UNIVERSITY of VIRGINIA
DEPARTMENT OF ECONOMICS

DEPARTMENT OF ECONOMICS
237 MONROE HALL
CHARLOTTESVILLE, VA 22904-4182
TELEPHONE: (434) 924-3177
FAX: (434) 982-2904
E-Mail: *denis@virginia.edu*

DENIS NEKIPELOV, ASSOCIATE PROFESSOR OF ECONOMICS
AND COMPUTER SCIENCE

June 11, 2014

Dear Chair and Members of the Recruiting committee,

I write this letter to express my strong support for Mr. Yongdong Liu. I first met Yongdong when he attended my course "Computational and Econometric Methods of Industrial Organization" which I taught in UC Berkeley three years ago where he was clearly among the brightest students that I have ever had a pleasure teaching. Further on, he became a collaborator and an active contributor on several projects (including my ongoing collaboration with Computer Science departments at Northwestern University and the University of Wisconsin at Madison) that we are still working on together. Yongdong is one of those rare students who combines excellent technical skills, ability to learn and a good amount of entrepreneurial spirit that allows him to scout new interesting problems and data sources. Yongdong was able to communicate with the Silicon Valley company AppMonsta, and convinced them to share their extensive dataset of mobile apps on iTunes and Google play. What is even more interesting, he was able to also convince them to share some of their proprietary data-mining algorithms and do all that free of charge.

Yongdong comes from an unusual background. Berkeley's Agricultural and Resource Economics (ARE) department. However, the structure of the graduate program is very well coordinated across fields between ARE, Haas and the Economics department. As a result, students can discover the love for the fields where they may not have thought of working in when they started the Graduate school. Thus, after discovering his affinity to Industrial organization and Econometrics, Yongdong started working with Minjung Park at Haas marketing and with me. I have worked as a dissertation co-advisor with several students at UC Berkeley and I think that Yongdong is compatible with the best of them. However, Yongdong has one big advantage that sets him apart: he mastered "Big data" technologies making him a unique "go to" person that many research institutions are currently looking for.

The work with AppMonsta required manipulation with very large (~50TB) datasets which included complicated imputation and linking procedures. Standard statistical software common in Marketing and Economics would not allow Yongdong even to run regressions. After very minor guidance, he was able to master Hadoop and Python. One of the serious data issues was that iTunes and Google play use very coarse definitions of markets such that in many cases the apps that are clearly not related would be put into the

same genre category. Yongdong implemented advanced classification procedures that transformed the text description and reviews of apps into the vectors of features, cluster the feature vectors to extract sub-markets for a smaller subset of data (where complex computation is feasible within a short amount of time) and then applied the Support Vector Machine to classify the rest of the data to identify individual markets. Text mining has an amazing potential to inform the researcher about the product attributes that are simply not available in the quantitative form. In many markets we do not have any quantitative information about the products except, possibly, prices and quantities. Text mining can significantly enrich the space of product characteristics allowing to create literally thousands of product characteristics in the market where product descriptions and consumer reviews of the products are available. Yongdong has definitely mastered this technique and I anticipate that he will produce many other successful research projects just based on these techniques.

Yongdong also used Hadoop to implement sparse regression procedures for efficient variable selection. The sets of features that come from "data rich" settings (such as unstructured text or user behavior histories) tend to be very large making it impossible to estimate meaningful statistical models. As a result, to make the model interpretable and, more importantly, robust, one needs to reduce the feature vector to a subvector of relevant features. A variety of Machine Learning techniques, such as regression trees, LASSO and deep learning procedure have been found effective for such a reduction (and some also have been proven having good statistical properties). Yongdong was able to successfully learn and apply these techniques. This is another skill that is vital in "Big data" settings.

The great thing about Yongdong's work is that behind all these amazing techniques there is true important substance: Yongdong studies the incentives for innovation in large dynamic market settings. This is a very important problem that until recently was hard to study empirically due to serious data limitation. The market for mobile apps does not have this problem!

The market for "apps" - programs that can be run under the operating systems that are installed on portable computer devices such as smartphones and tablets - has become extremely popular over the past few years, and it represents a dramatic showcase of trade-offs that arise for innovative firms in a highly competitive marketplace. The app market is represented by two dominant platforms that distribute apps for two competing operating systems: Android and Apple. The apps are not directly portable between the two platforms. Without a substantial change in the source code an app developed for Apple cannot run on an Android device. In his dissertation work Yongdong studies the decisions of developers to (a) choose the platform of the "first entry", and (b) make a decision to port apps that were successful on one platform to another platform (cross-platform entry). In a competitive marketplace such a decision can be complex. First of all, the app market (even within distinct categories) is a dynamic highly differentiated market where each product has a very rich set of characteristics. As result, standard hedonic approaches for demand prediction do not apply and the analysis of potential demand for the app on a new platform becomes a very complex problem for the developers. Second, even within small app categories there is a large number of developers who simultaneously make decisions to create new apps and port the apps

237 Monroe Hall • P.O. Box 400207 • Charlottesville, VA 22904-4182
Phone: 434-924-3177 • Fax: 434-982-2984 • E-mail: denis@virginia.edu

across platforms. In this environment many developers base their decisions to move across the platforms by using a combination of statistical analysis using historical data from other developers and the data from experimentation by rolling out the prototypes of their apps (which are usually less tested with lower functionality than the full versions of their apps).

One of Yongdongs main goals is to study such a versioning phenomenon in a competitive marketplace. If there are multiple competing developers who have apps that serve similar functions, the effort to introduce a quality app can be undermined by competitors' entry which can divert the user base and reduce the potential profits from entry. In this case, a developer making such a decision faces a tradeoff between finishing the beta-testing to create a quality app at the risk of losing demand to competitors and introducing an app quickly to preempt competitors without adequate testing at the risk of receiving poor user ratings as well as incurring an increased cost of fixing a running app. Yongdong's work provides evidence that the timing and quality decision of entry in the app market is strategic and that the firms making decisions about porting their apps to another platform take into account the actions of their competitors.

The entry and prototyping decisions in the app market are strategic dynamic decisions under uncertainty with a high-dimensional state space (in our data, the state space exceeds 100,000 dimensions). The structural model that Yongdong has estimated is based on the Bayes-Nash equilibrium where the developers make inferences regarding the app development processes of their competitors. Then, on the basis of this prediction they make entry decision that trades off the quality of the app that they want to introduce on the new platform and the duration of the development process. The Bayes-Nash equilibrium is characterized by the equilibrium strategies (corresponding to the entry and versioning decisions) and the beliefs of developers regarding their competitors. To recover developers' beliefs, one needs to determine the set of "potential competitors" for each developer. There is no such information that comes directly from the marketplace: the app platforms use broad definitions of markets (genres) that pool together thousands of apps that are clearly not in direct competition with each other. Yongdong provided a new econometric technique that allowed him us to estimate the set of potential competitors by combining the existing techniques from the natural language processing, support vector machines, and the standard tools for non-parametric regression estimation. It is interesting to note that the only information that is available through the marketplace which characterizes the potential market for the app comes as verbal information in the app description and the user reviews. Since this information is not quantitative, we cannot apply standard Econometric techniques. To make the inference possible, Yongdong "normalized" the verbal information available for each app. This is done by forming a dictionary of words used in each app's description and user reviews and removing the "stop" words (such as prepositions and articles). Then each app becomes represented by a very large vector of descriptive words. Then Yongong defined the distance in the space of such non-numeric vectors which is based on relative overall frequency of descriptor words in the marketplace and the frequency of such words in the description of a given app. Using the constructed distance, he clustered the apps using the k-means clustering approach. He then used the support vector machine to estimate the boundary of each cluster. The support vector machine (SVM) is powerful computational technology for binary classification. The idea behind SVM is to construct a boundary between two

237 Monroe Hall • P.O. Box 400207 • Charlottesville, VA 22904-4182
Phone: 434-924-3177 • Fax: 434-982-2984 • E-mail: denis@virginia.edu

classes of points in an arbitrary semi-metric space such that the distance from this boundary to each class is maximal. It turns out that the location of the boundary will be determined only by a small number of points from each class, called the support vectors. For our analysis these support vectors are the subsets of words from the app description that essentially define markets. This leads to a substantial reduction in the dimension of the vectors describing apps. For each app than we can identify the "market" of competing apps based on their descriptors.

The developer's strategy then incorporates the entry decision and the decision to upgrade the prototype within each market. Yongdong assumed that these decisions are made on the basis of quantitative characteristics of the dynamic demand in each market. The set of such characteristics is very large. To estimate the "reduced form" of the developers' strategy Yongdong included the features of demand (i.e. the full history of the number of downloads) for each competitor. Then Yongdong estimated the structural parameters of the app developers' payoffs using plug-in estimates of the reduced form strategies for estimated potential competitors.

Yongdong is currently finishing a set of very interesting counterfactual experiments, such as an increase in demand for certain app categories, transfers from the platform to the developers as well as 'A+B'-type contracts between the platform and the developers, on the resulting app quality.

To conclude, I think that Yongdong is a great young researcher with good prospects in academia. He is working hard to improve his every academic quality. For instance, two years ago he joined a discussion club just to make sure that he can be persuasive and effective in his academic presentations. He will make a great colleague and a collaborator. I believe that research Universities interested to have a person with very strong quantitative skills who thinks hard about important economic problems should give Yongdong a strong consideration for an Assistant Professor position.

Please, do not hesitate to contact me if you have any questions regarding Mr. Liu and his research at denis@virginia.edu or call me at (919) 308-9035.

Yours sincerely,

Denis Nekipelov

237 Monroe Hall • P.O. Box 400207 • Charlottesville, VA 22904-4182
Phone: 434-924-3177 • Fax: 434-982-2984 • E-mail: denis@virginia.edu