Does the Network Matter?

Pritha Dev
ITAM - School of Business

February 9, 2011

Abstract

Network formation theory is based on the assumption that the benefits of belonging a network depend on the number of people a person is linked to which includes direct links, links of links, links of links of links, etc. Most empirical work on the other hand assumes that person a's decision to link to person b is not affected by the other links of person b. This paper seeks to bridge the gap between theory and empirical work by allowing links of link to enter each person's link formation decision.

We use a rich dataset from the Eastern Region of Ghana to test our hypothesis and we find that links of links does indeed have an effect on the individuals link formation decision.

Keywords: Link formation, Binary Choice Model, Endogeneity

JEL Codes: D85, C25, C26

1 Introduction

In most networks studied by economic theory, the higher the number the connections a node has, the more is its the profitability. Consider networks of information where information is transferred from node to node. In these networks, more links, links of links, etc, a person has; the more information he gets and thus the more profitable his position. All this implies that anyone choosing to form a link would consider (among other things) the shape of network or in particular, the connectivity of the person he is contemplating linking with. The more links this person has, network formation theory would imply, the more likely it is that the link is formed. Most previous empirical work has implicitly assumed the contrary and that the probability of a link depends only on the direct benefits of the link and not on the links of the link. This paper, on the other hand, posits that players might also be taking into account the indirect benefits of being linked to another person. In other words each person takes the shape of the network into account when making his decisions to link and this paper allows the links of links to enter into each person's decision formation. We hope to see that the more connected a player is, the more beneficial it is for another player to link to him.

Note: This analysis might still be thought partial in the sense that it does not allow more indirect link like links of links of links to enter into the decision, but we believe that decay in benefits is high enough to wipe off any benefits from more indirect links and moreover such an analysis would require observations from numerous networks.

We think of the network as the result of everyone simultaneously deciding who to link with. If we then allow the total links of j to affect the decision of i to link to j, we have an endogeneity problem because the total links of j are determined simultaneously and are affected by $i's$ decisions. Since the decision to link is binary, we have endogeneity in a binary choice problem. I use the control function approach to tackle with the endogeneity and model $j's$ total links as function of his characteristics and the characteristics of the representative/mean individual in that network. The rationale for using this control is as follows: Previous literature has modeled the decision to link as depending on the individual characteristics of the two people. A link between i and j will be cheaper, the more similar they are in their characteristics. In this paper I model $i's$ decision to form a link with j as depending the characteristics of i and j, as well as on the total number of links j has. In this case, each one of a person's decisions will depend on how similar his characteristics are with the other person, and since his total links is a sum of all these decisions, we can think of a person's total links as a function of how similar his characteristics are to the representative/mean individual in that network.

Another issue that still remains is that these link decisions are inherently correlated. For instance, $i's$ decision to link with j and i decision to link with k are presumably correlated and similarly, $j's$ decision to form a link with i and $k's$ decision to form a link with i are again correlated. This kind of spatial correlation has been studied by Conley() and we use the corrected standard errors he suggests in our section on robustness.

To test the hypothesis, I use an unusually rich data set collected in four clusters of villages in the Eastern Region of Ghana collected by Chris Udry and Markus Goldstein. The data was collected over the course of two years and fifteen modules in a four village clusters in Eastern Region of Ghana. In each village 60 couples/triples were questioned. The network data used here was collected by asking each individual in the sample about seven randomly selected (without replacement) from the sample and three focal village residents.

Related Literature: The theoretical literature in economics on network formation follows two main strands - one follows Jackson and Wolinsky (1996) and the other follows Bala and Goyal (2000). A recent survey is Jackson (2005). Among empirical literature dealing investigating networks Krishnan and Sciubba (2006) investigate the effect of the number of links as well as the structure on labour sharing networks. They present a model of network formation, where each person decides who wants to share his labour with, and labourers have different productivities. They test the predictions of their model on data from rural Ethiopia. Durlauf and Fafchamps (2005) is a recent review of literature dealing with social capital and networks.

Goldstein and Udry (1999) give a detailed description of the data. Udry and Conley (2004) use the same data to analyse the information, capital, labour and land networks in the same data set. Bandiera and Rasul (2006) study the relationship between the network and probability of adopting a new production technology by farmers in Mozambique. They find that the probablity of a farmer adopting a new technology is increasing in the number of adoptees in his network, if that number is small and decreasing if the total adoptees in his network is large. This paper is also related to work by De Weerdt (2002), De Weerdt and Dercon (2006), Fafchamps and Lund (2003), Foster and Rosenzweig (1995), Munshi (2004) and Besley and Case (1997).

2 Identifying Network Effects

I first present a standard theoretical model and then propose a way to estimate it. I assume that the observed network is the equilibrium of a one-shot game of network formation played by a set of individuals denoted by $N = \{1, ..., n\}$. For any player, the benefit from the network is the number of individuals accessed and the costs are from making direct links. Individuals maybe accessed through direct links and also by indirect links. The cost is assumed to depend on the characteristics of the two individuals linked.

Let $g_i = \{g_{i,1}, ...g_{i,j}, ...g_{i,n}\}$ denote player $i's$ strategy where $g_{i,j} = 1$ denotes that i links with j. Let $g = \{g_1, ..., g_i, ..., g_n\}$ denote the network. The distance, d, between two individuals in a network is defined as the minimum number of links connecting them. Let $\mathbf{X}$ denote the characteristics matrix, $X_i = \{x_{i1}, ..., x_{id}, ..., x_{im}\}$ denotes the vector of individual characteristics along m-dimensions of identity. Let Π denote the profit function, where

$$\begin{aligned} \Pi_i(g) &= \Phi(n_i(g), n_i^d(g), \mathbf{X}), \text{ where} \\ n_i(g) &= \text{number of people accessed by i in g} \\ n_i^d(g) &= \text{number of people with whom i forms links in g} \end{aligned}$$

The above model assumes that benefit from a link is the same regardless of the distance. A more general version of the model will allow for some decay in the passage of information and so indirectly accessed links will not be as valuable as directly accessed links. Such a model will have a variable δ_d which denotes the decay in benefit if the accessed individual is at a distance d.

To simplify the model and make it more easy to estimate, lets assume:

A.1 The profit function is linear in benefits and cost, i.e. profit = benefit - cost

A.2 Benefits from all links at distance d are δ_d and do not depend on individual characteristics.

A.3 Cost of forming any link depends only on the characteristics of the two individuals involved in the link.

A.4 To simplify the model further, assume that i can access $k's$ information only if $g_{ik} = 1$ or there is some other player j such that $g_{ij} = g_{ji} = 1$.

Under these assumptions, the profits function is now additive in each link formed. A link will be formed if it yields positive profits given that its not accessed by some other link. Let π_{ij} be the profit to i from forming a link with j and $-c(X_i, X_j)$ be the cost of a link given the individual characteristics. Define m_j to be the total links of agent j;

$$m_j = \sum_{k \in N} g_{j,k}$$

The decision by i to form a link with j, i.e. $g_{i,j}$ can then be seen as the following[1]:

$$
\begin{aligned}
g_{i,j} &= 1(\pi_{ij}(g) > 0) \\
\pi_{ij}(g) &= \delta m_j + c(X_i, X_j) + \varepsilon_{ij}
\end{aligned}
$$

Empirical work till now has implicitly assumed $\delta = 0$ or that the network does not matter, and so they did not have any endogeneity problem. But in the above model, $g_{i,j}$ is modeled as a function of m_j where both are determined simultaneously and hence, ε_{ij} is not independent of m_j. Since theoretical work does assume a positive δ, this paper investigates if shape of the network matters or if players just take into account who they are directly linked to.

Economic theory and empirical work suggests that $c(X_i, X_j)$ should actually be a function of the social distance between the two people, in particular it could be linear in social distance or that

$$c(X_i, X_j) = d(X_i, X_j)'\beta; \text{ where } d(X_i, X_j) \text{ is some measure of social distance.}$$

In particular we will use

$$
\begin{aligned}
d(X_i, X_j) &= \{d(x_{i1}, x_{j1})..., d(x_{id}, x_{jd}), ..., d(x_{im}, x_{jm})\} && (1) \\
d(x_{id}, x_{jd}) &= x_{id} - x_{jd} \qquad \text{if dimension-}d\text{ is continuous} && (2) \\
&= 1(x_{ki} = x_{kj}) \quad \text{if dimension-}d\text{ is discrete}
\end{aligned}
$$

To estimate the model where we include the endogenous variable m, we propose to use the control function method. We will model m as a function of exogenous regressors and an error term. The endogeneity will then be assumed to be a result of the relation between the error in m_j and the error in π_{ij}. Since m is a sum of indicator functions, it is not possible to derive the exact functional form of it. But if we model a single link decision as a function of social distance, then we can think of the m (which is the sum of single link decisions) as being a function of the social distance of the of each individual from the representative/average individual. Denote the average individual's characteristics by X_A. Now the model to be estimated (assuming all equations to be linear) is:

$$
\begin{aligned}
g_{ij} &= 1(\pi_{ij}(g) > 0) \\
\pi_{ij}(g) &= \delta m_j + d(X_i, X_j)'\beta + \varepsilon_{ij} \\
m_j &= d(X_j, X_A)'\gamma + \eta_j \\
\varepsilon_{ij} &= \rho\eta_j + \nu_{ij}
\end{aligned}
$$

[1] Ideally the link benefit should also depend on a indicator taking value 1 if i accesses j through some other link, but that variable is suppressed here.

where ν_{ij} and η_j are independent of all the regressors and all errors are assumed to be normal.

A remaining problem is that the errors might be spatially correlated. The error terms their errors ε_{ij} and ε_{ik}. might be correlated because they capture the error in g_{ij} and g_{ik}, both of which are $i's$ decision to form a link with different players and might be correlated. Similarly, g_{ji} and g_{ki} reflect $j's$ and $k's$ decision, respectively, to form a link with i. Since both of these decisions depend on the characteristics of i, they are not independent and neither are the errors ε_{ji} and ε_{ki}. This spatial correlation will be dealt with in the section on robustness.

3 Data

The data was collected by Chris Udry and Markus Goldstein over the course of two years and fifteen modules in a four village clusters in Eastern Region of Ghana. In each village 60 couples/triples were questioned. The network data used here was collected by asking each individual in the sample about seven randomly selected (without replacement) from the sample and three focal village residents. The questions asked were:

Could you go to ___ if you had a problem with unhealthy crops?

Could you go to ___ for advice about when to apply a new kind of fertilizer?

Could you go to ___ if you wanted to discuss changing your method of planting?

Could you go to ___ if you wanted to find a buyer for any of your crops?

The answer to these questions implies the presence (or not) of an informational link. Table 1 presents the summary statistics for these four types of informational links. Table 2 presents the summary statistics for the endogenous variables which take the value of the total link each respondent has for each of these link types. Given our sample, the maximum number of links possible is ten and minimum is zero.

If we think of the village residents as the population participating the network formation game, then the randomly selected 60 couples and further their links with randomly selected seven individuals from within that sample, allows us to see a randomly selected portion of the network. Analysing the structure of connections within this portion of the network would give us a good idea of the actual network.

I also use data on identity or individual characteristics for both the respondent and the match and this includes information on their religion, clan, gender, age, wealth, primary occupation, soil type, if they are the first of their family to reside in that village, school level and the experience in different crops grown. Using information on individual characteristics, I construct variable measuring the distance between two individuals. In particular, for discrete characteristics like religion, clan, gender, occupation, soil type, if they are the first of their family to reside in that village and school level; I construct a variable taking value 1 if both have the same characteristic and 0 otherwise. For continuous characteristics like age, wealth and experience in different crops, I construct a variable taking on the difference between the respondent's and match's characteristic's values. The only variables for which I don't do this is the variable Office (indicator variable taking value 1 if respondent holds office and 0 otherwise) and Moffice (indicator variable taking value 1 if match holds office and 0 otherwise). Whether the respondent/match holds an office might be correlated with the perceived value of information to be gained from them. We can easily imagine that villager assume that office holders might have more access to information, or might have been given the office because of their knowledge; in either case the fact that someone holds an office would mean that they have more information to give. Moreover, the holder of the office might have less need to form information links, since he possibly has more information than his links. For that reason we include directly Office and Moffice, instead of a derived variable. The summary statistics for these variables is presented in Table 3 .

Another set of variables are used to control for the total links of the match. The variables measure the distance of the match from the representative/average individual for that village cluster. For discrete variables like religion, clan, gender, occupation, soil type, if they are the first of their family to reside in that village and school level, I construct a variable taking value 1 if the match has the modal characteristic of the village and 0 otherwise. For continuous variables like age, wealth and experience in different crops grown I construct two variable each. The first takes the difference between the match's characteristic and the village mean for that characteristic if this difference is positive and zero otherwise and the second takes the difference between the match's characteristic and the village mean for that characteristic if this difference is negative and zero otherwise. The reason for doing this is that the effect might not be symmetric and we allow that. The distribution for these variables might be skewed and then it would matter whether the difference is positive or negative. The variables are presented in **Table 4**.

4 Estimating Network Effects

We first run a simple probit including the endogenous regressor in Table 5. In all of these regressions, we see that the effect of the match's total links if positive and significant.

In Table 6, we present the results of regressing the total links of the match on the difference between the match's characteristics and the representative individual (the variables from Table 4). We see that these variables do have explanatory power, in particular, the crop experience seems to matter in the total links. This makes sense considering all these networks are informational and information is crop based.

Table 7 finally presents the results using the control function approach. Again we see that the endogenous regressor, even after controlling for the endogeneity has a positive and significant impact on the decision to form a link.

References

BALA, V., AND S. GOYAL (2000): "A Noncooperative Model of Network Formation," *Econometrica.*

BANDIERA, O., AND I. RASUL (2006): "Social Networks and Technology Adoption in Northern Mozambique," *The Economics Journal.*

BESLEY, T. J., AND A. CASE (1997): "Diffusion as a learning process: evidence from HYV cotton," *mimeo, Princeton University.*

DE WEERDT, J. (2002): "Risk-Sharing and Endogenous Network Formation," *Discussion Paper 2002/57, WIDER, Helsinki.*

DE WEERDT, J., AND S. DERCON (2006): "Risk-Sharing Networks and Insurance against Illness," *Journal of Development Economics, Elsevier.*

DURLAUF, S. N., AND M. FAFCHAMPS (2005): "Social Capital," *Handbook of Economic Growth.*

FAFCHAMPS, M., AND S. LUND (2003): "Risk-sharing networks in rural Philippines," *Journal of Development Economics.*

FOSTER, A. D., AND M. R. ROSENZWEIG (1995): "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *Journal of Political Economy*, 103(6), 1176–1209.

GOLDSTEIN, M., AND C. UDRY (1999): "Agricultural Innovation and Resource Management in Ghana," .

JACKSON, M. O. (2005): "The Economics of Social Networks," *Proceedings of the 9th World Congress of the Econometric Society, edited by Richard Blundell, Whitney Newey, and Torsten Persson, Cambridge University Press, forthcoming.*

JACKSON, M. O., AND A. WOLINSKY (1996): "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 71(1), 44–74.

KRISHNAN, P., AND E. SCIUBBA (2006): "Links and Architecture in Village Networks," *Birkbeck Working Papers in Economics and Finance 0614.*

Munshi, K. (2004): "Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution," *Journal of Development Economics*, 73(1), 185–213.

Udry, C. R., and T. G. Conley (2004): "Social Networks in Ghana," .

Table 1: Variables Measuring Presence of Link

Variable	Definition	Mean (Std Dev)
Askprob	0-1 variable taking value 1 if respondent would ask match if they had a problem with unhealthy crop	0.3266078 (0.4690316)
Askfert	0-1 variable taking value 1 if respondent would go to match for advice on new fertilizer	0.3029004 (0.4595705)
Askplant	0-1 variable taking value 1 if respondent would go to match to discuss planting method	0.3092055 (0.4622244)
Askbuyer	0-1 variable taking value 1 if respondent would go to match for find a buyer	0.2530272 (0.4348013)

Table 2: Variables Measuring Total Links

Variable	Definition	Mean (Std Dev)
ttprob	Variable representing the number of matches a respondent would ask about a problem with unhealthy crop	3.264313 (2.756746)
ttfert	Variable representing the number of matches a respondent would go to for advice on new fertilizer	3.027743 (2.583175)
ttplant	Variable representing the number of matches a respondent would go to discuss planting method	3.090542 (2.675547)
ttbuyer	Variable representing the number of matches a respondent would go to match for find a buyer	2.526608 (2.755896)

Mttprob	Variable representing the number of matches the 'match' would ask about a problem with unhealthy crop	3.273902 2.761207
Mttfert	Variable representing the number of matches the 'match' would go to for advice on new fertilizer	3.015504 2.58929
Mttplant	Variable representing the number of matches the 'match' would go to discuss planting method	3.093023 2.696501
Mttbuyer	Variable representing the number of matches a respondent would go to match for find a buyer	2.503876 2.755468

Table 3: Variables Measuring Distance between Respondent and Match

Variable	Definition	Mean	Std. Dev.
Shhn	0-1 variable taking value 1 if respondent and match have are from the same household and 0 o.w.	0.0071	0.0842
Shomeregion	0-1 variable taking value 1 if respondent and match have the same homeregion and 0 o.w.	0.7131	0.4524
Slanguage	0-1 variable taking value 1 if respondent and match have the same language and 0 o.w.	0.3844	0.4865
Sfirsthere	0-1 variable taking value 1 if either respondent and match were both first from their families in the village, or both not the first in the village and 0 o.w.	0.65538	0.47533
Sresprel	0-1 variable taking value 1 if respondent and match have the same religion and 0 o.w.	0.2715	0.4448
Ssex	0-1 variable taking value 1 if respondent and match have the same sex and 0 o.w.	0.4985	0.5001
Sschoollevel	0-1 variable taking value 1 if respondent and match have the same level of schooling and 0 o.w.	0.364	0.4812
Sclan	0-1 variable taking value 1 if respondent and match belong to the same clan and 0 o.w.	0.3045	0.4603
Sagyapong	0-1 variable taking value 1 if respondent and match have the same soil type and 0 o.w.	0.103	0.304
Stotwealth	Continuous variable measuring the difference in the wealth of the respondent and match	-379416	2714457
Smaizeyrs	Continuous variable measuring the difference in the years of experience in maize farming of the respondent and match	-1.1964	18.9637

Scassayrs	Continuous variable measuring the difference in the years of experience in cassava farming of the respondent and match	-1.1967	19.0373
Spineyrs	Continuous variable measuring the difference in the years of experience in pineapple farming of the respondent and match	-0.9935	5.8541
Scocoayrs	Continuous variable measuring the difference in the years of experience in cocoa farming of the respondent and match	0.0842	11.3626
Syamyrs	Continuous variable measuring the difference in the years of experience in yam farming of the respondent and match	-0.9118	17.8441
Socc	0-1 variable taking value 1 if respondent and match have the same occupation 0 o.w.	0.5063	0.5
Off	0-1 variable taking value 1 if respondent holds office and 0 o.w.	0.2059	0.4045
Pineyes	0-1 variable taking value 1 if respondent has experience in pineapple farming and 0 o.w.	0.4196	0.4936
Cocoayes	0-1 variable taking value 1 if respondent has experience in cocoa farming and 0 o.w.	0.2529	0.4347
Moff	0-1 variable taking value 1 if match holds office and 0 o.w.	0.2629	0.4403
Mpineyes	0-1 variable taking value 1 if match has experience in pineapple farming and 0 o.w.	0.5132	0.4999
Mcocoayes	0-1 variable taking value 1 if match has experience in cocoa farming and 0 o.w.	0.2522	0.4343

Table 4: Variables Measuring Distance of Match from Average Respondent

Variable	Definition	Mean	Std. Dev.
Mdmoderesprel	0-1 variable taking value 1 if match has the modal religion of his village and 0 o.w.	0.4304	0.49522
Mdmodeschool	0-1 variable taking value 1 if match has the modal education of his village and 0 o.w.	0.53436	0.49891
Mdmodeocc1	0-1 variable taking value 1 if match has the modal occupation of his village and 0 o.w.	0.75798	0.42838
Mdmodeclan	0-1 variable taking value 1 if match belongs to the modal clan of his village and 0 o.w.	0.47786	0.4996
Mdmodehomeregion	0-1 variable taking value 1 if match has the modal home region of his village and 0 o.w.	0.8548	0.35237
Mdpmeantotwealth	Difference in wealth between match and average for the village if match has more wealth than average, 0 otherwise	327144	330751
Mdnmeantotwealth	Difference in wealth between match and average for the village if match has less wealth than average, 0 otherwise	697260	1995031
Mdpmeanage	Difference in age between match and average for the village if match has more age than average, 0 otherwise	4.98393	5.92479
Mdnmeanage	Difference in age between match and average for the village if match has less age than average, 0 otherwise	5.2116	8.30507
Mdpmeanmaizeyrs	Difference in experience with maize (in years) between match and average for the village if match has more experience than average, 0 otherwise	4.60295	6.63978
Mdnmeanmaizeyrs	Difference in experience with maize (in years) between match and average for the village if match has less experience than average, 0 otherwise	5.91893	9.32316

Mdpmeancassayrs	Difference in experience with cassava (in years) between match and average for the village if match has more experience than average, 0 otherwise	4.61727	6.68501
Mdnmeancassayrs	Difference in experience with cassava (in years) between match and average for the village if match has less experience than average, 0 otherwise	5.93284	9.34766
Mdpmeanpineyrs	Difference in experience with pineapple (in years) between match and average for the village if match has more experience than average, 0 otherwise	1.04005	1.05353
Mdnmeanpineyrs	Difference in experience with pineapple (in years) between match and average for the village if match has less experience than average, 0 otherwise	2.03952	3.96562
Mdpmeancocoayrs	Difference in experience with cocoa (in years) between match and average for the village if match has more experience than average, 0 otherwise	2.35537	1.56624
Mdnmeancocoayrs	Difference in experience with cocoa (in years) between match and average for the village if match has less experience than average, 0 otherwise	2.36621	6.65376
Mdpmeanyamyrs	Difference in experience with yam (in years) between match and average for the village if match has more experience than average, 0 otherwise	3.82191	4.69297
Mdnmeanyamyrs	Difference in experience with yam (in years) between match and average for the village if match has less experience than average, 0 otherwise	4.69776	10.4069

Table 5: Probit Results

	Askprob	Askfert	Askplant	Askbuyer
Mttprob	0.064			
	(3.38)**			
Mttfert		0.077		
		(3.72)**		
Mttplant			0.081	
			(4.17)**	
Mttbuyer				0.105
				(6.21)**
Shhn	1.436	1.397	1.452	1.093
	(3.00)**	(3.00)**	(3.09)**	(2.61)**
Shomeregion	0.091	0.116	0.071	0.085
	(0.68)	(0.85)	(0.53)	(0.61)
Ssex	0.236	0.078	0.123	0.287
	(2.30)*	(0.76)	(1.20)	(2.74)**
Sclan	0.166	0.153	0.159	0.22
	(1.49)	(1.37)	(1.45)	(1.99)*
off	-0.67	-0.317	-0.443	-0.392
	(4.68)**	(2.24)*	(3.15)**	(2.80)**
Moff	0.334	0.073	0.189	0.256
	(2.44)*	(0.53)	(1.38)	(1.79)
Sfirsthere	0.233	0.296	0.284	0.247
	(1.96)*	(2.45)*	(2.37)*	(2.01)*
Sresprel	0.028	-0.024	-0.117	-0.083
	(0.25)	(0.21)	(1.04)	(0.72)
Sschoollevel	-0.065	-0.015	-0.026	0.078
	(0.60)	(0.14)	(0.24)	(0.69)
Sage	0.006	-0.001	0.005	0.01
	(1.14)	(0.22)	(1.03)	(1.90)
Smaizeyrs	0.021	0.019	0.021	0.037
	(0.96)	(0.86)	(0.94)	(1.61)
Scassayrs	-0.02	-0.017	-0.021	-0.029
	(0.93)	(0.79)	(0.98)	(1.29)
Spineyrs	-0.043	-0.071	-0.048	-0.01
	(4.88)**	(7.66)**	(5.40)**	(1.14)
Scocoayrs	-0.007	0.011	0.006	0.006
	(1.71)	(2.55)*	(1.59)	(1.38)
Syamyrs	-0.011	-0.009	-0.01	-0.019
	(3.39)**	(2.86)**	(3.30)**	(5.66)**
Socc	0.156	0.227	0.134	0.251
	(1.27)	(1.78)	(1.08)	(1.92)
Constant	-1.201	-1.376	-1.232	-1.552
	(7.23)**	(7.97)**	(7.40)**	(9.09)**
Observations	781	781	781	781

Absolute value of t statistics in brackets.
* significant at 5%; ** significant at 1%

Table 6: First Stage OLS Regressions For Total Links of Match

	Mttprob	Mttfert	Mttplant	Mttbuyer
Mdmoderesprel	0.374	0.429	0.335	-0.128
	(0.369)	(0.345)	(0.356)	(0.346)
Mdmodeschool_level	0.352	0.163	0.172	0.299
	(0.370)	(0.346)	(0.358)	(0.347)
Mdmodeocc1	0.537	0.481	0.417	0.61
	(0.425)	(0.398)	(0.411)	(0.398)
Mdmodeclan	0.476	0.178	0.304	0.357
	(0.354)	(0.332)	(0.342)	(0.332)
Mdmodehomeregion	0.383	0.227	0.272	0.501
	(0.552)	(0.516)	(0.533)	(0.517)
Mdpmeanage	-0.008	0.031	0.025	-0.006
	(0.042)	(0.039)	(0.041)	(0.039)
Mdnmeanage	-0.021	-0.012	-0.01	0
	(0.034)	(0.032)	(0.033)	(0.032)
Mdpmeanmaizeyrs	0.01	0.012	0.01	0.003
	(0.144)	(0.135)	(0.140)	(0.135)
Mdnmeanmaizeyrs	-6.059**	-7.028***	-6.496***	-12.107***
	(2.447)	(2.288)	(2.364)	(2.291)
Mdpmeancassayrs	-0.004	-0.02	-0.009	0.006
	(0.142)	(0.133)	(0.138)	(0.133)
Mdnmeancassayrs	6.071**	7.017***	6.506***	12.091***
	(2.447)	(2.289)	(2.364)	(2.291)
Mdpmeanpineyrs	0.387*	0.397*	0.454**	0.446**
	(0.227)	(0.212)	(0.219)	(0.212)
Mdnmeanpineyrs	-0.015	-0.022	-0.013	0.004
	(0.067)	(0.063)	(0.065)	(0.063)
Mdpmeancocoayrs	-0.386***	-0.375***	-0.394***	-0.361***
	(0.144)	(0.135)	(0.139)	(0.135)
Mdnmeancocoayrs	-0.008	0.023	0.019	0.011
	(0.030)	(0.028)	(0.029)	(0.028)
Mdpmeanyamyrs	0.094**	0.119***	0.106**	0.120***
	(0.043)	(0.040)	(0.041)	(0.040)
Mdnmeanyamyrs	-0.03	-0.006	-0.022	-0.012
	(0.023)	(0.022)	(0.023)	(0.022)
Constant	2.394***	2.101***	2.241***	1.366*
	(0.853)	(0.798)	(0.824)	(0.799)
Observations	268	268	268	268
R-squared	0.16	0.2	0.18	0.25

Absolute value of t statistics in brackets.
* significant at 5%; ** significant at 1%

Table 7: Two Step IV Probit Results

	Askprob	Askfert	Askplant	Askbuyer
Mttprob	0.213***			
	(0.056)			
Mttfert		0.190***		
		(0.054)		
Mttplant			0.222***	
			(0.055)	
Mttbuyer				0.258***
				(0.045)
Shhn	0.883*	0.966*	0.924*	0.747*
	(0.517)	(0.494)	(0.507)	(0.451)
Shomeregion	0.025	0.066	0.01	0.038
	(0.147)	(0.146)	(0.147)	(0.154)
Ssex	0.231**	0.068	0.112	0.299***
	(0.111)	(0.110)	(0.111)	(0.115)
Sclan	0.142	0.145	0.137	0.204*
	(0.120)	(0.117)	(0.118)	(0.121)
off	-0.616***	-0.305**	-0.397***	-0.386**
	(0.153)	(0.150)	(0.152)	(0.152)
Moff	0.255*	0.017	0.095	0.067
	(0.148)	(0.146)	(0.148)	(0.165)
Sfirsthere	0.252*	0.304**	0.297**	0.310**
	(0.129)	(0.128)	(0.129)	(0.136)
Sresprel	0.013	-0.03	-0.131	-0.085
	(0.121)	(0.120)	(0.122)	(0.126)
Sschool_level	-0.06	0.001	-0.025	0.074
	(0.118)	(0.117)	(0.118)	(0.124)
Sage	0.005	-0.002	0.004	0.011*
	(0.005)	(0.005)	(0.005)	(0.006)
Smaizeyrs	-0.007	-0.005	-0.003	0.01
	(0.026)	(0.026)	(0.026)	(0.028)
Scassayrs	0.007	0.006	0.003	-0.003
	(0.026)	(0.026)	(0.026)	(0.027)
Spineyrs	-0.048***	-0.077***	-0.055***	-0.023**
	(0.010)	(0.010)	(0.010)	(0.010)
Scocoayrs	-0.004	0.014***	0.010**	0.007
	(0.005)	(0.005)	(0.005)	(0.005)
Syamyrs	-0.014***	-0.011***	-0.013***	-0.020***
	(0.004)	(0.003)	(0.003)	(0.004)
Socc	0.089	0.157	0.075	0.172
	(0.134)	(0.135)	(0.134)	(0.143)
Constant	-1.058***	-1.210***	-1.053***	-1.521***
	(0.175)	(0.178)	(0.177)	(0.191)
Observations	769	769	769	769

Absolute value of t statistics in brackets.
* significant at 5%; ** significant at 1%