Title
Society to 2050 AD: Anthropological Forecasts Extrapolating Correlates of Modernization

Journal
World Cultures eJournal, 16(1)

Author
Denton, Trevor

Publication Date
2008-03-04

Peer reviewed

Society to 2050 AD: Anthropological Forecasts Extrapolating Correlates of Modernization[1]

Trevor Denton
657 Water Street, Peterborough, Ontario, CANADA K9H 3N2: dentont@cogeco.ca

1. INTRODUCTION

A measure of modernization $m \geq 0$ is created applicable to both preindustrial and contemporary societies. A sample of 174 preindustrial societies are coded for $m \geq 0$, time t, $1800 \leq t \leq 1965$ AD, 25 binary constructs X = 1, 0 (e.g. X = high, low divorce rate) and one continuous construct X > 0 (population density). A sample of 189 contemporary countries at 2000 AD is coded for the same constructs. For the 25 binary constructs parametric logistic regression functions P(X = 1) = f(m, t) are fitted to the combined sample of 363 societies. The candidate predictor set is powers from 1 to 3 of m, t and mt. Backward selection (α = 0.05) is used to reduce the candidate predictor set where appropriate. Since the 174 preindustrial societies are equally distributed over the 19th and 20th centuries it is assumed that the fitted models hold over 1800 AD $\leq t \leq$ 2000 AD, although not necessarily prior to 1800 AD. Functions P(X = 1) = f(m) are fitted to the sample of 189 countries at 2000 AD. For the continuous construct X > 0 regression models X = h(m, t), h(m) are fitted. Tests for monotonic trend over m are fitted to both samples for all 26 constructs - either a Cochran-Armitage test (for binary constructs) or a Spearman rho test (for the continuous construct). Forecasts are made by extrapolating the fitted models to out of sample values of the predictor variable(s) up to 2050 AD, or by linear approximation. Internal checks are created to enhance forecast validity. Extension of the forecast method to new subject matters is considered. Business and policy applications are suggested. Strengths and limitations of the method are outlined.

2. FORECASTS EXTRAPOLATING BEHAVIORAL RELATIONS P(X = 1) = f(m, t) OR X = h(m, t)

A behavioral relations forecast (Harvey 1993) of a Bernoulli random variable X = 1, 0 starts by modeling how X "behaves" in the presence of a predictor set $\mathbf{z}` = <z_1, z_2, \ldots, z_p>$. The behavioral relation connecting predictor set **z`** to Bernoulli random variable X = 1, 0 might be stated as an equation for the probability P(X = 1) = f(**z`**). We discover whether P(X = 1) really is equal to f(**z`**) by fitting the equation P(X = 1) = f(**z`**) to data. Such a behavioral relations model might be a fully specified causal model or simply a curve fitted to data. There are few fully specified causal models in Cultural Anthropology at the moment. We will focus on curve fitting.
A behavioral relations forecast to some future time t using the fitted P(X = 1) = f(**z`**) is done in two stages. First, we forecast **z`** at time t. Secondly, we extrapolate the behavioral relation f(**z`**) to **z`** at time t. Most constructs currently found in data bases of Cultural Anthropology

are multinomial random variables X = a, b, …, r which may be collapsed into binary random variables X = 1, 0. If X is a continuous random variable, $-\infty < X < \infty$, a behavioral relations forecast may be made in analogous fashion by extrapolating a fitted behavioral relation X = h(**z**`) to **z**` at time t.

Is there a predictor set **z**` such that we might hope to fit functions P(X = 1) = f(**z**`) or X = h(**z**`) to data, thereby discovering the requisite behavioral relation to extrapolate as a forecast? The answer is yes. It has been found that, in preindustrial societies, many subject matters of Cultural Anthropology are functions, generally logistic regression functions (Denton 2003, 2004a), of a measure of modernization m to be considered in the next paragraph. To *m* we add time *t* so as indirectly to measure the effects of unknown predictors beyond *m*. For example, if the candidate predictor set is **z**`= $<m, m^2, m^3, t, t^2, t^3, mt, m^2t^2, m^3t^3>$, we have models P(X = 1) = f(**z**`) or X = h(**z**`) which just might be used to make forecasts, provided we know **z**` at time *t* and are able to fit to data the behavioral relation model P(X = 1) = f(**z**`) or X = h(**z**`). We will discover such behavioral relations by fitting equations to data sets to be created in section 3.

The measure of modernization we will use is an updated form of the cultural complexity M-scale of Murdock and Provost (1973), *m* = 0, 1, 2, …, 40. Suppose we draw a sample of *n* preindustrial societies such as 1950's Bushmen, ancient Babylon, 17th century BC, Huron, …, *n*. If we measure each society *j* = 1, 2, …, *n* using all 10 subscales of Table 1, and for each society denote the sum of the 10 subscales $m_j = \Sigma s_i$, *i* = 1, 2, …, 10, *j* = 1, 2, …, *n*, then we have an established measure of the modernization of each society.[2] Murdock and Provost (1973) selected the 10 subscales of Table 1 because these constructs were considered best able to distinguish between forms of societies along a continuum from the most recent form to the most ancient form.

It is worth emphasizing that the degree of modernization of a society is defined here - à la Murdock and Provost (1973) – to be, "the degree of recency of the form of the society as measured by Table 1, updated in such a way as to measure both preindustrial and industrial societies." According to Murdock and Provost (1973:379), Table 1 distinguishes between societies, "with reference to … classificatory criteria which have been postulated to correlate with different levels or stages in cultural development." (present writer's underlining)

The sequence 0, 1, 2, 3, 4 of each of the subscales of Table 1 (Page 65) is also the approximate sequence by which the states of the subscale appeared over time. For any given subscale, not every society progresses from 0 to 1, 2, 3, 4 in that sequence. But, where a society is measured at 4 (or 3, or 2, or 1) for such a subscale, Murdock and Provost (1973:379) imply that there is a direct, or indirect, historical connection to a prior, lower level of the subscale. By this definition, societies ancient in time (e.g. ancient Rome) may be more modern than recent societies rating lower on Table 1 (e.g. Kalahari Desert Kung of the 1950's). In short, modernization denotes the recency of the form of a society, not its chronological recency.

Perusal of Table 1 will show that the M-scale $m = 0, 1, 2, \ldots, 40$ is able to distinguish among preindustrial peoples of varying modernization m but is unable to distinguish, for example, early 20th century China from the contemporary United States of America (USA), nor contemporary USA from Bulgaria; on the scale of Table 1 all would be scored 40. We need to re-jig the M-scale in this regard.

By focusing on division of labor we may extend the M-scale of Table 1 to make it apply to preindustrial, industrial and future societies up to 2050 AD or beyond. Consider the number of specialized occupations $w = 0, 1, 2, \ldots$ present in a society. Subscale 5 of Table 1 and number of specialized occupations $w = 0, 1, 2, \ldots$ both measure division of labor but the latter does so in a much more thorough way. It is shown in Appendix 1 that, for a sample of preindustrial societies, w is a univariate function of m, $w = \mathrm{g}(m) = \exp(0.00138m^2 + 0.09040m)$. Since this function is monotone it may be inverted to obtain $m = \mathrm{g}^{-1}(w)$. Division of labor is an established construct in the social sciences (Denton 1996). The constructs causing number of specialized occupations $w = 0, 1, 2, \ldots$ appear to be approximately the same in industrial and preindustrial societies - only the magnitudes of the predictors differ producing higher w in the former (Denton 1996). Hence, if we count the number of specialized occupations $w = 0, 1, 2, \ldots$ in a contemporary society its degree of modernization m may be approximated (Appendix 1) as $m = \mathrm{g}^{-1}(w)$, the non-negative solution to the quadratic equation $0.00138m^2 + 0.09040m - \ln(w) = 0$. Such a measure preserves the scale of Table 1, extending it to contemporary and future societies.

Since $m = \mathrm{g}^{-1}(w)$, if we know number of specialized occupations $w = 0, 1, 2, \ldots$ at some future time t we have a reasonable approximation of $m \geq 0$ at time t. If we also know the function $\mathrm{P}(\mathrm{X} = 1) = \mathrm{f}(m, t)$, or $\mathrm{X} = \mathrm{h}(m, t)$, we have the rudiments of a method for making forecasts of some construct X correlated with modernization. In case there are additional, unknown predictors of Bernoulli $\mathrm{X} = 1, 0$ or continuous X, $-\infty < \mathrm{X} < \infty$, we will fit to data functions $\mathrm{P}(\mathrm{X} = 1) = \mathrm{f}(m, t)$, $\mathrm{X} = \mathrm{h}(m, t)$ where the candidate predictor set includes interaction terms and powers up to 3 of m, t and mt.

Chatfield (1996:68) suggests that time series forecasts (extrapolating a model of a time series of observations) are preferable to forecasts extrapolating a fitted curve. The reason given is that several curves may occasionally be fitted to the same data set, each curve resulting in a different forecast. For reasons to be stated in the conclusion to this paper, anthropological time series of observations going back thousands of years are currently problematic. If anthropological time series forecasts are currently problematic, anthropological behavioral relations forecasts may be an alternative. Internal checks will be created below to enhance the validity of forecasts slightly outside the ranges of the predictors for which a curve is fitted.

Table 1. Subscales of Murdock and Provost (1973) Preindustrial Modernization M-scale

Subscale 1 Writing and Records 4 True writing; records 3 True writing; no records 2 Nonwritten records 1 Mnemonic devices 0 None	**Subscale 6 Land Transport** 4 Automotive vehicles 3 Animal drawn vehicles 2 Draft animals 1 Pack animals 0 Human only
Subscale 2 Fixity of Residence 4 Sedentary 3 Sedentary; impermanent 2 Semisedentary 1 Seminomadic 0 Nomadic	**Subscale 7 Money** 4 True money 3 Elementary forms 2 Alien currency 1 Domestically usable articles 0 None
Subscale 3 Agriculture (Intensification) 4 Intensive 3 Primary; not intensive 2 more than 10%; secondary 1 less than 10% food supply 0 None	**Subscale 8 Density of Population** 4 Greater than 100 persons per sq. mile 3 26 - 100 2 5.1 - 25 1 1 - 5 0 less than 1
Subscale 4 Urbanization (Mean Size of Local Communities) 4 greater than 1000 persons 3 400 - 999 2 200 - 399 1 100 - 199 0 fewer than 100	**Subscale 9 Level of Political Integration** 4 3 or more admin. levels above loc. com. 3 2 levels 2 1 levels 1 Autonomous local communities 0 None
Subscale 5 Technological Specialization 4 At least smiths, weavers and potters 3 Metalwork only 2 Loom weaving only 1 Pottery only 0 None	**Subscale 10 Social Stratification** 4 3 social classes or castes 3 2 social classes, castes/slavery 2 2 social classes, no castes/slavery 1 Hereditary slavery 0 Egalitarian

Definitions of subscales are from Divale (2004b), VAR149 – VAR158 rescaled from 1 – 5 to 0 – 4. More exact definitions appear in Murdock and Provost (1973).

3. DATA

In section 4 we will discover behavioral relations $P(X = 1) = f(m, t)$ (Bernoulli $X = 1, 0$), $X = h(m, t)$ (continuous X, $-\infty < X < \infty$) by fitting equations to data. In section 3 we create the data.

The writer's goal is to construct a data base having the following three properties. First, the data base should consist of observations of constructs related to, "how societies work."[3] Secondly, measurement of the observations must have adequate validity and reliability. Thirdly, the conditional observations $X|(m, t)$ (X Bernoulli or continuous) must "sufficiently" span modernization *m* and time *t*. Societies from the least to most modern must be included. For a 45-year forecast horizon to about 2050 AD, observations ought to span at least the era 1800 AD to present.

We start by finding a sample of preindustrial societies. After that, we will create a sample of contemporary societies. We will obtain an overall sample by adding the preindustrial and contemporary samples together.

3.1. A Sample of 174 Preindustrial Societies: The Standard Cross-Cultural Sample

For a sample of preindustrial societies we will use the Standard Cross-Cultural Sample (SCCS). The SCCS is the most widely used ethnographic data base today. It was created by the cultural anthropologists Murdock and White (1969; see also Barry and Schlegel 1980). Technically, the SCCS is not a sample. Rather, it is a frame of size 186 each element of which is focused to a particular time and place.[4] For example, SCCS society number 1 is the African people Nama Hottentot focused to the Gei//Khauan tribe (27^o 30'S17^o E) at 1860AD with established ethnographic sources (Divale 2004a). If one wishes to create a conditional sample of observations of a random variable X (categorical or continuous), conditioning on modernization *m* and time of observation *t*, one reads each of the 186 SCCS ethnographies and measures each society for X, *m* and *t*. For example, Murdock and Provost (1973) coded all 186 SCCS societies for each of the 10 constructs of Table 1. They measured m of the Nama Hottentot to be 8 at 1860 AD. Over the years, well over 1000 constructs have been measured for SCCS societies. These measurements are readily available in electronic form (Divale 2004b).

Most coded data currently available for the SCCS (Divale 2004b) consist of codes for the modal value of a construct for SCCS societies. For example, Nama Hottentot residence after marriage is coded patrilocal (with or near the husband's parents) even though some Nama Hottentot marriages reside elsewhere. Such global characterization of a society obscures intra-societal variation. In spite of this shortcoming, cross-cultural research has been able to recognize sound behavioral theory (Levinson and Malone 1980).

For our purposes, societies of the SCCS sufficiently span both modernization m and time t. SCCS societies are all preindustrial with modernization m ranging from 0 to 40. In the hope of fitting simpler models we will delete the 12 SCCS societies predating 1800 AD, the earliest being Babylonia at 1750 BC.[5] The remaining 174 SCCS societies are relatively equally spaced over m and the 1800s and 1900s. The most recent SCCS time of ethnographic observation is 1965 AD, for example the #71 Burmese. Behavioral relations discovered for SCCS societies appear to hold back to 1000 AD (Denton 2004b).

3.2. A Sample of 189 Contemporary Societies: UN Member Countries at 2000 AD

Currently, there is no usable data base of observations of contemporary countries.[6] We will create a data base of all member countries of the United Nations at 2000 AD. 2000 AD is a convenient time of observation since it facilitates comparison with times such as 1500 AD, 1000 AD, and the like. While we would wish contemporary observations at more time points than 2000 AD, observations at a single time point 2000 AD are the most the writer was able currently to approximate.

In 2000 AD there were the 189 member countries of the United Nations listed in column 2 of Appendix 2. The UN list of countries is a political list. Taiwan is excluded since it is considered part of China. So also, various dependencies such as Puerto Rico, Hong Kong and the like are excluded even though the UN may compile separate data for them. Suppose we posit a Bernoulli random variable X = 1, 0 or a continuous random variable X, $-\infty < X < \infty$. For each country we measure m and X at time t = 2000 AD. Statistically, each conditional observation $X|(m, t)$ will be considered a random realization of X given both m and t.

The UN codes member countries as least developed, less developed (from which we will exclude least developed) and developed. For least developed, less developed and developed the writer used grouped number of specialized occupations w = 2000, 6000 and 18,000 (Denton 1996). Grouped modernization m-values of m = 48.368, 53.135, 57.65 were calculated as $m = g^{-1}(w)$, the non-negative solutions to the equation $0.00138m^2 + 0.09040m - \ln(w) = 0$ of Appendix 1.

How might we make observations of social constructs such as divorce rates of the 189 countries of Appendix 2 column 2? For basic demographic data we will use UN statistics. For most other constructs of significant relevance to "society" no contemporary coded data are currently available by country.

Data about contemporary countries at 2000 AD are few and far between. In the absence of a better data base, the writer used Ember and Ember (2001, abbreviated below as CC). CC is a four-volume encyclopedia describing the cultures of countries of the world at about 2000 AD. If we limit CC entries to UN members at 2000 AD and add Liechtenstein (somehow excluded from CC) we have the 189 countries of Appendix 2 column 2. CC is of marginal use since its descriptions of national cultures are both very terse and of very uneven quality. It seems that country "experts" were unable to meet the informational demands which CC editors (Ember and Ember 2001) made of them. The writer coped with this constraint by imposing three criteria. First, each construct to be coded must be pertinent to the neo-Malinowskian model of society sketched elsewhere by the writer (Denton 1998). Secondly, there must be SCCS coded data for the construct (Divale 2004b). Thirdly, CC country descriptions of the construct must be of usable quality.

In practice, the writer began by pruning the bank of SCCS coded data (Divale 2004b) for conceptually viable societal constructs. All those constructs left for which CC country descriptions were unusable were further pruned. SCCS coding definitions were re-jigged in Appendix 3 to create Bernoulli ($X = 1, 0$) or continuous ($-\infty < X < \infty$) random variables for which behavioral relations models $P(X = 1) = f(m, t)$, $X = h(m, t)$ might be sought by fitting equations to data. Then, CC country descriptions were coded - and subsequently re-coded - by the writer. The result is the coded data of Appendix 2. That coding for Appendix 2 was done solely by the writer detracts from the reliability of the data so obtained. Several, independent coders with measures of intercoder reliability would have been preferable. That coding for Appendix 2 uses CC data of marginal quality detracts from the validity of the data so obtained. As a result, data analyses in this paper must be considered exploratory and provisional.

After having coded the 189 CC societies for the constructs of Appendix 3 writer formed the conclusion that he may have misapplied Murdock and Provost' (1973) coding rules for V151Rec intensive agriculture = 1, 0 (present, absent). This matter is discussed in Appendix 1 section 1.3.

In Appendix 2 "develop" is the UN designation of a country as 1, 2, 3 (least developed, less developed excluding least developed, developed). V1130Rec is SCCS variable 1130 (density) redefined in Appendix 3 to approximate a continuous random variable V1130Rec > 0. The remaining random variables for which coded data appear in Appendix 2 are Bernoulli $X = 1, 0$ coded using the definitions of Appendix 3.

3.3. Combined Data Base of 363 Societies

For each construct to be modeled a sample of 363 observations was obtained by adding the 189 observations of Appendix 2 to the 174 observations of SCCS coded data at or after 1800 AD (from Divale 2004b) converted into the redefined constructs of Appendix 3. In cases where there are missing SCCS data the sample size is correspondingly reduced. There are no missing CC data in Appendix 2.

4. FORECASTS OF SOCIETY TO 2050 AD

In section 4 we will fit parametric logistic regression and regression models P(X = 1) = f(m, t), X = h(m, t) to the data bases of section 3. We will make forecasts by extrapolating the behavioral relations models f(m, t), h(m, t), so discovered, to m and t at 2050 AD. We will also test for monotonic trend in X = 1, 0, $-\infty < X < \infty$ over $m \geq 0$. Such trend will lead to a complementary variety of forecasts. Given the discoveries currently ongoing in biotechnology, the writer is unwilling to attempt forecasts past 2050 AD. The methods of section 4 may be used to make forecasts of any society from the currently least to most developed to any time t from the immediate present to 2050 AD.

The model selection-model checking process of Kutner et al. (2004) was used to fit parametric logistic regression models P(X = 1) = f(m, t), or regression models X = h(m, t), to data. Model selection (selection of predictor variables) proceeded as follows. In order to avoid fitting models which capitalize on chance sampling variation rather than sound behavioral relations, prior behavioral theory is used to suggest candidate predictor variables. Modernization $m \geq 0$ is such a predictor (Denton 2003; 2004a). Any unknown predictors beyond $m \geq 0$ will hopefully (see below) be measured indirectly by a surrogate predictor time t, $1800 \leq t \leq 2000$ AD. Powers up to 3 of m, t and mt are used. This predictor set will detect trend. Behavioral theory of modernization and exploratory data analyses both suggest that there is no reason to believe a cycle is at work beyond increasing-then-decreasing logistic regression or regression curves. Such curves will be detected by the candidate predictor set fitted to data. No forecast will be made unless a curve is successfully fitted to data. Even fitted curves will be subjected to careful scrutiny.

For each Bernoulli random variable the process of fitting equations to data began by graphing X = 1, 0 as a function of m. Based on these results, the writer attempted to fit a logistic regression model P(X = 1) = f(m, t) to all 363 observations – fewer observations where there are missing SCCS data. Predictors investigated included powers up to three of m, t and mt. Backward selection ($\alpha = 0.05$) was used to reduce the candidate predictor set $\mathbf{z}$\` = $<m, m^2, m^3, t, t^2, t^3, mt, m^2t^2, m^3t^3>$ to ones significantly impacting the criterion variable. In order to facilitate checking goodness of fit, and because the m values of the 189 CC countries are grouped, SCCS codes for modernization m and time t were grouped.[7] An intercept model was used where there was good fit even if the p-value of the intercept exceeded 0.05.

The model selection method of the preceding paragraphs would be more convincing with greater variability in time of observation of contemporary countries than a single time t at 2000 AD. A data base with a single contemporary time of 2000 AD is the best the writer was able to attain. Models fitted to the data currently available may be re-fitted when additional data become available. In the meantime, time depth is provided by the SCCS where $1800 \leq t \leq 1965$ AD (Murdock and White 1969).

After a model with the fewest possible number of predictors was found, goodness of fit was checked using all three of Hosmer-Lemeshow, deviance and Pearson residual tests ($\alpha = 0.05$) (Kutner et al. 2004; Hosmer and Lemeshow 2000). Only where a logistic regression model (applied to grouped data) met all three tests was the model retained.

In Table 2 (Page 72) columns 1 – 3 show the subject matter of each of 26 constructs forecast. In column 6 parametric logistic regression models are fitted to 11 of 25 Bernoulli random variables X = 1, 0.[8] Of these a graph of the fitted model for V744Rec (high divorce rate) appears in Figure 1 (page 71) as an example. In all 14 cases where no logistic regression model was fitted the reason was quasi-complete separation of data points preventing unique maximum likelihood (ML) estimation. Inclusion of SCCS data cases predating 1800 AD would reduce quasi-separation of data points. The 12 SCCS societies predating 1800 AD were excluded for the reasons stated above.

In Table 2 population density, V1130Rec > 0, is treated as a continuous random variable.[9] The same modeling process (selection of predictor set followed by checking goodness of fit) was followed for V1130Rec as in the preceding paragraph with the exception that the initial model fitted was the linear regression model ln(V1130Rec) = h(m, t). The attempt to fit a linear regression model was abandoned when a singular matrix prevented making a test for goodness of the model fitted.

In column 4 of Table 2 tests for monotonic trend are shown. Of the 26 constructs forecast 25 are binary constructs X = 1, 0. For each of these a 2-sided Cochran-Armitage (abbreviated CA) test ($\alpha = 0.05$) (Kotz and Johnson 1988:334-336; Agresti 2002) was used to test for monotonic trend. The CA test for trend in a binary construct is hard to beat in cases where values of the predictor are reasonable precise. CA tests for any variety of monotonic trend, not simply linear trend (Kotz and Johnson 1988:334-336). The CA hypotheses tested are

$$H_o : P(X = 1|m) = k, \text{ k is a constant}$$
$$H_1 : P(X = 1|m_j) > P(X = 1|m_i),\ m_j > m_i, \text{ or } P(X = 1|m_j) < P(X = 1|m_i),\ m_j > m_i$$

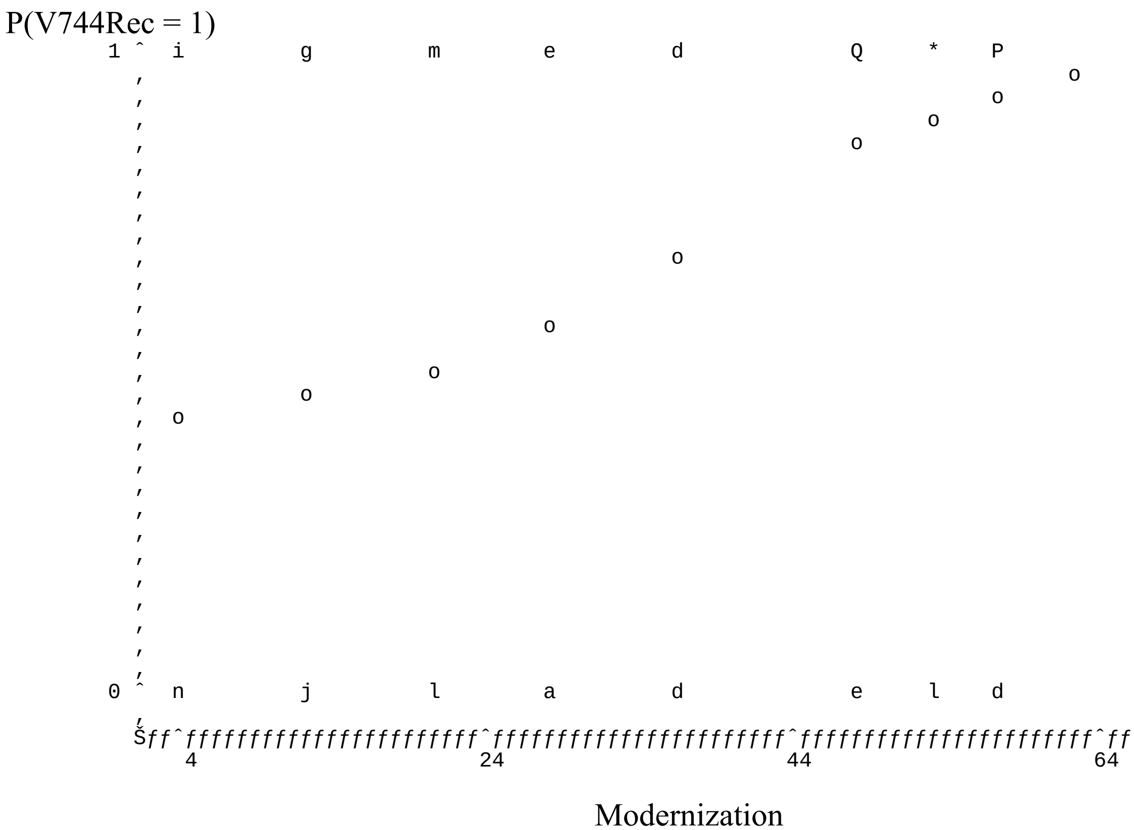
The trend tested is trend over $m \geq 0$. The results are shown in Table 2 column 4. CA tests show increasing trend for 23 Bernoulli random variables in Table 2, decreasing trend for V1648Rec (frequency of war) and a constant process for V121Rec (females predominate in cooking). A graph of observations of V754Rec (wife beating) appears in Figure 2.

In biomedical research, where the CA trend test is used, an experimental design usually controls for extraneous variables thus making the test reasonable even though variables beyond the predictor variable might influence P(X = 1) of a Bernoulli random variable X = 1, 0 under investigation. In the present research a Cochran-Armitage test was performed using modernization m as the sole predictor of trend. Mean modernization $m \geq 0$ also increases over time.

In Table 2 V1130Rec (population density) is treated as a continuous random variable. The test used to detect trend is a 2-sided test ($\alpha = 0.05$) for Spearman's rho.

Forecasts for the 11 constructs for which logistic regression models were fitted[8] might be made as follows. Estimate the values of the predictor(s) at 2050 AD.[10] Put these predictor values into the fitted model.[8] Calculate a model based point prediction and 90% confidence interval (CI). Use the point prediction and 90% CI as a point forecast and 90% prediction interval (PI). The problem with all but one of the fitted models is that the predictor set[8] includes powers and interactions of modernization *m* and time *t*.

Figure 1. Probability that a Society Has a High Divorce Rate (PV744Rec = 1)



Graph of P(V744Rec = 1), the expected logistic regression probability that a society has a high divorce rate, shown by o on the y-axis. Modernization m appears on the x-axis. Number of observations V744Rec = 1, 0 is shown on the y-axis as a letter: 1 - 26 observations by a - z; 27 - 52 observations by A - Z. * denotes 83 observations. The probability graphed at $m = 62$ is the predicted expectation at 2050 AD when $m \approx 62$ (see text). Observations have been grouped into intervals of *m* to facilitate model checking (see text). The model fitted is $P(V744Rec = 1) = f(m^2)$, n = 268.

Figure 2. Graph of Observations of 257 Societies: y-axis Is V754Rec = 1 (Wife Beating Absent or Rare), 0 (Otherwise); x-axis Is Time

```
V754Rec
     1 ˆ                         b    ba aaabaa  aa         *
       ‚
       ‚
       ‚
       ‚
       ‚
       ‚
       ‚
       ‚
       ‚
       ‚
       ‚
       ‚
       ‚
       ‚
       ‚
       ‚
       ‚
       ‚
       ‚
     0 ˆ  a      a     a aa ba b abdb caba efdacg aa        n
       ‚
       Šƒƒˆƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒˆƒƒ
          0                       100                      200

                               Time
```

Number of observations is shown as follows: 1 to 26 by a to z. Star (*) denotes 175. Time is not grouped. Time is scaled so $t = 0$ at 1800 AD.

Table 2. Forecasts to 2050 AD Based on Sample of 363 Societies $0 \le m \le 57.65$, at Times $1800 \le t \le 2000$ AD

Subject	Codes	Forecast	Trend? Yes/No ↑,↓, k	PI Forecast Due to Trend or k to m = 62 at 2050 AD	Parametric model fitted Yes/No/*	90% PI Forecast based on parametric model to m = 62 at 2050 AD
Intercommunity Trade as Food Source	V1Rec 1≥50% food from outside community 0=otherwise	P(V1Rec=1)	Yes ↑ CA	P(P(.)>0.87)>0.90	*	
Credit Source	V18Rec 1=money lending specialists present 0=personal loans between friends or relatives	P(V18Rec=1)	Yes ↑ CA	P(P(.)>0.94)>0.90	Yes LR	Not calculated
Compactness of Settlement	V62Rec 1=compact 0=otherwise	P(V62Rec=1)	Yes ↑ CA	P(P(.)>0.94)>0.90	Yes LR	Not calculated
Community Size	V63Rec 1=average community ≥5000 population 0=otherwise	P(V63Rec=1)	Yes ↑ CA	P(P(.)>0.81)>0.90	*	
Large or Impressive Structures	V66Rec 1=present 0=otherwise	P(V66Rec=1)	Yes ↑ CA	P(P(.)>0.94)>0.90	*	
Form of Family	V68Rec 1=modal family independent 0=otherwise	P(V68Rec=1)	Yes ↑ CA	P(P(.)>0.87)>0.90	Yes LR	Not calculated
Marital Residence	V69Rec 1=neolocal residence predominates 0=otherwise	P(V69Rec = 1)	Yes ↑ CA	P(P(.) > .87) > .90	Yes LR	Not calculated

Table 2. Forecasts to 2050 AD Based on Sample of 363 Societies $0 \leq m \leq 57.65$, at Times $1800 \leq t \leq 2000$ AD (continued)

Subject	Codes	Forecast	Trend? Yes/No ↑,↓, k	PI Forecast Due to Trend or k to m = 62 at 2050 AD	Parametric model fitted Yes/No/*	90% PI Forecast based on parametric model to m = 62 at 2050 AD
Descent-Membership in Corporate Kinship Groups	V70Rec 1=bilateral kinship 0=otherwise	P(V70Rec=1)	Yes ↑CA	P(P(.)>0.84)>0.90	Yes LR	Not calculated
Polygamy	V79Rec 1=monogamy predominates , polygamy rare 0=otherwise	P(V79Rec=1)	Yes ↑CA	P(P(.)>0.94)>0.90	*	
Cooking	V121Rec 1=females predominate 0=otherwise	P(V121Rec=1)	No k CA	P(0.94<P(.)<1)=0.9 0	*	
Smelting	V129Rec 1=present 0=otherwise	P(V129Rec=1)	Yes ↑ CA	P(P(.)>0.94)>0.90	Yes LR	Not calculated
Laundering	V143Rec 1=females predominate 0=otherwise	P(V143Rec=1)	Yes ↑ CA	P(P(.)>0.94)>0.90	*	
Writing and Records	V149Rec 1=true writing present 0=otherwise	P(V149Rec=1)	Yes ↑ CA	P(P(.)>0.94)>0.90	*	
Fixity of Residence	V150Rec 1=sedentary permanent 0=otherwise	P(V150Rec=1)	Yes ↑ CA	P(P(.)>0.94)>0.90	*	
Agriculture	V151Rec 1=mainly primary intensive 0=otherwise	P(V151Rec=1)	Yes ↑ CA	P(P(.)>0.94)>0.90	Yes LR	Not calculated
Technological Specialization	V153Rec 1=smiths, weavers, potters present 0=otherwise	P(V153Rec=1)	Yes ↑ CA	P(P(.)>0.94)>0.90	*	
Money	V155Rec 1=true money present 0=otherwise	P(V155Rec=1)	Yes ↑ CA	P(P(.)>0.94)>0.90	*	

Table 2. Forecasts to 2050 AD Based on Sample of 363 Societies $0 \leq m \leq 57.65$, at Times $1800 \leq t \leq 2000$ AD (continued)

Subject	Codes	Forecast	Trend? Yes/No ↑,↓, k	PI Forecast Due to Trend or k to m = 62 at 2050 AD	Parametric model fitted Yes/No/*	90% PI Forecast based on parametric model to m = 62 at 2050 AD
Political Integration	V157Rec 1=state (above local community) 0=otherwise	P(V157Rec=1)	Yes ↑ CA	P(P(.)>0.94)>0.90	*	
Social Stratification	V158Rec 1=non-egalitarian 0=otherwise	P(V158Rec=1)	Yes ↑ CA	P(P(.)>0.94)>0.90	*	
Food Stress or Hunger	V678Rec 1=food supply constant 0=otherwise	P(V678Rec=1)	Yes ↑ CA	P(P(.)>0.87)>0.90	Yes LR	Not calculated
Frequency of Divorce	V744Rec 1=universal, frequent, common, not uncommon 0=otherwise	P(V744Rec=1)	Yes ↑ CA	P(P(.)>0.81)>0.90	Yes LR	(0.92, 0.97)
Wife-Beating	V754Rec 1=absent or rare 0=otherwise	V(754Rec=1)	Yes ↑ CA	P(P(.)>0.94)>0.90	Yes LR	Not calculated
Adult Mobility	V786Rec 1=movement to other community as adult common or occasional 0=individuals generally attached to particular community throughout life especially after marriage	P(V786Rec=1)	Yes ↑ CA	P(P(.)>0.94)>0.90	*	
Population Density	V1130Rec Use UN 2000 AD density (persons per square km)	V1130Rec	Yes ↑ Sp	P(E[.]>89)>0.90	**	

Table 2. Forecasts to 2050 AD Based on Sample of 363 Societies $0 \leq m \leq 57.65$, at Times $1800 \leq t \leq 2000$ AD (continued)

Subject	Codes	Forecast	Trend? Yes/No ↑,↓, k	PI Forecast Due to Trend or k to $m = 62$ at 2050 AD	Parametric model fitted Yes/No/*	90% PI Forecast based on parametric model to $m = 62$ at 2050 AD
Overall Frequency of War	V1648Rec 1=war once each 3-10 years or more frequently 0=war absent or rare	P(V1648Rec=1)	Yes ↓ CA	P(P(.)<0.22)>0.90	Yes LR	Not calculated
Presence of Formal Education within Local Community	V1738Rec 1=formal education present at least for some 0=absent	P(V1738Rec=1)	Yes ↑ CA	P(P(.)>0.94)>0.90	*	

See Appendix 2. for the substantive meaning of the Variables of column 1. Meanings of abbreviations follow in parentheses: LR (logistic regression), *m* (measure of modernization outlined in Appendix 3.), CA (Cochran-Armitage trend test, Sp (Spearman rho trend test), k (constant), PI (prediction interval), * (quasi-separation of points prevented fitting a LR model owing to indeterminate maximum likelihood estimates), ** (a singular matrix prevented tests on a regression model). All trend tests are 2-sided $\alpha = 0.05$. Where tests show no trend a constant k is assumed to be implied. See text for methods used to fit models and method of calculating PIs. Fitted LR models are stated in the text. Probability prediction intervals are to two decimal places. Where the phrase, "Not calculated" appears it denotes that the predictors in a fitted logistic regression model included powers of time *t* or interactions of time *t* and modernization *m*. Such predictors made the writer reluctant to calculate an "out-of-sample" PI. Such fitted models are evidence of monotonic trend. V1130Rec PI is rounded to an integer. LR point forecasts may be calculated from models stated in the text.

In curve fitting the curve fitted is valid only over the range of the predictors for which the model is fitted (Kutner et al. 2004). It might be reasonable to make a model based forecast where the predictor is close to the interval of the predictor for which the model is fitted. This is done in Table 2 for V744Rec where the predictor is in *m* only. Denton (1996) estimates the number of specialized occupations at 2050 AD to be 50,000 or more from which $m \geq 62$ is estimated by solving for the positive root of the equation $0.00138m^2 + 0.09040m - \ln(50{,}000) = 0$ from Appendix 1. The forecast for V744Rec is made by putting $m = 62$ into the fitted[8] logistic regression model $P(X = 1) = [1 + \exp(-\mathbf{b}`\mathbf{m})]^{-1}$. A 90% prediction interval (PI) forecast appears in Table 2 column 7. Forecasts to $m < 62$ may also be made by the methods of this paragraph.

Many of the 11 logistic regression curves $P(X = 1) = f(\mathbf{z}`)$ fitted in Table 2 may suffer from overfitting to the candidate predictor set $\mathbf{z}` = \langle m, m^2, m^3, t, t^2, t^3, mt, m^2t^2, m^3t^3 \rangle$. Overfitting occurs when a fitted curve reflects chance variation in the predictors rather than behavioral relations between predictors and criterion variable. In Appendix 1 gross overfitting of a regression curve is corrected by applying jackknife tests to standard definitions of error sums of squares (Latin et al. 2003: 73; Kutner et al. 2004:360). Currently, testing for overfitting in logistic regression models is not as straightforward as for regression models. There are jackknife diagnostics available for logistic regression but these are not as definitive as jackknife regression tests. Hence, overfitting of the logistic regression curves fitted in Tables 2 and 3 is an unresolved issue.

As an alternative to simply putting m = 62, t = 2050 into the fitted logistic regression equation one might make a forecast using the linear approximation (Spiegel 1981) estimator $P(X = 1) = P(57.65, 200) + \partial P/\partial m|_{57.65,200}\,(62 - 57.65) + \partial P/\partial t|_{57.65,200}\,(250 - 200)$. Such a forecast might be made for all 11 fitted models of Table 2. Such a forecast method seems reasonable even where the predictors are in powers and interactions of m and t.

The reader is reminded that the model of Appendix 1 - $0.00138m^2 + 0.09040m - \ln(w) = 0$ - is fitted to only 24 data cases. Model-based confidence intervals are calculated in Appendix 1. These confidence intervals are so large that the forecasts of Table 2 are not based on them. Efforts (outlined in Appendix 1) to use tangent line approximation to estimate m lead the writer to the opinion that, at 2050 AD, $m \approx 62$ estimated as the non-negative solution to the model $0.00138m^2 + 0.09040m - \ln(50000) = 0$ outlined in Appendix 1, is not out of line.

Forecasts based on the monotonic trend models of Table 2 might be made as follows. For the Cochran-Armitage trend models a forecast to 2050 AD is made as follows. The methods of this paper may be used to make forecasts to any future time t from any base modernization m. We will make forecasts for societies currently designated by the United Nations (2005) as "developed." Let the superscript "```," as in proportion p```, modernization m```, refer to most developed countries at 2000 AD. The 2000 AD proportion of developed societies p``` = 1/n``` $\Sigma I_{X|(m```,t)}$, $I_{X|(m```,t)}$ is the Indicator function of X|(m```, t), n``` = 46, the proportion p``` is viewed as a random realization of an underlying Bernoulli probability parameter θ|(n = 46, p```, m```, t = 2000 AD) for which a 90% binomial confidence may be calculated. Such a binomial confidence interval may be used where the predictor trend variable is m or t. If a Cochran-Armitage test shows increasing trend a PI forecast is that the probability that θ|(m > 57.65) > lower 90% confidence interval limit for θ|(m = 57.65) is greater than 90%. If a CA test shows decreasing trend a PI forecast is that the probability that θ|(m > 57.65) < upper 90% confidence interval limit for θ|(m = 57.65) is greater than 90%. If a CA test shows no trend, i.e. a constant value k, a PI forecast is that the probability that θ|(m > 57.65) falls within the 90% confidence interval limits for θ|(m = 57.65) is equal to 90%. Such forecast PIs appear in Table 2 column 5; they are conditioned on the results of the CA tests.

A forecast based on a fitted Spearman rho test for trend may be made using the method of the preceding paragraph for forecasts based on a fitted CA test for trend.

In many instances there is extraneous, conceptual or empirical evidence supporting the forecasts of Table 2. As examples, there is extraneous conceptual information that neolocal residence (V69Rec) and bilateral kinship (V70Rec) are associated with industrial economies

(Ember and Ember 2004). There is extraneous empirical information from the CC encyclopedia (Ember and Ember 2001) which reports many instances where there is a trend toward neolocal residence and bilateral kinship in developing countries.

The same process by which Table 2 was constructed was repeated to construct Table 3 for the 189 CC countries where $m = 57.65$, $t = 2000$ AD.[11]

Table 3. Forecasts to 2050 AD Based on Sample of 189 Societies $48.368 \leq m \leq 57.65$ at 2000 AD

Subject	Variable	Forecast	Trend? Yes/No ↑,↓, k	PI Forecast Due to Trend or k to m = 62 at 2050 AD	Parametric model fitted Yes/No/*	90% PI Forecast based on parametric model to m = 62 at 2050 AD
Intercommunity Trade as Food Source	V1Rec 1≥50% food from outside community 0=otherwise	P(V1Rec=1)	Yes ↑ CA	P(P(.>0.87)>0.90	Yes	(0.99, 1.0)
Credit Source	V18Rec 1=money lending specialists present 0=personal loans between friends or relatives	P(V18Rec=1)	No k All X=1	P(0.94<P(.)<1)=0.90	All X=1	
Compactness of Settlement	V62Rec 1=compact 0=otherwise	P(V62Rec=1)	No k CA	P(0.94<P(.)<1)=0.90	*	
Community Size	V63Rec 1=average community ≥ 5000 population 0=otherwise	P(V63Rec=1)	Yes ↑ CA	P(P(.)>0.81)>0.90	Yes	(0.97, 1.0)
Large or Impressive Structures	V66Rec 1=present 0=otherwise	P(V66Rec=1)	No k All X=1	P(0.94<P(.)<1)=0.90	All X=1	
Form of Family	V68Rec 1=modal family independent 0=otherwise	P(V68Rec=1)	Yes ↑ CA	P(P(.)>0.87)>0.90	Yes	(0.93, 0.99)

Table 3. Forecasts to 2050 AD Based on Sample of 189 Societies 48.368 ≤ m ≤ 57 .65 at 2000 AD (continued)

Subject	Variable	Forecast	Trend? Yes/No ↑,↓, k	PI Forecast Due to Trend or k to *m* = 62 at 2050 AD	Parametric model fitted Yes/No/*	90% PI Forecast based on parametric model to *m* = 62 at 2050 AD
Marital Residence	V69Rec 1=neolocal residence predominates 0=otherwise	P(V69Rec=1)	Yes ↑ CA	P(P(.)>0.87)>0.90	Yes	(0.98, 1.0)
Descent-Membership in Corporate Kinship Groups	V70Rec 1=bilateral kinship 0=otherwise	P(V70Rec=1)	Yes ↑ CA	P(P(.)>0.84)>0.90	No LR fit	
Polygamy	V79Rec 1=monogamy predominates , polygamy rare 0=otherwise	P(V79Rec=1)	Yes ↑ CA	P(P(.)>0.94)>0.90	*	
Cooking	V121Rec 1=females predominate 0=otherwise	P(V121Rec=1)	No k All X=1	P(0.94<P(.)<1)=0.9 0	All X=1	
Smelting	V129Rec 1=present 0=otherwise	P(V129Rec=1)	No k All X=1	P(0.94<P(.)<1)=0.9 0	All X=1	
Laundering	V143Rec 1=females predominate 0=otherwise	P(V143Rec=1)	No k All X=1	P(0.94<P(.)<1)=0.9 0	All X=1	
Writing and Records	V149Rec 1=true writing present 0=otherwise	P(V149Rec=1)	No k All X=1	P(0.94<P(.)<1)=0.9 0	All X=1	
Fixity of Residence	V150Rec 1=sedentary permanent 0=otherwise	P(V150Rec=1)	No k All X=1	P(0.94<P(.)<1)=0.9 0	All X=1	
Agriculture	V151Rec 1=mainly primary intensive 0=otherwise	P(V151Rec=1)	Yes ↑ CA	P(P(.)>0.94)>0.90	*	
Technological Specialization	V153Rec 1=smiths, weavers, potters present 0=otherwise	P(V153Rec=1)	No k All X=1	P(0.94<P(.)<1)=0.9 0	All X=1	

Table 3. Forecasts to 2050 AD Based on Sample of 189 Societies 48.368 ≤ m ≤ 57 .65 at 2000 AD (continued)

Subject	Variable	Forecast	Trend? Yes/No ↑,↓, k	PI Forecast Due to Trend or k to m = 62 at 2050 AD	Parametric model fitted Yes/No/*	90% PI Forecast based on parametric model to m = 62 at 2050 AD
Money	V155Rec 1=true money present 0=otherwise	P(V155Rec=1)	No k All X=1	P(0.94<P(.)<1)=0.90	All X=1	
Political Integration	V157Rec 1=state (above local community) 0=otherwise	P(V157Rec=1)	No k All X=1	P(0.94<P(.)<1)=0.90	All X=1	
Social Stratification	V158Rec 1=non-egalitarian 0=otherwise	P(V158Rec=1)	No k All X=1	P(0.94<P(.)<1)=0.90	All X=1	
Food Stress or Hunger	V678Rec 1=food supply constant 0=otherwise	P(V678Rec=1)	Yes ↑ CA	P(P(.)>0.87)>0.90	No LR fit	
Frequency of Divorce	V744Rec 1=universal, frequent, common, not uncommon 0=otherwise	P(V744Rec=1)	No k CA	P(0.81<P(.)<097)=0.90	**	
Wife-Beating	V754Rec 1=absent or rare 0=otherwise	V(754Rec=1)	No k CA	P(0.94<P(.)<1)=0.90	*	
Adult Mobility	V786Rec 1=movement to other community as adult common or occasional 0=individuals generally attached to particular community throughout life especially after marriage	P(V786Rec=1)	No k All X=1	P(0.94<P(.)<1)=0.90	All X=1	

Table 3. Forecasts to 2050 AD Based on Sample of 189 Societies 48.368 ≤ m ≤ 57 .65 at 2000 AD (continued)

Subject	Variable	Forecast	Trend? Yes/No ↑,↓, k	PI Forecast Due to Trend or k to m = 62 at 2050 AD	Parametric model fitted Yes/No/*	90% PI Forecast based on parametric model to m = 62 at 2050 AD
Population Density	V1130Rec Use UN 2000 AD density (persons per square km)	V1130Rec	Yes ↑ Sp	P(E[.]>89)>0.90	***	
Overall Frequency of War	V1648Rec 1=war once each 3-10 years or more frequently 0=war absent or rare	P(V1648Rec=1)	Yes ↑ CA	P(P(.)>0.04)>0.90	*	
Presence of Formal Education within Local Community	V1738Rec 1=formal education present at least for some 0=absent	P(V1738Rec=1)	No k All X=1	P(0.94<P(.)<1)=0.90	All X=1	

See Appendix 2. for the substantive meaning of the Variables of column 1. Meanings of abbreviations follow in parentheses: LR (logistic regression), *m* (measure of modernization outlined in Appendix 3.), CA (Cochran-Armitage trend test, Sp (Spearman rho trend test), k (constant), (PI (prediction interval), * (quasi-separation of points prevented fitting a LR model owing to indeterminate maximum likelihood estimates). ** (backward selection removed all predictors from a LR model) *** (a singular matrix prevented tests on a regression model). All trend tests are 2-sided $\alpha = 0.05$. Where tests show no trend a constant k is assumed to be implied. The phrase, "All X = 1" denotes that all data cases were code X = 1; no test was performed but a 90% binomial confidence interval was calculated and used as a PI. See text for methods used to fit models and method of calculating PIs. Fitted LR models are stated in the text. Probability prediction intervals are to two decimal places. V1130Rec PI is rounded to an integer. LR point forecasts may be calculated from models stated in the text.

It is an accepted rule that models fitted to a data set are valid only for the range of predictors observed in the data set (Kutner et al. 2004). All 26 forecasts of Tables 2, 3 extrapolate effects outside the range of predictors of the fitted models. This is a defect of the method of this paper. It is true that there are relatively small increments to modernization *m* (62 versus the fitted range $0 \leq m \leq 57.65$) and time *t* (250 at 2050 AD versus the fitted range $0 \leq t \leq 200$ from 1800 to 2000 AD). Nevertheless, forecasts based on powers and interactions of predictors warrant caution. Steps must be taken to avoid invalid forecasts. It is to such steps that we now turn.

Of the subject matters forecast in Tables 2, 3 one cries out for clarification – V1648Rec Frequency of War. For V1648Rec the conservative, conditional (on monotonic trend) 90%

trend forecasts of Tables 2, 3 column 3 are in opposite directions leading to a PI P(P(V1648Rec = 1) < .22) > .90 in Table 2 and a PI P(P(V1648Rec = 1) > .04) > .90 in Table 3. The long run trend is for decreased war. The short run trend is for increased war.

We have up to six varieties of information on which to base forecasts: The first and second varieties of information are parametric models fitted over both $48.368 \leq m \leq 57.65$, $t = 200$ and $0 \leq m \leq 57.65$, $0 \leq t \leq 200$. In Appendix 1 it is shown that the candidate predictor set **z**`= $<m, m^2, m^3, t, t^2, t^3, mt, m^2 t^2, m^3 t^3>$ may result in parametric linear models which arise from overfitting rather than from behavioral relations. Some of the fitted logistic regression models in Tables 2 and 3 may result from overfitting. Currently, testing for overfitting of logistic regression models is more complicated than testing for overfitting of regression models. Where there is overfitting a simpler model in powers 1 to 3 of *m* should be sought. Latin et al. (2003:71-75) and Kutner et al. (2004:353-361, 591-601) discuss overfitting and summarize diagnostics and remedial measures.

The third and fourth varieties of information to use in forecasting are Cochran-Armitage and Spearman tests for trend over both $48.368 \leq m \leq 57.65$, $t = 200$ and $0 \leq m \leq 57.65$, $0 \leq t \leq 200$.

The fifth variety of information on which to base forecasts is extraneous, conceptual or empirical information such as the known theoretical and empirical connection of neolocal residence to industrialization.

Finally, if we have fitted parametric model(s) we may extrapolate total differentials (Spiegel 1981) by calculating partial derivatives at or near $m = 57.65$ and $t = 200$. Maxima and minima near $m = 57.65$ and $t = 200$ may also be examined.

If all available varieties of information converge to approximately the same forecast, such a forecast seems well founded. In the writer's opinion, forecasts based on Cochran-Armitage or Spearman trend tests are the safest to use.

In Tables 2 and 3 parametric model forecasts are not calculated where the predictors are in powers of *m* or *t* or are interactions of *m* and *t*. Powers and interactions exaggerate the degree to which forecast predictor values are outside the range of values for which the models are fitted. Nevertheless, such parametric forecast predictions are information which the forecaster might use. For the reasons stated above models in powers and/or interactions of time *t* ought to be checked for overfitting and a simpler model in *m* used if overfitting so justifies. The forecast methods of this paper are not automatic; they demand thought and judgment.

From Tables such as 2, 3 a variety of forecasts may be considered. For example, one might consider forecasts such as: trend (increase or decrease) in P(X = 1), certainty or near certainty (P(X = 1) ≈ 1, P(X = 1) ≈ 0)), equal likelihood (P(X = 1) ≈ P(X = 0) ≈ 0.5), relative probability (P(X = 1) > (<) P(X = 0)) or even approximate, 90% (or other %) parametric model-based PIs. Just as in medical research, all such forecasts may find use.

5. DISCUSSION

New, behavioral relations methods have been created in this paper for forecasting society up to about 2050 AD. Forecasts are made by extrapolating a fitted behavioral relations model to future modernization m and time t at times up to 2050 AD (rescaled $t = 250$). Parametric logistic regression and regression models were fitted to data. Cochran-Armitage and Spearman rho trend models were also fitted to data. We get modernization $m = g^{-1}(w)$ at time t using the known connection (Appendix 1) of $m \geq 0$ to number of specialized occupations $w = 0, 1, 2, \ldots$. Internal checks were developed to enhance the validity of "out of sample" forecasts to $m = 62$ at $t = 250$ (2050 AD) slightly beyond the domain $0 \leq m \leq 57.65$, $0 \leq t \leq 200$ (1800 AD $\leq t \leq$ 2000 AD) of the predictors used to fit the model to data.

In order to fit models to data two samples were used. The first sample consisted of 363 preindustrial and contemporary societies, $0 \leq m \leq 57.65$, with an effective time domain of $1800 \leq t \leq 2000$ rescaled to $0 \leq t \leq 200$. The second sample, a subset of the first, consisted of 189 least developed, less developed (excluding least developed) and developed societies ($48.368 \leq m \leq 57.65$) at 2000 AD (rescaled to $t = 200$).

The forecast methods created here are rooted in behavioral relations connecting the degree of modernization $m \geq 0$ of a society to the sorts of behaviors likely to occur in it (Denton 2004a). The degree of modernization of a society is defined - à la Murdock and Provost (1973) - as the recency of the form of the society as measured by Table 1 updated by Appendix 1. By this definition, societies ancient in time may be more modern than recent societies rating lower on Table 1 updated by Appendix 1. Reflection on the nature of the subscales of Table 1 and on the correlates of division of labor (Denton 1996) may convince the reader that the methods of this paper have merit. The methods work because the impacts of the components of modernization are momentous. Modernization is not a single construct (Denton 2004a). Modernization is best thought of a process whereby correlated constructs increase together. When any one of the 10 subscales of Table 1 increases, the conditional expectation of each of the others increases, conditioned on the first. This process continues today with the proviso that demographic constructs of Table 1 such as population density and mean community size are not good measures of modernization in the industrial era. In order to apply to the contemporary world Murdock and Provost' (1973) Table 1 components of preindustrial modernization must be extended to apply to the contemporary world. For example, Table 1 subscale 5 Technological Specialization becomes number of specialized occupations. Table 1 subscale 7 Money becomes per capita GDP.

Currently, human behavior not leaving interpretable clear material remains is not archaeologically observable (Fagan and DeCorse 2005). As a result, time series forecasts (Chatfield 2001) - extrapolating a model of a time series of observations - appear currently to be problematic for most subject matters a cultural anthropologist might wish to model.[12] For many subject matters forecasts based on a method of behavioral relations appear to be a viable alternative.

A picture of life in the most modernized societies at 2050 AD emerges. For the most part, the trends which distinguish developed societies from less developed and least developed societies today are the same trends – carried forward in time - which will distinguish modernized societies at 2050 AD from developed societies today; "much like today, only more so." There will be state political organization, social classes, greater division of labor, intensification of agriculture, new forms of monetary exchange, etc.

The methods of this paper are suited to detect trends correlated with modernization. Consequently, where a subject matter is not correlated with modernization the methods of this paper are not applicable. Whether or not a particular subject matter is correlated with modernization is left to data to decide, subject to guidance by theoretical conceptualization.

Quantitative methods for modeling the great transitions in societies from least to most modern are developed in this paper. Many of the forecasts of Tables 2 - 3 are what the reader might intuitively expect, as with the presence of large impressive structures (V66Rec). Some forecasts of Table 2, such as of V149Rec – V158Rec, are of constructs known to be characteristics of modern societies. Some of the models fitted in Table 2 are the first quantitative evidence for some of the narrative (i.e., in words) trends which many cultural anthropologists (Levinson and Malone 1980:28, 32, 296) have claimed to find in preindustrial societies. Where data permitted, such trends were quantitatively tested here with a database which includes contemporary societies. Other forecasts of Table 2 are by no means obvious as with divorce rates, food stress and war. The methods of this paper may be applied to many new subject matters with results that may not be obvious.

This paper has numerous applications to business and public policy. The constructs of Tables 2 - 3 have implications for understanding social problems of the future. Where future social problems are forecast mitigation might be sought. Here, food stress (V678Rec), wife beating (V754Rec) and war (V1048Rec were investigated. If a goal of policy analysis is to describe the future characteristics of society then a Table such as 2 or 3 has much to say about gender roles (V121Rec, V143Rec), political organization (V157Rec) and many matters describing how people will make exchanges to meet their basic needs (Denton 1998). There appear to be many potential applications of this paper to business. Characteristics of commodities, bases of exchange (Denton 1998), sectors of demand and the like may all be forecast, provided a model is successfully fitted to a suitable data base of observations of meaningful constructs.

The criticisms which may be made of cross-cultural research are not, in the writer's opinion, sufficient grounds to discard the opportunities such data offer to forecasting. Most coded data for the SCCS are modal values for societies, for example whether the divorce rate of a society is high (X = 1) or low (X = 0). Many modal values for contemporary countries appear in Appendix 2. Such measures of central tendency obscure the dispersion about the central tendency which may be found in any society. This criticism seems an opportunity to improve, rather than abandon, the approach outlined here.

Beyond getting measures of both central tendency and dispersion, how might the forecasting approach of this paper be improved? Better, or simply different, measures of modernization are needed applicable to both preindustrial and industrial societies. It would be helpful to obtain a larger sample of number of specialized occupations to estimate $m = g^{-1}(w)$. It would be reasonable to fit models to contemporary data with predictor $w = 0, 1, 2, \ldots$ rather than $m \geq 0$. On the other hand, since most SCCS societies are not coded for $w = 0, 1, 2, \ldots$, to omit using $m \geq 0$ would necessitate dropping use of the SCCS.

Since the model $w = g(m)$ is fitted in Appendix 1 to a sample of only 24 societies conservative forecasts might make the following precautions. Give more credence to CA and Spearman rho trend forecasts. Construe parametric model-based forecasts $P(X = 1) = f(m, t)$, $X = h(m, t)$ as simply directional changes when m, t increase.

Beyond finding a better, or simply different, measure of modernization spanning preindustrial and industrial societies the main need is for the systematic collection of conceptually meaningful data by country. Beyond excellent demographic data there are few United Nations data pertinent to society. World Bank World Development Indicators (WDI) are data created to meet the needs of economic development. Available electronically (http://www.worldbank.org/data/countrydata/countrydata.html at the time of writing) or in hard copy, these data are of limited use as descriptors of society. Kurian (2001) is a useful collection of data but is now out of date. Inglehart (2004; author of many related works) has created world attitudinal databases by country. The NationMaster.com website (http://www.nationmaster.com/ at the time of writing) assembles contemporary data, by country on diverse subject matters much like those of Kurian (2001).

There appear to be some uncollated, survey sample data available in individual, developed countries. Consider, for example, kinship relatedness V70Rec = 1 (bilateral kinship) or 0 (descent based kinship). It might be possible to assemble results of sample surveys where kinship is observed in most of the currently developed countries. If number of specialized occupations $w = 0, 1, 2, \ldots$ is estimated for each developed country then gradations in $m \geq 0$ will become discernible. Finer measures of $m \geq 0$ (or $w = 0, 1, 2, \ldots$) may result in improvement to the models fitted in this paper where all developed countries are grouped together as $m``` = 57.65$. A model such as P(V70Rec) = f(w) (for bilateral kinship) might be fitted to such data. While it would be expensive to code SCCS societies for a new construct X (binary, multinomial or continuous) this might be considered where funds are available.[13]

It seems preferable to retain preindustrial societies in any data base to which models are fitted. However, investigators might opt to gather data over a narrow span of contemporary m (or w, or simply development d) and time t. While forecasts based on both preindustrial and contemporary data are compelling, forecasts resting only on contemporary data might be attempted. This is done here in Table 3. Since the forecast methods of this paper are non-automatic, careful thought and judgment must be used.

6. NOTES

1. The research reported in this paper was begun when the writer was a faculty member at Brandon University. Data analyses were designed and executed by the writer using SAS V8.2 supplied by Brandon University to facilitate completion of the research. Excel 2003 was used for Appendix 2. A Windows XP-Pro platform was used.

2. Denton (2004a) examines measurement properties of the Table 1 M-scale. It is shown that each of the 10 subscales of Table 1 is an exponential measure of the underlying construct it measures.

3. What is worth forecasting? The answer depends on the goals of the forecaster. If a forecaster wishes to focus on the needs of business, or public policy, or idiosyncratic interests then each of these goals may suggest different subject matters deserving forecasts. The writer's goal is to forecast how society will work. The key construct in this regard is "society."

"Society" is defined to be, "a set of individuals who are interdependent in regard to meeting basic zoological needs (for food, shelter, clothing, safety, mobility, sex, health, learning, leisure), but are relatively independent of individuals outside this set." This is a standard, anthropological definition of society (Malinowski 1988). Even today a society may operationally be equated with independent political unit. A country such as France may be considered to be a society. A politically independent "Bushmen" band of hunters and gatherers from the 1950's may also be considered to be a society. The definition of "society" accounts for its usefulness as a unit of analysis and observation. If a set of individuals is interdependent in regard to how they meet basic zoological needs, they will develop common customs which facilitate meeting these needs (Denton 1998). A focus on "societies" will determine the sorts of subject matters we will be able to forecast by the methods of this paper.

4. Denton (2003) shows that SCCS cultural units may be treated as societies.

5. In the data analyses which follow, the 12 SCCS societies predating 1800 AD (Denton 2004b) are excluded. The reason for excluding these 12 societies is the hope of fitting models simpler than might result from a time depth extending back to 1750 BC. Also, 12 societies spanning the era [1750 BC, 1759 AD] are insufficient to reflect that era. One might nevertheless consider using the methods of this paper with societies extending back in time beyond 1800 AD. Inclusion of the 12 SCCS peoples predating 1800 AD would facilitate model fitting by reducing occasional quasi-complete separation of data cases in fitting logistic regression models.

6. For social science research where contemporary countries are the unit of observation see Jackman (1985), Kohn (1987).

7. For the need to group data where a logistic regression predictor is continuous see, *inter alia*, Hosmer and Lemeshow (2000). Here are the SCCS groupings used for modernization *m*

and time t. Time is recoded to start at 1800 AD by subtracting 1800 from the 1800 – 1965 AD time of ethnographic observation.

Modernization m: [0, 8] → 4, [9, 16] → 12.5, [17, 24] → 20.5, [25, 32] → 28.5, [33, 40] → 36.5

Time t: [0, 25] → 12, [25, 50] → 37, [50, 75] → 62, [75, 100] → 87, [75, 99] → 87, [100, 124] → 112, [125, 165] → 146

In the preceding groupings the arrow → points to the group value given to a bracketed interval of m or t. Brackets [,] include the interval boundaries shown. Example: If $0 \leq m \leq 8$ grouped $m = 4$. Brackets (,] include the upper boundary but exclude the lower boundary. Example: If $25 < d \leq 50$ grouped $d = 37.5$. All contemporary countries are coded at 2000 AD which, upon subtracting 1800, becomes 200.

8. A logistic regression model of a Bernoulli random variable X = 1, 0 is defined to be $P(X = 1) = [1 + \exp(-\mathbf{b`z})]^{-1}$, **b** is the vector of coefficients, **z** is the vector of predictors, **b`** is the transpose of the coefficient vector **b**. It may be shown that the natural logarithm of the odds ratio $\ln(P/(1 - P)) = \mathbf{b`z}$, which makes **b`z** interpretable. The odds ratio interpretation of **b`z** does not hold where interaction predictors are fitted, for example mt and powers thereof.

For V1Rec, V63Rec, V66Rec, V79Rec, V121Rec, V143Rec, V149Rec, V150Rec, V153Rec, V155Rec, V157Rec, V158Rec, V786Rec and V 1738Rec quasi-complete separation of data cases prevented using maximum likelihood estimation to fit a logistic regression model. Inclusion of data cases preceding 1800 AD would reduce quasi-complete separation of data cases mentioned in the text. In the case of V1130Rec the attempt to fit a regression model was abandoned when a singular matrix prevented using a test to assess goodness of fit.

The models which follow were all fitted to the combined sample of 363 societies, minus any missing SCCS cases. For V62Rec the following model was fitted: $P(X = 1) = [1 + \exp(0.4991 - 0.3287mr + 0.0179mr^2 - 0.000331mr^3 + 0.00069mrfr)]^{-1}$, b_o, b_1, b_2, b_3, b_4 :α = 0.4184, 0.0117, 0.0046 , 0.0005, 0.0174; c = 0.878, R-Square = 0.3372. Hosmer-Lemeshow = 0.9537, Deviance = 0.8802, Pearson = 0.9636; n = 363.

For V68Rec the following model was fitted: $P(X = 1) = [1 + \exp(1.0798 - 0.1737mr + 0.0109mr^2 - 0.000146mr^3 - 2.33\text{E-}7fr^3)]^{-1}$, b_o, b_1, b_3, b_4 :α = 0.0684, 0.0321, 0.0014 , 0.0001, 0.0410; c = 0.750, R-Square = 0.1832; Hosmer-Lemeshow = 0.3619, Deviance = 0.7451, Pearson = 0.8105; n = 363.
For V69Rec the following model was fitted: $P(X = 1) = [1 + \exp(-0.4945 + 0.1683fr - 0.00191fr^2 + 6.56\text{E-}6fr^3 - 7.98\text{E-}12\text{mr}^3fr^3)]^{-1}$, b_o, b_1, b_2, b_3, b_4: α = 0.7830, 0.0349, 0.0257, 0.0109, < 0.0001; c = 0.912, R-Square = 0.4547; Hosmer-Lemeshow = 0.7902, Deviance = 0.8913, Pearson = 0.5450; n = 362.

For V70Rec the following model was fitted: $P(X = 1) = [1 + \exp(0.6816 - 0.000062mr^3 - 0.000711fr^2 + 3.83\text{E-}6fr^3 + 0.00301mrfr - 4.59\text{E-}7mr^2fr^2 + 219\text{E-}13mr^3fr^3)]^{-1}$, $b_o, b_1, b_2, b_3, b_4, b_5, b_6, b_7 : \alpha = 0.2090, 0.0326, < 0.0001, = 0.0004, < 0.0001, = 0.0008, 0.0029$; c = 0.797; R-Square = 0.2610; Hosmer-Lemeshow = 0.5580, Deviance = 0.1595, Pearson = 0.3153; n = 363.

For V129Rec the following model was fitted: $P(X = 1) = [1 + \exp(4.4122 - 0.0212mr^2 + 0.00047mr^3 + 2.27\text{E-}7mr^2fr^2 - 3.98\text{E-}11mr^3fr^3)]^{-1}$, $b_o, b_1, b_2, b_3, b_4 : \alpha = <0.0001, < 0.0001, = 0.0001, 0.0090, 0.0004$; c = 0.964; R-Square = 0.5715; Hosmer-Lemeshow = 0.9330, Deviance = 0.5842, Pearson = 0.7469; n = 363.

For V151Rec the following model was fitted: $P(X = 1) = [1 + \exp(0.0232mr^2 - 0.000642mr^3 + 0.0518fr - 0.00581mrfr + 6.64\text{E-}7mr^2fr^2)]^{-1}$, $b_1, b_2, b_3, b_4 : \alpha = < 0.0001$; c = 0.897; R-Square = 0.4569; Hosmer-Lemeshow = 0.9557, Deviance = 0.1461, Pearson = 0.0992; n = 363.

For V678Rec the following model was fitted: $P(X = 1) = [1 + \exp(0.7878 - 2{,}26\text{E-}12\ mr^3fr^3)]^{-1}$, $b_o, b_1 : \alpha < 0.0001$, .; c = 0.815, R-Square = 0.2829; Hosmer-Lemeshow = 0.5463, Deviance = 0.5553, Pearson = 0.7250; n = 320.

For V744Rec the following model was fitted: $P(X = 1) = [1 + \exp(0.3467 - 0.000856mr^2)]^{-1}$, $b_o, b_1 : \alpha = 0.1779, < 0.0001$; c = 0.754; R-Square = 0.1672; Hosmer-Lemeshow = 0.9041, Deviance = 0.5929, Pearson = 0.6208; Odds ratio: $mr^2fr^2 = 1.001$; n = 268.

For V754Rec the following model was fitted: $P(X = 1) = [1 + \exp(2.8670 - 6.665\text{E-}7fr^3)]^{-1}$, $b_o, b_1 : \alpha < 0.0001$, .; c = 0.864, R-Square = 0.3897; Hosmer-Lemeshow = 0.1085, Deviance = 0.4584, Pearson = 0.3826; Odds ratio: $fr^3 = 1.000$; n = 257.

For V1648Rec the following model was fitted: $P(X = 1) = [1 + \exp(-4.7004 - 0.00378mr^2 + 0.0395fr + 8.79\text{E-}8mr^2fr^2)]^{-1}$, $b_o, b_1, b_2, b_3 : \alpha < 0.0001, = 0.0002, < 0.0001, = 0.0004$; c = 0.899, R-Square = 0.4174; Hosmer-Lemeshow = 0.2203, Deviance = 0.7408, Pearson = 0.9262; n = 337.

In fitting models for V149Rec – V158Rec the global predictor $m_j = \Sigma s_i$, $i = 1, 2, \ldots, 9$, $i \neq j$, was calculated over all subscales of Table 1 with the exception of the subscale being modeled. The resulting m_j was then multiplied by 10/9 to give it the same range $0 \leq m \leq 40$ used to fit the remainder of the models of Tables 2 and 3.

9. Monaco was deleted from model fitting as an outlier where density is 22,403 per square kilometer. Monaco is also deleted from model fitting as an outlier in Table 3.

10. Exclusion of time t from a model such as Bernoulli P(X = 1) = f(m) or continuous X = h(m) means one of two things: either time t has no impact on P(X = 1), X or the power of the test is insufficient to detect the impact. While power analysis for linear regression is reasonably well understood (Cohen 1988) it is rudimentary for logistic regression (Hsieh 1989). However, by limiting the forecast horizon to 2050 AD extrapolation of a model P(X = 1) = f(m) when the true model is P(X = 1) = f(m, t) will reduce the impact.

11. For V18Rec, V66Rec, V121Rec, V129Rec, V143Rec, V149Rec, V150Rec, V153Rec, V155Rec, V157Rec, V158Rec, V786Rec and V1738Rec all data cases were code X = 1; no logistic regression model could be fitted. For V62Rec, V79Rec, V151Rec, V754Rec and V1648Rec quasi-complete separation of data cases prevented using maximum likelihood estimation to fit a logistic regression model. Inclusion of data cases preceding 1800 AD would reduce quasi-complete separation of data cases. In the case of V678Rec there was no goodness of fit for the model with predictors obtained by backward selection ($\alpha = 0.05$). In the case of V744Rec attempts to fit a logistic regression model by backward selection ($\alpha = 0.05$) removed all predictors. For this a constant process k is assumed in Table 3. In the case of V1130Rec the attempt to fit a regression model was abandoned when a singular matrix prevented using a test to assess goodness of fit.

The following logistic regression models were fitted to the sample of 189 member countries of the UN at 2000 AD. For V1Rec the following model was fitted: $P(X = 1) = [1 + \exp(31.2664 - 0.6053mr)]^{-1}$, b_o, b_1:$\alpha < 0.0001$; c = 0.842, R-Square = 0.3770; Hosmer-Lemeshow = 0.3696, Deviance = 0.3969, Pearson = 0.3696; Odds ratio: mr = 1.832; n = 189.

For V63Rec the following model was fitted: $P(X = 1) = [1 + \exp(25.9238 - 0.4950mr)]^{-1}$,$b_o$, b_1:$\alpha < 0.0001$; c = 0.867, R-Square = 0.3201; Hosmer-Lemeshow = 0.4811, Deviance = 0.4865, Pearson = 0.4811; Odds ratio: mr = 1.640; n = 189.

For V68Rec the following model was fitted: $P(X = 1) = [1 + \exp(6.0072 - 0.00254mr^2)]^{-1}$,$b_o$, b_1: $\alpha < 0.0001$; c = 0.706; R-Square = 0.1196; Hosmer-Lemeshow = 0.1807, Deviance = 0.1653, Pearson = 0.1807; odds ratio mr^2 = 1.0003; n = 189.

For V69Rec the following model was fitted: $P(X = 1) = [1 + \exp(13.7798 - 0.00503mr^2)]^{-1}$, b_o, b_1: $\alpha < 0.0001$; c = 0.817; R-Square = 0.3386; Hosmer-Lemeshow = 0.7888, Deviance = 0.7869, Pearson = 0.7888; odds ratio mr^2 = 1.005; n = 189.

12. For a Bernoulli random variable X = 1, 0 the time series forecast method of Denton (2003) calculates

$$P(X_t = 1) = \Sigma_y [\Sigma_m P(X_t = 1|M_t = m)P(M_t = m)]P(Y_t = y), y = 1, 2, 3, m \in \{m_y\}$$

at 500-year intervals from about 10,000 BC to 1500 AD. The result is a time series of observations. In the preceding equation *m* denotes the scale of Table 1, $y = 1, 2, 3$ denotes band forager, pre-state food producers and state food producers. A time series forecast is made by applying a forecast method such as Holt-Winters non-seasonal (Chatfield 2001) to the time series of observations so created. In the preceding equation $P(X_t = 1|M_t = m)$ is estimated by fitting a curve to the SCCS. Denton (2003) denotes such use of the SCCS as Reconstruction of Prehistory by Quantitative Ethnographic Analogy (RPQEA). Denton (2004b) shows that such RPQEA estimates have error prior to 1000 AD. Denton (2003) uses the archaeological record to estimate $P(M_t = m)$ and $P(Y_t = y)$ in the above equation. Such estimates are subject to error in the archaeological record. Errors arising from demographic reconstructions of regional world populations coded $y = 1, 2, 3, m \in \{m_y\}$, may also arise in estimates of $P(M_t = m)$, $P(Y_t = y)$ in the above equation. Estimates of each of the elements of the RHS of the above equation are subject to the possible errors stated. These errors are currently difficult to estimate. None of these errors appears in the forecast method of the present paper. Errors appearing in the methods of the present paper are outlined in section 5.

Suppose we wish to minimize the error of estimates of behavioral relations $P(X = 1) = f(m)$ fitted to SCCS data and applied to times prior to 1000 BC. Denton (2004b) reviews errors arising from RPQEA. One solution might be to analyze RPQEA time series of observations using a time series forecast method such as Holt-Winters non-seasonal (Chatfield 2001) which puts greater weight on more recent observations. Future research, modeling how RPQEA differs from archaeological reconstruction of the same subject matters, may extend the range of RPQEA back beyond 1000 AD.

13. See Levinson and Malone (1980) for a critique of the sort of cross-cultural research facilitated to date by the SCCS.

7. REFERENCES

Agresti, A.
2002 *Categorical Data Analysis*. 2nd Edition. New York: Wiley.
Barry, H., III and A. Schlegel, eds.
1980 *Cross-cultural Samples and Codes*. Pittsburgh: University of Pittsburgh Press.
Chatfield, C.
1996 *The Analysis of Time Series*. 5th Edition. London: Chapman & Hall.
2002 *Time Series Forecasting*. Boca Raton: Chapman & Hall/CRC.
Cohen, J.
1988 *Statistical Power Analysis for the Behavioral Sciences*. 2nd Edition. Hillsdale: NJ: Lawrence Erlbaum.
Denton, T.
1996 Social differentiation: A speculative birth and death process hypothesis. *Journal of Quantitative Anthropology* 6:273-302.

1998 Social and structural differentiation: Conceptualization and measurement. *Cross-Cultural Research* 32:37-78.
1999 Long range forecasts of culture and society: Four quantitative methods from cultural anthropology. *Anthropologica* 41:177-193.
2003 Indirect reconstruction of prehistory by quantitative ethnographic analogy. *World Cultures* 13:110-146.
2004a Reconstruction of prehistory by quantitative ethnographic analogy: Data based tests. *World Cultures* 15:2-27.
2004b Cultural complexity revisited. *Cross-Cultural Research* 38:3-26.
nd Human society to 2050 AD: Forecasts. Manuscript.

Divale, W. T.
2004a Focused ethnographic bibliography for the Standard Cross-Cultural Sample. *World Cultures* CD data disk. *World Cultures* 14:194.
2004b *World Cultures* CD data disk. *World Cultures* 14:194.

Ember, C. R. and M. Ember
2004 *Cultural Anthropology*. 11th Edition. Upper Saddle River, NJ: Prentice-Hall.

Ember, M. and C. R. Ember, eds.
2001 *Countries and their Cultures*. Four volumes. New York: Macmillan Reference USA.

Fagan, B. M. and C. DeCorse
2005 *In the Beginning*. 11th Edition. Upper Saddle River, NJ: Prentice Hall.

Harvey, A. C.
1993 *Time Series Models*. 2nd Edition. London: Harvester-Wheatsheaf.

Hosmer, D. W. and S. Lemeshow
2000 *Applied Logistic Regression*. 2nd edition, New York: Wiley.

Hsieh, F. Y.
1989 Sample size tables for logistic regression. *Statistics in Medicine* 8:795-802.

Inglehart, R.
2004 *Human Beliefs and Values: A Cross-cultural Sourcebook Based on the 1999-2002 Values Surveys*. Mexico City: Siglo XXI.

Jackman, W.
1985 Cross-national research and the study of comparative politics. *American Journal of Political Science* 29:161-182.

Kohn, M. L.
1987 Cross-national research as an analytic strategy. *American Sociological Review* 52:713-731.

Kotz, S., and N. L. Johnson
1988 *Encyclopedia of Statistical Sciences,* Vol 9. New York: Wiley. Pp 334-336.

Kurian, G. T.
2001 *The Illustrated Book of World Rankings*. 5th edition. Armonk New York: Sharpe Reference.

Kutner, M. H., C. J. Nachtscheim, and J. Neter
2004 *Applied Linear Regression Models*. 4th Edition. Boston: McGraw-Hill.

Lattin, J., J. D. Carroll, and P. E. Green
2003 *Analyzing Multivariate Data*. Pacific Grove, CA: Thomson Learning.
Levinson, D., and M. J. Malone
1980 *Toward Explaining Human Culture*. New Haven: HRAF Press.
Lockyer, H.
1969 *All the Trades and Occupations of the Bible*. Grand Rapids, MI: Zondervan Publishing House.
Malinowski, B.
1988 The group and the individual in functional analysis. In, *High Points in Cultural Anthropology*. 2nd Edition. (P. Bohannan and M. Glazer, eds.). New York: Knopf. Pp. 275-293.
Murdock, G. P.
1981 *Atlas of World Cultures*. Pittsburgh: University of Pittsburgh Press.
Murdock, G. P., and C. Provost
1973 Measurement of cultural complexity. *Ethnology* 12:379-392.
Murdock, G. P., and D. R. White
1969 The standard cross-cultural sample. *Ethnology* 8:329-369.
Sarkees, M. R.
2000 The correlates of war data on war: An update to 1997. *Conflict Management and Peace Science* 18:123-144 (http://cow2.la.psu.edu/).
Spiegel. M. R.
1981 *Theory and Problems of Advanced Calculus*. New York: Schaum's Outline Series, McGraw-Hill.
Tatje, T. A., and R. Naroll
1970 Two measures of societal complexity. In, *A Handbook of Method in Cultural Anthropology*. (R. Naroll and R. Cohen, eds.). Garden City: NY, Natural History Press. Pp. 766-833.
United Nations
2005 *World Population Prospects: The 2004 Revision*. New York: United Nations. (http://esa.un.org/unpp/index.asp at the time of writing).

APPENDIX 1. DIVISION OF LABOR: PREINDUSTRIAL DATABASE AND LINEAR MODEL

In Appendix 1 we will create a preindustrial sample of coded data for number of specialized occupations $w = 0, 1, 2, \ldots$ with the ultimate goal of fitting to the data a function $w = g(m)$ or $g(m, t)$. First we will create such a database. After that, we will attempt to fit the function $w = g(m, t)$ to the database. The database will be a sub-set of the Standard Cross-Cultural Sample (SCCS) of preindustrial societies discussed elsewhere in this paper. While SCCS societies have been coded for $m \geq 0$ there are currently no usable SCCS coded data for number of specialized occupations $w = 0, 1, 2, \ldots$.

Appendix 1.1. A Preindustrial Database of Observations m = 1, 2, …, 37; 0.55 ≤ w ≤ 331.10

A "specialized occupation" is defined to exist in a society if there is significant production of a commodity such that producers make exchanges for the commodity based on the structured role "producer of the commodity (Denton 1998)." Commodities may be goods or services. Specialized occupation roles may be part-time or full-time. "Significance" may be in the number or importance of producers. Denton (1996) reviews the construct division of labor. The construct, "structured role" (Denton 1998) denotes an idealized role (e.g. chef) rather than the idiosyncratic performance of the role by particular individuals (e.g. how Arthur behaves as a chef).

Suppose a mother or father makes a grilled cheese sandwich for his or her child. If asked, he or she would explain the basis of the exchange in terms of the parent/child relation. Hence, the basis of the exchange is a kinship role rather than a specialized production role. Suppose, on the other hand, a chef in a restaurant makes a grilled cheese sandwich for a customer. The basis of the latter exchange would be explained in terms of being the producer of the commodity. Hence, "chef" meets the definition of a specialized occupation.

Tatje and Naroll (1970) give pertinent coded data for number of "durable artefact" occupations present in each of 58 preindustrial societies, many of which are also in the SCCS. Smiths leave durable artefacts whereas soothsayers do not. Hence, smith is a durable artefact occupation whereas soothsayer is not. Tatje and Naroll (1970) Raw Scores Column C are for full-time occupations leaving durable artefacts. From these occupations D = 0, 1, 2, … we will construct a crude set of observations of number of specialized occupations W = 0, 1, 2, …

Tatje and Naroll's (1970) data $d = 0, 1, 2, \ldots$ are for full-time durable artifact occupations whereas the construct W, W = 0, 1, 2, … of the present paper includes both full time and part time occupations. We will add a correction to estimate a component for part-time durable artifact occupations not having full-time counterparts.

Suppose we eliminate from Tatje and Naroll's (1970) roster of 58 societies all those which are not part of the SCCS or not pinpointed to the same SCCS time and geographic focus. Suppose we further prune the Tatje and Naroll (1970) societies by eliminating all those for which data are missing as to number of durable artifacts. Suppose we eliminate SCCS cultures 156 (Miskito) and 29 (Fur) on the grounds that Tatje and Naroll (1970) say that their codes for these societies are suspect. We are left with the first 23 SCCS societies of Table 4. From each of Tatje and Naroll's (1970) codes of number of durable artifact occupations, we delete two gender roles in order to make their codes conform to the definition of specialized occupation stated above.

Table 4 lists 23 SCCS societies. In Table 4 columns 1-3 appear the SCCS societal ID, SCCS societal name and number of observed durable artifact occupations D = 0, 1, 2, … from Tatje and Naroll (1970) Column C with their 2 gender roles deleted. In Table 4 column 4 appears the Murdock and Provost' (1973) cultural complexity code *m*. Table 4 Column 7 lists the date of ethnographic observation of each society. Columns 5 and 6 of Table 4 list W|*d* and W|*d*Adj which are, respectively, a provisional and final (adjusted) estimate of the total number of specialized occupations present in the society. These estimates will be derived in a moment. They will be treated as observations. The 24th society listed in Table 4 is Palestine about 0 AD which is included here since data are available for it (Lockyer 1969). The m-code for Palestine in Table 4 is the writer's estimate.

Estimation of number of specialized occupations, W|*d*Adj in Table 4, proceeds as follows. Logically, W|*d* = XD|*d* + *d*, where D denotes durable-artifact occupations and XD denotes non-durable artifact occupations. We have *d* from Tatje and Naroll (1970). If we are able to estimate XD we will have a crude, marginally usable, but usable estimate W|*d*.

W|*d* IN SOCIETIES CODED $0 \leq D \leq 6$: The basis of inference is conceptualization. Simple societies will need medico/religious practitioners. As they become more complex they will need political officials. Hence, using the symbol $\rightarrow$ to denote, "leads the writer to assume an expected W|*d* of …," ($d = 0$) $\rightarrow$ (W|$d = 0.5$), ($d = 1$) $\rightarrow$ (W|$d = 1.5$), ($d = 2$) $\rightarrow$ (W|$d = 3$), ($d = 3$) $\rightarrow$ (W|$d = 4.5$), ($d = 4$) $\rightarrow$ (W|$d = 6$), (D = 5) $\rightarrow$ (W|$d = 7$), ($d = 6$) $\rightarrow$ (W|$d = 8$).

W|*d* IN SOCIETIES CODED D > 6: Here also the basis of inference is conceptualization but of a different variety. From theory we will construct a differential equation for XD from which we will calculate W|*d*. The rate of change of the expected number of non-durable occupations XD in regard to durable occupations D will vary as the number of non-durable occupations D:

$$d(XD)/dD \;\; \alpha \;\; \mathrm{D},\ \mathrm{XD} = c\mathrm{D}^2/2 + k$$

We estimate *c*, *k* from boundary conditions. When D = 6 XD = 2 (Copper Inuit, Denton 1998). When D = 37 XD = 184 (Denton 1996 for Palestine at about 0 AD). Solving the two equations, $2 = ((c)(6^2)/2 + k)$, $184 = ((c)(37^2)/2 + k)$, in two unknowns, $c = 0.273068267$, $k = -2.915228806$. We will use the formula $\mathrm{XD} = c\mathrm{D}^2/2 + k$ to calculate XD rounded to two decimals. We will obtain W|*d* from the equality W|*d* = XD + D.

We need to adjust W|*d* in Table 4 so as to account for three varieties of error. First, the durable artifact D codes of Tatje and Naroll (1970) omit sub-specialized occupations. For example, real estate lawyers and criminal lawyers would appear as lawyers even though they produce different commodities. Secondly, we need to adjust for occupations omitted (for whatever reason) from the sources used by Tatje and Naroll (1970) to make their observations of D. Thirdly, we need to adjust for part time, durable artifact occupations omitted from Tatje and Naroll's (1970) Column C observations of D.

Let S denote number of sub-specialized occupations. We might calculate XD sub-specializations from D or we might work with the total W|*d* which we have already calculated. The writer chose to do the latter, making the assumption that branching for D and XD is also that of W|D. Using the same reasoning as with D, XD, the rate of change of S in regard to W|d, will vary as W|*d* which we abbreviate W:

$$dS/dW \ \alpha \ W, \ dS/dW = hW \ , \ S = hW^2/2 + g$$

We estimate h, g from boundary conditions. When $W \leq 8$ we will assume that S = 0. The codes for Palestine for W calculated by Denton (1996) from Lockyer (1969) are much like those of Tatje and Naroll (1970) for D in that they omit sub-specializations. The writer estimates that about 80 sub-specialized occupations are omitted from Lockyer (1969) especially for military ranks, civil service occupations, priests and the like. For the Copper Inuit $S = 0 = ((h)(8^2)/2 + g)$. For Palestine $80 = ((h)(221^2)/2 + g)$. Solving the two equations in two unknowns, $h = 0.003280235$, $g = -0.104967505$ provisionally left at nine decimals. Preliminary estimates of W preliminary = S + W are calculated using the formula $S = hW^2/2 + g$. To each preliminary estimate of W so obtained we add 10% to allow for underreporting of D in the sources used by Tatje and Naroll (1970). The result is W|*d*Adj as reported in Table 4 column 6 to two decimals.

In Table 4 D is an observation. W|*d*Adj of column 6 will be treated as an observation but is in fact the writer's estimate E[WAdjusted|*d*]. This is a marginally usable, but usable, database.

Table 4. Number of Specialized Occupations W in 24 Preindustrial Societies

SCCS #	Society	D	M	W|d	W|dAdj adj	Time AD
1	Hottentot	2	8	3.00	3.30	1860
77	Semang	1	2	1.50	1.65	1925
79	Andamanese	0	4	0.50	0.55	1860
119	Gilyak	2	9	3.00	3.30	1890
124	Copper Eskimo	6	6	8.00	8.80	1915
180	Aweikowa	0	1	0.50	0.55	1932
3	Thonga	6	20	8.00	8.80	1895
14	Mongo	4	20	6.00	6.60	1930
16	Tiv	4	19	6.00	6.60	1920
20	Mende	11	22	24.61	28.05	1945
23	Tallensi	8	22	13.82	15.55	1934
28	Azande	3	21	4.50	4.95	1905
60	Gond	5	17	7.00	7.70	1938
65	Kazak	2	20	3.00	3.30	1885
87	Toradja	6	18	8.00	8.80	1910
92	Orokaiva	3	10	4.5	4.95	1925
94	Kapauku	3	19	4.5	4.95	1955
100	Tikopia	10	18	20.74	23.47	1930
121	Chukchee	3	8	4.5	4.95	1900
151	Papago	3	17	4.5	4.95	1910
37	Amhara	14	32	37.85	44.10	1953
71	Burmese	26	38	115.38	150.82	1965
153	Nahua	32	34	168.90	237.14	1520
	Palestine	37	37	221.00	331.10	0

D denotes number of occupations leaving durable artifacts. W|d denotes preliminary, unadjusted estimate of total Number of Specialized Occupations given a particular d of D. W|dAdj denotes final, adjusted estimate of Number of Specialized Occupations given d. M denotes the M-scale value. Time denotes time of ethnographic observation of the society.

Appendix 1.2. The Function w = g(m, t) Fitted to the Database of Table 4

Next we will attempt to fit, to the data of Table 4, the equation w = g(m, t). In so doing we will retain the last two data cases of Table 4 Nahua (or Aztec, at 1520 AD) and Palestine (0 at AD) even though they predate 1800 AD. The reason for retaining these data cases is that they, of all the societies of Table 4, have the largest number of specialized occupations. Any model of number of specialized occupations $w = 0, 1, 2, \ldots$ in preindustrial societies must be fitted to data spanning the range of preindustrial w.

The graph (not shown) of Table 4 $w = 0, 1, 2, \ldots$ as a function of $m \geq 0$ suggests exponential growth. The writer started by fitting to data the regression model:

$$y = f(\mathbf{z}`), \ \mathbf{z}` = <m, m^2, m^3, t, t^2, t^3, mt, m^2 t^2, m^3 t^3>$$
$$y = \ln(w)$$

Backward selection ($\alpha = 0.05$) reduced the candidate predictor set to $\mathbf{z}` = <m^2, m^3, t, t^2, mt, m^3 t^3>$ with $R^2 = 0.9635$, SSE = 6.40, MSE = 0.36, RMSE = 0.59. Because predictors in powers and interactions of t were intended to be exploratory devices to measure unspecified predictors, the writer investigated a variety of additional models in m and powers thereof. When the regression model:

$$y = f(\mathbf{z}`), \ \mathbf{z}` = <m, m^2, m^3, t>$$

was fitted to data backward selection ($\alpha = 0.05$) reduced the candidate predictor set to $\mathbf{z}` = <m>$. However, Mallows $C_p$ was not optimal. Mallows $C_p$ becomes optimal if the predictor set is $\mathbf{z}` = <m, m^2>$. Even though the p-value of the coefficient of m^2 is $p \approx 0.10$, that is the model tentatively selected as a viable alternative to the 6-predictor model. P-values of predictors are only one of several diagnostics to be used in curve fitting. The 2-predictor model is shown in Table 5. A no-intercept model is fitted in Table 5 for two reasons. First, exclusion of an intercept significantly increases R^2. Second, the model fitted is conceptually more interpretable. With the 2-predictor model of Table 5 $R^2 = 0.9355$, SSE = 11.29, MSE = 0.51, RMSE = 0.7165. Models comparison tests (Latin et al. 2003) show that R^2 of the 6-predictor model is significantly higher ($\alpha = 0.05$) than R^2 of the 2-predictor model.

Table 5. Linear Regression Model for Number of Specialized Occupations W in the 24 Preindustrial Societies of Table 4

Model	$\ln(W_i) = am + bm^2 + \varepsilon_i$	**Residuals**	
$\mathbf{R^2}$	0.9355	Null H_o: $\varepsilon_i \sim N$	P<W=0.9064 P>D>0.15 P>Wsqu.>0.25 P>Asqu.>0.25
Coefficients	a = 0.09040 P < 0.0008	Null H_o: $\varepsilon_i \sim$ random	Runs Test p>0.05
	b = 0.00138 P < 0.1037	Null H_o: $\varepsilon_i \sim$ indpt.	Yes DW=1.822 1st ord ρ=0.061
Mallow's C_p	2.00	Root Mean Square Error	0.71650

W is W\|d adj of Table 4. Residual tests are from SAS V8.2 PROC REG, PROC UNIVARIATE. See text for model fitting.

The candidate predictor set **z`**= <m, m^2, m^3, t, t^2, t^3, mt, m^2t^2, m^3t^3> runs the risk of overfitting (Kutner et al. 2004; Latin et al. 2003). Overfitting occurs when a fitted model capitalizes on random characteristics of the data set rather than on true behavioral relations. Overfitting is especially likely to occur where there are exploratory candidate predictors as is the case with **z`**= <m, m^2, m^3, t, t^2, t^3, mt, m^2t^2, m^3t^3>. To test for overfitting the jackknife method was used (Latin et al. 2003:73; denoted PRESS_p by Kutner et al. 2004:360) for which a PRESS option is available in SAS PROC REG. Application of the jackknife method proceeded as follows. Using the SAS PROC REG PRESS option each of the two models (6-predictor and 2-predictor) was refitted (with appropriate model predictors) as many times as there are data cases. At each fit a different data case was omitted. The refitted residual $\hat{y}_i - y_i$ was calculated, $\hat{y}_i$ is the predicted value based on the model with the data case omitted, y_i is the observed value of y for the omitted data case. Then, the residuals are squared. SSE, MSE and R^2 are recalculated. If the jackknife method results in MSE for the 6-predictor model which is higher than that of the 2-predictor model we assume that the 6-predictor model is based on overfitting. This is exactly what happened.

The jackknife method produced the following results. For the 6-predictor model jackknifed SSE = 55.2394698, MSE = 3.068859433, R^2 = 0.6847. Jackknifing reduced R^2 of the 6-predictor model by some 28%. For the 2-predictor model jackknifed SSE = 13.2532369, MSE = 0.602419859, R^2 = 0.9243. The conclusion is that the 6-predictor model is the product of overfitting. That the jackknife R^2 of the 2-predictor model is approximately equal to the R^2 of the original 2-predictor model fitted to all 24 data cases suggests that this model arises from true behavioral relations rather than from overfitting. There is a lesson to be learned here. If a model such as y = f(**z`**), **z`**= <m, m^2, m^3, t, t^2, t^3, mt, m^2t^2, m^3t^3>, with exploratory candidate predictors in t and interactions, is fitted to data and backward selection retains powers and/or interactions of time t, the fitted model may arise from overfitting.

Table 5 shows the results of fitting to data the function y = f(**z`**), **z`**= <m, m^2>, $y = \ln(w)$. There is good fit. Since the fitted model is a monotone function of m it may be inverted to find $m = g^{-1}(w)$ as a function of w, the non-negative solution to the quadratic equation $0.00138m^2 + 0.09040m - \ln(w) = 0$. Denton (1996) finds the causes of w = 0, 1, 2, … in preindustrial and industrial societies appear to be approximately the same – only the magnitude of the causes differs. If we know w in an industrial society we ought to be able to calculate an approximate measure m in the society as the non-negative solution to the quadratic equation $0.00138m^2 + 0.09040m - \ln(w) = 0$. If we estimate mean number of specialized occupations in least developed, less developed (excluding least developed) and developed countries at 2000 AD we will be able to solve for grouped modernization m in these countries. Based on Denton (1996) mean number of specialized occupations in least developed, developing (excluding least developed) and developed countries at 2000 AD is estimated to be 2000, 6000 and 18,000. Putting these values of w into $0.00138m^2 +$

$0.09040m - \ln(w) = 0$, the three positive solutions for grouped m are $m \approx 48.368$, 53.135, 57.65 in such countries.

A 90% Bonferroni joint family confidence interval (Kutner et al. 2004) for the coefficients *a* and *b* of Table 5 is $0.04208 \le a \le 0.13872$, $0.00000 \le b \le 0.00306$. The model fitted in Table 5 is based on only 24 observations of marginal quality. Assuming about 25,000 specialized occupations in developed nations such as Canada at about 2000 AD (Denton 1996), the upper and lower 90% Bonferroni joint family confidence interval is $39.16 \le m \le 240.65$ with a model-based expectation $E[M|w = 25000] = 58.96$. In the following section tangent line approximation will be used to provide evidence that the model of Table 5 gives viable estimates of $m = g^{-1}(w)$ in contemporary countries.

Appendix 1.3. Supporting Tangent Line Approximations

Tangent line approximations support the conclusion that a grouped $m \approx 48.368$, 53.135, 57.65 are viable estimates for grouped modernization in least developed, less developed (excluding least developed) and developed nations at 2000 AD. The writer was able to use tangent line approximation to make an independent estimate of the magnitude of modernization in contemporary societies which the United Nations (2005) deems to be least developed, $d = 1$. This independent estimate supports the estimate of modernization $m \approx 48.368$ in least developed countries made by inverting the function $w = g(m)$ and solving $m = g^{-1}(w)$ at $w = 2000$. Here is how validation by tangent line approximation was done.

Tangent line approximation (Spiegel 1981) was applied to a logistic regression model of agricultural intensification - subscale 3 of Table 1 - for which there are coded SCCS data (V151 of Divale 2004b). We will rescale Divale's (2004b) to range over the integers 0, 1, 2, 3, 4 rather than 1, 2, 3, 4, 5. The rescaling conforms to Murdock and Provost' (1973) original Table 1 scale 3. From Murdock and Provost (1973) we have independent estimates of modernization *m* in these societies. The writer transformed scale 3 of Table 1 (SCCS V151) into a binary random variable X = 1 (if scale 3 = 4), X = 0 (if scale 3 = 0 OR 1 OR 2 OR 3). The result is a Bernoulli random variable X = 1 (intensive agriculture is the predominant form of food getting in the society), 0 (otherwise). Modernization m^* was calculated as the sum of the nine scales of Table 1 excluding scale 3. The latter precaution makes the measure m^* independent of the measure of binary random variable X = 1, 0.

The logistic regression equation $P(X = 1) = [1 + \exp(4.1713 - .1955m`)]^{-1}$ was successfully fitted to the 174 SCCS data cases from 1800 – 1965 AD. Using the model selection-model fitting strategy of Kutner et al. (2004) the writer first selected the predictor set (*infra*) by backward selection ($\alpha = 0.05$). The fit of the resulting model was tested using all three of Pearson residual, deviance residual and Hosmer-Lemeshow tests (Kutner et al. 2004).

The fitting of a logistic regression model where interval predictors are used necessitates grouping predictors in order to test fit (Hosmer and Lemeshow 2000). Time of ethnographic description of SCCS societies (V838 of Divale 2004b) from 1800 to 1965 AD inclusive was reset to a zero point at 1800 AD by the transformation t = V838 – 1800, following which six grouped intervals of time were created and the median group time $t`$ assigned to data cases in the group. Using $m^*$ to denote the Table 1 modernization measure obtained be excluding scale 3, five grouped intervals of $m^*$ were created and the median group modernization $m`$ assigned to data cases in the group. The grouped intervals of m^* and t were:

if $m^* \geq 0$ and $m^* \leq 7$ then $m` = 3.5$;
if $m^* \geq 8$ and $m^* \leq 15$ then $m` = 11.5$;
if $m^* \geq 16$ and $m^* \leq 22$ then $m` = 19$;
if $m^* \geq 23$ and $m^* \leq 29$ then $m` = 26$;
if $m^* \geq 30$ and $m^* \leq 36$ then $m` = 33$;

if $t \geq 0$ and $t \leq 24$ then $t` = 12$;
if $t \geq 25$ and $t \leq 49$ then $t` = 37$;
if $t \geq 50$ and $t \leq 74$ then $t` = 62$;
if $t \geq 75$ and $t \leq 99$ then $t` = 87$;
if $t \geq 100$ and $t \leq 124$ then $t` = 112$;
if $t \geq 125$ and $t \leq 165$ then $t` = 146$;

The model predictor set for $P(X = 1) = h(.)$ was obtained using backward selection (selection size $\alpha = 0.05$) which eliminated candidate predictors time $t`$, $t`^2$, $t`^3$, modernization $m`^2$, $m`^3$ and interactions $m`t`$, $m`^2 t`^2$ and $m`^3 t`^3$. The intention of positing candidate predictors in powers and interactions of time t was to use time t as a surrogate indicator of unobserved predictors. However, backward selection ($\alpha = 0.05$) eliminated all predictors in time t. The fitted model passed the standard tests of the Hosmer-Lemeshow statistic (Prob > ChiSquared = 0.7507), deviance residual (Prob > ChiSquared = 0.7473) and Pearson residual (Prob > ChiSquared = 0.7507) suggesting adequate fit. For the fitted model the ROC (receiver operating characteristic) is 0.843, R-Square = 0.3217 and Max-rescaled R-Square = 0.4604, all very acceptable. In logistic regression the definition of R-Square is different from that of linear regression and is generally much lower.

The fitted equation $P(X = 1) = h(m`) = [1 + \exp(4.1713 - 0.1955m`)]^{-1}$ is monotonic, as are all logistic regression equations. Since the fitted model is monotonic, it is invertible and $m` = h^{-1}(P(X = 1))$. Using tangent line approximation at the upper grouped SCCS value $m`_o = 33$ where $P_o \mid m`_o = 0.90722$, we might attempt to estimate $m$ in least developed countries from the tangent line approximation $(P - P_o) = dh/dm|_{m`_o} (m - m`_o)$. Rearranging the preceding equation, $m = (P - P_o)/(dh/dm)|_{m`_o} + m_o$. Setting P = 1 at m today the resulting estimate is $m = 36.8194$ which, after multiplying by 10/9 to give it the same measurement domain as that of Table 1 which includes scale 3 agricultural intensification, becomes $m` = 40.91044$. The latter estimate of m for least developed countries agrees approximately with the estimate 48.368 based on inverting the model of number of specialized occupations, $m = g^{-1}(w)$.

In the preceding paragraph the proportion (probability) of least developed countries at 2000 AD where there is intensive agriculture is set to P = 1. Let $d = 1, 2, 3$ denote developed, less developed (excluding least developed) and developed countries at 2000 AD. In Appendix 2 the numbers of countries coded $d = 1, 2, 3$ are 48, 95 and 46. The conditional proportions of countries where V151Rec = 1 (intensive agriculture) given that $d = 1, 2, 3$ are p(V151Rec = 1|d = 1) = 13/48 = 0.27, p(V151Rec = 1|d = 2) = 80/95 = 0.91, p(V151REc = |d = 3) = 46/46 = 1.0. In order to solve for m in least developed countries the tangent line approximation of the preceding paragraph uses P = p(V151Rec = 1|d = 1) = 1.0. How is this justified?

After coding the 189 CC countries of Appendix 2 for intensive agriculture V151Rec = 1, 0 the writer decided that he had misapplied Murdock and Provost (1973) coding practices for Agriculture (scale 3 of Table 1. What does one code – the food getting method producing the greatest proportion of food, the food getting method used by the largest proportion of food producers, or some other estimate? How intensive does agriculture have to be to be coded V51Rec = 1? If the code is based on the food getting method producing the greatest proportion of food the writer stands by the tangent line estimate for least developed countries coded as having intensive agriculture as p(V151Rec = 1|(d = 1) ≈ 1.0. The codes of Appendix 2 are left as is as a monument to the difficulties of meshing SCCS coded data (Divale 2004b) with coded data for contemporary countries. There is a need for additional methods of measuring modernization in such a way that a single measure is applicable to both preindustrial and industrial societies. Currently, this is no simple matter.

If we limit ourselves to the 174 SCCS societies predating 1800 AD the proportion p(V151Rec = 1|33 ≤ m ≤ 40) of such societies coded V151Rec = 1 by Murdock and Provost (1973) is 19/20 = 0.95. The writer's original (Appendix 2) coding for V151Rec results in a proportion of less developed societies (excluding least developed) where there is intensive agriculture of p(V151Rec = 1|d = 2) = 80/95 = 0.91. That p(V151Rec = 1|33 ≤ m ≤ 40) = 0.95 while p(V151Rec = 1|d = 2) = 0.91 leads the writer to the conclusion that his original coding for V151Rec in Appendix 2 is wrong. Setting p(V151Rec = 1|d = 1) ≈ 1.0 in the tangent line approximation used to solve for m in least developed countries seems appropriate.

We may get a better tangent line approximation if we use a logistic regression model in which modernization $m^* = 0, 1, 2,\ldots, 36$ remains ungrouped. The model $P(X = 1) = h(m^*) = [1 + \exp(4.5569 - .2159m^*)]^{-1}$ was successfully fitted to the 174 SCCS data cases. Backward selection reduced the predictor set of the preceding paragraph to the logistic regression model $P(X = 1) = h(m^{*2})$. However, the latter model is not significantly different from $P(X = 1) = h(m^*) = [1 + \exp(4.5569 - .2159m^*)]^{-1}$. Since an ungrouped, interval predictor is used in the fitted model, goodness of fit statistics have no meaning (Hosmer and Lemeshow 2000). The validity of the model fitted rests on successful fit of the model using grouped data.

The fitted logistic regression model of the preceding paragraph is monotonic – as are all logistic regression models - and is therefore invertible. The tangent line approximation is $(P - P_o) = dh/dm^*|_{m^*_o} (m^* - m^*_o)$ where the differential equation is calculated at $m^*_o = 36$ and $P_o|m^*_o \approx 0.96$. Rearranging the preceding equation, $m^* = (P - P_o)/(dh/dm^*) + m^*_o$. The linear approximation of m^* at $m^*_o = 36$ is 40.1160 which becomes 44.57 when multiplied by 10/9 to give it the same measurement domain as that of modernization m using all 10 indicators of Table 1. The adjusted estimate $m = 44.57$ agrees reasonably well with the estimate 48.368 based on inverting the model of number of specialized occupations $w = \mathrm{g}(m)$. In the fitted logistic regression model from which the $m = 44.57$ estimate was calculated, the 90% confidence interval for $\mathrm{P}(\mathrm{X} = 1)|m^* = 36$ is (0.88978, 0.98715) suggesting, possibly, a closer conformity to the estimated $m = 48.368$ in least developed countries obtained by inverting the model of number of specialized occupations $w = \mathrm{g}(m)$. Any use we make of estimates of the magnitude of modernization at 2050 AD must recognize the uncertainty of the estimates.

APPENDIX 2. CODED DATA FOR 189 UN MEMBER COUNTRIES AT 2000 A.D.

Soc Num	Society Name	Dev elop	V1 Rec	V18 Rec	V62 Rec	V63 Rec	V66 Rec	V68 Rec	V69 Rec	V70 Rec	V79 Rec
1	Afghanistan	1	0	1	1	0	1	0	0	0	1
2	Albania	3	1	1	1	0	1	0	0	0	1
3	Algeria	2	1	1	1	1	1	1	1	0	1
4	Andorra	3	1	1	1	1	1	1	1	1	1
5	Angola	1	1	1	1	0	1	1	1	0	0
6	Antigua	2	1	1	1	0	1	1	1	1	1
7	Argentina	2	1	1	1	1	1	1	1	1	1
8	Armenia	2	1	1	1	1	1	1	1	0	1
9	Australia	3	1	1	1	1	1	1	1	1	1
10	Austria	3	1	1	1	1	1	1	1	1	1
11	Azerbaijan	2	1	1	1	1	1	1	1	0	1
12	Bahamas	2	1	1	1	1	1	1	1	1	1
13	Bahrain	2	1	1	1	1	1	1	1	0	1
14	Bangladesh	1	0	1	1	0	1	0	0	0	1
15	Barbados	2	1	1	1	0	1	1	1	1	1
16	Belarus	3	1	1	1	1	1	1	1	1	1
17	Belgium	3	1	1	1	1	1	1	1	1	1
18	Belize	2	1	1	1	0	1	1	1	1	1
19	Benin	2	0	1	1	0	1	1	0	0	0
20	Bhutan	1	0	1	1	0	1	0	0	0	1
21	Bolivia	2	1	1	1	1	1	1	1	1	1
22	Bosnia Herzecegov	3	1	1	1	0	1	1	1	1	1
23	Botswana	2	1	1	1	1	1	0	0	0	1
24	Brazil	2	1	1	1	1	1	1	0	1	1
25	Brunei Darussalam	2	1	1	1	1	1	1	1	1	1
26	Bulgaria	3	1	1	1	1	1	1	1	1	1
27	Burkina Faso	1	0	1	1	0	1	1	0	0	0
28	Burma	1	0	1	1	0	1	1	1	1	1
29	Burundi	1	0	1	1	0	1	0	0	0	0
30	Cambodia	1	0	1	1	0	1	1	0	1	1
31	Cameroon	2	0	1	1	0	1	1	0	0	0
32	Canada	3	1	1	1	1	1	1	1	1	1
33	Cape Verde	1	1	1	1	1	1	1	1	1	0
34	Central Af Republic	1	0	1	1	0	1	1	0	1	1
35	Chad	1	0	1	1	0	1	0	0	0	0
36	Chile	2	1	1	1	1	1	1	1	1	1
37	China	2	0	1	1	0	1	1	1	0	1
38	Colombia	2	1	1	1	1	1	1	1	1	1
39	Comoros	1	0	1	1	0	1	1	0	0	1
40	Demrepub Congo	1	0	1	1	0	1	1	0	0	0
41	Repub Congo	2	1	1	1	1	1	0	0	0	0
42	Costa Rica	2	1	1	1	1	1	0	1	1	1
43	Cote D’Ivoire	2	0	1	1	0	1	0	0	0	1

Soc Num	Society Name	Dev elop	V1 Rec	V18 Rec	V62 Rec	V63 Rec	V66 Rec	V68 Rec	V69 Rec	V70 Rec	V79 Rec
44	Croatia	3	1	1	1	1	1	0	0	1	1
45	Cuba	2	1	1	1	1	1	0	0	1	1
46	Cyprus	2	1	1	1	1	1	1	1	1	1
47	Czech Republic	3	1	1	1	1	1	1	1	1	1
48	Denmark	3	1	1	1	1	1	1	1	1	1
49	Djibouti	1	1	1	1	1	1	1	0	0	1
50	Dominica	2	1	1	1	1	1	1	1	1	1
51	Dominican Republic	2	1	1	1	1	1	0	0	1	0
52	Ecuador	2	1	1	1	1	1	1	1	1	1
53	Egypt	2	0	1	1	0	1	1	1	0	1
54	El Salvador	2	1	1	1	1	1	1	1	1	1
55	Equatorial Guinea	1	0	1	1	0	1	0	0	0	0
56	Eritrea	1	0	1	1	0	1	1	0	0	1
57	Estonia	3	1	1	1	1	1	1	1	1	1
58	Ethiopia	1	0	1	0	0	1	0	0	1	1
59	Fiji	2	0	1	1	0	1	0	0	0	1
60	Finland	3	1	1	1	1	1	1	1	1	1
61	France	3	1	1	1	1	1	1	1	1	1
62	Gabon	2	1	1	1	1	1	0	0	0	0
63	Gambia	1	0	1	1	0	1	1	0	0	0
64	Georgia	2	1	1	1	1	1	1	1	1	1
65	Germany	3	1	1	1	1	1	1	1	1	1
66	Ghana	2	0	1	1	0	1	0	0	0	1
67	Greece	3	1	1	1	1	1	1	1	1	1
68	Grenada	2	1	1	1	0	1	1	1	1	1
69	Guatemala	2	0	1	1	0	1	1	1	1	1
70	Guinea	1	0	1	1	1	1	0	0	0	0
71	Guinea Bissau	1	0	1	1	0	1	0	0	0	0
72	Guyana	2	0	1	1	0	1	0	0	1	1
73	Haiti	1	0	1	1	0	1	1	1	1	0
74	Honduras	2	0	1	1	0	1	1	1	1	1
75	Hungary	3	1	1	1	1	1	1	1	1	1
76	Iceland	3	1	1	1	1	1	1	1	1	1
77	India	2	0	1	1	0	1	0	0	0	1
78	Indonesia	2	0	1	1	0	1	1	0	1	1
79	Iran	2	1	1	1	1	1	1	1	0	1
80	Iraq	2	1	1	1	1	1	0	0	0	0
81	Ireland	3	1	1	1	1	1	1	1	1	1
82	Israel	2	1	1	1	1	1	1	1	1	1
83	Italy	3	1	1	1	1	1	1	1	1	1
84	Jamaica	2	1	1	1	1	1	0	1	1	1
85	Japan	3	1	1	1	1	1	1	1	1	1
86	Jordan	2	1	1	1	1	1	0	1	0	0
87	Kazakhstan	2	1	1	1	1	1	0	1	0	0
88	Kenya	2	0	1	1	0	1	1	1	1	0
89	Kiribati	1	0	1	1	0	1	1	0	0	1

Soc Num	Society Name	Dev elop	V1 Rec	V18 Rec	V62 Rec	V63 Rec	V66 Rec	V68 Rec	V69 Rec	V70 Rec	V79 Rec
90	North Korea	2	1	1	1	1	1	1	1	0	1
91	South Korea	2	1	1	1	1	1	1	1	0	1
92	Kuwait	2	1	1	1	1	1	0	0	0	0
93	Kyrgystan	2	0	1	1	0	1	1	0	0	1
94	Laos	1	0	1	1	0	1	1	0	1	1
95	Latvia	3	1	1	1	1	1	1	1	0	1
96	Lebanon	2	1	1	1	1	1	1	1	0	0
97	Lesotho	1	0	1	1	0	1	0	0	0	0
98	Liberia	1	0	1	1	0	1	0	0	0	0
99	Libya	2	1	1	1	1	1	1	1	0	0
100	Liechtenstein	3	0	1	1	0	1	1	1	1	1
101	Lithuania	3	1	1	1	1	1	1	1	1	1
102	Luxembourg	3	1	1	1	1	1	1	1	1	1
103	Macedonia	3	1	1	1	1	1	1	1	0	1
104	Madagascar	1	0	1	1	0	1	1	0	0	0
105	Malawi	1	0	1	1	0	1	1	0	0	0
106	Malaysia	2	1	1	1	1	1	1	1	1	1
107	Maldives	1	0	1	1	0	1	1	0	0	1
108	Mali	1	0	1	1	0	1	0	0	0	0
109	Malta	3	1	1	1	1	1	1	1	1	1
110	Marshall Islands	2	1	1	1	1	1	1	1	0	1
111	Mauritania	1	1	1	1	1	1	0	0	0	0
112	Mauritius	2	1	1	1	0	1	0	0	0	1
113	Mexico	2	1	1	1	1	1	1	1	1	1
114	Fed Micronesia	2	0	1	1	0	1	1	0	0	1
115	Moldova	3	0	1	1	0	1	1	1	1	1
116	Monaco	3	1	1	1	1	1	1	1	1	1
117	Mongolia	2	1	1	1	1	1	1	1	1	1
118	Morocco	2	1	1	1	1	1	1	1	0	1
119	Mozambique	1	0	1	1	0	1	0	0	0	1
120	Namibia	2	0	1	1	0	1	0	0	0	1
121	Nauru	2	1	1	1	1	1	1	0	0	1
122	Nepal	1	0	1	1	0	1	0	0	0	1
123	Netherlands	3	1	1	1	1	1	1	1	1	1
124	New Zealand	3	1	1	1	1	1	1	1	1	1
125	Nicaragua	2	1	1	1	1	1	0	0	1	1
126	Niger	1	0	1	1	0	1	0	0	0	0
127	Nigeria	2	0	1	1	0	1	1	0	0	0
128	Norway	3	1	1	1	1	1	1	1	1	1
129	Oman	2	1	1	1	1	1	0	0	0	0
130	Pakistan	2	1	1	1	0	1	0	0	0	0
131	Palau	2	1	1	1	1	1	1	0	0	1
132	Panama	2	1	1	1	1	1	1	1	1	1
133	Papua	2	0	1	1	0	1	1	0	0	1
134	Paraguay	2	1	1	1	1	1	1	1	1	1
135	Peru	2	1	1	1	1	1	1	1	1	1

Soc Num	Society Name	Develop	V1 Rec	V18 Rec	V62 Rec	V63 Rec	V66 Rec	V68 Rec	V69 Rec	V70 Rec	V79 Rec
136	Philippines	2	1	1	1	1	1	0	0	1	1
137	Poland	3	1	1	1	1	1	1	1	1	1
138	Portugal	3	1	1	1	1	1	1	1	1	1
139	Qatar	2	1	1	1	1	1	0	0	0	0
140	Romania	3	1	1	1	1	1	1	1	1	1
141	Russia	3	1	1	1	1	1	1	1	1	1
142	Rwanda	1	0	1	1	0	1	1	0	0	1
143	Saint Kitts Nevis	2	1	1	1	0	1	1	1	1	1
144	Saint Lucia	2	1	1	1	0	1	1	1	1	1
145	Saint Vincentgren	2	1	1	1	1	1	1	1	1	1
146	Samoa	1	0	1	1	0	1	1	0	0	1
147	San Marino	3	1	1	1	1	1	1	1	1	1
148	Saotomeprincipe	1	0	1	1	0	1	1	1	1	0
149	Saudi Arabia	2	1	1	1	1	1	1	1	0	1
150	Senegal	1	0	1	1	0	1	1	0	0	0
151	Serbia Montenegro	3	1	1	1	1	1	1	1	1	1
152	Seychelles	2	1	1	1	0	1	1	1	1	1
153	Sierra Leone	1	0	1	1	0	1	0	0	0	0
154	Singapore	2	1	1	1	1	1	1	1	0	1
155	Slovakia	3	1	1	1	1	1	1	1	1	1
156	Slovenia	3	1	1	1	1	1	1	1	1	1
157	Solomon islands	1	0	1	1	0	1	0	0	0	1
158	Somalia	1	0	1	1	0	1	1	0	0	0
159	South Africa	2	0	1	1	0	1	1	1	0	1
160	Spain	3	1	1	1	1	1	1	1	1	1
161	Sri Lanka	2	0	1	1	0	1	1	0	1	1
162	Sudan	1	0	1	1	0	1	1	0	0	1
163	Suriname	2	1	1	1	1	1	1	0	1	1
164	Swaziland	2	0	1	1	0	1	1	0	0	1
165	Sweden	3	1	1	1	1	1	1	1	1	1
166	Syria	2	1	1	1	1	1	0	1	0	0
167	Tajikistan	2	0	1	1	0	1	1	0	0	1
168	Tanzania	1	0	1	1	0	1	0	0	0	1
169	Thailand	2	0	1	1	0	1	1	1	1	1
170	Togo	1	0	1	1	0	1	0	0	0	0
171	Tonga	2	0	1	1	0	1	1	0	1	1
172	Trinidad	2	1	1	1	1	1	1	1	1	1
173	Tunisia	2	1	1	1	1	1	1	1	0	1
174	Turkey	2	1	1	1	1	1	1	1	0	1
175	Turkmenistan	2	1	1	1	0	1	1	0	0	1
176	Tuvalu	1	1	1	1	1	1	1	1	1	1
177	Uganda	1	0	1	1	0	1	0	0	0	0
178	Ukraine	3	1	1	1	1	1	1	1	1	1
179	United Arab Emir	2	1	1	1	1	1	1	1	1	1
180	United kingdom	3	1	1	1	1	1	1	1	1	1
181	USA	3	1	1	1	1	1	1	1	1	1

Soc Num	Society Name	Dev elop	V1 Rec	V18 Rec	V62 Rec	V63 Rec	V66 Rec	V68 Rec	V69 Rec	V70 Rec	V79 Rec
182	Uruguay	2	1	1	1	1	1	1	1	1	1
183	Uzbekistan	2	1	1	1	0	1	1	0	0	1
184	Vanuatu	1	0	1	1	0	1	1	0	0	1
185	Venezuela	2	1	1	1	1	1	1	1	1	1
186	Vietnam	2	0	1	1	0	1	0	0	0	1
187	Yemen	1	0	1	1	0	1	0	0	0	0
188	Zambia	1	0	1	1	0	1	0	0	0	1
189	Zimbabwe	2	0	1	1	0	1	0	0	0	0

Soc Num	Society Name	V 121 Rec	V 129 Rec	V 143 Rec	V 149 Rec	V 150 Rec	V 151 Rec	V 153 Rec	V 155 Rec	V 157 Rec	V 158 Rec
1	Afghanistan	1	1	1	1	1	1	1	1	1	1
2	Albania	1	1	1	1	1	1	1	1	1	1
3	Algeria	1	1	1	1	1	1	1	1	1	1
4	Andorra	1	1	1	1	1	1	1	1	1	1
5	Angola	1	1	1	1	1	0	1	1	1	1
6	Antigua	1	1	1	1	1	1	1	1	1	1
7	Argentina	1	1	1	1	1	1	1	1	1	1
8	Armenia	1	1	1	1	1	1	1	1	1	1
9	Australia	1	1	1	1	1	1	1	1	1	1
10	Austria	1	1	1	1	1	1	1	1	1	1
11	Azerbaijan	1	1	1	1	1	1	1	1	1	1
12	Bahamas	1	1	1	1	1	1	1	1	1	1
13	Bahrain	1	1	1	1	1	1	1	1	1	1
14	Bangladesh	1	1	1	1	1	1	1	1	1	1
15	Barbados	1	1	1	1	1	1	1	1	1	1
16	Belarus	1	1	1	1	1	1	1	1	1	1
17	Belgium	1	1	1	1	1	1	1	1	1	1
18	Belize	1	1	1	1	1	1	1	1	1	1
19	Benin	1	1	1	1	1	1	1	1	1	1
20	Bhutan	1	1	1	1	1	1	1	1	1	1
21	Bolivia	1	1	1	1	1	1	1	1	1	1
22	Bosnia Herzecegov	1	1	1	1	1	1	1	1	1	1
23	Botswana	1	1	1	1	1	1	1	1	1	1
24	Brazil	1	1	1	1	1	1	1	1	1	1
25	Brunei Darussalam	1	1	1	1	1	1	1	1	1	1
26	Bulgaria	1	1	1	1	1	1	1	1	1	1
27	Burkina Faso	1	1	1	1	1	0	1	1	1	1
28	Burma	1	1	1	1	1	1	1	1	1	1
29	Burundi	1	1	1	1	1	0	1	1	1	1
30	Cambodia	1	1	1	1	1	0	1	1	1	1
31	Cameroon	1	1	1	1	1	1	1	1	1	1
32	Canada	1	1	1	1	1	1	1	1	1	1
33	Cape Verde	1	1	1	1	1	1	1	1	1	1
34	Central Af Republic	1	1	1	1	1	0	1	1	1	1
35	Chad	1	1	1	1	1	0	1	1	1	1
36	Chile	1	1	1	1	1	1	1	1	1	1
37	China	1	1	1	1	1	1	1	1	1	1
38	Colombia	1	1	1	1	1	1	1	1	1	1
39	Comoros	1	1	1	1	1	1	1	1	1	1
40	Demrepub Congo	1	1	1	1	1	0	1	1	1	1
41	Repub Congo	1	1	1	1	1	1	1	1	1	1
42	Costa Rica	1	1	1	1	1	1	1	1	1	1
43	Cote D'Ivoire	1	1	1	1	1	1	1	1	1	1

Soc Num	Society Name	V 121 Rec	V 129 Rec	V 143 Rec	V 149 Rec	V 150 Rec	V 151 Rec	V 153 Rec	V 155 Rec	V 157 Rec	V 158 Rec
44	Croatia	1	1	1	1	1	1	1	1	1	1
45	Cuba	1	1	1	1	1	1	1	1	1	1
46	Cyprus	1	1	1	1	1	1	1	1	1	1
47	Czech Republic	1	1	1	1	1	1	1	1	1	1
48	Denmark	1	1	1	1	1	1	1	1	1	1
49	Djibouti	1	1	1	1	1	1	1	1	1	1
50	Dominica	1	1	1	1	1	1	1	1	1	1
51	Dominican Republic	1	1	1	1	1	1	1	1	1	1
52	Ecuador	1	1	1	1	1	1	1	1	1	1
53	Egypt	1	1	1	1	1	1	1	1	1	1
54	El Salvador	1	1	1	1	1	1	1	1	1	1
55	Equatorial Guinea	1	1	1	1	1	1	1	1	1	1
56	Eritrea	1	1	1	1	1	1	1	1	1	1
57	Estonia	1	1	1	1	1	1	1	1	1	1
58	Ethiopia	1	1	1	1	1	0	1	1	1	1
59	Fiji	1	1	1	1	1	0	1	1	1	1
60	Finland	1	1	1	1	1	1	1	1	1	1
61	France	1	1	1	1	1	1	1	1	1	1
62	Gabon	1	1	1	1	1	1	1	1	1	1
63	Gambia	1	1	1	1	1	0	1	1	1	1
64	Georgia	1	1	1	1	1	1	1	1	1	1
65	Germany	1	1	1	1	1	1	1	1	1	1
66	Ghana	1	1	1	1	1	0	1	1	1	1
67	Greece	1	1	1	1	1	1	1	1	1	1
68	Grenada	1	1	1	1	1	1	1	1	1	1
69	Guatemala	1	1	1	1	1	1	1	1	1	1
70	Guinea	1	1	1	1	1	0	1	1	1	1
71	Guinea Bissau	1	1	1	1	1	0	1	1	1	1
72	Guyana	1	1	1	1	1	0	1	1	1	1
73	Haiti	1	1	1	1	1	0	1	1	1	1
74	Honduras	1	1	1	1	1	0	1	1	1	1
75	Hungary	1	1	1	1	1	1	1	1	1	1
76	Iceland	1	1	1	1	1	1	1	1	1	1
77	India	1	1	1	1	1	1	1	1	1	1
78	Indonesia	1	1	1	1	1	1	1	1	1	1
79	Iran	1	1	1	1	1	1	1	1	1	1
80	Iraq	1	1	1	1	1	1	1	1	1	1
81	Ireland	1	1	1	1	1	1	1	1	1	1
82	Israel	1	1	1	1	1	1	1	1	1	1
83	Italy	1	1	1	1	1	1	1	1	1	1
84	Jamaica	1	1	1	1	1	1	1	1	1	1
85	Japan	1	1	1	1	1	1	1	1	1	1
86	Jordan	1	1	1	1	1	1	1	1	1	1
87	Kazakhstan	1	1	1	1	1	1	1	1	1	1
88	Kenya	1	1	1	1	1	0	1	1	1	1
89	Kiribati	1	1	1	1	1	0	1	1	1	1

Soc Num	Society Name	V 121 Rec	V 129 Rec	V 143 Rec	V 149 Rec	V 150 Rec	V 151 Rec	V 153 Rec	V 155 Rec	V 157 Rec	V 158 Rec
90	North Korea	1	1	1	1	1	1	1	1	1	1
91	South Korea	1	1	1	1	1	1	1	1	1	1
92	Kuwait	1	1	1	1	1	1	1	1	1	1
93	Kyrgystan	1	1	1	1	1	0	1	1	1	1
94	Laos	1	1	1	1	1	0	1	1	1	1
95	Latvia	1	1	1	1	1	1	1	1	1	1
96	Lebanon	1	1	1	1	1	1	1	1	1	1
97	Lesotho	1	1	1	1	1	0	1	1	1	1
98	Liberia	1	1	1	1	1	0	1	1	1	1
99	Libya	1	1	1	1	1	1	1	1	1	1
100	Liechtenstein	1	1	1	1	1	1	1	1	1	1
101	Lithuania	1	1	1	1	1	1	1	1	1	1
102	Luxembourg	1	1	1	1	1	1	1	1	1	1
103	Macedonia	1	1	1	1	1	1	1	1	1	1
104	Madagascar	1	1	1	1	1	0	1	1	1	1
105	Malawi	1	1	1	1	1	0	1	1	1	1
106	Malaysia	1	1	1	1	1	1	1	1	1	1
107	Maldives	1	1	1	1	1	0	1	1	1	1
108	Mali	1	1	1	1	1	0	1	1	1	1
109	Malta	1	1	1	1	1	1	1	1	1	1
110	Marshall Islands	1	1	1	1	1	1	1	1	1	1
111	Mauritania	1	1	1	1	1	0	1	1	1	1
112	Mauritius	1	1	1	1	1	1	1	1	1	1
113	Mexico	1	1	1	1	1	1	1	1	1	1
114	Fed Micronesia	1	1	1	1	1	0	1	1	1	1
115	Moldova	1	1	1	1	1	1	1	1	1	1
116	Monaco	1	1	1	1	1	1	1	1	1	1
117	Mongolia	1	1	1	1	1	1	1	1	1	1
118	Morocco	1	1	1	1	1	1	1	1	1	1
119	Mozambique	1	1	1	1	1	0	1	1	1	1
120	Namibia	1	1	1	1	1	0	1	1	1	1
121	Nauru	1	1	1	1	1	1	1	1	1	1
122	Nepal	1	1	1	1	1	0	1	1	1	1
123	Netherlands	1	1	1	1	1	1	1	1	1	1
124	New Zealand	1	1	1	1	1	1	1	1	1	1
125	Nicaragua	1	1	1	1	1	0	1	1	1	1
126	Niger	1	1	1	1	1	0	1	1	1	1
127	Nigeria	1	1	1	1	1	0	1	1	1	1
128	Norway	1	1	1	1	1	1	1	1	1	1
129	Oman	1	1	1	1	1	1	1	1	1	1
130	Pakistan	1	1	1	1	1	1	1	1	1	1
131	Palau	1	1	1	1	1	1	1	1	1	1
132	Panama	1	1	1	1	1	0	1	1	1	1
133	Papua	1	1	1	1	1	0	1	1	1	1
134	Paraguay	1	1	1	1	1	0	1	1	1	1
135	Peru	1	1	1	1	1	1	1	1	1	1

Soc Num	Society Name	V 121 Rec	V 129 Rec	V 143 Rec	V 149 Rec	V 150 Rec	V 151 Rec	V 153 Rec	V 155 Rec	V 157 Rec	V 158 Rec
136	Philippines	1	1	1	1	1	0	1	1	1	1
137	Poland	1	1	1	1	1	1	1	1	1	1
138	Portugal	1	1	1	1	1	1	1	1	1	1
139	Qatar	1	1	1	1	1	1	1	1	1	1
140	Romania	1	1	1	1	1	1	1	1	1	1
141	Russia	1	1	1	1	1	1	1	1	1	1
142	Rwanda	1	1	1	1	1	0	1	1	1	1
143	Saint Kitts Nevis	1	1	1	1	1	1	1	1	1	1
144	Saint Lucia	1	1	1	1	1	1	1	1	1	1
145	Saint Vincentgren	1	1	1	1	1	1	1	1	1	1
146	Samoa	1	1	1	1	1	0	1	1	1	1
147	San Marino	1	1	1	1	1	1	1	1	1	1
148	Saotomeprincipe	1	1	1	1	1	1	1	1	1	1
149	Saudi Arabia	1	1	1	1	1	1	1	1	1	1
150	Senegal	1	1	1	1	1	0	1	1	1	1
151	Serbia Montenegro	1	1	1	1	1	1	1	1	1	1
152	Seychelles	1	1	1	1	1	1	1	1	1	1
153	Sierra Leone	1	1	1	1	1	1	1	1	1	1
154	Singapore	1	1	1	1	1	1	1	1	1	1
155	Slovakia	1	1	1	1	1	1	1	1	1	1
156	Slovenia	1	1	1	1	1	1	1	1	1	1
157	Solomon islands	1	1	1	1	1	0	1	1	1	1
158	Somalia	1	1	1	1	1	0	1	1	1	1
159	South Africa	1	1	1	1	1	1	1	1	1	1
160	Spain	1	1	1	1	1	1	1	1	1	1
161	Sri Lanka	1	1	1	1	1	1	1	1	1	1
162	Sudan	1	1	1	1	1	0	1	1	1	1
163	Suriname	1	1	1	1	1	1	1	1	1	1
164	Swaziland	1	1	1	1	1	0	1	1	1	1
165	Sweden	1	1	1	1	1	1	1	1	1	1
166	Syria	1	1	1	1	1	1	1	1	1	1
167	Tajikistan	1	1	1	1	1	1	1	1	1	1
168	Tanzania	1	1	1	1	1	0	1	1	1	1
169	Thailand	1	1	1	1	1	1	1	1	1	1
170	Togo	1	1	1	1	1	0	1	1	1	1
171	Tonga	1	1	1	1	1	1	1	1	1	1
172	Trinidad	1	1	1	1	1	1	1	1	1	1
173	Tunisia	1	1	1	1	1	1	1	1	1	1
174	Turkey	1	1	1	1	1	1	1	1	1	1
175	Turkmenistan	1	1	1	1	1	1	1	1	1	1
176	Tuvalu	1	1	1	1	1	0	1	1	1	1
177	Uganda	1	1	1	1	1	1	1	1	1	1
178	Ukraine	1	1	1	1	1	1	1	1	1	1
179	United Arab Emir	1	1	1	1	1	1	1	1	1	1
180	United kingdom	1	1	1	1	1	1	1	1	1	1
181	USA	1	1	1	1	1	1	1	1	1	1

Soc Num	Society Name	V 121 Rec	V 129 Rec	V 143 Rec	V 149 Rec	V 150 Rec	V 151 Rec	V 153 Rec	V 155 Rec	V 157 Rec	V 158 Rec
182	Uruguay	1	1	1	1	1	1	1	1	1	1
183	Uzbekistan	1	1	1	1	1	1	1	1	1	1
184	Vanuatu	1	1	1	1	1	0	1	1	1	1
185	Venezuela	1	1	1	1	1	1	1	1	1	1
186	Vietnam	1	1	1	1	1	1	1	1	1	1
187	Yemen	1	1	1	1	1	1	1	1	1	1
188	Zambia	1	1	1	1	1	0	1	1	1	1
189	Zimbabwe	1	1	1	1	1	1	1	1	1	1

Soc Num	Society Name	V678 Rec	V744 Rec	V754 Rec	V786 Rec	V1130 Rec	V1648 Rec	V1738 Rec
1	Afghanistan	0	0	1	1	36	0	1
2	Albania	0	1	1	1	107	0	1
3	Algeria	1	1	1	1	13	0	1
4	Andorra	1	1	1	1	146	0	1
5	Angola	0	1	1	1	11	0	1
6	Antigua	1	1	1	1	174	0	1
7	Argentina	1	1	1	1	13	0	1
8	Armenia	1	1	1	1	103	0	1
9	Australia	1	1	1	1	2	0	1
10	Austria	1	1	1	1	97	0	1
11	Azerbaijan	1	1	1	1	94	0	1
12	Bahamas	1	1	1	1	22	0	1
13	Bahrain	1	1	1	1	991	0	1
14	Bangladesh	1	0	1	1	895	0	1
15	Barbados	1	1	0	1	619	0	1
16	Belarus	1	1	1	1	48	0	1
17	Belgium	1	1	1	1	338	0	1
18	Belize	1	1	1	1	11	0	1
19	Benin	0	1	0	1	64	0	1
20	Bhutan	1	1	1	1	41	0	1
21	Bolivia	1	1	1	1	8	0	1
22	Bosnia Herzecegov	0	1	1	1	75	0	1
23	Botswana	1	1	1	1	3	0	1
24	Brazil	1	1	1	1	20	0	1
25	Brunei Darussalam	1	1	1	1	58	0	1
26	Bulgaria	1	1	1	1	72	0	1
27	Burkina Faso	0	1	1	1	41	0	1
28	Burma	1	1	1	1	71	0	1
29	Burundi	1	1	1	1	233	0	1
30	Cambodia	1	1	1	1	70	0	1
31	Cameroon	1	1	1	1	31	0	1
32	Canada	1	1	1	1	3	1	1
33	Cape Verde	1	1	1	1	112	0	1
34	Central Af Republic	1	1	1	1	6	0	1
35	Chad	1	1	1	1	6	0	1
36	Chile	1	0	1	1	20	0	1
37	China	1	1	1	1	133	0	1
38	Colombia	1	1	1	1	37	0	1
39	Comoros	1	1	1	1	313	0	1
40	Demrepub Congo	0	1	1	1	21	0	1
41	Repub Congo	1	1	1	1	10	0	1
42	Costa Rica	1	1	1	1	77	0	1
43	Cote D'Ivoire	1	0	1	1	52	0	1

Soc Num	Society Name	V678 Rec	V744 Rec	V754 Rec	V786 Rec	V1130 Rec	V1648 Rec	V1738 Rec
44	Croatia	1	1	1	1	80	0	1
45	Cuba	1	1	1	1	100	0	1
46	Cyprus	1	1	1	1	85	0	1
47	Czech Republic	1	1	1	1	130	0	1
48	Denmark	1	1	1	1	124	0	1
49	Djibouti	0	1	1	1	31	0	1
50	Dominica	1	1	1	1	104	0	1
51	Dominican Republic	0	1	1	1	170	0	1
52	Ecuador	1	1	1	1	43	0	1
53	Egypt	1	1	1	1	67	1	1
54	El Salvador	1	1	1	1	298	0	1
55	Equatorial Guinea	1	1	0	1	16	0	1
56	Eritrea	1	1	1	1	30	0	1
57	Estonia	1	1	1	1	30	0	1
58	Ethiopia	0	1	1	1	62	0	1
59	Fiji	1	1	0	1	44	0	1
60	Finland	1	1	1	1	15	0	1
61	France	1	1	1	1	107	1	1
62	Gabon	1	0	1	1	5	0	1
63	Gambia	1	1	1	1	117	0	1
64	Georgia	1	1	1	1	68	0	1
65	Germany	1	1	1	1	231	0	1
66	Ghana	1	1	1	1	83	0	1
67	Greece	1	1	1	1	83	0	1
68	Grenada	1	1	1	1	295	0	1
69	Guatemala	1	1	0	1	103	0	1
70	Guinea	1	1	1	1	34	0	1
71	Guinea Bissau	1	1	1	1	38	0	1
72	Guyana	1	1	1	1	3	0	1
73	Haiti	0	1	1	1	286	0	1
74	Honduras	1	1	1	1	57	0	1
75	Hungary	1	1	1	1	110	0	1
76	Iceland	1	1	1	1	3	0	1
77	India	1	1	1	1	311	0	1
78	Indonesia	1	1	1	1	110	0	1
79	Iran	1	1	1	1	40	0	1
80	Iraq	0	1	1	1	57	1	1
81	Ireland	1	1	1	1	54	0	1
82	Israel	1	1	1	1	275	0	1
83	Italy	1	1	1	1	192	1	1
84	Jamaica	1	1	1	1	235	0	1
85	Japan	1	0	1	1	336	0	1
86	Jordan	1	1	1	1	56	0	1
87	Kazakhstan	1	1	1	1	6	0	1
88	Kenya	1	1	1	1	53	0	1
89	Kiribati	1	0	1	1	123	0	1

Soc Num	Society Name	V678 Rec	V744 Rec	V754 Rec	V786 Rec	V1130 Rec	V1648 Rec	V1738 Rec
90	North Korea	0	1	1	1	181	0	1
91	South Korea	1	1	1	1	472	0	1
92	Kuwait	1	1	1	1	125	1	1
93	Kyrgystan	0	1	1	1	25	0	1
94	Laos	1	1	1	1	22	0	1
95	Latvia	1	1	1	1	37	0	1
96	Lebanon	1	0	1	1	327	0	1
97	Lesotho	0	1	1	1	59	0	1
98	Liberia	1	1	1	1	28	0	1
99	Libya	1	1	1	1	3	0	1
100	Liechtenstein	1	1	1	1	205	0	1
101	Lithuania	1	1	1	1	54	0	1
102	Luxembourg	1	1	1	1	168	0	1
103	Macedonia	1	1	1	1	78	0	1
104	Madagascar	1	1	1	1	28	0	1
105	Malawi	1	1	1	1	97	0	1
106	Malaysia	1	1	1	1	70	0	1
107	Maldives	1	1	1	1	974	0	1
108	Mali	1	1	1	1	9	0	1
109	Malta	1	0	1	1	1240	0	1
110	Marshall Islands	1	1	0	1	288	0	1
111	Mauritania	1	0	1	1	3	0	1
112	Mauritius	1	1	1	1	581	0	1
113	Mexico	1	0	1	1	51	0	1
114	Fed Micronesia	1	1	1	1	153	0	1
115	Moldova	1	1	1	1	127	0	1
116	Monaco	1	0	1	1	22403	0	1
117	Mongolia	1	0	1	1	2	0	1
118	Morocco	1	1	1	1	65	1	1
119	Mozambique	1	1	1	1	22	0	1
120	Namibia	1	1	1	1	2	0	1
121	Nauru	1	1	1	1	581	0	1
122	Nepal	1	1	1	1	174	0	1
123	Netherlands	1	1	1	1	389	0	1
124	New Zealand	1	1	1	1	14	0	1
125	Nicaragua	0	1	1	1	38	0	1
126	Niger	0	1	1	1	9	0	1
127	Nigeria	1	1	0	1	127	0	1
128	Norway	1	1	1	1	14	0	1
129	Oman	1	1	1	1	11	1	1
130	Pakistan	0	1	1	1	179	0	1
131	Palau	1	1	1	1	42	0	1
132	Panama	1	1	1	1	39	0	1
133	Papua	1	1	1	1	11	0	1
134	Paraguay	1	1	1	1	13	0	1
135	Peru	1	1	1	1	20	0	1

Soc Num	Society Name	V678 Rec	V744 Rec	V754 Rec	V786 Rec	V1130 Rec	V1648 Rec	V1738 Rec
136	Philippines	0	1	1	1	253	0	1
137	Poland	1	1	1	1	120	0	1
138	Portugal	1	1	1	1	111	0	1
139	Qatar	1	1	1	1	55	1	1
140	Romania	1	1	1	1	93	0	1
141	Russia	1	1	1	1	9	0	1
142	Rwanda	0	1	1	1	305	0	1
143	Saint Kitts Nevis	1	1	1	1	155	0	1
144	Saint Lucia	1	1	1	1	248	0	1
145	Saint Vincentgren	1	1	1	1	299	0	1
146	Samoa	1	1	1	1	63	0	1
147	San Marino	1	0	1	1	442	0	1
148	Saotomeprincipe	0	1	1	1	145	0	1
149	Saudi Arabia	1	1	1	1	10	1	1
150	Senegal	0	1	0	1	53	0	1
151	Serbia Montenegro	1	1	1	1	103	0	1
152	Seychelles	1	1	0	1	170	0	1
153	Sierra Leone	1	1	1	1	63	0	1
154	Singapore	1	1	1	1	6501	0	1
155	Slovakia	1	1	1	1	110	0	1
156	Slovenia	1	1	1	1	97	0	1
157	Solomon islands	1	1	1	1	14	0	1
158	Somalia	0	1	1	1	11	0	1
159	South Africa	1	1	0	1	37	0	1
160	Spain	1	1	1	1	81	0	1
161	Sri Lanka	1	0	1	1	303	0	1
162	Sudan	0	1	1	1	13	0	1
163	Suriname	1	1	1	1	3	0	1
164	Swaziland	1	1	1	1	59	0	1
165	Sweden	1	1	1	1	20	0	1
166	Syria	1	1	1	1	91	1	1
167	Tajikistan	0	1	0	1	43	0	1
168	Tanzania	1	1	1	1	37	0	1
169	Thailand	1	1	0	1	120	0	1
170	Togo	0	1	1	1	94	0	1
171	Tonga	1	1	1	1	134	0	1
172	Trinidad	0	1	0	1	250	0	1
173	Tunisia	1	1	1	1	58	0	1
174	Turkey	1	0	1	1	88	0	1
175	Turkmenistan	1	0	1	1	9	0	1
176	Tuvalu	1	1	0	1	392	0	1
177	Uganda	1	1	1	1	103	0	1
178	Ukraine	1	1	1	1	81	0	1
179	United Arab Emir	1	1	1	1	39	1	1
180	United kingdom	1	1	1	1	240	1	1
181	USA	1	1	1	1	30	1	1

Soc Num	Society Name	V678 Rec	V744 Rec	V754 Rec	V786 Rec	V1130 Rec	V1648 Rec	V1738 Rec
182	Uruguay	1	0	1	1	19	0	1
183	Uzbekistan	1	0	1	1	55	0	1
184	Vanuatu	1	1	1	1	16	0	1
185	Venezuela	1	1	1	1	27	0	1
186	Vietnam	1	1	1	1	237	0	1
187	Yemen	1	0	1	1	34	0	1
188	Zambia	1	1	1	1	14	0	1
189	Zimbabwe	1	0	1	1	32	0	1

APPENDIX 3. DEFINITIONS OF DATA CODED IN APPENDIX 2

In this Appendix 25 multinomial SCCS constructs (Divale 2004b) are redefined to be binary constructs X = 1, 0. For example, SCCS V1, INTERCOMMUNITY TRADE AS FOOD SOURCE has seven categories the last two of which are, “6 = < 50% of food, and more than any single local source; 7 = > 50% of food.” V1 becomes a binary variable V1Rec = 1, 0 by recoding it V1Rec = 1 if SCCS V1 = 7 , V1Rec = 0 if SCCS V1 = 1 OR 2 OR 3 OR 4 OR 5 OR 6. One additional SCCS construct is redefined (V1130Rec) to approximate a continuous construct V1130Rec > 0.

DEVELOP = 1, 2, 3 is the United Nations development codes least developed, less developed excluding least developed, and developed. Excluding DEVELOP, SCCS coded data for the constructs of this Appendix may be found in Divale (2004b).

The definitions used to recode SCCS variables are applied in Appendix 2 to a sample consisting of all 189 member countries of the United Nations as of 2000 AD. Codes for the 189 UN member countries at 2000 AD were created by the writer by applying the definitions of this Appendix to descriptions of the 189 countries in Ember and Ember’s (2001) *Countries and their Cultures* (CC). Any exceptions are noted below. The result of applying the recoded SCCS definitions to both SCCS societies and the 189 UN member countries at 2000 AD is a combined sample of 363 societies – 174 SCCS societies (from 1800 to 1965 AD) plus 189 UN member countries at 2000 AD. There are no missing data for the 189 UN member countries in Appendix 2. SCCS sample elements for which there are missing data may be determined from Divale (2004b).

Coding Rule 1: V1648Rec (war) codes for the 189 UN member countries at 2000 AD are based on SCCS code definitions applied to the Correlates of War (COW) data set V3.0 (Sarkees 2000) for the first complete 10-year interval preceding 2000 AD (http://cow2.la.psu.edu/). United Nations (2005) data (http://esa.un.org/unpp/index.asp) are used to code the 189 UN member countries at 2000 AD for the following constructs: DEVLOP, V63Rec, V1130Rec.

Coding Rule 2: Suppose diversity exists in a society. For example, suppose X = 1 denotes independent family, X = 0 denotes extended family, and both forms exist in a country. In such a case the form in which the greatest proportion of people lives is coded. If several ethnic groups are present the largest ethnic group is coded.

Coding Rule 3: The unit of observation is, “society” (see text). If a construct is a trait of local community (e.g. settlement size) code the type of local community in which the greatest proportion of individuals in the society live. First apply Rule 2, then Rule 3.

Coding Rule 4: If data for UN member countries at 2000 AD are lacking in Ember and Ember (2001, abbreviated CC here) check CIA Country Studies (http://lcweb2.loc.gov/frd/cs/cshome.html). UN data are used for DEVLOP, V63Rec and V1130Rec. Murdock (1981) was used occasionally to resolve presence/absence of bilateral kinship (V70Rec). Lichtenstein, a UN member state strangely omitted from Ember and Ember (2001), is coded the same as France and Germany for all but DEVLOP, V63Rec and V1130Rec (UN data) and V1648Rec (COW data).

Coding Rule 5: An event is defined to be rare or uncommon if it occurs less frequently than 20% of the time.

For definitions of anthropological constructs see, *inter alia*, Ember and Ember (2004).

The writer is of the opinion that he may have misapplied the coding practices of Murdock and Provost (1973) when coding CC countries for V151Rec (intensive agriculture). This matter is discussed in Appendix 1, section 1.3. Based on reading Ember and Ember (2001) the writer is of the opinion that, for V70Rec (bilateral kinship), there appears to be a time lag between an increase in modernization $m \geq 0$ in a contemporary, developing country and the ultimate appearance of the form of kinship suited to it.

Column Content	SCCS Variable Number; Recoded Number	Recoding of SCCS Variable	Meaning of Recoded Categories V.Rec = 1,0	Notes
SOCNO				Country Number
SOCNAME				Country Name
DEVELOP				UN codes 1 = least developed country 2 = less developed country 3 = developed country
Intercommunity Trade as Food Source	V1 V1Rec	X = 1 if V1=7 Else X = 0	1 = >50% food from outside community 0 = otherwise	X = 0 denotes preponderance of food locally produced
Credit Source	V18 V18Rec	X = 1 if V18 = (2 OR 3 OR 5) Else X = 0	1 = money lending specialists present 0 = personal loans between friends or relatives	
Compactness of Settlement	V62 V62Rec	X=1 if V62 = 1 Else X = 0	1 = compact 0 = otherwise	
Community Size	V63 V63Rec	X=1 if V63 = (7 OR 8) Else X = 0	1 = average community $\geq$ 5000 population 0 = otherwise	UN data used to code CC countries, Barbados %urban=50.0. Coded V63=0. Flipping a coin resulted in V63=1

Column Content	SCCS Variable Number; Recoded Number	Recoding of SCCS Variable	Meaning of Recoded Categories V.Rec = 1,0	Notes
Large or Impressive Structures	V66 V66Rec	X=1 if V66 = (2 OR 3 OR 4 OR 5 OR 6) Else X = 0	1 = present 0 = otherwise	
Form of Family	V68 V68Rec	X=1 if v68 = (1 OR 2 OR 3 OR 4) Else X = 0	1 = modal family independent 0 = otherwise	Independent family is a one husband-wife unit, possibly polygamous. X = 0 denotes several such related units in a single household.
Marital Residence	V69 V69Rec	X=1 if V69 = 5 Else X = 0	1 = neolocal 0 = otherwise	Neolocal denotes residence at marriage not based primarily on location of kin. Codes exclude special residence in first year of marriage
Descent-Membership in Corporate Kinship Groups	V70 V70Rec	X=1 if V70 = 5 Else X = 0	1 = bilateral kinship 0 = otherwise	Bilateral denotes kin relatedness on both mo and fa sides.
Polygamy	V79 V79Rec	X=1 if V79 = (1 OR 2 OR 3) Else X = 0	1 = monogamy predominates, polygamy rare 0 = otherwise	
Cooking	V121 V121Rec	X=1 if V121 = (4 OR 5) Else X = 0	1 = females predominate 0 = otherwise	
Smelting	V129 V129Rec	X=1 if V129 = (. OR 1 OR 2 OR 3 OR 4 OR 5) Else X = 0	1 = present 0 = otherwise	SCCS V129 code dot (.) denotes task present
Laundering	V143 V143Rec	X=1 if V121 = (4 OR 5) Else X = 0	1 = females predominate 0 = otherwise	
Writing and Records	V149 V149Rec	X=1 if V149 = 5 Else X = 0	1 = true writing present 0 = otherwise	
Fixity of Residence	V150 V150Rec	X=1 if V150 = 5 Else X = 0	1 = sedentary permanent 0 = otherwise	X = 1 denotes local community does not change location
Agriculture	V151 V151Rec	X=1 if V151 = 5 Else X = 0	1 = mainly primary intensive 0 = otherwise	

Column Content	SCCS Variable Number; Recoded Number	Recoding of SCCS Variable	Meaning of Recoded Categories V.Rec = 1,0	Notes
Technological Specialization	V153 V153Rec	X=1 if V153 = 5 Else X = 0	1 = smiths, weavers, potters present 0 = otherwise	X = 1 denotes all 3 activities
Money	V155 V155Rec	X=1 if V155 = 5 Else X = 0	1 = true money present 0 = otherwise	X = 0 denotes true money absent
Political Integration	V157 V157Rec	X=1 if V157 = (3 OR 4 OR 5) Else X = 0	1 = state (above local community) 0 = otherwise	X = 1 denotes political integration above local community, i.e. state
Social Stratification	V158 V158Rec	X=1 if V158 = (2 OR 3 OR 4 OR 5) Else X = 0	1 = non-egalitarian 0 = otherwise	X = 1 denotes classes, castes or slavery present
Food Stress or Hunger	V678 V678Rec	X=1 if V678 = 1 Else X = 0	1 = food supply constant 0 = otherwise	X = 0 denotes occasional famine, chronic famine, etc.
Frequency of Divorce	V744 V744Rec	X=1 if V744 = 1 OR 2 Else X = 0	1 = universal, frequent, common, not uncommon 0 = otherwise	Dissolution of stable, common law unions counts as divorce
Wife-Beating	V754 V754Rec	X=1 if V754 = 1 Else X = 0	1 = absent or rare 0 = otherwise	X = 0 denotes more than rare occurrence
Adult Mobility	V786 V786Rec	X = 1 if V786 = (2 OR 3) Else X = 0	1 = movement to other community as adult common or occasional 0 = individuals generally attached to particular community throughout life especially after marriage	

Column Content	SCCS Variable Number; Recoded Number	Recoding of SCCS Variable	Meaning of Recoded Categories V.Rec = 1,0	Notes
Population Density	V1130 V1130Rec	If V1130 = 3 then X = .3861((4.9 - 1)/2 + 1); If V1130 = 4 then X = .3861((24.9 - 5)/2 + 5); If V1130 = 5 then X = .3861((99.9 - 25)/2 + 25); If V1130 = 6 then X = .3861((499.9 - 100)/2 + 100); If V1130 = 7 then X = .3861(750); If V1130 = 2 then X = .3861(.5); If V1130 = . then V1130Rec = .;	Use UN 2000 AD density (persons per square km)	UN density is a ratio variable X ≥ 0. SCCS "persons per square mile" multiplied by .3861 are converted to persons per square km.
Overall Frequency of War	V1648 V1648Rec	X = 0 if V168 = (1 OR 2 OR 3 OR 4 OR 5) Else X = 1	1 = war once each 3-10 years or more frequently 0 = war absent or rare	SCCS code V1648 = 0 excluded
Presence of Formal Education within Local Community	V1738 V1738Rec	X = 1 if V1738 = (2 OR 3 OR 4 OR 5) Else X = 0	1 = formal education present at least for some 0 = absent	