Discussion of “Modeling the Top Tail of the Medical Care Spending Distribution” by Matt Harris and Jennifer Kohn

Andrew Ching
University of Toronto

September 25, 2014

Motivation: The Top Tail Matters

- The top 5% of medical care consumers account for nearly 50% of medical care spending.
  - The top 1% of consumers account for nearly 25% of spending.
  - The top 10% of consumers account for nearly 65% of spending.
- Significant heterogeneity exists among individuals in the top 5% of medical care consumers.
  - Approximately 45% of individuals in the top 5% are over 65 years old.
  - Roughly 1/3 of individuals in the top 5% are in “poor health”
- **Predicting participation in the right tail matters when forecasting medical care expenditures.**
- **It is not inherently obvious who will be in the right tail.**

Not clear if the existing models can help [Why?]

- Theoretical Models of investment in health as human capital
  - Grossman (1972), Ehrlich and Chuma (1990), Galama (2011), et al.
- Empirical methods to address selection issues and skewness of the distribution of medical care
  - Cameron and Trivedi (1986), Pohlmeier and Uhlrich (1995), Deb and Trivedi (1997), Gurmu (1997), Cameron and Johansson (1997).
- Grossman-based dynamic discrete choice models or microsimulation models:
  - Yang, Gilleskie, and Norton (2009); Khwaja (2010)
- Dynamic lifecycle models of joint demand for health and wealth:
  - Edwards (2008), Hall and Jones (2009), Yogo(2009), Hugonnier et al.(2013)

In particular,

Existing models have trouble in explaining:

- Why do individuals spend so much on health care at end-of-life when expected value from health gains are low?
- Why are young and relatively healthy people observed in the top tail?

Conjecture: "change in health" may play a role in determining who comprises the top 5% of medical care consumers.

What the authors do ...

Use a new theoretical model by Kohn and Patrick (2012) to guide the empirical work. Kohn and Patrick (2012) extend the Grossman model in the following way:

- Include "change in health" as a state variable in utility function.
  - Empirically, this is equivalent to including health history in the utility function. This has been done in macroeconomic models by including habit in consumption in utility functions
  - Intuition: large declines in health from habitual levels decrease utility from any observed level of health.
- The health transition function is generalized to allow health to affect the productivity of medical care.
  - Intuition: Co-morbidities may reduce the efficacy of medical care for those in particularly poor health.
- Health depreciation is modeled as a periodic shock rather than a multiplicative rate.
  - Intuition: Should better fit the observation that health declines rapidly at end of life rather than a gradual slope.

Hypothesis of Interest

The authors wish to test the following:

- The greater the decline in health from some habitual level, the greater the demand for medical care, conditional on the state of health.
- The higher the level of lagged health, the smaller the effect of a decline in health on medical care expenditures.
  - medical care has a higher productivity when health state is higher.
- The demands for medical care and other consumption are not separable.

My understanding is that the the 2nd and 3rd points are the assumptions needed in order for the model to predict the 1st point.

To achieve this goal

- Test the theoretical implications of the model by estimating a joint system of semi-parametric expressions for:
  - demand for medical care
  - demand for aggregate consumption
  - health state transition
  - probability of death
- Utilize Conditional Density Estimation (Gilleskie & Mroz, 2004) for each expression, enabling us to:
  - match any moment of the distribution of the variable of interest (i.e., top 5%, 10%, 15% of medical care consumers)
  - allow the marginal effect of a variable of interest to vary over the support of the dependent variable.
- Compare the predictions of this theoretical model with a more Grossman-consistent version, with the goal of matching the upper part of the distribution of medical care consumers.

A 3-state Dynamic Deterministic Optimal Control Model

$$\max_{\{z,m,T\}} LU \equiv \int_0^T e^{-rt} U\left[z(t), H(t), A(t)\right] dt$$

$$\dot{H} = \alpha\left(t, m(t), H(t)\right) \qquad \frac{\partial \alpha}{\partial H}, \frac{\partial^2 \alpha}{\partial H \partial m} > 0$$

Healthy bodies can "heal thyself"

Comorbidities hurt prognosis from medical care

$$\dot{A} = \ddot{H} = \dot{\alpha}$$

$$\dot{R} = rR(t) + w(H(t)) - P_m(t)m(t) - P_z(t)z(t)$$

$$H(0) = H_0 > H_{\min},\ H(T) = H_{\min}$$

$$A(0) \text{ free},\ A(T) < 0$$

$$R(0) = R_0 \geq 0,\ R(T) \geq 0$$

Can neither be born nor die in debt

$$T \leq T_{\max}.$$

Maximum biological lifespan

Implications and Testable Hypotheses

$$\frac{U_H}{\lambda^R(T)} + w_H = g[r + \alpha_H - \tilde{g}]$$

$$g = \frac{\lambda^H}{\lambda^R}$$

Shadow price of health capital

Health decline increases the marginal benefits from health

Health decline decreases the cost of health capital

Inevitable disequilibrium

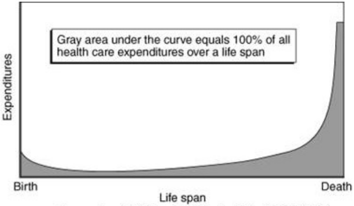
Hypothesis:

The greater the decline in health, the greater the disequilibrium and therefore the greater the demand for medical care

✓ The lower the state of health (older age) the larger the disequilibrium from any decline.

✓ The larger the decline the larger the disequilibrium at any state of health (younger age).

Implications and Testable Hypotheses



Source: Rand, "Living Well At the End of Life" (2003)

Time path for medical care demand

$$\dot{m} = \frac{\overset{+}{\lambda^H}\,\overset{+}{\alpha_m} - \overset{+}{\lambda^R}\,\overset{+}{P_m}}{-\overset{+/-}{\lambda^A}\,\overset{-}{\alpha_{mm}}} + \frac{\left(\overset{+}{\alpha_{mH}}\,\overset{+/-}{\dot{H}} + \overset{+}{\alpha_{mt}}\right)}{-\overset{-}{\alpha_{mm}}} \overset{?}{>} 0$$

Time path for non-medical care demand

$$\dot{z} = \frac{\left[U_{zH}\dot{H} + U_{zA}\left(\alpha_t + \alpha_m \dot{m} + \alpha_H \dot{H} + \alpha_\delta \dot{\delta}\right)\right]}{-U_{zz}}$$

Joint hypothesis:

If the change in health is relevant then the demands for medical care and non-medical care consumption should be estimated jointly.

Transition to Discrete Time

Each period, the individual enters knowing:

- $\mathbf{X_t} \equiv$ vector of exogenous characteristics (in this specification, include income/wealth as exogenous characteristics)
- $H_t \equiv$ endogenous health state
- $\{H_{t-k}, m_{t-k}, z_{t-k}\} \forall k \in \{1, ..., t\} \equiv$ historic values of health state, medical care, and other consumption.

Comment: Why not present a discrete time version of the model?

Empirical Challenges

- Make model match theory.
- Capture the tail of highly skewed distributions.
- Address non-separability of demand for medical and non-medical consumption.
- Account for unobserved heterogeneity.

Expressions for joint demand and health transition

$$\begin{aligned} Z_t^* &= z(H_t, \Delta H_t, M_{t-1}, Z_{t-1}, \mathbf{X}_t, \epsilon_t^Z) \\ M_t^* &= m(H_t, \Delta H_t, M_{t-1}, Z_{t-1}, \mathbf{X}_t, \epsilon_t^M) \end{aligned} \quad (1)$$

where $\Delta H_t = H_t - \frac{1}{p}\sum_{i=1}^{p} H_{t-i}$.

Comment: Note that the authors have not included M_{t-1}, Z_{t-1} in the utility function in a continuous time model.

$$H_{t+1} = h(H_t, M_t, Z_t, \mathbf{X_t}, \epsilon_t^H) \quad (2)$$

Comment: It is important to model why consumers care about ΔH_t. A reduced form utility function with ΔH_t should be used carefully.

Comment:

- Suppose that H_{t+1} measures "visible health" (e.g., whether one has visible illness), and ϵ_t^H is the true health state which consumers are uncertain about.
- Decline in health can be an informative signal about ϵ_t^H. A low ϵ_t^H could imply H_{t+1} is more likely to decline.
- Suppose that an increase in M can improve the true health state. This can explain why ΔH_t can affect one's choice on M and Z even if ΔH_t does not enter the utility function directly. Then, ΔH_t would enter the health transition equation.

$$H_{t+1} = h(\Delta H_t, H_t, M_t, Z_t, \mathbf{X_t}, \epsilon_t^H) \quad (3)$$

Identification

Three sources:

- Exclude lagged consumption and prior health history from the health transition equation.
  - Model implies all effects of historical variables should be captured by demand expressions. Contemporaneous medical and non-medical consumption thusly enter the health transition equation
  - **Comment:** But a more fundamental theory could imply that the exclusion restrictions are not valid.
- The timing of the model.
  - Requires modeling endogenous initial conditions. Use employment history and occupational demands as exclusion restrictions for initial health state, medical care, and consumption
- Some identification is achieved through non-linearities in the model.

Conditional Density Estimation - Gilleskie and Mroz(2004)

Uses a sequence of conditional logit probability functions to approximate the density of the outcome of interest.

- Divide each variable of interest y into K quantiles containing equal numbers of observations in each "cell'.'
- For each interval, the k^{th} interval is defined by $[y_{k-1}, y_k)$ and $y_k = \infty$
- The conditional probability that Y falls into the 1^{st} interval can be expressed as:

$$\lambda(1, x) = p[y_0 \leq Y < y_1 | x] = \int_{y_0}^{y_1} f(y|x)dy \tag{4}$$

Conditional Density Estimation - cont'd

The probability that Y falls into the k^{th} interval can be expressed as:

$$p[y_{k-1} \leq Y < y_k | x] = \int_{y_{k-1}}^{y_k} f(y|x)dy \tag{5}$$

The conditional probability that the dependent variable is observed in the k^{th} interval, given that Y was not observed in intervals one through $k-1$ can be expressed as:

$$\lambda(k, x) = p[y_{k-1} \leq Y < y_k | x, Y \geq y_{k-1}] = \frac{\int_{y_{k-1}}^{y_k} f(y|x)dy}{1 - \int_{y_0}^{y_{k-1}} f(y|x)dy} \tag{6}$$

Conditional Density Estimation - cont'd (3)

The $\lambda(k, x)$ serves as a discrete hazard function. Thus, the probability that Y falls into the k^{th} quantile is given by:

$$p[y_{k-1} \leq Y < y_k | x] = \lambda(k, x) \prod_{j=1}^{j-1} [1 - \lambda(j, x)] \tag{7}$$

Following Gilleskie and Mroz (2004),

- use logit probabilities for the hazard functions.
- interact each covariate with a function of the cell number, $\gamma_k = -ln(K - k)$
- write for each variable H_t, M_t, Z_t and initial conditions; and for each cell $k \in \{1, \dots, K\}$:

$$g^j(k, x) = X^j \beta_1 + X^j \gamma_k \beta_2 + X^j \gamma_k^2 \beta_3 + \epsilon^j \tag{8}$$

$$\lambda^j(k, x) = \frac{e^{g^j(k,x)}}{1 + e^{g^j(k,x)}} \tag{9}$$

Discrete Factor Random Effects

- The model permits permanent and time-varying heterogeneity to affect each expression without imposing distributional assumptions on the error term.
- The authors follow Heckman and Singer (1984) and Mroz (1999) by approximating the joint distribution of both types of heterogeneity with a step function.
- Each epsilon term can be decomposed into three components:

$$\epsilon_t^j = \mu^j + \nu_t^j + e_t^j \qquad \forall j \in z, m, H; \quad \forall t \in \{1, \dots, T\} \tag{10}$$

where:

- μ captures permanent heterogeneity for each expression
- ν represents the time-varying component
- e_t^j represents the remaining i.i.d. Type 1 Extreme Value Error for the logit hazard probabilities.

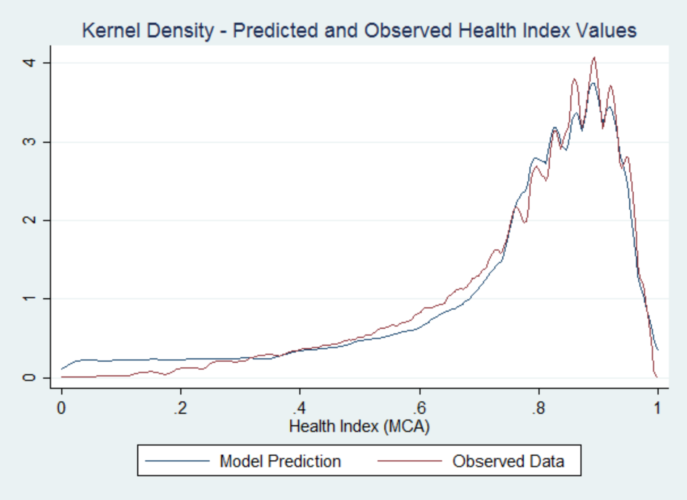
ν terms in the initial conditions are restricted to zero.

The Likelihood Function

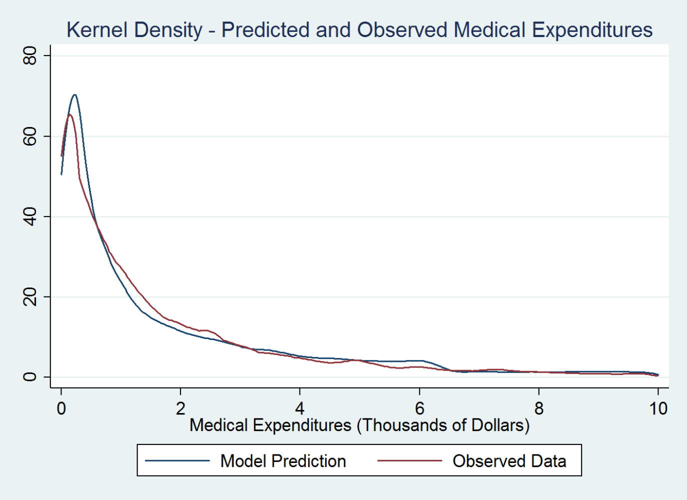
$$
\begin{aligned}
L_i(\Theta,\psi,\pi) \quad = \quad & \textstyle\sum_{k=1}^{K} \pi_k \Big\{ \prod_{j_f=1}^{J_f} Pr(H_1 = j_f|\mu_k^F)^{\mathbf{1}(H_1=j_f)} \prod_{j_s=1}^{J_s} Pr(H_2 = j_s|\mu_k^F)^{\mathbf{1}(H_2=j_s)} \\
& \textstyle\prod_{j_m=1}^{J_m} Pr(M_2 = j_m|\mu_k^{MI})^{\mathbf{1}(M_2=j_m)} \prod_{j_z=1}^{J_z} Pr(Z_2 = j_z|\mu_k^{ZI})^{\mathbf{1}(M_2=j_z)} \\
& \textstyle\times \prod_{t=2}^{T} \sum_{l=1}^{L} \psi_l \Big[\prod_{j_m=1}^{J_m} Pr(M_t = j_m|\mu_k^M, \nu_{lt}^M)^{\mathbf{1}(M_t=j_m)} \\
& \textstyle\prod_{j_z=1}^{J_z} Pr(Z = j_z|\mu_k^Z, \nu_{lt}^Z)^{\mathbf{1}(Z_t=j_z)} \prod_{j_h=1}^{J_h} Pr(H_{t+1} = j_h|\mu_k^H, \nu_{lt}^H)^{\mathbf{1}(H_{t+1}=j_h)} \\
& Pr(death|H_t, M_t, Z_t, \mu_k^D, \nu_{lt}^D)^{\mathbf{1}(death)}(1 - Pr(death|H_t, M_t, Z_t, \mu_k^D, \nu_{lt}^D)
\end{aligned}
\tag{11}
$$

where Θ is the vector of parameters on the covariates for each expression, the π_k terms are the mixing probabilities on the permanent heterogeneity, and the ψ terms are the mixing probabilities on the time-varying heterogeneity.

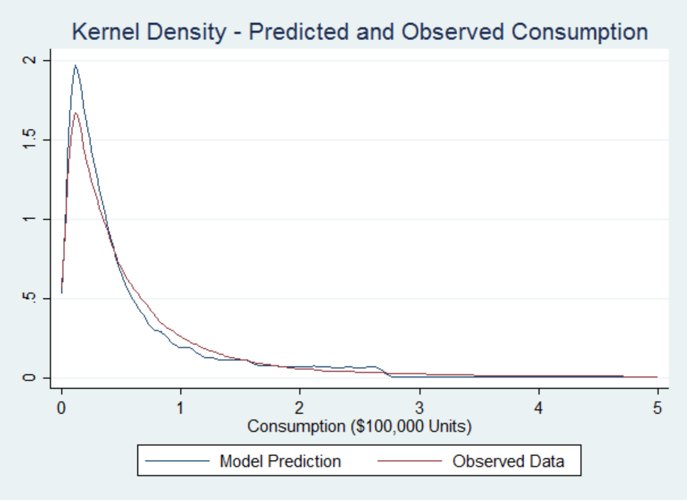
Predicted & Observed Distribution of Health Index Values



Predicted & Observed Distribution of Medical Care Exp.



Predicted & Observed Distribution of Consumption



The Effect of Change in Health

- A 10% decline in health, holding contemporaneous health constant, increases the probability that the individual is observed in the top 5% by 22%.
  - A 10% decline in health increases the probability of being in the top 10% of medical care consumers by 18%
- A similar decline in health *and* lagged health increases the probability that in individual is observed in the top 5% by 44%.
- The primary driver of medical care consumption is contemporaneous health, but habituated levels of health do matter.

Comment: How about the effect of **increasing** in health? If I follow my suggested modification earlier (i.e., ΔH changes consumer's expectation about how H_{t+1} evolves over time, and that utility is concave in H, I would expect the marginal effect is asymmetric.

Other Key Results

- The demand for consumption and medical care are not separable.
  - The estimates for the parameters for the unobserved heterogeneity are significant at the 1% level.
- A 10% decline in health (current and historic) leads to a 5.7% reduction in consumption.
- A 10% decline in health, holding contemporaneous health constant, increases mean consumption by 6%, but the effect is heterogeneous.

Comment: The last two findings are very useful for guiding us what features a better structural model needs.

Model Prediction vs. Observed Data

Table : Market Level Results: Average Minutes of Moderate Exercise, Single Women

Summary Statistics of Top 5% of Medical Care Consumers			
Variable	**Observed Data**	**Preferred Model**	**Grossman**
Age	70.70	**70.77**	73.39
Health Index	0.636	**0.594**	0.35
Change in Health	-0.078	**-0.073**	-0.004
Years of Schooling	12.42	**12.13**	11.67
Female	0.635	**0.646**	0.611
Married	0.575	0.487	**0.546**
Income	0.497	0.460	0.432
Lagged Med. Care	0.037	**0.033**	0.055

Matching Individuals In the Top 5 %

For a benchmark, a log-linear regression with the same covariates yields an R^2 of 0.06

- This model generates a match rate of 12.5% in the top 5% of medical care consumers.
- The Grossman consistent model generates a match rate of 8.5% of the top 5%
  - The theoretical innovation of change in health represents a 47% improvement in matching the top 5%.
- The estimated model generates a match of 36% over the top quartile.
- It generates a match of 41% over the top quartile when the model accurately predicts health index within one contiguous quartile.
  - The theoretical model does not represent a significant improvement over Grossman here. The improvement is greatest in the upper tail of the distribution, as the authors expected.

Conclusion

- Change in health matters. When a person's health deteriorates, the estimates indicate that 1/3 of the increased demand for medical care is due to health history.
- Given what we know about the skewness of medical care distributions, income distributions, and health problems – semi-parametric estimators such as these are useful in forming expectations over future expenditures.
  - Predicting poor health is one thing - predicting who will spend the resources to try to improve their health is another.
  - The search for explanatory variables needs to continue. As variables with greater predictive power are identified, tools such as joint CDE estimation with unobserved heterogeneity can glean the most information from them.

Comment: Might want to consider other conditional distribution estimation methods as well (see Jones, Lomas and Rice, 2014, presented in this conference).