Heterogeneity, State Dependence and Health

by Timothy J. Halliday,*

Department of Economics and John A. Burns School of Medicine,

University of Hawaii at Manoa

Abstract

This paper investigates the evolution of health over the life-cycle using the Panel Study of Income Dynamics. We allow for two sources of persistence: unobserved heterogeneity and state dependence. The former is modeled by discrete "types." Estimation indicates that there are at least four types suggesting that there is a large degree of heterogeneity governing health dynamics. We find that the degree of state dependence is near unity for over half of the population. The implications of these findings are twofold. First, health inequalities in adulthood have antecedents in childhood. Second, policies that improve health care and its delivery may be an effective means of mitigating the gradient.

Key Words: Health, Dynamic Panel Data Models, Gradient
JEL Codes: I1, C5

1 Introduction

This paper explores the notion that an individual's health follows a persistent stochastic process. We concern ourselves with two tasks. The first is to gain a better understanding of the appropriate way of modeling the evolution of health over the life-course. This is important because, while many empirical studies have investigated the dynamics of both the level of earnings (Lillard and Willis 1978; Abowd and Card 1989) and, more recently, the variance of earnings (Meghir and Pistaferri 2004), few have investigated the dynamics of health.[1] It has been noted by many researchers such as Deaton (1992) and Caballero (1990) that different assumptions about the stochastic process governing income can imply very different types of life-cycle consumption behavior. Accordingly, as health status becomes a more common state variable in structural models, it is becoming increasingly more important that researchers arrive at a better understanding of its dynamics.[2] The second task of this paper is to quantify the relative contributions of unobserved heterogeneity and state dependence in the determination of health. Doing so is important as this will have implications for the effectiveness of health policy.

To analyze health dynamics, we utilize data on Self-Reported Health Status (SRHS) from the Panel Study of Income Dynamics (PSID). We observe that SRHS is highly persistent in the PSID. Indeed, in our data, the first order auto-correlation of a dummy variable indicating

---

[1] Contoyannis, Jones and Rice (2004) and Contoyannis, Jones and Leon-Gonzalez (2004) are notable exceptions.

[2] For examples of structural model using health as a state variable, see Rust and Phelan (1997), French (2005) and Arcidiacono, Heig and Sloan (2007).

that an individual reports fair or poor health is 0.5661 and 0.5643 for men and women, respectively. However, while these correlations do indicate a high degree of persistence, they are not informative of the underlying stochastic properties of the health process.

To gain additional insight, we model the evolution of health over the life-cycle as a first order Markov process. The model allows for two sources of persistence. The first source is unobserved heterogeneity which models an individual's (unobserved) ability to cope with idiosyncratic health shocks such as accidents or exposure to disease-causing agents. We allow all of the parameters of our linear index to vary across individuals. Our model incorporates heterogeneity by allowing for a distribution of unobserved discrete "types." The second source of persistence in the model is state dependence which models the degree to which an individual's ability to cope with a given health shock depends on her health status. State dependence in health captures the idea that people who are ill are less able to cope with health shocks than people who are well. We model state dependence in a flexible manner by allowing for heterogeneity in the coefficient on lagged health. This approach contrasts with much of the applied literature on dynamic panel data models which, typically, only allows for heterogeneity in the constant term and, thus only models state dependence as a homogeneous function of the lagged state.[3]

Estimation yields many interesting findings. First, we find that the number of types in our data is at least four indicating that unobserved heterogeneity is an important factor governing the evolution of health. This suggests that much of what determines health in adulthood can be traced back to childhood. Similar results using a more straightforward approach can be found in Case, Paxson and Lubotsky (2002). Second, we find a very large degree of state dependence

---

[3]For example, see Magnac (2000), Chay, Hoynes and Hyslop (1999), Contoyannis, Jones and Rice (2004) and Hyslop (1999).

for over half of our population. In fact, it is close to unity! Taken at face value, this suggests that medical interventions and attempts to expand access to health care are likely to have large effects on population health. Moreover, to the extent that a large portion of the ubiquitous correlation between health and wealth or the gradient is determined by the causal effect of health on income, as has been asserted by Smith (1999), the large degree of state dependence that we uncover strengthens the case for health policies as a means of mitigating the gradient. However, while we certainly do view this as a provocative result, we advise the reader to interpret this finding with caution for technical reasons which we discuss in the body of the paper.

The balance of this paper is organized as follows. Section 2 describes the data. Section 3 sets up our model of health dynamics. Section 4 describes our estimation procedure. Section 5 discusses our findings. In Section 6, we discuss the relevance of our findings for health policy. Section 7 concludes.

2   Data

We use data from the PSID spanning the years 1984 to 1997. The variables that we employ are SRHS, age and gender. The SRHS question was only asked of heads of household and their spouses and, thus, our sample is restricted to these individuals. We do not employ data prior to 1984 as the SRHS question was not asked in these years. The PSID contains an over-sample of low-income families called the Survey of Economic Opportunity (SEO). Because the sample was chosen based on income, we follow Lillard and Willis (1978) and drop it due to endogenous

selection.[4]

SRHS is a categorical variable that takes on integer values between 1 and 5. 1 means that the individual perceives that their health is excellent; 2 is very good; 3 is good; 4 is fair; 5 is poor. While these data are subjective measures, there is an extensive literature that has shown a strong link between SRHS and more objective health outcomes such as mortality and the prevalence of disease (Mossey and Shapiro 1982; Kaplan and Camacho 1983; Idler and Kasl 1995; Smith 2004). It should also be mentioned that many supposedly objective health measures are not without their limitations. For example, self-reports of specific morbidities such as diabetes or cancer are often inaccurate since many people are unaware that they even have these conditions due to low consumption of medical services. In addition, these measures typically do not account for the severity of the condition. To lower the number of parameters that we estimate, we map all self-reports of fair or poor health into unity and all other reports into zero. This is the conventional way of partitioning the SRHS variable in the health economics, public health and epidemiology literatures.

We restrict our sample to individuals between ages 22 and 60. We do not include people younger than age 22 because there are not that many household heads younger than this age. We do not include people older than age 60 to mitigate any possible bias resulting from attrition due to mortality. We drop individuals whose age declines across successive survey years. We also drop individuals whose age increases by more than two years across successive survey years. Finally, we restrict our sample to white men and women. Table 1 reports the descriptive statistics

---

[4]Meghir and Pistaferri (2004) include the SEO in their work. Their reason for its inclusion is that by estimating linear models in differences, the unobserved heterogeneity is purged from the regression equation and, consequently the initial condition problem posed by the endogenous selection of the SEO is solved. However, this will not work in our case as we work with a non-linear model in which it is impossible to purge the model of any fixed effects.

from the resulting sample.

3   The Empirical Model

We let $h_{i,t} \in \{0,1\}$ denote the health of individual i at age t. Throughout this paper, we refer to individuals for whom $h_{i,t} = 1$ as "ill" and individuals for whom $h_{i,t} = 0$ as "well." We assume that health evolves according to the following process:

$$h_{i,t} = 1(\alpha_i + \gamma_i h_{i,t-1} + \boldsymbol{\rho}_i \mathbf{T} + \varepsilon_{i,t} \geq 0) \tag{1}$$

where $\mathbf{T} = [t, t^2]'$. The model allows for four determinants of health: idiosyncratic risk, aging, state dependence and heterogeneity.

Idiosyncratic risk is represented by $\varepsilon_{i,t}$ and models illness-causing agents or events that affect individual i when he is of age t. Throughout this paper, we will refer to $\varepsilon_{i,t}$ as a "health shock." Examples of these shocks are accident occurrence, disease onset or exposure to bacteria and viruses. Because the effects of these agents can vary considerably in their intensity, $\varepsilon_{i,t}$ can assume a continuum of values. We assume that $\varepsilon_{i,t}$ is independent of $(\alpha_i, \gamma_i, \boldsymbol{\rho}'_i, h_{i,0})$ and that it is distributed *i.i.d.* across time with a logistic distribution.[5] These assumptions imply that

$$P(h_{i,t} = 1 | h_{i,t-1}, ..., h_{i,0}, \theta_i) = \frac{\exp(\boldsymbol{\theta}'_i \mathbf{Z}_{i,t-1})}{1 + \exp(\boldsymbol{\theta}'_i \mathbf{Z}_{i,t-1})} \tag{2}$$

---

[5]Attempts have been made by Hyslop (1999) and Contoyannis, Jones and Rice (2004) to relax the *i.i.d.* assumption by allowing for serial correlation in $\varepsilon_{i,t}$. However, both attempts resulted in negative estimates of the serial correlation in $\varepsilon_{i,t}$. Both papers conclude that this finding is odd and probably reflects an identification issue in these more complicated models.

where $\boldsymbol{\theta}_i \equiv (\alpha_i, \gamma_i, \boldsymbol{\rho}'_i)'$ and $\mathbf{Z}_{i,t-1} = (1, h_{i,t-1}, \mathbf{T}')'$.

The coefficients in $\boldsymbol{\rho}'_i$ are our aging coefficients. They allow the effects of health shocks to increase with age. In addition, they account for the fact that many diseases like Alzheimer's, cancer, hypertension and heart disease are more likely to manifest later in life. Within the context of the Grossman model of health investment (Grossman 1972), these coefficients can be interpreted as the rate at which the health capital stock depreciates.

The coefficient γ_i is our state dependence coefficient. It models the notion that a person who is ill may be less able to cope with a given health shock than when that same person is well. To give a concrete (albeit extreme) example, exposure to a flu virus is more likely to affect a person's health if she is HIV positive than if she is HIV negative.[6]

Our model allows for a large degree of flexibility in heterogeneity. We do this by letting the vector, $\boldsymbol{\theta}_i$, vary across individuals. So, not only is there heterogeneity in the "constant" term, α_i, but there is also heterogeneity in all of the model's parameters. This contrasts with the majority of the dynamic panel data literature in which, typically, only α_i is individual-specific. Unobserved heterogeneity models an individual's ability to resist health shocks.[7]

While our motivation of the model is rooted in epidemiology, two important ties to the economics literature can be made. The first is that many recent investigations into life-cycle

---

[6]It is important to contrast our model with an obvious alternative formulation in which health is determined by a continuous index given by $H_{i,t}$ which follows an AR(1) process and $h_{i,t} = 1 \, (H_{i,t} \geq 0)$. While this alternative does allow health shocks to have persistent effects, it does not allow for state dependence. In other words, in this model, the effects of a shock on future health outcomes are not conditioned by the agent's health status.

[7]Note that we can write part of the index in equation (1) as

$$\alpha_i + \gamma_i h_{i,t-1} = \kappa_{i,0}(1 - h_{i,t-1}) + \kappa_{i,1} h_{i,t-1}$$

where $\gamma_i \equiv \kappa_{i,1} - \kappa_{i,0}$ and $\alpha_i \equiv \kappa_{i,0}$. What this calculation tells us is that, as long as we expect to see heterogeneity in both the persistence of illness, or the transition from $h_{i,t-1} = 1$ to $h_{i,t} = 1$, and in the onset of illness, or the transition from $h_{i,t-1} = 0$ to $h_{i,t} = 1$, then we should expect heterogeneity in both the constant coefficient (α_i) and the state dependence coefficient (γ_i).

consumer behavior such as Arcidiacono, Sieg and Sloan (2004), French (2005) and Rust and Phelan (1997), have incorporated exogenous uncertainty over health states. Our empirical investigation will provide additional insights into how this uncertainty should be modeled. The second is that our empirical model can be interpreted as a reduced form version of a model of health investment. For example, a simple model in which health is modeled by discrete states and the costs of health investment are state dependent will imply that health will follow a persistent process. One interpretation of equation (1) is that it is a reduced form version of this process.

Finally, it is important to emphasize the ties between the identification of state dependence in health and in labor market outcomes. As pointed out by Hyslop (1999), many sources of state dependence in labor force participation have been cited including intertemporally nonseparable preferences for leisure (Hotz, Kydland and Sedlacek 1988) and search costs which depend on participation states (Eckstein and Wolpin 1990). However, it is important to recognize that, regardless of the underlying source, understanding the magnitude of state dependence in labor force participation will have policy implications since it tells us about the effectiveness of policies that alleviate short-term unemployment. Similarly, the magnitude of state dependence in health will be informative of the relative importance of unobserved individual characteristics *vis-a-vis* idiosyncratic health shocks. To the extent that the effects of these shocks can be mitigated by improvements in health care and its delivery, understanding the magnitude of state dependence in health will have implications for many health policy debates. In both the cases of labor and health economics, the statistical properties of the data will contain information that is pertinent for the conduct of policy.

4 MLE

We estimate the model in equation (1) using an MLE procedure which has been discussed in Heckman (1981a and 1981b). Individual i ($i = 1, ..., N$) experiences $h_{i,t}$ at time $t \in \{0, ..., T_i\}$. However, the econometrician only observes $h_{i,t}$ for $t \in \{\tau_i, ..., T_i\}$ where $\tau_i \geq 0$, and, thus, we have an initial conditions problem. The procedure that we use accounts for this.

We now construct the likelihood function. The likelihood of a sequence of health outcomes conditional on $(\boldsymbol{\theta}_i', h_{i,\tau_i})$ for individual i for $t = \tau_i, ..., T_i$ is given by

$$P(h_{i,T_i}, ..., h_{i,\tau_i+1} | h_{i,\tau_i}, \boldsymbol{\theta}_i') =$$
$$\prod_{t=\tau_i+1}^{T_i} \Lambda \left(\boldsymbol{\theta}_i' \mathbf{Z}_{i,t-1} (2h_{i,t} - 1) \right). \tag{3}$$

We assume that the heterogeneity vector has a discrete support where it can take on one of A values so that $\boldsymbol{\theta}_i \in \{\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_A\}$. The probability weight that is associated with each point of support is π_a. This approach is similar to Heckman and Singer (1984) who use a discrete distribution to approximate the distribution of unobserved heterogeneity when estimating duration models in the presence of heterogeneity via nonparametric maximum likelihood.[8] Let $P_{\tau_i}(h_{i,\tau_i} | \boldsymbol{\theta}_a')$ denote the probability of the first observation conditional on $\boldsymbol{\theta}_i = \boldsymbol{\theta}_a$. We can now

---

[8]Heckman and Singer (1984) verify that Kiefer-Wolfowitz conditions are satisfied for a general class of duration models with unobserved heterogeneity. These conditions ensure consistent estimation of the distribution of unobserved heterogeneity and structural parameters in this class of duration models. In addition, they cite a theorem from Lindsay (1983) that says that the nonparametric maximum likelihood estimate of the structural parameters and distribution of the unobserved heterogeneity will be such that the estimate of the heterogeneity distribution is discrete. Our approach contrasts with Heckman and Singer since we assume that the population distribution of the heterogeneity is discrete and, hence, we do not need to verify the Kiefer-Wolfowitz conditions.

obtain the unconditional likelihood of observing $(h_{i,\tau_i}, ..., h_{i,T_i})$ via

$$P(h_{i,T_i}, ..., h_{i,\tau_i}) =$$

$$\sum_{a=1}^{A} P(h_{i,T_i}, ..., h_{i,\tau_i}|\boldsymbol{\theta}'_a)\pi_a = \tag{4}$$

$$\sum_{a=1}^{A} \prod_{t=\tau_i+1}^{T_i} \Lambda(\boldsymbol{\theta}'_a \mathbf{Z}_{i,t-1}(2h_{i,t}-1)) P_{\tau_i}(h_{i,\tau_i}|\boldsymbol{\theta}'_a)\pi_a.$$

Summing over the heterogeneity addresses the incidental parameters problem (Neyman and Scott 1948).

Our model implies a recursive definition for $P_{\tau_i}(h_{i,\tau_i}|\boldsymbol{\theta}'_a)$. To compute this probability, first, we let the probability of being well in $t = 0$ conditional on $\boldsymbol{\theta}_a$ be given by $p_a \equiv P_0(h_{i,0} = 0|\boldsymbol{\theta}'_a)$. The probability of observing $h_{i,t}$ conditional on $\boldsymbol{\theta}_a$ in any subsequent period is then given by

$$P_t(h_{i,t}|\boldsymbol{\theta}'_a) = \sum_{d=0}^{1} P_t(h_{i,t}|h_{i,t-1} = d, \boldsymbol{\theta}'_a) P_{t-1}(h_{i,t-1} = d|\boldsymbol{\theta}'_a) \tag{5}$$

$$= \sum_{d=0}^{1} \Lambda((\boldsymbol{\theta}'_a \mathbf{Z}_{i,t-1})(2h_{i,t}-1)) P_{t-1}(h_{i,t-1} = d|\boldsymbol{\theta}'_a).$$

Substituting, we get

$$P_t(h_{i,t}|\boldsymbol{\theta}'_a) = \sum_{d=0}^{1} \Lambda((\boldsymbol{\theta}'_a \mathbf{Z}_{i,t-1})(2h_{i,t}-1))$$

$$\sum_{d=0}^{1} \Lambda((\boldsymbol{\theta}'_a \mathbf{Z}_{i,t-2})(2h_{i,t-1}-1))... \tag{6}$$

$$\sum_{d=0}^{1} \Lambda((\boldsymbol{\theta}'_a \mathbf{Z}_{i,0})(2h_{i,1}-1)) * (p_a)^{1-d} (1-p_a)^d.$$

Using the above recursive formulation, we can calculate $P_{\tau_i}(h_{i,\tau_i}|\boldsymbol{\theta}'_a)$.[9] Of course, this is a burdensome task if τ_i is large since computation will involve calculating the sum of the probabilities of all possible sequences of health outcomes that could have led to h_{i,τ_i}. Fortunately, the above recursive definition simplifies matters greatly.

Our treatment of the initial condition in (6) imposes no additional parametric assumptions on the model beyond the assumptions that health evolves according to equation (1) and that the heterogeneity distribution is discrete. The reason is that because equation (1) contains no time varying regressors, we can roll the model back to $t = 0$. In addition, because the heterogeneity has a discrete distribution, we can treat the probabilities p_a as additional parameters in the model to be estimated. In other words, because of these two assumptions, we do not need to make an additional parametric assumption about $P_0(h_{i,0} = 0|\boldsymbol{\theta}'_a)$.

Using these probabilities, taking logs and summing over individuals, we obtain the likelihood function:

$$L(\boldsymbol{\beta}) = \tag{7}$$

$$\sum_{i=1}^{N} \log\left(\sum_{a=1}^{A}\prod_{t=\tau_i+1}^{T_i} \Lambda(\boldsymbol{\theta}'_a \mathbf{Z}_{i,t-1}(2h_{i,t}-1))P_{\tau_i}(h_{i,\tau_i}|\boldsymbol{\theta}'_a)\pi_a\right).$$

---

[9]Heckman (1981a) proposes using this method which involves using the underlying statistical model to calculate $P_{\tau_i}(h_{i,\tau_i}|\boldsymbol{\theta}'_a)$ which can in turn be used to calculate $P(h_{i,T_i}, ..., h_{i,\tau_i})$. This procedure addresses the initial condition problem that occurs when the stochastic process has been running prior to τ_i. Since our underlying statistical model does not have any time varying regressors, we do not need to concern ourselves with the distribution of the time varying regressors for $t < \tau_i$. However, in the presence of time varying regressors, auxiliary distributional assumptions must be made. In addition, the computations become rather involved. An alternative to this is provided by Wooldridge (2005) who proposes modeling the distribution of the heterogeneity conditional on h_{i,τ_i} and any time varying regressors that may be present. Doing this does not require internal consistency with the underlying statistical model nor does it require computations that are as involved as the previous method, but it does require additional distributional assumptions. A third solution to the initial conditions problem assumes that the process has been running sufficiently long prior to the sampling period and that the process is in equilibrium. It then uses the stationary distribution for the process as the probability of the first observation. However, this will not work in our case as health is non-stationary process.

where $\boldsymbol{\beta} \equiv (\boldsymbol{\theta}'_1, ..., \boldsymbol{\theta}'_A, \pi_1, ..., \pi_{A-1}, p_1, ..., p_A)$. $\boldsymbol{\beta}$ is of dimension $7A - 1$. $L(\boldsymbol{\beta})$ was maximized using the Fletcher-Powell algorithm, a variant of Newton's Method, which only requires the computation of the the gradient vector $\nabla L(\boldsymbol{\beta})$.[10] Since evaluating the likelihood in (7) can be time-consuming, we calculated analytical gradients, as opposed to numerical gradients.[11]

Often times, when the number of support points for the mixing distribution exceeds two, attempts to estimate the probabilities π_a directly will result in trivial probabilities for some support points so that the number of support points effectively collapses to two or (sometimes) three. To avoid this complication, we follow Arcidiacono and Jones (2003) and note that the MLE of π_a is given by

$$\widehat{\pi}_a = \frac{1}{N} \sum_{i=1}^{N} \frac{f_{i,a}}{q_i} \tag{8}$$

where $f_{i,a} \equiv \prod_{t=\tau_i+1}^{T_i} \Lambda(\boldsymbol{\theta}'_a \mathbf{Z}_{i,t-1}(2h_{i,t} - 1)) P_{\tau_i}(h_{i,\tau_i} | \boldsymbol{\theta}'_a) \pi_a$ and $q_i \equiv \sum_{a=1}^{A} f_{i,a}$. This insight suggests the following iterative strategy. First, choose a set of values for the mixing distribution probabilities and call these values π_a^1. Similarly, choose initial values for $\varpi \equiv (\boldsymbol{\theta}'_1, ..., \boldsymbol{\theta}'_A, p_1, ..., p_A)$ and call these values ϖ^1. Next, calculate the gradient with respect to ϖ using the probabilities π_a^1 and ϖ^1 and iterate to get ϖ^2. Then, evaluate equation (8) using π_a^1 and ϖ^1 to obtain π_a^2. Repeat the process. To verify that this procedure does in fact work, using two support points, we calculated the MLE using this method and using the alternative method in which the probabilities π_a were estimated directly (*i.e.* we differentiated the likelihood function with respect to π_a as well). Both procedures yielded the same estimates.

---

[10]Typically, when the distribution of unobserved heterogeneity is treated in the spirit of Heckman-Singer, the EM algorithm is used for optimization. However, Newton's method has also been used (see Baker and Melino 2000, for example).

[11]All computer programs and data used are available upon request from the author.

5   Estimation Results

5.1   Model Selection

We investigate model selection along two dimensions. The first is the specification of the index inside equation (1) and the second is the number of support points.[12]   To accomplish this, we employ the Akaike Selection Criterion (AIC) which is given by

$$AIC = -\frac{2}{N}\log L\left(\widehat{\beta}\right) + \frac{2p}{N} \propto -\log L\left(\widehat{\beta}\right) + p$$

where $p = \dim(\beta)$ (Amemiya 1985). When comparing models, the preferred model has the lowest AIC.

5.1.1   Index

We consider three indices. The first is $\boldsymbol{\rho}_a = [\rho^1, 0]$ for all a. We call this model the linear model. The second is $\boldsymbol{\rho}_a = [\rho^1, \rho^2]$ for all a. We this the homogeneous quadratic model. The third model is unrestricted so that we have that $\boldsymbol{\rho}_a = [\rho_a^1, \rho_a^2]$. We call this model the heterogeneous quadratic model. In all three models, we allow for heterogeneity in both α_a and γ_a.

The results of this exercise are reported in Table 2 for $A = 2$ and $A = 3$. When we only have two points of support, we see that the AIC slightly favors the homogeneous quadratic model for both men and women. When we move to three points, the AIC still favors the homogeneous

---

[12]When testing for the number of support points, likelihood-based test statistics are inappropriate because under the null hypothesis one of the probabilities π_a must be set to zero. This places the parameter vector at the edge of a compact set and, thus, violates the regularity conditions of MLE. Consequently, the resulting test statistic will not be χ^2. However, because model selection criteria do not require that the true parameter lie in the interior of a compact set, they are appropriate. For a formal derivation of this, we refer the reader to Leroux (1992).

quadratic model for men, but now favors the heterogeneous quadratic model for women. What is important to note, however, is that our choice of index does not alter the AIC too much.

5.1.2 Support Points

Table 3 reports the AIC results for the number of support points. We consider up to four points of support. We did not venture beyond four points due to computational limitations. For each value of A that we considered, we estimated the model with the homogeneous quadratic function of age.[13]

What we see is that, for both men and women, the AIC increases with the number of support points, but at a decreasing rate. In contrast to altering the index, adding support points has a dramatic effect on the AIC. The preferred model has four points of support suggesting that there is a tremendous amount of heterogeneity in health dynamics. Presumably, if we had continued to add support points, we would have found evidence of even more heterogeneity. However, because the relationship between the AIC and A appears to be concave, we conjecture that eventually the selection criterion would have started to decline. The results of this table stand in contrast to results in Deb and Trivedi (1997) who find that only two points of support were necessary when estimating a model for the demand of medical care.

5.2 Health Dynamics

---

[13]When choosing the number of support points we did not concern ourselves with the index selection for two reasons. The first is that, as indicated by Table 2, changing the index did not alter the AIC tremendously. The second is that the computations in this exercise were quite intensive. Utilizing more complicated indices, such as the heterogeneous quadratic model, would only have made it worse.

Tables 4 and 5 report the parameter estimates and their standard errors for the homogeneous quadratic model with four support points for men and women.[14] This model had the lowest AIC of all the models that we considered. Each column of the tables corresponds to a separate support point which we call a "type." We have defined each according to the magnitude of α_a. The lowest value (*i.e.* most negative) of α_a is defined to be "Type 1" and the highest is "Type 4."[15]

There are several salient features of these tables. The first is that type 1 men and women appear to have a tremendous degree of state dependence since the estimates of α_1 are very low and the estimates of γ_1 are very high. The second is that many of the standard errors are quite large. We presume that this is a consequence of the large number of support points that we have. The third is that the probabilities associated with the support points, π_a, are all reasonably large so that there appear to be non-trivial numbers of each type.

In Figures 1 through 4, we take the parameter estimates for men in Table 4 and map them into health transition probabilities.[16] Each figure corresponds to a separate type and plots two profiles. The first is the probability of being ill today conditional on having been ill yesterday. We call this profile the persistence of illness. The second is the probability of being ill today conditional on having been well yesterday. We call this profile the onset of illness. Each profile

---

[14]Standard errors were calculated using the "sandwich" standard errors. The gradient vector from the likelihood function was used to calculated the average of its outer product. To calculate the Hessian, we numerically differentiated the gradient vector.

[15]The high estimate of p_4, or the probability of being well at age 22 for type 4 men, in Table 4 is a bit of a curiosity. In Table 5, we see that the estimates of p_a are declining with the type. In Table 4, the probabilities decline only until type 3. We believe that the reason for this anomaly is that these probabilities can be difficult to pin down. Moreover, the facts that we have fewer men than women in our sample, men have lower morbidity and the probability of being type 4 is somewhat low suggest that the probability, p_4, might be hard to estimate for men.

[16]The figures for women were similar and, so we do not report them.

displays health dynamics for ages 23 to 60. We plot 95% confidence bands around each profile.[17] It is important to realize that because many of the standard errors in Tables 4 and 5 are quite large, some of these confidence bands include zero or unity.

The figures show a large degree of heterogeneity in health dynamics. Figure 1, which corresponds to Type 1 men, shows that the persistence of illness is close to unity and that the onset of illness is close to zero at all ages. This suggests that Type 1 men exhibit a tremendous degree of state dependence. Figures 2 through 4 correspond to Types 2 through 4 and provide what we believe to be more reasonable estimates. Type 2 men are the healthiest and Type 4 men are the unhealthiest. These figures show more far more muted degrees of state dependence than Figure 1.

Figures 5 through 12 display the degree of state dependence, which is defined to be the difference between the persistence and onset profiles, for men and women. Each figure corresponds to a separate type. Figures 5 through 8 correspond to men and the remaining figures correspond to women. Each figure includes a 95% confidence band which as before might include either zero or unity due to the high standard errors of our parameter estimates. The degree of state dependence is close to unity for Type 1 men and women. It is important to note that over 50% of men and women are estimated to be Type 1. If we take this result literally then it suggests that medical interventions that keep people healthy are likely to be extremely effective. The degree of state dependence for Type 2 men and women is very low - below 10% for most ages. For Types 3 and 4, we see a more intermediate degree of state dependence that is somewhere between 10% and 20%.

---

[17]The δ-method was used to calculate the standard errors.

At this point, we must subject the reader to the caveat that we believe that the high degree of state dependence that we uncovered for Type 1 people is somewhat suspicious. We conjecture that this may reflect an undesirable property of working with mixing distributions with large number of support points. We cite two reasons for our suspicions. First, Halliday (2007) using the same data, but alternative semi-parametric tests, did not find strong evidence of state dependence. However, it is important to be aware that this finding may be a consequence of the low power of his test statistic. Second, Baker and Melino (2000) investigated the properties of duration models with unobserved heterogeneity and showed that certain pathologies arose as the number of support points increased. We believe that it is reasonable to expect that similar pathologies might exist in our context as well. As a testament to this, in an earlier draft of this paper where we only allowed for two support points, we found a far more muted degree of state dependence.

6 Discussion

Can the estimates in this paper inform us about health policy and if so, what do they tell us? The answer to this question, to a large degree, depends on how good our health measures are and how well our modelling decisions and estimation procedures capture individual level health dynamics. While many criticisms can be made, including that our approach is too "reduced form" in that it does not model individual decisions, we do believe that some lessons can be learned from our results. In addition, we would like to re-emphasize a point made earlier that our approach, which is focused on deepening our understanding of the statistical properties of our data while make parametric restrictions that are as weak as possible, can be informative of policy.

In fact, because measuring health is so difficult and incorporating it into life-cycle consumption models often results in models that are very hard to estimate and potentially fragile in the face of mis-specified distributional and modeling assumptions, many authors such as Adams, Hurd, McFadden, Merrill and Ribeiro (2003), Adda, Banks and Van Gaudecker (2006) and Halliday (2007) have also adopted less structural approaches in health applications.

So, what can we learn? If we can take our estimates literally, we can make the following conclusions. First, the high degree of state dependence that we find for Type 1 men and women suggests that there is ample room for improvements in health care and its delivery to reduce morbidity in society. This is especially true given that Type 1 people constitute over 50% of the population. Second, the relatively low degrees of state dependence that we find for Types 2, 3 and 4 suggest that policies that target these adults will be substantially less effective than for the Type 1 people.

Of course, the types are unobserved to the policy makers and, thus, it is difficult to know precisely who should be targeted. This can be circumvented, to some extent, if we extend the estimation procedure to allow the probability of being a certain type to depend on additional observed characteristics. We leave this task for future work. However, it should also be mentioned that the fact that the types are unobserved is not necessarily problematic. The reason is that we have predicted a large degree of state dependence for roughly half of the population. Moreover, while the other half has more muted degrees of state dependence, it is still positive and, typically, statistically different from zero. Accordingly, our results suggest that even if a policy targets the entire population, it is still likely to have large effects on population health.

7   Conclusions

This paper investigated the evolution of health over the life-course by estimating several specifications of a flexible model of health dynamics which allowed for two sources of persistence: unobserved heterogeneity and state dependence. Our analysis suggested that altering the linear index of our model did little to improve its fit. In contrast, however, adding support point to the mixing distribution improved the model's fit tremendously. We found that at least four support points were necessary indicating that there is a large degree of heterogeneity in our data. This suggests that much of what determines health in adulthood can be traced back to childhood and is consistent with recent work by Case, Paxson and Lubotsky (2002). In addition, we found that for over half of the population the degree of state dependence was close to unity. However, we are cautious to warn readers that we find this result to be somewhat suspicious. Further research on this topic using alternative methods, health measures and data sets is necessary to investigate the robustness of this result.

The results of this paper shed light on the gradient: the much-studied but little-understood statistical correlation between health and socioeconomic status (Adams, *et al* 2003; Adler, *et al* 1994). If it is the case that the gradient is largely determined by the causal impact of health status on earnings and wealth - as suggested by Smith (1999) - then the relevant policy prescription is to directly target health *via* improvements in health care and its delivery (Deaton 2002). The argument for health policies is further strengthened if health exhibits a high degree of state dependence as this implies that interventions will have large dynamic effects which operate through the causal effect of health on itself. If we take the results of this paper at face value, they strengthen the case for targeting health as a means of mitigating the gradient.

This paper suggests several further research topics. First, additional work should be done to incorporate mortality into the existing framework, while addressing the selection bias that it induces. As discussed by Wooldridge (2000), correcting selection bias due to non-random attrition - which encompasses mortality - is difficult, particularly, when the econometrician is unwilling to make stringent assumptions. The reason is that while the sample may be random in the initial time period, the sample is, generally, non-random for every subsequent period. Consequently, we expect any estimates of a transition probability in these subsequent periods to suffer from selection bias even if mortality were to be added as a third (absorbing) state.

Second, further work should estimate models with higher orders of state dependence. The main challenge here concerns the initial condition. When the model only contains first order state dependence, only two initial condition probabilities need to be estimated for each support point (*i.e.* that p_a's in the model). However, with second order state dependence then four of these probabilities must be estimated for each support point. Generally, in the case of Kth order state dependence, we will need to estimate $A * K^2$ probabilities.

Finally, additional work should estimate structural models that incorporate health as a state variable. While there have been some studies that have done so such as Rust and Phelan (1997), French (2003) and Arcidiacono, Heig and Sloan (2007), this is still a relatively new field. Particular attention should be paid to how assumptions about the health process affect estimation results. While many researchers working on consumption have strived to better understand how assumptions on the underlying income process affect identification, similar work has not been done in life-cycle models with health uncertainty.

References

[1] Abowd, J. and D. Card (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57, 411-445.

[2] Adams, H.P., M.D. Hurd, D. McFadden, A. Merrill and T. Ribeiro (2003): "Healthy, Wealthy and Wise? Tests for Direct Causal Pathways between Health and Socioeconomic Status," *Journal of Econometrics*, 112, 3-56.

[3] Adda, J., J. Banks and H.M. von Gaudecker (2006): "The Impact of Income Shocks on Health: Evidence from Cohort Data," unpublished manuscript.

[4] Adler, N.E., T. Boyce, M.A. Chesney, S. Cohen, S. Folkman, R.L. Kahn and S.L. Syme (1994): "Socioeconomic Status and Health: The Challenge of the Gradient," *American Psychologist*, 49, 15-24.

[5] Amemiya, T. (1985): *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

[6] Arcidiacono, P. and J.B. Jones (2003): "Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm," *Econometrica*, 71, 933-946.

[7] Arcidiacono, P., H. Sieg and F. Sloan (2007): "Living Rationally Under the Volcano? Heavy Drinking and Smoking Among the Elderly," *International Economic Review*, 48, 37-65.

[8] Baker, M. and A. Melino (2000): "Duration Dependence and Non-parametric Heterogeneity: A Monte Carlo Study," *Journal of Econometrics*, 96, 357-393.

[9] Caballero, R.J. (1990): "Consumption Puzzles and Precautionary Savings," *Journal of Monetary Economics*, 25, 113-136.

[10] Case, A., D. Lubotsky and C. Paxson (2002): "Economic Status and Health in Childhood: The Origins of the Gradient," *American Economic Review*, 92, 1308-1334.

[11] Chay, K.Y., H. Hoynes and D. Hyslop (1999): "A Non-experimental Analysis of True State Dependence in Monthly Welfare Participation Sequences," *Proceedings of the American Statistical Association*, 9-17.

[12] Contoyannis, P., A.M. Jones and R. Leon-Gonzalez (2004): "Using Simulation-based Inference with Panel Data in Health Economics," *Health Economics*, 13, 101-122.

[13] Contoyannis, P., A.M. Jones, and N. Rice (2004): "Simulation-based Inference in Dynamic Panel Probit Models: An Application to Health," *Empirical Economics*, 29, 49-77.

[14] Deaton, A. (1992): *Understanding Consumption.* Oxford: Oxford University Press.

[15] Deaton, A. (2002): "Policy Implications of the Gradient of Health and Wealth," *Health Affairs,* 21, 13-30.

[16] Deb, P. and P. Trivedi (1997): "Demand for Medical Care by the Elderly: A Finite Mixture Approach," *Journal of Applied Econometrics*, 12, 313-336.

[17] Eckstein, Z. and K. Wolpin (1989): "Dynamic Labor Force Participation of Married Women and Endogenous Work Experience," *Review of Economic Studies*, 56, 375-390.

[18] Grossman, M. (1972): "On the concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 80, 223-255.

[19] French, E. (2005): "The Effects of Health, Wealth and Wages on Labour Supply and Retirement Behavior," *Review of Economic Studies*, 72, 395-427.

[20] Halliday, T. (2007): "Income Risk and Health," unpublished manuscript.

[21] Halliday, T. (2007): "Testing for State Dependence with Time Variant Transition Probabilities," forthcoming *Econometric Reviews.*

[22] Heckman, J.J. (1981a): "Statistical Models for Discrete Panel Data," in *Structural Analysis of Discrete Data*, ed. by Charles Manski and Daniel McFadden. Cambridge, MA: MIT Press.

[23] Heckman, J.J. (1981b): "Heterogeneity and State Dependence," in *Structural Analysis of Discrete Data*, ed. by Charles Manski and Daniel McFadden. Cambridge, MA: MIT Press.

[24] Heckman, J.J. and B. Singer (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, 271-320.

[25] Hotz, J.V., Kydland, F.E. and G.L. Sedlacek (1988): "Intertemporal Preferences and Labor Supply," *Econometrica*, 56, 335-360.

[26] Hyslop, D. (1999): "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women," *Econometrica*, 67, 1255-1294.

[27] Idler, E.L. and S.V. Kasl (1995): "Self-Ratings of Health: Do They Also Predict Changes in Functional Ability?" *Journal of Gerontology*, 50, S344-S353.

[28] Kaplan, G.A. and T. Camacho (1983): "Perceived Health and Mortality: A 9 Year Follow-up of the Human Population Laboratory Cohort," *American Journal of Epidemiology*, 177, 292.

[29] Leroux, B.G. (1992): "Consistent Estimation of a Mixing Distribution," *Annals of Statistics*, 20, 1350-1360.

[30] Lillard, E.L., and R. Willis (1978): "Dynamic Aspects of Earnings Mobility," *Econometrica*, 46, 985-1012.

[31] Lindsay, B. (1983): "The Geometry of Mixture Likelihoods, Part I" *Annals of Statistics*, 11, 86-94.

[32] Magnac, T. (2000): "Subsidised Training and Youth Employment: Distinguishing Unobserved Heterogeneity from State Dependence in Labour Market Histories," *Economic Journal*, 110, 805-837.

[33] Meghir, C. and L. Pistaferri (2004): "Income Variance Dynamics and Heterogeneity," *Econometrica*, 72, 1-32.

[34] Mossey, J.M. and E. Shapiro (1982): "Self-Rated Health: A Predictor of Mortality Among the Elderly," *American Journal of Public Health*, 71, 100.

[35] Neyman, J. and E. Scott (1948): "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1-32.

[36] Rust, J. and C. Phelan (1997): "How Social Security and Medicare Affect Retirement Behavior In a World of Incomplete Markets," *Econometrica*, 65, 781-831.

[37] Smith, J. (1999): "Healthy Bodies and Thick Wallets: The Dual Relation between Health and Economic Status," *Journal of Economic Perspectives*, 13, 145-166.

[38] Smith, J. (2004): "Health and SES Over the Life-Course," unpublished manuscript, RAND.

[39] Wooldridge, J. (2000): *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.

[40] Wooldrdige, J. (2005): "Simple Solutions to the Initial Conditions Problem for Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity," forthcoming *Journal of Applied Econometrics.*
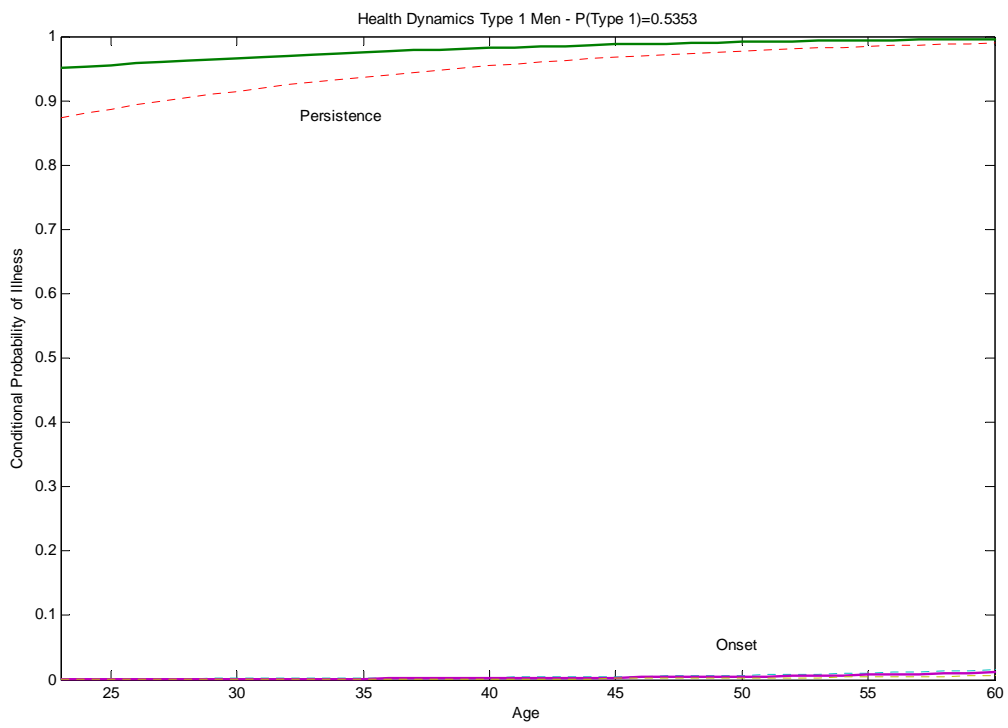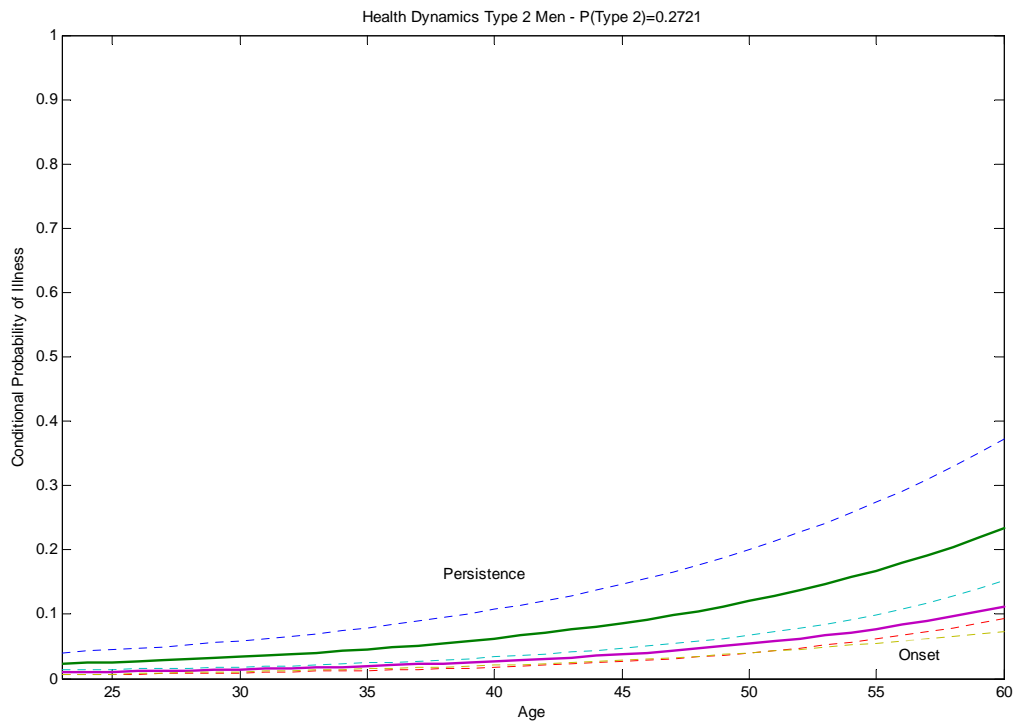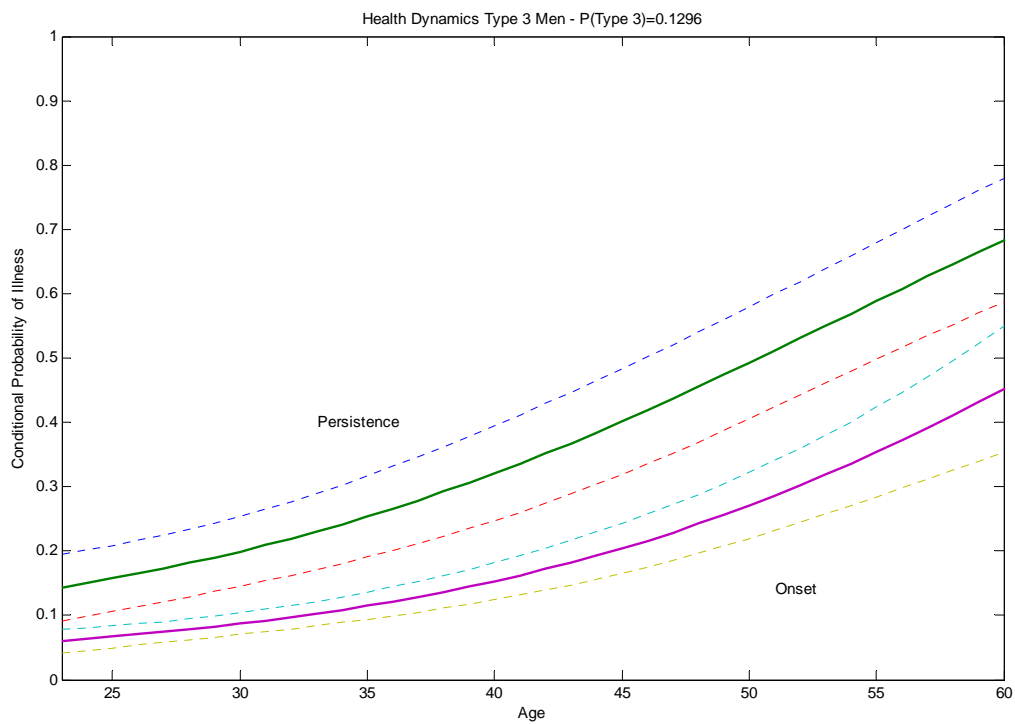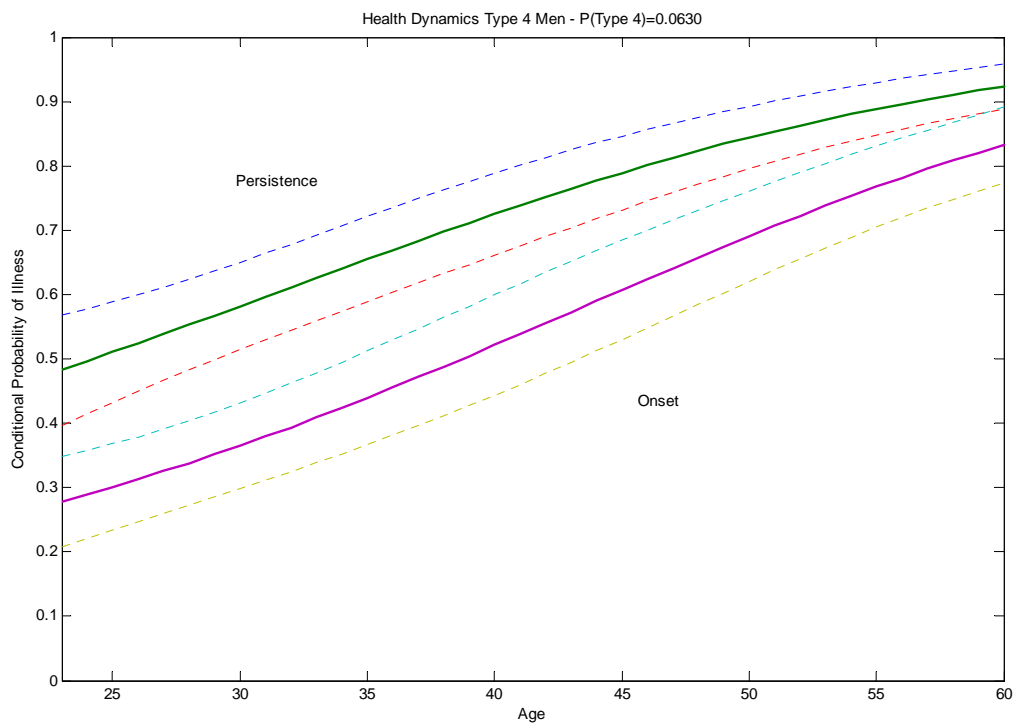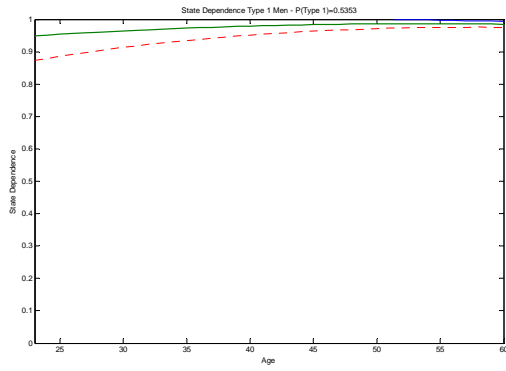
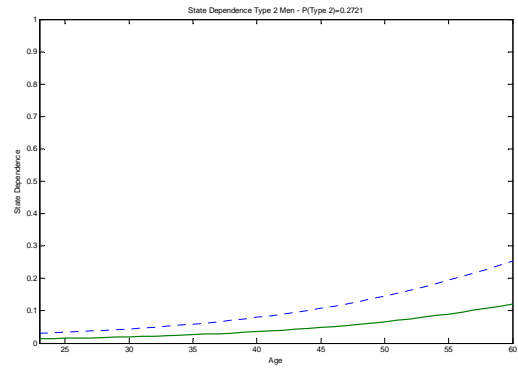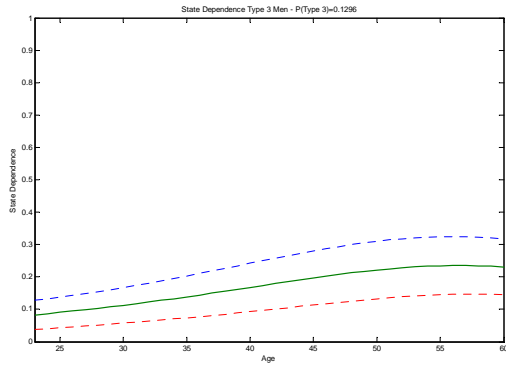Figure 1

Figure 2

Figure 3

Figure 4

27

State Dependence Type 1 Men - P(Type 1)=0.5353

Figure 5

State Dependence Type 2 Men - P(Type 2)=0.2721

Figure 6

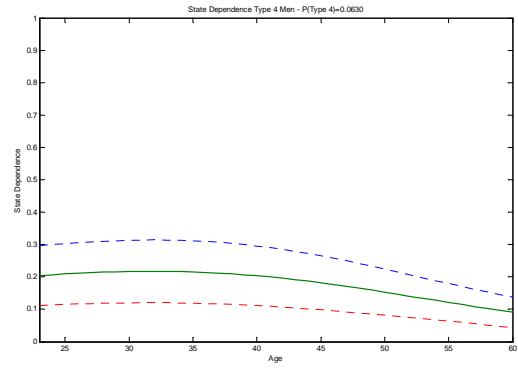State Dependence Type 3 Men - P(Type 3)=0.1296

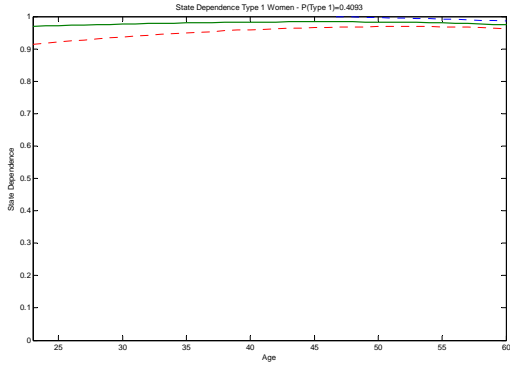Figure 7

State Dependence Type 4 Men - P(Type 4)=0.0630

Figure 8

Figure 9
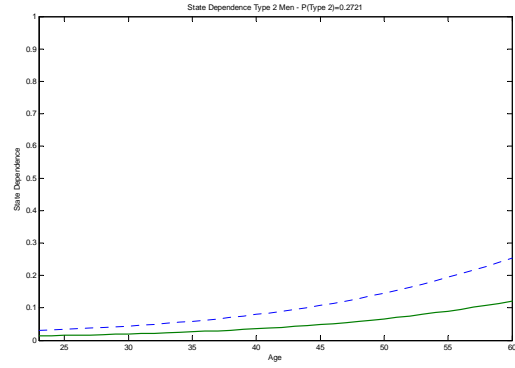

Figure 10
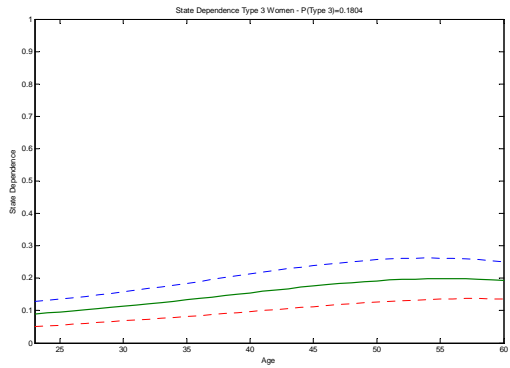
Figure 11
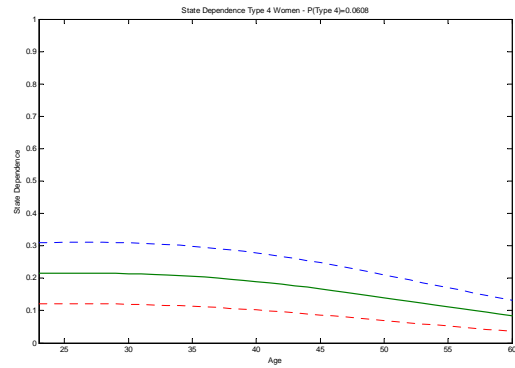

Figure 12

Table 1: Descriptive Statistics

Women

	Mean	25% Quantile	75% Quantile	Standard Deviation
SRHS (5-Point)	2.22	1	3	0.99
SRHS (2-Point)	0.10	0	0	0.30
Age	39.10	31	46	9.82
Panel Duration*	8.21	4	14	4.45
$N = 4186^{**}$				

Men

	Mean	25% Quantile	75% Quantile	Standard Deviation
SRHS (5-Point)	2.10	1	3	0.98
SRHS (2-Point)	0.08	0	0	0.27
Age	39.34	32	46	9.56
Panel Duration*	8.44	4	14	4.46
$N = 3923^{**}$				

*Panel duration refers to the length of time that the individual was in the panel.

**N is the number of individual observations, not individual-time observations.

Table 2: AIC for Index Selection

	$A = 2$		$A = 3$	
	Men	Women	Men	Women
Linear Model	6163.8	7564.5	6084.9	7453.9
Homogeneous Quadratic Model	6163.1	7563.2	6083.5	7452.9
Heterogeneous Quadratic Model	6163.9	7564.7	6085.6	7450.9

Table 3: AIC for Selection of the Number of Support Points

Points of Support	Men	Women
$A = 1$	11798.0	13631.0
$A = 2$	6163.1	7563.2
$A = 3$	6083.5	7452.9
$A = 4$	6062.9	7422.3

The homogeneous quadratic model was employed in the estimation.

Table 4: Parameter Estimates for Preferred Model - Men

	Type 1	Type 2	Type 3	Type 4
α_a	-8.0789	-5.6632	-3.7868	-1.9916
	(0.7920)	(0.7537)	(0.7400)	(0.7238)
γ_a	9.9901	0.8776	0.9597	0.8335
	(0.9093)	(0.3874)	(0.1915)	(0.2075)
ρ^1	0.3598	0.3598	0.3598	0.3598
	(0.3657)	(0.3657)	(0.3657)	(0.3657)
ρ^2	3.9879	3.9879	3.9879	3.9879
	(4.3498)	(4.3498)	(4.3498)	(4.3498)
p_a	0.9907	0.9578	0.8757	0.9999
	(0.0055)	(0.0655)	(0.1387)	(0.0001)
π_a	0.5353	0.2721	0.1296	0.0630

Table 5: Parameter Estimates for Preferred Model - Women

	Type 1	Type 2	Type 3	Type 4
α_a	-6.8666	-5.5826	-3.1537	-1.4090
	(0.7049)	(0.6721)	(0.6584)	(0.6586)
γ_a	9.5925	0.7514	0.8067	0.8779
	(1.0865)	(0.5424)	(0.1340)	(0.2033)
ρ^1	0.2494	0.2494	0.2494	0.2494
	(0.3267)	(0.3267)	(0.3267)	(0.3267)
ρ^2	4.4630	4.4630	4.4630	4.4630
	(3.8702)	(3.8702)	(3.8702)	(3.8702)
p_a	0.9997	0.9587	0.8874	0.7432
	(0.0054)	(0.0285)	(0.0868)	(0.2245)
π_a	0.4093	0.3495	0.1804	0.0608