Estimating diffusion and adoption parameters in networks

New estimation approaches for the latent-diffusion-observed-adoption model

Stephan, L.S.

Publication date
2021

Chapter 5

Summary

In recent years, economists have started challenging the idea that all economic actors interact with one another in an anonymous way in favour of the more realistic view that socio-economic networks matter for many economic interactions. Various models of network formation and interaction have been put forward. Yet fitting these models to data entails challenges not previously encountered by econometricians. Individual observations are interdependent, the network models are highly non-linear and there is a lack of pre-designed estimation software. In this thesis, titled "Estimating diffusion and adoption parameters in Networks – new estimation approaches for the latent-diffusion-observed-adoption model", I propose three estimation methods for a tractable, yet widely applicable model of network interaction: the "latent-diffusion-observed-adoption model". In this model, a signal propagates through the network exposing the agents to a choice whenever the signal newly reaches them. The network, the signal initiation and affirmative choices are observable, yet the signal spreading remains hidden to the researcher, whose aim it is to estimate the rate of diffusion (i.e. the probability of the signal spreading from one agent to the next) and the rate of adoption (i.e. the probability to affirm upon signal reception). The challenge stems from the unobservability of the diffusion process, which entails a curse of dimensionality: an absence of an affirmative choice could be a deliberate action or a simple lack of alternative. As a consequence, the number of possibilities of signal spread that are in accordance with the observed data grows exponentially in the number of agents involved. In each chapter of the thesis, a distinct strategy to tackle this problem is proposed. Given the importance of network models and the need to develop methods to estimate these models, this thesis makes a contribution that is both novel and important.

Throughout the thesis, I employ real-world sampled data on social networks, made

publicly available at *https://dataverse.harvard.edu* and *https://web.stanford.edu/ jacksonm/Data.html*. This data was gathered in a joint venture between researchers from MIT and Stanford and a microfinance organization operating in rural India. Here, the "signal" is the information on the existence of the microfinance project and the "adoption" is the decision to become a micro-borrower. The workhorse model of this thesis, the "latent-diffusion-observed-adoption" model, was also developed by these researchers. The subsequent paper Banerjee et al. (2013) received tremendous attention from the scientific community. Chapter two works exclusively with the adoption data as sampled, while in chapter three and four, simulation techniques are used as well. Simulating the adoption on the sampled network matrices has the advantage that the true parameters of the data-generating process are known and the performance of the estimators can better be evaluated, which is in particular important for the estimators in chapter three and four, the properties of which are less well-known from standard econometric theory.

Chapter two restricts the modelled time horizon to two periods. With only one round of signal exchange, the possibilities of signal propagation are limited, rendering exact Maximum Likelihood (ML) estimation feasible. This enables me to estimate a series of nested models and compare them using the Likelihood ratio test. The baseline model imposes that both the diffusion and the adoption rate are homogeneous across agents. The first extension includes the number of friends that made an affirmative choice in the first period as a determinant for an individual's own chance to do so in the second period. I find that having friends who adopted previously significantly increases the individual's own adoption probability. The second extension includes relevant household characteristics as determinants of the individual adoption rates. The two determinants included are a dummy variable indicating whether the household is Hindu as well as the number of rooms in the house (used as an indicator for household wealth). Both characteristics have a significant negative impact, implying that richer households and Hindus are less likely to adopt. Both extensions are estimated first with homogeneous diffusion rates and a second time allowing diffusion rates to depend on the sender's participation status. All estimations indicate that participants are significantly more likely to transmit the signal, regardless of whether adoption rates are modelled as being homogeneous (baseline model), as being dependent on the number of friends that previously adopted (first extension) or as varying with household religion and wealth (second extension). These findings have important policy implications. First, they enable decision-makers to distinguish whether low diffusion or low adoption rates are the driver behind sub-optimally low take-up rates and here, the latter is the case. Second, across all specifications, participants are more likely to pass on the signal than nonparticipants. This finding, combined with the significant and positive “endorsement” effect (i.e. past adopters nudging their friends into the same choice) implies that increasing early adoption has a twofold benefit by increasing both diffusion and adoption rates. Finally, the fact that poorer households are more likely to adopt is encouraging as this is a desired feature of the project.

The third chapter of the thesis makes use of the fact that the probability to observe the data at hand is established as a sum over the respective probabilities of a particular signal-spreading scenario. This leads to the insight that it is possible to establish and maximise the probability to observe not all, but only a subset of all possible signal propagation-scenarios, while neglecting the remaining ones. A Monte Carlo study reveals that the estimator that results from maximizing the approximate Log-likelihood function converges to the maximum likelihood estimator (MLE) that uses all signal propagation scenarios when approximately two third of the scenarios are taken into account. Despite the fact that it is not feasible to be in this range for the real villages, an estimation using four real villages indicates a steady convergence towards the MLE. I further analyse theoretically how certain network characteristics impact this “trimmed” estimator.

While chapter two and three are based on maximum likelihood estimation, chapter four presents a moment-based estimation for the model at hand. Moment-based estimation has the advantage that moments that are computationally hard to evaluate can be neglected under the condition that the remaining moments are sufficient to identify the parameters. The key insight for the identification strategy is that an individual’s probability to receive the information at her first possible opportunity to do so can relatively easily be established using some shorthand formulas. Two estimators that use these individual “first-opportunity moments” are proposed and their properties are investigated both theoretically and in a simulation exercise. With individual-specific moments being correlated within the networks, the estimator resulting from minimizing squared deviations of individual outcomes from their unconditional expectations outperforms an estimator that first aggregates the individual deviations before squaring them. The former estimator also shows a clear superiority as compared to the two-period maximum likelihood estimator from chapter two, which is intuitive as additional information (the outcomes in later time periods) is taken into account. Both moment-based estimators are employed for the real data.

This thesis proposes multiple ways to estimate parameters in the “latent-diffusion-observed-adoption model”. The model has been established in Banerjee et al. (2013) for purpose of investigating the spread of information about and participation in a microfinance project in rural India, but it lends itself to modelling various other economic phenomena. As a consequence, an investigation into the best possibilities to estimate the model parameters is highly useful and important. Furthermore, the estimators that are proposed (and the software I wrote) can be adjusted to different assumptions regarding the spreading and adoption processes. Therefore, I hope to make a novel and important contribution relevant for the econometrics of network interaction.