Long-term Health Spending Persistence among the Privately Insured: Exploring Dynamic Panel Estimation Approaches

UNIVERSITY OF MIAMI

Sebastian Calonico, PhD
Assistant Professor of Economics
University of Miami

Co-authors
Richard Hirth, PhD, University of Michigan
Teresa Gibson, PhD, Arbor Research Collaborative for Health
Helen Levy, PhD, University of Michigan
Jeffrey Smith, PhD, University of Michigan
Anup Das, BA, University of Michigan

Introduction

- We study long-term spending patterns in the under-65 population in the US.
  - Target of the Affordable Care Act (ACA) coverage expansions.
- We use 6 recent years of claims (2003-2008) to examine trends in persistence of spending.
  - Relevant to improve coverage and control cost.
- Contribution to the literature in terms of:
  - Timeliness.
  - Length of follow-up.
  - Sample size.

Introduction

- Persistence in health spending has implications for:
  - Insurers concerned about adverse selection.
  - Regulators attempting to detect and manage risk-selection by insurers.
  - Identification of effective cost-control measures.
  - Distributional impact of out-of-pocket spending under high deductible plans.
- Predicting spending patterns can help design appropriate insurance products and public policies to ensure adequate coverage. Examples:
  - Low persistence: healthy individuals may opt out of ACA-mandated coverage and face risk of unexpected short-term spending spike.
  - High persistence: risk selection could threaten the functioning of health insurance exchanges.
- Success of cost control measures depends on ability to identify people likely to be (or become) high spenders and modify their care trajectories.

Introduction

- Hirth et al. (2015) describes long-term concentration and persistence of spending in the U.S. privately-insured, under-65 population.
- Key Findings:
  - Fairly high persistence of spending at both ends of the spending distribution. Over 6 years:
  - Low end —› 70% enrollees never had annual spending in the top 10%

    —› Bottom 50% of spenders accounted for less than 10% of spending.
  - High end: those in top 10% almost as likely (34%) to be in the top 5 years later as 1 year later (43%).
  - Many comorbid conditions retained predictive power even 5 years later.

Objective

- We build upon these findings, exploring econometric strategies to analyze persistence in spending.
- Highlight a range of modeling options available to researchers for different contexts.
- Focus on:
    - Types of questions that can be answered.
    - Types of inferences that could be drawn using each class of model.
- Models for discrete and continuous dependent variables.
- It addresses three main limitations found in the literature:
    - Representative data sets with limited follow-up (MEPS, a 2 year panel).
    - Focus on single employer or single insurer datasets, or involved Medicare (>65) population.
    - Use of older data (1980s or 1990s).

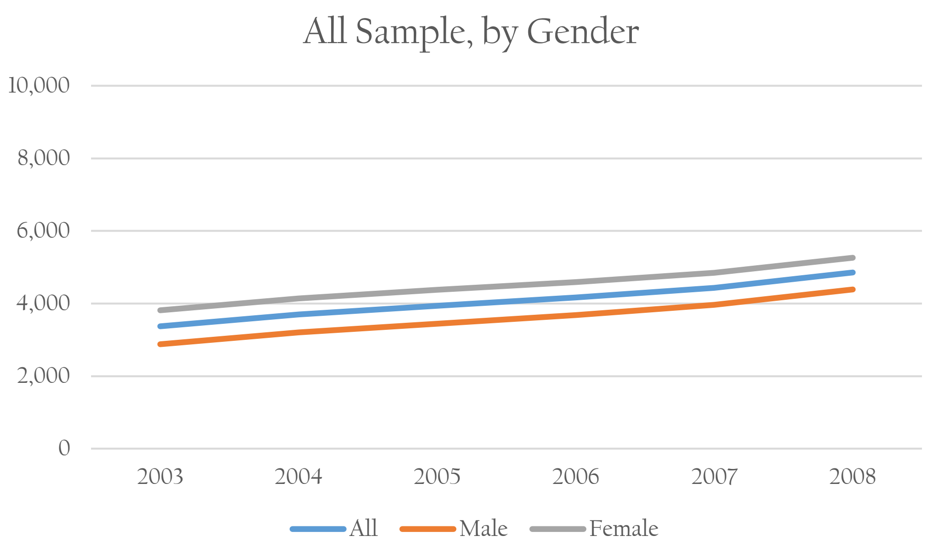
Data

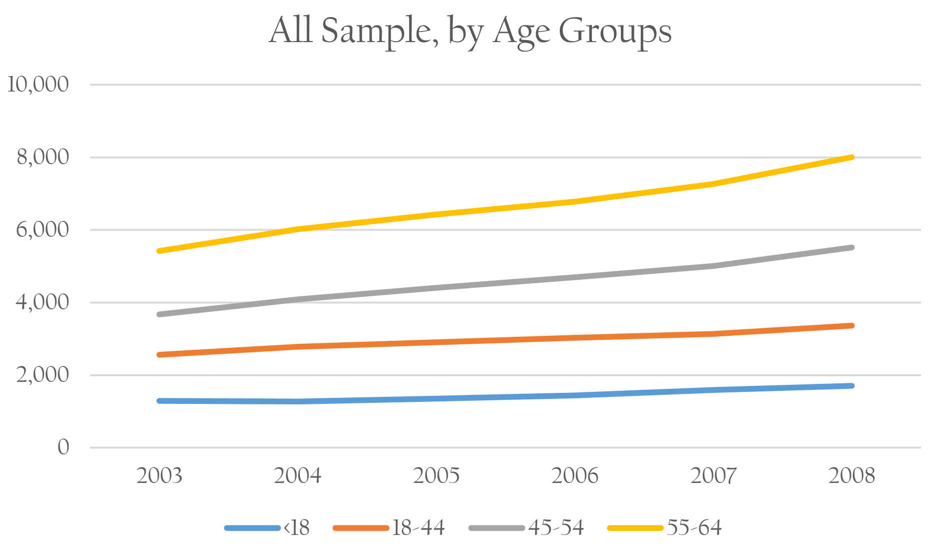
- 2003-2008 health care claims in MarketScan Research Database, collected by Truven Health.
  - ~2.5 million people can be followed for the entire 6 year period 2003-2008.
- Employees and dependents receiving coverage through 100+, mainly self-insured, medium/large firms.
  - ✓ All carve-outs (e.g., prescription drug, mental health).
  - ✓ Claims satisfying the deductible and falling below the threshold (if deductible imposed).
  - ✗ Out-of-plan spending (over the counter drugs) and patient-borne costs (travel to appointments).
- Also includes:
  - ✓ Demographic characteristics.
  - ✓ Psychiatric diagnosis: depressions, anxiety disorders, schizophrenia, etc.
  - ✓ Comorbidities: heart failure, chronic pulmonary disease, diabetes, tumors, dementia, etc.

Data: Summary Statistics

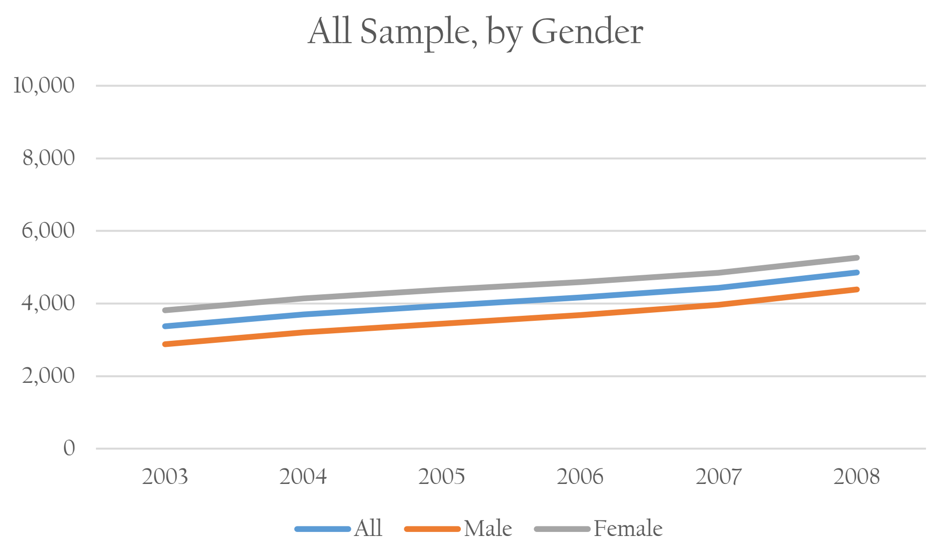
	All	Male	Female	<18	18-44	45-54	55-64
Age	35.2	34.4	35.9	6.7	28.1	44.7	53.9
Gender (Male)	47%	-	-	51%	47%	45%	46%
Urban Area Indicator	78.60%	79%	78%	80%	79%	78%	77%
Region							
Northeast	9%	9%	9%	10%	9%	9%	8%
North Central	29%	29%	28%	27%	27%	28%	32%
South	34%	33%	36%	32%	35%	36%	35%
West	27%	27%	27%	29%	28%	27%	24%
Median HH income at zip code (thousands $)	49.3	49.7	48.9	51.2	49.4	49.2	47.8

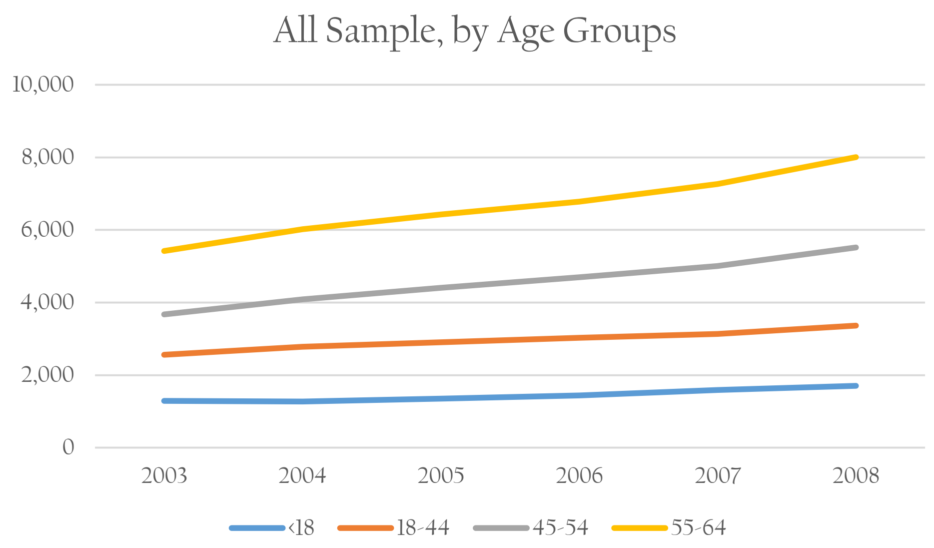
Data: Evolution of Medical Expenses (in 2008 US dollars).

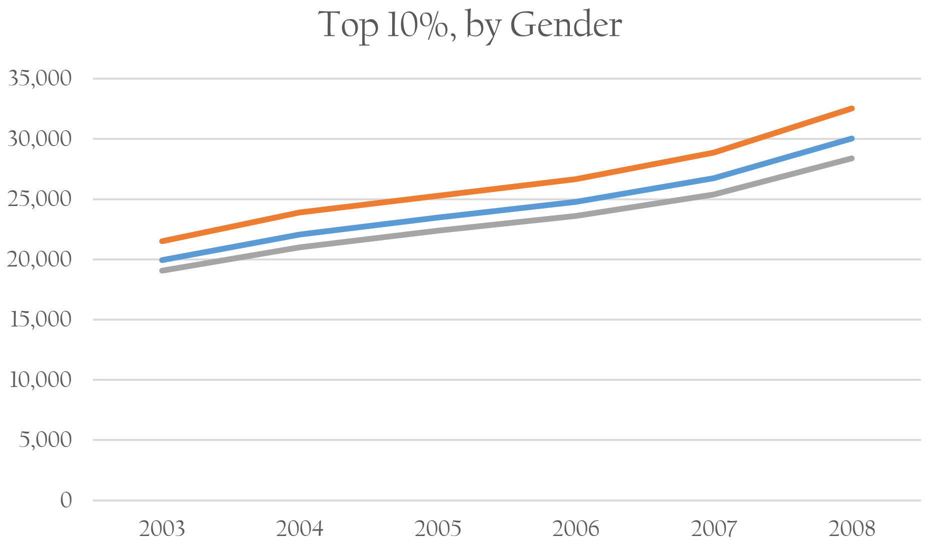


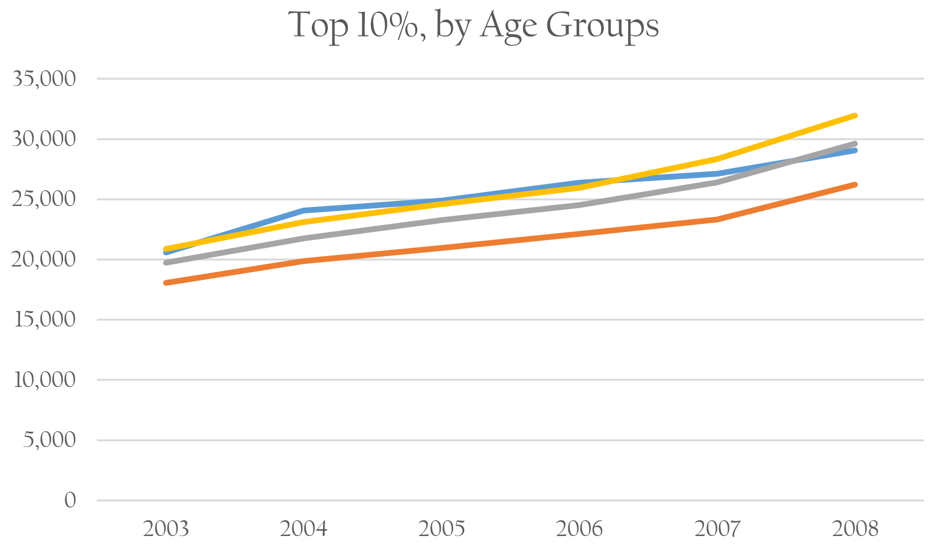


Data: Evolution of Medical Expenses (in 2008 US dollars).









Data: Persistence of High Spending Across Years

High (Top 10%)	High 1 year later	High 2 years later	High 3 years later	High 4 years later	High 5 years later
2003	43.4%	39.6%	37.7%	35.9%	34.4%
2004	44.0%	40.2%	37.9%	36.3%	
2005	44.3%	40.3%	38.2%		
2006	44.9%	40.9%			
2007	45.4%				

Fairly large persistence of spending at the high end of the distribution.

- Those in top 10% almost as likely (34.4%) to be in the top 10% five years later as one year later (43.4%).

Modeling Health Expenditures: Discrete Outcomes

- Models for discrete dependent variables.
- Useful to analyze consumption patterns when spending only reported categorically.
- Or based on discrete utilization measures.
  - Hospitalizations.
- Less subject to concern about influence of outliers.
  - Relevant given highly skewed nature of health spending.
- We focus on dynamic discrete choice models.

Probability of High Health Expenditure

- Probability of being in the highest decile of health expenditure with dynamic RE probit model:

$$Pr(y_{it} = 1 | y_{i,t-1}, \ldots, y_{i0}, x_{it}, \mu_i) = \Phi(\rho y_{i,t-1} + \gamma x_{it} + \mu_i)$$

- Wooldridge (2005): under certain distributional assumptions the likelihood function has same structure as standard RE probit model.
- Once we estimate the parameters we can calculate average partial effects (APE):

$$\widehat{APE} = \frac{1}{n} \sum_{i=1}^{n} \Phi\left(\hat{\alpha}_0 + \hat{\alpha}_1 y_{i0} + \hat{\gamma} x_{it} + \hat{\rho} y_{i,t-1}\right)$$

- We can also compute changes or derivatives with respect to x_{it} or $y_{i,t-1}$.
- Standard errors via Delta Method.
- We focus on $\mathrm{E}[Pr(y_{it} = 1 | \boldsymbol{y_{i,t-1} = 1}, x_{it} = x)]$ and $\mathrm{E}[Pr(y_{it} = 1 | \boldsymbol{y_{i,t-1} = 0}, x_{it} = x)]$

Probability of high health expenditure: **Key Findings**

- Persistence by demographic characteristics generally lower than persistence by comorbidities.
- Presence of a comorbid condition:
  - More likely to be in the top 10% in year *t* regardless of whether they were in the top 10% in *t -1*.
  - Also more likely to be in top 10% at *t* if they were also in *t-1*.
  - Example: APE of chronic pulmonary disease on being in the top 10% = 0.21 conditional on $\boldsymbol{y_{i,t-1} = 1}$ vs. 0.14 conditional on $\boldsymbol{y_{i,t-1} = 0}$.
  - On average: being in top 10% increases the probability of remaining in the top 10% by 5 to 11 p.p.
- Most likely to be and remain in the top 10% are those with cardiovascular disease or rheumatologic conditions, especially youngest (less than 18) and oldest (55 through 64) groups.

Modeling Health Expenditures: Continuous Outcomes

- Models for continuous dependent variables.
  - Expenditures.
- Could help make projections about actual insurance program outlays.
- Understand the role of observed characteristics versus previous spending.
- Error Components Models:
  - Temporal dependence modeled via AR and MA terms.
- Linear Panel Data Models:
  - Static Fixed Effects and Random Effects models.
  - Dynamic panel models: lagged expenditures as independent variables.

Error Component Models

- Study the autocorrelation structure of the health cost process.
- To understand the intertemporal persistence of health care costs: $h_{it} = \beta x_{it} + \varepsilon_{it}$
- Decomposition of the residual term ε_{it} as the sum of a permanent (α_i) and a transitory (v_{it}) component:

$$\varepsilon_{it} = p_t \alpha_i + \lambda_t v_{it}$$

$$\alpha_i \sim (0, \sigma_\alpha^2) \quad v_{it} \sim (0, \sigma_{vt}^2)$$

- p_t, λ_t: factor loadings. Allow variances to change over time in a way that is common across individuals.
- Objective:
  - Identify roles played by permanent and transitory shocks.
  - Understand how these roles change over time.

Error Component Models: Implementation

- First stage: estimate β by regressing health costs on demographic and health insurance variables that households can use to forecast future health costs.

- Second stage: estimate the covariance matrix of the residuals, fit the model using GMM approach.

	2003	2004	2005	2006	2007	2008
2003	1.32	0.49	0.43	0.39	0.36	0.33
2004	0.66	1.35	0.49	0.43	0.39	0.36
2005	0.57	0.67	1.35	0.49	0.43	0.39
2006	0.52	0.58	0.67	1.37	0.49	0.43
2007	0.48	0.53	0.58	0.68	1.37	0.49
2008	0.44	0.49	0.53	0.59	0.68	1.39

- Different way

  - AR(1) p
  - ARMA(1,1) process: $v_{it} = \rho v_{i,t-1} + \theta u_{i,t-1} + u_{it}$

Error Component Models: Key Findings

VARIABLES	All		Male		Female	
	AR(1)	ARMA(1,1)	AR(1)	ARMA(1,1)	AR(1)	ARMA(1,1)
ρ	0.154***	0.669***	0.154***	0.648***	0.154***	0.690***
	(0.001)	(0.006)	(0.001)	(0.006)	(0.001)	(0.007)
σ_{α}^2	0.444***	0.370***	0.449***	0.383***	0.440***	0.357***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.003)
σ_{vt}^2	0.873***	0.948***	0.865***	0.932***	0.880***	0.963***
	(0.001)	(0.003)	(0.001)	(0.002)	(0.001)	(0.003)
σ_{u}^2	0.408***	0.824***	0.399***	0.807***	0.416***	0.840***
	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)
θ		-0.442***		-0.427***		-0.456***
		(0.004)		(0.004)		(0.004)

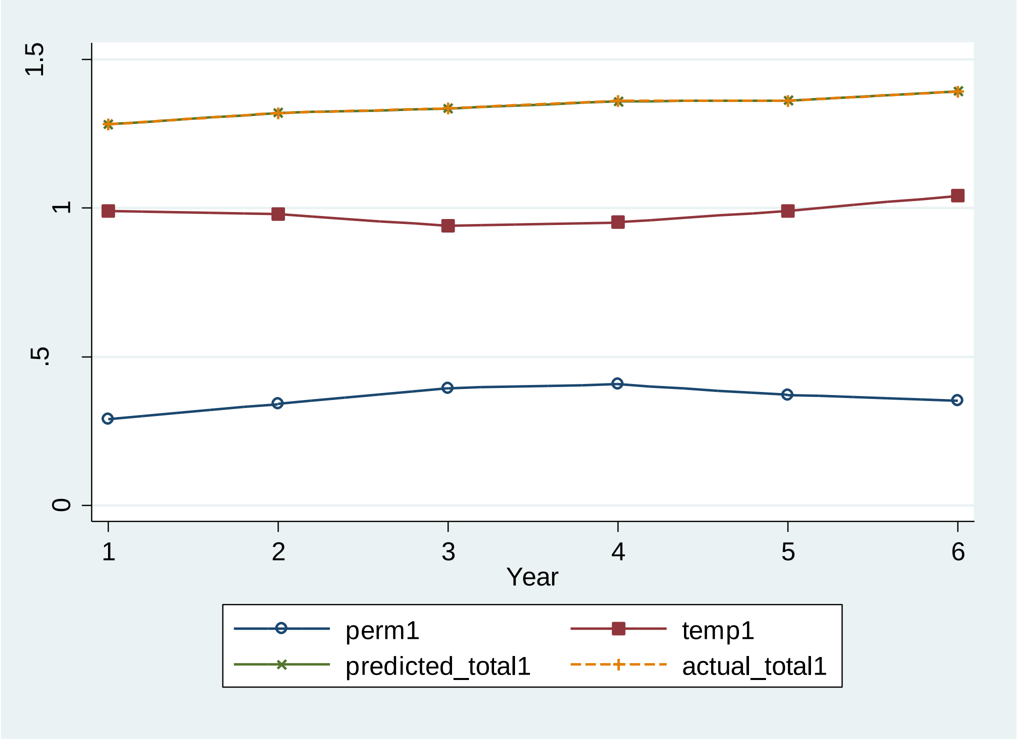
- Both models capture important features of expenditure dynamics: time varying parameters and serial correlation of the transitory shocks.
- Low persistence in the transitory shock.
- ρ and θ parameter very similar in magnitude across groups (also by age).
- Factor loadings (not reported): relatively constant transitory and permanent variances over time.
- Pending: Wald test to assess model fit.

Error Component Models

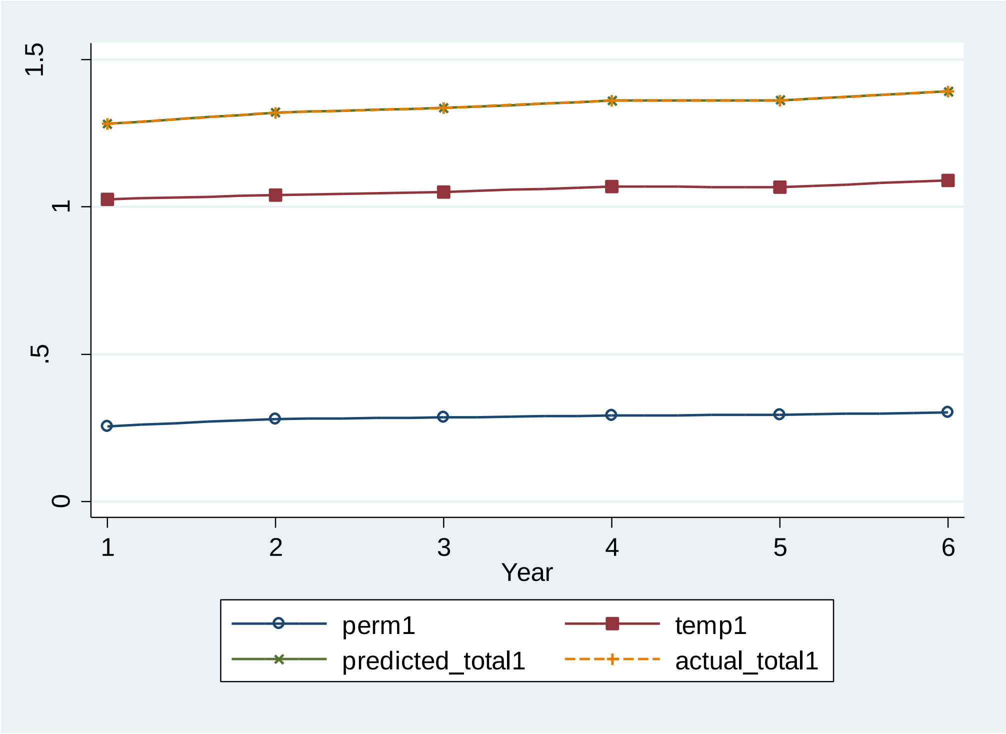
- Decomposition of observed variability between predicted transitory and permanent components:

$$\varepsilon_{it} = p_t \alpha_i + \lambda_t v_{it}$$

AR(1)



ARMA(1,1)



- Larger contribution of the transitory component on observed variability.

Static Linear Panel Models

$$h_{it} = \mu_i + \beta x_{it} + \varepsilon_{it}$$

- h_{it} health expenditures for individual i at time t
- x_{it} vector of explanatory variables
- μ_i unobserved individual characteristics
- ε_{it} is the residual term

- Two main specifications: fixed (FE) and random effects (RE) panel data models.
- Best suited for answering a "cross-sectional" question: what predicts spending level?
- Could help make projections about actual insurance program outlays.
- Useful if we know individual characteristics (demographics, prior medical conditions), but not prior spending history.

Static Linear Panel Models: **Key Findings**

- FE models: estimates based on within person changes over time
    - e.g., regional effects are based on enrollees who re-locate.
- Most comorbid conditions are associated with higher spending.
- The magnitudes of the higher spending associated with the psychiatric conditions tend to be lower than those associated with most physical comorbidities.
- Hausman specification test strongly rejects the hypothesis that individual-level effects are adequately modeled by RE model.

Dynamic Linear Panel Models

$$h_{it} = \delta h_{i,t-1} + \beta x_{it} + \varepsilon_{it}$$

- h_{it} health expenditures for individual i at time t,
- $h_{i,t-1}$ health expenditures for the same individual i in the previous period, $t - 1$,
- x_{it} vector of explanatory variables including time-varying characteristics.
- We also consider also a simple version of the model with no covariates.

Dynamic Linear Panel Models

VARIABLES	All		Male		Female	
Medical Spending (in t-1)	0.398***	0.342***	0.386***	0.339***	0.407***	0.345***
	(0.000304)	(0.000328)	(0.000451)	(0.000487)	(0.000411)	(0.000444)
Initial Medical Spending (in 2003)	0.130***	0.101***	0.156***	0.125***	0.110***	0.0833***
	(0.000342)	(0.000347)	(0.000545)	(0.000553)	(0.000438)	(0.000442)

- Initial (baseline) spending and lagged spending are substantive predictors of future spending.
- Full sample (no covariates): $1 increase in baseline predicts $0.13 higher spending, while $1 increase in lagged spending predicts $0.40 higher.
- With covariates: magnitudes decrease slightly (2nd columns for each group).
- Females: slightly lower short term persistence than males, but notably lower long term persistence.
- Age groups: persistence greatest for those under age 18 and smallest for 18-44 years.

Conclusions

- We study long-term spending patterns in the under-65 population in the US.
- Using claims from 2003-2008 to examine trends in persistence of spending.
- Contribution to the literature in terms of timeliness, length of follow-up, and sample size.
- Highlight a range of modeling options available to researchers for different contexts.
- Focus on the types of questions that can be answered and types of inferences that could be drawn using each class of model.
    - Models for discrete and continuous dependent variables.

Conclusions

- High expenditures: dynamic discrete choice models reveals relevant role of prior spending category and comorbid conditions.
- Intertemporal persistence of health spending shocks. Using error components models we find
  - Large role for temporary components.
  - Similar patterns among age groups.
  - Factor are relatively stable over time.
- Linear panel data models:
  - Static: useful to predict spending based on observed characteristics.
  - Dynamic: estimates of persistence only slightly affected by the inclusion of covariates.