Evaluating a Markov regime-switching model of health care expenditure dynamics

Course: MACS 40200 Structural Estimation

Presenters: Bobae Kang and DP Zhang

Agenda

- Research Question
- Data
- Model
- Estimation strategy and results
- Evaluation of labeling scheme
- Summary

Research Question

- Can a Markov regime-switching model explain/simulate individual monthly health care expenditures behaviors?
    - We attempt to replicate Evans, Humpherys and Taylor’s unpublished work on estimating health care cost dynamics and test/validate their identification/labeling of health care patterns.

Data

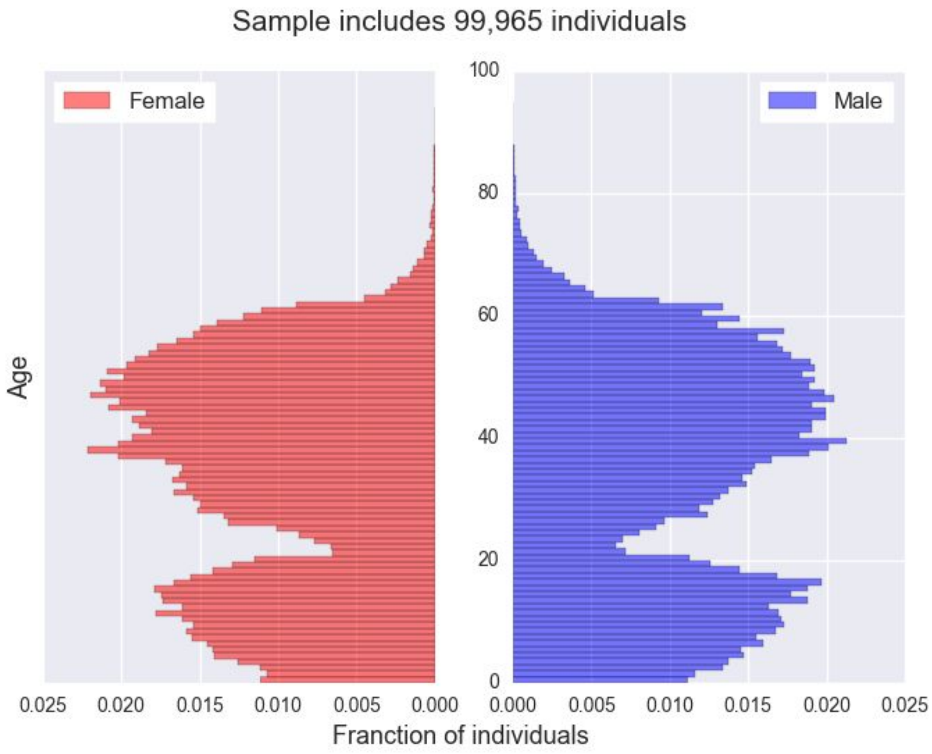
- Monthly health care expenditure data for a sample of 99965 individuals
  - Removing 35 observations with negative monthly health expenditure

Table 1: Descriptive statistics of the monthly health care expenditures

	Count	Mean	Median	Max	Min	St. Dev.
Monthly exp., total	2399160	305.31	0.0	422855.84	0.0	2235.93
Monthly exp., zero	1375975	-	-	-	-	-
Monthly exp., positive	1023185	715.88	173.30	422855.84	0.01	3380.62
Monthly exp., female	1228872	328.18	0.0	355495.02	0.0	2008.85
Monthly exp., male	1170288	281.29	0.0	422855.84	0.0	2451.61
Age, total	99965	33.20	36	95	0	18.45
Age, female	51203	33.30	36	94	0	18.18
Age, male	48762	33.09	36	95	0	18.73

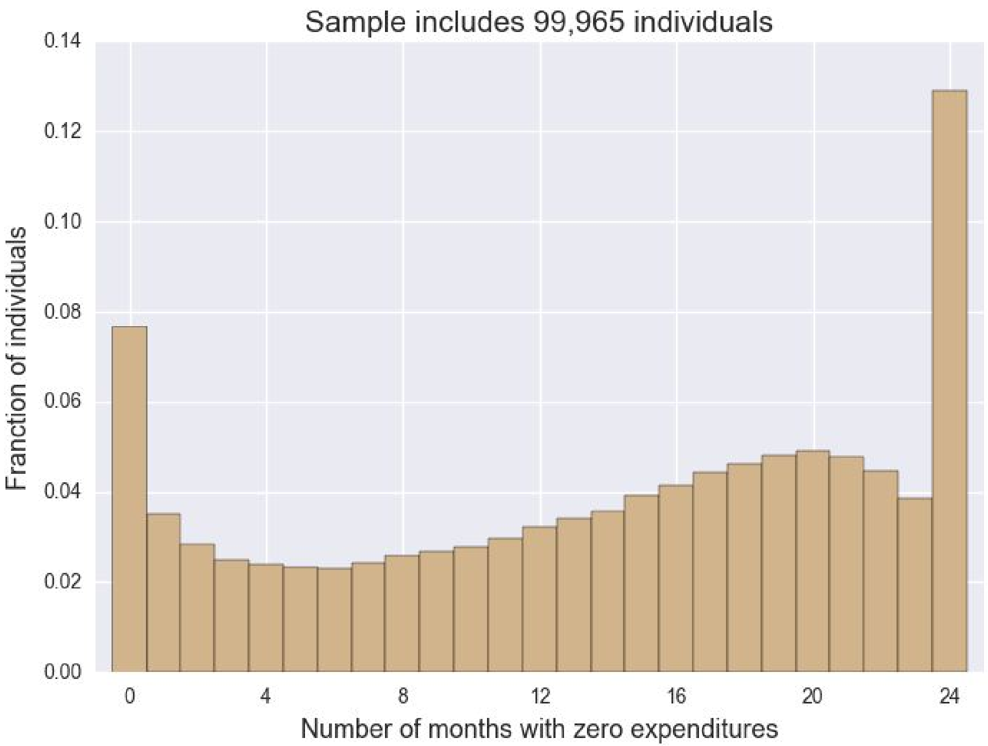
Data (cont.)

- Distribution of male and female individuals by age
  - Bimodal distribution



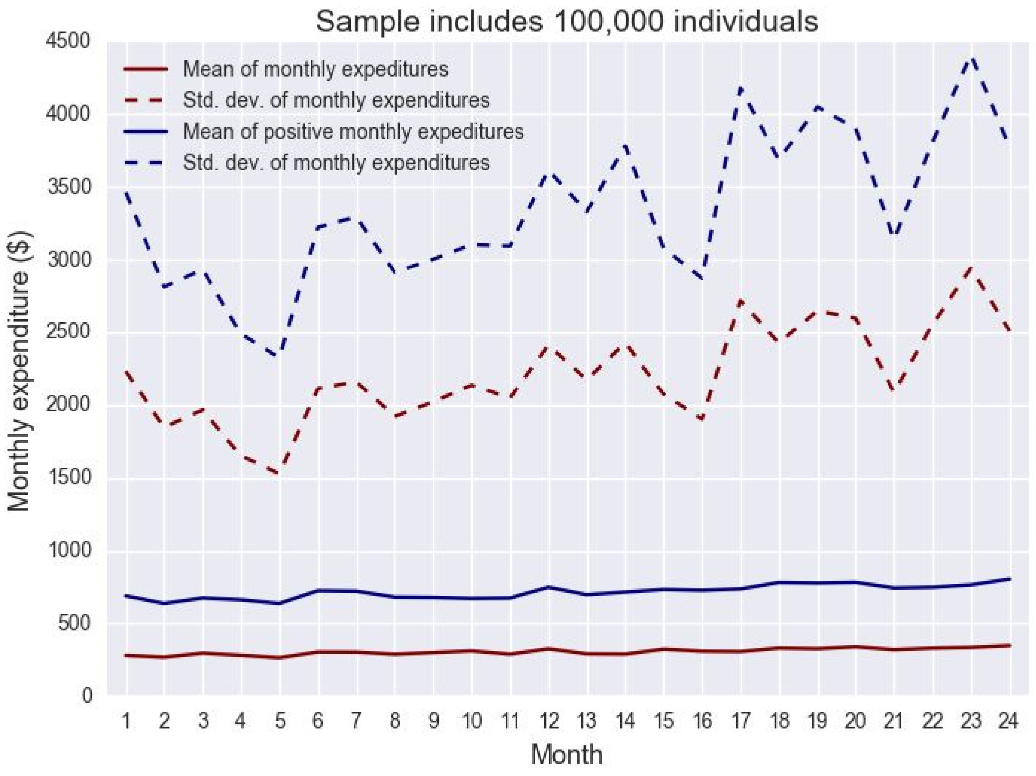
Data (cont.)

- Histogram of the fraction of individuals with a given number of zero-expenditure months



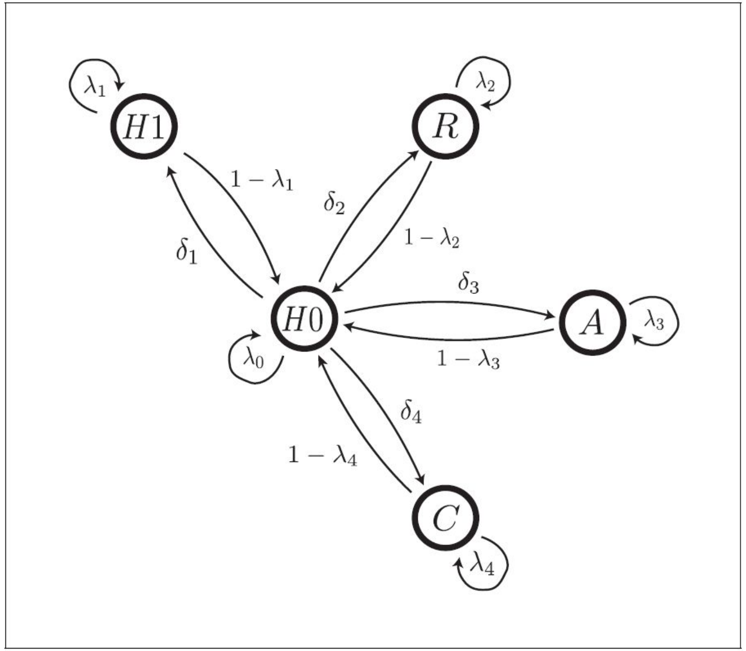
Data (cont.)

- Mean and standard deviation of both all and positive monthly expenditures by month



Model

- Data Generating Process (DGP)
- Markov regime-switching process (5 states)
- Five states $s \in \{H1, H0, R, A, C\}$
- Proportion vector v = $[v_0, v_1, v_2, v_3, v_4]^T$
  - Assuming the ergodic distribution
- Transition matrix **A** $\mathbf{v}^T\mathbf{A} = \mathbf{v}^T$



Estimation strategy

- Generalized Methods of Moments (GMM)

Estimation strategy (cont.)

- Data moments
- b_k : The unconditional probability that the current period and the previous k periods all had **positive** expenditures (24 moments; 0 =< k =< 23)
- $a_k = Pr(h_{i,t} > 0 | h_{i,t-1} > 0, \dots, h_{i,t-k} > 0)$

$$
\begin{aligned}
b_k &= Pr(h_{i,t} > 0, h_{i,t-1} > 0, \dots, h_{i,t-k} > 0) \\
&= Pr(h_{i,t} > 0 | h_{i,t-1} > 0, \dots, h_{i,t-k} > 0) \cdot Pr(h_{i,t-1} > 0, \dots, h_{i,t-k} > 0) \\
&= a_k b_{k-1} = a_k a_{k-1} b_{k-2} = a_k a_{k-1} a_{k-2} b_{k-3} \\
&= \prod_{j=0}^{k} a_j \quad \text{because} \quad b_0 = a_0
\end{aligned}
$$

Estimation strategy (cont.)

- Data moments
- d_k : The unconditional probability that the current period and the previous k periods all had **zero** expenditures (24 moments; 0 =< k =< 23)
- $c_k = Pr(h_{i,t} = 0 | h_{i,t-1} = 0, \ldots, h_{i,t-k} = 0)$

$$\begin{aligned} d_k &= Pr(h_{i,t} = 0, h_{i,t-1} = 0, \ldots, h_{i,t-k} = 0) \\ &= Pr(h_{i,t} = 0 | h_{i,t-1} = 0, \ldots, h_{i,t-k} = 0) \cdot Pr(h_{i,t-1} = 0, \ldots, h_{i,t-k} = 0) \\ &= c_k d_{k-1} = c_k c_{k-1} d_{k-2} = c_k c_{k-1} c_{k-2} d_{k-3} \\ &= \prod_{j=0}^{k} c_j \quad \text{because} \quad d_0 = c_0 \end{aligned}$$

Estimation strategy (cont.)

- Model moments
- b_k : The unconditional probability that the current period and the previous k periods all had **positive** expenditures (24 moments; 0 =< k =< 23)
- $$v_2 + v_3 + v_4 = b_0$$

$$\lambda_2 v_2 + \lambda_3 v_3 + \lambda_4 v_4 = b_1$$

$$\lambda_2^2 v_2 + \lambda_3^2 v_3 + \lambda_4^2 v_4 = b_2$$

$$\vdots$$

$$\lambda_2^k v_2 + \lambda_3^k v_3 + \lambda_4^k v_4 = b_k$$

Estimation strategy (cont.)

- Model moments
- d_k : The unconditional probability that the current period and the previous k periods all had **zero** expenditures (24 moments; 0 =< k =< 23)
- $$d_k = p_{k-1,0}(\lambda_0 + \delta_1) + p_{k-1,1},$$

$$p_{k,0} = p_{k-1,0}\lambda_0 + p_{k-1,1}(1 - \lambda_1),$$

$$p_{k,1} = p_{k-1,0}\delta_1 + p_{k-1,1}\lambda_1 \quad \text{for} \quad k = 1, 2, \ldots$$

$$\text{and} \quad d_0 = v_0 + v_1, \quad p_{0,0} = v_0, \quad p_{0,1} = v_1.$$

Estimation strategy (cont.)

- Vectors of data moments: $\mathbf{b} = [b_0, b_1, ...b_{23}]^T$ and $\mathbf{d} = [d_0, d_1, ...d_{23}]^T$
- Model moments: $B(\vec{\lambda})\mathbf{v}$ and $D(\vec{\lambda}, \delta_1, \mathbf{v})$

$$B(\vec{\lambda}) = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & \lambda_2 & \lambda_3 & \lambda_4 \\ 0 & 0 & \lambda_2^2 & \lambda_3^2 & \lambda_4^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \lambda_2^{T-1} & \lambda_3^{T-1} & \lambda_4^{T-1} \end{bmatrix} \quad \text{and} \quad D(\vec{\lambda}, \delta_1, \mathbf{v}) = \begin{bmatrix} v_0 + v_1 \\ p_{0,0}(\lambda_0 + \delta_1) + p_{0,1} \\ p_{1,0}(\lambda_0 + \delta_1) + p_{1,1} \\ \vdots \\ p_{22,0}(\lambda_0 + \delta_1) + p_{22,1} \end{bmatrix}.$$

- Minimization:

$$\min_{\{\lambda_i\}_{i=0}^4, \{\delta_j,\}_{j=1}^3} \|B(\vec{\lambda})\mathbf{v} - \mathbf{b}\|_2 + \|D(\vec{\lambda}, \delta_1, \mathbf{v}) - \mathbf{d}\|_2$$

$$\text{s.t.} \quad \mathbf{v}'A(\vec{\lambda}, \vec{\delta}) = \mathbf{v}' \quad \text{and} \quad \delta_4 = 1 - \lambda_0 - \delta_1 - \delta_2 - \delta_3$$

$$\text{and} \quad \lambda_i, \delta_j \in (0, 1) \quad \forall i, j$$

Estimation results

- Estimated parameters and ergodic distribution

Table 2: Estimated values of $\vec{\lambda}$ and $\vec{\delta}$ and corresponding v

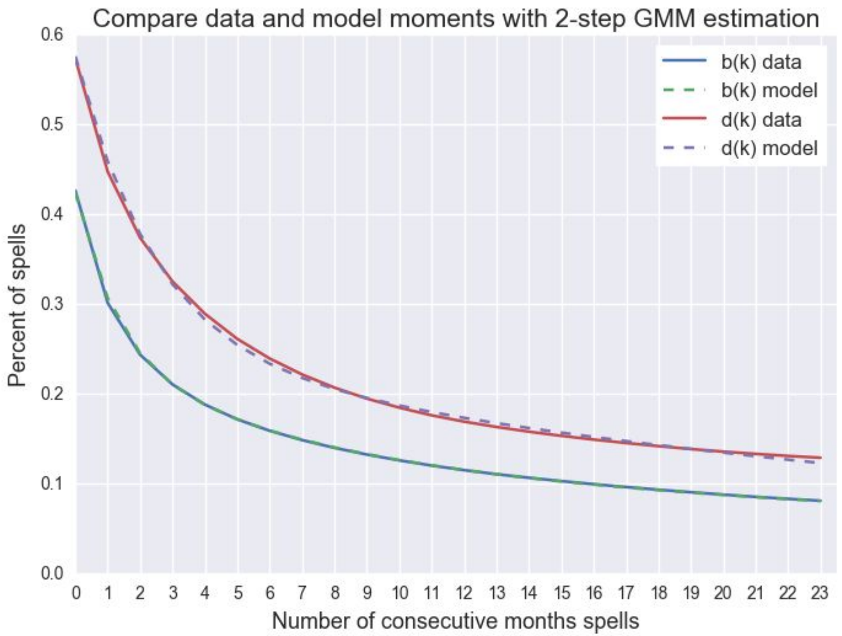
	λ_s	δ_s	v_s
$H0\ s=0$	0.6698		0.3743
$H1\ s=1$	0.9690	0.0167	0.2017
$H2\ s=2$	0.8465	0.0525	0.1281
$H3\ s=3$	0.9796	0.0069	0.1256
$H4\ s=4$	0.4415	0.2541	0.1703

Estimation results (cont.)

- Comparison of model moments to data moments

	model	data		model	data		model	data
b_0	0.4265	0.4240	b_8	0.1400	0.1405	b_{16}	0.0996	0.0991
b_1	0.3014	0.3066	b_9	0.1326	0.1329	b_{17}	0.0964	0.0959
b_2	0.2438	0.2455	b_{10}	0.1261	0.1264	b_{18}	0.0935	0.0930
b_3	0.2106	0.2104	b_{11}	0.1204	0.1206	b_{19}	0.0906	0.0902
b_4	0.1882	0.1879	b_{12}	0.1152	0.1154	b_{20}	0.0880	0.0877
b_5	0.1718	0.1718	b_{13}	0.1107	0.1107	b_{21}	0.0855	0.0853
b_6	0.1591	0.1593	b_{14}	0.1067	0.1065	b_{22}	0.0833	0.0830
b_7	0.1487	0.1491	b_{15}	0.1030	0.1026	b_{23}	0.0812	0.0809

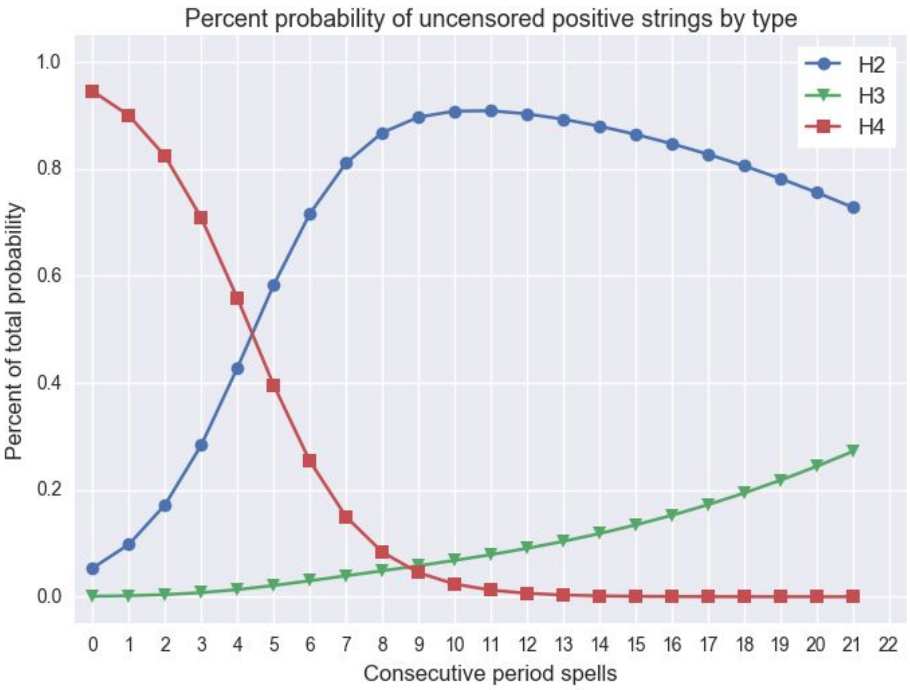
	model	data		model	data		model	data
d_0	0.5735	0.5760	d_8	0.2072	0.2056	d_{16}	0.1493	0.1522
d_1	0.4474	0.4587	d_9	0.1949	0.1955	d_{17}	0.1454	0.1475
d_2	0.3739	0.3781	d_{10}	0.1848	0.1871	d_{18}	0.1420	0.1431
d_3	0.3253	0.3222	d_{11}	0.1762	0.1799	d_{19}	0.1387	0.1388
d_4	0.2893	0.2828	d_{12}	0.1692	0.1734	d_{20}	0.1358	0.1347
d_5	0.2613	0.2545	d_{13}	0.1633	0.1675	d_{21}	0.1333	0.1307
d_6	0.2395	0.2337	d_{14}	0.1581	0.1621	d_{22}	0.1310	0.1269
d_7	0.2217	0.2179	d_{15}	0.1534	0.1570	d_{23}	0.1290	0.1232

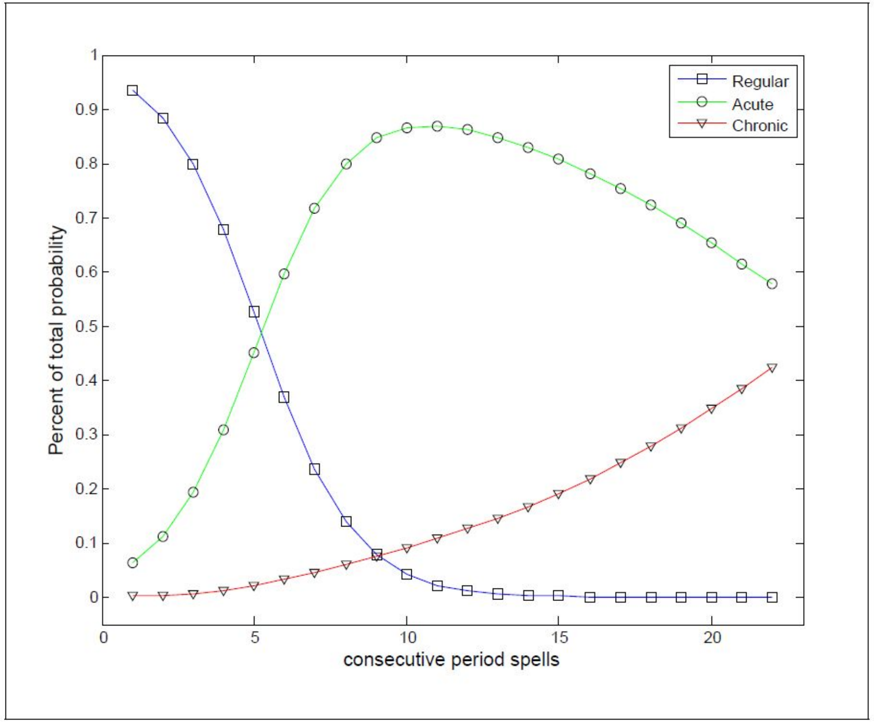


Evaluation of labeling scheme

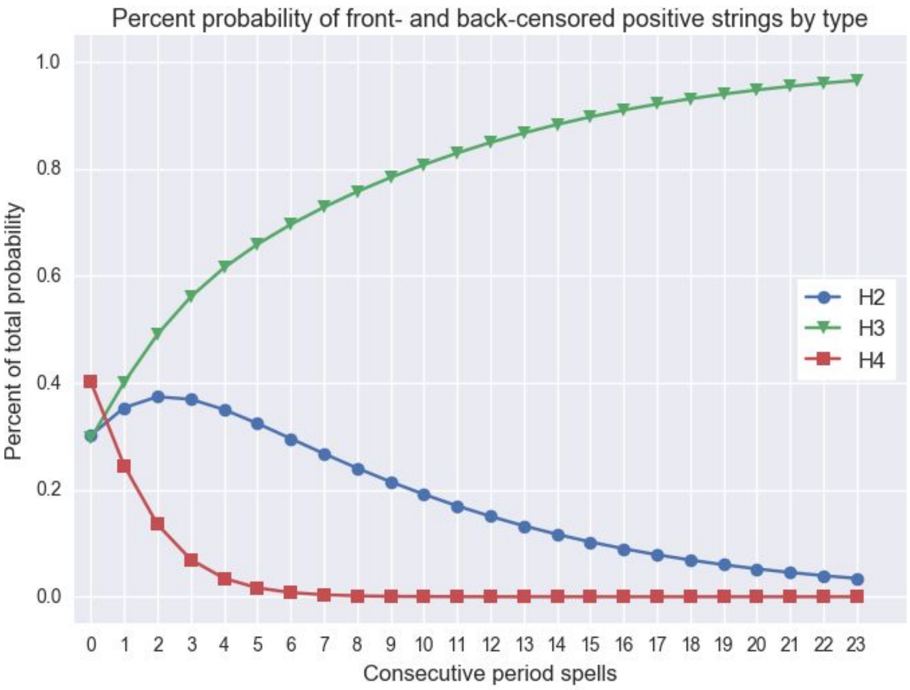
- Estimating probabilities of positive expenditure spells:
- $Pr\left(h_{i,t}=0, \{h_{i,j}\}_{j=t-1}^{t-n}>0, h_{i,t-n-1}=0\right)=\sum_{s=2}^{4}\delta_s\lambda_s^{n-1}(1-\lambda_s) \quad \text{for} \quad 1\leq n\leq 10$
- $Pr\left(h_{i,n+1}=0, \{h_{i,j}\}_{j=1}^{n}>0\right)=\sum_{s=2}^{4}v_s\lambda_s^{n-1}(1-\lambda_s) \quad \text{for} \quad 1\leq n\leq 11$
- $Pr\left(\{h_{i,j}\}_{j=12-n+1}^{12}>0, h_{i,12-n}=0\right)=\sum_{s=2}^{4}\delta_s\lambda_s^{n-1} \quad \text{for} \quad 1\leq n\leq 11$
- $Pr\left(\{h_{i,j}\}_{j=1}^{n}>0\right)=\sum_{s=2}^{4}v_s\lambda_s^{n-1} \quad \text{for} \quad n=11$
- Any string of the positive expenditures has a positive probability of being any of the three health states with positive expenditure

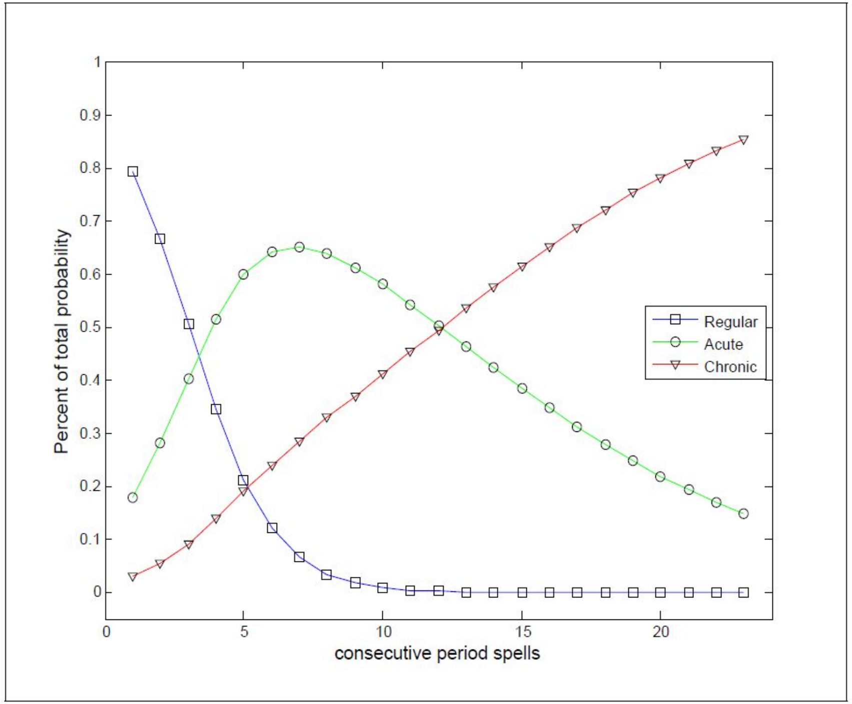
Evaluation of labeling scheme (cont.)





Evaluation of labeling scheme (cont.)





Evaluation of labeling scheme (cont.)

- H2, H3, and H4 corresponds to A (acute), C (chronic), and R (regular) in the original study
- Original labels may misrepresent the characteristics of H2 and H4 types
- Alternative labeling scheme:
  - M (moderately ill)
  - S (severely ill), and
  - C (chronically ill)

Table 4: Comparison of positive expenditure patterns across gender and age

		*H*2	*H*3	*H*4	Total
Count	All	277,056	283,213	462,916	1,023,185
	Female	164,017	169,934	250,234	584,185
	Male	113,039	113,279	212,682	439,000
	Age (< 21)	50,216	27,231	164,519	241,966
	Age (< 45)	101,061	82,189	153,296	336,546
	Age (< 65)	119,435	161,555	139,454	420,444
	Age (≥ 65)	6,344	12,238	5,647	24,229
Mean	All	798.71	1,073.61	447.46	715.88
	Female	753.09	1,010.54	431.79	690.35
	Male	864.89	1,168.23	465.89	749.86
	Age (< 21)	605.37	1,107.95	342.77	483.38
	Age (< 45)	742.09	859.14	486.15	654.09
	Age (< 65)	917.64	1,157.64	520.78	878.23
	Age (≥ 65)	991.94	1,328.38	636.15	1078.95
St. Dev.	All	3,771.12	4,650.51	1,833.88	3,380.62
	Female	3,059.81	3,897.57	1,625.46	2,859.95
	Male	4,611.30	5,591.74	2,052.03	3,947.60
	Age (< 21)	3,606.08	6,251.56	1,538.57	2,950.69
	Age (< 45)	3,212.72	3,741.47	1,893.12	2,735.59
	Age (< 65)	4,232.99	4,731.42	1,999.29	3,875.27
	Age (≥ 65)	3,966.34	4,851.03	3,193.50	4,287.45

Summary

- Markov process to model the dynamic process of health care expenditure
- GMM: Successfully matching target moments with estimated Markov-process parameters
- Alternative labeling scheme

Reference

- Evans, Richard W., Jeffery Humpherys and Sara Taylor. “A Markov Regime-switching approach to Estimating Health Cost Dynamics.” January 2014 (version 14.01.d).