Sida Peng

www.sidapeng.com

Contact Information

Department of Economics
429 Uris Hall, Ithaca, NY 14853, Cornell University

+1 (434) 249-5434
sp947@cornell.edu

Concentration

Econometrics, Empirical Industrial Organization, Applied Microeconomic Theory

Education

Ph.D. in Economics, Cornell University	Expected 2017
M.S. in Statistics, University of Virginia	2012
B.S. in Economics, University of Virginia	2011
B.S. in Mathematics, University of Virginia	2011

Working Papers

Heterogeneous Endogenous Effects in Networks [Job Market Paper]

Local Regression Smoothers with Set Valued Outcome Data (with Qiyu Li, Ilya Molchanov and Francesca Molinari)

On Testing Continuity and the Detection of Failures (with Matthew Backus)

Mostly Harmless Regulation? Health Warnings, Electronic Cigarettes, and Consumer Welfare (with Donald Kenkel, Mike Pesko and Hua Wang)

It's Not What You Know But Who Who-You-Know Knows (with Kyle Rozema, submitted, *Journal of Law, Economics and Organization*)

Research Experience

Research Assistant for Prof. Donald Kenkel — 2015-2016
Department of Policy Analysis and Management, Cornell University

Research Assistant for Prof. Francesca Molinari — 2012-2015
Department of Economics, Cornell University

Research Assistant for Dr. Lars Vilhuber — 2014
Labor Dynamics Institute, Cornell University

Research Assistant for Prof. Edward Morrison — 2011
Columbia Law School, Columbia University

Research Assistant for Prof. Douglas Lake — 2010
Cardiology Department, UVA Hospital

Programming Skills

MATLAB, R, STATA, SAS, Linux, Latex

Teaching Experience

Introductory Microeconomics, Cornell University — 2016
Teaching Assistant for Prof. Thomas Lyons

High-Performance Computing for Economists, Cornell University — 2014-2015
Instructor

An Introduction to Statistics, University of Virginia — 2012
Teaching Assistant for Prof. Dan Spitzner

Fellowships and Awards

Ernest Liu 64, Ta-Chung and Ya-Chao Liu Memorial Fellowship, Cornell University — 2016
Mabel A. Rollins Scholarship, Cornell University — 2015
Ruth Ada Birk Eastwood Fellowship, Cornell University — 2015
Jacobs Fellowship, Cornell University — 2015
Sage Fellowship, Cornell University — 2012
Pi Mu Epsilon National Honorary Mathematics Society — 2011
Undergraduate Academic Award, University of Virginia — 2011
Ingrassia Research Grant, University of Virginia — 2010
Putnam Mathematical Competition Team, University of Virginia — 2008

References

Prof. Francesca Molinari (Committee Chair) — +1 (607) 255-6367
Department of Economics, Cornell University — fm72@cornell.edu

Prof. Matthew Backus (Committee Member) — matthew.backus@columbia.edu
Graduate School of Business, Columbia University

Prof. David Easley (Committee Member) — +1 (607) 255-6283
Department of Economics, Cornell University — dae3@cornell.edu

Prof. Marten Wegkamp (Committee Member) — marten@math.cornell.edu
Department of Statistical Science, Cornell University

Prof. Donald Kenkel — +1 (607) 255-2594
Department of Policy Analysis and Management — dsk10@cornell.edu
Cornell University

Prof. Thomas Lyons (Teaching Reference) — +1 (607) 255-9534
Department of Economics, Cornell University — tpl4@cornell.edu

Placement Information

Prof. Karl Shell (Placement Chair) — +1 (607) 255-5277
Department of Economics, Cornell University — karl.shell@cornell.edu

Rachel Lukens (Placement Coordinator) — +1 (607) 255 4893
Department of Economics, Cornell University — rll35@cornell.edu

Last update: October 28, 2016

Columbia Business School
AT THE VERY CENTER OF BUSINESS™

Columbia University
Uris Hall
3022 Broadway
New York, NY 10027

Matthew Backus
Assistant Professor
Finance & Economics

T. 212-851-2446
matthew.backus@columbia.edu

November 26, 2016

Dear Committee Members,

I am writing to recommend Sida Peng to all economics departments without reservation. I have known Sida since I moved to Cornell in the Fall of 2013. He sat in on my graduate IO class; I advised him on several projects, and we began a joint collaboration which I will discuss below.

Besides my overall assessment above, there are two things I want to do with this letter. I'll tell you about our collaboration in a moment, but first I want to emphasize that Sida, while he is a theoretical econometrician with a strong prior background in statistics, *likes applied economics*. I stress this because there is perhaps a stereotype that technically strong econometricians can be somewhat removed from applied research, and in this case, it is wrong. In fact, Sida switched from statistics to economics precisely because he wanted to see a direct connection to applied work.

This is reflected in his job market paper. While I will leave it to his primary advisor, Francesca Molinari, to give you a tour of its technical virtuosity, I want to highlight the fact that developing an empirical framework for studying contagion on multi-layer networks is a first-order economic problem. Which networks should one use to stop a contagious disease or encourage the spread of an idea? What degree of connections? It is a signature of Sida's research agenda that the economic questions are as important as the technical contribution.

Sida and I are coauthors on a project we've provisionally titled "On the Detection and Estimation of Discontinuities." The project endeavors to formalize a LASSO regression-based approach to detecting discontinuities. The context is simple: consider a function that is everywhere continuous but for a finite set of discontinuities. These discontinuities may be of unknown number, unknown location, and unknown type (point, jump, kink, etc).

This problem lends itself to LASSO regression for two reasons: first, because both the discontinuities and the continuous part can be cast in a linear regression (using an unpenalized sieve approximation for the latter), and second, because the number of *potential* discontinuities one

Columbia Business School
AT THE VERY CENTER OF BUSINESS™

Columbia University
Uris Hall
3022 Broadway
New York, NY 10027

Matthew Backus
Assistant Professor
Finance & Economics

T. 212-851-2446
matthew.backus@columbia.edu

could look for is arbitrarily large. It is a linear model selection problem in which one could easily posit more models than observations.

In this setting, Sida showed that, subject to some limitations, the design matrix generated by the large set of discontinuities will satisfy a property called *irreducibility*. This property says, informally, that the regressors are not "too correlated," and is the condition under which LASSO regression is guaranteed to provide consistent model selection.

Of course, in addition to consistency, we would also like to make inferential guarantees about the selected model. Borrowing new results from the statistics literature, Sida showed two things: first, that we can select the LASSO penalty parameter lambda in a way that controls the false-discovery rate. In other words, since post-selection t-tests will not be uniformly valid, Sida builds the inferential guarantee into the model selection procedure itself. Second, he also showed that, using the method of van der Geer 2016, one could generate uniformly valid confidence intervals for the size of the discontinuities selected by the model.

We see many potential applications for our method. On the one hand, discontinuities are often of direct economic interest: discontinuities may represent tax sheltering (Rees Jones 2014) or tipping points in models of residential segregation (Card, Mas, and Rothstein 2008). In my own work (Backus, Blake, and Tadelis 2016), we studied the use of round numbers in bargaining as cheap-talk signals -- formalized as point discontinuities in expected outcomes. Finally, as a diagnostic tool, the method offers an even more agnostic placebo test for RD designs.

I should stress that Sida is a full coauthor on the project and deserves credit for most, if not all, of the technical accomplishments. He has been a pleasure to collaborate with and imperturbably hard-working, even during his job-market preparation. We expect a public draft to be ready imminently (the project has been on hold while we wait on data for one of the applications).

I hope I've made clear that Sida is an excellent researcher with impressive technical skills. I think that he is well-positioned to take advantage of the current enthusiasm for incorporating "big data" methods in econometrics. In addition, Sida is personable and has been a great contributor to public goods both for his cohort as well as for the greater economics community at Cornell.

Columbia Business School
AT THE VERY CENTER OF BUSINESS™

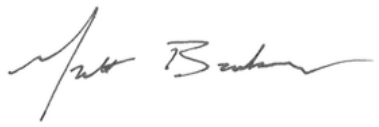
Columbia University
Uris Hall
3022 Broadway
New York, NY 10027

Matthew Backus
Assistant Professor
Finance & Economics

T. 212-851-2446
matthew.backus@columbia.edu

In summary, I believe that *all* economics departments should consider hiring Sida. He will generate positive externalities wherever he lands. Please feel free to reach out to me with any remaining questions at matthew.backus@columbia.edu or 212-851-2446.

Sincerely,



Matt Backus

Cornell University
Department of Economics

Francesca Molinari
Professor
458 Uris Hall
Ithaca, New York 14853
t. 607.255.6367
f. 607.255.2818
molinari.economics.cornell.edu

November 20, 2016:

Dear Colleague:

I am writing to enthusiastically endorse **Sida Peng**'s application for an Assistant Professor position at your Department. I first started to interact with Sida when he was at the end of his first year in the Ph.D. program in Economics at Cornell. In light of Sida's promise as a student, I soon after agreed to chair Sida's dissertation committee, and it has been a pleasure to see him develop into an excellent economist, specializing in microeconometrics and with particular expertise in machine learning methods and the econometrics of network data. Sida's technical skills are excellent: He came to Cornell with a double major in Economics and Mathematics and an MS in Statistics from Virginia. Most importantly, Sida's work is always motivated by economic questions that are policy relevant, as I will soon exemplify by discussing Sida's job market paper. I am very impressed with the quality and quantity of Sida's output, and I expect his production of first rate econometric work to continue in time. As such, I recommend Sida Peng to the very best economics departments without reservations.

Sida's interest in first order economic questions as well as his technical skills are well represented in his job market paper. Here Sida proposes a method to estimate a social interactions model where each individual may influence the outcome of her neighbors differently – something that Sida calls "heterogeneous endogenous effects." To be clear, in this model the connections between individuals (the network) are taken as exogenously given, and their formation is not modeled. Social interaction models with exogenously given networks have received an enormous amount of attention in empirical economics (as well as in sociology), with applications in many fields, including labor, development, public finance, political economy, health economics, finance and many others. But in these applications, each individual influences her neighbors at an identical rate, or individuals with pre-specified characteristics all influence their neighbors at an identical rate (which might differ from the rate with which individuals with different pre-specified characteristics influence their neighbors, but the heterogeneity in effects is based on observable pre-specified exogenous characteristics).

Diversity and Inclusion are a part of Cornell University's heritage. We are an employer and educator recognized for valuing AA/EEO, Protected Veterans, and Individuals with Disabilities.

In Sida's paper, each individual may affect her neighbors at her own rate, and it is these heterogeneous endogenous effects that Sida's method is designed to estimate. With these estimates at hand, a social planner can target the most influential individuals in the network (i.e., in a community) to diffuse information, or to stop the spread of a disease, or to promote healthy behaviors, or to reduce criminality, or to stimulate labor productivity, etc. These "most influential individuals" are referred to as "key players" in the social interactions literature (see, e.g., Ballester, Calvó-Armengol, and Zenou, *Econometrica* 2006)). To date, key players have been identified through measures of agents' "centrality" based on the network structure, and on the outcomes predicted with a homogeneous endogenous effect model when each agent in turn is removed from the network. Sida's work departs from this literature by identifying the key players through observed behavior on the network.

Individuals in a community can often be connected through multiple channels. For example, individuals in rural villages in India may be friends, or lend/borrow rice or kerosene or money from each other, or ask each other advice or medical help, or go to temple together. Each individual may affect those connected to her differently, and may have an effect on her friends that differs from her effect on the people to whom she lends money. Again, from the perspective of a planner who has limited resources and the goal (say) to quickly transmit information or promote healthy behavior, it is important to learn who are the most influential individuals in each network (friendship, borrowing/lending rice, etc.), and through which network the social influence passes more effectively. Sida's work allows one to estimate both heterogeneous individual endogenous effects as well as heterogeneous endogenous effects through different connections.

To set some notation, a simplified standard model of social interaction takes the spatial autoregressive (SAR) form:

$$Y=\lambda MY+X\beta+\varepsilon$$

Where Y is an n x 1 vector collecting each individual's outcome, M is a n x n adjacency matrix with $\{M\}_{ij} = 1$ if individuals i and j are connected, and zero otherwise, and X is a n x k matrix of covariates. In the standard SAR model, λ is a parameter – the social multiplier – representing the effect of i's neighbors on her outcome. Importantly, λ is a scalar: each individual has the same effect on the individuals she is connected to.

Sida significantly extends the scope of the SAR model, to allow for: (1) λ to be a n x 1 vector with each entry λ_j, j=1,…, n, equal to the effect of individual j on her

neighbors; (2) The presence of multiple types of connections (i.e., networks 1,…,q), yielding a model of the form

$$Y=\lambda^1 M^1 Y+\ldots+\lambda^q M^q Y+X\beta+\varepsilon;$$

(3) A combination of homogeneous and heterogeneous endogenous effects.

Looking at the model, it becomes quickly apparent that it contains more parameters than observations, in particular it has $nq+k$ parameters. However, it seems natural to assume that the number of key players in a population be small relative to the size of the entire population – in jargon, the model is arguably *sparse*.

Sida's fundamental insight is that the problem is then mathematically similar to a "big data" scenario with $p > n$ parameters. As such, methodologies for model selection and estimation are applicable. And indeed, Sida shows that the (square root) least absolute shrinkage and selection operator (LASSO) method can be used to consistently learn who the influential individuals in the network are, and to consistently estimate their effect on their neighbors. Specifically, Sida achieves this through a novel two-stage LASSO procedure that he proposes, where in the first stage he exploits instrumental variables within the LASSO approach to deal with the endogeneity of individuals' outcomes to their neighbors outcomes, and in the second stage the "fitted first stage regressors" are used again in the LASSO to learn who the influential individuals are and to estimate their effect. Sida also shows that the square root sparse LASSO can be used to learn the most effective networks in passing information. Finally, Sida provides a method to build uniformly valid confidence intervals for the estimated social interaction effects.

This part of Sida's paper is a technical tour-de-force, where he is able to bring together and build on state-of-the-art results from Belloni, Chernozhukov and Wang (*Biometrika*, 2011), Belloni, Chernozhukov and Hansen (*Review of Economic Studies*, 2014), Belloni, Chenozhukov and Kato (*Biometrika*, 2014), van de Geer, Bühlmann, Ritov and Dezeure (*Annals of Statistics*, 2014) and Bunea, Lederer and She (2013) in a seemingly effortless manner, to extend them to the network context and obtain results that are very important for empirical work. Sida also derives asymptotic statistical properties for the square-root sparse group LASSO estimator that are novel to the literature. He applies his method to a dataset previously analyzed by Banerjee, Chandrasekhar, Duflo, and Jackson (*Science*, 2013), to study villagers' decisions to participate in micro-finance programs in rural areas of India. Sida's results suggest that key players in these villages significantly affect others' decisions to join the micro finance program, and that the type of connections that the villagers have also significantly impact their decision to join the program. Interestingly, the influential

individuals selected by the LASSO, while well connected, are not necessarily the individuals with highest centrality measures.

Sida has been extremely productive during his time at Cornell. He has 3 nearly completed working papers co-authored with faculty members, and one completed working paper co-authored with a fellow graduate student and already submitted to a journal. Of the papers co-authored with faculty, one is joint work with Donald Kenkel, Mike Pesko and Hua Wang. In this paper the authors study the tradeoffs involved in regulating electronic cigarettes. My colleague Don Kenkel is writing a letter in support of Sida's candidacy at your Department – and I will leave it to him to discuss in detail the paper. On my side, I want to emphasize that I have discussed the paper often with Sida, and I have seen him present it at a Cornell seminar. I was extremely impressed with Sida's knowledge of the institutional details, his ability to clearly explain the challenges facing identification of the parameters of interest, and the experimental set-up used for data collection. I believe this paper will be submitted for publication early next year.

Sida has also co-authored a paper with Matthew Backus at Columbia GSB. The paper proposes a LASSO regression-based approach to detecting discontinuities. Matt Backus is writing a letter in support of Sida's candidacy at your Department – and again I will leave it to him to discuss in detail the paper. But again I will emphasize something. As I said in the opening of this letter, while technically very skilled Sida is always motivated by economic questions. The paper with Backus is another great example of this: a while ago Matt Backus presented a paper of his at Cornell (Backus, Blake, and Tadelis 2016), where the authors study the use of round numbers in bargaining as cheap-talk signals -- formalized as point discontinuities in expected outcomes. The original Backus, Blake and Tadelis paper stimulated a conversation between Matt and Sida on strategies to learn the location/size of discontinuities in a "big data" environment (the dataset used in Backus, Blake and Tadelis is very large), which eventually led to their joint paper. I believe this paper will be submitted for publication early next year.

Finally, Sida has co-authored a paper with Qiyu Li, Ilya Molchanov and me on local linear regression smoothing with interval valued outcome data. The paper builds on earlier work in Beresteanu and Molinari (*Econometrica* 2008), where the authors proposed the use of random set theory to test hypothesis and build confidence intervals in partially identified models. One of the examples in Beresteanu and Molinari is that of ordinary least squares with interval outcome data. The Li, Molchanov, Molinari and Peng (2016) paper extends that method to the case of local linear regression. It turns out that certain technical subtleties have to be addressed in

this case, to avoid certain types of finite sample bias. The paper is nearly complete and I expect that we will submit it for publication by the end of the year.

The excellence of Sida's work has been recognized repeatedly at Cornell, where he has been awarded several fellowship and one of our signature awards, the Liu Memorial Fellowship. While I cannot independently attest to Sida's teaching skills because he has not served as teaching assistant in any of the courses I teach, I understand that my colleague Tom Lyons is providing a teaching evaluation letter on Sida's behalf. On my side, I expect Sida to be a very good teacher.

In addition to being my coauthor, Sida has also worked for me as a research assistant. He wrote the code for many of the Monte Carlo simulations in my joint work with Hiroaki Kaido (at Boston University) and Joerg Stoye (also at Cornell) on "Confidence Intervals for Projections of Partially Identified Parameters", and wrote the code for Monte Carlo simulations of my joint work with Haim Bar (at the University of Connecticut, Department of Statistics) on "Computational Methods for Partially Identified Models via Data Augmentation and Support Vector Machines." Sida's programming skills are outstanding, and he understands very well the econometrics literature on moment inequalities and its subtleties and challenges, as well as the statistical learning theory literature on support vector machines. I view these as additional extremely important skills that will continue to prove useful in Sida's research.

In sum, I give Sida Peng my highest recommendation for any department, including the very best. He is extremely creative and hard working. He has already made a number of contributions to econometric theory in his analysis of questions of first order importance, and I expect he will continue to do so. He will be a great addition to your department.

Please, do not hesitate to contact me if I can provide further information on Sida Peng's suitability for a position at your Department. My email address is fm72@cornell.edu and my phone number is 607-255-6367.

Sincerely,



Francesca Molinari

Cornell CIS
Statistical Science

Marten Wegkamp
Faculty of Computing and Information Science
Professor and Chair

1194 Comstock Hall
Ithaca, New York 14853-4201
Phone: 607-255-4360
Fax: 607-255-9801
Email mhw73@cornell.edu

November 21, 2016

To Whom It May Concern:

This letter is in response to your request for an evaluation of Mr. ***Sida Peng*** who has applied for a position at your department.

I am a minor member of Sida's PhD Committee. He took my math 7740 Statistical Learning class. This is a small PhD level mathematics class where I discuss classification and other high dimensional statistics topics. Sida did a project on the *square root lasso* and gave a nice presentation. He deservedly got an A.
For his thesis, *Heterogenous Endogenous Effects in Networks,* Sida studies the use of networks in a regression setting. Rather than a standard regression setting, I find Sida's setting fascinating since it is highly relevant from a practical perspective, while from a technical perspective, the modeling and dealing with all dependencies involved is very challenging. Sida dealt with the identification issues, developed a two-stage lasso estimator and shows that certain de-biased versions can be used for valid post-model selection inference. The inclusion of the network and the fact that the dependent and independent variables are dependent in a complicated way, makes this analysis so interesting. Sida is very knowledgable of the relevant literature and has done an admirable job. He is a modest, integer and smart person with an easy going nature. I strongly recommend Mr. Peng.

Sincerely,



Marten Wegkamp
Professor, Department of Mathematics, Cornell University
Professor and Chair, Department of Statistical Science, Cornell University