Online Appendix for: "Growing Old in Rural America: Measuring Late-life Health and Economic Well-being"

Contents

List of Tables

List of Figures

A Forecasting model

In this appendix we briefly detail the forecasting model proposed by Miller and Bairoliya (2023) and utilized in this study.

A.1 Panel VAR representation

Although the model can accommodate multiple lags, focusing on the $VAR(1)$ reveals its core aspects. Denoting Y_{it} as the outcome vector for individual i at time t, including log consumption (c), self-rated health (s), retirement indicator (r), cube root of wealth (w), and a set of $n = 9$ morbidity states represented by M, the outcomes follow the structural $VAR(1)$ model:

$$AY_{it} = BY_{it-1} + \varepsilon_{it}. \tag{1}$$

Here, ε comprises independent and identically distributed (iid) shocks with a mean of zero, and matrix A has diagonal elements normalized to one. Model parameters remain consistent across individuals and over time, implying $A_{it} = A$ for all i and t. Estimation occurs in five distinct "blocks" of outcomes: morbidity, self-rated health, labor supply, consumption, and wealth. The unrestricted model is expressed in block matrix form as:

$$
\begin{array}{r}
\\ n \\ \\ \\ 4 \\ \\
\end{array}
\begin{array}{c}
\begin{array}{c|cccc} \overbrace{\phantom{-A_{11}}}^{n} & \overbrace{\phantom{-A_{12}\ -A_{13}\ -A_{14}\ -A_{15}}}^{4} \end{array} \\
\left[\begin{array}{c|cccc}
-A_{11} & -A_{12} & -A_{13} & -A_{14} & -A_{15} \\
\hline
-A_{21} & 1 & -a_{23} & -a_{24} & -a_{25} \\
-A_{31} & -a_{32} & 1 & -a_{34} & -a_{35} \\
-A_{41} & -a_{42} & -a_{43} & 1 & -a_{45} \\
-A_{51} & -a_{52} & -a_{53} & -a_{54} & 1
\end{array}\right]
\end{array}
\left[\begin{array}{c}
M_{it} \\ \hline s_{it} \\ r_{it} \\ c_{it} \\ w_{it}
\end{array}\right]
=
\begin{array}{c}
\begin{array}{c|cccc} \overbrace{\phantom{B_{11}}}^{n} & \overbrace{\phantom{B_{12}\ B_{13}\ B_{14}\ B_{15}}}^{4} \end{array} \\
\left[\begin{array}{c|cccc}
B_{11} & B_{12} & B_{13} & B_{14} & B_{15} \\
\hline
B_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\
B_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\
B_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\
B_{51} & b_{52} & b_{53} & b_{54} & b_{55}
\end{array}\right]
\end{array}
\left[\begin{array}{c}
M_{it-1} \\ \hline s_{it-1} \\ r_{it-1} \\ c_{it-1} \\ w_{it-1}
\end{array}\right]
+
\left[\begin{array}{c}
\epsilon_{1,it} \\ \hline \epsilon_{2,it} \\ \epsilon_{3,it} \\ \epsilon_{4,it} \\ \epsilon_{5,it}
\end{array}\right],
$$

with $n \times n$ matrix A_{11} having main diagonal terms scaled to one. Our proposed causal pathways establish a block recursive system, assuming that causal relationships start with morbidities, then affect self-rated health, retirement, consumption, and ultimately, wealth.[1] Health outcomes and retirement influence all future outcomes through lagged effects, while lagged consumption affects future wealth. We assume that consumption and wealth do not have lagged effects otherwise.[2] This system triangulation eliminates simultaneity across blocks, facilitating block-by-block estimation.

A.2 Exogenous characteristics

A vector of fixed external individual traits, denoted as X_{it}, is integrated into the model as exogenous predictors of the evolution of outcomes. The $VAR(1)$ model with exogenous regressors, can be expressed as follows:

$$AY_{it} = BY_{it-1} + CX_{it} + \varepsilon_{it}. \tag{2}$$

[1]This assumption is represented in the $VAR(1)$ model by setting $A_{12} = A_{13} = A_{14} = A_{15} = 0$ in the morbidity block, $a_{23} = a_{24} = a_{25} = 0$ in the self-rated health block, $a_{34} = a_{35} = 0$ in the retirement block, and $a_{45} = 0$ in the consumption block.

[2]i.e., $B_{14} = B_{15} = b_{24} = b_{25} = b_{34} = b_{35} = b_{45} = 0$.

The vector X_{it} comprises dummy variables for age, education, gender, race, urban[3], census division, census occupation code, birth cohort, a post-2008 indicator to account for the great recession, and a linear trend for the calendar year. Further included is a time-invariant individual fixed effect in both the consumption equation (π^c) and the wealth equation (π^w). The unobserved individual effects help maintain the proper variance in consumption and wealth over time by functioning as a person-specific drift in the autoregressive process. Time-invariant exogenous regressors are excluded from the consumption and wealth equations due to collinearity with the fixed effect. The exogenous effects can be written:

$$CX_{it} = \begin{array}{r} n\{ \\ \\ \\ \\ \\ \end{array} \underbrace{\left[\begin{array}{cccccccccccc} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} & C_{17} & C_{18} & C_{19} & C_{110} & 0 & 0 \\ \hdashline c_{21} & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} & c_{27} & c_{28} & c_{29} & c_{210} & 0 & 0 \\ c_{31} & c_{32} & c_{33} & c_{34} & c_{35} & c_{36} & c_{37} & c_{38} & c_{39} & c_{340} & 0 & 0 \\ c_{41} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{49} & c_{410} & 1 & 0 \\ c_{51} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{59} & c_{510} & 0 & 1 \end{array}\right]}_{(n+4)\times k} \underbrace{\begin{bmatrix} Age_{it} \\ Education_i \\ Gender_i \\ Race_i \\ Urban_i \\ Division_i \\ Occupation_i \\ Cohort_i \\ Year_t \\ Post_t \\ \pi_i^c \\ \pi_i^w \end{bmatrix}}_{k\times 1}.$$

A.3 Consumption

The explicit equation for forecasting consumption, as provided in system (2), is as follows:

$$\begin{aligned} c_{it} = {} & A_{41}M_{it} + a_{42}s_{it} + a_{43}r_{it} + B_{41}M_{it-1} + b_{42}s_{it-1} + b_{43}r_{it-1} \\ & + b_{44}c_{it-1} + c_{41}Age_{it} + c_{49}Year_t + c_{410}Post_t + \pi_i^c + \varepsilon_{4,it}. \end{aligned} \tag{3}$$

Equation (3) is a standard linear dynamic panel data model that incorporates a lagged dependent variable and individual-level fixed effects (π_i^c). Given the block recursive system, equation (3) can be estimated independently of other blocks, while ensuring that all structural parameters are identified, including the variance of $\varepsilon_{4,it}$.

A.4 Wealth

Similar to the equation for consumption forecasting given in equation (3), the equation for wealth can be written as:

$$\begin{aligned} w_{it} = {} & A_{51}M_{it} + a_{52}s_{it} + a_{53}r_{it} + a_{54}c_{it} + B_{51}M_{it-1} + b_{52}s_{it-1} + b_{53}r_{it-1} \\ & + b_{54}c_{it-1} + b_{55}w_{it-1} + c_{51}Age_{it} + c_{59}Year_t + c_{510}Post_t + \pi_i^w + \varepsilon_{5,it}. \end{aligned} \tag{4}$$

[3]In our forecasting model we split urbanicity into urban (Beale code 1), suburban (Beale code 2), and rural (Beale code >2) to add precision to the forecast.

Just like equation (3), the wealth equation is a linear dynamic panel data model with a lagged dependent variable and individual-level fixed effects (π_i^w). This equation can also be estimated independently of other blocks.

A.5 Retirement

Retirement is considered as a binary outcome with a continuous latent variable, designated as r^*, forming the basis for the observed outcome. To be precise, r_{it} is defined as follows:

$$r_{it} = 0 \quad if \quad r_{it}^* \leq 0$$
$$r_{it} = 1 \quad if \quad r_{it}^* > 0.$$

Assuming that the individual was employed during the previous period (with $b_{33} = 0$), the retirement model, as described in system (1), can be expressed as follows:

$$r_{it}^* \quad = \quad A_{31}M_{it} + a_{32}s_{it} + B_{31}M_{it-1} + b_{32}s_{it-1} + [c_{31}, \ldots, c_{310}]X_{it} + \varepsilon_{3,it}. \tag{5}$$

In equation (5), retirement is influenced by both the current and past values of self-rated health, specific health conditions, and exogenous individual characteristics. It is assumed that ε_3 follows an iid shock with a standard normal distribution, indicating that the retirement model has a standard probit structure.

A.6 Self-rated health

Self-rated health is measured on a five-point scale, ranging from poor (one) to excellent (five) in the HRS. Consequently, it is assumed that a continuous latent variable, denoted as s^*, underlies the observed outcome. The self-rated health model, as defined in system (2), can be expressed as follows:

$$s_{it}^* \quad = \quad A_{21}M_{it} + B_{21}M_{it-1} + b_{22}s_{it-1} + b_{23}r_{it-1} + [c_{21}, \ldots, c_{210}]X_{it} + \varepsilon_{2,it}. \tag{6}$$

The observed health state is defined by the following equation:

$$s_{it} = \delta \;\; if \; \kappa_{\delta-1} < s_{it}^* < \kappa_\delta \;\; for \; \delta = 1, \ldots, 5.$$

Here, when $\delta = 1$, it signifies the poorest health state (poor), while $\delta = 5$, it signifies the best health state (excellent). To account for the persistence of general health shocks over the life-course, it is assumed that latent self-rated health is influenced by the prior observed self-rated health category. Additionally, it is assumed that ε_2 is iid with a standard normal distribution, resulting in a standard ordered probit model.

A.7 Morbidities

In contrast to consumption, wealth, retirement, and self-rated health, the block triangulation within the system does not facilitate the direct identification of the structural parameters within the morbidity block, mainly due to the presence of nine distinct outcomes. Instead, the morbidity block is estimated using a reduced-form VAR. The reduced-form system is derived by pre-multiplying the structural system block with the inverse of matrix A_{11} as follows:

$$M_{it}^* = -A_{11}^{-1}B_{11}M_{it-1} - A_{11}^{-1}B_{12}s_{it-1} - A_{11}^{-1}B_{13}r_{it-1} - A_{11}^{-1}\left[C_{11},\ldots,C_{110}\right]X_{it} - A_{11}^{-1}\varepsilon_{1,it}.$$

Denoting $-A_{11}^{-1}B_{1j} = \hat{B}_j$, $-A_{11}^{-1}\left[C_{11},\ldots,C_{110}\right] = \hat{C}$ and $-A_{11}^{-1}\varepsilon_{1,t} = e_t$ yields the following reduced form system:

$$M_{it}^* = \hat{B}_1 M_{it-1} + \hat{B}_2 s_{it-1} + \hat{B}_3 r_{it-1} + \hat{C}X_{it} + e_{it}.$$

All right-hand-side variables are predetermined at time t and morbidity states do not have a direct contemporaneous effect on each other in the reduced from VAR. However, the error terms e_t are combinations of structural shocks specific to each morbidity, and they may be correlated across different morbidity states. This possibility allows for contemporaneous correlation in the probability of experiencing different morbidity states. For instance, the onset of heart disease might be correlated with the onset of hypertension or stroke due to contemporaneous shocks that affect them simultaneously.

It is again assumed that there exists a continuous latent variable denoted as m^* underlying each observed morbidity state, which can be expressed as follows:

$$\begin{aligned} m_{j,it} &= 0 \quad if \quad m_{j,it}^* \leq 0 \\ m_{j,it} &= 1 \quad if \quad m_{j,it}^* > 0 \end{aligned}$$

This assumption allows us to estimate the morbidity block of equations using the following model:

$$\begin{bmatrix} m_{1,it}^* \\ \vdots \\ m_{n,it}^* \end{bmatrix} = \begin{bmatrix} \hat{b}_{11} & \cdots & \hat{b}_{1n} \\ \vdots & \ddots & \vdots \\ \hat{b}_{n1} & \cdots & \hat{b}_{nn} \end{bmatrix} \begin{bmatrix} m_{1,it-1} \\ \vdots \\ m_{n,it-1} \end{bmatrix} + \hat{B}_2 s_{it-1} + \hat{B}_3 r_{it-1} + \hat{C}X_t + \begin{bmatrix} e_{1,it} \\ \vdots \\ e_{n,it} \end{bmatrix}. \tag{7}$$

It is assumed that contemporaneous morbidity shocks follow a standard multivariate normal distribution with an $n \times n$ covariance matrix denoted as Σ. Thus the morbidity block of equations is in the form of a multivariate probit model.

A.8 Higher order lags

In order to avoid autocorrelation within the structural error terms of the model, it may be necessary to consider additional outcome lags. An extension of the VAR(1) model to higher orders is straightforward, as seen with the following VAR(2) version of our model:

$$AY_{it} = BY_{it-1} + DY_{it-2} + CX_{it} + \varepsilon_{it},$$

with the block matrix form of DY_{it-2} given by:

$$\left[\begin{array}{c:cccc} D_{11} & D_{12} & D_{13} & D_{14} & D_{15} \\ \hdashline D_{21} & d_{22} & d_{23} & d_{24} & d_{25} \\ D_{31} & d_{32} & d_{33} & d_{34} & d_{35} \\ D_{41} & d_{42} & d_{43} & d_{44} & d_{45} \\ D_{51} & d_{52} & d_{53} & d_{54} & d_{55} \end{array}\right] \left[\begin{array}{c} M_{it-2} \\ \hdashline s_{it-2} \\ r_{it-2} \\ c_{it-2} \\ w_{it-2} \end{array}\right].$$

A.9 Survival

As all other outcomes in the system depend on survival, mortality probabilities are estimated independently of the VAR system described in equation (2). The probability of survival to the following period of life is estimated as follows:

$$\psi_{it} = I\left(\sum_{k=1}^{K}\left[\gamma_k^M M_{it-k} + \gamma_k^s s_{it-k} + \gamma_k^r r_{it-k}\right] + \delta X_{it} + u_{it} > 0\right) \tag{8}$$

where $\psi = 1$ represents survival, X is the vector of individual characteristics previously defined, and u_{it} is an iid random shock following a standard normal distribution. This general specification allows for the influence of K lags of morbidity states, self-rated health, and retirement on the probability of survival.

A.10 Estimation

We estimate the forecasting model using a pooled sample of individuals born before 1966, all of whom were aged fifty or older at the time of the survey. This gives us 40,973 unique individuals and a total of 269,299 individual-year observations. Following the biennial structure of the HRS, a model period corresponds to two calendar years and individuals are grouped in two-year age intervals. Table 1 presents descriptive statistics for each cohort in the HRS. The full set of estimation results are shown in Tables 2-4.

Since there is no simultaneity across blocks in the system, the model is estimated one block at a time. Both the consumption equation (3) and the wealth equation (4) are standard single-equation linear dynamic panel data models with lagged dependent variables and individual-level fixed effects.[4] To ensure that shocks are serially uncorrelated in the consumption equation, we include a one period lag of health and retirement and two lags of consumption itself. We use the same lags in the wealth equation with the addition of two lags of wealth itself. Similarly, we incorporate two lags of outcomes in all retirement, health, and survival equations and set $K = 2$ in the survival model. The ordered probit model for the self-rated health equation (6) is estimated independently of other VAR blocks using maximum likelihood. The retirement equation (5) and the survival equation (8) are estimated independently using standard probit regressions. Finally, we estimate the multivariate probit morbidity block via a series of bivariate probit estimators recommended by Mullahy (2016) because of the large number of outcomes and observations in the HRS.[5]

A.11 Simulations

After estimating the parameters of the forecasting model, outcomes paths are simulated from age sixty onward for individuals in the HRS. Simulations require data at age fifty-eight and sixty as "initial" conditions. This leaves five cohorts with required data for simulations: the EHRS, LHRS,

[4] We apply the bootstrap-based method of Everaert and Pozzi (2007) to correct for the Nickell (1981) bias that arises from OLS estimates of such models.

[5] See Miller and Bairoliya (2023) on issues around quasi-complete separation in this model given that morbidities are abosorbing states. We follow their approach by constraining the infinite coefficients to large values in the bivariate probit model to work around this issue.

War Babies, and early and mid Baby Boomers. Using age sixty data as initial ($t = 0$) conditions[6], remaining life outcomes for each individual i are simulated as follows:

1. Draw a survival shock, denoted as u_{i1}. Using equation (8), determine an individual's survival up to time $t = 1$ (corresponding to age 62). If the individual survives, they proceed to the next step.

2. Draw the morbidity shock vector, denoted as e_{i1}, from a standard multivariate normal distribution with the estimated covariance matrix Σ. This shock vector, in conjunction with the model described in equation (7), is used to compute the simulated morbidity vector at age 62, represented as M_{i1}.

3. Given the age 62 morbidities (M_{i1}), draw the general health shock $\varepsilon_{2,i1}$ and compute the age 62 self-rated health (s_{i1}) using equation (6).

4. Given age 62 self-rated health (s_{i1}) and morbidities (M_{i1}), draw the retirement shock $\varepsilon_{3,i1}$ and determine age 62 retirement (r_{i1}) using equation (5).

5. Given age 62 retirement (r_{i1}), self-rated health (s_{i1}), and morbidities (M_{i1}), draw consumption shock $\varepsilon_{4,i1}$ to determine age 62 consumption (c_{i1}) using equation (3).

6. Given all age 62 consumption (c_{i1}), retirement (r_{i1}), self-rated health (s_{i1}), and morbidities (M_{i1}), draw wealth shock $\varepsilon_{5,i1}$ to compute age 62 wealth (w_{i1}) using equation (4).

7. Steps 1-6 are repeated for $t = 2, 3, \ldots$ until death or $t = 30$ (age 120).

The above process is repeated 5,000 times for each individual in the simulation sample.

Figures 1-4 display a comparison between average simulated life-cycle profiles and those derived from HRS data by urbanicity for the EHRS cohort. The simulations closely align with the aggregated data, suggesting that the forecasting model provides a reasonable approximation of the underlying data generation processes. Notably, the data and simulations coincide at age sixty by design. However, even up to 26 years later, as the EHRS cohort hits age 86, the simulations exhibit a strong match with the data.

To further illustrate the precision of the model, we compare consumption and health utility means and standard deviations of the data with simulated life-cycle profiles for each birth cohort in Figures 5-6. The simulations closely mirror the data across birth cohorts, underscoring the efficacy of the VAR approach in accurately forecasting joint dynamics.

A.12 Weighting

The HRS is designed to represent the non-institutionalized US population aged fifty and above, providing respondent-level analysis weights for each wave. For aggregate welfare calculations, we utilized base year weights from specific waves when cohorts were around sixty years old—1996 weights for the EHRS, 2000 for the LHRS, 2006 for War Babies, 2012 for early Baby Boomers, and 2018 for mid Baby Boomers. Simulations were not possible for respondents with missing

[6]Initial conditions also include unobserved consumption and wealth endowments $\hat{\pi}$ estimated using the prediction method of De Vos et al. (2015).

data at ages 58-59 or 60-61 given (lagged) data requirements. For instance, if a member of the EHRS cohort was interviewed at age 60 in 1996 but was missing from the 1994 wave, they would be excluded from the simulation sample while remaining part of the 1996 nationally representative sample. Table 5 provides a comparison between the weighted representative sample and the weighted sample used in our simulations after excluding these missing cases. The simulation sample showed slightly higher proportions of females, educated individuals, whites, and urban residents compared to the representative sample, although the differences were small and generally consistent across all cohorts.

A.13 Figures and tables

Table 1: Estimation sample descriptive statistics by cohort

	AHEAD	CODA	EHRS	LHRS	WB	EBB	MBB	LBB
Individuals	7,758	4,242	5,371	5,135	3,655	4,834	5,185	4,793
Observations	37,372	29,693	48,886	50,213	32,786	31,579	25,594	13,176
Age (mean)	81.82	75.78	68.54	63.92	61.96	60.13	57.02	54.40
Hypertension (%)	54.71	57.76	54.56	52.31	52.10	52.61	50.48	48.73
Diabetes (%)	15.47	19.09	19.80	19.15	20.08	22.57	22.17	22.56
Cancer (%)	16.73	17.83	14.67	11.78	11.85	9.46	8.44	7.44
Lung disease (%)	9.44	10.25	9.82	8.92	7.85	7.55	8.21	8.42
Heart disease (%)	35.32	31.51	24.15	20.45	18.46	16.29	13.89	12.20
Stroke (%)	15.32	12.34	7.81	6.57	6.38	5.52	4.98	4.82
Psyche problem (%)	11.81	11.97	11.33	13.18	17.56	20.04	20.57	22.20
Arthritis (%)	55.98	60.71	58.18	54.10	54.07	49.44	42.86	37.61
Difficulty with ADLs (%)	40.58	29.79	25.35	23.03	23.54	23.34	21.61	18.81
Self-rated health (%)								
Poor	14.25	10.38	9.27	7.82	6.56	7.45	6.86	7.11
Fair	25.73	22.02	19.57	19.13	17.37	19.96	21.55	22.65
Good	30.86	32.16	31.86	31.21	31.39	31.15	32.23	32.13
Very good	21.39	26.39	28.02	28.85	31.82	30.17	29.33	28.06
Excellent	7.76	9.07	11.28	12.99	12.87	11.27	10.03	10.05
Retired (%)	95.50	91.82	78.44	67.19	63.39	56.56	48.68	40.00
Annual consumption ($1000s, mean)	22.31	24.64	24.76	25.92	26.41	23.27	20.01	19.15
Male (%)	37.52	46.27	44.81	45.20	37.74	42.36	42.58	41.58
Education (%)								
<HS	41.60	32.17	30.80	27.96	21.04	20.12	21.78	21.38
HS	29.65	31.48	32.69	32.82	30.77	24.33	24.64	23.21
Some college	16.36	17.88	18.69	20.63	24.45	28.45	29.44	29.53
College	12.39	18.47	17.83	18.59	23.74	27.10	24.13	25.88
Race (%)								
White	84.96	86.92	80.37	79.93	80.07	66.76	59.83	52.49
Black	12.93	9.64	16.25	15.93	14.91	22.03	26.49	27.79
Other	2.11	3.44	3.38	4.14	5.02	11.21	13.67	19.73
Location (%)								
Urban	43.78	43.06	42.70	43.47	43.12	48.71	55.51	58.11
Suburban	31.31	30.85	28.89	28.31	27.74	28.11	25.72	24.20
Rural	24.91	26.09	28.41	28.21	29.15	23.18	18.77	17.70

Notes: Children of the Depression denoted by CODA, War Babies by WB, early Baby Boomers by BB, mid Baby Boomers by MBB, and late Baby Boomers by LBB. Consumption is reported in real 2010 dollars. Source: HRS.

Table 2: Model estimates for ADLs, self-rated health, retirement, consumption, and mortality

	ADLs		Self-rated health		Retirement		Consumption		Mortality	
Variable	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	SE	SE
Hyper			-0.282	0.013	0.070	0.033	0.008	0.012	0.102	0.025
Diab			-0.263	0.016	0.062	0.042	-0.001	0.015	0.096	0.030
Cancer			-0.685	0.018	0.177	0.048	0.025	0.018	0.662	0.024
Lung			-0.466	0.021	0.172	0.064	-0.016	0.015	0.408	0.029
Heart			-0.481	0.014	0.109	0.042	-0.005	0.011	0.194	0.022
Stroke			-0.489	0.020	0.472	0.069	-0.065	0.019	0.240	0.027
Psych			-0.415	0.019	0.385	0.054	-0.035	0.020	0.216	0.027
Arthritis			-0.221	0.013	0.031	0.032	0.016	0.011	-0.026	0.023
ADL			-0.672	0.012	0.404	0.036	-0.042	0.014	0.346	0.018
Health 2					-0.561	0.043	0.041	0.014	-0.334	0.015
Health 3					-0.711	0.044	0.057	0.016	-0.530	0.017
Health 4					-0.728	0.046	0.084	0.016	-0.645	0.020
Health 5 (best)					-0.714	0.050	0.112	0.022	-0.644	0.030
Lag Hyper	0.033	0.029	0.159	0.018	-0.023	0.045	-0.008	0.009	-0.046	0.024
Lag Diab	0.095	0.035	0.104	0.023	-0.026	0.060	-0.005	0.012	0.061	0.031
Lag Cancer	0.046	0.040	0.531	0.026	-0.121	0.074	-0.007	0.018	-0.449	0.026
Lag Lung	0.197	0.044	0.214	0.030	0.040	0.096	0.001	0.019	-0.122	0.031
Lag Heart	0.068	0.031	0.289	0.020	-0.150	0.063	0.004	0.011	-0.034	0.023
Lag Stroke	0.377	0.043	0.369	0.029	-0.259	0.115	-0.005	0.017	-0.052	0.029
Lag Psych	0.327	0.040	0.233	0.027	-0.121	0.079	0.023	0.017	-0.133	0.029
Lag Arthritis	0.232	0.024	0.117	0.017	0.048	0.042	-0.005	0.012	-0.076	0.022
Lag ADL			0.335	0.017	-0.202	0.053	-0.005	0.013	-0.127	0.018
Lag Health 2	-0.244	0.028	0.629	0.013	-0.010	0.056	0.018	0.010	-0.061	0.017
Lag Health 3	-0.486	0.029	1.144	0.013	-0.052	0.057	0.018	0.013	-0.094	0.019
Lag Health 4	-0.658	0.031	1.687	0.014	-0.087	0.059	0.019	0.013	-0.133	0.021
Lag Health 5	-0.738	0.038	2.306	0.017	-0.082	0.062	0.021	0.014	-0.144	0.029
Time	-0.049	0.006	0.018	0.003	-0.002	0.009	0.006	0.009	-0.014	0.004
2008+	0.026	0.023	0.003	0.011	-0.070	0.030	-0.057	0.009	0.039	0.020
Suburban	0.001	0.015	0.000	0.007	0.007	0.019			0.001	0.014
Rural	0.007	0.015	-0.008	0.007	-0.011	0.018			0.011	0.013
CODA	0.094	0.030	0.020	0.015	0.062	0.075			-0.009	0.022
Early HRS	0.140	0.042	0.019	0.021	0.072	0.086			-0.044	0.031
Late HRS	0.156	0.053	0.011	0.026	-0.008	0.098			-0.061	0.040
War Babies	0.202	0.066	-0.000	0.032	0.033	0.114			-0.130	0.050
Early Boomers	0.295	0.080	-0.063	0.039	0.035	0.134			-0.155	0.060
Mid Boomers	0.342	0.095	-0.100	0.046	-0.052	0.152			-0.189	0.073
Late Boomers	0.418	0.113	-0.086	0.054	0.089	0.173			-0.298	0.096
Black	0.088	0.017	-0.053	0.008	0.045	0.021			0.051	0.014
Other	0.011	0.028	-0.089	0.012	-0.026	0.030			-0.089	0.026
Female	-0.003	0.013	0.035	0.006	0.119	0.016			-0.214	0.011
HS grad	-0.090	0.015	0.085	0.007	-0.031	0.021			0.029	0.013
Some college	-0.037	0.018	0.118	0.008	-0.052	0.023			0.021	0.015
College grad	-0.099	0.020	0.202	0.009	-0.070	0.025			-0.010	0.018
Retired							-0.045	0.012	0.190	0.030
Lag Retired	0.115	0.026	-0.021	0.012			-0.032	0.012	-0.014	0.026
Lag2 Retired	-0.011	0.025	-0.017	0.012						
Lag Con							0.164	0.005		
Lag2 Con							0.080	0.005		
Constant	-0.939	0.072			-0.909	0.176			-1.706	0.241

Notes: Dependent variable across columns. Multivariate probit results reported for ADLs as dependent outcome. Standard (ordered) probit results reported for self-rated health, mortality, and retirement as dependant outcomes. Linear dynamic panel estimates reported for consumption as outcome. All regressions also include dummies for age. Regressions for ADLs, self-rated health, mortality, and retirement also include dummies for occupation and census division. Regressions for ADLs and self-rated health also includes second lag for all health outcomes.

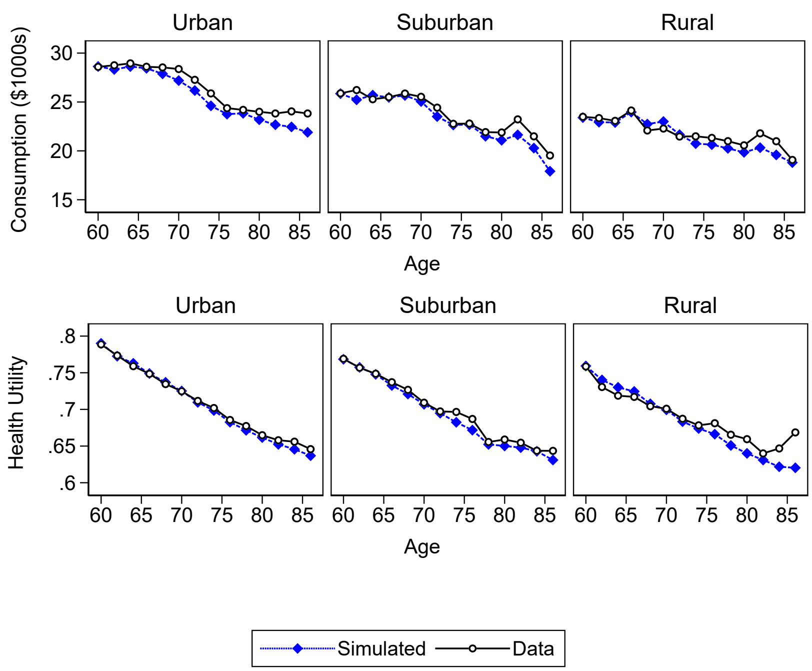
Table 3: Model estimates for morbidities

	Hypertension		Diabetes		Cancer		Lung disease		Heart disease		Stroke		Psych		Arthritis	
Variable	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	SE	SE
Lag Hyper			0.269	0.031	-0.036	0.036	0.073	0.037	0.146	0.030	0.102	0.038	0.121	0.036	0.075	0.031
Lag Diab	0.253	0.045			0.026	0.042	0.057	0.045	0.039	0.039	0.025	0.048	0.055	0.044	0.035	0.041
Lag Cancer	-0.061	0.050	0.002	0.049			0.043	0.054	-0.052	0.046	-0.022	0.053	-0.071	0.056	0.053	0.047
Lag Lung	0.121	0.054	0.079	0.051	0.124	0.053			0.261	0.046	0.034	0.056	0.133	0.056	0.197	0.058
Lag Heart	0.088	0.041	0.092	0.037	0.028	0.038	0.207	0.037			0.165	0.037	0.085	0.038	0.090	0.039
Lag Stroke	0.072	0.063	-0.047	0.058	0.020	0.053	-0.020	0.056	0.095	0.049			0.279	0.047	-0.071	0.058
Lag Psych	0.056	0.050	0.061	0.047	-0.072	0.054	0.119	0.050	0.086	0.044	0.119	0.050			0.266	0.049
Lag Arthritis	0.077	0.028	0.013	0.030	-0.034	0.032	0.144	0.034	0.060	0.028	0.005	0.034	0.110	0.033		
Lag ADL	0.058	0.031	0.028	0.032	0.002	0.032	0.098	0.033	0.069	0.028	0.184	0.030	0.224	0.030	0.163	0.034
Lag Health 2	0.025	0.033	-0.009	0.030	-0.048	0.031	-0.086	0.030	-0.106	0.027	-0.121	0.030	-0.174	0.028	-0.063	0.034
Lag Health 3	0.005	0.034	-0.026	0.032	-0.076	0.033	-0.163	0.032	-0.170	0.029	-0.214	0.032	-0.281	0.031	-0.102	0.035
Lag Health 4	-0.030	0.035	-0.107	0.034	-0.106	0.035	-0.310	0.036	-0.252	0.031	-0.249	0.036	-0.378	0.035	-0.145	0.037
Lag Health 5 (best)	-0.118	0.039	-0.228	0.042	-0.154	0.042	-0.429	0.050	-0.307	0.037	-0.334	0.046	-0.451	0.047	-0.245	0.041
Lag2 Hyper			0.013	0.030	0.076	0.036	-0.086	0.037	0.031	0.030	0.043	0.038	-0.063	0.035	0.026	0.031
Lag2 Diab	-0.088	0.048			-0.043	0.045	-0.136	0.047	0.101	0.041	0.096	0.050	-0.055	0.047	-0.025	0.043
Lag2 Cancer	0.053	0.053	-0.009	0.052			0.045	0.057	0.076	0.048	-0.004	0.057	0.088	0.059	0.000	0.051
Lag2 Lung	-0.182	0.059	-0.077	0.055	0.011	0.057			-0.100	0.051	0.013	0.061	-0.023	0.060	-0.111	0.065
Lag2 Heart	-0.055	0.043	-0.012	0.039	0.011	0.039	-0.094	0.039			-0.006	0.038	-0.066	0.040	-0.022	0.041
Lag2 Stroke	-0.055	0.070	0.053	0.063	-0.028	0.058	0.056	0.062	0.028	0.054			-0.193	0.053	0.067	0.064
Lag2 Psych	-0.030	0.053	-0.080	0.050	0.063	0.057	0.018	0.053	-0.029	0.046	-0.019	0.053			-0.118	0.053
Lag2 Arthre	-0.041	0.028	-0.019	0.030	0.072	0.032	-0.035	0.033	0.038	0.028	-0.013	0.034	-0.016	0.032		
Lag2 ADL	-0.059	0.033	0.039	0.033	-0.018	0.034	-0.025	0.034	0.022	0.029	-0.098	0.031	-0.069	0.031	-0.077	0.037
Lag2 Health 2	-0.016	0.034	-0.051	0.031	-0.055	0.033	-0.076	0.031	-0.004	0.029	-0.064	0.031	-0.073	0.031	0.030	0.037
Lag2 Health 3	-0.023	0.035	-0.051	0.032	-0.019	0.034	-0.114	0.034	-0.013	0.031	-0.055	0.034	-0.115	0.033	0.046	0.038
Lag2 Health 4	-0.036	0.037	-0.097	0.035	-0.005	0.036	-0.162	0.037	-0.033	0.033	-0.040	0.037	-0.196	0.037	0.018	0.040
Lag2 Health 5	-0.065	0.040	-0.140	0.041	0.008	0.042	-0.263	0.048	-0.077	0.038	-0.084	0.045	-0.264	0.046	-0.053	0.043
Time	0.018	0.007	0.005	0.007	-0.002	0.007	0.003	0.008	-0.017	0.006	-0.034	0.007	-0.010	0.007	-0.030	0.007
2008+	0.001	0.026	0.004	0.027	0.028	0.027	0.029	0.031	-0.006	0.024	0.045	0.030	-0.044	0.030	-0.000	0.026
Suburban	0.026	0.017	0.008	0.018	0.006	0.018	-0.021	0.022	0.018	0.016	-0.022	0.021	0.013	0.021	-0.004	0.017
Rural	-0.013	0.016	-0.015	0.017	-0.035	0.017	0.014	0.020	0.030	0.016	-0.009	0.020	0.048	0.020	0.021	0.017
CODA	-0.016	0.037	0.002	0.041	-0.024	0.037	0.016	0.043	-0.019	0.034	0.018	0.037	0.078	0.040	-0.079	0.037
Early HRS	-0.077	0.050	-0.027	0.054	-0.061	0.051	-0.034	0.058	0.009	0.046	0.001	0.051	0.047	0.055	-0.080	0.050
Late HRS	-0.065	0.063	-0.018	0.068	-0.086	0.065	-0.007	0.074	0.035	0.058	0.022	0.066	0.075	0.069	0.007	0.063
War Babies	-0.061	0.077	0.033	0.084	-0.044	0.081	-0.028	0.092	0.061	0.071	0.104	0.082	0.214	0.085	0.117	0.077
Boomers	-0.130	0.094	0.053	0.102	-0.108	0.099	-0.065	0.112	0.081	0.087	0.116	0.100	0.287	0.103	0.168	0.094
Mid Boomers	-0.227	0.110	0.082	0.119	-0.082	0.118	-0.026	0.133	0.146	0.103	0.164	0.119	0.280	0.121	0.172	0.109
Late Boomers	-0.262	0.129	0.123	0.140	-0.066	0.141	-0.010	0.158	0.148	0.123	0.319	0.146	0.297	0.143	0.280	0.127
Black	0.200	0.020	0.070	0.019	-0.042	0.020	-0.121	0.023	-0.124	0.018	0.049	0.022	-0.177	0.023	-0.007	0.019
Other	0.054	0.027	0.210	0.028	-0.161	0.036	-0.099	0.037	-0.110	0.030	-0.137	0.041	-0.057	0.034	-0.044	0.027
Female	0.008	0.014	-0.092	0.015	-0.191	0.016	-0.037	0.018	-0.176	0.014	-0.070	0.017	0.135	0.017	0.158	0.014
HS grad	-0.038	0.017	-0.080	0.018	0.008	0.019	-0.084	0.021	0.030	0.017	0.028	0.020	-0.058	0.020	-0.032	0.018
Some college	-0.090	0.020	-0.082	0.021	0.048	0.021	-0.065	0.024	0.055	0.019	0.043	0.024	0.004	0.023	-0.004	0.020
College grad	-0.123	0.022	-0.137	0.024	0.054	0.024	-0.185	0.029	-0.011	0.022	0.025	0.028	-0.013	0.027	-0.053	0.022
Lag Retired	-0.029	0.028	0.045	0.029	0.033	0.031	0.064	0.038	0.018	0.029	0.050	0.040	0.082	0.035	0.007	0.027
Lag2 Retired	0.017	0.027	-0.044	0.028	-0.004	0.030	0.011	0.035	-0.000	0.028	0.034	0.037	-0.017	0.033	-0.023	0.026
Constant	-1.498	0.081	-1.934	0.085	-1.918	0.086	-1.955	0.099	-1.688	0.078	-2.430	0.102	-1.792	0.090	-1.297	0.082

Notes: Multivarite probit results with dependent variable across columns. Regressions also include dummies for age, occupation, and census division.

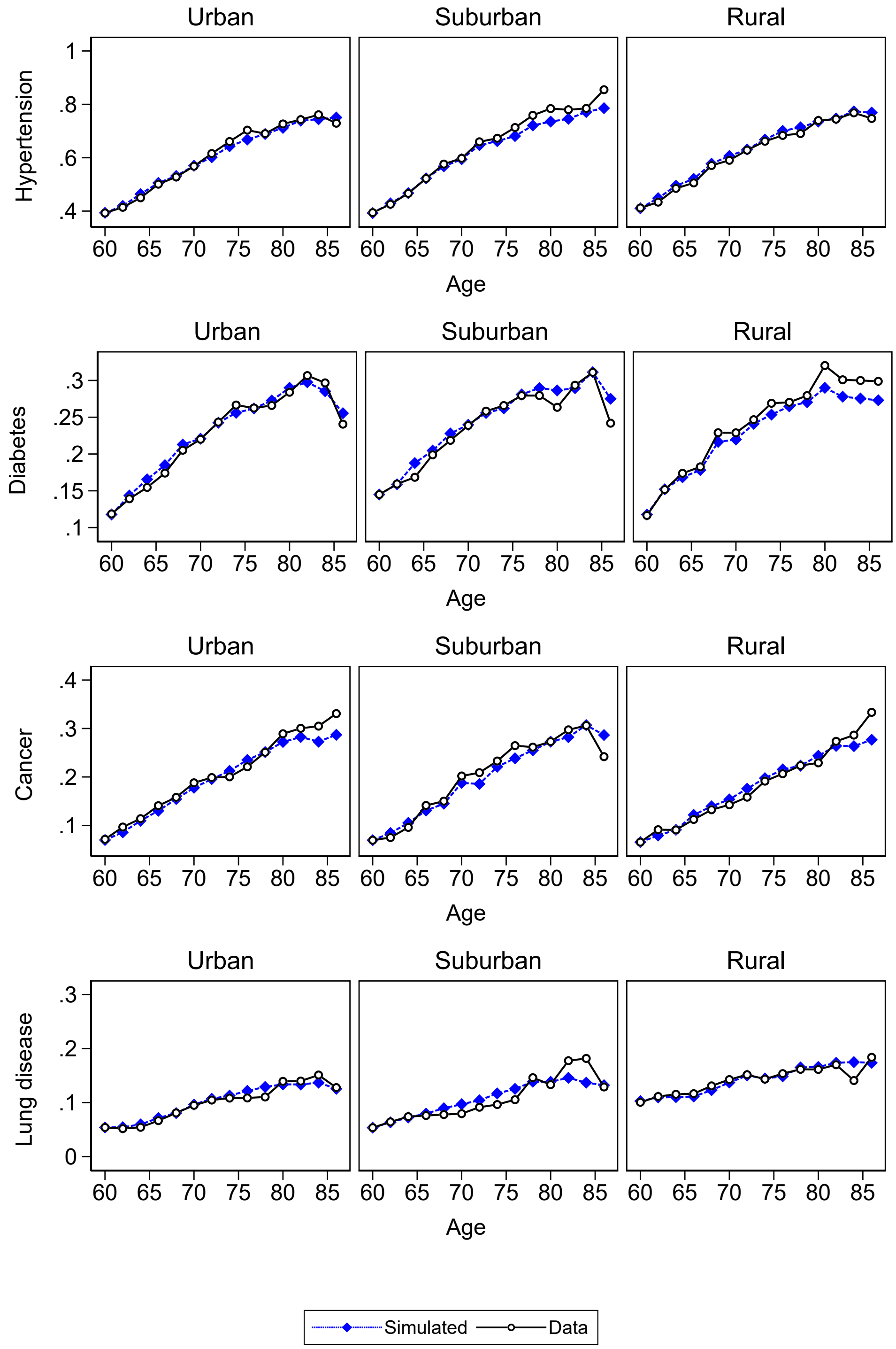
Table 4: Morbidity shock covariance matrix (Σ)

	Hyper	Diabetes	Cancer	Lung	Heart	Stroke	Psych	Arthritis	ADLs
Hyper	1.00	0.26	0.05	0.07	0.28	0.28	0.14	0.09	0.10
Diabetes	0.26	1.00	0.06	0.04	0.10	0.13	0.06	0.05	0.06
Cancer	0.05	0.06	1.00	0.13	0.05	0.06	0.11	0.05	0.13
Lung	0.07	0.04	0.13	1.00	0.23	0.10	0.17	0.09	0.18
Heart	0.28	0.10	0.05	0.23	1.00	0.28	0.16	0.10	0.14
Stroke	0.28	0.13	0.06	0.10	0.28	1.00	0.20	0.09	0.39
Psych	0.14	0.06	0.11	0.17	0.16	0.20	1.00	0.16	0.29
Arthritis	0.09	0.05	0.05	0.09	0.10	0.09	0.16	1.00	0.25
ADLs	0.10	0.06	0.13	0.18	0.14	0.39	0.29	0.25	1.00



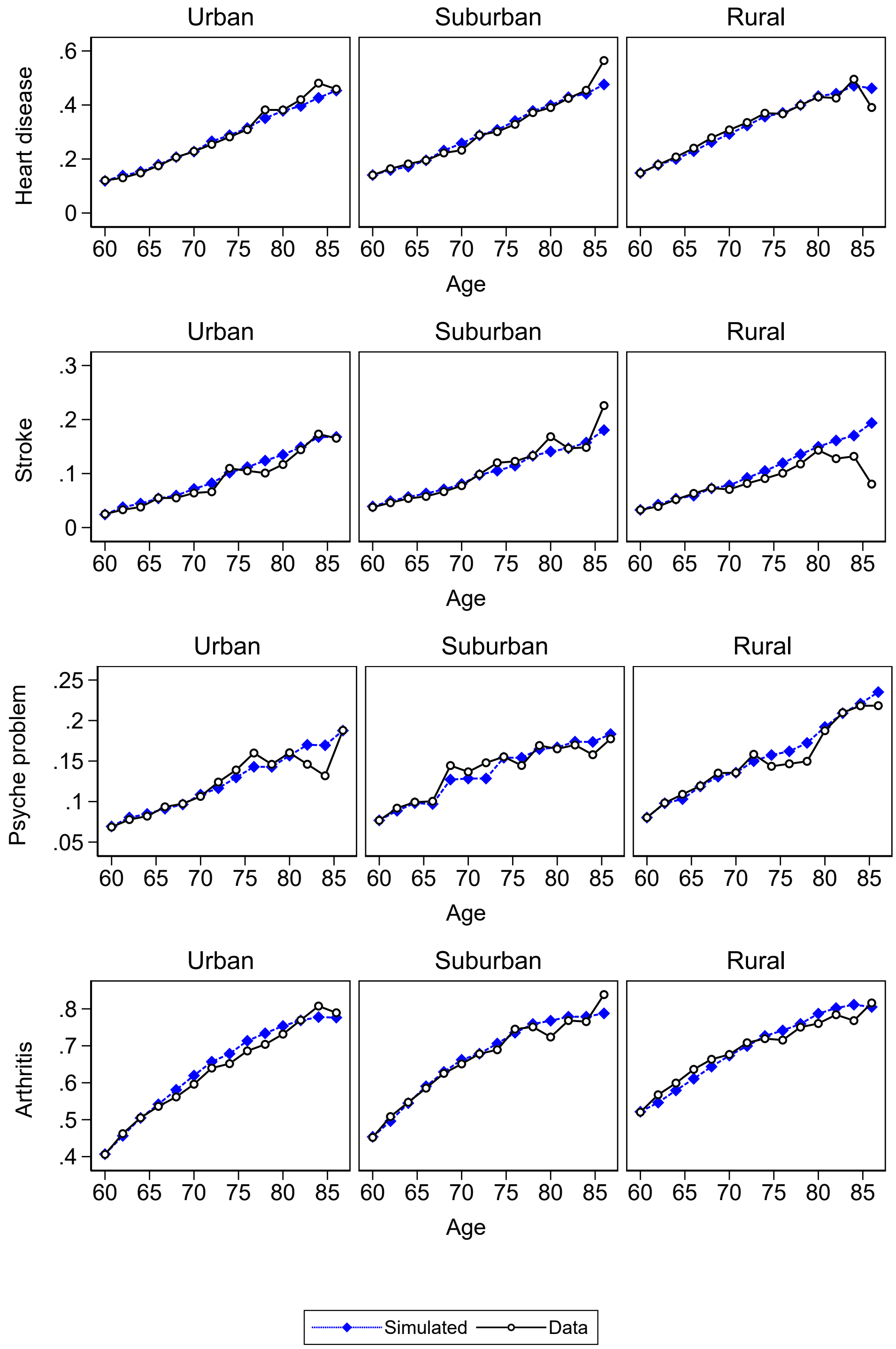
Notes: "Data" plots mean of all available data (inclusive of imputed missing values) in the EHRS cohort by two-year age interval. "Simulated" plots mean of expected simulated outcome for each observation in the data (i.e. the expected outcome for each person-year observation in the data).

Figure 1: Mean of life-cycle consumption and health utility profiles by location



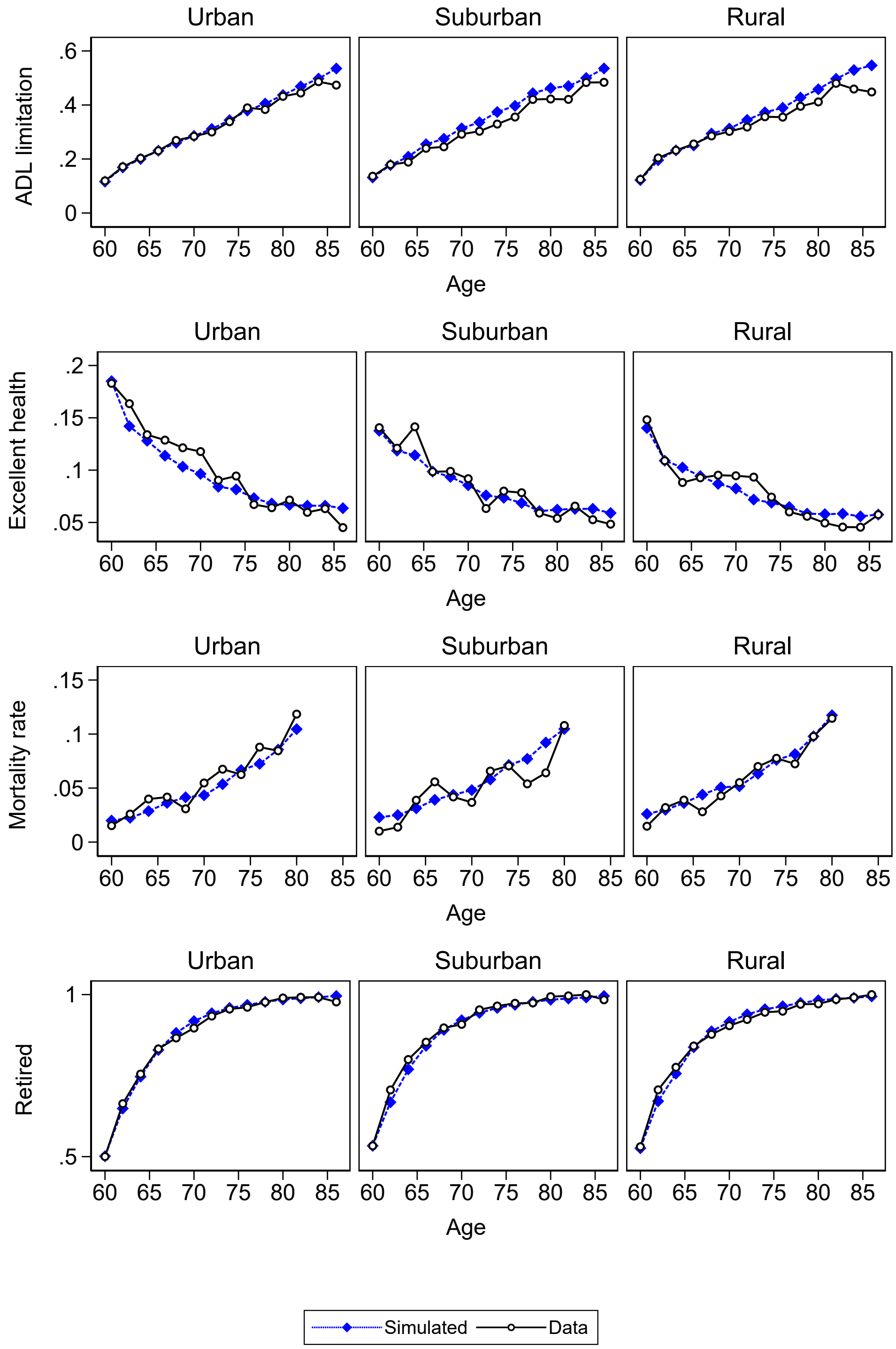
Notes: "Data" plots mean of all available data (inclusive of imputed missing values) in the EHRS cohort by two-year age interval. "Simulated" plots mean of expected simulated outcome for each observation in the data (i.e. the expected outcome for each person-year observation in the data).

Figure 2: Mean of life-cycle morbidity profiles by location



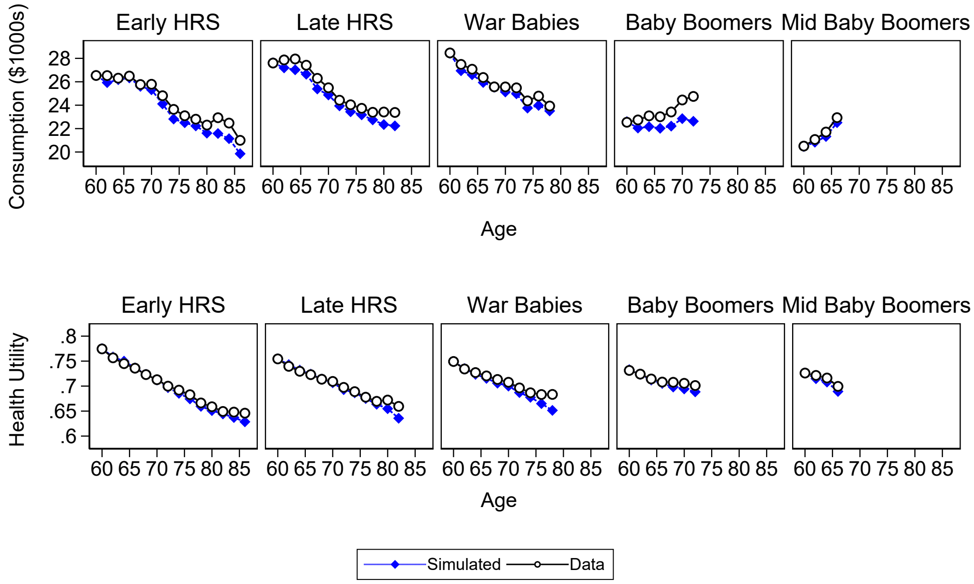
Notes: "Data" plots mean of all available data (inclusive of imputed missing values) in the EHRS cohort by two-year age interval. "Simulated" plots mean of expected simulated outcome for each observation in the data (i.e. the expected outcome for each person-year observation in the data).

Figure 3: Mean of life-cycle morbidity profiles by location



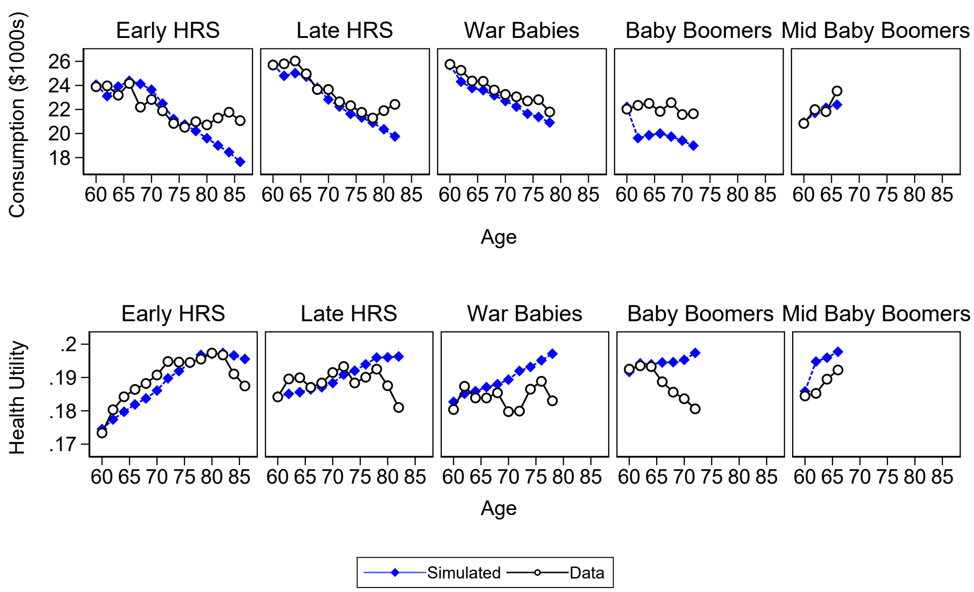
Notes: "Data" plots mean of all available data (inclusive of imputed missing values) in the EHRS cohort by two-year age interval. "Simulated" plots mean of expected simulated outcome for each observation in the data (i.e. the expected outcome for each person-year observation in the data).

Figure 4: Mean of life-cycle health, mortality, and retirement profiles by location



Notes: "Data" plots mean of all available data (inclusive of imputed missing values) in HRS by two-year age interval and cohort. "Simulated" plots mean of expected simulated outcome for each observation in the data (i.e. the expected outcome for each person-year observation in the data).

Figure 5: Mean of life-cycle consumption and health utility profiles by cohort



Notes: "Data" plots standard deviation of all available data (inclusive of imputed missing values) in HRS by two-year age interval and cohort. "Simulated" plots mean of standard deviations of simulated outcome (i.e. the mean of standard deviations calculated for each of the 5,000 simulation runs).

Figure 6: Standard deviation of consumption and health utility life-cycle profiles by cohort

Table 5: Representative and simulation sample comparison

	EHRS		LHRS		WB		EBB		MBB	
	Rep	Sim	Rep	Sim	Rep	Sim	Rep	Sim	Rep	Sim
	0	1	2	3	4	5	6	7	8	9
Individuals	3,158	3,093	3,808	3,599	2,699	2,575	3,874	3,260	3,405	3,182
Male (%)	47.17	46.32	46.93	46.71	47.77	47.82	47.71	47.14	47.67	47.67
Education (%)										
<HS	29.12	28.92	25.28	25.38	18.77	18.46	14.20	12.97	14.51	14.34
HS	33.56	33.75	32.05	32.30	30.43	30.32	23.27	23.48	22.19	22.19
Some college	19.30	19.25	21.64	21.49	24.31	24.40	29.05	28.77	30.37	30.05
College	18.01	18.07	21.03	20.82	26.48	26.82	33.48	34.79	32.93	33.42
Race (%)										
White	86.30	86.54	86.17	86.54	85.51	85.95	80.78	83.68	79.03	79.42
Black	10.38	10.26	9.98	9.96	9.68	9.22	11.10	9.05	12.11	11.96
Other	3.31	3.20	3.85	3.50	4.81	4.84	8.13	7.28	8.86	8.63
Location (%)										
Urban	44.83	45.17	43.99	44.21	42.13	41.77	47.78	47.23	50.24	50.23
Suburban	28.13	27.87	28.32	27.76	28.22	28.32	27.40	27.21	26.09	25.99
Rural	27.04	26.96	27.69	28.03	29.65	29.91	24.82	25.56	23.67	23.78

Notes: War Babies denoted by WB and Baby Boomers by BB. EHRS cohort inclues those under age 60 in 1992. "Rep" indicates representative sample based on HRS respondent analysis weights. "Sim" indicates simulation sample weighted by the same analysis weights.

B Multiple imputation of consumption and other missing data

We followed the procedure outlined in Miller and Bairoliya (2023) to impute missing consumption data using the EM-bootstrapping algorithm suggested by Honaker and King (2010), implemented through Amelia II software (Honaker et al. 2011). This method produces multiple complete datasets (set as $m = 12$) for independent analysis, with results combined into a single estimate. The approach assumes multivariate normality of complete data and missing data being at random, with related covariates included in the imputation model to address nonrandom missingness. Variables from the RAND HRS data file are utilized, including household size, age, age squared, cubed root of total wealth, log household income, hours worked, and an alternate measure of consumption that included health spending. We also include dummy indicators for urban, cohort, labor force status, retired, gender, race, education, marital status, census division, 1980 census occupation code for longest reported tenure, self-reported health, ADLs, and eight doctor diagnosed health conditions. Lags and leads of consumption, wealth, income, and hours worked were also included in our imputation model. While primarily focusing on imputing consumption data, Amelia II also provides imputed values for other missing variables in the model. Diagnostic tests were conducted to assure the credibility of the imputation model.

C Health utility weights

In the calibration of health utility weights it was assumed that the HUI3 measures relative utility across health states holding consumption and leisure fixed. Table 6 provides the health utility weights γ estimated via a linear regression of HUI3 utility score on health outcomes.

Table 6: Estimated health utility weights (γ)

Measure	Weight	SE
Self-rated health		
Fair	0.226	0.026
Good	0.312	0.026
Very good	0.402	0.028
Excellent	0.420	0.031
Hypertension	0.005	0.012
Diabetes	-0.002	0.018
Cancer	0.010	0.017
Lung disease	-0.026	0.022
Heart disease	-0.030	0.015
Stroke	-0.076	0.022
Psych problem	-0.070	0.020
Arthritis	-0.062	0.012
Diff with ADL	-0.162	0.016
Constant	0.517	0.028

Notes: Results from regression of adjusted HUI3 score on self-rated health and morbidities. SE denotes standard error. $R^2 = 0.497$. N = 1,088.

Some scholars have raised doubts about whether respondents fully consider changes in health states without also factoring in changes in other life aspects, such as consumption and leisure, when completing the HUI3 instrument (Feeny et al. 2018). In such scenarios, a more appropriate representation of the HUI3 instrument could be formulated as follows:

$$\gamma h\left[\bar{u}+log\left(c\right)+\nu\left(l\right)\right] = HUI3\times h_{best}\left[\bar{u}+log\left(c_{best}\right)+\nu\left(l_{best}\right)\right],$$

where c_{best} and l_{best} are unobserved consumption and leisure that would arise in the best health state. By rearranging terms and normalizing $h_{best}=1$, we obtain:

$$\gamma h = HUI3\frac{\bar{u}+log\left(c_{best}\right)+\nu\left(l_{best}\right)}{\bar{u}+log\left(c\right)+\nu\left(l\right)}. \tag{9}$$

Using the forecasting model, we can predict the expected values for c_{best} and l_{best} for each individual in the sample. This allows us to compute the right-hand side of equation (9) for every HUI3 respondent in our simulation sample. Subsequently, we conducted a regression of this value on self-rated health and all morbidity indicators to derive alternative utility weights γ for our robustness check (see results in Table 7).

Table 7: Estimated alternate health utility weights (γ)

Measure	Weight	SE
Self-rated health		
Fair	0.270	0.034
Good	0.332	0.035
Very good	0.412	0.036
Excellent	0.406	0.041
Hypertension	-0.003	0.017
Diabetes	0.011	0.023
Cancer	0.003	0.023
Lung disease	-0.036	0.029
Heart disease	-0.048	0.020
Stroke	-0.059	0.029
Psych problem	-0.058	0.027
Arthritis	-0.063	0.017
Diff with ADL	-0.139	0.021
Constant	0.509	0.038

Notes: Results from regression of adjusted HUI3 score on self-rated health and morbidities. SE denotes standard error. $R^2 = 0.444$. N = 759.

D Additional results

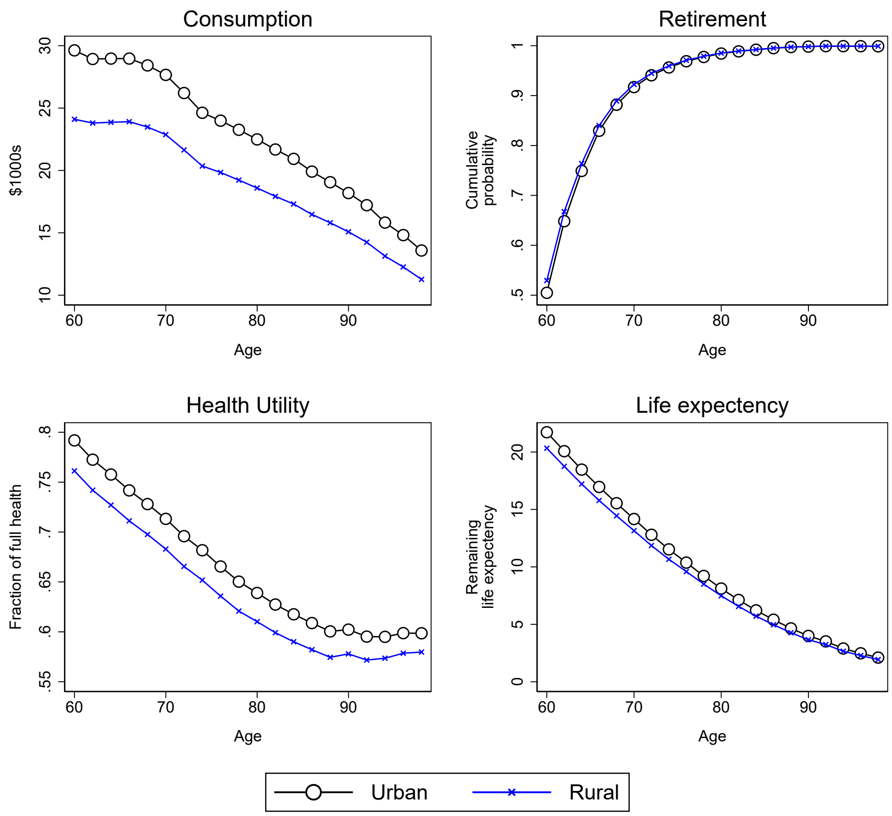


Figure 7: Average life cycle profiles by rural/urban

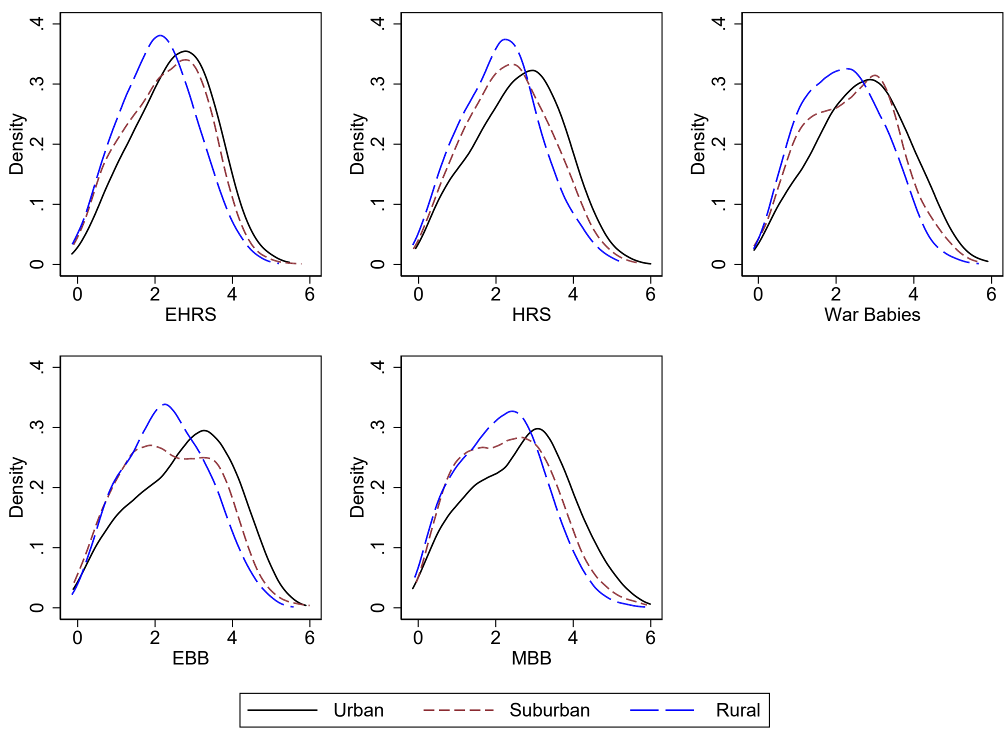


Figure 8: Distribution of log welfare by location

References

De Vos, I., Everaert, G., and Ruyssen, I. (2015). Bootstrap-based bias correction and inference for dynamic panels with fixed effects. *The Stata Journal*, 15(4):986–1018.

Everaert, G. and Pozzi, L. (2007). Bootstrap-based bias correction for dynamic panels. *Journal of Economic Dynamics and Control*, 31(4):1160–1184.

Feeny, D., Furlong, W., and Torrance, G. W. (2018). What were they thinking when providing preference measurements for generic health states? The evidence for HUI3. *Health and Quality of Life Outcomes*, 16:1–10.

Honaker, J. and King, G. (2010). What to do about missing values in time-series cross-section data. *American Journal of Political Science*, 54(2):561–581.

Honaker, J., King, G., Blackwell, M., et al. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, 45(7):1–47.

Miller, R. and Bairoliya, N. (2023). Health, longevity, and welfare inequality of older americans. *Review of Economics and Statistics*, 105(5):1145–1160.

Mullahy, J. (2016). Estimation of multivariate probit models via bivariate probit. *Stata Journal*, 16(1):37–51.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, pages 1417–1426.