UNIVERSITY OF CALIFORNIA, BERKELEY

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO   SANTA BARBARA • SANTA CRUZ

WALTER A. HAAS SCHOOL OF BUSINESS   Telephone: (510) 643 1898
Minjung Park, Assistant Professor of Marketing   Email: mpark@haas.berkeley.edu

10/17/2014

To Whom It May Concern:

   I am writing to recommend Yongdong Liu for a position at your institution. I have known
Yongdong for the past 3 years, and have closely worked with him on his dissertation during the
past couple of years. Below, I shall briefly describe some of Yongdong's research
accomplishments. However, the bottom line is that Yongdong is a talented young researcher
with very strong quantitative skills with a particular focus on big-data-related techniques.

   Yongdong is a Ph.D. candidate in the department of Agricultural and Resource Economics at
the University of California, Berkeley. I first got to know him when he contacted me three and a
half years ago asking for permission to attend marketing seminars at Haas. Since then, he
regularly came to my office hours in order to discuss his research progress, and I have been
closely involved in his dissertation work during the past two years.

 One of Yongdong's strongest assets is his technical and quantitative skills, especially his ability
to work with Big Data using sophisticated machine learning tools. With the advent of massive
consumer-level, government and business datasets, the econometric tools that have been
traditionally used in empirical research may not be sufficient and new techniques may be called
for. In working with massive datasets, there could be a need to effectively reduce the
dimensionality of the data without losing valuable information and a researcher might need to
deal with unstructured and hard-to-categorize information, such as texts that show up in
customer reviews and product description. Traditional econometric tools are not adequate to
address these challenges, and I expect that such data challenges and the toolkits required to
address them would become relevant for more and more researchers due to the arrival of Big
Data. Yongdong is well versed in machine learning tools, such as text classification algorithms
and model selection techniques, which are crucial in meeting those challenges. This would make
him a great addition to any school that wants to strengthen its quantitative group and he would
also be a wonderful colleague to have for anyone but especially those who are interested in doing
research with massive datasets.

 Yongdong has written a couple of papers on the mobile apps market. One of them is his job
market paper and the other is a joint project with Denis Nekipelov and me. In these papers, he
asks the following questions: What determines app developers' choice of platform and how does
competition affect the platform choice? Typically app developers initially launch their apps on
one of the two platforms (Android and iTunes; simultaneously launching on both platforms is
rare), and might decide to enter the other platform later depending on the performance of their
apps on the first entered platform. His job market paper examines the first part (which platform

to enter initially) and our joint project examines the second part (what influences the timing of entry into the second platform and how firms might use versioning strategy in their entry decision - entering early with a relatively unpolished product in order to preempt competitors and upgrading after entry). Since Denis Nekipelov will discuss the joint project in depth in his letter, I will focus on discussion of Yongdong's job market paper in my letter.

In his job market paper, Yongdong examines app developers' platform choices, with a particular focus on how the quality distribution of potential and existing competitors on the two platforms influences a firm's platform choice. He models the platform choice using a static game framework with incomplete information. There are some data challenges and estimation challenges he faced, and the way he met those challenges highlights his strength and perseverance. Below, I will discuss those challenges and how he addressed them.

In his job market paper, Yongdong uses a huge and unique dataset on the universe of mobile apps on Android and iTunes. Yongdong played a key role in obtaining the dataset. He was the one who reached out to the CEO of a company that collects the apps data, and he prepared proposals to convince the CEO to share the data as well as the company's proprietary data mining algorithms with us. This clearly demonstrates that he is the type of person who has initiative and makes things happen.

The dataset is not easy to work with, to say the least, due to its size and complexity. The raw data are almost 50 TB and there are a few billion observations in the database. One cannot even run a simple regression using standard statistical software on the dataset, and researchers often hire professional programmers to manage datasets like this. However, Yongdong quickly learned necessary tools such as Python and SQL to work with the dataset, optimized the initial data structure, and has been performing analyses on the data all by himself. He is extremely good at figuring out what tools would be necessary to address the problems at hand and quickly mastering them. For instance, to prepare the data for empirical analysis, he had to merge two datasets, one for Android and another for iTunes, and one of the difficulties in doing so is that there are inconsistencies in naming of apps between the two platforms (such challenges often arise when researchers need to combine multiple datasets using a common identifier). To do the correction and successfully merge the datasets, he mastered sophisticated tools from the machine learning literature that allowed him to perform matching by creating a measure of distance in the non-metric space (in his case, app name).

Another data challenge he met was that he had to re-define markets for his analysis since the genre classification in the original datasets is extremely coarse (there are thousands of apps that belong to the game genre, but clearly not all of them directly compete with each other as there are various kinds of game that appeal to different sets of users). He mastered text mining techniques to extract from non-metric variables, such as product descriptions and customer reviews, information that can be used for classification. In the typical empirical analysis the researchers only consider purely numeric characteristics, or string characteristics that have been converted to the numeric form. However, in many instances, the string data may not be easily convertible to the numeric form but still contain information that is a lot more important than the observed consumer actions. For instance, the product review by a consumer reveals a lot more about this consumer's satisfaction with the product than the consumer's choice to purchase the product. Similarly, the sequence of searches typed into the command line of the search engine may be very important for understanding how the thought process leads a particular consumer to the action of interest. Extraction and subsequent use of the string information in conjunction with the traditional numeric information could significantly enrich empirical analyses, and Yongdong's knowledge of the relevant tools will allow him to be at an advantage in these areas of importance.

A key estimation challenge he faces comes from the fact that there are many players in a given market (even after refining the market definition as discussed above). Since it is almost impossible to solve a game with more than a hundred players, he uses a two-step pseudo maximum likelihood estimator proposed by Aguirregabiria and Mira (2007). Quality of an app is not directly observed, and he recovers app quality by estimating a demand model that is simple but nicely reflects the key feature of consumers' app choice – consumers' limited knowledge of available apps due to high search costs and the resulting heavy reliance on the ranking charts.

One of the key findings from his analysis is that while the presence of high-quality competitors on a platform is viewed as detrimental by an app deciding which platform to enter, the presence of low-quality apps on a platform actually makes the platform more attractive to other potential entrants with low quality. This result seems a bit surprising at first, since it is hard to imagine why presence of competitors, regardless of their quality levels, would attract more firms to enter the market. Yongdong explains that this could be due to the network effects that naturally arise in this two-sided market. Although having more competitors is clearly bad because of business stealing effects, more apps on a platform might induce more consumers to adopt the platform (in this case, adoption of a platform by a consumer would mean purchase of a mobile device that is compatible with that platform), making the platform more attractive to other potential entrants (network effects). As a result, the net effects of having more competitors on the entry probability of the focal firm are not clear a priori, and Yongdong finds that the benefits of having more competitors due to network effects outweigh the costs of having more competitors due to business stealing effects in case the competitors are low quality.

Overall, this is a very interesting paper. Yongdong shows deep knowledge of the industry and strong technical skills. Despite the impressive technical skills Yongdong possesses, he is not just a technician. He learned all those tools because he wanted to answer important questions using exciting datasets.

Before concluding the letter, I want to note that Yongdong went on the marketing market this summer. His job market paper and other related working papers examine topics that are of significant interest to marketing researchers, so he decided to try the marketing job market. Unfortunately, his outcome wasn't as great as hoped for. I think a major factor that might have worked against him on the marketing market is his communication skills as well as his somewhat unusual background (a Ph.D. candidate from an ARE department isn't easy to find on the marketing job market). Although Yongdong clearly communicates his ideas without any problem whatsoever, business schools have a somewhat tougher requirement for presentation skills than econ departments typically do due to the need to teach MBAs, and some schools might have worried that he might not be an effective instructor for MBAs. To Yongdong's defense, he has been working very hard to improve his presentation skills (for example, he has been actively participating in a discussion club) and I could very clearly see his significant progress over the past couple of years. I know that he will continue to make efforts to improve even more (he is very tenacious), and I have no doubt whatsoever that he will be an effective teacher.

I encourage you to talk to him at ASSA. I am sure that you will greatly enjoy learning about his research and will be impressed by his passion. If I can provide any additional information, please do not hesitate to contact me.

Best regards,
Minjung Park