The Economics of Exogamous Marriage in Small-Scale Societies

Gregory K. Dow, Clyde G. Reed, and Simon Woodcock

Department of Economics

Simon Fraser University

September 2015

Abstract. Marriage is a core institution in almost every human society, including small-scale societies based on foraging or subsistence agriculture. A crucial dimension of the marriage systems in such societies involves endogamy and exogamy: that is, the choice of a marriage partner from within one's own community or from an outside community. We develop a model in which the exogamy rate is higher when good local matches are scarce due to small community sizes, and when productivity differs across communities due to environmental shocks**.** These theoretical predictions are supported by econometric analysis of data from the Standard Cross Cultural Sample.

JEL Codes. D13, J11, N30, Z13.

Acknowledgements. We are grateful for research assistance from Colleen Cunningham, Ola Heldal, Leanna Mitchell, and Scott Skjei. We thank Chris Muris and two anonymous referees for helpful comments and suggestions. The Social Sciences and Humanities

Research Council of Canada and the Human Evolutionary Studies Program at Simon Fraser University provided financial support. All opinions are those of the authors.

1. Introduction

Until about 13,000 years ago, all humans lived in hunter-gatherer societies (Dow and Reed, 2011). An adequate understanding of human biological and cultural evolution requires an understanding of marriage patterns in such societies. For example, the degree of genetic relatedness within early human groups depended on the extent to which people preferred local partners. Even after the Neolithic shift to agriculture (Dow et al., 2009), for thousands of years most people still lived in small villages. Kinship was the principal institutional glue that held these societies together (Johnson and Earle, 2000; Diamond, 2012). Thus marriage systems had important implications for risk sharing, public goods supply, human capital investments, warfare, trade, and inequality.

In the extensive literature on the economics of marriage, little attention has been given to understanding endogamy (in which marriage partners come from the same local community) and exogamy (in which the partners come from different communities). The anthropological literature does address this issue (Parkin, 1997; Stone, 2001) and reveals that small-scale foraging and agricultural societies vary considerably on this dimension. Here we show that much of this variation can be explained by economic factors.

In our approach, there are two advantages of exogamy. First, households in which spouses come from different communities are more mobile. One reason is that spouses can share knowledge about the natural environment at two different locations. Another is that they have kin in two places and can expect to be welcome in each of two communities (see the discussion of 'social boundary defense' in Kelly, 2007: 193-197).

Environmental shocks that generate productivity differences across communities tend to induce the formation of mobile households.

A second benefit from exogamy is that the pool of potential marriage partners is larger. When agents are heterogeneous and groups are small, better matches can often be achieved by searching outside one's home community. This effect is less important when groups become large because then there are more opportunities for good local matches.

Exogamy also has costs. The most obvious is that a marriage between partners from different communities requires at least one partner to live at a distance from family and friends. This is inconvenient and may weaken that partner's bargaining power within the household. Individuals unfamiliar with local conditions may also face disadvantages with respect to food production or childcare. We do not model these costs explicitly but we do take them into account by assuming that when other things are equal, the lowest equilibrium rate of exogamy is adopted.

Our formal model involves two symmetric communities subject to productivity shocks. Households with a partner from each community are mobile and can produce food at either site. Households with two partners from the same community can only produce food at their site of origin. All agents are risk-neutral so exogamy is not a form of insurance. Rather, natural shocks create income differences across communities that can be exploited by the formation of mobile households. In equilibrium we require that such households cannot gain by changing locations.

We also allow random variation in agent qualities. For a given food income, households in which both adults have 'good' parenting skills can produce more or better adult offspring than households in which one or both of the adults have 'bad' parenting skills. Agents are ultimately interested in the quantity or quality of the offspring they produce, and for a given match quality seek the highest possible food income in order to achieve these goals. In matching equilibrium, no pair of agents can gain by deviating to form a new household, given the prevailing food incomes at the two communities. Our comparative static analysis shows that the expected exogamy rate is an increasing function of the productivity gap between communities and a decreasing function of community size.

We test these empirical predictions using data from the Standard Cross Cultural Sample (SCCS), a representative sample of 186 well-documented and culturally independent pre-modern societies. OLS estimates indicate that both community size and productivity variation are significant determinants of exogamy.

Our theoretical model treats community size as exogenous. However, there may be factors outside our model, such as technology or cultural institutions like religion, that jointly determine community size and the exogamy rate. We estimate IV specifications that address this possibility, and obtain results very similar to our baseline regressions. Our estimates are robust to a variety of alternative specifications, instruments, and sample restrictions.

Our theoretical model assumes monogamous marriage, which in many societies was never the norm. While this may limit the generality of the model, we have no a priori reason to expect our theoretical model's predictions not to hold for polygamous societies. In our estimation sample, only about 10 percent of societies are strictly monogamous. However, nearly 70 percent are predominantly monogamous and have an incidence of polygamy below 20 percent. Our empirical estimates are robust to the inclusion or exclusion of societies with a high (above 20 percent) incidence of polygamy.

Our model also assumes symmetry across genders. In pre-modern societies, it is more common for married women to move away from their birth communities than for married men to do so. However, in our sample this asymmetry is not extreme: 51% of the societies are patrilocal (couples reside with the husband's kin), 30% are matrilocal (couples reside with the wife's kin), and the remaining 19% practice some other form of marital location. Baker and Jacobsen (2007) use a bargaining framework to argue that patrilocality should be more common when the husband's human capital is relatively location-specific compared to the wife's, and conversely for matrilocality. In a sample related to ours, they find that the existence of fixed postmarital residence rules is weakly responsive to a set of environmental, technological, and economic variables.

The main alternative explanation for exogamy in the economics literature is risk aversion and the use of marriage relationships for consumption smoothing (Rosenzweig, 1988a, 1988b; Rosenzweig and Stark, 1989). Using longitudinal data from surveys of up to 400 households in six agricultural villages in rural India during 1975/76 - 1984/85, Rosenzweig and Stark (1989) report that the variance of rainfall is positively correlated with the variance of household farm profit, and that the variance of household profit is positively correlated with the variance of household food consumption.

Exogamous marriage provides one way of smoothing household consumption. In this society daughters generally move away from their home village to join their spouse's household. This facilitates income transfers across villages when rainfall shocks occur, which gives some insurance. The covariance of rainfall across villages declines with the distance between villages. Rosenzweig and Stark (1989) find that higher profit variance for a household is associated with greater distance between the home villages of the two marital partners, presumably in order to generate better insurance. This effect is smaller for wealthy households that can more readily self-insure.

Our data set differs from that of Rosenzweig and Stark in several ways. First, our units of observation are communities rather than households. Our dependent variable is the fraction of the marriages in the community that involve a partner from outside the community. In Rosenzweig and Stark's sample there is almost universal exogamy in our sense, because almost every marriage in a village involves one partner from outside that village. Their dependent variable is the distance between the home villages for partners within a household, which can be taken as a measure of the degree of exogamy within the household. Another difference is that our data set, while cross-sectional, spans a large set of pre-modern societies from around the world. Some of these societies rely on foraging rather than agriculture, and some have access to a cash economy while others do not.

Both our theory and that of Rosenzweig and Stark assume that rainfall variation leads to differences in productivity across communities. Both also predict that greater rainfall variation should lead to more exogamy (whether measured at the household or community level). Similar correlations are reported in the anthropological literature. Kelly (2007: 274-275), for example, notes that bilocal residence, a system in which couples can move back and forth between their parental families, is more common in natural environments with greater uncertainty.

Our theory differs from that of Rosenzweig and Stark (1989) by assuming risk neutrality. In our framework, exogamous households are more mobile and better able to exploit differences in production opportunities across communities through migration. When these differences are larger and more frequent, exogamy should be more common. We are unable to test Rosenzweig and Stark's alternative explanation both because our observations are on communities rather than households, and because our data set does not include information about the distances between communities or transfers of food or funds between communities.

Although anthropologists have written extensively on exogamy, this literature advances only a few explanations for its existence in small-scale societies (see Harris, 1997, 253-7). The hypothesized benefits of exogamy include income gains due to increased household mobility; achievement of gender balance; avoidance of incest; avoidance of military conflict through political alliances; and exploitation of comparative advantage through trade. The role of household mobility is central to our model and will be discussed extensively below. Here we ignore the issue of gender imbalance, which is unlikely to be important except in very small groups. The remaining factors will be considered in sequence.

Perhaps the most popular hypothesis about exogamy involves incest avoidance, motivated by the potential for negative health outcomes for the children of genetically related parents. Recent research has found that, in many cases, the size of the health cost associated with these marriages is quite modest (Bennett et al., 2002; Joseph, 2007; Callaway, 2008). There is even evidence of reproductive benefits from marriage between genetic relatives, at least up to a point. An analysis of Icelandic genealogical data from 1800 to 1965 found that couples who shared a great-great-grandparent had more children and grandchildren than more distantly related couples (Helgason et al., 2008). Research on consanguineous marriages in South Asia suggests a benefit in terms of reduced agency costs within families, and helps to explain why such marriages are often preferred despite some evidence of increased child mortality (Do et al., 2013).

Repeated intermarriage within a small group of closely related people might be expected to yield poor health consequences, but highly endogamous societies often have institutionalized constraints that reduce this risk. An illustrative case involves the Kel Ahaggar, semi-nomadic pastoralists in southern Algeria. The vast majority of marriages are endogamous within descent groups so it is "virtually impossible" for the Kel Ahaggar to marry someone who is not a cousin of some kind. The preferred union for a son is with the mother's brother's daughter. Nevertheless, there are numerous prohibitions: a man is forbidden from marrying his mother, his father's and mother's uterine sisters, his uterine sister, his daughter, the daughters of his uterine brothers and sisters, and his mother-in-law, sister-in-law, and daughter-in-law (Keenan, 1977a, 1977b).

If there are large health costs associated with marriage between close relatives, the most direct solution would be to prohibit such marriages, as in the example from the preceding paragraph. This is likely to be a lower cost solution than a requirement that marriages be exogamous with respect to the community as a whole. To the extent that such prohibitions strongly decrease the size of the matching pool within a small local community, and this leads to exogamy, our empirical work will pick up this effect via our community size variable.

To address possible concerns that exogamy may be motivated by incest avoidance in our data, we undertake several robustness checks. We find no evidence of a systematic relationship between exogamy and prohibitions on marriage between close cousins.

Moreover, when we add controls for prohibitions on marriage between close cousins to our main empirical specification, they have no statistically significant effect on the exogamy rate, while the coefficients on our measures of community size and productivity variation remain unchanged.

Proposed relationships in the anthropological literature involving warfare and trade are primarily about the effects of exogamy rather than its causes, although if exogamy has beneficial effects this may help to explain why communities adopt it. We believe the anthropological literature has not fully come to grips with the difficult causality problems associated with empirical work along these lines. Does the practice of exogamy decrease the probability of warfare and/or improve the prospects for trade? Answers to these questions require a credibly exogenous source of variation in the exogamy rate and adequate controls for the other determinants of warfare and trade. This is not the approach that has been taken in the existing literature. Overall, we are skeptical about the potential for empirical research along these lines, given the general absence of critical control variables in data sets like the SCCS.

With regard to warfare, various authors have argued that exogamy promotes peaceful co-existence by creating divided loyalties within and across social groups; reducing competition for marriage partners within groups; and building solidarity across groups (see Kang, 1979, 85-6). However, studies by Berndt (1964), Hayano (1973), and Kang (1976, 1979) all fail to establish a positive correlation between exogamy and the absence of conflict. This is unsurprising because these studies do not control for the level of threat, do not consider other potential responses to threats, and do not consider the possibility that exogamy could promote warfare through the formation of alliances that enhance offensive capabilities.

We are not aware of empirical studies that attempt to establish a relationship between exogamy and trade. But as with warfare, inferring a causal link would be problematic in the absence of data on other relevant variables. These include the size of the potential gains from trade as well as alternative mechanisms for reducing transaction costs, such as bond posting or efficiency wages (for examples, see Greif, 2006a, 2006b).

Anthropologists would define marriage institutions as part of 'culture'. Our paper can then be regarded as exploring one facet of a larger question: what determines culture? Some anthropologists (Johnson and Earle, 2000: 16-22) argue that at least in the long run, culture tends to adapt to population, environment, and technology. Our paper fits into this tradition by arguing that exogamy (a culture trait) adapts to population (in the sense of community size) and the environment (in the sense that nature affects productivity).

One culture trait (e.g., exogamy) could be influenced by other culture traits (e.g., religion). Such ideas are common in the anthropological literature (Parkin, 1997; Stone, 2001). Exogamy might also be influenced by technological variables that we do not observe. Our IV strategy deals with the possibility that such variables could influence exogamy both directly and through community size. We also control for the technology of food production, the presence of food storage, sedentism, and whether the territory and population of a society are expanding or shrinking (a possible indication that culture or institutions do not reflect long run equilibrium conditions). We also take into account issues of social structure; specifically, whether a society has a class system or slavery.

We lack good data on religious beliefs that might be related to community size or exogamy. In general, we are reluctant to include a large number of cultural variables as potential determinants of exogamy, both because we lack strong guidance from theory about how one culture trait influences others, and we lack confidence about the aspects of culture (if any) that can safely be regarded as exogenous. In the conclusion, we return to the question of whether economic variables are important determinants of culture.

The rest of the paper is organized as follows. Section 2 develops a model of two communities subject to productivity shocks. This section defines locational equilibrium and derives a labor allocation for each fixed exogamy level. Section 3 introduces agent heterogeneity and studies household formation. Here we define a matching equilibrium and characterize the exogamy levels that are supportable in equilibrium. Assuming that the lowest supportable exogamy level will be observed, section 4 shows that exogamy is increasing in the productivity gap across communities and decreasing in community size.

Section 5 describes the Standard Cross Cultural Sample, defines variables, and presents our empirical methodology. Section 6 presents our econometric results. Section 7 summarizes the paper and discusses possible extensions. Proofs of formal propositions are in an appendix available from the authors.

2. Locational Equilibrium

Consider a production site with population $n > 0$ and one unit of land. Each agent is endowed with a unit of work time and maximizes food income. We ignore leisure. All agents are risk-neutral and the population is divided equally between males and females. Food is shared equally among the agents who work at a given site.

In our empirical work, we study communities ranging in size from a few dozen to several thousand people. We assume that in this population range an individual household ignores its own effect on the average product of labor for the community. Thus an agent or household can be treated as having negligible size relative to the population at a site.

The production function for food (measured in calories) has the form

$$Y = \theta f(L) \tag{1}$$

where $L \geq 0$ is labor and $\theta > 0$ is the natural productivity of the site.

A1 The production function satisfies $f(0) = 0$; $f'(L) > 0$ for $L > 0$ with $f'(0) = \infty$ and $f'(\infty) = 0$; and $f''(L) < 0$ for $L > 0$.

There are two symmetric sites $s = 1, 2$ with this technology. Households consist of two agents, one of each sex. The aggregate population of the two sites is $2n$ so there are n households. Let $m \in [0, n]$ be the number of households that have one partner from each site. This leaves $(n-m)/2$ local matches where both partners are from site 1, and an equal number of local matches where both partners are from site 2. Households that have local matches are immobile because neither partner is well informed about the natural or social environment at the other site, so the household cannot produce food at that site. Mixed households are mobile and can produce food at either site.

A2 The productivity state is $\underline{\theta} \equiv (\theta_1, \theta_2)$ where $\theta_s \in \{\theta_A, \theta_B\}$ for $s = 1, 2$ with $0 < \theta_B \leq \theta_A$. Each of the four states {AA, AB, BA, BB} has positive probability.

The sequence of events runs as follows. First, the productivity state is revealed. Agents form households after seeing this state, and then mobile households decide where to locate. In this section we take the number of mobile households as given and consider their locational decisions. Household formation is examined in section 3.

Food income per capita at each site will be denoted by $y_1 \equiv \theta_1 f(L_1)/L_1$ and $y_2 \equiv \theta_2 f(L_2)/L_2$ where $L_1 + L_2 = 2n$ is the aggregate labor endowment. The labor allocation $(L_1, L_2) \geq 0$ must satisfy the following requirement.

<u>Locational equilibrium</u>. Choose any $m \in [0, n]$. The allocation (L_1, L_2) is a locational equilibrium for m if no mobile household can increase its income by changing sites.

To see the implications of this condition, let $(m_1, m_2) \geq 0$ be the distribution of mobile households across sites, where $m_1 + m_2 = m$. First suppose that productivity is equal across sites ($\theta_1 = \theta_2 = \theta$). For $m = 0$ we have $L_1 = L_2 = n$ because all matches are local, there are n/2 households at each site, and each household has two units of labor. This gives $y_1 = y_2 = \theta f(n)/n$. For $m > 0$ it can be shown from A1 that equilibrium requires $y_1 = y_2$. This is only possible when mobile households are divided equally across sites so that $L_1 = L_2 = n$, $m_1 = m_2 = m/2$, and $y_1 = y_2 = \theta f(n)/n$. Mobile households are indifferent between sites when incomes are equal, so this is a locational equilibrium.

Next, suppose productivity is unequal across sites so that one site has θ_A and the other has $\theta_B < \theta_A$. In this context it is convenient simply to refer to the sites as A and B. Define m* to be the number of mobile households required to equalize incomes across sites when all of the mobile households are at site A:

(2) $$\theta_A f(n + m^*)/(n + m^*) \equiv \theta_B f(n - m^*)/(n - m^*)$$

A1 guarantees that $m^* \in (0, n)$ is unique. It is easy to verify that m^* is increasing in the site productivity ratio θ_A/θ_B with $m^* = 0$ when $\theta_A/\theta_B = 1$ and $m^* \to n$ as $\theta_A/\theta_B \to \infty$. We assume $\theta_A/\theta_B > 1$ and therefore $m^* > 0$ in what follows.

It can be shown in this case that equilibrium requires $y_A \geq y_B$. Now consider how the income ratio y_A/y_B varies with m. If $m = 0$ then $L_A = L_B = n$. This implies that $y_A = \theta_A f(n)/n > \theta_B f(n)/n = y_B$ and thus $\theta_A/\theta_B = y_A/y_B > 1$. Therefore when there are no mobile households, the productivity ratio equals the income ratio. When $m \in (0, m^*)$, there are not enough mobile households to drive y_A down to y_B even when all mobile households go to site A. As a result the unique locational equilibrium is $L_A = n + m$ and $L_B = n - m$. This yields $\theta_A/\theta_B > y_A/y_B > 1$ where y_A/y_B is decreasing in m.

For higher exogamy levels $m \in [m^*, n]$, there are enough mobile households to equalize income across sites. This is achieved by having $(m^* + m)/2$ mobile households at site A, and the rest at site B. As a result the unique locational equilibrium is $L_A = n + m^*$ and $L_B = n - m^*$. This yields $\theta_A/\theta_B > y_A/y_B = 1$.

We summarize these results as follows.

Lemma 1 (locational equilibrium).

(a) If productivities are equal ($\theta_1 = \theta_2 = \theta$) then for every $m \in [0, n]$ we have $L_1 = L_2 = n$; $m_1 = m_2 = m/2$; and $y_1 = y_2 = \theta f(n)/n$.

(b) If productivities are unequal ($\theta_A > \theta_B$) then

  (i) $m \in [0, m^*)$ implies $L_A = n + m$ and $L_B = n - m$ with $y_A > y_B$.

(ii) $m \in [m^*, n]$ implies $L_A = n + m^*$ and $L_B = n - m^*$ with $y_A = y_B$.

3. Matching Equilibrium

Exogamy may partly be a response to productivity differences across sites, but it may also lead to improved matches through a larger pool of potential marriage partners. To model this idea, we introduce heterogeneity among agents.

A3 The number of surviving adult children for a household is $2\gamma y$ where $2y$ is the total food income of the household and $\gamma > 0$ is the rate at which the household converts food into adult offspring. The type of a household is $\gamma = \min \{\gamma^F, \gamma^M\}$ where $\gamma^F \in \{\gamma^b, \gamma^g\}$ is the type of the female partner and $\gamma^M \in \{\gamma^b, \gamma^g\}$ is the type of the male partner, with $0 < \gamma^b < \gamma^g$.

We will call agents with γ^b the 'bad' type and those with γ^g the 'good' type. The general idea is that 'good' agents have better parenting skills than 'bad' agents. In some societies, such skills may have a direct effect on the survival prospects of children and therefore affect the quantity of adult offspring. In other societies, these skills may affect the health or skills of children after they reach adulthood, and thus influence the quality of adult offspring. There is no difference between good and bad agents with respect to food production. In our model the parenting skills of mothers and fathers are perfect complements, so both parents must be 'good' in order to obtain any quantity or quality benefits with respect to children. It does not matter for our purposes whether the source of this heterogeneity is genetic or involves random life experiences, such as having good adult role models in childhood.

The simplest interpretation of A3 is that individual agents maximize the quantity of their surviving adult offspring, and (for a given match quality) maximize their food income as a means to this end. However, A3 can also be interpreted as a statement about the quality of children. In this case parental skill (γ) is multiplied by parental income ($2y$) to obtain child quality ($2\gamma y$). One can devise extensions in which parents care about both the quantity and quality of their children, but we do not pursue this issue here.

The agent types are determined by nature before households are formed, and are publicly observed. The number of good agents in each site/gender category is given by the quality state $\underline{n} \equiv (n_1^F, n_1^M, n_2^F, n_2^M) \geq 0$ where the subscripts denote the initial site of the agent and superscripts denote gender. No entry in this vector can exceed $n/2$, which is the total number of agents in each site/gender category. The probability distribution over quality states ($\underline{n}$) is independent of the distribution over productivity states ($\underline{\theta}$).

A household allocation is a vector $\underline{h} = (h_1^g, h_{12}^g, h_1^b, h_{12}^b;\ h_2^g, h_{21}^g, h_2^b, h_{21}^b) \geq 0$ where superscripts indicate household quality (b = bad, g = good). A single subscript refers to households involving local matches and says where the households are located. A double subscript refers to mobile households, where the first digit indicates where the households engage in production (site 1 or 2).

The set of feasible household allocations depends on the state $\underline{n}$. For example, we cannot have a positive number of good local matches at site 1 ($h_1^g > 0$) if all of the agents from that site are bad ($n_1^F = n_1^M = 0$). The notation $\underline{h} \in H(\underline{n})$ indicates that $\underline{h}$ is physically feasible when the quality state is $\underline{n}$.

A household allocation determines the labor allocation as follows:

$$L_1 = 2(h_1^g + h_{12}^g + h_1^b + h_{12}^b) \qquad \text{(total labor at site 1)} \tag{3}$$

$$L_2 = 2(h_2^g + h_{21}^g + h_2^b + h_{21}^b) \qquad \text{(total labor at site 2)}$$

A household allocation also determines the level of exogamy:

$$(4) \qquad m = h_{12}^g + h_{12}^b + h_{21}^g + h_{21}^b \qquad \text{(total mobile matches)}$$

$$(n\text{-}m)/2 = h_1^g + h_1^b \qquad \text{(total local matches at site 1)}$$

$$(n\text{-}m)/2 = h_2^g + h_2^b \qquad \text{(total local matches at site 2)}$$

We write $m = M(\underline{h})$ for the level of exogamy induced by $\underline{h}$ through (4).

The sequence of events remains as in section 2. First the productivity state $\underline{\theta}$ and quality state $\underline{n}$ are revealed. Agents then form households, and finally mobile households choose sites. At the last step all of the mobile households choose site A if $y_A > y_B$ and are indifferent between sites if $y_1 = y_2$. This is true regardless of the qualities of individual matches. Therefore we continue to require locational equilibrium and Lemma 1 applies.

Next consider a given allocation $\underline{h} \in H(\underline{n})$ and the reasons why an agent might want to deviate from it. First, the agent may want to change partners in order to gain a higher income. For example, if $y_A > y_B$ an agent in a local match at site B would prefer a mobile match with an agent from site A, holding quality constant. Second, the agent may want to change partners in order to raise household quality even though income does not change. For example, a good agent in a bad local match prefers a good partner from the same site. Because an agent cannot form a new household without a willing partner, we require that any deviation make both participants strictly better off.

This motivates the following series of definitions.

Immunity to deviation. For a given quality state $\underline{n}$ and given incomes (y_1, y_2), we say that $\underline{h} \in H(\underline{n})$ is immune to deviation if no pair of agents of opposite sex both become strictly better off by leaving their current partners and forming a new household.

Because our model describes the formation of two-person coalitions, the usual Nash idea that no individual can gain by deviating does not apply. Instead we require that no pair of agents (of different genders) can gain by deviating. Notice that immunity from deviation does not require that $\underline{h}$ actually generate the given incomes (y_1, y_2), which are regarded as parametric by pairs of agents. We add this further requirement to obtain our definition of matching equilibrium.

Matching equilibrium. For a given productivity state $\underline{\theta}$ and quality state $\underline{n}$, we say that $\underline{h} \in H(\underline{n})$ is a matching equilibrium if at the incomes (y_1, y_2) generated by the locational equilibrium for $m = M(\underline{h})$, the allocation $\underline{h}$ is immune to deviation.

Supportable exogamy levels. For a given productivity state $\underline{\theta}$ and quality state $\underline{n}$, we say that $m \in [0, n]$ is supportable if there is some $\underline{h} \in H(\underline{n})$ with $m = M(\underline{h})$ that is a matching equilibrium.

Lemmas 2 and 3 below characterize household allocations that are immune to deviation for given incomes (y_1, y_2). Propositions 1 and 2 determine supportable exogamy levels.

Let $n^F \equiv n_1^F + n_2^F$ and $n^M \equiv n_1^M + n_2^M$ be the aggregate supplies of good females and good males. The maximum number of good matches is

$$(5) \qquad h_{max}^g \equiv \min \{n^F, n^M\}$$

If $n^F \leq n^M$ we will say that good females are globally scarce and conversely for $n^M \leq n^F$. This theoretical maximum is closely linked to the concept of immunity to deviation.

Lemma 2 (equal incomes across sites).

Let $y_1 = y_2$ and fix some quality state $\underline{n}$ such that $n^F \leq n^M$.

A household allocation $\underline{h} \in H(\underline{n})$ is immune to deviation iff it achieves $h_{max}{}^g$.

For a given exogamy level $m \in [0, n]$, there is some $\underline{h} \in H(\underline{n})$ that is immune to deviation and yields $m = M(\underline{h})$ iff $m \in [\mu_{min}(\underline{n}), \mu_{max}(\underline{n})]$. This interval is non-empty with

$$\mu_{min}(\underline{n}) \quad = \quad 2 \max \{0; n_1{}^F - n_1{}^M; n_2{}^F - n_2{}^M\}$$

$$\mu_{max}(\underline{n}) \quad = \quad n - 2 \max \{0; n_1{}^F - n_2{}^M; n_2{}^F - n_1{}^M\}$$

At most one of the arguments of the max operator can be strictly positive in each case. When $n^F \geq n^M$, the roles of F and M in Lemma 2 are reversed.

Efficient matching must occur because otherwise there would be a good female in a bad match and a good male in a bad match. All matches yield equal incomes, so these two agents could both gain by leaving their current partners and creating a new household of higher quality. If good females are globally scarce ($n^F \leq n^M$), the minimum exogamy level $\mu_{min}(\underline{n})$ depends on whether there is a site where good females are locally abundant ($n_1{}^F > n_1{}^M$ or $n_2{}^F > n_2{}^M$). If neither inequality holds (good females are scarce at both sites), then $h_{max}{}^g$ can be achieved through local matching alone and $m = 0$ is feasible. If one of these two inequalities holds, some exogamy is needed in order to match surplus good females from one site with surplus good males from the other. The factor 2 arises in Lemma 2 because there must be equal flows of males and females in each direction to maintain gender balance. The case of maximum exogamy $\mu_{max}(\underline{n})$ is similar except that now cross-site pairings are relevant. If $n_1^F \leq n_2^M$ and $n_2^F \leq n_1^M$ then all good females can be assigned to good partners through exogamous matching and m = n is feasible. If not, some local matching is needed.

When incomes are unequal across sites, agents may have to consider trade-offs between a higher income and a better match. In what follows, we assume match quality dominates. This is guaranteed by the restriction

A4 $\gamma^g/\gamma^b > \theta_A/\theta_B$.

Recall from the paragraph following (2) that $\theta_A/\theta_B \geq y_A/y_B$ because the income ratio never exceeds the productivity ratio. Combined with A4, this yields $\gamma^g/\gamma^b > y_A/y_B$ so that match quality trumps income when these objectives conflict. In particular, good agents from the lower income site will always prefer a good local match to a bad mobile match. Thus an allocation that is immune to deviation must still achieve the maximum feasible number of good matches h_{max}^g as in Lemma 2.

Lemma 3 (unequal incomes across sites).

Let $y_A > y_B$ and adopt A4. Fix some quality state $\underline{n}$ such that $n^F \leq n^M$.

A household allocation $\underline{h} \in H(\underline{n})$ is immune to deviation iff $\underline{h}$ satisfies both EM and IC:

EM (efficient matching) $\underline{h}$ achieves h_{max}^g

IC (incentive compatibility) $\underline{h}$ either has no good local matches at site B; no bad matches involving a good male from site A; or both.

For a given exogamy level $m \in [0, n]$, there is some $\underline{h} \in H(\underline{n})$ that is immune to deviation and yields $m = M(\underline{h})$ iff $m \in [\eta_{min}(\underline{n}), \mu_{max}(\underline{n})]$. This interval is non-empty with

$$\begin{aligned} \eta_{min}(\underline{n}) &= 2(n_A^F - n_A^M) && \text{if } n_A^M \leq n_A^F; \\ &= 2(n_A^M - n_A^F) && \text{if } n_A^F \leq n_A^M \leq n^F; \\ &= 2n_B^F && \text{if } n^F \leq n_A^M \\ \mu_{max}(\underline{n}) &= n - 2 \max \{0; n_A^F - n_B^M; n_B^F - n_A^M\} \end{aligned}$$

At most one of the arguments of the max operator can be strictly positive.

When $n^F \geq n^M$ the roles of F and M in Lemma 3 are reversed.

From efficient matching (EM), an allocation $\underline{h}$ that is immune to deviation must achieve h_{max}^g. Given that good females are globally scarce by assumption ($n^F \leq n^M$), EM implies that all good females have good partners.

The incentive compatibility restriction (IC) arises as follows. Suppose we have a good local match at site B and a bad match involving a good male from site A. The good female from site B would deviate with a good male from site A to improve income from y^B to y^A while preserving match quality. The good male from site A would deviate with a good female from site B to improve match quality from γ^b to γ^g while preserving income y^A through the mobility of the household. IC rules out mutual gains of this kind.

The first line in the description of the lower bound η_{min} in Lemma 3 for the case $n_A^M \leq n_A^F$ is the same as for μ_{min} in Lemma 2, apart from the change in site labels. In this situation all of the good males from site A can be matched locally to good females and IC does not bind. For the other two lines in the description of η_{min} IC binds and we have the inequality $\mu_{min} < \eta_{min}$. IC does not affect the upper bound $\mu_{max}(\underline{n})$ in Lemma 3, which is the same as in Lemma 2 aside from the change in labels.

The next task is to find the exogamy levels that can be supported by a matching equilibrium for a given productivity state $\underline{\theta}$ and quality state $\underline{n}$. When the sites have equal productivities, from Lemma 1 they must also have equal incomes. Lemma 2 then gives a straightforward result: if $m \in [\mu_{min}(\underline{n}), \mu_{max}(\underline{n})]$ then there is a matching equilibrium that supports m, while for other values of m there is not. We record this result below.

Proposition 1 (supportable exogamy levels with equal site productivities).
Let $\theta_1 = \theta_2$. Fix the quality state $\underline{n}$. An exogamy level $m \in [0, n]$ is supportable by a matching equilibrium iff $m \in [\mu_{min}(\underline{n}), \mu_{max}(\underline{n})]$, with $y_1 = y_2$ for all such m.

The analysis is more complex for unequal site productivities. The productivity ratio $\theta_A/\theta_B > 1$ determines a level of exogamy $m^* \in (0, n)$ through equation (2) such that incomes are unequal across sites when $m < m^*$ and equal when $m \geq m^*$. For a fixed $\underline{n}$ we need to study the relationship between m* and the boundaries $\mu_{min}(\underline{n})$, $\eta_{min}(\underline{n})$, and $\mu_{max}(\underline{n})$ derived in Lemmas 2 and 3. Proposition 2 lists the cases that can arise.

Proposition 2 (supportable exogamy levels with unequal site productivities).
Let $\theta_1 \neq \theta_2$ with $0 < \theta_B < \theta_A$. Fix the quality state $\underline{n}$. We have $\mu_{min}(\underline{n}) \leq \eta_{min}(\underline{n}) \leq \mu_{max}(\underline{n})$.

(i) If $m^* \leq \mu_{min}(\underline{n})$ then an exogamy level m is supportable by a matching equilibrium iff $m \in [\mu_{min}(\underline{n}), \mu_{max}(\underline{n})]$, with $y_A = y_B$ for all such m.

(ii) If $\mu_{min}(\underline{n}) \leq m^* \leq \eta_{min}(\underline{n})$ then an exogamy level m is supportable by a matching equilibrium iff $m \in [m^*, \mu_{max}(\underline{n})]$, with $y_A = y_B$ for all such m.

(iii) If $\eta_{min}(\underline{n}) \leq m^* \leq \mu_{max}(\underline{n})$ then an exogamy level m is supportable by a matching equilibrium iff $m \in [\eta_{min}(\underline{n}), \mu_{max}(\underline{n})]$, with $y_A > y_B$ for $\eta_{min}(\underline{n}) \leq m < m^*$ and $y_A = y_B$ for $m^* \leq m \leq \mu_{max}(\underline{n})$.

(iv) If $\mu_{max}(\underline{n}) < m^*$ then an exogamy level m is supportable by a matching equilibrium iff $m \in [\eta_{min}(\underline{n}), \mu_{max}(\underline{n})]$, with $y_A > y_B$ for all such m.

The general message of Proposition 2 is that for a fixed quality state $\underline{n}$, lower values of m* are associated with matching equilibria that have equal incomes across sites, while higher values of m* are associated with equilibria that have unequal incomes. At small m* values as in case (i), the result is the same as in Proposition 1. At larger m* values as in case (ii), the situation is identical except that the previous lower bound on exogamy $\mu_{min}(\underline{n})$ is replaced by m*. Once m* exceeds $\eta_{min}(\underline{n})$ as in case (iii), there are equilibrium exogamy levels consistent with equal or unequal incomes. Finally, when m* is large as in case (iv), any equilibrium exogamy level must yield unequal incomes.

4. Comparative Statics

In the empirical part of the paper, we will be interested in the expected exogamy rate E(M) where $M \equiv m/n$ is the fraction of agents who marry outside their community. This expectation depends on the variation in natural productivity captured by θ_A/θ_B and community size captured by n. We will show that an increase in the productivity ratio increases E(M) while an increase in community size decreases E(M).

Section 3 characterized intervals for the exogamy levels that can be supported in a matching equilibrium. In order to generate predictions that are useful for empirical work, we adopt the following equilibrium selection rule.

A5 When multiple exogamy levels are supportable, the lowest is selected.

A5 uses the preference for local matches as a tie-breaking rule. Such preferences could arise from a desire for closeness to family and friends, for a partner from the same birth community, or site-specific human capital. This assumption rules out heterogeneous individual preferences about such matters (e.g., the possibility that some agents want their families to be nearby while other agents want their families to be far away).

The expected exogamy rate has two components: the expectation when the sites have the same productivity, and the expectation when the site productivities differ. Due to the independence of productivity shocks and agent qualities, we can write

$$\begin{aligned} (6) \qquad E(M) \quad = \quad & \text{Prob (AA or BB) } E(M \mid \theta_1 = \theta_2) \\ & + \text{Prob (AB or BA) } E(M \mid \theta_1 \neq \theta_2) \end{aligned}$$

The probability in each line is strictly positive by A2. We confine attention to the more interesting case $\theta_B < \theta_A$ with obvious simplifications when $\theta_B = \theta_A$ holds.

In the first line of (6), there is no need to distinguish states AA and BB because the exogamy rate is not affected by the absolute productivity level. From Proposition 1 and A5, the exogamy rate is $M = \mu_{min}(\underline{n})/n$ where $\mu_{min}(\underline{n})$ is defined in Lemma 2. The conditional expected exogamy rate is therefore

$$(7) \qquad E(M \mid \theta_1 = \theta_2) = E[\mu_{min}(\underline{n})/n]$$

In the second line in (6), the threshold exogamy level $m^* \in [0, n]$ from section 2 becomes relevant. When $m < m^*$, locational equilibrium requires a higher food income at the higher-productivity site ($y_A > y_B$). When $m \geq m^*$, these incomes are equal ($y_A = y_B$). For a fixed community size, m^* is an increasing function of the productivity ratio θ_A/θ_B. As above, we work with the fractions $M \equiv m/n$ and $M^* \equiv m^*/n$.

To determine the exogamy rate as a function of the agent quality state $\underline{n}$, we use Proposition 2 and A5. In combination, these imply that the exogamy rate M is equal to the intermediate value among the set $\{\mu_{\min}(\underline{n})/n, M^*, \eta_{\min}(\underline{n})/n\}$, where $\mu_{\min}(\underline{n})$ is defined in Lemma 2 and $\eta_{\min}(\underline{n})$ is defined in Lemma 3. Recall that $\mu_{\min}(\underline{n}) \leq \eta_{\min}(\underline{n})$ for all agent quality states $\underline{n}$. We therefore write the conditional expected exogamy rate as

$$(8) \qquad E(M \mid \theta_1 \neq \theta_2)$$

$$= E[\text{intermediate value among } \{\mu_{\min}(\underline{n})/n, M^*, \eta_{\min}(\underline{n})/n\}]$$

Now consider the probability distribution over quality states $\underline{n}$. Again we express each variable relative to the community size n so that $N_A^F \equiv 2n_A^F/n$ is the fraction of site-A females who belong to the good type, and likewise for N_A^M, N_B^F, and N_B^M. All of these variables are confined to the unit interval. We define $\underline{N} \equiv (N_A^F, N_A^M, N_B^F, N_B^M)$.

A6 An agent is good with probability $\lambda \in (0, 1)$. These draws are independent across agents and sites. We use the following approximation to the binomial distribution for large n. The random variables for the fraction of good agents in each of the four site/gender categories $(N_A^F, N_A^M, N_B^F, N_B^M)$ are independent and normally distributed with mean λ and variance $\sigma^2 = 2\lambda(1-\lambda)/n$.

We assume that for communities of several dozen people or more, the normal distribution in A6 is a reasonable approximation. For our comparative static results in Propositions 3 and 4, $M(N_A^F, N_A^M, N_B^F, N_B^M)$ must be extended from the domain $[0, 1]^4$ to $(-\infty, \infty)^4$. We adopt an extension that preserves the non-negativity and continuity of M. The details are provided in the mathematical appendix. This extension plays no substantive role because improper quality states have negligible probability at high values of n.

Now consider the effect of an increase in the productivity ratio θ_A/θ_B on E(M) with community size n held constant. As explained above, this ratio drops out of the first line in (6) so we only need to study the second line in which $\theta_1 \neq \theta_2$. For a quality state $\underline{n}$ with $\mu_{min}(\underline{n}) = \eta_{min}(\underline{n})$, the value of M* is irrelevant and M is constant for that quality state regardless of θ_A/θ_B. For any quality state with $\mu_{min}(\underline{n}) < \eta_{min}(\underline{n})$, M is constant at $\mu_{min}(\underline{n})/n$ for low values of θ_A/θ_B; equal to M* at moderate values of θ_A/θ_B; and constant at $\eta_{min}(\underline{n})/n$ for high values of θ_A/θ_B. Under the normal approximation in A6, there are always parts of the state space on which M is equal to M* and thus increasing in θ_A/θ_B. Because M is constant elsewhere, E(M) is increasing in θ_A/θ_B. We formalize this result as follows.

<u>Proposition 3</u> (effect of productivity ratio on expected exogamy rate). Fix $n > 0$ and let there be two productivity vectors (θ_A^0, θ_B^0) and (θ_A^1, θ_B^1) such that $\theta_A^0/\theta_B^0 < \theta_A^1/\theta_B^1$. The approximation from A6 implies $E(M \mid \theta_A^0/\theta_B^0) < E(M \mid \theta_A^1/\theta_B^1)$.

The second comparative static question involves the effect of an increase in the community size n on the expected exogamy rate E(M), holding θ_A/θ_B constant. Intuition suggests that when n is large most of the gains from good matches are available within an individual community, and it is less necessary to seek good partners elsewhere. For the first line of (6) in which productivities are equal, this intuition can be validated by using A6 to solve explicitly for $E(M \mid \theta_1 = \theta_2)$ and showing that the result is decreasing in n.

For the second line of (6) in which productivities are unequal, matters are more complicated. The problem is that $M^* \equiv m^*/n$ is influenced by both n and θ_A/θ_B through equation (2) in section 2, which makes it difficult to obtain unambiguous results when n changes. This problem disappears under the following technological restriction.

<u>A7</u> The production function has the form $f(L) = L^\alpha$ where $0 < \alpha < 1$.

The reasons for A7 are pragmatic. The functional form in A7 is the simplest one that is consistent with our technological assumptions in A1, and the constant output elasticity implies that scale effects are irrelevant for the determination of M*. More specifically, A7 implies that $M^* \equiv m^*/n$ is increasing in θ_A/θ_B but does not vary with the absolute community size n. Thus we can separate the productivity effect involving θ_A/θ_B, which operates only through M*, from the matching effect involving n, which operates only through the variance of the quality distribution in A6. This delivers clear comparative static predictions.

Under A7 it can be shown that the previous intuition carries over to the situation of unequal productivity. As n increases, A6 ensures that the quality distributions for the site/gender variables $(N_A^F, N_A^F, N_B^F, N_B^M)$ collapse around a common mean. As a result N_A^F is likely to be very close to N_A^M and similarly for site B, so there is little need for exogamous marriage in order to achieve efficient matching.

Proposition 4 (effect of community size on expected exogamy rate). Fix $\theta_A/\theta_B \geq 1$ and use the approximation in A6.

(a) $E(M \mid \theta_1 = \theta_2) = \sigma K$ where K is a positive constant and $\sigma = [2\lambda(1-\lambda)/n]^{1/2}$.

(b) $E(M \mid \theta_1 \neq \theta_2)$ is decreasing in n for sufficiently large n with $E(M) \rightarrow 0$ as $n \rightarrow \infty$.

Part (a) does not require the restriction A7 but part (b) does.

Propositions 3 and 4 provide the foundations for our empirical work. In general we expect to observe more exogamy when natural productivity differences across sites are larger, and less exogamy when communities are larger. The next two sections offer support for these theoretical predictions.

5. Empirical Methods

We test the predictions of our theoretical model using data from the Standard Cross Cultural Sample (SCCS). The SCCS (Murdock and White 1969, 2006; Cashdan 2003) is widely used for cross-cultural studies, and has been used by other economists to study small-scale economies (see for example Pryor, 1986; Baker, 2008). It comprises a representative sample of 186 well-documented and culturally independent pre-modern societies. Each society is "pinpointed" to a specific date and locality. The pinpointing locality corresponds to "the local community where the principal authority conducted his most intensive field research" (Murdock and White 1969, p. 330). The pinpointing date is the earliest date for which reliable ethnographic data exist "to avoid insofar as possible the acculturative effects of contacts with Europeans" (ibid., p. 340). Most pinpointing dates, therefore, are shortly after a society's first contact with Europeans.[1] The SCCS contains detailed ethnographic, geographic, and economic characteristics of each society, compiled from a variety of primary sources, that pertain specifically to that society's pinpointing date and locality.

We restrict the SCCS sample to observations with no missing data (178 observations), and societies without slavery (82 observations). We impose the latter restriction because our theoretical model describes the behavior of individuals making decisions on their own behalf, which is not the case for all individuals in societies with slavery. Moreover, for such societies we cannot be certain which populations (slaves, non-slaves, or both) are described by the variables we use.

Our baseline empirical specification is:

$$Y_i = CS_i\delta + V_i\eta + X_i'\beta + \varepsilon_i \tag{9}$$

where Y_i is the exogamy rate in society i; CS_i is a measure of community size; V_i is a measure of variation in site productivity; X_i is a vector of observable societal and environmental characteristics (controls); ε_i is statistical error; and δ, η, and β are parameters to be estimated. We discuss key variables and specification issues below, and define all variables in Table 1.

Our measure of the exogamy rate, Y_i, is derived from the SCCS measure of the frequency of intercommunity marriage (see Table 1). This variable takes five categorical values ranging from highly endogamous (0-10% of marriages are intercommunity) to

---

[1] Appendix A of Murdock and White (1969) lists the 186 societies of the SCCS, along with their pinpointing dates and localities, and identifies the primary sources on which this information is based. By way of example, the Aztecs are pinpointed to "The city and environs of Tenochtitlan … in 1520, the date of the arrival of the Spaniards," and the principal authority is Sahagun (Murdock and White 1969, p. 366); the Kung ("bushmen") are pinpointed to "the Agau Kung of the Nyae Nyae region … in 1950, when the Marshalls began their study of this still unacculturated group," and the principal authority is L. Marshall (Murdock and White 1969, p. 354).

highly exogamous (90-100% of marriages are intercommunity). We construct a continuous measure of the exogamy rate from this categorical variable that equals the midpoint of the range in the reported category. All specifications reported in the main text are for linear regression models based on this continuous measure of the exogamy rate. For robustness, we also estimate ordered probit specifications based on the categorical measure, and report those estimates in an Appendix. The probit and linear model specifications yield qualitatively very similar results.

Our measure of community size, CS_i, is also based on a categorical SCCS variable (see Table 1). For ease of interpretation, our main results are based on a continuous measure of community size, defined as the natural logarithm of the midpoint of the range in the reported category (the category cutoffs increase at a roughly exponential rate). We also report estimates based on a more flexible functional form where we include dummies for each community size category.

We report estimates for two different measures of variability in site productivity, V_i. The first is the coefficient of variation in mean annual rainfall at the society's pinpointed location. It is natural to proxy for productivity in food production using rainfall, because the vast majority of SCCS societies are agrarian or based on hunting and gathering.[2,3] Our second measure is an index equal to the first principal component of the correlation between our rainfall measure, elevation, and the intra-year range of mean

---

[2] Baker (2003) also uses rainfall variation to proxy for productivity variation in small-scale societies.

[3] Note that our marriage model concerns spatial variation in productivity (between sites) at a point in time. In contrast, our rainfall measure concerns intertemporal variation in site productivity at a point in space (the pinpointing location). We rely on the latter as a proxy for the former, because there is no convincing measure of spatial variation in productivity in the SCCS.

monthly temperature.[4] We expect productivity in agrarian and hunter-gatherer societies to depend on environmental factors such as these. Our index is designed to capture the (statistically) most important dimension of common variation in these three factors.

We report estimates for specifications with and without control variables, X_i. We include controls to improve the precision of our estimates, and to reduce the risk of bias due to omitted factors that may be correlated with the exogamy rate and our variables of primary interest (community size, in particular) for reasons outside our theoretical model. To wit, we control for social stratification (egalitarian vs. class-based society) since we expect this to determine marriage opportunities within societies.[5] We also control for the likely determinants of average productivity in food production (region, method of food production, sedentism/nomadism, food storage technology, mean annual rainfall, mean annual temperature), because average productivity almost certainly determines community size, but could also have a direct effect on the exogamy rate if our theoretical model is incomplete or incorrect. Finally, we control for whether a society's territory and population are in a period of long run growth or decline. We have in mind that our theoretical model describes a long run demographic equilibrium (which would entail a steady state population size, and hence no growth). However, societies might be out of long run equilibrium as of the pinpointing date, and this could conceivably disrupt equilibrium marriage practices.

---

[4] The rainfall measure is SCCS variable v1914 "Coefficient of variation in mean annual rainfall," and the elevation measure is SCCS variable v183 "Altitude in meters." The intra-year temperature range is the difference between SCCS variable v187 "Hottest month mean temperature (°C)" and SCCS variable v188 "Coldest month mean temperature (°C)." All three variables receive positive factor loadings (0.60, 0.52, and 0.77, respectively), and the first principal component explains 42 percent of their covariation.

[5] Our measures of social stratification are derived from SCCS variable 158. This variable is defined in Murdock and Provost (1973). Egalitarian societies are defined as "… essentially egalitarian, lacking social classes, castes, hereditary slavery, and important wealth distinction," whereas class-based societies are "… stratified into two social classes of freemen but lacks both caste distinctions and hereditary slavery."

Summary statistics are reported in Table 2. The mean exogamy rate in our estimation sample is slightly more than 50 percent. Communities are generally small: more than 20 percent of societies have communities of fewer than 50 people, and 90 percent have communities of fewer than 400. Most societies (about 60 percent) are agricultural, about 5 percent are pastoral, and the remainder rely on other methods of food production (hunting, gathering, and fishing, primarily). Relative to the entire SCCS sample, societies in our estimation sample (i.e., those without slavery) generally have smaller communities, are less reliant on agriculture, are less likely to be located in the Mediterranean basin, are less likely to have fixed settlements and have adequate food storage, and were pinpointed about 50 years later.

Endogeneity Concerns. Our theoretical model treats variation in site productivity (θ) and community size (n) as exogenous. If our empirical measures of these quantities satisfy this condition,[6] then parameter estimates based on eq. (9) have a causal interpretation. This assumption is plausible for our measures of productivity variation, because these are based on environmental factors, but less so for our measure of community size. In particular, if there are unobserved factors outside of our theoretical model that influence both community size and the exogamy rate (e.g., cultural institutions, technology), then community size may be endogenous in eq. (9).

To address this possibility, we also report instrumental variables (IV) estimates based on eq. (9). Our instrument for community size is the society's pinpointing year. From the society's perspective, the pinpointing year is clearly exogenous.[7] We expect

---

[6] Or more precisely, if they satisfy a conditional independence condition given X_i; see Angrist and Pischke (2009, p. 51).

[7] Recall that pinpointing year is the earliest date for which reliable ethnographic data exist, and is usually shortly after first contact with Europeans.

pinpointing year to be negatively correlated with community size, because exogenous factors that are likely to increase community size (e.g., abundance of natural resources, proximity to the coast, a temperate latitude and climate, etc.) are also likely to have brought the society into contact with Europeans earlier. In effect, pinpointing year acts as a convenient proxy for a host of plausibly exogenous environmental factors that are likely to determine community size.

6. Empirical Results

Table 3 presents our main regression estimates.[8] Columns 1-3 are based on our rainfall measure of variation in site productivity, and the remaining columns are based on our index. Estimates based on the two measures are very similar, but the estimates based on the index are less precise. We therefore restrict our discussion to estimates based on the rainfall measure. The estimates strongly support the predictions of our theoretical model. From column 1, we see that the exogamy rate declines monotonically with community size and increases with variation in site productivity. The relationship with community size is nearly log-linear, so we prefer the more parsimonious specification based on our continuous measure of community size in columns 2 and 3. Controlling for observable characteristics X_i (column 3) increases the magnitude and precision of both effects. The baseline estimates in column 3 indicate that a one standard deviation increase in (log) community size reduces the exogamy rate by 15.9 percentage points, and a one standard deviation increase in rainfall variability increases the exogamy rate by 10.4 percentage points. This does not necessarily imply that matching considerations are a more important determinant of exogamous marriage than productivity shocks, however,

---

[8] Corresponding estimates based on the ordered probit specification are in Appendix Table 1. They are qualitatively very similar to the estimates in Table 3.

because estimates based on our rainfall measure provide a lower bound on the effect of productivity variation on exogamy.[9]

Table 4 presents our IV estimates. Visual inspection of the data suggests that the unconditional relationship between (log) community size and our instrument is non-monotone. To avoid mis-specifying our first-stage regressions, we therefore parameterize the relationship between pinpointing year and community size using a set of four indicator variables (pinpointing year before 1850, 1850-1899, 1900-1949, and 1950 or later). The cutoff years for the indicators correspond roughly to turning points in the unconditional relationship between pinpointing year and community size. The first three columns of Table 4 are based on our rainfall measure of variation in site productivity, and the remaining three columns are based on our index. Again, estimates are very similar for the two measures, so we focus our discussion on the rainfall measure.

The two-stage least squares (2SLS) estimates in column 1 are similar to our baseline estimates in column 3 of Table 3, although our 2SLS estimates are somewhat larger in absolute value. We test the hypothesis that community size is exogenous via the usual Durbin-Wu-Hausman test, and fail to reject this hypothesis. We likewise cannot reject the model's overidentifying restrictions (i.e., validity of the instruments) via the usual *J* test. There is some indication, however, that the instruments are weak.[10] It is well known that weak instruments bias IV estimates toward the corresponding OLS estimate, and that exogeneity/validity tests can be unreliable in the presence of weak

---

[9] Our rainfall measure is a proxy for variation in site productivity, and hence measures productivity variation with error, which generates attenuation bias.

[10] The first stage F-statistic on the excluded instruments is 3.02, which is considerably less than the "rule of thumb" value of 10. The Cragg-Donald F-statistic for the test of weak identification is 3.56, which is considerably less than Stock and Yogo (2005) critical values (e.g., 9.08 for 10% maximal bias, and 9.54 for 20% maximal size).

instruments. We therefore present Limited Information Maximum Likelihood (LIML) estimates in column 2. LIML is known to be less biased than overidentified 2SLS in the presence of weak instruments, but it is also less precise. The LIML estimates are very similar to the 2SLS estimates in column 1, which suggests that any bias due to the weak instruments is probably very small.

Following the prescription of Angrist and Pischke (2009), we report exactly-identified 2SLS estimates for our strongest instrument (the indicator for a pinpointing year after 1949) in column 3. Exactly-identified 2SLS is median-unbiased, and hence minimizes concerns about weakness of our instrument. Moreover, there is less evidence in this specification that our single instrument is weak.[11] Hence we prefer these estimates to our overidentified IV estimates.[12] The estimates in column 3 remain very similar to our OLS estimates, but larger in magnitude, and we easily reject that community size has no effect on the exogamy rate using a test that is robust to weak instruments.[13] For this specification, a one standard deviation increase in (log) community size reduces the exogamy rate by 28.0 percentage points and a one standard deviation increase in rainfall variability increases the exogamy rate by 11.2 percentage points. Overall, our IV estimates reinforce our earlier conclusion that the SCCS data support the predictions of our theoretical model.

---

[11] The Cragg-Donald F-statistic for the test of weak identification is 7.44, which lies between the Stock and Yogo (2005) critical values for 15% and 20% maximal size (8.96 and 6.66, respectively). As for validity of the instrument, we conduct a rough test by adding the single instrument to the OLS specifications reported in columns 3 and 6 of Table 3. In both cases, we fail to reject the hypothesis that the instrument can be excluded from the outcome equation (p-values for the t-test are .268 and .318 for the two specifications), indicating that the data support the exclusion restriction.

[12] Once again, we fail to reject the hypothesis that community size is exogenous in this specification. Since our single instrument is reasonably strong, we have some confidence in this result, and consequently, in the OLS estimates in Table 3 as well.

[13] p-values for the Anderson-Rubin weak-instrument robust Wald test are given in square brackets in Table 4.

Robustness. Our estimates in Tables 3 and 4 are robust to a variety of alternate specifications. As noted previously, we report ordered probit estimates in Appendix Table 1; they are very similar to estimates in Table 3. We have also estimated specifications, not reported here but available on request, using alternate instruments based on the natural environment: mean temperature, mean rainfall, the number of nearby natural habitats, and proximity to the ocean. Each of those instruments is weaker than pinpointing year, but yields estimates that are qualitatively similar to those reported in Table 4.

Our theoretical model concerns monogamous marriage. Some societies in our estimation sample, however, practice polygamous marriage. Although we have no a priori reason to suspect that our model's predictions do not apply to polygamous societies, we verify this by restricting our sample to predominantly monogamous societies and re-estimating the specifications reported in Tables 3 and 4.[14] The resulting estimates are reported in Appendix Tables 2 and 3. The community size coefficients are very similar to our main estimates and remain highly statistically significant. Coefficients on our measures of productivity variation have the same sign as in Tables 3 and 4, but are much less precisely estimated. We attribute the lower precision to the reduced sample size.

Finally, we investigate the possibility that exogamy may be motivated by incest avoidance in our data. As pointed out in the introduction, incest can readily be avoided by directly prohibiting marriages among close relatives, without requiring exogamy for the community as a whole. Most societies in our sample (about 61 percent) prohibit

---

[14] We drop 25 observations where the incidence of polygamy exceeds 20 percent, as reported in SCCS variable 79 "Polygamy."

marriages between close cousins (second cousins or closer). However, one might still be concerned that such prohibitions indirectly lead to a higher rate of exogamy, especially in small communities. If so, one might expect a positive association between the presence of prohibitions on close cousin marriage and the exogamy rate. In fact, we find no statistically significant relationship between them.[15] When we add an indicator for a prohibition on marriage between close cousins and its interaction with community size to our main OLS specification (column 3 of Table 3), neither coefficient is statistically significant. The coefficients on our community size and rainfall measures are little changed and remain highly significant.[16] Thus prohibitions on marriage between cousins seem to have no discernable effect on the exogamy rate, even in small communities. It therefore seems unlikely that exogamy is being driven by incest avoidance in our data.

7. Conclusion

We have developed a theoretical model in which the exogamy rate in pre-modern societies is higher when community sizes are small, and when sites differ in their natural productivity. These predictions are strongly supported by our empirical work. Baseline estimates indicate that a one standard deviation increase in (log) community size reduces the exogamy rate by 15.9 percent, and a one standard deviation increase in rainfall variability increases the exogamy rate by 10.4 percent. IV estimates that address the possible endogeneity of community size also support our theoretical model. Further

---

[15] A simple regression of the exogamy rate on an indicator for a prohibition on marriage between close cousins has R^2=0.004, and the coefficient on the indicator is 3.7 with a p-value of 0.58.

[16] R^2=0.45 in this regression. The coefficient on the indicator for a prohibition on marriage between close cousins is -5.74 (p-value=0.836), and the coefficient on its interaction with log community size is 1.68 (p-value=0.75). Coefficients on log community size and our rainfall measure are -14.4 (p-value=0.002) and 0.801 (p-value=0.018), respectively.

reinforcement is provided through robustness checks involving specifications, procedures, and control variables.

As we noted in the introduction, exogamy is an important variable in the study of human prehistory. During the late Pleistocene (until about 11,600 years ago), the world climate system was highly volatile, even on decadal time scales. Natural shocks to food productivity were therefore frequent and severe. Since the onset of the Holocene, world climate has been much more stable (Richerson, Boyd, and Bettinger, 2001; Woodward, 2014, chs. 8-9). The transition from the Pleistocene to the Holocene was accompanied by a transition from mobile foraging to sedentary foraging, and then to agriculture, in many parts of the world. This led to larger community sizes (Bellwood, 2005; Dow and Reed, 2015). Our results suggest that these environmental and technological developments may have led to a shift from mainly exogamous marriages across small foraging bands to mainly endogamous marriages within farming communities.

Our empirical estimates are based on the subset of societies in the Standard Cross Cultural Sample that do not have slavery. Some societies in our sample are egalitarian while others are stratified by economic class. This distinction was included among our control variables. In both cases, exogamy has a straightforward geographic meaning. However, in complex societies marriages that are exogamous with respect to geography may be endogamous with respect to class, caste, religion, or other social boundaries. Feudal societies, states, and empires provide many examples in which elite marriages were used to reinforce multi-country military or commercial alliances. There is much research to be done on the evolution of marriage systems in societies where stratification is in its early stages (Dow and Reed, 2013).

Aside from the costs and benefits for individual agents, marriage practices could influence the growth or decline of entire populations. This could be true, for example, if genetic diversity is especially valuable in particular disease environments (we thank an anonymous referee for this point). We have constructed models involving Malthusian population dynamics and competition among societies to study whether marriage rules have long run survival benefits for the societies that adopt them. It was difficult to derive clear predictions using this framework, and we abandoned it in favor of the comparative static analysis in section 4. However, this question could be pursued in future work.

There is a long-standing debate among anthropologists about the importance, or even the relevance, of economic variables in explaining the social organization of small-scale societies. Some (Harris, 1980) insist that social institutions like marriage systems are strongly influenced by the natural environment, technology, and related economic variables. Others (Hallpike, 1988) insist that these material factors impose quite loose constraints on institutions, which are shaped mainly by autonomous cultural evolution.

Our results support the role of economic variables. Marriage systems are central to small-scale societies and we have shown that an important dimension of such systems (exogamy) responds to the environment and community size in the ways that economic reasoning would lead one to expect. Recent research by Do et al. (2013) likewise shows that the incidence of marriage between close genetic relatives is responsive to economic factors, despite widespread arguments in favor of religious or cultural explanations.

Future research might use the SCCS or similar data sets to test explanations based on economic variables against explanations based on common cultural descent (proxied, for example, by membership in a common ethnic or linguistic group). A serious attempt to adjudicate this dispute would need to consider a wider range of institutions and cultural practices, but our results suggest that economic variables have a substantial role to play.

References

Angrist, J. D. and J.-S. Pischke, 2009, Mostly Harmless Econometrics: An Empiricist's Companion, Princeton University Press, Princeton, NJ.

Baker, Matthew J., 2003, An equilibrium conflict model of land tenure in hunter-gatherer societies, Journal of Political Economy 111(1), February, 124-73.

Baker, Matthew J., 2008, A structural model of the transition to agriculture, Journal of Economic Growth 13(4), December, 257-92.

Baker, Matthew J. and Joyce P. Jacobsen, 2007, A human capital-based theory of postmarital residence rules, Journal of Law, Economics, and Organization 23(1), April, 208-41.

Bellwood, Peter, 2005, First Farmers: The Origins of Agricultural Societies, Blackwell, Oxford.

Bennett, R.L., A.G. Motulsky, A. Bittles, L. Hudgins, S. Uhrich, D.L. Doyle, K. Silvey, C.R. Scott, E. Cheng, B. McGillivray, R.D. Steiner, and D. Olson, 2002, Genetic counseling and screening of consanguineous couples and their offspring: Recommendations of the National Society of Genetic Counselors, Journal of Genetic Counseling, 11(2), April, 97-199.

Berndt, R. M., I964, Warfare in the New Guinea highlands, American Anthropologist 66, 183- 203.

Callaway, Ewen, 2008, All in the family, http://sciencenews.org/view/feature/id/9570, 15(173), April 12.

Cashdan, E., 2003, Ethnic diversity, habitat diversity, and rainfall codes: STDS88-90, World Cultures 13, 180-94.

Diamond, Jared, 2012, The World Until Yesterday: What Can We Learn From Traditional Societies?, Viking, New York.

Do, Quy Toan, Sriya Iyer, and Shareen Joshi, 2013, The economics of consanguineous marriage, Review of Economics and Statistics, forthcoming.

Dow, Gregory K. and Clyde G. Reed, 2011, Stagnation and innovation before agriculture, Journal of Economic Behavior and Organization 77(3), March, 339-350.

Dow, Gregory K. and Clyde G. Reed, 2013, The origins of inequality: Insiders, outsiders, elites, and commoners, Journal of Political Economy 121(3), June, 609-641.

Dow, Gregory K. and Clyde G. Reed, 2015, The origins of sedentism: Climate, population, and technology, Journal of Economic Behavior and Organization 119, November, 56-71; online, Aug. 27, 2015, DOI 10.1016/j.jebo.2015.07.007.

Dow, Gregory K., Clyde G. Reed and Nancy Olewiler, 2009, Climate reversals and the transition to agriculture, Journal of Economic Growth 14(1), March, 27-53.

Greif, Avner, 2006a, The family, institutions, and economic growth, American Economic Review Papers and Proceedings 96(2), May, 308-12.

Greif, Avner, 2006b, Institutions and the Path to the Modern Economy: Lessons from Medieval Trade, Cambridge University Press, New York.

Hallpike, C. R., 1988, The Principles of Social Evolution, Clarendon Press, Oxford.

Harris, Marvin, 1997, Culture, People, Nature, 7th edition, Longman, New York.

Hayano, D. M., I973, Sorcery death, proximity, and the perception of out-groups: The Tauna Awa of New Guinea, Ethnology 12, 179-191.

Helgason, Agnar, Snæbjörn Pálsson, Daníel Guðbjartsson, þórður Kristjánsson, and Kári Stefánsson, 2008, An association between the kinship and fertility of human couples, Science 8(310), no. 5864, 813-816 (see http://www.sciencedaily.com/releases/2008/02/080207140855.htm).

Johnson, Allan and Timothy Earle, 2000, The Evolution of Human Societies, 2nd ed., Stanford University Press, Stanford, CA.

Joseph, Suzanne, 2007, Kissing cousins: Consanguineous marriage and early mortality in a reproductive isolate, Current Anthropology 48(5), 756-764.

Kang, G. E., 1976, Conflicting loyalties theory: A cross-cultural test, Ethnology 15, 201-210.

Kang, G. E., 1979, Exogamy and peace relations of social units: A cross-cultural test, Ethnology 18(1), January, 85-99.

Keenan, J. 1977a, Power and wealth are cousins: Descent, class and marital strategies among the Kel Ahaggar (Tuareg. Sahara), Part I, Africa: Journal of the International African Institute 47(3), 242-252.

Keenan, J., 1977b, Power and wealth are cousins: Descent, class and marital strategies among the Kel Ahaggar (Tuareg. Sahara), Part II, Africa: Journal of the International African Institute 47(4), 333-343.

Kelly, Robert L., 2007, The Foraging Spectrum: Diversity in Hunter-Gatherer Lifeways, reprinted by Percheron Press, previously published in 1995 by Smithsonian Institution Press, Washington, DC.

Murdock, G. P. and C. Provost, 1973, Measurement of cultural complexity, Ethnology 12(4), 379-392.

Murdock, G. P. and D. R. White, 1969, Standard cross-cultural sample, Ethnology 8(4), 329-369.

Murdock, G. P. and D. R. White, 2006, Standard cross-cultural sample: on-line, Social Dynamics and Complexity Working Papers, UC Irvine, http://escholarship.org/uc/item/62c5c02n

Parkin, Robert, 1997, Kinship: An Introduction to the Basic Concepts, Blackwell, Oxford.

Pryor, F. L., 1986, The adoption of agriculture: Some theoretical and empirical evidence, American Anthropologist 88(4), December, 879-897.

Richerson, Peter J., Robert Boyd, and Robert L. Bettinger, 2001, Was agriculture impossible during the Pleistocene but mandatory during the Holocene? A climate change hypothesis, American Antiquity 66(3), 387-411.

Rosenzweig, Mark R., 1988a, Risk, private information, and the family, American Economic Review Papers and Proceedings 78(2), May, 245-250.

Rosenzweig, Mark R., 1988b, Risk, implicit contracts and the family in rural areas of low-income countries, The Economic Journal 98(393), December, 1148-1170.

Rosenzweig, Mark R. and Oded Stark, 1989, Consumption smoothing, migration, and marriage: Evidence from rural India, Journal of Political Economy 97(4), August, 905-926.

Stock, J. H. and M. Yogo (2005), Testing for weak instruments in linear IV regression, in Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg, D. W. Andrews and J. H. Stock (eds.), 80-108, Cambridge University Press, Cambridge.

Stone, Linda, 2001, ed., New Directions in Anthropological Kinship, Rowman and Littlefield, Oxford.

Woodward, Jamie, 2014, The Ice Age: A Very Short Introduction, Oxford University Press, Oxford, UK.

Table 1: Variable Definitions

Variable	Definition
Exogamy Rate	Frequency of intercommunity marriage. Derived from SCCS variable 72. The discrete measure has 5 categories (local exogamy rate: 1 = 0-10%; 2 = 11-39%; 3 = 40-60%; 4 = 61-89%; 5 = 90-100%). The continuous measure is the midpoint of the range indicated by the categorical value.
Community Size	Community Size. Derived from SCCS variable 63. The discrete measure has 7 categories, as reported in Table 2. The continuous measure is the natural logarithm of the midpoint of the range indicated by the categorical value.
Coefficient of Variation of Annual Rainfall	Coefficient of (year-to-year) variation in mean annual rainfall. Equals SCCS variable 1914.
Temperature Range	Difference between hottest month mean temperature and coldest month mean temperature. Derived from SCCS variables 187 and 188.
Elevation	Altitude in meters of closest weather station. Equals SCCS variable 183.
Index of Variation in Site Productivity	First principal component of correlation matrix of the coefficient of variation in annual rainfall, temperature range, and elevation
Social Stratification:	
Egalitarian Society*	Binary indicator: society is egalitarian. Derived from SCCS variable 158.
Two-Class Society without slavery	Binary indicator: society has 2 social classes but no castes or slavery. Derived from SCCS variable 158.
Multiple Classes with slavery	Binary indicator: society has castes/slavery and/or 3 or more social classes.
Method of Food Production:	
Agriculture*	Binary indicator: contritbution of agriculture to local food supply > 10%. Derived from SCCS variable 3.
Pastoral*	Binary indicator: contribution of animal husbandry to local food supply > 50%. Derived from SCCS variable 5.
Other	Binary indicator: contribution of agriculture to local food supply ≤ 10% and contribution of animal husbandry to local food supply ≤ 50%.
Region:	
Africa*	Binary indicator: region is Africa, excl. North Africa, Madagascar and the Sahara. Derived from SCCS variable 200.
Circum-Mediterranean*	Binary indicator: region is North Africa, Europe, Turkey, Caucasus, or Semitic Near East. Derived from SCCS variable 200.
East Eurasia*	Binary indicator: region is East Eurasia, incl. Madagascar and islands in the Indian Ocean. Derived from SCCS variable 200.
Insular Pacific*	Binary indicator: region is Insular Pacific, incl. Australia, Indonesia, Formosa, and the Phillipines. Derived from SCCS variable 200.
South America*	Binary indicator: region is South America, incl. Antilles, Yucatan, and Central America. Derived from SCCS variable 200.
North America	Binary indicator: region is North America, to the Isthmus of Tehuantepec. Derived from SCCS variable 200.
Fixed Settlements*	Binary indicator: fixity of settlement is permanent (vs. migratory or nomadic). Derived from SCCS variable 61.
Adequate Food Storage*	Binary indicator: food surplus via storage is "simple or adequate," or "complex or more than adequate" (vs. "none or barely adequate"). Derived from SCCS variable 21.
Mean Annual Rainfall**	Mean yearly annual rainfall (cm). Equals SCCS variable 1913.
Mean Annual Temperature**	Mean annual temperature. Equals SCCS variable 186.
Territory/Population Expanding*	Binary indicator: boundaries/population are expanding in the long run. Derived from SCCS variable 908.
Territory/Population Shrinking*	Binary indicator: boundaries/population are shrinking in the long run. Derived from SCCS variable 908.
Pinpointing Year	Year of pinpointing date. Equals SCCS variable 838.
Pinpointing Year:	
<1850	Binary indicator: Pinpointing Year < 1850.
1850-1899	Binary indicator: Pinpointing Year 1850-1899.
1900-1949	Binary indicator: Pinpointing Year 1900-1949.
>1949	Binary indicator: Pinpointing Year > 1949.

Notes: one asterisk (*) indicates the variable is among control variables, X_i, in regressions; two asterisks (**) indicates the variable and its square are included in X_i.

Table 2: Summary Statistics

	Full SCCS Sample		Restricted Sample		
Variable	Mean	Std. Dev	Mean	Std. Dev	pval
Exogamy Rate	53.5	28.0	50.9	29.4	.128
ln(Community Size)	5.40	1.61	4.84	1.20	.000
Community Size:					
1-49	.157	.364	.220	.414	.980
50-99	.152	.360	.195	.399	.928
100-199	.253	.436	.293	.458	.868
200-399	.174	.380	.195	.399	.749
400-999	.152	.360	.049	.217	.000
1000-4999	.073	.261	.049	.217	.121
>5000	.039	.195	.000		.004
Coefficient of Variation of Annual Rainfall	23.8	18.2	22.7	12.7	.217
Temperature Range	11.8	11.0	11.2	11.0	.249
Elevation	458.3	677.7	406.7	498.8	.166
Index of Variation in Site Productivity			.000	1.00	
Social Stratification:					
Egalitarian Society	.360	.481	.780	.416	1.00
Two-Class Society without slavery	.101	.302	.220	.416	1.00
Multiple Classes with slavery	.539	.498	0	0	1.00
Method of Food Production:					
Agriculture	.674	.470	.598	.493	.023
Pastoral	.084	.279	.049	.217	.053
Other	.242	.428	.354	.478	1.00
Region:					
Africa	.146	.354	.146	.356	.504
Circum-Mediterranean	.152	.360	.049	.217	.000
East Eurasia	.180	.385	.122	.329	.030
Insular Pacific	.163	.370	.207	.408	.927
South America	.180	.385	.256	.439	.992
North America	.180	.384	.220	.416	.895
Fixed Settlements	.534	.500	.378	.488	.000
Adequate Food Storage	.618	.487	.488	.503	.001
Mean Annual Rainfall	138.9	107.3	152.4	118.5	.936
Mean Annual Temperature	19.1	9.25	19.9	8.51	.862
Territory/Population Expanding	.225	.419	.146	.356	.009
Territory/Population Shrinking	.315	.466	.378	.488	.952
Pinpointing Year	1854	363.7	1902	47.7	.959
Pinpointing Year:					
<1850	.107	.310	.085	.281	.195
1850-1899	.242	.429	.317	.468	.984
1900-1949	.478	.501	.463	.502	.365
>1949	.174	.380	.134	.343	.095
N	178		82		

Notes: The "Full SCCS Sample" comprises all societies in the SCCS with non-missing values of community size (missing for one society), and mean annual temperature (missing for six societies). The "Resticted Sample" is the subset of societies in the Full SCCS Sample without slavery. The column titled "pval" is the p-value for a t-test of mean differences (with unequal variances) between the Restricted Sample and the sample of societies with slavery. All reported regression specifications are estimated on the Resrticted Sample.

Table 3: Regression Estimates

Variable	(1)	(2)	(3)	(4)	(5)	(6)
ln(Community Size)		-11.2***	-13.3***		-10.9***	-13.6***
		(1.97)	(3.06)		(2.06)	(3.14)
Community Size 50-99	-6.65			-4.46		
	(9.99)			(9.94)		
Community Size 100-199	-13.3			-11.9		
	(8.33)			(8.45)		
Community Size 200-399	-20.5**			-20.2**		
	(9.10)			(9.60)		
Community Size 400-999	-42.9***			-42.7***		
	(11.3)			(11.6)		
Community Size 1000-4999	-57.0***			-53.5***		
	(7.26)			(8.39)		
CV of Annual Rainfall	.321*	.333*	.816***			
	(.183)	(.169)	(.309)			
Index of Variation in Site Productivity				-.353	-.317	12.6*
				(2.92)	(2.63)	(6.37)
Controls for:						
Social Stratification			YES			YES
Method of Food Production			YES			YES
Region			YES			YES
Fixed Settlements & Food Storage			YES			YES
Mean Temperature & Rainfall			YES			YES
Territory/Population Growth			YES			YES
N	82	82	82	82	82	82
R^2	.24	.22	.45	.22	.20	.43

Source: Authors' calculations based on the SCCS database. See Tables 1 and 2 for variable definitions. Robust Standard errors in parentheses. * indicates statistically significant at the 10 percent level, ** indicates statistically significant at the 5 percent level, *** indicates statistically significant at the 1 percent level.

Table 4: IV Estimates

Variable	(1) 2SLS	(2) LIML	(3) 2SLS	(4) 2SLS	(5) LIML	(6) 2SLS
ln(Community Size)	-16.0**	-16.8*	-23.4***	-15.6**	-16.2*	-22.3***
	(6.93)	(8.74)	(8.86)	(7.09)	(8.97)	(8.15)
	[.099]	[.099]	[.013]	[.111]	[.111]	[.019]
CV of Annual Rainfall	.833***	.838**	.878***			
	(.276)	(.279)	(.296)			
Index of Variation in Site Productivity				13.1**	13.3**	14.7**
				(6.06)	(6.27)	(6.05)
Number of Excluded Instruments	3	3	1	3	3	1
First-stage F (excluded instruments)	3.02	3.02	5.46	2.85	2.85	5.21
Cragg-Donald Weak Identification Test (F-statistic)	3.56	3.56	7.44	3.38	3.38	7.13
H_0: Community Size is exogenous (p-value)	.882	.882	.212	.842	.842	.266
Hansen J Overidentification test (p-value)	.188	.191		.168	.168	
N	82	82	82	82	82	82
R^2	.44	.44	.35	.43	.43	.36

Source: Authors' calculations based on the SCCS database. Excluded instruments in columns 1-2 and 4-5 are indicators for pinpointing year before 1850, 1850-1899, and 1900-1949; excluded instrument in columns 3 and 6 is an indicator for pinpointing year 1950 or later. All specifications include the full set of controls for Social Stratification, Method of Food Production, Region, Fixed Settlements, Food Storage, Mean Temperature and Rainfall, and Territory/Population Growth as in columns 3 and 6 of Table 3. Robust Standard errors in parentheses, Anderson-Rubin weak-instrument robust p-values in brackets. * indicates statistically significant at the 10 percent level based on robust standard errors, ** indicates statistically significant at the 5 percent level, *** indicates statistically significant at the 1 percent level.

Appendix Table 1: Ordered Probit Estimates

Variable	(1)	(2)	(3)	(4)	(5)	(6)
ln(Community Size)		-.493***	-.682***		-.477***	-.690***
		(.097)	(.153)		(.100)	(.157)
Community Size 50-99	-.229			-.133		
	(.385)			(.376)		
Community Size 100-199	-0.531*			-.468		
	(.314)			(.314)		
Community Size 200-399	-0.806**			-.799**		
	(.359)			(.376)		
Community Size 400-999	-1.92***			-1.91***		
	(.644)			(.641)		
Community Size 1000-4999	-2.91***			-2.70***		
	(.638)			(.677)		
CV of Annual Rainfall	.011	.011*	.037***			
	(.007)	(.007)	(.015)			
Index of Variation in Site Productivity				-.054	-.059	.556*
				(.117)	(.107)	(.299)
Controls for:						
Social Stratification			YES			YES
Method of Food Production			YES			YES
Region			YES			YES
Fixed Settlements & Food Storage			YES			YES
Mean Temperature & Rainfall			YES			YES
Territory/Population Growth			YES			YES
N	82	82	82	82	82	82
Pseudo-R^2	.10	.09	.20	.10	.08	.19

Source: Authors' calculations based on the SCCS database. See Tables 1 and 2 for variable definitions. Robust Standard errors in parentheses. * indicates statistically significant at the 10 percent level, ** indicates statistically significant at the 5 percent level, *** indicates statistically significant at the 1 percent level.

Appendix Table 2: Regression Estimates on Monogamous Subsample

Variable	(1)	(2)	(3)	(4)	(5)	(6)
ln(Community Size)		-11.4***	-13.7***		-10.8***	-12.1**
		(1.95)	(4.67)		(2.08)	(4.50)
Community Size 50-99	-1.90			.735		
	(10.7)			(10.4)		
Community Size 100-199	-12.2			-9.65		
	(9.22)			(9.11)		
Community Size 200-399	-29.1*			-18.2		
	(10.5)			(11.70)		
Community Size 400-999	-38.2**			-38.4**		
	(18.0)			(18.3)		
Community Size 1000-4999	-57.5***			-53.2***		
	(7.48)			(8.73)		
CV of Annual Rainfall	.225	.271*	.475			
	(.163)	(.150)	(.503)			
Index of Variation in Site Productivity				-1.71	-1.17	.506
				(3.74)	(3.15)	(9.12)
Controls for:						
Social Stratification			YES			YES
Method of Food Production			YES			YES
Region			YES			YES
Fixed Settlements & Food Storage			YES			YES
Mean Temperature & Rainfall			YES			YES
Territory/Population Growth			YES			YES
N	57	57	57	57	57	57
R^2	.29	.26	.43	.28	.25	.42

Source: Authors' calculations based on the SCCS database. See Tables 1 and 2 for variable definitions. The monogamous subsample excludes societies where the incidence of polygamy exceeds 20 percent. Robust Standard errors in parentheses. * indicates statistically significant at the 10 percent level, ** indicates statistically significant at the 5 percent level, *** indicates statistically significant at the 1 percent level.

Appendix Table 3: IV Estimates on Monogamous Subsample

Variable	(1) 2SLS	(2) LIML	(3) 2SLS	(4) 2SLS	(5) LIML	(6) 2SLS
ln(Community Size)	-14.7**	-14.9*	-24.4***	-13.1*	-13.3*	-20.9***
	(7.49)	(8.72)	(8.75)	(6.98)	(7.79)	(7.43)
	[.230]	[.230]	[.050]	[.193]	[.193]	[.047]
CV of Annual Rainfall	.520	.529	.957*			
	(.501)	(.540)	(.553)			
Index of Variation in Site Productivity				1.10	1.20	5.72
				(8.62)	(8.92)	(9.07)
Number of Excluded Instruments	3	3	1	3	3	1
First-stage F (excluded instruments)	3.93	3.93	4.81	4.36	4.36	5.68
Cragg-Donald Weak Identification Test (F-statistic)	3.72	3.72	5.40	4.37	4.37	6.47
H_0: Community Size is exogenous (p-value)	.790	.790	.271	.900	.900	.290
Hansen J Overidentification test (p-value)	.221	.220		.248	.248	
N	57	57	57	57	57	57
R^2	.43	.43	.35	.42	.42	.36

Source: Authors' calculations based on the SCCS database. The monogamous subsample excludes societies where the incidence of polygamy exceeds 20 percent. Excluded instruments in columns 1-2 and 4-5 are indicators for pinpointing year before 1850, 1850-1899, and 1900-1949; excluded instrument in columns 3 and 6 is an indicator for pinpointing year 1950 or later. All specifications include the full set of controls for Social Stratification, Method of Food Production, Region, Fixed Settlements, Food Storage, Mean Temperature and Rainfall, and Territory/Population Growth as in columns 3 and 6 of Table 3. Robust Standard errors in parentheses, Anderson-Rubin weak-instrument robust p-values in brackets. * indicates statistically significant at the 10 percent level based on robust standard errors, ** indicates statistically significant at the 5 percent level, *** indicates statistically significant at the 1 percent level.