Discrete dynamic models: A HMM approach to estimation and forecasting using panel survey data

Tiago Ribeiro[1]

November 6, 2002

[1]This paper is part of my dissertation. I would like to express my gratitude to my dissertation adviser, Prof. Daniel McFadden, for innumerous discussions, guidance and teachings. I would also like to thank Peter Bickel, Paul Gertler, Bronwyn Hall, Paul Ruud and Till von Wachter for helpful discussions and suggestions. All errors are of my own responsibility. This work was supported by scholarship PRAXIS XXI/BD/16014/98 from the Portuguese Fundação para a Ciência e a Tecnologia

Contents

1 Introduction

This paper describes a discrete vector valued dynamic model and its implementation. This work is motivated by problems arising in the use of individual survey data to forecast the simultaneous behavior of certain characteristics of a population.

The contribution of this paper is to propose a model of the joint distribution of a high dimensional vector of discrete variables and of its evolution over time. We do so by means of a Hidden Markov Model (HMM). The model combines features of multiple indicator-multiple cause (MIMIC) models[1] and LISREL factor-analytic models[2] in the sense that some interpretable structure is imposed on the relations between a given number of latent variables that generate the observations with features of filtering algorithms like the Kalman filter which allow us to estimate the evolution over time of the latent variables. In fact, in its most simple form, the model proposed here can be seen as the discrete counterpart of a Kalman filter, where both the latent and observed variables are discrete.

The model proposed performs a dimension reduction of chosen groups of variables, which are each represented by a latent variable whose dynamics are modelled. It can also constitute a building block of a larger model used in forecasting and policy analysis. It further has the advantages that it can both be estimated without use of simulation (the EM-algorithm is used) and can also incorporate directly missing observations hence removing the need for separate imputation procedures.

To the best of our knowledge this is the first application, in economics, of an HMM to panel survey data with the above characteristics. Most of the applications of the HMM framework in economics are in the area of finance and constitute mostly time series applications of HMM to a single series or applications to switching regression models.

For the purposes of illustration we model the evolution of health and wealth of the oldest old population using and HMM and present some preliminary and mostly descriptive results based on data from the AHEAD survey. The relevant characteristics of the population under study are the health status as characterized by the presence or absence of certain health conditions and the wealth status of the oldest old individuals as defined by their portfolio. The HMM formulation tries to capture the effect of multiple health risks and allows for the possibility of testing causality between different health and wealth components.

This paper is organized as follows: section 2 describes a simple HMM and uses this simplified model to derive the algorithms used in estimation and testing and to discuss general approximating properties of hidden Markov models; section 3 presents the formulation of a problem in the area of economics of aging in terms

[1]See Borsch-Supan, McFadden & Schnabel (1996) for an application of MIMIC models in the economics of aging literature

[2]See Joreskog (1973) for an extensive discussion

of HMM and discusses interpretation, testing, incorporation of missing values and forecasting; section 4 gives a brief description of the AHEAD data, with particular focus on the patterns of missing data for the wealth related questions of the survey; section 5 illustrates the methodology proposed by describing the evolution of wealth portfolios composed of 12 different types of assets over a 8 year period using an HMM as well as the evolution of 12 health conditions.

2 Hidden Markov Models - HMM

A Hidden Markov model (HMM) is generally in the following way: if $\{X_t\}_{t=1}^{\infty}$ is a (first order) Markov process in a finite state space and $\{Y_t\}_{t=1}^{\infty}$ is a deterministic or stochastic function of $\{X_t\}_{t=1}^{\infty}$ then $\{Y_t\}_{t=1}^{\infty}$ is a HMM. Usually $\{Y_t\}_{t=1}^{\infty}$ depends on $\{X_t\}_{t=1}^{\infty}$ only locally. Moreover $\{Y_t\}_{t=1}^{\infty}$ is generally not Markov and can have a complicated dependence structure. Still, the conditional distribution of $\{X_t\}_{t=1}^{\infty}$ given $\{Y_t\}_{t=1}^{\infty}$ is simple in the sense that it has a first order Markov structure[3]. Generally $\{Y_t\}_{t=1}^{T}$ is an observed sequence whereas $\{X_t\}_{t=1}^{T}$ is assumed to exist but is not observed. The possibility of modelling complex structures of $\{Y_t\}_{t=1}^{\infty}$ through a simple formulation makes HMMs attractive and they have been been applied in a wide range of areas: speech recognition (Rabiner (1989)), neurophysiology (Fredkin & Rice (1986)) and biology (Leroux & Puterman (1992)). A well known tool in the economics and econometrics literature with essentially the same structure is the Kalman filter. The HMM is basically the same model but both the state and observed variables are discrete. As we describe the estimation procedures for HMM we will frequently refer to the Kalman filter in order to make this analogy more concrete. Another frequent application in economics of similar models occurs in time-series under the name of Markov-switching or regression-switching models[4] .

2.1 Notation

The following notation will be used throughout the paper: $\{X_t\}_{t=1}^{\infty}$ will denote a finite-state, homogeneous, discrete time Markov chain. This chain is not observed. The state space of X_t has K elements which will be identified with the set $S_X = \{e_1, \ldots, e_K\}$ where e_i are unit vectors in $\mathbf{R}^K$ with unity in the i^{th} component. We may write $X_t = i$ by which we mean $X_t = e_i$ and X_t^i will denote the i^{th} component of X_t (i.e. $X_t = i$ means $X_t^i = 1$ and $X_t^j = 0\ \forall j \neq i$) . The sequence $\{X_t\}_{t=1}^{\infty}$ will be first order Markov, that is

$$P(X_{t+1}|X_1, \ldots, X_t) = P(X_{t+1}|X_t)$$

[3]It also usually assumed that $\{X_t\}_{t=1}^{\infty}$ is an homogenous first order Markov chain. $\{X_t\}_{t=1}^{\infty}|\{Y_t\}_{t=1}^{\infty}$ will still be first order Markov but not homogeneous

[4]See Hamilton for a survey of that literature

and the transition matrix will be denoted by $A = [a_{ji}]$ where a_{ji} is defined as

$a_{ji} = P(X_{t+1} = j|X_t = i)$

Note that by definition the columns of A add to 1 (i.e. $\underline{1}'A = \underline{1}'$ where $\underline{1}$ is a conformable vector of 1's) and that

$E(X_{t+1}|X_t) = AX_t$

Define the residual V_t by difference:

$V_t = X_t - E(X_t|X_{t-1}) = X_t - AX_{t-1}$

Similarly $\{Y_t\}_{t=1}^T$ will denote the sequence of observations. Y_t takes values on a discrete set of M elements. This set will be identified with the set $S_Y = \{f_1, \ldots, f_M\}$ where f_i are unit vectors in $\mathbf{R}^M$ with unity in the i^{th} component. Y_t depends only on current X_t that is

$P(Y_t|X_1, \ldots, X_t, Y_1, \ldots, Y_{t-1}) = P(Y_t|X_t)$

for all t and we let the matrix $C = [c_{ji}]$ define the mapping from X_t to Y_t, where

$c_{ji} = P(Y_t = j|X_t = i)$

With this notation we also have that

$E(Y_t|X_t) = CX_t$

and by definition the columns of C sum to 1 ($\underline{1}'C = \underline{1}'$). Similarly define the residual U_t as:

$U_t = Y_t - E(Y_t|X_t) = Y_t - AX_t$

The marginal distribution of X_1 will be denoted by π.

In order to shorten the notation in some long expressions we will replace $X_1, \ldots, X_t$ by $X_{1:t}$.

2.2 Graphical representation

The model just described is represented in the diagram of figure 1 using graphical models' schematics. We follow the convention of denoting observed variables by shaded nodes and latent (unobserved) variables by empty nodes. This representation is useful in two ways: it is a succinct representation of the joint distribution of $(X_1, \ldots, X_T, Y_1, \ldots, Y_T)$ and it provides an economical way of stating substantive assumptions. This follows because we can, in a mechanical way, recover the conditional independence relations of the joint distribution from the graphical representation (Pearl (2000, Theorem 1.2.5, page 18)). This representation will become useful in the description of our extension of model to characterize the evolution of health and wealth among the oldest old and especially in the discussion of the meaning of the causal relations implied by it.

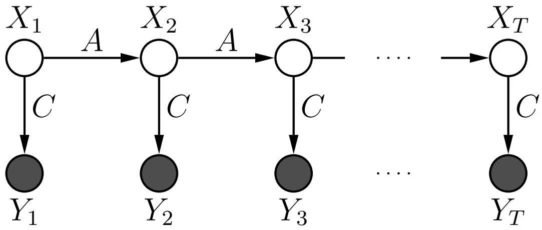


Figure 1: **HMM**

2.3 Estimation

We review the most common estimation procedures used in HMM. The present exposition of the estimation algorithms used for this simple HMM carries over virtually unchanged to the models discussed in section 3. We assume that we are given N i.i.d. observations of the paths $Y_{1:T}$. The full likelihood (that is the likelihood if we observed both X and Y) of a given sequence can be factorized as[5]:

$$\begin{aligned} P(X_{1:t}, Y_{1:t}, \theta) &= P(X_1) \prod_{t=1}^{T} P(Y_t|X_t) \prod_{t=2}^{T} P(X_t|X_{t-1}) & (1) \\ &= \prod_{i=1}^{K} \pi_i^{X_1^i} \prod_{t=1}^{T} \prod_{i=1}^{K} \prod_{j=1}^{M} c_{ji}^{Y_t^j X_t^i} \prod_{t=2}^{T} \prod_{i=1}^{K} \prod_{j=1}^{K} a_{ji}^{X_t^j X_{t-1}^i} & (2) \end{aligned}$$

where θ is a vector containing the elements of π, A and C. To obtain the likelihood of the observed data one should integrate (sum in this case) the above expression with respect to the distribution of the unobserved X. This operation can be performed in a efficient way as will be described. One can then use maximization routines to obtain estimates of the parameters of interest. Another approach is to use the EM-algorithm (Dempster, Laird & Rubin (1977), McLachlan & Krishnan (1997)) which operates directly with the full likelihood which has a much simpler form and whose maximization is much simpler. The EM algorithm, however, has the disadvantage of progressing only linearly close to the maximum of a function yielding a slow convergence. We will therefore use a combination of both optimization routines as suggested in Ruud (1991). Initially the EM will be used and when convergence slows down a gradient method is employed. We describe both approaches in turn.

[5]We omit the sample index n for ease of notation

The full log-likelihood of an observation is:

$$
\begin{aligned}
l(\theta, X, Y) &= \sum_{i=1}^{K} X_1^i \ln(\pi_i) + \\
& \sum_{i=1}^{K}\sum_{j=1}^{M}\left[\sum_{t=1}^{T} Y_t^j X_t^i\right] \ln(c_{ji}) + \\
& \sum_{i=1}^{K}\sum_{j=1}^{K}\left[\sum_{t=2}^{T} X_t^j X_{t-1}^i\right] \ln(a_{ji}) \\
&= \sum_{i=1}^{K} X_1^i \ln(\pi_i) + \sum_{i=1}^{K}\sum_{j=1}^{M} m_{ji}^{YX} \ln(c_{ji}) + \sum_{i=1}^{K}\sum_{j=1}^{K} m_{ji}^{XX} \ln(a_{ji})
\end{aligned}
$$

In order to maximize the log-likelihood of the observed data the via the EM algorithm to the full log-likelihood $\mathcal{L}(\theta, X, Y) = \sum_{n=1}^{N} l_n(\theta, X, Y)$. The EM-algorithm is an iterative procedure in which each iteration is composed of two steps: an E-step in which a conditional expectation is calculated and a M-step, a maximization step. In the E-step of iteration $k+1$ we calculate $\mathcal{Q}(\theta|Y, \theta_k) = E\left(\mathcal{L}(\theta, X, Y)\,|Y, \theta_k\right)$, where θ_k are the parameters estimated in iteration k, and in the M-step we maximize $\mathcal{Q}(\theta|Y, \theta_k)$ with respect to θ to obtain θ_{k+1}. In the present case $\mathcal{Q}(\theta|Y, \theta_k)$ has a very simple form:

$$
\begin{aligned}
\mathcal{Q}(\theta|Y, \theta_k) &= \sum_{i=1}^{K}\sum_{n=1}^{N} E\left(X_{1,n}^i\,|Y, \theta_k\right) \ln(\pi_i) + & (3) \\
& \sum_{i=1}^{K}\sum_{j=1}^{M}\left[\sum_{n=1}^{N}\sum_{t=1}^{T} E\left(Y_{t,n}^j X_{t,n}^i\,|Y, \theta_k\right)\right] \ln(c_{ji}) + & (4) \\
& \sum_{i=1}^{K}\sum_{j=1}^{K}\left[\sum_{n=1}^{N}\sum_{t=2}^{T} E\left(X_{t,n}^j X_{t-1,n}^i\,|Y, \theta_k\right)\right] \ln(a_{ji}) & (5) \\
&= \sum_{i=1}^{K} \hat{X}_1^i \ln(\pi_i) + \sum_{i=1}^{K}\sum_{j=1}^{M} \hat{m}_{ji}^{YX} \ln(c_{ji}) + \sum_{i=1}^{K}\sum_{j=1}^{K} \hat{m}_{ji}^{XX} \ln(a_{ji}) & (6)
\end{aligned}
$$

Once $E\left(X_{1,n}^i\,|Y, \theta_k\right)$, $E\left(Y_{t,n}^j X_{t,n}^i\,|Y, \theta_k\right)$ and $E\left(X_{t,n}^j X_{t-1,n}^i\,|Y, \theta_k\right)$ are computed (the E-step) maximization is straightforward. In order to be able to calculate these quantities all we need is the posterior distribution of X given the observed data Y. More specifically we need to calculate $P(X_t|Y_1, \ldots, Y_T, \theta)$ and $P(X_{t-1}, X_t|Y_1, \ldots, Y_T, \theta)$ for $t = 1, \ldots, T$. Since the algorithms to calculate these are important in themselves we will describe them in detail.

2.3.1 Filtered estimates of the latent states

A recursion, forward in time, to compute the filtered estimates of the latent state X_t defined as $P(X_t|Y_1,\ldots,Y_t)$ can be derived as follows:

$$
\begin{aligned}
\alpha(X_{t+1}) &\triangleq P(Y_{1:t+1}, X_{t+1}) \\
&= \sum_{X_t \in S_X} P(Y_{1:t+1}, X_t, X_{t+1}) \\
&= \sum_{X_t \in S_X} P(Y_{1:t+1}, X_{t+1}|X_t)P(X_t) \\
&= \sum_{X_t \in S_X} P(Y_{1:t}|X_t)P(X_{t+1}|X_t)P(Y_{t+1}|X_{t+1}, X_t)P(X_t) \\
&= \sum_{X_t \in S_X} \alpha(X_t)P(X_{t+1}|X_t)P(Y_{t+1}|X_{t+1}) \qquad (7)
\end{aligned}
$$

The quantity $\alpha(X_t)$ can be interpreted as an unnormalized conditional probability function of X_t where the conditioning is on information available up to time t. The filtered estimate is directly obtained by normalization, $P(X_t|Y_1,\ldots,Y_t) = \alpha(X_t)/\sum_{X_t}\alpha(X_t)$

In vector notation the above recursion translates into $\alpha_{t+1} = \text{diag}(A\cdot\alpha_t)C'Y_{t+1}$ where the i^{th} entry of the vector α_t is $\alpha(e_i)$, $\alpha_1 = \text{diag}(\pi)C'Y_1$ and 'diag$(\cdot)$' denotes a diagonal matrix with the argument in the main diagonal. We note that this recursion is the discrete time analog of the Kalman filtering recursions. One important aspect of this recursion is that $\sum_{X_T\in S_X}\alpha(X_T) = P(Y_1,\ldots,Y_T)$, so at the end of the recursion we obtain, at no extra cost, the likelihood of the observed data. This again parallels with the more familiar use of the Kalman filter to evaluate the likelihood of the observed data in an ARMA model (see for example Hamilton (1994) chapter 13).

2.3.2 Smoothed estimates of the latent states

To finally obtain $P(X_t|Y_1, \dots, Y_T)$ we decompose it as follows:

$$
\begin{aligned}
\gamma(X_t) &\triangleq P(X_t|Y_{1:T}) \\
&= \sum_{X_{t+1} \in S_X} P(X_t, X_{t+1}|Y_{1:T}) \\
&= \sum_{X_{t+1} \in S_X} P(X_t|X_{t+1}Y_{1:T})P(X_{t+1}|Y_{1:T}) \\
&= \sum_{X_{t+1} \in S_X} \frac{P(X_t, X_{t+1}, Y_{1:t})}{\sum_{X_t \in S_X} P(X_t, X_{t+1}, Y_{1:t})} \gamma(X_{t+1}) \\
&= \sum_{X_{t+1} \in S_X} \frac{P(X_{t+1}|X_t)\alpha(X_t)}{\sum_{X_t \in S_X} P(X_{t+1}|X_t)\alpha(X_t)} \gamma(X_{t+1})
\end{aligned}
$$

This derivation establishes a recursion which allows the calculation of $\gamma(X_t) = P(X_t|Y_1, \dots, Y_T)$ for $t = 1, \dots, T$. This recursion begins with $\gamma(X_T) = \alpha(X_T)/\sum_{X_T} \alpha(X_T)$ by definition of γ and α. In a similar way one can obtain a recursion to compute $P(X_t, X_{t+1}|Y_1, \dots, Y_T)$.

$$
\begin{aligned}
P(X_t, X_{t+1}|Y_{1:T}) &= \frac{P(Y_{1:T}|X_t, X_{t+1})P(X_{t+1}|X_t)P(X_t)}{P(Y_{1:T})} \\
&= \frac{P(Y_{1:t}|X_t)P(Y_{t+1:T}|X_{t+1})P(X_{t+1}|X_t)P(X_t)}{P(Y_{1:T})} \\
&= \frac{\alpha(X_t)P(Y_{t+1}|X_{t+1})P(Y_{t+2:T}|X_{t+1})P(X_{t+1}|X_t)}{P(Y_{1:T})} \\
&= \alpha(X_t)P(Y_{t+1}|X_{t+1})P(X_{t+1}|X_t)\frac{\gamma(X_{t+1})}{\alpha(X_{t+1})}
\end{aligned}
$$

We have thus derived a forward-backward algorithm that computes the posterior probability of the latent variables given the observations. This algorithm coupled with the maximization step described earlier forms what is know in the HMM literature as the Baum-Welsh updates (Baum & Petrie (1966), Baum, Petrie, Soules & Weiss (1970)). A more detailed discussion of these algorithms is given for example in Rabiner (1989).

Inspecting the algorithms described it can be seen that the complexity of the computation of the log-likelihood is linear in the number of observations, number of periods, observed variables and quadratic in the number of underlying states, so implicit integration over the unobserved X variables can be performed efficiently.

2.3.3 Gradient methods

Using similar procedures one can also compute the score of the observed likelihood from the complete data. It is well known that the score of the observed log-likelihood is related to the score of the full log-likelihood by $s(Y,\theta) = E\left(s(X,Y,\theta)\,|Y,\theta\right)$ (Ruud (1991) lemma pag. 318). Since the dependence of $s(X,Y,\theta)$ on X is linear on the sufficient statistics computed above the calculation of the observed score is straightforward and is equal to $\partial\mathcal{Q}(\theta'|Y,\theta)/\partial\theta'|_\theta$ where differentiation is with respect to the first argument. The Hessian of the observed log-likelihood can be computed using a general identity for models with missing data (see Louis (1992) page 227),

$$D^2 l(Y,\theta) = E\left[D^2 l(X,Y,\theta)\big|\,Y,\theta\right] + E\left[s(X,Y,\theta)s(X,Y,\theta)'|\,Y,\theta\right] - s(Y,\theta)s(Y,\theta)'$$

where D denotes the differential operator. It is generally computationally intensive to calculate the second term on the right hand side of the expression above, since it involves estimating higher order moments of the latent variables. So for purposes of maximization the BHHH algorithm is particularly convenient.

It is nevertheless of interest to calculate the Hessian for purposes of variance estimation. To do this it may be useful to turn to methods that allow a simple estimation of general functions of latent variables.

2.3.4 A more general filter

The alternative methods described in this section allow us to to construct the expectation of more general functions of the latent variables. These methods are based on the so-called reference probability methods. Elliott, Aggoun & Moore (1995) provide an extensive treatment of the topic. The use of these techniques avoids the forward-backward recursion of Baum-Welsh type algorithms. The following exposition is based on Elliott et al. (1995, Theorem 5.3, page 31).

It is useful to note in advance that:

$$\begin{aligned}
\beta(X_{t-1}) &\triangleq P(X_{t-1}|X_t, Y_{1:t}) \\
&= \frac{P(X_{t-1}, X_t, Y_{1:t})}{\sum_{X_{t-1}\in S_X} P(X_{t-1}, X_t, Y_{1:t})} \\
&= \frac{P(X_{t-1}, Y_{1:t-1})P(X_t|X_{t-1})}{\sum_{X_{t-1}\in S_X} P(X_{t-1}, Y_{1:t-1})P(X_t|X_{t-1})} \\
&= \frac{\alpha(X_{t-1})P(X_t|X_{t-1})}{\sum_{X_{t-1}\in S_X} \alpha(X_{t-1})P(X_t|X_{t-1})}
\end{aligned}$$

Now let H be a general functional of the type $H_t(X_{1:t}) = \sum_{s=1}^{t} h_s(X_s)$. In order to compute $E\left(H_T|\,Y_{1:T}\right)$ a recursion is defined on $E\left(H_t|\,X_t, Y_{1:t}\right)$. Having computed $E\left(H_T|\,X_T, Y_{1:T}\right)$ one can obtain $E\left(H_T|\,Y_{1:T}\right)$ by integrating against

$\gamma(X_T) = P(X_T|Y)$. The calculation of $E\left(H_t|\, X_t, Y_{1:t}\right)$ proceeds as follows:

$$
\begin{aligned}
E\left(H_t|\, X_t, Y_{1:t}\right) &= \sum_{X_{t-1}\in S_X} E\left(H_t|\, X_t, X_{t-1}, Y_{1:t}\right) P(X_{t-1}|X_t, Y_{1:t}) \\
&= \sum_{X_{t-1}\in S_X} E\left(H_{t-1}|\, X_{t-1}, Y_{1:t-1}\right) \beta(X_{t-1}) + h_t(X_t)
\end{aligned}
$$

We illustrate the use of the above technique by computing the sufficient statistics for π, A and C, respectively: $P(X_1|Y_{1:T})$, $E\left(\sum_{s=2}^T X_{s-1}X_s\,|, Y_{1:T}\right)$ and $E\left(\sum_{s=1}^T X_sY_s\,|, Y_{1:T}\right)$.

To compute the sufficient statistic for π take $h_1(X_1) = X_1$ and $h_t = 0$ for $t > 1$:

$$
\begin{aligned}
P(X_1|X_t, Y_{1:t}) &= \sum_{X_{t-1}\in S_X} P(X_1|X_t, X_{t-1}, Y_{1:t})P(X_{t-1}|X_t, Y_{1:t}) \\
&= \sum_{X_{t-1}\in S_X} P(X_1|X_{t-1}, Y_{1:t-1})\beta(X_{t-1})
\end{aligned}
$$

Similarly taking $h_t(X_t) = X_{t-1}X_t$ and $h_1 = 0$ we compute the sufficient statistic for A:

$$
\begin{aligned}
E\left(\sum_{s=2}^{t} X_{s-1}X_s\,|X_t, Y_{1:t}\right) &= \sum_{X_{t-1}\in S_X} E\left(\sum_{s=2}^{t} X_{s-1}X_s\,|X_t, X_{t-1}, Y_{1:t}\right) P(X_{t-1}|X_t, Y_{1:t}) \\
&= \sum_{X_{t-1}\in S_X} E\left(\sum_{s=2}^{t-1} X_{s-1}X_s\,|X_{t-1}, Y_{1:t-1}\right) \beta(X_{t-1}) + X_t
\end{aligned}
$$

Finally the sufficient statistics for C requires taking $h_t(X_t) = Y_tX_t$:

$$
\begin{aligned}
E\left(\sum_{s=1}^{t} X_sY_s\,|X_t, Y_{1:t}\right) &= \sum_{X_{t-1}\in S_X} E\left(\sum_{s=1}^{t} X_sY_s\,|X_t, X_{t-1}, Y_{1:t}\right) P(X_{t-1}|X_t, Y_{1:t}) \\
&= \sum_{X_{t-1}\in S_X} E\left(\sum_{s=1}^{t-1} X_sY_s\,|X_{t-1}, Y_{1:t-1}\right) \beta(X_{t-1}) + X_tY_t
\end{aligned}
$$

2.3.5 Missing values

Missing values can be directly integrated into the estimation procedure. From the description in section 2 of the forward-backward recursions one can see that there is only one step in which actual data is used. This step is the recursion that calculates $\alpha(X_t)$, the forward recursion. The quantity $\alpha(X_t)$ can be interpreted as the (unnormalized) conditional distribution $P(X_t|Y_{1:t})$ and each step of the recursion can be loosly seen as calculation the expected value of X in the next period by applying the transition matrix A and also by using new information that has become available at time $t+1$ that is Y_{t+1}. If Y_{t+1} is missing this means that there is no new information to update $\alpha(X_t)$ to $\alpha(X_{t+1})$ so the only change in α should be the one that reflects the fact the one period has passed. This is captured by the diagonal matrix $\mathrm{diag}(A \cdot \alpha_t)$ in the recursion $\alpha_{t+1} = \mathrm{diag}(A \cdot \alpha_t)C'Y_{t+1}$. This result can be readily obtained by integrating Y_{t+1} out of expression 7. The computation of the sufficient statistics with missing Y is also straightforward since the quantity $E\left(Y^j_{t,n}X^i_{t,n}\,|i, Y_{1:s}\right)$ equivalent to $E(Y^j_{t,n}|X^i_{t,n})E\left(X^i_{t,n}\,|a, Y_{1:n}\right)$d the first term of this product corresponds to a column of the appropriate C matrix.

2.4 Statistical properties

Consistency of MLE of HMM parameters has been studied by Leroux (1992)and Leroux & Puterman (1992) and the conditions under which asymptotic normality holds have been established in Bickel & Ritov (1996) and Bickel, Ritov & Rydén (1998). The asymptotic theory in the work just referenced deals with the case where $T \to \infty$. In the present application we are interested in the case where $n \to \infty$. In the case where the dimension of the latent states is assumed to be known, standard asymptotic theory applies. If on the other hand we allow the size of X to grow with sample size then our model is in fact a MLE sieve whose properties can be studied using the tools of empirical processes theory. Wong & Shen (1995) and more recently Genovese & Wasserman (2000) have established rates of convergence for MLE sieves and Gaussian mixture sieves respectively using entropy methods. These same methods can be applied to the present case. We leave this for future work.

3 Modelling health and wealth evolution using HMM

We wish to study the dynamics of health and wealth among the oldest old (70 years or older).We will do so by modelling the joint dynamics of health status as well as the that of socio-economic status (SES), taking into account how the dynamics of the former affect the latter and vice-versa. The purpose of building such a model is that one can then use it to forecast the evolution of a population and to study the effects of a policy intervention on the dynamics of health and wealth. To be suitable for these tasks the model should pass some tests that verify if certain causal relationships are present and that the model is valid in different time periods. We will return to these points below when we discuss forecasting and testing within our framework.

We think of an individual at each point in time as being characterized by an health state and a SES state. Each possible health state is characterized by the presence or absence of certain health conditions. Over time, individuals transition between different health states and eventually fall into a terminal health state which we identify as death. Similarly a SES state is characterized, for example, by the presence of certain assets in the individual's portfolio and their value. Transition between SES states also occurs during the lifetime of the individual and one would expect, according to the life cycle hypothesis, that at older ages a substantial decrease in wealth is observed which would be represented by transitions from states associated with higher wealth to states associated with lower wealth. Interplay between health and SES dynamics occurs in the sense that being in a certain health state has a certain effect on the transition between SES states and likewise each SES state the individual might be, affects in a different way the transition between health states. The existence of these cross-effects is testable within the framework developed here.

The *true* states are not observed. We observe instead the elements that characterize these states, that is the health conditions of an individual and his portfolio composition and other SES variables. We interpret health and SES variables as providing noisy information about the underlying health and SES state.

The modelling approach just described can be implemented as an HMM. It is important to note that this approach is not as restrictive as it might seem since, as Künsch, Geman & Kehagias (1995) have shown, any discrete stationary process can be approximated, in distribution, as close as desired by an HMM.[6]

The AHEAD panel survey used for this study will be described in some detail in section 4. For the moment we will think of SES as being characterized by a set of categorical variables (wealth quartiles, income quartiles, ownership of certain types of assets, etc) and health as characterized by the presence of certain diseases or conditions (cancer, stroke, diabetes, high blood pressure, etc).

Missing data is a pervasive problem, specially in questions pertaining to wealth and income, so imputation plays an important role in the construction of wealth and income variables when analyzing AHEAD data. Within the HMM framework neither imputation nor aggregation of categorical wealth and income variables are necessary. The method can deal with these problems directly. We will comment on this aspect in more detail below.

The construction of a model that analyzes the evolution of health and SES has been done in McFadden, Hurd, Adams, Merrill & Ribeiro (2002). In that work each health condition is modelled directly as a function of exogenous variables and previous health conditions. The cost of modelling in this way the evolution of a high dimensional vector is that certain restrictions on the correlation structure of health

[6] Their result is actually more general since it also applies to functions of discrete Markov random fields. In both cases the approximation is obtained by increasing the number of hidden states

conditions have to be imposed and consequently the multiple risk structure of health conditions and its implications for duration pattern of certain health states may not be correctly accounted for. However, in that study, tests for the off diagonal terms of the covariance matrix of the latent variables that define each health condition are not statistically different from zero. The present approach is more generous in accounting for the multiple risk structure of incidence of health conditions at the cost of some interpretability of the underlying health states. It also has the potential of providing a more flexible characterization of the dynamics of wealth, and more importantly it provides an integrated approach to handling missing data which was previously dealt with with imputation.

In what follows we formalize the model just discussed and comment on some of its aspects in more detail.

3.1 A probabilistic model

We assume the existence of two discrete latent variables X_t^W and X_t^H that describe the "state of the world" for SES and health respectively. Let $X_t^W \in S_{X^W} = \{e_1, \ldots, e_{K_W}\}$ and $X_t^H \in S_{X^H} = \{e_1, \ldots, e_{K_H}\}$, that is there are K_W possible states that characterize SES and K_H possible states that characterize health. There is no particular ordering to these states, but we will provide an interpretation below. We assume that the processes evolve over time as first order Markov chains and further that, conditional on the past, X_t^W and X_t^H are independent. Bellow we provide ways to test for these hypothesis. In essence we have:

$$\begin{aligned} P(X_t^W, X_t^H | X_{1:t-1}^W, X_{1:t-1}^H) &= P(X_t^W, X_t^H | X_{t-1}^W, X_{t-1}^H) \\ &= P(X_t^W | X_{t-1}^W, X_{t-1}^H) P(X_t^H | X_{t-1}^W, X_{t-1}^H) \end{aligned} \tag{8}$$

for all t. These state variables in turn generate the observations. For example, an individual in a given state of health ($X^H = e_i$, for some i) is more likely to have certain health conditions than an individual in another state. A similar interpretation applies to X^W. We observe D_H variables that characterize health status denoted by $Y^{H_1}, \ldots, Y^{H_{D_H}}$ and D_W variables that characterize SES which we denote by $Y^{W_1}, \ldots, Y^{W_{D_W}}$. Since neither X_t^W nor X_t^H are observed we think of the observations as providing noisy information about the state in which the individual might be. We further assume that conditional on the current underlying state X the current observations are mutually independent and independent over time. Formally, this assumption is stated as:

$$\begin{aligned} P(Y_t^W, Y_t^H | X_{1:T}^W, X_{1:T}^H, Y_{1:T\setminus t}^W, Y_{1:T\setminus t}^H) &= P(Y_t^W, Y_t^H | X_t^W, X_t^H) \\ &= P(Y_t^W | X_t^W) P(Y_t^H | X_t^H) \\ &= \prod_{i=1}^{D_W} P(Y_t^{W,i} | X_t^W) \prod_{i=1}^{D_H} P(Y_t^{H,i} | X_t^H) \end{aligned} \tag{9}$$

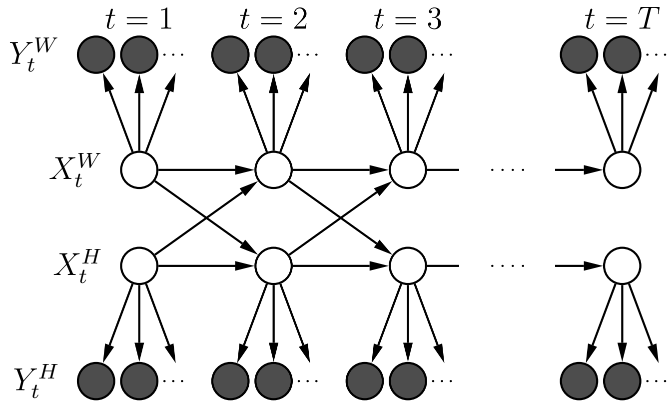


Figure 2: **A HMM model for health and wealth transitions**

In the notation introduced in section 2 (also see figure 1) the assumption expressed in (8) is simply a restriction on the elements of the matrix A. The restriction in (9) states that there is a matrix C for each of the observed Ys. All these assumptions are captured in the representation of this model in figure 2.

3.1.1 Interpretation

We cannot a priori give an interpretation to the various latent states (values of X). We can nevertheless, after estimating our model, give an interpretation analogous to the interpretation given in factor models. In factor or principal component analysis, factors are interpreted by inspecting their correlation with the observed variables. In this case we will estimate $P(Y_t|X_t)$ for all observed variables Y and for all possible states of X - in our notation these are the C matrices. An individual in a given state (say with $X_t^H = e_i$) will be more likely to have certain health conditions than an individual in a different state (say with $X_t^H = e_j$), so an interpretation of states by associating them with the set of conditions they will most likely give rise to is possible. Restrictions can also be imposed on the matrix A that will imply a certain interpretation for the latent states. For example if the matrix A corresponding to the wealth transitions is set to be lower triangular then each state will represent a lower level of wealth. The dynamics of this system will imply that households will ultimately converge to one state (the low wealth state) and the transitions can be interpreted as running down one's wealth which occurs at older ages.

3.1.2 Identification

This will obviously raise the question of identification of the parameters of our models, and whether there is an alternative structure which will equally fit the data. Related literature on identification in HMM models has tackled this problem. Ito, Amari & Kobayashi (1992) give necessary and sufficient conditions for identification in the case where C defines a deterministic map. At this point we note that our formulation is equivalent to that case. For the purpose of studying identification, the model presented can then be translated into the case where C defines a deterministic map. We do this by redefining the hidden state as (X, Y) redefining C to be the projection $f(X, Y) = Y$ (see for example Baum & Petrie (1966)). More recently Rydén (1996) extends the identification results to continuous time and Larget (1998) provides a canonical representation for HMM.

3.2 Specification Tests

One straightforward way to test the restrictions we have imposed is to use LM-type tests. Despite the fact that these tests don't require re-estimating the parameters they do require evaluating the gradient of the alternative hypothesis with the existing parameters. The new gradient in turn requires computing new sufficient statistics under the alternative which can by computationally demanding in particular for the statistics involving the latent variables. Standard asymptotic theory of LM tests would apply in these cases (Newey & McFadden 1994). We propose tests based on some moment conditions implied by the assumptions we have imposed. These tests have the advantage that they can be computed with the output available from the maximization routines and can, for the most part, be interpreted as conditional tests for independence in contingency tables.

3.2.1 Testing independence of Y's given X

Within a given time period our assumptions imply:

$$\begin{aligned} & E(Y_t^1 - C_1X_t)(Y_t^2 - C_2X_t)' = 0 \\ \Leftrightarrow \quad & E(Y_t^1Y_t^{2'} - C_1X_tY_t^{2'} - Y_t^1X_t'C_2' + C_1X_tX_t'C_2') = 0 \\ \Leftrightarrow \quad & E(Y_t^1Y_t^{2'} - C_1X_tX_t'C_2') = 0 \end{aligned}$$

Note that the zeros in the equality refer to matrices of zeros. The two last expression have empirical counterparts which can be computed. First we condition on the data to obtain and estimate of the latent variables for each observation. For the integration over Y we take the empirical version, that is we want to test if the following means are jointly zero:

$$\frac{1}{n}\sum_{i=1}^{n} Y_{it}^1Y_{it}^{2'} - \hat{C}_1E(X_{it}X_{it}'|Y_{1:T})\hat{C}_2'$$

The first term in this sum is just the contingency table between variables Y^1 and Y^2 at period t. If the dimension of X collapses to 1 (i.e. there is no latent variable) the above test reduces to testing if the cell probabilities are equal to the product of the marginal probabilities. Note also that the quantities needed to compute the above statistic are already available from the optimization routine described earlier.

Our assumptions also imply that, conditional on X, a given Y is independent over time. The procedure just described is also applicable in this situation replacing Y_t^2 with Y_{t+1}^1, C_2 with C_1 and the second X_t with X_{t+1}.

3.2.2 Testing the first order Markov assumption

One can also test for the validity of the Markov assumption using the principle just described. The following moment is implied by that assumption:

$$\begin{aligned} & E(X_{t+1} - AX_t)(X_t - AX_{t-1})' = 0 \\ \Leftrightarrow \quad & E(X_{t+1}X_t' - AX_tX_t' - X_{t+1}X_{t-1}'A' + AX_tX_{t-1}'A') = 0 \end{aligned}$$

With the exception of the term $X_{t+1}X_{t-1}'$, the empirical counterparts of the terms in the last expression can be computed from the output of the algorithms described in section 2. Again the procedure amounts to testing if the following sample means are jointly zero:

$$\frac{1}{n}\sum_{i=1}^{n} \quad \begin{aligned} & [E\left(X_{it+1}X_{it}' \left| Y_{1:T}\right.\right) - \hat{A}E\left(X_{it}X_{it}' \left| Y_{1:T}\right.\right) - \\ & E\left(X_{it+1}X_{it-1}' \left| Y_{1:T}\right.\right)A' + AE\left(X_{it}X_{it-1}' \left| Y_{1:T}\right.\right)A'] \end{aligned}$$

3.2.3 Causality

Causality tests can be interpreted as a further restriction to the ones in expression (8). Lack of causality from health to wealth, for example, can be formulated as $P(X_t^W|X_{t-1}^W, X_{t-1}^H) = P(X_t^W|X_{t-1}^W)$. In terms of the graphical representation of figure 2, lack of causality is equivalent to removing the upward diagonal edges from the X^H to the X^W nodes. There is an important underlying assumption about what our model truly represents in this definition of causality tests. Usually causality tests are stated in terms of conditional independence of certain observed variables. This is in particular the case of McFadden et al. (2002). If we interpret our model as a structural model, that is, if we assume that the true underlying structure is one where there are underlying states that give rise to observations that merely give indications as to which state we might be in, then this definition of causality is the relevant one. However if the model just described is seen as just a parsimonious way to describe the high dimensional joint distribution of Y, then the test one should be doing is $P(Y_t^W|Y_{t-1}^W, Y_{t-1}^H) = P(Y_t^W|Y_{t-1}^W)$. Unfortunately the fact that we might have $P(X_t^W|X_{t-1}^W, X_{t-1}^H) = P(X_t^W|X_{t-1}^W)$ does not imply that we will also have $P(Y_t^W|Y_{t-1}^W, Y_{t-1}^H) = P(Y_t^W|Y_{t-1}^W)$ except in the particular case where there are only 2 periods (i.e. $T = 2$).

3.3 Estimation

Estimation is similar to the procedures described in section 2. The few differences from that setting are that now we observe multiple variables at each moment in time and that we would like to include exogenous covariates in the specification of the transition matrix A. More specifically we expect that factors like age might affect the transition of some states to others. Also the time between interviews should reasonably affect the transition matrix. For each observation, these changes leave the calculation of the posterior expectations defined in expressions (3), (4) and (5) unchanged. Since each column of A is, by definition, a probability function, the maximization with exogenous variables can simply be parameterized as a set of multinomial logits (one for each column) with appropriately calculated weights and dependent variables.Moreover, there is nothing in the formulation of the problem that prevents us from using directly disaggregated wealth and income information coded as categorical variables.

It should also be noted that the easy incorporation of missing values can be used as a modelling tool to deal with the varying inter-wave periods. One could assume that the transition matrix A referred to, say, a three month transition period. Since data is collected every 24-36 months the time periods in between would correspond to missing data for all variables.

3.4 Forecasting, policy analysis and further studies

To forecast the evolution of a population over time using an estimated HMM model one starts with a sample of the population and uses each observation to make draws from the posterior distribution of X given Y. Thus for each individual one makes multiple imputations of his or her possible underlying state. Once simulated values of X are obtained, the dynamics of X are fully described by the model as well as the transition from forecasted values of X back to predicted Y. This can be seen as more than an instrument of prediction. Since subsequent waves of AHEAD will become available in the near future, a test of the model should be how well it can predict the observed outcomes in new data. This of course can be done with the current data by estimating the model on a subset of the waves and predicting the outcomes of the remaining waves.

Careful modelling of the transition matrix A with exogenous variables that also constitute variables suitable of policy manipulation will make the model an instrument of policy analysis since now different forecasts can be produced under different policy scenarios. For example making the transition matrix of wealth states depend of the existence to prescription drug financing would help predict the effect of different Medicare polices with respect to drug financing. Also, the effect new discovery can be analyzed. For example a cure for a certain health condition would mean that

that condition would no longer be present in the starting population. In turn, in the forecasting procedure described above, states that gave rise to that condition should be under sampled and the subsequent evolution of the system would reflect that fact.

One event we will analyze in the future is the effect of eligibility to Medicare on health and wealth outcomes especially of those individuals who were uninsured previously. At the present time a significant part of an earlier cohort (born in 1931-1941) which was being interviewed as part of the HRS panel survey is now meeting eligibility criteria Medicare. The methodology proposed in this paper will be used to investigate if eligibility to Medicare creates a structural change in the evolution of health and/or SES and if so what is its impact.

4 The AHEAD Panel Data

4.1 Sample Characteristics

Our data come from the Asset Dynamics among the Oldest-Old (AHEAD) study[7]. This is a panel of individuals born in 1923 or earlier and their spouses. At baseline in 1993 the AHEAD panel contained 8222 individuals representative of the non-institutionalized population, except for over-samples of blacks, Hispanics and Floridians. Of these subjects 7638 were over the age of 69 and the remainder were younger spouses. There were 6052 households, including individuals living alone or with others, in the sample. The Wave 1 surveys took place between October 1993 and August 1994 with half the total completed interviews finished before December 1993. The Wave 2 surveys took place approximately 24 months later, between November 1995 and June 1996, with half the total completed interviews finished before February 1996. The Wave 3 surveys took place approximately 27 months after Wave 2, between January and December 1998, with half the total completed interviews finished before March 1998. In each wave there was a long but thin tail of interviews heavily populated by subjects who had moved, or required proxy interviews due to death or institutionalization. AHEAD is a continuing panel but it has now been absorbed into the larger Health and Retirement Survey (HRS) which is being conducted on a three year cycle.

4.2 Wealth data

AHEAD individuals and couples are asked for a complete inventory of assets and debts. Subjects are asked first if they have any assets in a specified category, and if

[7]The AHEAD survey is conducted by the University of Michigan Survey Research Center for the National Institute on Aging (see Soldo, Hurd, Rodgers & Wallace (1997))

so, they are asked for the amount. A non-response is followed by unfolding bracket questions to bound the quantity in question and this may result in complete or incomplete bracket responses. We consider 10 asset categories plus debt and mortgages. The ownership and missing data patterns are summarized in Tables 1 and 2. Non-response to the ownership question is small. In Wave 1 94% of the households had complete responses to all 12 categories considered. For waves 2 and 3 this value is 92% and 88% respectively. More than 75% of the households declared to own between 1 and 5 assets in all waves. For each asset considered individually the percentage of missing observations is between 1.2% and 3.8% for wave 1, between 0.8% and 4% for wave 2 and between 1% and 6.2% for wave 3.

The problem of non-response is more severe in amount questions. Tables 3-5 summarize the patterns of missing data for the amount variable for each asset. In each wave only 40% of the households had complete amount information for all assets considered. Roughly 28% had a non response in one asset only. The rate of non-response to the amount question is between 15% and 26% for the most commonly held assets (checking account, transportation, CD's stocks and housing). The use of unfolding brackets reduces complete non-response to less than 4% in most cases. With this rate of missing data, imputation usually plays an important role in the construction of wealth variables. The methods described here circumvent the need for imputation.

5 Illustrative results

For the purpose of illustration we estimate models for health and wealth evolution separately. We first estimate the evolution of each asset in isolation, controlling for transitions from two person household to single households. These results are presented in tables W.1.1-W.1.12. For reference table W.0 has the breakpoints of each asset category. Finally we estimate an HMM model of the evolution of health conditions and mortality (tables H.1-H.2). We comment on these models in turn.

When analyzing the evolution of each asset separately we have a different latent variable controlling the evolution of holdings of that asset as well as spouse mortality. For the latent variable we specified 8 categories. The first 4 we associate with single households and the last 4 with couples. We do this by restricting the appropriate matrix C. Since for some of the assets most households don't change their portfolio by significantly since they tend to concentrate on one category (no ownership) the latent states capture not only different holdings of those assets but also different spouse mortality across households. This is in particular the case of real estate, business, IRAs, Stocks, Bonds, CDs and debt. For checking accounts, transportation and housing assets the latent variable captures mostly different levels of asset holdings. In most assets a transition from states with higher holdings to states with lower holdings occurs and this rate of decrease is usually different in singles and couples. The death of a spouse is also associated with reductions in certain assets in particular real estate and mobile homes.

To summarize the information contained in the health variables we assume the existence of a single underlying health variable with 8 categories which generates the observed health conditions. By restriction of the parameters one of categories of the latent variable is identified with death. The HMM procedure groups individuals by the prevalence of certain health conditions with each group (category of the latent variable having its own mortality rate. For example groups 4 and 8 represent relatively healthy individuals at the beginning of the survey whose health deteriorates over time. For individuals in group 8 the deterioration in health can occur due to the appearance of cancer. Group 3 is highly associated with diabetes and high blood pressure, with high blood pressure, incontinence and cognitive impairment, group 6 with lung problems and group 7 with cancer. For the most part there is little evolution in health conditions except when an individual dies and as mentioned before when a respondent belongs to a group of relatively health people. We note the these results are just meant to be illustrative. The transition matrix has been restricted to be a bordered diagonal capturing the idea that health deteriorates by moving to more healthy states to states that are similar in the health conditions that are relevant but represent poorer health. Also no use of other exogenous covariates, such as age, race and education, has been made. These are likely to affect the transition from state to state and most certainly the mortality rate.

A finally note of the specification tests that were proposed is due. Tests of independence of Y in a given period conditional on X pass for the most part in the health conditions. Tests of independence of Y over time usually fail. The failure is more severe in the asset case. Given the simplified models used and the lack of use of other control variables this is not to surprising. There is still room to develop the models and to analyze the interaction between health and wealth. Had the tests passed then we would have captured all the asset and wealth evolution is separate and simple models so any hope of finding any interaction between health and wealth would be gone.

6 Final remarks

We have proposed and described an empirical model which can be used to study the joint evolution of a relatively large vector of discrete variables over time. The model draws on factor analytic and filtering literature and can in its most simple form be described as discrete Kalman filter. We illustrated with simple examples how the model operates. We have also proposed future work which will have has building blocks the models described here.

References

Baum, L. E. & Petrie, T. (1966), 'Statistical inference for probabilistic functions of finite state Markov chains', *Ann. Math. Statist.* **37**, 1554–1563.

Baum, L. E., Petrie, T., Soules, G. & Weiss, N. (1970), 'A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains', *Ann. Math. Statist.* **41**, 164–171.

Bickel, P. J. & Ritov, Y. (1996), 'Inference in hidden Markov models. I. Local asymptotic normality in the stationary case', *Bernoulli* **2**(3), 199–228.

Bickel, P. J., Ritov, Y. & Rydén, T. (1998), 'Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models', *Ann. Statist.* **26**(4), 1614–1635.

Borsch-Supan, A., McFadden, D. L. & Schnabel, R. (1996), Living arrangements: Health and wealth effects, *in* D. Wise, ed., 'Advances in the Economics of Aging', The University of Chicago Press, chapter 6, pp. 193–218.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977), 'Maximum likelihood from incomplete data via the EM algorithm', *J. Roy. Statist. Soc. Ser. B* **39**(1), 1–38. With discussion.

Elliott, R. J., Aggoun, L. & Moore, J. B. (1995), *Hidden Markov models*, Springer-Verlag, New York. Estimation and control.

Fredkin, D. R. & Rice, J. A. (1986), 'On aggregated Markov processes', *J. Appl. Probab.* **23**(1), 208–214.

Genovese, C. R. & Wasserman, L. (2000), 'Rates of convergence for the Gaussian mixture sieve', *Ann. Statist.* **28**(4), 1105–1127.

Ito, H., Amari, S.-i. & Kobayashi, K. (1992), 'Identifiability of hidden Markov information sources and their minimum degrees of freedom', *IEEE Trans. Inform. Theory* **38**(2, part 1), 324–333.

Joreskog, K. G. (1973), A general method for estimating a linear structural equation system, *in* A. S. Goldberger & O. D. Duncan, eds, 'Structural Equation Models in the Social Sciences', Seminar Press, Inc, chapter 5, pp. 85–112.

Künsch, H., Geman, S. & Kehagias, A. (1995), 'Hidden Markov random fields', *Ann. Appl. Probab.* **5**(3), 577–602.

Larget, B. (1998), 'A canonical representation for aggregated Markov processes', *J. Appl. Probab.* **35**(2), 313–324.

Leroux, B. G. (1992), 'Maximum-likelihood estimation for hidden Markov models', *Stochastic Process. Appl.* **40**(1), 127–143.

Leroux, B. G. & Puterman, M. L. (1992), 'Maximum penalized likelihood estimation for independent and Markov-dependent mixture models', *Biometrics* **98**, 545–558.

McFadden, D., Hurd, M., Adams, P., Merrill, A. & Ribeiro, T. (2002), 'Health wealth and wise? tests for direct causal paths between health and socioeconomic status', *Forthcoming Journal of Econometrics*.

McLachlan, G. J. & Krishnan, T. (1997), *The EM algorithm and extensions*, John Wiley & Sons Inc., New York. A Wiley-Interscience Publication.

Newey, W. K. & McFadden, D. (1994), Large sample estimation and hypothesis testing, *in* 'Handbook of econometrics, Vol. IV', Vol. 2 of *Handbooks in Econom.*, North-Holland, Amsterdam, pp. 2111–2245.

Pearl, J. (2000), *Causality*, Cambridge University Press, Cambridge. Models, reasoning, and inference.

Rabiner, L. R. (1989), 'A tutorial on hidden markov models and selected applications in speech recognition', *Proceedings of the IEEE* **77**(2), 257–286.

Ruud, P. A. (1991), 'Extensions of estimation methods using the EM algorithm', *Journal of Econometrics* **49**, 305–341.

Rydén, T. (1996), 'On identifiability and order of continuous-time aggregated Markov chains, Markov-modulated Poisson processes, and phase-type distributions', *J. Appl. Probab.* **33**(3), 640–653.

Soldo, B., Hurd, M., Rodgers, W. & Wallace, R. (1997), 'Asset and health dynamics among the oldest old: An overview of the ahead study', *The Journal of Gerontology, Series B* (52B), 1–20.

Wong, W. H. & Shen, X. (1995), 'Probability inequalities for likelihood ratios and convergence rates of sieve MLEs', *Ann. Statist.* **23**(2), 339–362.

A Tables

Table 1: **Ownership non-response and asset ownership**

Asset	Missing			Owned		
	Wave 1	Wave 2	Wave 3	Wave 1	Wave 2	Wave 3
Checking account	3.4	2.6	2.9	73.2	81.4	80.2
Transportation	1.2	0.8	1.0	69.6	64.8	61.3
CDs/Sav. Bonds/T. Bills	3.8	3.4	4.2	20.5	29.7	28.8
IRA	2.5	2.0	2.6	14.2	16.6	16.9
Bonds	3.3	2.4	3.4	5.3	8.5	7.7
Stocks	3.1	2.6	3.2	18.2	27.6	27.3
House	1.2	3.3	5.9	70.6	71.3	71.1
Real Estate	2.3	1.4	1.4	17.7	14.6	11.4
Business	1.6	1.1	1.3	3.9	5.8	5.2
Other assets	2.9	2.1	2.9	9.4	8.2	8.7
Mortgage	1.5	4.0	6.2	11.2	8.5	9.1
Debt	2.5	1.6	2.1	13.9	13.7	12.1

Table 2: **Number of ownership non-responses and number of assets**

	Missing			Owned		
# of assets	Wave 1	Wave 2	Wave 3	Wave 1	Wave 2	Wave 3
0	93.5	91.8	88.2	7.3	5.8	6.5
1	2.1	2.1	2.8	12.2	11.3	11.5
2	1.1	3.5	5.7	16.9	14.0	14.4
3	0.4	0.4	0.6	20.1	18.3	19.3
4	0.3	0.4	0.5	18.7	19.0	18.3
5	0.4	0.2	0.4	12.1	14.3	14.0
6	0.2	0.3	0.3	6.9	9.8	9.3
7	0.2	0.1	0.3	3.5	4.8	4.1
8	0.3	0.1	0.2	1.4	1.9	2.2
9	0.1	0.1	0.3	0.7	0.5	0.4
10	0.6	0.2	0.3	0.2	0.2	0.1
11	0.0	0.4	0.0	0.0	0.0	0.0
12	0.7	0.3	0.6	0.0	0.0	0.0

Table 3: **Breakdown of amount non-response - Wave 1**

	Wave 1			
Asset	Regular B.	Irreg. B.	Non-resp.	Total
Checking account	16.3	6.8	3.4	26.6
Transportation	14.6	2.4	1.2	18.2
CDs/Sav. Bonds/T. Bills	4.7	2.7	3.8	11.3
IRA	2.5	1.1	2.5	6.2
Bonds	1.5	0.7	3.3	5.5
Stocks	6.1	2.1	3.1	11.3
House	18.2	3.1	1.2	22.6
Real Estate	4.6	1.2	2.3	8.2
Business	2.1	0.3	1.5	3.9
Other assets	2.3	0.6	2.9	5.8
Debt	1.7	0.3	2.5	4.4

Table 4: **Breakdown of amount non-response - Wave 2**

	Wave 2			
Asset	Regular B.	Irreg. B.	Non-resp.	Total
Checking account	20.8	7.4	2.6	30.8
Transportation	11.8	10.6	0.8	23.2
CDs/Sav. Bonds/T. Bills	6.6	3.7	3.4	13.7
IRA	1.4	3.9	2.0	7.3
Bonds	2.2	1.4	2.4	6.1
Stocks	9.6	3.2	2.5	15.4
House	17.3	3.3	3.3	23.9
Real Estate	3.9	0.8	1.4	6.1
Business	1.9	0.9	1.1	3.9
Other assets	1.8	0.6	2.1	4.5
Debt	2.1	0.4	1.6	4.1

Table 5: **Breakdown of amount non-response - Wave 3**

	Wave 3			
Asset	Regular B.	Irreg. B.	Non-resp.	Total
Checking account	13.9	9.2	2.9	26.1
Transportation	13.1	2.5	1.0	16.6
CDs/Sav. Bonds/T. Bills	4.5	4.5	4.2	13.2
IRA	1.8	3.6	2.6	8.0
Bonds	1.6	1.3	3.4	6.3
Stocks	6.9	4.8	3.2	14.9
House	8.0	8.4	5.9	22.3
Real Estate	2.6	1.0	1.4	5.0
Business	1.3	0.9	1.3	3.4
Other assets	1.6	0.8	2.9	5.3
Debt	1.4	0.5	2.1	4.0

Table 6: **Number of assets with amount non-response**

	% of Households		
# of assets	Wave 1	Wave 2	Wave 3
0	41.77	34.77	41.00
1	27.99	28.90	28.32
2	15.28	17.22	14.45
3	7.25	8.70	7.60
4	3.60	5.39	4.17
5	1.55	2.39	2.07
6	0.85	1.05	1.04
7	0.41	0.45	0.31
8	0.36	0.23	0.39
9	0.18	0.23	0.29
10	0.41	0.21	0.27
11	0.36	0.47	0.10

TABLE W.0

Asset category lower breakpoints

	1	2	3	4	5	6
rlst	0	1	2.5K	125K	500K	1000K
bsns	0	1	5K	10K	100K	1000K
ira	0	1	10K	25K	100K	400K
stck	0	1	2.5K	25K	125K	400K
chck	0	1	5K	50K	150K	300K
bond	0	1	2.5K	10K	100K	400K
cd	0	1	2.5K	25K	125K	250K
trns	0	1	5K	25K	200K	
othr	0	1	5K	50K	100K	
debt	0	1	0.5K	5K	50K	
mobh	0	1	5K	10K	20K	100K
hous	0	1	15K	50K	150K	500K

TABLE W.1.1
ASSET: Real Estate

Conditional Distribution of **Y** given **X**

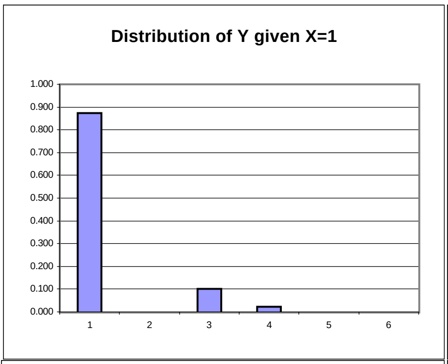
	X1	X2	X3	X4	X5	X6	X7	X8
Y1	0.875	0.976	0.995	0.174	0.875	0.976	0.995	0.174
Y2	0.001	0.004	0.003	0.035	0.001	0.004	0.003	0.035
Y3	0.102	0.017	0.002	0.475	0.102	0.017	0.002	0.475
Y4	0.021	0.001	0.000	0.228	0.021	0.001	0.000	0.228
Y5	0.001	0.001	0.000	0.057	0.001	0.001	0.000	0.057
Y6	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.030

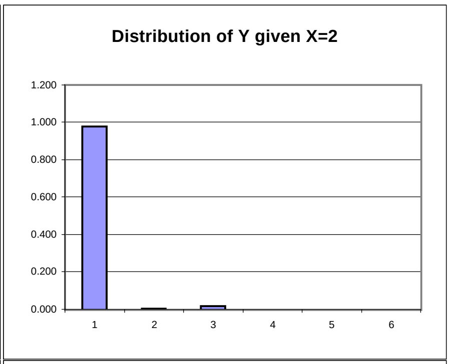
Transition matrix

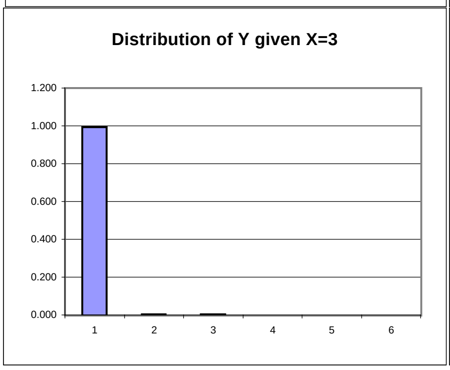
t+1\t	X1	X2	X3	X4	X5	X6	X7	X8
X1	0.620	0.003	0.000	0.095	0.000	0.228	0.000	0.008
X2	0.367	0.585	0.076	0.000	0.000	0.003	0.007	0.000
X3	0.000	0.411	0.924	0.002	0.000	0.319	0.001	0.021
X4	0.013	0.000	0.000	0.903	0.000	0.000	0.000	0.101
X5	0.000	0.000	0.000	0.000	0.074	0.110	0.000	0.061
X6	0.000	0.000	0.000	0.000	0.923	0.081	0.429	0.000
X7	0.000	0.000	0.000	0.000	0.000	0.259	0.550	0.011
X8	0.000	0.000	0.000	0.000	0.003	0.000	0.014	0.797

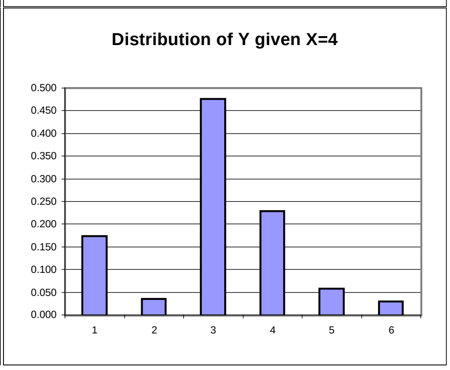
Initial distribution

P0	0.135	0.184	0.272	0.051	0.150	0.005	0.140	0.064









Notes: States 1-4 correspond to single member households and states 5-8 correspond to couples
Lower left block of transition matrix is restricted to be zero.
The conditional probability of Y given X=1,...,4 is retricted to be equal to the conditional distribution when X=5,...,8

TABLE W.1.2

ASSET: Business

Conditional Distribution of **Y** given **X**

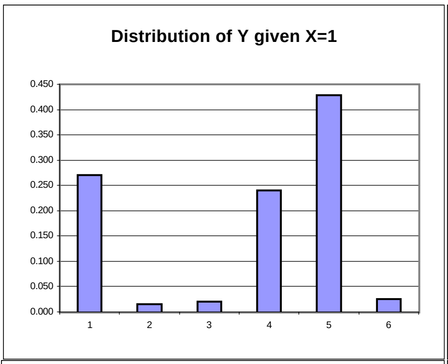
	X1	X2	X3	X4	X5	X6	X7	X8
Y1	0.271	0.988	0.995	0.994	0.271	0.988	0.995	0.994
Y2	0.015	0.001	0.001	0.001	0.015	0.001	0.001	0.001
Y3	0.020	0.000	0.000	0.000	0.020	0.000	0.000	0.000
Y4	0.241	0.007	0.004	0.001	0.241	0.007	0.004	0.001
Y5	0.428	0.004	0.000	0.004	0.428	0.004	0.000	0.004
Y6	0.025	0.001	0.000	0.001	0.025	0.001	0.000	0.001

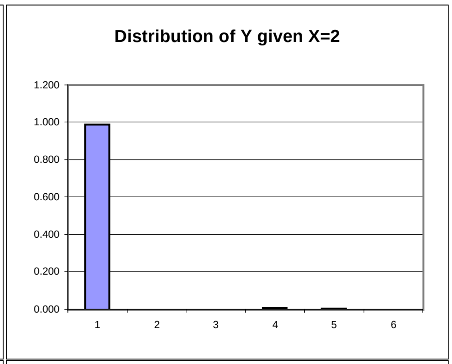
Transition matrix

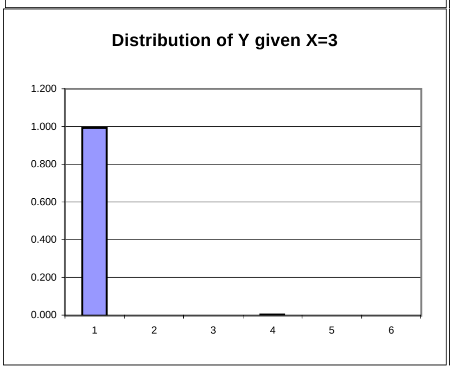
t+1\t	X1	X2	X3	X4	X5	X6	X7	X8
X1	0.849	0.019	0.000	0.001	0.084	0.012	0.000	0.000
X2	0.100	0.227	0.191	0.000	0.016	0.376	0.001	0.000
X3	0.000	0.754	0.277	0.419	0.000	0.376	0.001	0.144
X4	0.051	0.000	0.532	0.579	0.005	0.000	0.000	0.236
X5	0.000	0.000	0.000	0.000	0.724	0.000	0.000	0.001
X6	0.000	0.000	0.000	0.000	0.054	0.152	0.224	0.000
X7	0.000	0.000	0.000	0.000	0.000	0.084	0.266	0.261
X8	0.000	0.000	0.000	0.000	0.117	0.000	0.509	0.357

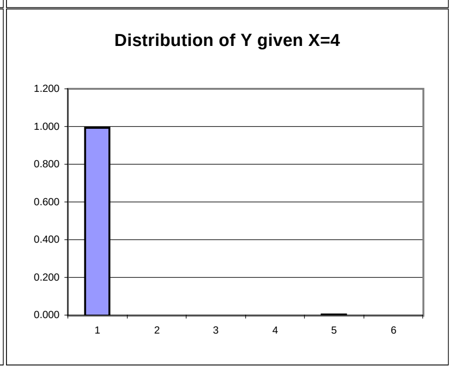
Initial distribution

P0	0.026	0.135	0.303	0.178	0.040	0.000	0.307	0.012









Notes: States 1-4 correspond to single member households and states 5-8 correspond to couples
Lower left block of transition matrix is restricted to be zero.
The conditional probability of Y given X=1,...,4 is retricted to be equal to the conditional distribution when X=5,...,8

TABLE W.1.3

ASSET: IRA

Conditional Distribution of **Y** given **X**

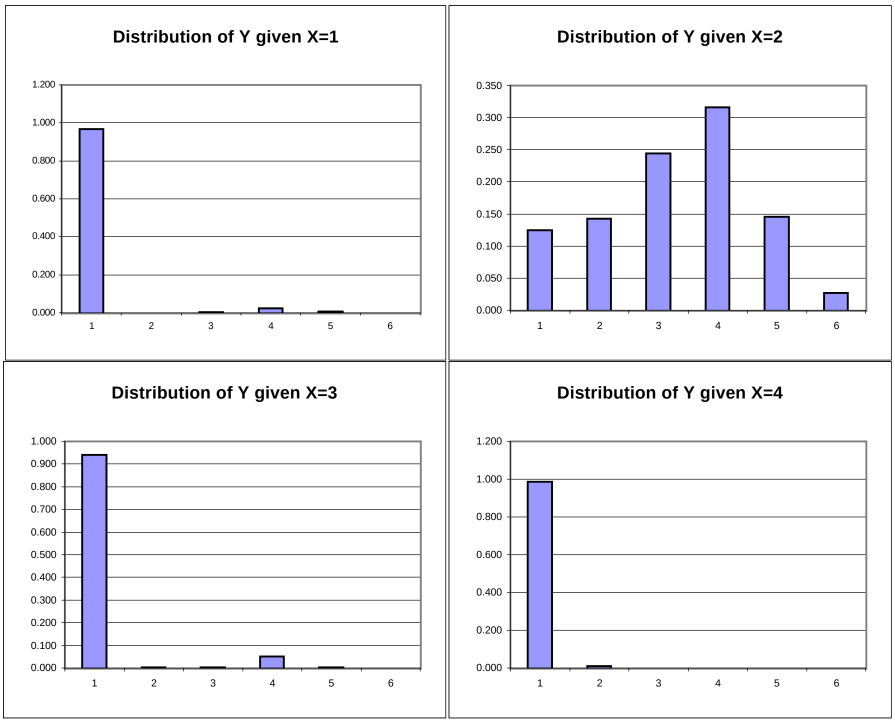
	X1	X2	X3	X4	X5	X6	X7	X8
Y1	0.966	0.124	0.941	0.988	0.966	0.124	0.941	0.988
Y2	0.000	0.143	0.002	0.009	0.000	0.143	0.002	0.009
Y3	0.004	0.244	0.002	0.001	0.004	0.244	0.002	0.001
Y4	0.022	0.317	0.051	0.002	0.022	0.317	0.051	0.002
Y5	0.008	0.146	0.004	0.000	0.008	0.146	0.004	0.000
Y6	0.000	0.027	0.000	0.000	0.000	0.027	0.000	0.000

Transition matrix

t+1\t	X1	X2	X3	X4	X5	X6	X7	X8
X1	0.445	0.040	0.000	0.325	0.121	0.002	0.000	0.001
X2	0.013	0.926	0.001	0.000	0.000	0.086	0.000	0.000
X3	0.000	0.034	0.695	0.278	0.000	0.005	0.271	0.009
X4	0.542	0.000	0.304	0.398	0.879	0.000	0.079	0.005
X5	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.306
X6	0.000	0.000	0.000	0.000	0.000	0.816	0.000	0.000
X7	0.000	0.000	0.000	0.000	0.000	0.079	0.534	0.297
X8	0.000	0.000	0.000	0.000	0.000	0.000	0.117	0.382

Initial distribution

P0	0.051	0.057	0.000	0.533	0.000	0.113	0.002	0.244



Notes: States 1-4 correspond to single member households and states 5-8 correspond to couples
Lower left block of transition matrix is restricted to be zero.
The conditional probability of Y given X=1,...,4 is retricted to be equal to the conditional distribution when X=5,...,8

TABLE W.1.4

ASSET: Cheking

Conditional Distribution of **Y** given **X**

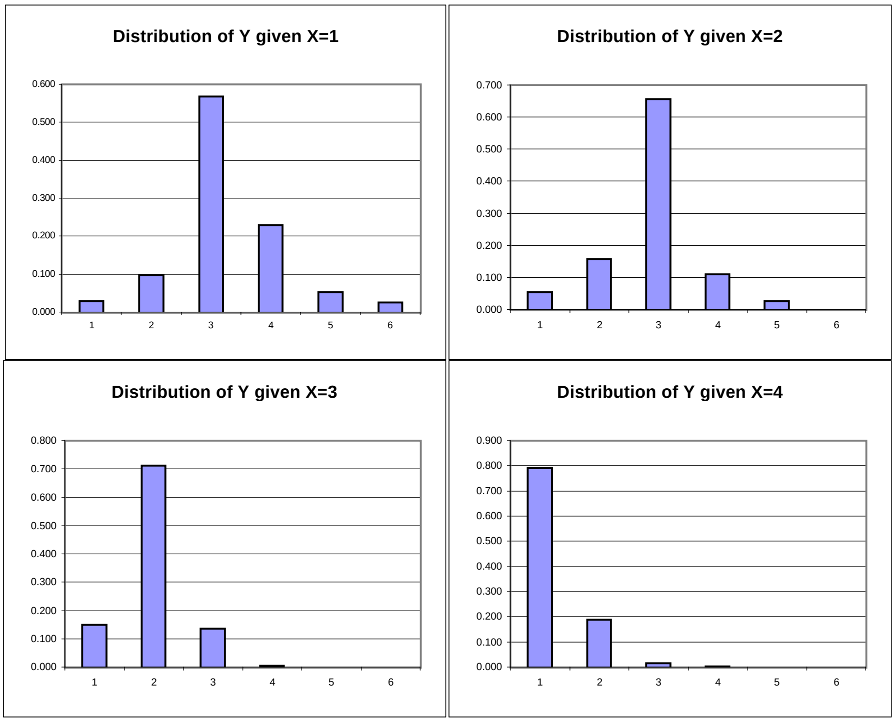
	X1	X2	X3	X4	X5	X6	X7	X8
Y1	0.028	0.053	0.149	0.791	0.028	0.053	0.149	0.791
Y2	0.098	0.157	0.712	0.189	0.098	0.157	0.712	0.189
Y3	0.568	0.656	0.135	0.015	0.568	0.656	0.135	0.015
Y4	0.230	0.109	0.003	0.003	0.230	0.109	0.003	0.003
Y5	0.052	0.025	0.000	0.000	0.052	0.025	0.000	0.000
Y6	0.025	0.000	0.000	0.001	0.025	0.000	0.000	0.001

Transition matrix

t+1\t	X1	X2	X3	X4	X5	X6	X7	X8
X1	0.954	0.000	0.000	0.000	0.005	0.390	0.000	0.000
X2	0.001	0.913	0.000	0.000	0.005	0.495	0.000	0.000
X3	0.000	0.087	1.000	0.000	0.000	0.115	0.111	0.022
X4	0.045	0.000	0.000	1.000	0.000	0.000	0.000	0.238
X5	0.000	0.000	0.000	0.000	0.646	0.000	0.000	0.000
X6	0.000	0.000	0.000	0.000	0.328	0.000	0.103	0.000
X7	0.000	0.000	0.000	0.000	0.000	0.000	0.755	0.000
X8	0.000	0.000	0.000	0.000	0.016	0.000	0.031	0.740

Initial distribution

P0	0.120	0.160	0.222	0.138	0.250	0.000	0.080	0.030



Notes: States 1-4 correspond to single member households and states 5-8 correspond to couples
Lower left block of transition matrix is restricted to be zero.
The conditional probability of Y given X=1,...,4 is retricted to be equal to the conditional distribution when X=5,...,8

<u>TABLE W.1.5</u>

ASSET: Stock

Conditional Distribution of **Y** given **X**

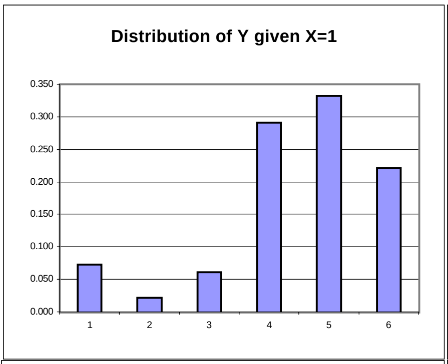
	X1	X2	X3	X4	X5	X6	X7	X8
Y1	0.073	0.181	0.976	0.975	0.073	0.181	0.976	0.975
Y2	0.022	0.095	0.002	0.011	0.022	0.095	0.002	0.011
Y3	0.061	0.372	0.010	0.008	0.061	0.372	0.010	0.008
Y4	0.291	0.318	0.012	0.001	0.291	0.318	0.012	0.001
Y5	0.333	0.030	0.000	0.006	0.333	0.030	0.000	0.006
Y6	0.221	0.004	0.000	0.000	0.221	0.004	0.000	0.000

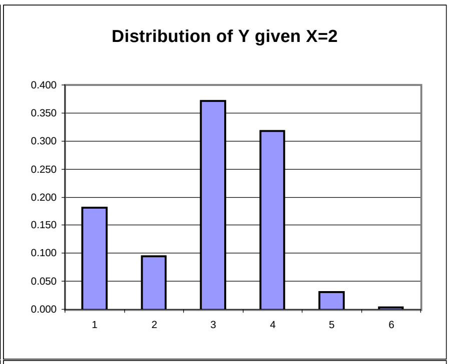
Transition matrix

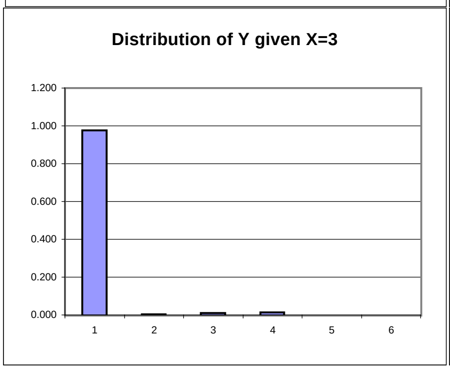
t+1\t	X1	X2	X3	X4	X5	X6	X7	X8
X1	0.951	0.001	0.000	0.018	0.158	0.001	0.000	0.000
X2	0.000	0.932	0.007	0.000	0.046	0.000	0.041	0.000
X3	0.000	0.067	0.308	0.593	0.000	0.050	0.516	0.014
X4	0.049	0.000	0.685	0.389	0.001	0.000	0.000	0.000
X5	0.000	0.000	0.000	0.000	0.794	0.233	0.000	0.000
X6	0.000	0.000	0.000	0.000	0.000	0.602	0.037	0.000
X7	0.000	0.000	0.000	0.000	0.000	0.115	0.406	0.697
X8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.290

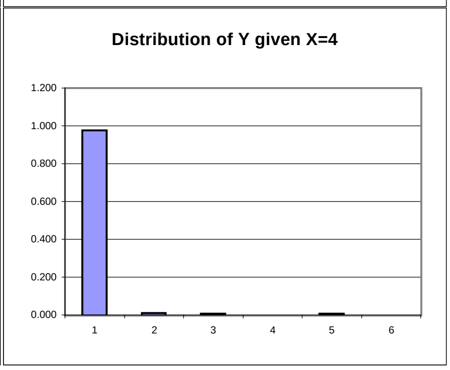
Initial distribution

P0	0.074	0.073	0.493	0.000	0.071	0.079	0.000	0.209









Notes: States 1-4 correspond to single member households and states 5-8 correspond to couples
Lower left block of transition matrix is restricted to be zero.
The conditional probability of Y given X=1,...,4 is retricted to be equal to the conditional distribution when X=5,...,8

TABLE W.1.6

ASSET: Bond

Conditional Distribution of **Y** given **X**

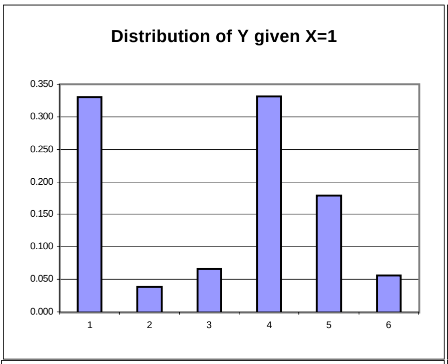
	X1	X2	X3	X4	X5	X6	X7	X8
Y1	0.330	0.985	0.988	0.692	0.330	0.985	0.988	0.692
Y2	0.038	0.001	0.004	0.003	0.038	0.001	0.004	0.003
Y3	0.066	0.008	0.002	0.013	0.066	0.008	0.002	0.013
Y4	0.331	0.006	0.006	0.152	0.331	0.006	0.006	0.152
Y5	0.179	0.000	0.000	0.098	0.179	0.000	0.000	0.098
Y6	0.056	0.000	0.000	0.042	0.056	0.000	0.000	0.042

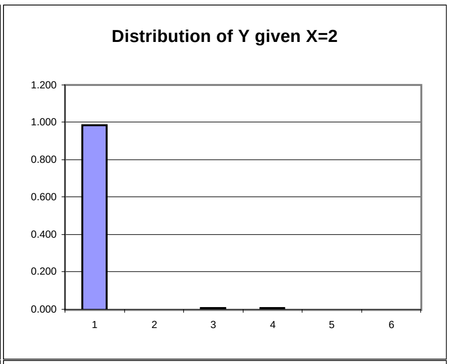
Transition matrix

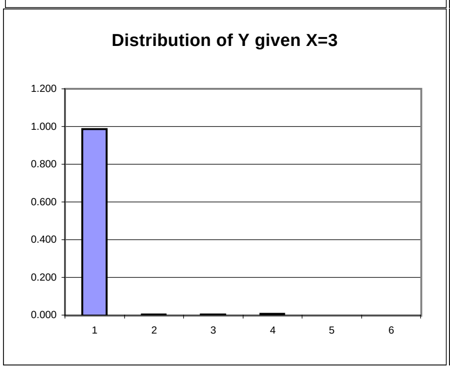
t+1\t	X1	X2	X3	X4	X5	X6	X7	X8
X1	0.802	0.003	0.000	0.304	0.002	0.001	0.000	0.122
X2	0.185	0.295	0.442	0.000	0.000	0.582	0.002	0.000
X3	0.000	0.703	0.528	0.024	0.000	0.047	0.016	0.060
X4	0.013	0.000	0.030	0.673	0.010	0.000	0.000	0.534
X5	0.000	0.000	0.000	0.000	0.621	0.045	0.000	0.000
X6	0.000	0.000	0.000	0.000	0.001	0.320	0.542	0.000
X7	0.000	0.000	0.000	0.000	0.000	0.004	0.425	0.261
X8	0.000	0.000	0.000	0.000	0.366	0.000	0.016	0.024

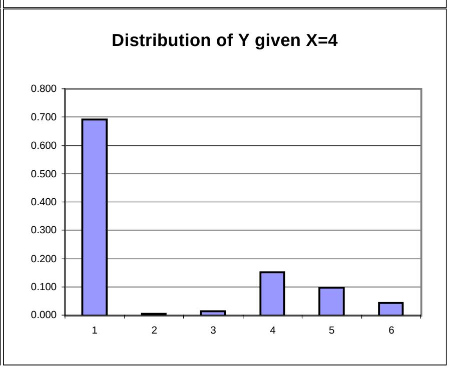
Initial distribution

P0	0.038	0.298	0.293	0.012	0.063	0.000	0.296	0.000









Notes: States 1-4 correspond to single member households and states 5-8 correspond to couples
Lower left block of transition matrix is restricted to be zero.
The conditional probability of Y given X=1,...,4 is retricted to be equal to the conditional distribution when X=5,...,8

TABLE W.1.7

ASSET: CD

Conditional Distribution of **Y** given **X**

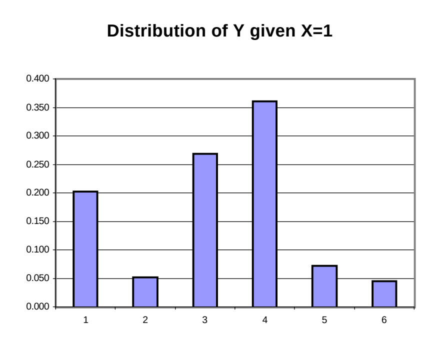
	X1	X2	X3	X4	X5	X6	X7	X8
Y1	0.202	0.979	0.911	0.185	0.202	0.979	0.911	0.185
Y2	0.052	0.001	0.018	0.064	0.052	0.001	0.018	0.064
Y3	0.268	0.015	0.051	0.287	0.268	0.015	0.051	0.287
Y4	0.360	0.005	0.020	0.352	0.360	0.005	0.020	0.352
Y5	0.072	0.000	0.000	0.081	0.072	0.000	0.000	0.081
Y6	0.045	0.000	0.000	0.031	0.045	0.000	0.000	0.031

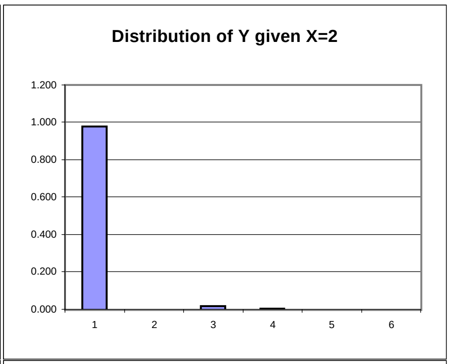
Transition matrix

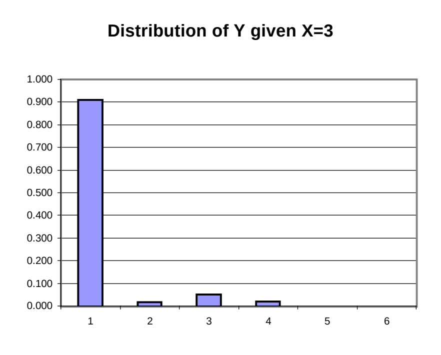
t+1\t	X1	X2	X3	X4	X5	X6	X7	X8
X1	0.369	0.015	0.000	0.433	0.000	0.054	0.000	0.189
X2	0.083	0.796	0.532	0.000	0.009	0.277	0.012	0.000
X3	0.000	0.189	0.424	0.151	0.000	0.226	0.002	0.085
X4	0.548	0.000	0.044	0.416	0.000	0.000	0.000	0.196
X5	0.000	0.000	0.000	0.000	0.299	0.046	0.000	0.244
X6	0.000	0.000	0.000	0.000	0.057	0.265	0.550	0.000
X7	0.000	0.000	0.000	0.000	0.000	0.133	0.340	0.007
X8	0.000	0.000	0.000	0.000	0.635	0.000	0.095	0.279

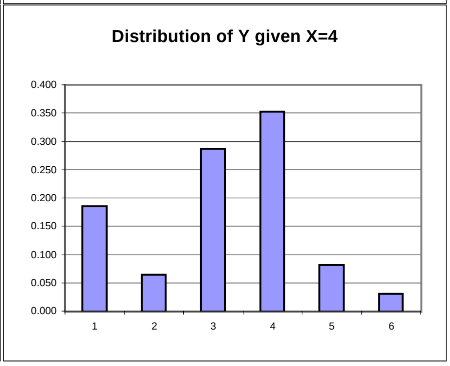
Initial distribution

P0	0.050	0.289	0.178	0.125	0.133	0.001	0.222	0.003









Notes: States 1-4 correspond to single member households and states 5-8 correspond to couples
Lower left block of transition matrix is restricted to be zero.
The conditional probability of Y given X=1,...,4 is retricted to be equal to the conditional distribution when X=5,...,8

TABLE W.1.8

ASSET: Tranportation

Conditional Distribution of **Y** given **X**

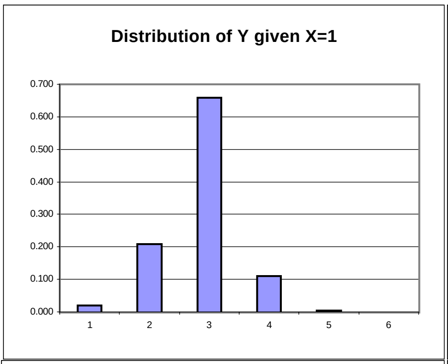
	X1	X2	X3	X4	X5	X6	X7	X8
Y1	0.020	0.032	0.542	0.977	0.020	0.032	0.542	0.977
Y2	0.208	0.797	0.308	0.018	0.208	0.797	0.308	0.018
Y3	0.659	0.169	0.145	0.005	0.659	0.169	0.145	0.005
Y4	0.109	0.002	0.004	0.001	0.109	0.002	0.004	0.001
Y5	0.003	0.000	0.000	0.000	0.003	0.000	0.000	0.000

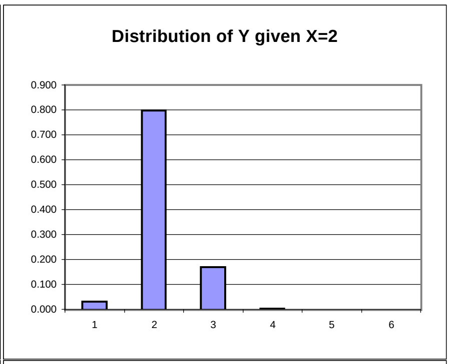
Transition matrix

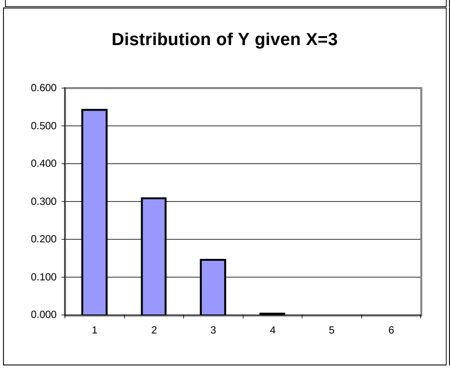
t+1\t	X1	X2	X3	X4	X5	X6	X7	X8
X1	0.877	0.001	0.000	0.000	0.107	0.000	0.000	0.002
X2	0.048	0.802	0.163	0.000	0.045	0.000	0.000	0.000
X3	0.000	0.197	0.296	0.000	0.000	0.000	0.000	0.058
X4	0.075	0.000	0.541	1.000	0.023	0.000	0.000	0.620
X5	0.000	0.000	0.000	0.000	0.819	1.000	0.000	0.000
X6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
X7	0.000	0.000	0.000	0.000	0.000	0.000	0.175	0.134
X8	0.000	0.000	0.000	0.000	0.006	0.000	0.825	0.186

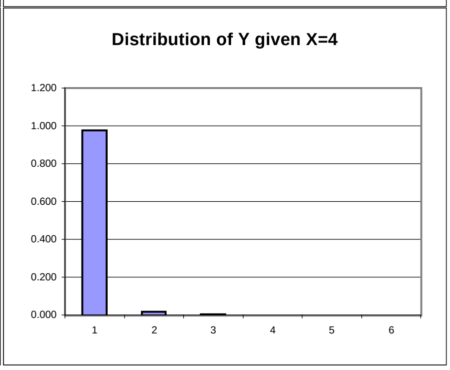
Initial distribution

P0	0.153	0.152	0.041	0.295	0.250	0.047	0.061	0.001









Notes: States 1-4 correspond to single member households and states 5-8 correspond to couples
Lower left block of transition matrix is restricted to be zero.
The conditional probability of Y given X=1,...,4 is retricted to be equal to the conditional distribution when X=5,...,8

TABLE W.1.9
ASSET: Other

Conditional Distribution of **Y** given **X**

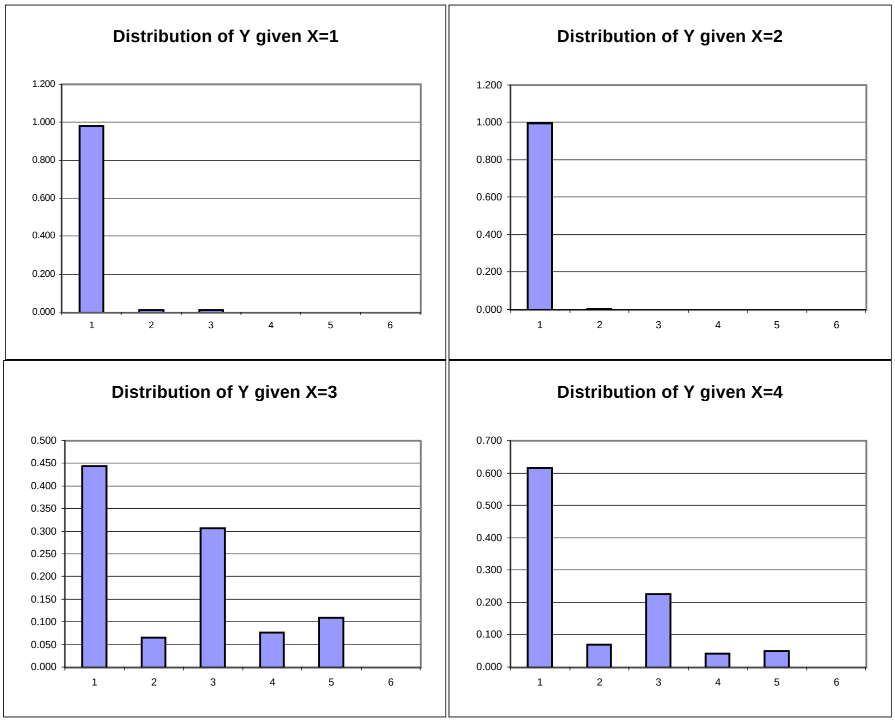
	X1	X2	X3	X4	X5	X6	X7	X8
Y1	0.981	0.996	0.444	0.614	0.981	0.996	0.444	0.614
Y2	0.011	0.003	0.064	0.069	0.011	0.003	0.064	0.069
Y3	0.008	0.000	0.307	0.226	0.008	0.000	0.307	0.226
Y4	0.000	0.001	0.076	0.041	0.000	0.001	0.076	0.041
Y5	0.000	0.000	0.109	0.050	0.000	0.000	0.109	0.050

Transition matrix

t+1\t	X1	X2	X3	X4	X5	X6	X7	X8
X1	0.263	0.364	0.000	0.635	0.000	0.155	0.000	0.188
X2	0.671	0.636	0.005	0.000	0.010	0.187	0.000	0.000
X3	0.000	0.000	0.967	0.191	0.000	0.000	0.000	0.321
X4	0.065	0.000	0.028	0.175	0.006	0.000	0.000	0.000
X5	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.044
X6	0.000	0.000	0.000	0.000	0.862	0.601	0.083	0.000
X7	0.000	0.000	0.000	0.000	0.000	0.057	0.365	0.000
X8	0.000	0.000	0.000	0.000	0.118	0.000	0.552	0.447

Initial distribution

P0	0.589	0.000	0.037	0.016	0.289	0.002	0.066	0.001



Notes: States 1-4 correspond to single member households and states 5-8 correspond to couples
Lower left block of transition matrix is restricted to be zero.
The conditional probability of Y given X=1,...,4 is retricted to be equal to the conditional distribution when X=5,...,8

TABLE W.1.10

ASSET: Debt

Conditional Distribution of **Y** given **X**

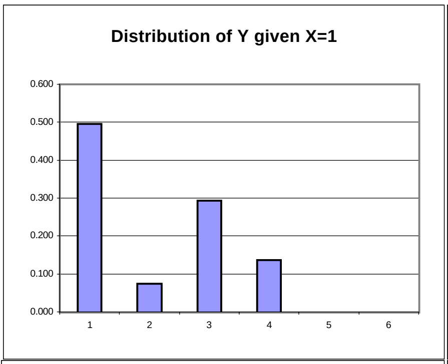
	X1	X2	X3	X4	X5	X6	X7	X8
Y1	0.495	0.958	0.977	0.434	0.495	0.958	0.977	0.434
Y2	0.074	0.015	0.015	0.065	0.074	0.015	0.015	0.065
Y3	0.294	0.017	0.006	0.202	0.294	0.017	0.006	0.202
Y4	0.137	0.010	0.000	0.123	0.137	0.010	0.000	0.123
Y5	0.000	0.000	0.002	0.175	0.000	0.000	0.002	0.175

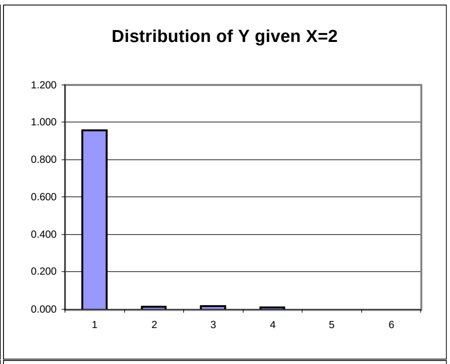
Transition matrix

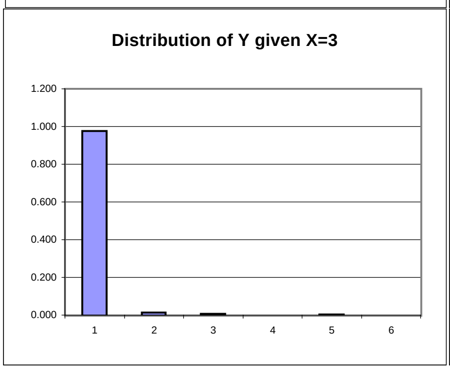
t+1\t	X1	X2	X3	X4	X5	X6	X7	X8
X1	0.866	0.000	0.000	0.572	0.194	0.000	0.000	0.020
X2	0.014	0.000	0.552	0.000	0.125	0.000	0.356	0.000
X3	0.000	1.000	0.448	0.334	0.000	0.013	0.006	0.000
X4	0.120	0.000	0.000	0.094	0.000	0.000	0.000	0.000
X5	0.000	0.000	0.000	0.000	0.494	0.000	0.000	0.972
X6	0.000	0.000	0.000	0.000	0.180	0.000	0.362	0.000
X7	0.000	0.000	0.000	0.000	0.000	0.987	0.277	0.007
X8	0.000	0.000	0.000	0.000	0.008	0.000	0.000	0.001

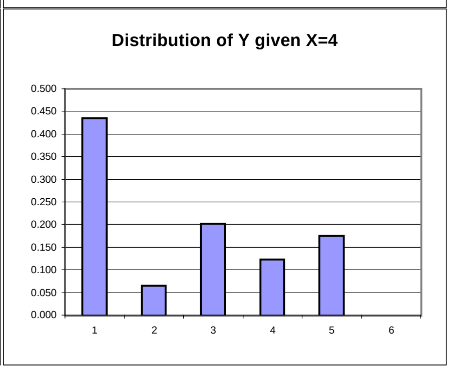
Initial distribution

P0	0.063	0.000	0.513	0.065	0.004	0.278	0.000	0.077









Notes: States 1-4 correspond to single member households and states 5-8 correspond to couples
Lower left block of transition matrix is restricted to be zero.
The conditional probability of Y given X=1,...,4 is retricted to be equal to the conditional distribution when X=5,...,8

TABLE W.1.11

ASSET: House

Conditional Distribution of **Y** given **X**

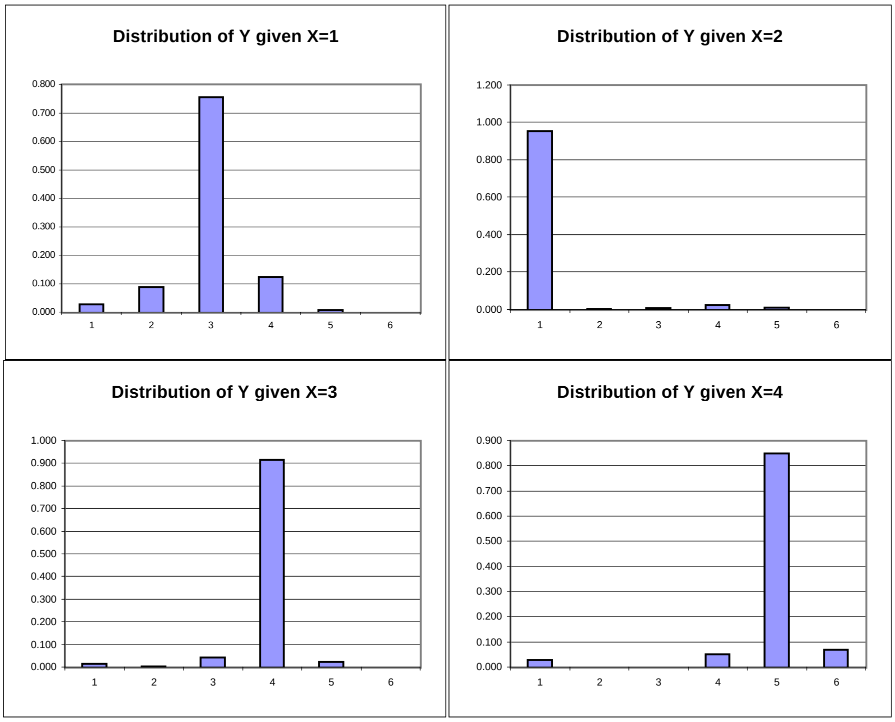
	X1	X2	X3	X4	X5	X6	X7	X8
Y1	0.027	0.954	0.014	0.028	0.027	0.954	0.014	0.028
Y2	0.089	0.003	0.003	0.000	0.089	0.003	0.003	0.000
Y3	0.755	0.008	0.044	0.001	0.755	0.008	0.044	0.001
Y4	0.124	0.023	0.915	0.052	0.124	0.023	0.915	0.052
Y5	0.006	0.011	0.023	0.850	0.006	0.011	0.023	0.850
Y6	0.000	0.000	0.001	0.070	0.000	0.000	0.001	0.070

Transition matrix

t+1\t	X1	X2	X3	X4	X5	X6	X7	X8
X1	0.864	0.004	0.000	0.007	0.170	0.000	0.000	0.000
X2	0.130	0.989	0.084	0.000	0.012	0.262	0.028	0.000
X3	0.000	0.007	0.863	0.000	0.000	0.000	0.120	0.001
X4	0.006	0.000	0.054	0.993	0.000	0.000	0.007	0.144
X5	0.000	0.000	0.000	0.000	0.761	0.000	0.000	0.000
X6	0.000	0.000	0.000	0.000	0.058	0.728	0.032	0.000
X7	0.000	0.000	0.000	0.000	0.000	0.010	0.771	0.013
X8	0.000	0.000	0.000	0.000	0.000	0.000	0.041	0.842

Initial distribution

P0	0.109	0.303	0.181	0.047	0.047	0.069	0.179	0.064



Notes: States 1-4 correspond to single member households and states 5-8 correspond to couples
Lower left block of transition matrix is restricted to be zero.
The conditional probability of Y given X=1,...,4 is retricted to be equal to the conditional distribution when X=5,...,8

TABLE W.1.12

ASSET: Mobile home

Conditional Distribution of **Y** given **X**

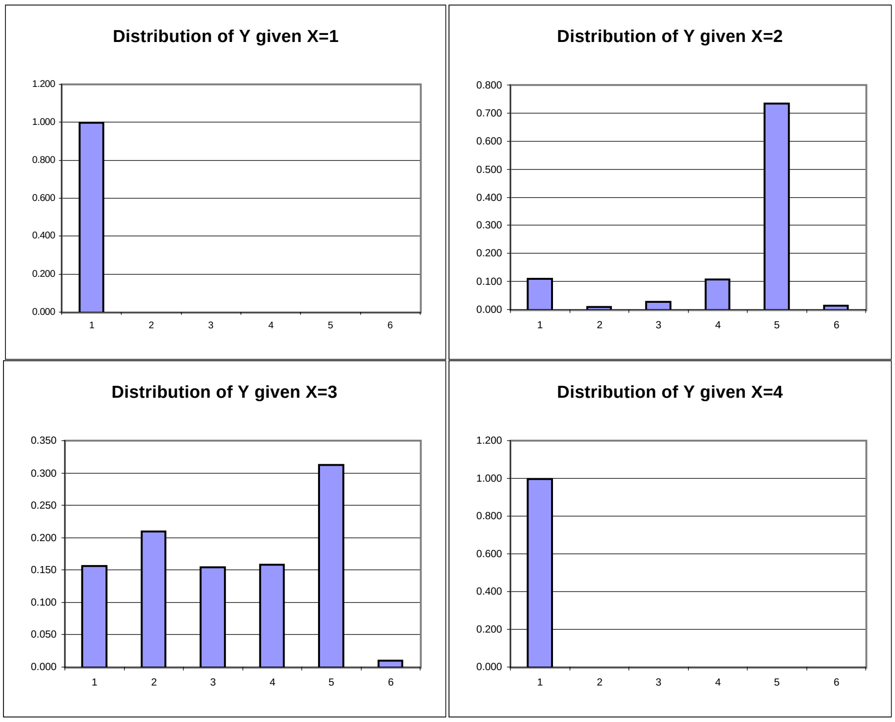
	X1	X2	X3	X4	X5	X6	X7	X8
Y1	0.999	0.110	0.156	0.997	0.999	0.110	0.156	0.997
Y2	0.000	0.008	0.209	0.000	0.000	0.008	0.209	0.000
Y3	0.000	0.027	0.154	0.001	0.000	0.027	0.154	0.001
Y4	0.000	0.108	0.158	0.001	0.000	0.108	0.158	0.001
Y5	0.001	0.734	0.313	0.000	0.001	0.734	0.313	0.000
Y6	0.000	0.013	0.010	0.001	0.000	0.013	0.010	0.001

Transition matrix

t+1\t	X1	X2	X3	X4	X5	X6	X7	X8
X1	0.693	0.052	0.000	0.001	0.125	0.000	0.000	0.008
X2	0.002	0.948	0.000	0.000	0.000	0.000	0.672	0.000
X3	0.000	0.000	0.894	0.008	0.000	0.000	0.000	0.000
X4	0.305	0.000	0.106	0.991	0.226	0.000	0.118	0.000
X5	0.000	0.000	0.000	0.000	0.647	0.059	0.000	0.988
X6	0.000	0.000	0.000	0.000	0.001	0.597	0.129	0.000
X7	0.000	0.000	0.000	0.000	0.000	0.344	0.081	0.004
X8	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000

Initial distribution

P0	0.608	0.016	0.017	0.000	0.012	0.023	0.000	0.325



Notes: States 1-4 correspond to single member households and states 5-8 correspond to couples
Lower left block of transition matrix is restricted to be zero.
The conditional probability of Y given X=1,...,4 is retricted to be equal to the conditional distribution when X=5,...,8

TABLE H.1
Health evolution

Probability of having health condition Y given X

Y\X	X1	X2	X3	X4	X5	X6	X7	X8
DIED	1	0	0	0	0	0	0	0
HEART	0.539	0.296	0.411	0.275	0.366	0.445	0.296	0.161
CANCER	0.328	0.015	0.139	0.015	0.062	0.163	0.969	0.007
STROKE	0.275	0.117	0.138	0.074	0.195	0.112	0.081	0.006
DIABET	0.630	0.004	0.995	0.008	0.034	0.028	0.030	0.014
HIGHBP	0.561	0.065	0.682	0.950	0.974	0.494	0.451	0.040
FALL	0.333	0.289	0.171	0.096	0.268	0.174	0.145	0.032
ARTHRT	0.692	0.285	0.327	0.262	0.394	0.314	0.216	0.164
HIPFRC	0.146	0.126	0.049	0.036	0.107	0.063	0.056	0.004
INCONT	0.363	0.517	0.297	0.010	0.794	0.317	0.304	0.014
LUNG	0.213	0.020	0.111	0.015	0.032	0.933	0.037	0.017
PSYCH	0.559	0.218	0.155	0.073	0.259	0.171	0.112	0.028
COGIMP	0.569	0.422	0.376	0.292	0.427	0.355	0.284	0.264
DEPRES	0.578	0.126	0.125	0.058	0.185	0.161	0.078	0.036

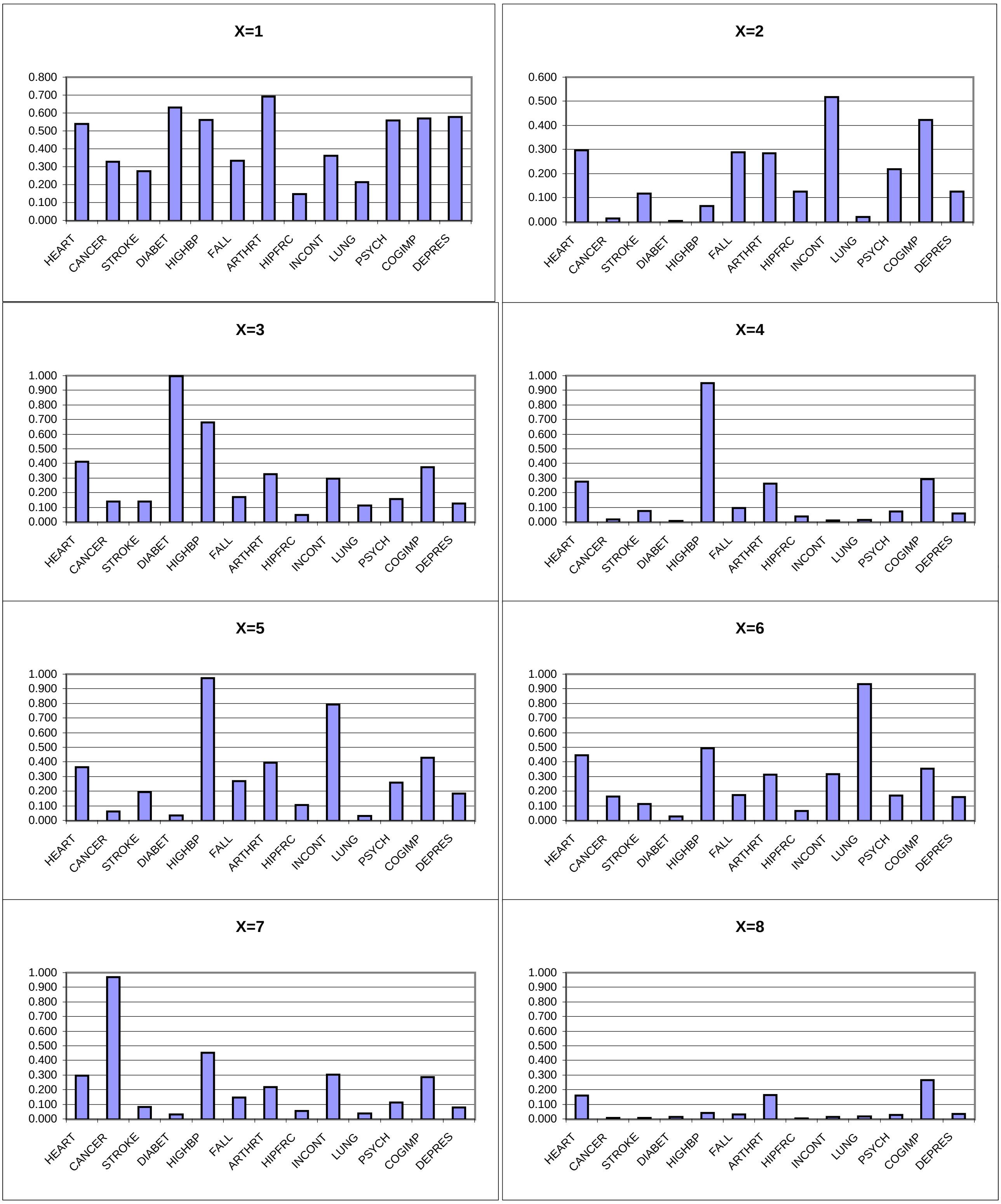
Transition matrix

t+1\t	X1	X2	X3	X4	X5	X6	X7	X8
X1	1	0.100	0.114	0.047	0.111	0.116	0.090	0.033
X2	-	0.876	0.000	-	-	-	-	0.099
X3	-	0.024	0.886	0.025	-	-	-	-
X4	-	-	0.000	0.825	0.000	-	-	-
X5	-	-	-	0.103	0.889	0.003	-	-
X6	-	-	-	-	0.000	0.881	0.000	-
X7	-	-	-	-	-	0.000	0.910	0.028
X8	-	0.000	-	-	-	-	0.000	0.841

Initial distribution

P0	-	0.125	0.120	0.253	0.093	0.085	0.098	0.227

TABLE H.2


X=1
X=2
X=3
X=4
X=5
X=6
X=7
X=8
HEART
CANCER
STROKE
DIABET
HIGHBP
FALL
ARTHRT
HIPFRC
INCONT
LUNG
PSYCH
COGIMP
DEPRES