariate effects, and there the ordering runs the other way. Across all four covariates, the three conditional estimators produce coefficients larger in magnitude than the naive and controls specifications, and column (4) produces the

largest. The larger magnitudes are not evidence that confounding was removed. In a logit model, omitted heterogeneity that is independent of the covariates shrinks pooled coefficients toward zero, so conditional estimates are larger in magnitude even without any confounding. Confounding does not have a predictable sign here either: the pooled coefficient combines the direction of the correlation between the omitted heterogeneity and the covariate with the attenuation from averaging over that heterogeneity, and positive correlation, as under the motivating story in which a taste for risk raises both alcohol use and smoking, is compatible with a pooled coefficient below the structural one. With the stability kernel omitted, the Γ remainder of equation (6) generally still depends on the level of the individual effect and of λ, so none of the conditional columns delivers the cancellation of Lemma 1, and columns (4) and (5), which do not use the network, cannot be credited with removing the class-specific part of the movement of λ. The differences between the columns are differences between objectives. The near-separation diagnostics give a further reason for caution: at the unpenalized maxima, half of the HK fitted probabilities and 76% of column (4)'s pair-level fitted probabilities lie outside $[0.01, 0.99]$, a configuration in which conditional logit coefficients are known to inflate. The conditional columns are therefore reported under Firth penalization, which shrinks every conditional coefficient, by roughly a third to a half in columns (3) and (4) and by more than half in column (5), while leaving the sign pattern intact. The unpenalized magnitudes should not be quoted as effect sizes, since odds-ratio translations of coefficients near 5 on 96 pairs are not meaningful.

The peer-average controls change the alcohol and state-dependence coefficients little, by 0.008 and 0.12, respectively. This parallels the simulations of Section 7, in which peer averages remove little of the confounding when neighbors' characteristics do not proxy for an agent's own: peer averages recover only the part of the confounding that is reflected in neighbors' *observables*, and small coefficient movements do not establish how much confounding remains.

Alcohol consumption is the strongest of the covariate results: the coefficient is positive in every column, moderate in the naive and controls specifications (0.80–0.81), and larger in the conditional fits (1.50, 3.25, and 1.70, under penalization). Its different magnitudes cannot by themselves establish a causal alcohol effect or the direction of omitted-variable bias, and alcohol use is among the covariates for which the strict exogeneity of Assumption 1(iv) is least plausible.

The remaining covariates show the same directional pattern with less strength. Leisure is small relative to its standard error in both pooled specifications and positive in the three conditional fits. Because the index rises as reported activity falls, the positive coefficient associates less frequent reported leisure activity with use. Romantic relationship is positive in the naive specification, smaller once peer averages are added, and positive in the three

conditional fits. The pocket money coefficient, per pound per month, is imprecisely estimated in every column. The differences between columns (3) and (4) are within the imprecision of the fits and are not evidence that cross-agent pairing detects effects that within-agent conditioning misses.

Among the additional covariates available to the naive and controls specifications, the female coefficient is negative in the naive logit and close to zero once peer averages are included. Age, which increases by one year per wave by construction, has a small coefficient in both specifications, which does not by itself establish whether age or other time-varying factors account for the prevalence trend. Smoking at home is positive but small relative to its standard error in both specifications.

Table 11 reports the robustness exercises that the initial-condition and outcome-definition assumptions call for. Under a tobacco-only outcome, the discordant sample falls from 22 to 16 students and the pair count from 96 to 48. The within-agent state dependence estimate drops to 1.00 (standard error 0.87) under penalization, although the unpenalized within-agent likelihood is quasi-completely separated, because only one of the 16 switching students changes romantic status between waves 1 and 2. The pairwise likelihood is completely separated, as a linear-programming certificate confirms, so its penalized estimate of -1.98 carries no information, and the class-conditional likelihood is completely separated as well, so its penalized estimate of -1.27 carries no information either. The positive state dependence in the conditional fits therefore rests on the composite tobacco-or-cannabis outcome, whose persistence may partly reflect the recording of past cannabis initiation. The initial-condition exercises are more informative about what the assumption does than about the estimates. Setting $y_{i0} = 1$ for the ten wave-1 regular smokers leaves the three conditional fits unchanged to three decimals, because none of those ten students is among the 22 who transition between waves 1 and 2, so the imputation never enters a conditional likelihood, and only the pooled columns move. Setting $y_{i0} = y_{i1}$ instead, the most pessimistic reading in which every wave-1 user already used a year earlier, leaves the pairwise estimator with no pairs at all: lag matching requires $y_{i0} = y_{j0}$ while discordance requires $y_{i1} \neq y_{j1}$, and the two cannot hold together once $y_{i0} = y_{i1}$. For the same reason, each lag stratum then contains switchers of a single ordering, and the class-conditional estimator is not defined. It also separates the within-agent likelihood completely: with $y_{i0} = y_{i1}$, the lagged-outcome regressor $y_{i0} - y_{i3}$, oriented by the observed switch, is never negative, and together with the covariates it orders every switching student correctly, so the within-agent value of 2.87 is a penalized value without an unpenalized counterpart. In this sense, the initial condition is load-bearing here. With three waves the estimator has exactly the four outcomes per student that the transition event needs, and the first of them must be supplied by assumption. An application in which

the paper's mechanism could be tested needs at least four waves, so that y_{i0} is observed, together with a network observed without a nomination cap.

Table 10: Empirical estimates: Any friend network.

	(1) Naive	(2) Controls	(3) HK	(4) Pairwise	(5) Class
Intercept	−4.931	−8.039	—	—	—
	(2.779)	(3.523)			
Pocket money	0.014	0.010	0.062	0.099	0.059
	(0.019)	(0.020)	(0.046)	(0.050)	(0.064)
Romantic relationship	0.815	0.543	2.423	4.943	2.741
	(0.346)	(0.381)	(1.731)	(2.272)	(2.365)
Alcohol	0.808	0.800	1.501	3.253	1.698
	(0.174)	(0.185)	(0.414)	(0.858)	(0.988)
Leisure	0.371	0.087	2.436	5.403	2.881
	(0.532)	(0.564)	(1.712)	(2.740)	(2.713)
Female	−0.773	−0.021	—	—	—
	(0.315)	(0.638)			
Age	0.070	0.285	—	—	—
	(0.195)	(0.245)			
Smoking at home	0.138	0.207	—	—	—
	(0.303)	(0.324)			
Lagged smoking ($\hat{\alpha}$)	2.222	2.101	1.827	1.413	0.893
	(0.420)	(0.444)	(0.668)	(1.368)	(1.536)
Peer average covariates	—	YES	—	—	—
Firth penalization	—	—	YES	YES	YES
Observations	288	288	22 agents	96 pairs	22 agents

Standard errors in parentheses, reported for scale only. No significance levels are attached to any column. Columns (1) and (2) are pooled dynamic logits with unclustered MLE standard errors on the 288 student-wave observations, which ignore dependence within students. Column (2) adds peer averages of all covariates and of lagged smoking, whose coefficients are not reported. Column (3) is a within-agent conditional fit in the spirit of Honoré and Kyriazidou (2000), omitting K_2, on the transition event $y_{i1} + y_{i2} = 1$ with regressors $X_{i1} - X_{i2}$ and $y_{i0} - y_{i3}$. Column (4) is the cross-agent pairwise conditional logit obtained as the unlocalized limit of the proposed estimator (all 96 discordant lag-matched pairs at unit weight, no codegree kernel, and no K_2). It does not use the network and is identical across network definitions. Column (5) is the class-conditional logit (19) at the same unlocalized bandwidth: the only feasible class is the lag stratum $y_{i0} = 0$, which contains all 22 switchers, 6 of them with the ordering B_1, and the likelihood conditions on that number, so it also does not use the network. Columns (3) and (4) use sandwich standard errors clustered on the 22 discordant students, and column (5) the inverse of the class-conditional information at the penalized estimate, so their precision is not comparable with that of columns (1) and (2). The conditional columns have no intercept and no coefficients on female or smoking at home, which are time-invariant and difference out. Age rises by one year per wave for every student, so its difference is a constant: it cancels in the cross-agent column (4) and in the class-conditional column (5), while column (3) omits it, which amounts to setting the common time effect of the within-agent ordering equation to zero. All three are reported with the Firth penalization described in the text. No unpenalized conditional likelihood is separated, as a linear-programming check confirms (for column 5, with the class intercept), so the unpenalized maxima exist. The HK and pairwise likelihoods are near-separated, with 50% of HK fitted probabilities and 76% of column (4)'s pair-level fitted probabilities outside $[0.01, 0.99]$, although extreme fitted probabilities alone do not establish separation. The unpenalized estimates of α are 3.884 (HK, standard error 2.018), 2.281 (pairwise, standard error 1.823), and 2.134 (class, standard error 2.411). Penalization shrinks every conditional coefficient, by roughly a third to a half in columns (3) and (4) and by more than half in column (5), and leaves the sign pattern intact.

Table 11: Robustness of the empirical estimates to the outcome definition and the initial condition.

	Baseline (any use, $y_{i0} = 0$)	Tobacco only	$y_{i0} = 1$ for wave-1 regular smokers	$y_{i0} = y_{i1}$
Students / discordant / pairs	96 / 22 / 96	97 / 16 / 48	96 / 22 / 96	96 / 22 / 0
Lagged smoking ($\hat{\alpha}$)				
Naive	2.222 (0.420)	2.816 (0.480)	2.438 (0.406)	3.893 (0.459)
Controls	2.101 (0.444)	2.447 (0.521)	2.299 (0.426)	3.743 (0.473)
HK	1.827 (0.668)	1.001‡ (0.874)	1.827 (0.668)	2.873† (0.502)
Pairwise	1.413 (1.368)	$-1.983^{\dagger}$ (1.683)	1.413 (1.368)	—
Class	0.893 (1.536)	$-1.274^{\dagger}$ (2.981)	0.893 (1.536)	—
Alcohol				
Naive	0.808 (0.174)	0.805 (0.182)	0.763 (0.179)	0.917 (0.221)
HK	1.501 (0.414)	1.308‡ (0.555)	1.501 (0.414)	1.241† (0.457)
Pairwise	3.253 (0.858)	2.304† (0.821)	3.253 (0.858)	—
Class	1.698 (0.988)	1.768† (1.258)	1.698 (0.988)	—

Standard errors in parentheses. HK, pairwise, and class rows are under Firth penalization as in Table 10, where the class fit uses the single unlocalized class. "Tobacco only" redefines the outcome as any tobacco use, dropping cannabis. One student with complete tobacco records but a missing cannabis record re-enters the complete-case sample, which is why that column has 97 students. The third column sets $y_{i0} = 1$ for the ten students who report regular tobacco use at wave 1 and zero otherwise, and the fourth sets $y_{i0} = y_{i1}$, so that every wave-1 user is assumed to have been a user one year earlier. †: the unpenalized likelihood is completely separated, as certified by a linear program (a direction orders every row correctly, and the largest minimum margin over unit-norm directions is 0.235 for the pairwise fit, 0.109 for the class fit with its class intercept, and 0.083 for the within-agent fit), so it has no maximum and the penalized value is reported for completeness only. ‡: the unpenalized within-agent likelihood is quasi-completely separated, because only one of the 16 switching students changes romantic status between waves 1 and 2, so its maximum is not attained, and the penalized value is reported for completeness only. The remaining conditional fits in the table are not separated by the same check, so their unpenalized maxima exist. —: no discordant lag-matched pair, or no class that contains both orderings, exists, so the estimator is not defined.

F Supplementary Tables

This appendix reports the Monte Carlo results that are summarized but not tabulated in Section 7: the finite-type design with lag-dependent links (Tables 12 and 13), the benchmark designs at $n = 1000$ (Table 14), the matching and inference diagnostics for the benchmark designs (Table 15), the comparisons on common successful replications (Table 16), the pairwise and class-conditional estimators in the benchmark and continuous designs (Table 17), the class-based alternatives (Table 18), and the continuous design (Tables 19 and 20). The simulation code, the population constants it uses, and the replication-level output are available with the paper.

Table 12: Monte Carlo results for β: finite-type design with lag-dependent links

n	Naive	Controls	HK	Proposed	Oracle	Class	Class orc.	Infeas.
Bias								
500	0.245	0.240	0.372	0.234	0.230	0.145	0.135	0.007
1000	0.239	0.233	0.285	0.143	0.143	0.084	0.084	0.000
2000	0.242	0.234	0.264	0.068	0.068	0.047	0.047	0.004
3000	0.239	0.231	0.251	0.043	0.043	0.033	0.033	0.000
Standard deviation								
500	0.069	0.069	0.468	0.632	0.619	0.528	0.523	0.073
1000	0.052	0.052	0.295	0.411	0.412	0.344	0.344	0.057
2000	0.035	0.035	0.200	0.262	0.262	0.227	0.227	0.038
3000	0.030	0.030	0.156	0.187	0.187	0.169	0.169	0.032
Root mean squared error								
500	0.254	0.250	0.598	0.673	0.660	0.547	0.540	0.074
1000	0.245	0.239	0.410	0.435	0.435	0.354	0.354	0.057
2000	0.245	0.237	0.331	0.270	0.271	0.232	0.232	0.038
3000	0.241	0.233	0.295	0.191	0.191	0.172	0.172	0.032
Coverage of nominal 95% interval								
500	0.050	0.068	0.964	0.965	0.967	0.984	0.981	0.956
1000	0.002	0.002	0.890	0.950	0.952	0.968	0.970	0.938
2000	0.000	0.000	0.812	0.948	0.948	0.960	0.960	0.954
3000	0.000	0.000	0.686	0.950	0.950	0.962	0.962	0.948
Replications without an estimate								
500	0	0	3	75	76	74	76	0
1000	0	0	1	1	1	1	1	0
2000	0	0	0	0	0	0	0	0
3000	0	0	0	0	0	0	0	0

Notes: $R = 500$ replications; $\beta_0 = 1$, $\alpha_0 = 0.5$. Naive: pooled dynamic logit. Controls: pooled dynamic logit with neighbors' average covariate and lagged outcome. HK: within-agent conditional logit of Honoré and Kyriazidou (2000). Proposed: estimator (15). Oracle: the same pairwise estimator with population matches. Class: the class-conditional estimator (19), with classes formed from the same estimated distances and bandwidth. Class orc.: the same likelihood with the population classes. Infeas.: pooled dynamic logit including the true unobserved term $\lambda(w_{it}) + \mu_i$. Bandwidths: $h_1 = 0.5\,\bar{\epsilon}\,(n/500)^{-1/4}$, with $\bar{\epsilon}$ the population separation gap of the design, a fixed constant computed from the block matrix and the population shares of the (type, y_{i0}) cells, and exact covariate stability ($h_2 = 0.5$ on the integer covariate). Coverage: share of replications in which the nominal 95% Wald interval contains the true value. Standard errors are MLE for the pooled logits, a sandwich for HK, the sandwich clustering pair-level scores by agent for the pairwise estimators, and the inverse of the negative Hessian of the class-conditional log likelihood for the class estimators. Statistics are computed over the replications with an estimate, and failures are defined in Section 7.2.

Table 13: Monte Carlo results for α: finite-type design with lag-dependent links

n	Naive	Controls	HK	Proposed	Oracle	Class	Class orc.	Infeas.
Bias								
500	0.715	0.666	0.430	−0.055	−0.053	−0.046	−0.046	−0.005
1000	0.711	0.650	0.425	0.008	0.008	0.005	0.005	−0.005
2000	0.711	0.643	0.400	0.000	0.000	0.001	0.001	0.001
3000	0.715	0.644	0.363	0.004	0.004	0.009	0.009	0.004
Standard deviation								
500	0.126	0.141	0.752	1.427	1.421	1.206	1.226	0.145
1000	0.093	0.111	0.498	0.917	0.916	0.792	0.792	0.102
2000	0.070	0.079	0.314	0.551	0.551	0.472	0.472	0.073
3000	0.056	0.064	0.269	0.415	0.415	0.391	0.391	0.060
Root mean squared error								
500	0.726	0.680	0.865	1.427	1.420	1.205	1.225	0.145
1000	0.717	0.659	0.655	0.917	0.915	0.791	0.791	0.103
2000	0.715	0.648	0.508	0.550	0.550	0.472	0.472	0.073
3000	0.717	0.647	0.452	0.415	0.415	0.391	0.391	0.060
Coverage of nominal 95% interval								
500	0.000	0.006	0.909	0.941	0.946	0.984	0.981	0.962
1000	0.000	0.000	0.876	0.942	0.942	0.952	0.952	0.954
2000	0.000	0.000	0.790	0.938	0.938	0.950	0.950	0.944
3000	0.000	0.000	0.702	0.946	0.946	0.950	0.950	0.956
Replications without an estimate								
500	0	0	3	75	76	74	76	0
1000	0	0	1	1	1	1	1	0
2000	0	0	0	0	0	0	0	0
3000	0	0	0	0	0	0	0	0

Notes: $R = 500$ replications; $\beta_0 = 1$, $\alpha_0 = 0.5$. Naive: pooled dynamic logit. Controls: pooled dynamic logit with neighbors' average covariate and lagged outcome. HK: within-agent conditional logit of Honoré and Kyriazidou (2000). Proposed: estimator (15). Oracle: the same pairwise estimator with population matches. Class: the class-conditional estimator (19), with classes formed from the same estimated distances and bandwidth. Class orc.: the same likelihood with the population classes. Infeas.: pooled dynamic logit including the true unobserved term $\lambda(w_{it}) + \mu_i$. Bandwidths: $h_1 = 0.5\,\bar{\epsilon}\,(n/500)^{-1/4}$, with $\bar{\epsilon}$ the population separation gap of the design, a fixed constant computed from the block matrix and the population shares of the (type, y_{i0}) cells, and exact covariate stability ($h_2 = 0.5$ on the integer covariate). Coverage: share of replications in which the nominal 95% Wald interval contains the true value. Standard errors are MLE for the pooled logits, a sandwich for HK, the sandwich clustering pair-level scores by agent for the pairwise estimators, and the inverse of the negative Hessian of the class-conditional log likelihood for the class estimators. Statistics are computed over the replications with an estimate, and failures are defined in Section 7.2.

Table 14: Benchmark designs at $n = 1000$

	HK			Proposed			Oracle			Class		
Design	Bias	RMSE	Cov.	Bias	RMSE	Cov.	Bias	RMSE	Cov.	Bias	RMSE	Cov.
β ($\beta_0 = 1$)												
Designs that satisfy the assumptions												
Baseline	0.296	0.433	0.88	0.151	0.458	0.95	0.151	0.459	0.95	0.085	0.381	0.97
Lag-dependent links	0.285	0.410	0.89	0.143	0.435	0.95	0.143	0.435	0.95	0.084	0.354	0.97
Fixed types	0.075	0.315	0.94	0.132	0.397	0.96	0.132	0.397	0.96	0.097	0.351	0.96
Weak network	0.277	0.403	0.92	–	–	–	0.133	0.451	0.95	–	–	–
Designs outside the assumptions												
Idiosyncratic movement	0.265	0.398	0.91	0.190	0.474	0.95	0.190	0.472	0.95	0.151	0.391	0.97
Covariates in links	0.257	0.380	0.91	0.223	0.733	0.97	0.132	0.453	0.95	0.119	0.511	0.97
Sparse network	0.306	0.438	0.88	0.857	1.336	0.85	0.166	0.467	0.93	0.445	0.671	0.98
α ($\alpha_0 = 0.5$)												
Designs that satisfy the assumptions												
Baseline	0.377	0.614	0.90	0.000	0.872	0.94	0.000	0.872	0.94	0.003	0.757	0.96
Lag-dependent links	0.425	0.655	0.88	0.008	0.917	0.94	0.008	0.915	0.94	0.005	0.791	0.95
Fixed types	−0.004	0.466	0.96	−0.037	0.719	0.97	−0.036	0.719	0.97	−0.041	0.645	0.97
Weak network	0.395	0.601	0.90	–	–	–	0.018	0.849	0.95	–	–	–
Designs outside the assumptions												
Idiosyncratic movement	0.365	0.588	0.90	0.137	0.786	0.95	0.136	0.784	0.95	0.132	0.686	0.97
Covariates in links	0.413	0.644	0.88	0.036	1.114	0.96	0.017	0.869	0.94	−0.004	0.875	0.98
Sparse network	0.400	0.616	0.88	0.512	2.026	0.85	0.005	0.890	0.94	0.387	1.063	0.97

Notes: $R = 500$ replications. Bias, root mean squared error, and coverage of the nominal 95% interval, computed over the replications with an estimate. HK: within-agent conditional logit of Honoré and Kyriazidou (2000). Proposed: estimator (15). Oracle: the same pairwise estimator with population matches. Class: the class-conditional estimator (19) with classes formed from the same estimated distances and bandwidth. The sparse design uses the 0.10 quantile of the estimated distances as h_1, and all other finite-type designs use $h_1 = 0.5\,\bar{\epsilon}\,(n/500)^{-1/4}$ with $\bar{\epsilon}$ the population separation gap of the design (Section 7.2). In the weak-network design at $n = 1000$, no pair has an estimated distance below the bandwidth, so the proposed and class estimators cannot be computed (–). The covariate-link and sparse designs violate the maintained network model (covariates in the link function and bounded degrees, respectively), and the idiosyncratic design violates Assumption 2(i). Failed fits are optimizer failures, all certified as separated, unless stated otherwise. Lag-dependent links: 1 each for Proposed, Oracle, HK, and Class. Weak network: Proposed 500 (fewer than three pairs) and Class 500 (fewer than three agents in classes with both orderings). Idiosyncratic movement: 1 each for Proposed, Oracle, and Class. Covariates in links: Proposed 18 and Class 12. Sparse network: Proposed 79 (78 certified as separated), Oracle 7, HK 1, and Class 21.

Table 15: Benchmark designs: matching, contributing-agent, and inference diagnostics for the proposed estimator

				Agents				Matching		Proposed estimator					
Design	n	Pairs	ESS	Elig.	Contrib.	Unm.	Min.	Prec.	Recall	Par.	Med. bias	SD	Mean SE	SD/SE	Coverage
Fixed types	1000	206	206	79	78	1	1.5	1.00	1.00	β	0.103	0.375	0.344	1.09	0.960
										α	−0.037	0.719	0.718	1.00	0.966
Fixed types	2000	830	830	159	159	0	3.5	1.00	1.00	β	0.024	0.219	0.213	1.03	0.958
										α	0.005	0.453	0.452	1.00	0.954
Weak network	1000	0	–	100	0	100	–	–	–	β	–	–	–	–	–
										α	–	–	–	–	–
Weak network	2000	346	262	200	163	36	1.0	1.00	0.45	β	0.076	0.346	0.299	1.15	0.940
										α	−0.060	0.719	0.656	1.10	0.926
Weak network	3000	1369	1214	299	289	10	1.0	1.00	0.79	β	0.040	0.219	0.200	1.09	0.924
										α	0.010	0.462	0.440	1.05	0.940
Idiosyncratic movement	1000	219	218	99	90	8	1.3	1.00	1.00	β	0.118	0.434	0.364	1.19	0.950
										α	0.101	0.775	0.776	1.00	0.952
Idiosyncratic movement	2000	889	889	198	195	3	2.0	1.00	1.00	β	0.104	0.256	0.232	1.10	0.926
										α	0.116	0.520	0.500	1.04	0.952
Covariates in links	1000	61	54	100	57	43	1.0	1.00	0.32	β	0.123	0.699	0.678	1.03	0.969
										α	−0.012	1.114	1.075	1.04	0.959
Covariates in links	2000	245	222	200	153	47	1.0	1.00	0.32	β	0.073	0.381	0.375	1.02	0.946
										α	−0.047	0.641	0.638	1.00	0.964
Sparse network	1000	122	85	100	47	53	1.0	0.17	0.11	β	0.640	1.027	0.685	1.50	0.850
										α	0.322	1.963	1.343	1.46	0.848
Sparse network	2000	496	344	201	103	98	1.0	0.17	0.11	β	0.449	0.622	0.420	1.48	0.783
										α	0.408	1.093	0.913	1.20	0.873

Notes: See the notes to Table 3. Failed fits of the proposed estimator are optimizer failures, all certified as separated, unless stated otherwise. Weak network: 500 at $n = 1000$, all with fewer than three pairs. Idiosyncratic movement: 1 at $n = 1000$. Covariates in links: 18 at $n = 1000$. Sparse network: 79 at $n = 1000$ (78 certified as separated) and 2 at $n = 2000$.

Table 16: Comparisons on common successful replications

		Successful fits			Own successes			Common with HK			Common with oracle		
Design	n	Prop.	HK	Orc.	Prop.	HK	Orc.	Fits	Prop.	HK	Fits	Prop.	Orc.
β													
Baseline	500	431	488	431	0.677	0.568	0.670	431	0.677	0.506	431	0.677	0.670
Lag-dependent links	500	425	497	424	0.673	0.598	0.660	425	0.673	0.491	424	0.674	0.660
Idiosyncratic movement	1000	499	500	499	0.474	0.398	0.472	499	0.474	0.395	499	0.474	0.472
Covariates in links	1000	482	500	500	0.733	0.380	0.453	482	0.733	0.364	482	0.733	0.426
Sparse network	1000	421	499	493	1.336	0.438	0.467	421	1.336	0.399	419	1.334	0.438
Sparse network	2000	498	500	500	0.833	0.292	0.239	498	0.833	0.292	498	0.833	0.239
Continuous	500	495	500	494	1.808	0.947	1.655	495	1.808	0.950	493	1.797	1.657
α													
Baseline	500	431	488	431	1.300	0.825	1.297	431	1.300	0.758	431	1.300	1.297
Lag-dependent links	500	425	497	424	1.427	0.865	1.420	425	1.427	0.758	424	1.426	1.420
Idiosyncratic movement	1000	499	500	499	0.786	0.588	0.784	499	0.786	0.587	499	0.786	0.784
Covariates in links	1000	482	500	500	1.114	0.644	0.869	482	1.114	0.629	482	1.114	0.829
Sparse network	1000	421	499	493	2.026	0.616	0.890	421	2.026	0.564	419	2.029	0.859
Sparse network	2000	498	500	500	1.195	0.494	0.501	498	1.195	0.492	498	1.195	0.502
Continuous	500	495	500	494	1.176	0.497	1.131	495	1.176	0.494	493	1.175	1.131

Notes: $R = 500$ replications. A fit succeeds if it returns finite estimates and standard errors for both parameters. Successful fits: number of successes. Own successes: RMSE over each estimator's own successes, as in the other tables. Common with HK (oracle): RMSE of the proposed estimator and HK (the oracle) over the replications in which both succeed (Fits). Only cells whose success sets differ are shown, as elsewhere the two comparisons coincide. No comparison is possible in the weak-network design at $n = 1000$, where the proposed estimator never succeeds.

Table 17: Pairwise and class-conditional estimators: benchmark and continuous designs

		Successful fits			RMSE for β			RMSE for α			Class: coverage		Classes		Time (ms)	
Design	n	Prop.	Class	Common	Prop.	Class	Class orc.	Prop.	Class	Class orc.	β	α	Exact	Used	Prop.	Class
Fixed types	1000	500	500	500	0.397	0.351	0.351	0.719	0.645	0.645	0.962	0.970	99.6	78	0.6	11.1
	2000	500	500	500	0.223	0.207	0.207	0.454	0.433	0.433	0.958	0.946	100.0	159	1.0	21.1
Weak network	1000	0	0	0	–	–	–	–	–	–	–	–	0.0	0	0.1	0.8
	2000	500	500	500	0.362	0.246	0.236	0.721	0.522	0.513	0.940	0.934	1.6	187	0.8	21.3
	3000	500	500	500	0.226	0.190	0.190	0.462	0.387	0.386	0.940	0.956	89.0	294	1.3	34.2
Idiosyncratic movement	1000	499	499	499	0.474	0.391	0.391	0.786	0.686	0.687	0.970	0.970	99.6	90	0.5	11.8
	2000	500	500	500	0.287	0.253	0.253	0.534	0.493	0.493	0.942	0.942	100.0	195	1.0	24.4
Covariates in links	1000	482	488	482	0.733	0.499	0.342	1.114	0.862	0.744	0.969	0.984	0.0	67	0.5	17.8
	2000	500	500	500	0.394	0.288	0.221	0.641	0.516	0.474	0.954	0.952	0.0	173	0.8	20.3
Sparse network	1000	421	479	419	1.334	0.604	0.362	2.029	1.014	0.763	0.977	0.967	0.0	53	0.6	15.6
	2000	498	500	498	0.833	0.426	0.210	1.195	0.730	0.457	0.904	0.912	0.0	114	0.9	16.5
Continuous	500	495	496	495	1.808	1.108	–	1.176	0.788	–	0.966	0.966	–	48	0.4	11.3
	1000	500	500	500	0.768	0.581	–	0.619	0.473	–	0.954	0.956	–	109	0.5	10.3
	2000	500	500	500	0.474	0.367	–	0.396	0.315	–	0.964	0.950	–	218	0.8	22.8
	3000	500	500	500	0.378	0.330	–	0.315	0.249	–	0.944	0.952	–	315	1.2	34.9

Notes: See the notes to Table 5. In the idiosyncratic-movement, covariate-link, and sparse designs, the class oracle uses the true type at $t = 1$ crossed with y_{i0}, as the pairwise oracle does. Failed fits are optimizer failures, all certified as separated, unless stated otherwise. Weak network, $n = 1000$: Proposed 500 (fewer than three pairs) and Class 500 (fewer than three agents in classes with both orderings). Idiosyncratic movement, $n = 1000$: 1 each for Proposed, Class orc., and Class. Covariates in links, $n = 1000$: Proposed 18 and Class 12. Sparse network: Proposed 79 (78 certified as separated), Class 21, and Class orc. 7 at $n = 1000$, and Proposed 2 at $n = 2000$. Continuous, $n = 500$: Proposed 5 and Class 4.

Table 18: Class-based alternatives: exact conditioning, joint class intercepts, and a single class

		β								α							
		Bias				RMSE				Bias				RMSE			
Design	n	Prop.	Class	Joint	One	Prop.	Class	Joint	One	Prop.	Class	Joint	One	Prop.	Class	Joint	One
Baseline	500	0.240	0.149	0.485	0.361	0.677	0.563	0.906	0.602	−0.076	−0.044	0.098	0.335	1.300	1.163	1.521	1.049
	1000	0.151	0.085	0.242	0.296	0.458	0.381	0.516	0.441	0.000	0.003	0.077	0.355	0.872	0.757	0.871	0.723
	2000	0.070	0.042	0.109	0.261	0.262	0.222	0.262	0.326	0.016	−0.005	0.028	0.343	0.543	0.477	0.510	0.548
	3000	0.026	0.009	0.051	0.229	0.201	0.174	0.190	0.277	0.027	0.024	0.046	0.350	0.419	0.380	0.399	0.485
Lag-dependent links	500	0.234	0.145	0.486	0.397	0.673	0.547	0.893	0.642	−0.055	−0.046	0.086	0.296	1.427	1.205	1.593	1.055
	1000	0.143	0.084	0.240	0.289	0.435	0.354	0.487	0.418	0.008	0.005	0.079	0.372	0.917	0.791	0.918	0.743
	2000	0.068	0.047	0.115	0.270	0.270	0.232	0.275	0.339	0.000	0.001	0.034	0.344	0.550	0.472	0.505	0.543
	3000	0.043	0.033	0.076	0.251	0.191	0.172	0.194	0.296	0.004	0.009	0.031	0.342	0.415	0.391	0.409	0.497
Fixed types	1000	0.132	0.097	0.260	0.084	0.397	0.351	0.492	0.326	−0.037	−0.041	0.028	−0.039	0.719	0.645	0.742	0.601
	2000	0.046	0.032	0.098	0.030	0.223	0.207	0.242	0.201	0.023	0.023	0.058	0.018	0.454	0.433	0.465	0.413

Notes: $R = 500$ replications, with statistics over each estimator's own successful fits. Prop.: pairwise estimator. Class: class-conditional estimator (19). Joint: logit with one intercept per class, estimated together with (β, α), on the eligible agents of the feasible classes that contain both orderings (classes with a single ordering are dropped, as in the class-conditional likelihood). One: the class-conditional likelihood with a single class per lag stratum, that is, the within-agent conditional logit with a common time effect conditioned out, which allows one common movement of λ for all eligible agents that share y_{i0} and uses no network information. All failed fits are optimizer failures certified as separated: Joint 65 and One 14 in the baseline design at $n = 500$, Joint 74 and One 9 with lag-dependent links at $n = 500$, and Joint 1 and One 1 with lag-dependent links at $n = 1000$.

Table 19: Monte Carlo results for β: continuous design

n	Naive	HK	Proposed	Oracle	Class	Infeas.
Bias						
500	0.973	0.437	0.431	0.392	0.226	0.011
1000	0.960	0.325	0.111	0.105	0.078	−0.003
2000	0.957	0.299	0.032	0.032	0.049	−0.007
3000	0.961	0.321	0.057	0.056	0.087	0.001
Standard deviation						
500	0.153	0.841	1.758	1.610	1.086	0.178
1000	0.104	0.529	0.761	0.772	0.576	0.125
2000	0.073	0.367	0.473	0.469	0.364	0.085
3000	0.055	0.316	0.375	0.378	0.319	0.065
Root mean squared error						
500	0.985	0.947	1.808	1.655	1.108	0.178
1000	0.966	0.620	0.768	0.778	0.581	0.125
2000	0.959	0.473	0.474	0.470	0.367	0.085
3000	0.963	0.450	0.378	0.381	0.330	0.064
Coverage of nominal 95% interval						
500	0.000	0.922	0.960	0.947	0.966	0.944
1000	0.000	0.922	0.964	0.958	0.954	0.948
2000	0.000	0.892	0.968	0.964	0.964	0.958
3000	0.000	0.818	0.958	0.956	0.944	0.966
Replications without an estimate						
500	0	0	5	6	4	0
1000	0	0	0	0	0	0
2000	0	0	0	0	0	0
3000	0	0	0	0	0	0

Notes: $R = 500$ replications. Steep beta-model links $f(u, v) = F(8(u + v - 1))$. The controls estimator is omitted. Bandwidths: $h_1 = c\,(n/500)^{-1/5}$, with $c = 5.13 \times 10^{-4}$ the population 0.10 quantile of the squared codegree distance, computed by numerical integration, and $h_2 = n^{-1/5}/5$. Oracle: the same pairwise estimator with the population codegree distances $\delta_1(w_{i1}, w_{j1})^2$, computed by quadrature, so free of link noise and of the sampling variation of the other agents' latent indices. Class: the class-conditional estimator, with the connected components of the thresholded graph as classes (not classes of equal movement here, Section 5.3) and the covariate-stability window as an indicator. With no distribution theory here, coverage is only a diagnostic. Failed fits at $n = 500$ are optimizer failures, all certified as separated except one of the oracle's: Proposed 5, Oracle 6, and Class 4.

Table 20: Monte Carlo results for α: continuous design

n	Naive	HK	Proposed	Oracle	Class	Infeas.
Bias						
500	0.667	0.157	0.145	0.148	0.067	−0.008
1000	0.670	0.142	0.052	0.052	0.025	0.002
2000	0.663	0.105	0.013	0.013	0.003	−0.006
3000	0.665	0.094	0.007	0.008	0.013	−0.003
Standard deviation						
500	0.119	0.472	1.168	1.123	0.793	0.130
1000	0.090	0.333	0.617	0.607	0.473	0.094
2000	0.063	0.242	0.396	0.401	0.315	0.066
3000	0.053	0.198	0.315	0.313	0.249	0.054
Root mean squared error						
500	0.678	0.497	1.176	1.131	0.795	0.131
1000	0.676	0.362	0.619	0.609	0.473	0.094
2000	0.666	0.264	0.396	0.401	0.315	0.066
3000	0.667	0.219	0.315	0.313	0.249	0.054
Coverage of nominal 95% interval						
500	0.000	0.960	0.964	0.964	0.966	0.974
1000	0.000	0.940	0.978	0.966	0.956	0.960
2000	0.000	0.938	0.954	0.944	0.950	0.954
3000	0.000	0.942	0.952	0.956	0.952	0.968
Replications without an estimate						
500	0	0	5	6	4	0
1000	0	0	0	0	0	0
2000	0	0	0	0	0	0
3000	0	0	0	0	0	0

Notes: $R = 500$ replications. Steep beta-model links $f(u, v) = F(8(u + v - 1))$. The controls estimator is omitted. Bandwidths: $h_1 = c\,(n/500)^{-1/5}$, with $c = 5.13 \times 10^{-4}$ the population 0.10 quantile of the squared codegree distance, computed by numerical integration, and $h_2 = n^{-1/5}/5$. Oracle: the same pairwise estimator with the population codegree distances $\delta_1(w_{i1}, w_{j1})^2$, computed by quadrature, so free of link noise and of the sampling variation of the other agents' latent indices. Class: the class-conditional estimator, with the connected components of the thresholded graph as classes (not classes of equal movement here, Section 5.3) and the covariate-stability window as an indicator. With no distribution theory here, coverage is only a diagnostic. Failed fits at $n = 500$ are optimizer failures, all certified as separated except one of the oracle's: Proposed 5, Oracle 6, and Class 4.

References

Albert, A. and Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, 71(1):1–10.

Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society: Series B*, 32(2):283–301.

Arduini, T., Patacchini, E., and Rainone, E. (2015). Parametric and semiparametric IV estimation of network models with selectivity. EIEF Working Paper No. 09/15.

Auerbach, E. (2021). Identification and estimation of a partially linear regression model using network data: Inference and an application to network peer effects. arXiv preprint arXiv:2105.10002.

Auerbach, E. (2022). Identification and estimation of a partially linear regression model using network data. *Econometrica*, 90(1):347–365.

Badev, A. (2021). Nash equilibria on (un)stable networks. *Econometrica*, 89(3):1179–1206.

Beugnot, J., Fortin, B., Lacroix, G., and Villeval, M. C. (2013). Social networks and peer effects at work. IZA Discussion Paper No. 7521.

Boistard, H., Lopuhaä, H. P., and Ruiz-Gazen, A. (2012). Approximation of rejective sampling inclusion probabilities and application to high order correlations. *Electronic Journal of Statistics*, 6:1967–1983.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184.

Bonhomme, S., Lamadon, T., and Manresa, E. (2022). Discretizing unobserved heterogeneity. *Econometrica*, 90(2):625–643.

Bramoullé, Y., Djebbari, H., and Fortin, B. (2020). Peer effects in networks: A survey. *Annual Review of Economics*, 12:603–629.

Breslow, N. E. and Day, N. E. (1980). *Statistical Methods in Cancer Research, Volume I: The Analysis of Case-Control Studies*. IARC Scientific Publications No. 32. International Agency for Research on Cancer, Lyon.

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

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *The Review of Economic Studies*, 47(1):225–238.

Chamberlain, G. (2010). Binary response models for panel data: Identification and information. *Econometrica*, 78(1):159–168.

Christakis, N. A. and Fowler, J. H. (2008). The collective dynamics of smoking in a large social network. *New England Journal of Medicine*, 358(21):2249–2258.

Conley, T. G. and Udry, C. R. (2010). Learning about a new technology: Pineapple in Ghana. *American Economic Review*, 100(1):35–69.

Dano, K., Honoré, B. E., and Weidner, M. (2025). Binary choice logit models with general fixed effects for panel and network data. arXiv preprint arXiv:2508.11556.

Darroch, J. N. (1964). On the distribution of the number of successes in independent trials. *The Annals of Mathematical Statistics*, 35(3):1317–1321.

De Giorgi, G., Frederiksen, A., and Pistaferri, L. (2020). Consumption network effects. *The Review of Economic Studies*, 87(1):130–163.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1):27–38.

Fletcher, J. M. (2010). Social interactions and smoking: Evidence using multiple student cohorts, instrumental variables, and school fixed effects. *Health Economics*, 19(4):466–484.

Gail, M. H., Lubin, J. H., and Rubinstein, L. V. (1981). Likelihood calculations for matched case-control studies and survival studies with tied death times. *Biometrika*, 68(3):703–707.

Gao, W. Y. and Niu, Y. (2026). Identification in dynamic dyadic network formation models with fixed effects. arXiv preprint arXiv:2604.07488.

Goldsmith-Pinkham, P. and Imbens, G. W. (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics*, 31(3):253–264.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica*, 85(4):1033–1063.

Gueyap Kounga, B. R. (2023). Identification and estimation of a semiparametric logit model using network data. arXiv preprint arXiv:2310.07151. First version October 2023, current version August 2026.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *The Annals of Mathematical Statistics*, 35(4):1491–1523.

Heckman, J. J. (1981a). Heterogeneity and state dependence. In S. Rosen (ed.), *Studies in Labor Markets*, pp. 91–140. University of Chicago Press.

Heckman, J. J. (1981b). The incidental parameters problem and the problem of initial conditions in estimating a discrete time–discrete data stochastic process. In C. F. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, pp. 179–195. MIT Press, Cambridge, MA.

Heinze, G. and Puhr, R. (2010). Bias-reduced and separation-proof conditional logistic regression with small or sparse data sets. *Statistics in Medicine*, 29(7–8):770–777.

Henderson, D. J., Carroll, R. J., and Li, Q. (2008). Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics*, 144(1):257–275.

Hjort, N. L. and Pollard, D. (1993). Asymptotics for minimisers of convex processes. Statistical Research Report, Department of Mathematics, University of Oslo. arXiv preprint arXiv:1107.3806.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30.

Honoré, B. E. and Kyriazidou, E. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68(4):839–874.

Honoré, B. E. and Powell, J. L. (2005). Pairwise difference estimators for nonlinear models. In D. W. K. Andrews and J. H. Stock (eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pp. 520–553. Cambridge University Press.

Honoré, B. E. and Tamer, E. (2006). Bounds on parameters in panel dynamic discrete choice models. *Econometrica*, 74(3):611–629.

Honoré, B. E. and Weidner, M. (2025). Moment conditions for dynamic panel logit models with fixed effects. *The Review of Economic Studies*, 92(5):3112–3137.

Honoré, B. E., Muris, C., and Weidner, M. (2025). Dynamic ordered panel logit models. *Quantitative Economics*, 16(3):899–945.

Hsieh, C.-S. and Lee, L. F. (2016). A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics*, 31(2):301–319.

Jackson, M. O., Rogers, B. W., and Zenou, Y. (2017). The economic consequences of social-network structure. *Journal of Economic Literature*, 55(1):49–95.

Johnsson, I. and Moon, H. R. (2021). Estimation of peer effects in endogenous social networks: Control function approach. *The Review of Economics and Statistics*, 103(2):328–345.

Kashaev, N. and Lazzati, N. (2025). Discrete choice with endogenous peer selection. arXiv preprint arXiv:2511.21446.

Magnac, T. (2004). Panel binary variables and sufficiency: Generalizing conditional logit. *Econometrica*, 72(6):1859–1876.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542.

McVittie, J. (2008). *National Qualifications: A Short History*. Research Report 3, Scottish Qualifications Authority.

Michell, L. and West, P. (1996). Peer pressure to smoke: The meaning depends on the method. *Health Education Research*, 11(1):39–49.

Muris, C., Raposo, P., and Vandoros, S. (2025). A dynamic ordered logit model with fixed effects. *The Review of Economics and Statistics*, 107(4):1104–1114.

Nakajima, R. (2007). Measuring peer effects on youth smoking behaviour. *The Review of Economic Studies*, 74(3):897–935.

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

Ouyang, F. and Yang, T. T. (2025). Semiparametric estimation of dynamic binary choice panel data models. *Econometric Theory*, 41(4):907–946.

Patacchini, E. and Zenou, Y. (2012). Juvenile delinquency and conformism. *The Journal of Law, Economics, and Organization*, 28(1):1–31.

Pearson, M. and Michell, L. (2000). Smoke Rings: Social network analysis of friendship groups, smoking and drug-taking. *Drugs: Education, Prevention and Policy*, 7(1):21–37.

Qian, J. and Wang, L. (2012). Estimating semiparametric panel data models by marginal integration. *Journal of Econometrics*, 167(2):483–493.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica*, 56(4): 931–954.

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *The Quarterly Journal of Economics*, 116(2):681–704.

Schaefer, D. R., Haas, S. A., and Bishop, N. J. (2012). A dynamic model of US adolescents' smoking and friendship networks. *American Journal of Public Health*, 102(6):e12–e18.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York.

Shin, S. (2025). Sufficient statistics for Markovian feedback processes and unobserved heterogeneity in dynamic panel logit models. arXiv preprint arXiv:2511.02816.

Starck, V. (2025). Improving control over unobservables with network data. arXiv preprint arXiv:2511.00612.

Steglich, C., Snijders, T. A. B., and West, P. (2006). Applying SIENA: An illustrative analysis of the coevolution of adolescents' friendship networks, taste in music, and alcohol consumption. *Methodology*, 2(1):48–56.

Steglich, C., Snijders, T. A. B., and Pearson, M. (2010). Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology*, 40(1):329–393.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

Xu, G. (2025). Networks as control functions: Nonparametric identification and estimation of partial effects. Job market paper, University of Washington.

Zeleneev, A. (2020). Identification and estimation of network models with nonparametric unobserved heterogeneity. Working paper, University College London. arXiv preprint arXiv:2602.06885.