



UvA-DARE (Digital Academic Repository)

Estimating diffusion and adoption parameters in networks

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

Stephan, L.S.

Publication date

2021

[Link to publication](#)

Citation for published version (APA):

Stephan, L. S. (2021). *Estimating diffusion and adoption parameters in networks: New estimation approaches for the latent-diffusion-observed-adoption model*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Bibliography

- Andrews, D. W. (1994). Empirical process methods in econometrics. In R. F. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 37, pp. 2247 – 2294. Elsevier.
- Anselin, L. (2001). Spatial econometrics. *A companion to theoretical econometrics 310330*, 111–222.
- Anselin, L. and A. K. Bera (1998). Introduction to spatial econometrics. *Handbook of applied economic statistics 237*, 111–222.
- Bandiera, O. and I. Rasul (2006). Social Networks and Technology Adoption in Northern Mozambique. *The Economic Journal 116 (514)*, 869–902.
- Banerjee, A., A. Chandrasekhar, E. Duflo, and M. Jackson (2013, Science 26 July). The diffusion of microfinance. Vol. 341 (Issue 6144).
- Bhagat, S., A. Goyal, and L. Lakshmanan (2012). Maximizing Product Adoption in Social Networks. *Proceedings of the fifth ACM international conference on Web search and data mining 5*, pp.603–612.
- Bickel, P. J., A. Chen, and E. Levina (2011, 10). The method of moments and degree distributions for network models. *Ann. Statist. 39(5)*, 2280–2301.
- Centola, D. (2015). The Social Origins of Networks and Diffusion. *American Journal of Sociology 120(5)*, 1295–1338.
- Christakis, N. A. and J. H. Fowler (2007). The Spread of Obesity in a Large Social Network over 32 Years. *The New England Journal of Medicine 357(4)*, pp.370–379.
- Coleman, J., E. Katz, and H. Menzel (1966). *Medical Innovation: A Diffusion Study*. Bobbs-Merrill Co.
- Conley, T. G. and C. R. Udry (2010). Learning about a New Technology: Pineapple in Ghana. *The American Economic Review 100(1)*, 35–69.

- Cowan, R. (2005). Network models of innovation and knowledge diffusion. In *Clusters, Networks and Innovation*, OUP E-Books, pp. 29–53. Oxford University Press New York.
- Craig, B. R., T. Phelan, J.-P. Siedlarek, and J. Steinberg (2020). Improving epidemic modeling with networks. *Economic Commentary 2020-23*, 111–222.
- de Paula, A. (2020). Econometric models of network formation. *Annual Review of Economics 12*(1), 775–799.
- Durlauf, S. N. and Y. M. Ioannides (2010). Social interactions. *Annual Review of Economics 2*(1), 451–478.
- Hansen, B. E. and S. Lee (2019). Asymptotic theory for clustered samples. *Journal of econometrics 210*(2), 268–290.
- Hansen, L. P. (2001). Generalized method of moments estimation: A time series perspective. In N. J. Smelser and P. B. Baltes (Eds.), *International Encyclopedia of Social and Behavioral Sciences*, pp. 9743–9751. Elsevier.
- Hethcote, H. W. (2000). The Mathematics of Infectious Diseases. *Society for Industrial and Applied Mathematics (SIAM) Review 42*(4), pp.599—653.
- Jackson, M., T. Rodriguez-Barraquer, and X. Tan (August 2012). Social capital and social quilts: Network patterns of favor exchanges. *American Economic Review Vol. 102, No 5*, pp.1857–97.
- Jackson, M. O. (2004). A survey of models of network formation: Stability and efficiency. In G. Demange and M. Wooders (Eds.), *Group Formation in Economics; Networks, Clubs and Coalitions*, Chapter 1, pp. 11–57. Cambridge University Press.
- Jackson, M. O. and L. Yariv (2005). Diffusion on social networks. In *Economie Publique*. Citeseer.
- Karaivanov, A. (2020). A social network model of covid-19.
- Katungi, E., S. Edmeades, and M. Smale (2008). Gender, Social Capital and Information Exchange in Rural Uganda. *Journal of International Development 20*, 35–52.
- Krisztin, T., P. Piribauer, and M. Woegerer (2020). The spatial econometrics of the coronavirus pandemic. *Letters in Spatial and Resource Sciences 13*(2), (accessible online).

- Liu, X. and L. fei Lee (2010). Gmm estimation of social interaction models with centrality. *Journal of Econometrics* 159(1), 99 – 115.
- Newey, W. K. and D. McFadden (1994). Chapter 36 large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 36, pp. 2111 – 2245. Elsevier.
- Neyman, J. and E. Pearson (1933). IX. On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. R. Soc. Lond.* 231(694-706), 289—337.
- Ryan, B. and N. Gross (1950). Acceptance and diffusion of hybrid corn seed in two Iowa communities. *Research Bulletin (Iowa Agriculture and Home Economics Experiment Station) Vol. 29 : No. 372*, Article 1.
- Singh, S. S., K. Singh, A. Kumar, H. K. Shakya, and B. Biswas (2019). A survey on information diffusion models in social networks. In A. K. Luhach, D. Singh, P.-A. Hsiung, K. B. G. Hawari, P. Lingras, and P. K. Singh (Eds.), *Advanced Informatics for Computing Research*, Singapore, pp. 426–439. Springer Singapore.
- Snijders, T. A. (2017). Stochastic actor-oriented models for network dynamics. *Annual Review of Statistics and Its Application* 4(1), 343–363.
- Su, L. and Z. Yang (2011). Instrumental variable quantile estimation of spatial autoregressive models. *Research Collection School Of Economics. Singapore Management University* https://ink.library.smu.edu.sg/soe_research/1074, 1–35.
- Todoa, Y., D. M. Yadateb, P. Matousc, and R. Takahashid (2011). Effects of Geography and Social Networks on Diffusion and Adoption of Agricultural Technology: Evidence from Rural Ethiopia. *Center for the Study of African Economics (CSAE) 25th Anniversary Conference*, 111–222.
- Valente, T. W. (1996). Social network thresholds in the diffusion of innovations. *Social networks* 18(1), 69–89.
- Valente, T. W. (2005). Network models and methods for studying the diffusion of innovations. *Models and methods in social network analysis* 28, 98.
- Wooldridge, J. (1994). Estimation and inference for dependent processes. In R. F. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 45, pp. 2639 – 2738. Elsevier.

Samenvatting

In de afgelopen jaren zijn economen begonnen met het ter discussie stellen van het idee dat alle interacties tussen economische actoren anoniem plaatsvinden, ten gunste van de meer realistische opvatting dat sociaal-economische netwerken van belang zijn voor veel economische interacties. Er zijn verschillende modellen van netwerkvorming en interactie voorgesteld. Echter, het fitten van deze modellen aan data brengt nieuwe uitdagingen voor econometristen met zich mee. Individuele waarnemingen zijn niet onafhankelijk, de netwerkmodellen in hoge mate niet-lineair en er is een gebrek aan vooraf ontworpen schattingssoftware. In dit proefschrift, getiteld “Estimating diffusion and adoption parameters in Networks – new estimation approaches for the latent-diffusion observed-adoption model” stel ik drie schattingsmethoden voor, voor een handelbaar, maar breed toepasbaar model van netwerkkinteractie: het latente-diffusie-waargenomen-adoptiemodel. In dit model verspreidt een signaal zich door het netwerk, waarbij het personen voor een keuze stelt wanneer het hen voor het eerst bereikt. Het netwerk, de signaalinitiatie en bevestigende keuzes zijn waarneembaar, maar de signaalverspreiding blijft verborgen voor de onderzoeker, wiens doel het is om de verspreidingsgraad (de kans dat het signaal zich van de ene persoon naar de andere verspreidt) en de adoptiegraad (de kans om te bevestigen bij signaalontvangst) te schatten. De uitdaging komt voort uit de onzichtbaarheid van het diffusieproces, wat een probleem van dimensionaliteit met zich meebrengt: de afwezigheid van een bevestigende keuze kan een opzettelijke actie zijn of een eenvoudig gebrek aan een alternatief. Als gevolg daarvan groeit het aantal mogelijkheden van signaalverspreiding dat in overeenstemming is met de waarnemingen exponentieel in het aantal betrokken personen. In elk hoofdstuk van dit proefschrift pak ik dit probleem aan met een andere strategie. Gezien het belang van netwerkmodellen en de noodzaak om methoden te ontwikkelen om deze modellen te schatten, levert dit proefschrift een bijdrage die zowel nieuw als belangrijk is. Op verschillende plekken in dit proefschrift maak ik gebruik van steekproeven van echte data over sociale netwerken, die openbaar beschikbaar zijn op <https://dataverse.harvard.edu> en <https://web.stanford.edu/jacksonm/Data.html>. Deze gegevens zijn verzameld in een

joint venture tussen onderzoekers van MIT en Stanford en een microfinancieringsorganisatie die actief is op het Indiase platteland. Het “signaal” in deze context, is de informatie over het bestaan van het microfinancieringsproject en de “adoptie” is de beslissing om microkredietnemer te worden. Het werkmodel van dit proefschrift, het “latente-diffusie-waargenomen-adoptie” model, is ook ontwikkeld door deze onderzoekers. Het daaropvolgende artikel Banerjee et al. (2013) kreeg heel veel aandacht van de wetenschappelijke gemeenschap. Hoofdstuk twee gebruikt alleen de data over adoptie, zoals verkregen uit de steekproeven. In de hoofdstukken drie en vier maak ik ook gebruik van simulatietechnieken. Het simuleren van adoptie op bemonsterde netwerkmatrices heeft als voordeel dat de echte parameters van het gegevensgeneratieproces bekend zijn en dat de prestaties van de schatters beter kunnen worden geëvalueerd. Dat is vooral belangrijk voor de schatters in de hoofdstukken drie en vier, omdat de eigenschappen daarvan minder bekend zijn uit de standaard econometrische theorie. Hoofdstuk twee beperkt de gemodelleerde tijdshorizon tot twee perioden. Met slechts één ronde signaaluitwisseling zijn de mogelijkheden van signaalverspreiding beperkt, waardoor een exacte Maximale Waarschijnlijkheid (ML) schatting mogelijk is. Hierdoor kan ik een reeks geneste modellen schatten en vergelijken met de waarschijnlijkheidsratio test. Het basismodel gaat ervan uit dat zowel de diffusie als de adoptiegraad homogeen zijn tussen personen. De eerste extensie voegt het aantal vrienden dat een bevestigende keuze heeft gemaakt in de eerste periode als een bepalende factor toe voor iemands eigen kans om dit te doen in de tweede periode. Als resultaat vind ik dat het hebben van vrienden die eerder de mogelijkheid tot het krijgen van microkrediet hebben aangegrepen (adoptie), de eigen kans op adoptie aanzienlijk vergroot. De tweede extensie voegt relevante kenmerken van het huishouden toe als determinanten van de individuele adoptiegraad. De twee opgenomen determinanten zijn een dummy variabele die aangeeft of het huishouden hindoe is en het aantal kamers van het huis (gebruikt als indicator voor het vermogen van het huishouden). Beide kenmerken hebben een aanzienlijk negatief effect, wat inhoudt dat rijkere huishoudens en hindoes minder geneigd zijn om te adopteren. Beide extensies zijn eerst geschat met homogene diffusiesnelheden en een tweede keer met diffusiesnelheden die mogen afhangen van de deelnamestatus van de afzender van het signaal. Alle schattingen geven aan dat deelnemers een significant grotere kans hebben om het signaal door te geven, ongeacht of de adoptiegraad wordt gemodelleerd als homogeen (basismodel), als afhankelijk van het aantal vrienden dat eerder adopteerde (eerste extensie) of als variërend met religie en rijkdom (tweede extensie). Deze bevindingen hebben belangrijke beleidsimplicaties. Ten eerste stelt het me in staat om te onderscheiden of een lage diffusie- of een lage adoptiegraad de drijvende kracht is achter suboptimaal lage benut-

tingsgraden, en hier is het laatste het geval. Ten tweede, onder alle omstandigheden geven deelnemers het signaal eerder door dan niet-deelnemers. Deze bevinding, gecombineerd met het significante en positieve 'goedkeuring'-effect (eerdere adoptanten die hun vrienden stimuleren dezelfde keuze te maken) impliceert dat het doen toenemen van vroege adoptie op twee manieren nuttig is. Het verhoogt zowel de verspreidings- als de adoptiegraad. Ten slotte is het feit dat armere huishoudens eerder geneigd zijn om te adopteren bemoedigend, aangezien dit een gewenst kenmerk van het project is. Het derde hoofdstuk van het proefschrift maakt gebruik van het feit dat de kans om juist deze data te observeren wordt gevormd door de som van de respectievelijke kansen van een bepaald scenario voor signaalverspreiding. Dit leidt tot het inzicht dat het mogelijk is om de kans vast te stellen en te maximaliseren om niet alle, maar slechts een deel van alle mogelijke signaalverspreidingen te observeren. De overige worden dan verwaarloosd. Een Monte Carlo-studie laat zien dat de schatter die het resultaat is van het maximaliseren van de benadering van de log-waarschijnlijkheids-functie convergeert naar de maximale waarschijnlijkheidsschatter (MLE) die alle signaalverspreidings-scenario's gebruikt wanneer ongeveer twee derde van de scenario's in aanmerking worden genomen. Ondanks het feit dat het niet haalbaar is om binnen dit bereik te zijn voor echte dorpen, duidt een schatting op basis van vier echte dorpen op een gestage convergentie naar de MLE. Ik analyseer verder theoretisch hoe bepaalde netwerkkenmerken van invloed zijn op deze "getrimde" schatter.

Terwijl hoofdstuk twee en drie gebaseerd zijn op schatting van de maximale waarschijnlijkheid, presenteert hoofdstuk vier een moment-gebaseerde schatting voor het gebruikte model. Moment-gebaseerde schatting heeft het voordeel dat momenten die rekenkundig moeilijk te evalueren zijn, kunnen worden verwaarloosd in het geval dat de resterende momenten voldoende zijn om de parameters te identificeren. Het belangrijkste inzicht voor de identificatiestrategie is dat de kans dat een individu informatie ontvangt bij de eerste mogelijke kans deze te ontvangen, relatief eenvoudig kan worden bepaald met behulp van enkele korte formules. Ik stel twee schatters voor die gebruik maken van deze individuele 'eerste-kans-momenten' en onderzoek hun eigenschappen zowel theoretisch als in een simulatieoefening. Met individueel-specifieke momenten die gecorreleerd zijn binnen de netwerken, presteert de schatter die het resultaat is van het minimaliseren van gekwadrateerde afwijkingen van individuele uitkomsten ten opzichte van hun onvoorwaardelijke verwachtingen, beter dan een schatter die eerst de individuele afwijkingen aggregereert alvorens ze te kwadrateren. De eerste schatter laat ook een duidelijke superioriteit zien ten opzichte van de twee-perioden maximale waarschijnlijkheidsschatter uit hoofdstuk twee, wat intuïtief is omdat er met aanvullende informatie (de uitkomsten in latere perioden) rekening wordt gehouden. Beide

moment-gebaseerde schatters worden gebruikt op de echte data.

In dit proefschrift stel ik meerdere manieren voor om parameters in het “latente-diffusie-waargenomen-adoptiemodel” te schatten. Het model is opgesteld in Banerjee et al. (2013) om de verspreiding van informatie over en deelname aan een microfinancieringsproject op het platteland van India te onderzoeken, maar het leent zich ook voor het modelleren van diverse andere economische verschijnselen. Bijgevolg is een onderzoek naar de beste manieren om de modelparameters te schatten zeer nuttig en belangrijk. Bovendien kunnen de schatters die ik voorstel (en de software die ik heb geschreven) worden aangepast aan verschillende aannames met betrekking tot de verspreidings- en adoptieprocessen. Daarom hoop ik een nieuwe en belangrijke bijdrage te leveren die relevant is voor de econometrie van netwerkinteractie.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

- 733. Z.M. LI, *Econometric Analysis of High-frequency Market Microstructure*
- 734. C.M. OOSTERVEEN, *Education Design Matters*
- 735. S.C. BARENDSE, *In and Outside the Tails: Making and Evaluating Forecasts*
- 736. S. SÓVÁGÓ, *Where to Go Next? Essays on the Economics of School Choice*
- 737. M. HENNEQUIN, *Expectations and Bubbles in Asset Market Experiments*
- 738. M.W. ADLER, *The Economics of Roads: Congestion, Public Transit and Accident Management*
- 739. R.J. DÖTTLING *Essays in Financial Economics*
- 740. E.S. ZWIERS *About Family and Fate: Childhood Circumstances and Human Capital Formation*
- 741. Y.M. KUTLUAY *The Value of (Avoiding) Malaria*
- 742. A. BOROWSKA *Methods for Accurate and Efficient Bayesian Analysis of Time Series*
- 743. B. HU *The Amazon Business Model, the Platform Economy and Executive Compensation: Three Essays in Search Theory*
- 744. R.C. SPERNA WEILAND *Essays on Macro-Financial Risks*
- 745. P.M. GOLEC *Essays in Financial Economics*
- 746. M.N. SOUVERIJN *Incentives at work*
- 747. M.H. COVENEY *Modern Imperatives: Essays on Education and Health Policy*
- 748. P. VAN BRUGGEN *On Measuring Preferences*
- 749. M.H.C. NIENKER *On the Stability of Stochastic Dynamic Systems and their use in Econometrics*
- 750. S. GARCIA MANDICÓ *Social Insurance, Labor Supply and Intra-Household Spillovers*
- 751. Y. SUN *Consumer Search and Quality*
- 752. I. KERKEMEZOS *On the Dynamics of (Anti) Competitive Behaviour in the Airline Industry*
- 753. G.W. GOY *Modern Challenges to Monetary Policy*
- 754. A.C. VAN VLODROP *Essays on Modeling Time-Varying Parameters*
- 755. J. SUN *Tell Me How To Vote, Understanding the Role of Media in Modern Elections*

756. J.H. THIEL *Competition, Dynamic Pricing and Advice in Frictional Markets: Theory and Evidence from the Dutch Market for Mortgages*
757. A. NEGRIU *On the Economics of Institutions and Technology: a Computational Approach*
758. F. GRESNIGT *Identifying and Predicting Financial Earth Quakes using Hawkes Processes*
759. A. EMIRMAHMUTOGLU *Misperceptions of Uncertainty and Their Applications to Prevention*
760. A. RUSU *Essays in Public Economics*
761. M.A. COTOFAN *Essays in Applied Microeconomics: Non-Monetary Incentives, Skill Formation, and Work Preferences*
762. B.P.J. ANDRÉE *Theory and Application of Dynamic Spatial Time Series Models*
763. P. PELZL *Macro Questions, Micro Data: The Effects of External Shocks on Firms*
764. D.M. KUNST *Essays on Technological Change, Skill Premia and Development*
765. A.J. HUMMEL *Tax Policy in Imperfect Labor Markets*
766. T. KLEIN *Essays in Competition Economics*
767. M. VIGH *Climbing the Socioeconomic Ladder: Essays on Sanitation and Schooling*
768. YAN XU *Eliciting Preferences and Private Information: Tell Me What You Like and What You Think*
769. S. RELLSTAB *Balancing Paid Work and Unpaid Care over the Life-Cycle*
770. Z. DENG *Empirical Studies in Health and Development Economics*
771. L. KONG *Identification Robust Testing in Linear Factor Models*
772. I. NEAMȚU *Unintended Consequences of Post-Crisis Banking Reforms*
773. B. KLEIN TEESELINK *From Mice to Men: Field Studies in Behavioral Economics*
774. B. TEREICK *Making Crowds Wiser: The Role of Incentives, Individual Biases, and Improved Aggregation*
775. A. CASTELEIN *Models for Individual Responses*
776. D. KOLESNYK *Consumer Disclosures on Social Media Platforms: A Global Investigation*
777. M.A. ROLA-JANICKA *Essays on Financial Instability and Political Economy of Regulation*
778. J.J. KLINGEN *Natural Experiments in Environmental and Transport Economics*
779. E.M. ARTMANN *Educational Choices and Family Outcomes*
780. F.J. OSTERMEIJER *Economic Analyses of Cars in the City*
781. T. ÖZDEN *Adaptive Learning and Monetary Policy in DSGE Models*
782. D. WANG *Empirical Studies in Financial Stability and Natural Capital*