Appendix for Accounting for selection bias due to death in estimating the effect of wealth shock on cognition for the Health and Retirement Study

Yaoyuan Vincent Tan, Carol A.C. Flannagan, Lindsay R. Pool, and Michael R. Ellio

February 25, 2021

A Appendix

A.1 Marginal Structural Models

The average treatment effect between two differing negative wealth shock profile $\bar{z}$ versus $\bar{z}'$ is thus $E[\bar{Y}_{\bar{z}} - \bar{Y}_{\bar{z}'}]$ (note that this estimand is not conditioned on the survival status). $E[\bar{Y}_{\bar{z}} - \bar{Y}_{\bar{z}'}]$ is obtained by maximizing the weighted likelihood of $\prod_{i=1}^{n} f(\bar{Y}_{i,\bar{z}(t)}|\theta_{it})^{w_{it}}$, where i indexes the subjects and θ_{it} are the parameters involved in the model for $\bar{Y}_{i,\bar{z}(t)}$ and

$$w_{it} = [\prod_{j=1}^{t} Pr\{Z_i(j) = z_i(j)|\bar{z}_i(j-1), \bar{y}_{i,\bar{z}(j-1)}, \bar{x}_{i,\bar{z}(t-1)}, \bar{w}_i(t-1), v_i; \tau(j)\}]^{-1}. \quad (1)$$

which is the inverse probability of receiving the observed exposure given all covariates and previous exposures. Under these four assumptions, inference about the exposure effects under a pseudo-population in which negative wealth shock is randomized can be obtained.

Similarly, this weighting method can be used to remove bias due to dropout. Let $R(t) = 1$ indicate that the subject's cognitive score is observed at time t and $R(t) = 0$ indicate that the subject's cognitive score is missing. The weight used to account for missing cognitive score is then

$$w_{it}^{r} = [\prod_{j=1}^{t} Pr(R_i(t) = r_i(t)|\bar{r}_i(j-1), \bar{z}_i(j-1), \bar{y}_{i,\bar{z}(j-1)}, \bar{x}_{i,\bar{z}(t-1)}, \bar{w}_i(t-1), v_i; \gamma(j))]^{-1}. \quad (2)$$

Finally, death is typically treated as equivalent to dropout in MSM. Let $D(t) = 1$ indicate that subject is dead at time t and $D(t) = 0$ indicate that the subject survived at time t (thus $D(t) = 1 - S(t)$). The weight for death censoring is then

$$w_{it} = [\prod_{j=1}^{t} Pr\{D_i(j) = d_i(j)|\bar{z}_i(j-1), \bar{y}_{i,\bar{z}(j-1)}, \bar{x}_{i,\bar{z}(t-1)}, \bar{w}_i(t-1), v_i; \lambda(j)\}]^{-1}. \quad (3)$$

Assuming that these three weights are independent of each other, the final weight to use becomes $w_{it}^f = w_{it} w_{it}^d w_{it}^r$. To stabilize the weights, the numerators of Equations 1, 2, and 3 are replaced by the marginal probabilities of negative wealth shock, dropout, and death at baseline given by $\prod_{j=1}^t Pr[Z_i(j) = z_i(j)|\bar{z}_i(j-1), v_i; \tau'(j)]$, $\prod_{j=1}^t Pr[R_i(j) = r_i(j)|\bar{r}_i(j-1), v_i; \gamma'(j)]$, and $\prod_{j=1}^t Pr[D_i(j) = d_i(j)|v_i; \lambda'(j)]$ respectively.

A.2 Penalized spline of propensity methods in treatment comparisons (PENCOMP)

We provide a detailed description of a particular implementation of PENCOMP here.

1. For $b = 1, \ldots, B$, generate a bootstrap sample $A^{(b)}$ from the original data A by sampling units with replacement, stratified on exposure group. For each sample b, carry out steps 2-7.

2. Estimate a logistic regression model for the distribution of $Z(1)$ given baseline covariates V with regression parameters $\gamma_{z(1)}$. Estimate the propensity of exposure $Z(1) = z(1)$ as $\hat{P}_{z(1)}(V) = Pr(Z(1) = z(1)|V; \hat{\gamma}_{z(1)}^{(b)})$, where $\hat{\gamma}_{z(1)}^{(b)}$ is the maximum likelihood (ML) estimate of $\gamma_{z(1)}$. Define $\hat{P}_{z(1)}^* = \log[\frac{\hat{P}_{z(1)}(V)}{1-\hat{P}_{z(1)}(V)}]$.

3. Using the cases assigned to exposure $Z(1) = z(1)$, estimate a normal linear regression of $Y_{z(1)}$ on V, with mean

$$E(Y_{Z(1)}|v, z(1), \theta_{z(1)}, \beta_{z(1)}) = s(\hat{P}_{z(1)}^*|\theta_{z(1)}) + g_{z(1)}(\hat{P}_{z(1)}^*, v; \beta_{z(1)}), \tag{4}$$

   where $s(\hat{P} *_{z(1)} |\theta_{z(1)})$ denotes a penalized spline with fixed knots and parameters $\theta_{z(1)}$ and $g_{z(1)}(.)$ represents a parametric function of other predictors of the outcome, indexed by parameters $\beta_{z(1)}$.

4. For $z(1) = 0, 1$, impute the values of $Y_{z(1)}$ for subjects in exposure group $1 - z(1)$ in the original data with draws from the predictive distribution of $Y_{z(1)}$ given V from the regression in Step 3, with the ML estimates $\hat{\theta}_{z(1)}^{(b)}, \hat{\beta}_{z(1)}^{(b)}$ substituted for the parameters $\theta_{z(1)}^{(b)}, \beta_{z(1)}^{(b)}$.

5. Estimate a logistic regression model for the distribution of $Z(2)$ given $V, Z(1), (Y_{Z(1)=0}, Y_{Z(1)=1})$, with regression parameters $\gamma_{z(2)}$ and missing values of (Y_0, Y_1) imputed from Step 4. Estimate the propensity of exposure $Z(2) = z(2)$ given $Z(1)$, $Y_{Z(1)}$, and V as $\hat{P}_{z(2)}(Z(1), Y_{Z(1)}, V) = Pr(Z(2) = z(2)|Z(1) = z(1), Y_{z(1)}, V; \hat{\gamma}_{z(2)}^{(b)})$, where $\hat{\gamma}_{z(2)}^{(b)}$ is the ML estimate of $\gamma_{z(2)}$. The probability of exposure profile $\{Z(1) = z(1), Z(2) = z(2)\}$ is denoted as $\hat{P}_{\bar{z}(2)} = \hat{P}_{z(1)}(V)\hat{P}_{z(2)}(Z(1), Y_{Z(1)}, V)$, and define $\hat{P}_{\bar{z}(2)}^* = \log[\frac{\hat{P}_{\bar{z}(2)}}{1-\hat{P}_{\bar{z}(2)}}]$.

6. Using the cases assigned to exposure $\{z(1), z(2)\}$, estimate a normal linear regression of $Y_{\bar{z}(2)}$ on $\bar{Z}(2)$, $Y_{Z(1)}$, and V with mean

$$E(Y_{\bar{Z}(2)}|v, y_{z(1)}, \bar{z}(2), \theta_{\bar{z}(2)}, \beta_{\bar{z}(2)}) = s(\hat{P}_{\bar{z}(2)}^*|\theta_{\bar{z}(2)}) + g_{\bar{z}(2)}(\hat{P}_{\bar{z}(2)}^*, Z(2), Z(1), Y_{Z(1)}, V; \beta_{\bar{z}(2)}). \tag{5}$$

7. For each combination of $\{z(1), z(2)\}$, impute the values of $Y_{\bar{z}(2)}$ for subjects not assigned this exposure profile in the original data with draws from the predictive distribution of $Y_{\bar{z}(2)}$ in Step 6, with ML estimates $\hat{\theta}^{(b)}_{\bar{z}(2)}, \hat{\beta}^{(b)}_{\bar{z}(2)}$ substituted for the parameters $\theta^{(b)}_{\bar{z}(2)}, \beta^{(b)}_{\bar{z}(2)}$. Let $\hat{\Delta}^{(b)}_{jk,lm} = E[Y_{Z(1)=j,Z(2)=k} - Y_{Z(1)=l,Z(2)=m}]$, denote the average exposure effects, with associated pooled variance estimates $W^{(b)}_{jk,lm}$, based on the observed and imputed values of Y for each exposure profile.

8. The MI estimate of $\Delta_{jk,lm}$ is then $\bar{\Delta}_{jk,lm,B} = \sum_{b=1}^{B} B^{-1}\hat{\Delta}^{(b)}_{jk,lm}$, and the MI estimate of the variance of $\bar{\Delta}_{jk,lm}$ is $T_B = \bar{W}_{jk,lm,B} + (1 + 1/B)D_{jk,lm,B}$, where $\bar{W}_{jk,lm,B} = \sum_{b=1}^{B} W^{(b)}_{jk,lm}/B$, $D_{jk,lm,B} = \sum_{b=1}^{B} \frac{(\hat{\Delta}^{(b)}_{jk,lm} - \bar{\Delta}_{jk,lm,B})^2}{B-1}$. The estimate $\Delta_{jk,lm}$ follows a t distribution with degree of freedom ν, $\frac{\Delta_{jk,lm} - \bar{\Delta}_{jk,lm,B}}{\sqrt{T_B}} \sim t_\nu$, where $\nu = (B-1)(1 + \frac{\bar{W}_{jk,lm,B}}{D_{jk,lm,B}(B+1)})^2$.

A.3 Simulation Setup

We set the size of our target population as 1 million. We then generate a single baseline variable V from a normal distribution. We set $T = 3$ and model our exposure, $Z(1)$, as

$$logit[P(Z(1) = 1|V)] = \gamma_0 + \gamma_1 V. \tag{6}$$

For the potential outcome at $t = 1$, $Y_{Z(1)}$, we model it as

$$Y_{Z(1)} = \beta_0 + \beta_Z I\{Z(1) = 1\} + \beta_V V + \beta_{VZ} V I\{Z(1) = 1\} + e, \tag{7}$$

where $e \sim N(0, 1)$.

We model the potential survival status at $t = 2$, $S_{Z(1)}$ as

$$\begin{aligned} logit(P[S_{Z(1)} = 1|V, Y_{Z(1)}]) = \alpha_0 + \alpha_{Y_1} Y_1 I\{Z(1) = 1\} + \alpha_{Y_0} Y_0 [1 - I\{Z(1) = 1\}] \\ + \alpha_Z I\{Z(1) = 1\} + \alpha_V V + \alpha_{VZ} V I\{Z(1) = 1\}. \end{aligned} \tag{8}$$

Monotonicity is imposed by setting $S(0) = 1$ if $S(1) = 1$. Because a negative wealth shock is an absorbing state, if $Z(1) = 1$, then $Z(2) = 1$. So when $Z(1) = 0$, we have

$$logit(P[Z(2) = 1|V, Y_0]) = \gamma_0 + \gamma_{Y_0,2} Y_0 + \gamma_2 V. \tag{9}$$

We model the potential outcome at $t = 2$, $Y_{\bar{Z}(2)}$ as

$$\begin{aligned} Y_{\bar{Z}(2)} = \beta_0 &+ \beta_{Z_{01}} I\{Z(1) = 0, Z(2) = 1\} + \beta_{Z_{11}} I\{Z(1) = 1, Z(2) = 1\} \\ &+ \beta_{Y_0 Z_{00}} Y_0 I\{Z(1) = 0, Z(2) = 0\} + \beta_{Y_0 Z_{01}} Y(0) I\{Z(1) = 0, Z(2) = 1\} \\ &+ \beta_{Y_1 Z_{11}} Y_1 I\{Z(1) = 1, Z(2) = 1\} + \beta_V V + \beta_{V Z_{01}} V I\{Z(1) = 0, Z(2) = 1\} \\ &+ \beta_{V Z_{11}} V I\{Z(1) = 1, Z(2) = 1\} + e, \end{aligned} \tag{10}$$

where $e \sim N(0, 1)$.

For the potential survival status at $t = 3$, $S_{\bar{Z}(2)}$, if $S_{Z(1)} = 0$, then $S_{\bar{Z}(2)} = 0$. When $S_{Z(1)} = 1$, we have

$$\begin{aligned} logit(P[S_{\bar{Z}(2)} = 1|X, Y_{\bar{Z}(2)}, S_{Z(1)} = 1]) = {} & \alpha_0 + \alpha_{Z_{01}} I\{Z(1) = 0, Z(2) = 1\} \\ & + \alpha_{Z_{11}} I\{Z(1) = 1, Z(2) = 1\} \\ & + \alpha_{Y_{00}Z_{00}} Y_{00} I\{Z(1) = 0, Z(2) = 0\} \\ & + \alpha_{Y_{01}Z_{01}} Y_{01} I\{Z(1) = 0, Z(2) = 1\} \\ & + \alpha_{Y_{11}Z_{11}} Y_{11} I\{Z(1) = 1, Z(2) = 1\} \\ & + \alpha_V V + \alpha_{VZ_{01}} V I\{Z(1) = 0, Z(2) = 1\} \\ & + \alpha_{VZ_{11}} V I\{Z_(1) = 1, Z(2) = 1\}. \end{aligned} \tag{11}$$

Again, we impose monotonicity by setting $S_{00} = S_{01} = 1$ if $S_{11} = 1$ and $S_{00} = 1$ if $S_{01} = 1$. For the exposure at $t = 3$, $Z(3)$, if $Z(1) = Z(2) = 0$, we have

$$logit(P[Z(3) = 1|X, \bar{Y}_{00}]) = \gamma_0 + \gamma_{Y_{00}} Y_{00} + \gamma_{Y_{0,3}} Y_0 + \gamma_3 V. \tag{12}$$

For the potential outcome at $t = 3$, $Y_{\bar{Z}(3)}$, we have

$$\begin{aligned} Y_{\bar{Z}(3)} = {} & \beta_0 + \beta_{Z_{001}} I\{Z(1) = 0, Z(2) = 0, Z(3) = 1\} + \beta_{Z_{011}} I\{Z_1 = 0, Z_2 = 1, Z_3 = 1\} \\ & + \beta_{Z_{111}} I\{Z(1) = 1, Z(2) = 1, Z(3) = 1\} + \beta_{Y_{00}Z_{000}} Y_{00} I\{Z(1) = 0, Z(2) = 0, Z(3) = 0\} \\ & + \beta_{Y_{00}Z_{001}} Y_{00} I\{Z(1) = 0, Z(2) = 0, Z(3) = 1\} \\ & + \beta_{Y_{01}Z_{011}} Y_{01} I\{Z(1) = 0, Z(2) = 1, Z(3) = 1\} \\ & + \beta_{Y_{11}Z_{111}} Y_{11} I\{Z(1) = 1, Z(2) = 1, Z(3) = 1\} + \beta_{Y_0 Z_0} Y_0 I\{Z(1) = 0\} \\ & + \beta_{Y_1 Z_1} Y_1 I\{Z(1) = 1\} + \beta_V V + \beta_{VZ_{001}} V I\{Z(1) = 0, Z(2) = 0, Z(3) = 1\} \\ & + \beta_{VZ_{011}} V I\{Z(1) = 0, Z(2) = 1, Z(3) = 1\} \\ & + \beta_{VZ_{111}} V I\{Z(1) = 1, Z(2) = 1, Z(3) = 1\} + e. \end{aligned} \tag{13}$$

Table 1 shows the parameters we used to achieve the three different simulation scenarios. Scenario 1 is achieved by setting γ_1, α_Z, γ_2, $\gamma_{Y_0,2}$, $\alpha_{Z_{01}}$, $\alpha_{Z_{11}}$, γ_3, $\gamma_{Y_0,3}$, and $\gamma_{Y_{00}}$ to be about 10 times smaller than the values in Scenarios 2 and 3. The rest of the differences between Scenario 1 versus 2 and 3 were to ensure the resulting simulated population would have enough deaths and subjects in the various different exposure profiles for the assumptions used by MSM and our proposed method to be valid. The difference between Scenario 2 versus 3 lie in β_{VZ}, α_{Y_1}, α_{Y_0}, $\beta_{Y_0 Z_{00}}$, $\beta_{Y_0 Z_{01}}$, $\beta_{Y_1 Z_{11}}$, $\alpha_{Y_0 Z_{00}}$, $\alpha_{Y_0 Z_{01}}$, $\alpha_{Y_1 Z_{11}}$, $\beta_{Y_{00} Z_{000}}$, $\beta_{Y_{00} Z_{001}}$, $\beta_{Y_{01} Z_{011}}$, and $\beta_{Y_{11} Z_{111}}$ where the values for Scenario 2 is about 10 times smaller compared to Scenario 3.

Table 1: Table of parameters for simulation

	Scenario 1	Scenario 2	Scenario 3
V	$N(0,2^2)$	$N(17,2^2)$	$N(17,2^2)$
γ_0	0	2	2
γ_1	-0.02	-0.2	-0.2
β_0	0	5.3	5.3
β_Z	-1.5	-1.5	-1.5
β_V	0.015	0.15	0.2
β_{VZ}	-0.005	-0.11	-0.05
α_0	0	1	0
α_{Y_1}	0.005	0.00625	0.0625
α_{Y_0}	0.01	0.0125	0.125
α_Z	-0.01	-0.2	-0.2
α_V	0.002	0.02	0.02
α_{VZ}	-0.002	-0.02	-0.02
γ_2	-0.002	-0.02	-0.02
$\gamma_{Y_0,2}$	-0.02	-0.2	-0.2
$\beta_{Z_{01}}$	-1.5	-1.5	-1.5
$\beta_{Z_{11}}$	-1	-1	-1
$\beta_{Y_0Z_{00}}$	0.015	0.02	0.3
$\beta_{Y_0Z_{01}}$	0.01	0.015	0.2
$\beta_{Y_1Z_{11}}$	0.005	0.01	0.1
$\beta_{VZ_{01}}$	-0.00011	-0.011	-0.011
$\beta_{VZ_{11}}$	-0.00005	-0.005	-0.005
$\alpha_{Z_{01}}$	-0.01	-0.2	-0.2
$\alpha_{Z_{11}}$	-0.015	-0.1	-0.1
$\alpha_{Y_0Z_{00}}$	0.01	0.0125	0.125
$\alpha_{Y_0Z_{01}}$	0.005	0.00625	0.0625
$\alpha_{Y_1Z_{11}}$	0.0025	0.003125	0.03125
$\alpha_{VZ_{01}}$	-0.0001	-0.02	-0.02
$\alpha_{VZ_{11}}$	-0.0005	-0.05	-0.05
γ_3	-0.0002	-0.002	-0.002
$\gamma_{Y_0,3}$	-0.002	-0.02	-0.02
$\gamma_{Y_{00}}$	-0.02	-0.2	-0.2
$\beta_{Z_{001}}$	-1.5	-1.5	-1.5
$\beta_{Z_{011}}$	-1	-1	-1
$\beta_{Z_{111}}$	-0.5	-0.5	-0.5
$\beta_{Y_{00}Z_{000}}$	0.015	0.02	0.3
$\beta_{Y_{00}Z_{001}}$	0.01	0.015	0.2
$\beta_{Y_{01}Z_{011}}$	0.005	0.01	0.1
$\beta_{Y_{11}Z_{111}}$	0.0025	0.005	0.05
$\beta_{Y_0Z_0}$	0.0008	0.08	0.08
$\beta_{Y_1Z_1}$	0.0003	0.03	0.03
$\beta_{VZ_{001}}$	-0.00011	-0.011	-0.011
$\beta_{VZ_{011}}$	-0.00005	-0.005	-0.005
$\beta_{VZ_{111}}$	-0.00003	-0.003	-0.003

To calculate the true parameters, we used the generated population data (size 1 million), and then took:

1. $\Delta_{1,0} = \bar{Y}_1 - \bar{Y}_0$;
2. $\Delta_{01,00} = \bar{Y}_{01} - \bar{Y}_{00}$ given $S_0 = 1$;
3. $\Delta_{11,00} = \bar{Y}_{11} - \bar{Y}_{00}$ given $S_0 = S_1 = 1$;
4. $\Delta_{11,01} = \bar{Y}_{11} - \bar{Y}_{01}$ given $S_0 = S_1 = 1$;
5. $\Delta_{001,000} = \bar{Y}_{001} - \bar{Y}_{000}$ given $S_{00} = 1$;
6. $\Delta_{011,000} = \bar{Y}_{011} - \bar{Y}_{000}$ given $S_{00} = S_{01} = 1$;
7. $\Delta_{111,000} = \bar{Y}_{111} - \bar{Y}_{000}$ given $S_{00} = S_{11} = 1$;
8. $\Delta_{011,001} = \bar{Y}_{011} - \bar{Y}_{001}$ given $S_{00} = S_{01} = 1$;
9. $\Delta_{111,001} = \bar{Y}_{111} - \bar{Y}_{001}$ given $S_{00} = S_{11} = 1$; and
10. $\Delta_{111,011} = \bar{Y}_{111} - \bar{Y}_{011}$ given $S_{01} = S_{11} = 1$.

A.4 Result for sample size 4000

Table 2: Simulation results for sample size 4,000

Scenario 1		Naïve				MSM				Proposed			
Parameter	True value	Bias	RMSE	95% Coverage	AIL	Bias	RMSE	95% Coverage	AIL	Bias	RMSE	95% Coverage	AIL
$\Delta_{1,0}$	-1.497	-0.001	0.032	95.4	0.123	-0.0002	0.032	95.1	0.123	-0.0001	0.032	97.0	0.143
$\Delta_{01,00}$	-1.499	-0.003	0.050	95.3	0.202	-0.003	0.050	95.3	0.202	-0.003	0.050	95.7	0.214
$\Delta_{11,00}$	-1.005	-0.003	0.049	95.0	0.189	-0.001	0.049	94.7	0.189	-0.001	0.049	99.2	0.262
$\Delta_{11,01}$	0.493	0.002	0.048	94.4	0.189	0.003	0.048	94.5	0.189	0.002	0.049	99.1	0.262
$\Delta_{001,000}$	-1.502	0.005	0.081	93.9	0.314	0.005	0.081	98.9	0.411	0.005	0.082	94.6	0.333
$\Delta_{011,000}$	-1.008	0.004	0.074	94.8	0.284	0.004	0.074	99.0	0.370	0.004	0.075	97.8	0.350
$\Delta_{111,000}$	-0.504	0.006	0.072	95.2	0.284	0.007	0.072	99.4	0.371	0.007	0.074	100.0	0.529
$\Delta_{011,001}$	0.495	-0.001	0.071	95.0	0.284	-0.0001	0.072	99.1	0.370	-0.0009	0.072	97.8	0.348
$\Delta_{111,001}$	1.000	-0.0001	0.072	95.6	0.284	0.001	0.072	99.0	0.371	0.001	0.074	99.9	0.528
$\Delta_{111,011}$	0.502	0.003	0.065	94.3	0.250	0.005	0.065	98.9	0.325	0.005	0.067	99.9	0.440
Scenario 2		Naïve				MSM				Proposed			
Parameter	True value	Bias	RMSE	95% Coverage	AIL	Bias	RMSE	95% Coverage	AIL	Bias	RMSE	95% Coverage	AIL
$\Delta_{1,0}$	-3.367	-0.047	0.061	78.5	0.154	0.002	0.041	93.8	0.160	0.002	0.041	96.1	0.177
$\Delta_{01,00}$	-1.727	-0.037	0.054	83.2	0.149	-0.032	0.051	86.9	0.150	-0.002	0.037	96.4	0.161
$\Delta_{11,00}$	-1.199	-0.136	0.146	24.5	0.202	-0.020	0.057	92.5	0.204	-0.004	0.053	96.5	0.229
$\Delta_{11,01}$	0.528	-0.098	0.111	49.2	0.199	0.013	0.054	93.7	0.201	-0.001	0.053	97.0	0.226
$\Delta_{001,000}$	-1.727	-0.029	0.062	91.9	0.220	-0.023	0.060	94.8	0.240	0.001	0.053	96.1	0.227
$\Delta_{011,000}$	-1.183	-0.065	0.082	75.0	0.199	-0.047	0.069	87.5	0.217	0.0004	0.048	97.8	0.220
$\Delta_{111,000}$	-1.169	-0.167	0.181	33.8	0.273	-0.042	0.084	93.2	0.305	-0.004	0.071	98.7	0.350
$\Delta_{011,001}$	0.544	-0.036	0.059	88.0	0.185	-0.024	0.053	94.4	0.202	-0.002	0.045	96.7	0.206
$\Delta_{111,001}$	0.558	-0.139	0.153	45.7	0.264	-0.019	0.071	96.3	0.294	-0.007	0.067	98.4	0.331
$\Delta_{111,011}$	0.013	-0.101	0.119	62.9	0.246	0.007	0.065	96.1	0.276	-0.002	0.063	98.1	0.299
Scenario 3		Naïve				MSM				Proposed			
Parameter	True value	Bias	RMSE	95% Coverage	AIL	Bias	RMSE	95% Coverage	AIL	Bias	RMSE	95% Coverage	AIL
$\Delta_{1,0}$	-2.347	-0.123	0.130	14.5	0.160	0.002	0.042	94.0	0.160	0.002	0.042	95.9	0.177
$\Delta_{01,00}$	-2.559	-0.114	0.122	23.9	0.165	-0.060	0.074	70.2	0.164	-0.001	0.038	96.4	0.163
$\Delta_{11,00}$	-3.062	-0.231	0.239	2.8	0.232	-0.033	0.068	89.9	0.226	-0.004	0.058	97.0	0.260
$\Delta_{11,01}$	-0.502	-0.118	0.132	48.9	0.233	0.026	0.065	92.2	0.227	-0.003	0.059	96.7	0.260
$\Delta_{001,000}$	-2.820	-0.125	0.139	47.9	0.242	-0.062	0.087	88.7	0.273	-0.0004	0.054	95.9	0.224
$\Delta_{011,000}$	-3.605	-0.143	0.152	19.6	0.198	-0.087	0.101	69.2	0.225	-0.006	0.045	96.4	0.202
$\Delta_{111,000}$	-4.032	-0.290	0.301	5.2	0.319	-0.082	0.117	89.4	0.376	-0.009	0.080	98.6	0.400
$\Delta_{011,001}$	-0.785	-0.019	0.060	93.3	0.225	-0.026	0.063	95.0	0.256	-0.006	0.052	96.7	0.226
$\Delta_{111,001}$	-1.217	-0.160	0.181	54.4	0.336	-0.015	0.087	97.2	0.396	-0.009	0.083	99.3	0.442
$\Delta_{111,011}$	-0.432	-0.141	0.160	54.9	0.306	0.011	0.080	97.4	0.363	-0.006	0.075	98.7	0.373

A.5 Descriptive statistics at baseline

Tables 3 to 4 show the descriptive statistics of the subjects at baseline by whether or not they experienced a negative wealth shock over the next six years regardless of survival status. At baseline, aside from whether the subject eventually survived until 2002 and health conditions like whether the subject ever had heart problems, high blood pressure, and stroke, all the other variables in Tables 3 to 4 were significantly associated with experiencing a negative wealth shock. A typical subject who would eventually experience a wealth shock would have a lower cognitive score at baseline; slightly higher BMI; lower opinion about his or her health; lower word recall score; likely still smoking; not insured; have depression; slightly lower income; either working, unemployed, or disabled; divorced or never married; lower wealth rank; have diabetes and/or psychological problems; younger; lesser years of education; and likely non-White.

Table 3: Descriptive statistics of 1996 Health and Retirement Study (baseline), part 1

Variables	No wealth shock Mean/Frequency (S.E./%)	Ever wealth shock Mean/Frequency (S.E./%)	*p*-value
Eventually survived?:			0.57
Yes	6,207 (94.7)	516 (94.0)	
No	350 (5.3)	33 (6.0)	
Cognitive score	17.07 (4.07)	16.26 (4.35)	< 0.01
BMI	27.21 (4.84)	27.73 (5.40)	0.03
Self-reported health			< 0.01
Excellent	1,207 (19.9)	83 (15.7)	
Very Good	2,126 (35.0)	128 (24.3)	
Good	1,715 (28.2)	163 (30.9)	
Fair	763 (12.6)	103 (19.5)	
Poor	261 (4.3)	50 (9.5)	
Current Smoking status:			< 0.01
Never	2,353 (40.0)	166 (32.4)	
Former	2,410 (41.0)	187 (36.5)	
Current	1,116 (19.0)	159 (31.1)	
Alcohol consumption:			< 0.01
Never	3,799 (62.9)	347 (66.1)	
Moderate	1,686 (27.9)	116 (22.1)	
Heavy	555 (9.2)	62 (11.8)	
Insured?:			< 0.01
No	1,014 (15.5)	120 (21.9)	
Yes	5,543 (84.5)	429 (78.1)	
Depression?:			< 0.01
No	4,922 (85.5)	361 (73.1)	
Yes	832 (14.5)	133 (26.9)	
Income (log transformed)	10.48 (1.21)	10.18 (1.45)	< 0.01
Labor force status:			< 0.01
Working	3,111 (51.2)	314 (59.6)	
Unemployed	96 (1.6)	13 (2.5)	
Retired	2,178 (35.9)	104 (19.7)	
Disabled	143 (2.4)	43 (8.2)	
Not in labor force	547 (9.0)	53 (10.1)	
Martial status:			< 0.01
Married	4,897 (80.8)	373 (70.8)	
Divorced	591 (9.7)	90 (17.1)	
Widowed	426 (7.0)	42 (8.0)	
Never Married	149 (2.5)	22 (4.2)	
Wealth rank in tertiles:			< 0.01
0	1,728 (26.4)	326 (59.4)	
1	2,360 (36.0)	124 (22.6)	
2	2,469 (37.7)	99 (18.0)	
Gender:			0.08
Male	3,113 (47.5)	239 (43.5)	
Female	3,444 (52.5)	310 (56.5)	

Table 4: Descriptive statistics of 1996 Health and Retirement Study (baseline), part 2

Variables	No wealth shock Mean/Frequency (S.E./%)	Ever wealth shock Mean/Frequency (S.E./%)	*p*-value
Ever had diabetes?:			< 0.01
No	5,474 (90.2)	451 (85.6)	
Yes	596 (9.8)	76 (14.4)	
Ever had heart problems?:			0.43
No	5,343 (88.0)	457 (86.7)	
Yes	730 (12.0)	70 (13.3)	
Ever had HBP?:			0.07
No	3,888 (64.0)	316 (60.0)	
Yes	2,183 (36.0)	211 (40.0)	
Ever had psych problems?:			< 0.01
No	5,691 (93.7)	469 (89.2)	
Yes	380 (6.3)	57 (10.8)	
Ever had stroke?:			0.1
No	5,912 (97.3)	506 (96.0)	
Yes	161 (2.7)	21 (4.0)	
Age	59.73 (3.19)	57.26 (2.18)	< 0.01
Number of education years centered	0.52 (2.93)	-0.17 (3.32)	< 0.01
Race:			< 0.01
Non-hispanic White	5,236 (79.9)	342 (62.3)	
Non-hispanic Black	759 (11.6)	120 (21.9)	
Hispanic	449 (6.8)	70 (12.8)	
Other	113 (1.7)	17 (3.1)	