Long-Term Health Spending Persistence among the Privately Insured in the US*

Richard A. Hirth,† Sebastian Calónico,‡ Teresa B. Gibson,§ Helen Levy,◊ Jeffrey Smith^ and Anup Das±

†University of Michigan
(rhirth@umich.edu)
‡University of Miami
(scalonico@bus.miami.edu)
§Truven Health
(tbgibson1@gmail.com)
◊University of Michigan
(hlevy@umich.edu)
^University of Michigan
(econjeff@umich.edu)
±University of Michigan
(anupdas@umich.edu)

*Submitted September 2015.

The authors gratefully acknowledge funding from the Agency for Healthcare Research and Quality under grant number AHRQ 1-R01-HS-017706-01-A1, the assistance of their project officer Melford Henderson and comments from seminar participants at the University of Colorado Denver.

Keywords: health care costs, health insurance.

JEL classification number: I10.

Abstract

There is little current information regarding the long-term persistence of health spending in the United States, in particular among the population aged under 65 (pre-Medicare eligibility). We describe and model the extent of persistence over a six-year period (2003–08) using medical and pharmacy claims for over 3 million employees, retirees and dependants derived from the Truven Health MarketScan database. Overall, substantial persistence in spending exists, particularly at the extremes of the distribution and for pharmaceutical spending. Error components models are estimated to separate transient from persistent variation in spending, and dynamic probit models are estimated to assess the predictive power of demographic and co-morbid conditions and prior high spending in determining the likelihood of future high spending. A better understanding of the persistence of health spending can inform the selection and evaluation of appropriate interventions to address high costs, and can help forecast the likelihood and severity of adverse selection in public and private programmes.

Policy points

- This study adds substantially to knowledge about the long-term persistence of health spending and the predictors of persistently high spending in the under-age-65 population with private insurance.
- The incidence of health spending, the potential for adverse selection in insurance markets and the optimal design of insurance features such as annual deductibles depend on the extent to which high spending persists over time at the individual level.
- Interventions to address high health spending would differ depending on the extent to which it is expected to persist over time.

I. Introduction

Given the low probability of high-cost health events over a short period of time, it is not surprising that the spending distribution is highly skewed within a single year. However, long-term spending patterns in the United States, particularly in the under-65 population which is the target of Affordable Care Act (ACA) coverage expansions, remain poorly understood. Existing studies based on broadly representative data sets, such as the Medical Expenditure Panel Survey (MEPS), have limited follow-up (the MEPS is a two-year panel). As a result, research on longer-term spending patterns has generally relied on single-employer or single-insurer data sets or has involved the Medicare population (aged 65 and over). In addition, most existing studies rely on data from the 1980s or 1990s.

Knowledge of the extent and correlates of persistence in health spending can inform the design of appropriate insurance products and public policies to ensure adequate coverage. First, such information is highly relevant to insurers who are concerned about adverse selection. For example, if high spending often arises quickly, low current expenditures will not strongly signal low future expenditures. Therefore, currently healthy individuals may hesitate to opt out of ACA-mandated coverage because of the risk that an unexpected short-term spending spike may occur before they are able to get coverage in the next open enrolment window. Conversely, if persistence is relatively high, risk selection could threaten the functioning of health insurance exchanges established under the ACA as a marketplace in which individuals without employer group coverage can purchase insurance and obtain income-based premium subsidies. Second, understanding which individuals are at greatest risk of persistently high spending can inform regulators attempting to detect and manage risk selection by insurers. Third, given the concentration of spending among a small percentage of patients, the success of cost-control measures strongly depends on the ability to identify people likely to become perpetually high spenders and modify their care trajectories. Moreover, the evaluation of interventions targeting high spenders (sometimes called 'hot spotters') depends critically on an accurate understanding on whether high spenders would have become low spenders over time even in the absence of any intervention. Fourth, the extent of household financial risk and the long-term incidence of medical spending depend strongly on the persistence of high spending. The rapid increase in the prevalence of high-deductible health plans in the US illustrates this potential impact. If high spending is transient, an individual might exceed the deductible in a particular year but retain the ability to diversify such risk over time. Conversely, persistent high spending can leave some individuals effectively underinsured over time under a high-deductible plan. Reduced cost sharing for high-value services used by chronically ill people could help address this financial risk and discourage the underuse of clinically important services that could result from high cost-sharing burdens.

Hirth et al. (2015) described the long-term concentration and persistence of spending in the US privately-insured, under-age-65 population, as well as cross-sectional (baseline) correlates of different patterns of spending dynamics. Their first key finding was fairly high persistence at both ends of the spending distribution. At the low end, over a six-year period, 69.8 per cent of enrollees never had annual spending in the top 10 per cent of the distribution and the bottom 50 per cent of spenders accounted for less than 10 per cent of total spending. At the high end, those in the top 10 per cent were almost as likely (34.4 per cent) to be in the top 10 per cent five years later as one year later (43.4 per cent). A second key finding was that many co-morbid conditions measured at baseline retained much of their predictive power even five years later.

This paper builds on these descriptive analyses and extends them in two main directions. First, we include an important part of the population aged 65 and over that has not been extensively studied – namely, those with employer-provided retiree supplements to standard Medicare coverage. Second, we explore more sophisticated econometric strategies to analyse persistence of health spending. The primary objective of the econometric analyses is not to test the relative performance of models in our particular data set (most of the models are not nested in a way that would readily facilitate such comparisons), but rather to demonstrate several modelling options that could each be appropriate in particular contexts. Such a demonstration will highlight the types of questions that can be answered and the types of inferences/interpretations of findings that could be drawn using each class of model.

Our primary econometric analyses employ two broad types of dynamic panel models. First, we consider error components models such as those in van Vliet (1992), designed for continuous dependent variables (for example, expenditures), in which the temporal dependence is modelled via autoregressive and moving-average terms within a composite error. Such models are of primary interest when a researcher (or an enrollee, insurer, employer or policymaker) wants to make projections about actual insurance programme outlays. The second type of model is designed for discrete dependent variables (for example, expenditure categories). In particular, we use dynamic discrete choice panel data models.[1] Such models are of primary interest when a researcher (or an enrollee, insurer, employer or policymaker) wants to draw more qualitative inferences about spending levels and their dynamics over time. They are also of interest for use in data sets where spending is only reported categorically, or in analyses of health care consumption patterns based on discrete utilisation measures (for example, hospitalisations). In addition, given the highly skewed nature of most health spending data, discrete models are less subject to concern over the influence of outlier cases on the estimates.

Section II reviews the related literature. Section III describes the data used in this study. Results are reported in Sections IV and V, with the dynamic modelling appearing in Section V. Finally, Section VI concludes.

II. Related literature

Several studies of spending persistence focus on the population aged 65 and over, using data from the traditional, fee-for-service Medicare programme. Garber, MaCurdy and McClellan (1998) found considerable turnover at the top of the spending distribution. Among enrollees who were in the top 5 per cent in one year, 15.2 per cent remained in the top 5 per cent the following year and only 8.8 per cent remained in that category two years later. However, expenditure growth was concentrated among the highest spenders. Rettenmaier and Wang (2006) used Medicare data from 1974 to 1997 to estimate a dynamic panel Tobit model, concluding that an additional dollar of spending in the prior year resulted in $0.19 higher spending in the current year. Finally, Riley (2007) documented time trends in spending persistence among Medicare enrollees from 1975 to 2004, with persistence increasing until the mid 1990s and then decreasing somewhat thereafter.

Other studies focused on privately-insured individuals or the general population. Using data from a single employer, Eichner, McClellan and Wise (1997) found that employees in the top decile of spending in 1989 spent over eight times the average in that year. While their spending declined in the subsequent two years, it remained high (about five times the average spending in 1990 and

[1]Honoré and Kyriazidou, 2000; Wooldridge, 2005.

three times the average spending in 1991). Using data from one health maintenance organisation (HMO), Chapman (1997) found that of those in the top 5 per cent of the 1989 spending distribution, 19 per cent remained in the top 5 per cent in 1990 and 14 per cent remained in the top 5 per cent in 1991, which exceeded the persistence seen in Garber et al.'s study of the Medicare population. Cohen and Yu (2012) used data on the non-institutionalised US population of all ages from the 2009–10 Medical Expenditure Panel Survey to analyse spending persistence over two years. They found that 40 per cent of those in the top decile of spending in 2009 remained in the top decile in 2010, somewhat higher than the one-third estimated by Monheit (2003) in a similar analysis of MEPS data from 1996–97. Finally, using an earlier sample from the same data set we use (MarketScan), Pauly and Zeng (2004) reported some limited information on the persistence of total spending (probability of remaining in the top 20 per cent of the spending distribution), finding that 46 per cent of those in the top quintile in 1994 remained in the top quintile in 1998.

A recent, relevant non-US study by Kohn and Liu (2013) used the British Household Panel Survey to examine the persistence of health care utilisation over an 18-year period among people aged 16 and over. The primary findings were that past use predicted future use even after controlling for health and other characteristics, past utilisation was more predictive of future utilisation at older ages and lower health status, and baseline utilisation retained some predictive power throughout the follow-up.

III. Data

The data for this project are health care claims and enrolment data for the years 2003–08 in the Truven Health MarketScan Research Database. Using six years of claims for a large sample of enrollees allows us to improve upon previous studies in terms of timeliness, length of panel and sample sizes. This study received an Institutional Review Board (IRB) exemption through the University of Michigan IRB due to the use of secondary data.

We use these data to examine trends in persistence of spending for commercially-insured individuals in the US. MarketScan represents the health care experience of employees and dependants receiving health insurance coverage through over 100, mainly self-insured, medium and large firms. These individuals have higher income, on average, than those not receiving insurance through an employer (for example, uninsured or Medicaid beneficiaries), so the results may not generalise to the entire under-65 population. However, because over 60 per cent of the under-65 US population receive insurance through an employer, the represented population is very significant in its own right and is the locus of many important interventions in health insurance design (for example, value-based insurance design) and cost containment methods (for example, disease management programmes). The number of individuals in the database rose from 8 million in 2003 to 41 million in 2008. Enrolment was distributed broadly across the country, with employees in all 50 states and each of the four Census regions having at least 6.5 million covered employees or dependants in 2008. The South was most heavily represented (38.8 per cent). Comparisons of the MarketScan data with estimates for the US population from the Medical Expenditure Panel Survey and the Kaiser Family Foundation State Health Facts reveal that the age distribution of the enrollees in the MarketScan data is similar to that in the population with employer-sponsored coverage. Comparisons with Census data reveal that the gender distribution of the MarketScan data is similar to the distribution of individuals with employer-sponsored primary or supplemental insurance.[2] Claims include all covered services (i.e. inpatient and outpatient care, prescription drugs and mental health services). Out-of-plan spending for items such as over-the-counter drugs and patient-borne costs such as travel to appointments is not represented. We follow more than 3 million enrollees for the entire six-year period from 2003 to 2008. If a deductible is imposed, claims satisfying the deductible and falling below the deductible threshold are included in the database. Spending has been adjusted to 2014 dollars using the personal consumption expenditure (PCE) price index. Firm identifiers were not available in the releasable data set.

Not surprisingly given the mobility of workers in the US and enrolment tied to employment at a given firm, attrition is common. For example, in the under-65 population, about two-thirds of the 2003 sample could not be followed for the entire six-year period. Some attrition arises for reasons not likely to be endogenous to health spending, such as censoring due to employers no longer providing data to MarketScan (i.e. the entire group exits rather than a self-selected subset of individuals within the group) or children ageing out of dependant status. Other exits, such as death, retirement (without continued coverage), loss of employment or taking a job with another employer, may be endogenous to health spending. Therefore, examining persistence of spending among those workers who do not exit for such reasons could cause biased estimates of spending persistence in the entire population of employees and dependants. Prior analyses of attrition by Hirth et al. (2015) in this sample concluded that the relationship with spending was relatively weak (while higher-spending employees were somewhat more likely to exit than employees with lower spending, higher-spending dependants were less likely to exit than dependants with lower spending, leaving little overall relationship between spending and exit). Although it is, by definition, not possible to examine the persistence of spending post-attrition, had there been a stronger relationship with the level of spending, concerns about such biases would have been heightened. In addition, similar results are obtained when we look at those persistently high spenders (over the first three years of the sample) in terms of attrition during the second three years.

Basic summary statistics are presented in Table 1 for the entire sample, and also grouped by sex and age in 2003 (under 25 years old, aged 25–64, and 65 and older). The 65-and-over population primarily represents individuals who receive a Medicare supplement policy through a firm contributing data

[2]Thomson Reuters, 2007; McKellar et al., 2012.

to MarketScan, but their spending represents the total amount paid by Medicare, the supplement and any out-of-pocket obligation.

[Table 1 about here]

Several notable spending trends emerge. Females had higher spending than males for both medical and prescription spending. There was substantial growth in total spending over time (58 per cent growth from 2003 to 2008 in total spending for the entire sample; 65 per cent growth for those aged 65 and over). This growth partly reflects the rise in overall health spending experienced in the US over that period, but should be interpreted cautiously as it also reflects the impact of five years of ageing in this continuously-enrolled sample. Spending growth was considerably higher (70 per cent) for medical spending than for prescription drug spending (30 per cent).

IV. The concentration of health spending in the cross-section and over time

1. Cross-sectional distribution

Table 2 shows the cross-sectional distribution of health spending by expenditure type. Individuals in the top 5 per cent of the total expenditure distribution spend $40,755 per year on average, almost eight times the overall average of $5,344, and constitute 45.8 per cent of all spending in our sample. The concentration of medical spending (the top 5 per cent account for 55.0 per cent of spending) is greater than the concentration of prescription drug spending (top 5 per cent account for 39.7 per cent). Spending is substantially less concentrated in the 65-and-over population. The top 5 per cent account for 34.7 per cent of spending, versus 49.3 per cent in the under-65 population. Likewise, the bottom 50 per cent account for nearly three times as large a share of total spending in the 65-and-over population as in the under-65 population. Prescription drug spending is almost as concentrated as medical spending in the under-65 population but is much less concentrated than medical spending in the 65-and-over population.

[Table 2 about here]

2. Correlation over time

To better understand long-term spending patterns, Table 3 shows correlations, in levels, of total spending and of types of spending in year *t* with spending in years *t*+1 to *t*+5. Overall, drug spending correlations are substantially higher than medical spending correlations, and decay more gradually over time (the drug spending correlation at five years is still two-thirds as large as it was at one year, while the medical spending correlation at five years is less than half as large as at one year).

[Table 3 about here]

Since correlations only capture a single linear measure of co-movement, in Table 4 we present transition matrices across spending quintiles, which enable us to observe more general relationships across time and spending categories. The table shows that the correlation of health spending is concentrated in the tails of the spending distribution. Focusing on the diagonal elements, which indicate the probability of remaining in the same spending quintile over time, persistence is consistently highest at the extremes. Nearly 60 per cent of those in the highest and lowest quintiles of 2003 spending remained in the same quintiles in 2004, substantially higher than the percentages remaining in the same quintile among those in the middle quintiles. Even five years out, nearly 50 per cent of those in the highest and lowest quintiles of 2003 spending remained in the same quintiles in 2008. Transitions between the extremes (top to bottom quintile and vice versa) were uncommon as one-year transitions (2.9 and 2.3 per cent, respectively) and remained almost as uncommon even as five-year transitions (5.6 and 4.0 per cent, respectively).

[Table 4 about here]

In Figure 1, we present another measure of concentration in medical spending, by displaying the cumulative distribution function (CDF) for total health spending averaged over one, two and six years. The graphs show that medical spending is highly concentrated even when the data are averaged across six years, which is consistent with health spending being persistent across time. Table 5 displays related measures of the concentration of health spending over different durations: the Gini coefficient and the shares of total spending for the top 1 per cent and top 10 per cent of spenders. Consistent with earlier comparisons of the 65-and-over and under-65 populations, the Gini coefficients show greater concentration in the younger group. In addition, the concentration of spending drops somewhat more rapidly when going from one to six years in the older population (for example, Gini coefficient decreases from 0.57 over one year to 0.46 over six years in the 65-and-over population versus 0.71 to 0.64 in the under-65 population).

[Figure 1 about here]

[Table 5 about here]

3. Average health spending over the life cycle

Figure 2 shows life-cycle profiles of mean total health spending, based on the synthetic cohort implicit in our data. Several notable trends emerge. Except at the extremes of the age distribution, there is a general increase in spending with age, accelerating as the person approaches Medicare eligibility at age 65. The discrete drop at age 65 reflects the transition of insurance coverage from a private, employer-sponsored plan to Medicare. There are at least three reasons such a drop could occur. First, not every person in the under-65 range would qualify for retiree coverage from these employers upon turning 65, so a healthier subset may be ageing into Medicare (because healthier people may be more likely to qualify for employer-related Medicare supplemental coverage because they amassed enough service at the firm to qualify). Second, benefits change at age 65, so the plan design also changes, which can affect covered utilisation. Third, private plans typically pay providers more than Medicare plans. Therefore, this drop may also reflect a change in prices more than any change in utilisation for those who do qualify for coverage (utilisation changes are more likely at age 65 for those transitioning from being uninsured to gaining Medicare coverage, but our sample is continuously insured).

[Figure 2 about here]

V. Modelling health expenditures

In this section, we implement different dynamic econometric models to better understand the factors behind our findings about the concentration of health expenditures. First, we model continuous measures of health expenditures using error components models that decompose the total variability of health expenditures into transitory and permanent components. Then, we apply dynamic models for categorical variables to the probability of being in the top decile of health expenditure. For all of these models, we incorporate in our analysis additional individual information, including trajectories of co-morbidities and trauma episodes.

These methods are based on normal or log-normal distributions. There is a debate in the literature regarding the best way to model the distribution of health costs, especially given some commonly-observed features such as skewness, excess zeros, multimodality and heavy right tails. For example, there could be a bias if the distribution of health expenditures is more skewed than what is assumed by the traditional log-normal assumption. In this case, the distribution may be fitted better by a Pareto or truncated log-normal with an attached Pareto. Feenberg and Skinner (1994) assume that the cross-sectional distribution of health care costs is log-normal and find that log health costs are well represented by an ARMA(1,1) process.[3] Rust and Phelan (1997) argue that the right tail of the health cost distribution is better represented by a Pareto distribution, which has a fatter right tail than a log-normal distribution, even though they do not formally test their Pareto specification against a log-normal alternative, nor do they account for the persistence of health costs. French and Jones (2004) find that the stochastic process for log health costs is well modelled as the sum of a white-noise process and a highly-persistent AR(1) process, with the innovations to this process modelled using a normal distribution adjusted to better capture the risk of extreme health cost shocks. Mihaylova et al. (2011) review several statistical methods for analysing health care resource use and costs and assess their ability to address skewness, excess zeros, multimodality and heavy right tails. They conclude that simple methods are preferred in large samples, where sample means are more likely to be normally distributed. Additionally, in some cases, methods able to deal with more specific data characteristics may be preferable, but checking sensitivity to assumptions is necessary. This would be the case, for example, when the data are skewed and/or heavy tailed, where we can model the costs using alternative distributions instead of normality. The authors recommend the use of inverse gamma and log-normal distributions, but checking the results for robustness to implementation options and outliers in the data.

1. Error components models

The objective here is to study the autocorrelation structure of the health spending process, h_{it}, in order to better understand the intertemporal persistence of health care spending. To do this, we employ a commonly-used error components model:

(1) $$h_{it} = \boldsymbol{\beta}\mathbf{x}_{it} + \varepsilon_{it}$$

where h_{it} is health expenditure for individual *i* at time *t*, x_{it} is a vector of explanatory variables including time-varying characteristics, and ε_{it} is the residual term. We focus on the decomposition of the residual term ε_{it} as the sum of a permanent component, α_i, and a transitory one, v_{it}:

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

where α_i and v_{it} are random variables with mean zero and variances σ_α^2 and σ_{vt}^2 respectively, and p_t and λ_t are factor loadings that allow these variances to change over time in a way that is common across individuals. Our main objective is to identify the separate roles played by the permanent and transitory shocks, and to examine how these roles may have changed over time.

[3]ARMA stands for autoregressive moving-average.

The estimation procedure consists of two stages. In the first stage, we estimate the parameter vector **β** by regressing health costs on demographic and health-related variables that forecast future health costs. In the second stage, we estimate the covariance matrix of the residuals from the first-step regression and fit it to the model using a generalised method of moments (GMM) approach.

We follow previous work in the literature[4] to model persistence in the transitory shocks v_{it}. First, we use an AR(1) process so that

(3) $$v_{it} = \rho v_{i,t-1} + u_{it}$$

where u_{it} is a random variable with variance σ_u^2 and the variance of $v_{i,t=1}$ is given by σ_{v1}^2.

A more elaborate specification that is also widely used in the literature models the transitory shock using an ARMA process with parameter θ. In this case,

(4) $$v_{it} = \rho v_{i,t-1} + \theta u_{i,t-1} + u_{it}.$$

The relatively simple models in (3) and (4) capture important features of expenditure dynamics – namely, time-varying parameters and serial correlation of the transitory shocks.

The model is estimated by GMM using the identity weighting matrix, where sample moments are matched to population moments. Given T periods of data, we have $T(T+1)/2$ moment conditions. The parameter vector to be estimated is given by

(5) $$\boldsymbol{\tau} = (\sigma_\alpha^2, \rho, \sigma_u^2, \sigma_{v_1}^2, p_1, \dots, p_T, \lambda_1, \dots, \lambda_T, \theta).$$

For the implementation of this model, we follow the approach in French and Jones (2004) to estimate the first step using the log of health expenditures as the dependent variable and recoding all health care costs below \$500 (including reports of no expenditures) to \$500.[5] For the second step, we estimate the autocorrelation structure of the health cost process residual using the GMM approach from Doris, O'Neill and Sweetman (2010).

Table 6 presents the empirical covariance matrix used to match the sample moments according to the chosen model. This matrix again reflects the persistence in health expenditures even after five years. Most of the observed drop occurs in the first year, with only gradual declines thereafter. Table 7 presents the parameter estimates obtained using the GMM approach. The two models yield similar conclusions. The estimate for ρ indicates low persistence in the transitory shock, though it is somewhat larger in the oldest age group than in the younger subsamples. The factor loadings λ_2 to

[4] For example, Feenberg and Skinner (1994).

[5] Our results are robust to other bottom-coding decisions (at \$250 and \$750).

λ_6 and p_2 to p_6 indicate relatively constant transitory and permanent variances over time, with somewhat lower values in the oldest group than in the younger subsamples. Still, we conduct a Wald test to evaluate the null hypothesis that the permanent factor loadings p_t are constant over time and we are able to reject this hypothesis at the 1 per cent significance level. Given the similarity of the parameters over time, the rejection likely results primarily from the large sample size.

[Table 6 about here]

[Table 7 about here]

2. Modelling the probability of high health expenditure

In this subsection, we analyse categories of health expenditures, by modelling the probability of being in the highest decile of spending using a dynamic probit model that accounts for individual-specific unobserved heterogeneity. That is,

(6) $$Pr\left(y_{it} = 1 \middle| y_{i,t-1}, \dots, y_{i0}, \mathbf{x}_{it}, \mu_i\right) = \Phi(\rho y_{i,t-1} + \boldsymbol{\gamma}\mathbf{x}_{it} + \mu_i)$$

where y_{it} is an indicator for being in the top 10 per cent of spenders, based on medical and drug payments in each year t. $\mathbf{x}_{it}$ is again a vector of explanatory variables, μ_i represents individual unobserved heterogeneity and $\Phi(\cdot)$ is the standard normal cumulative distribution function. Note that the model assumes that dynamic effects are of the first order, once $\mathbf{x}_{it}$ and μ_i are accounted for (so that only $y_{i,t-1}$ appears on the right-hand side of (6)). Predicted probabilities and other parameters of interest are based on the following conditional expectation function:

(7) $$\Upsilon(\mathbf{x}_{it}, y_{i,t-1}) = E[\Phi(\alpha + \rho y_{i,t-1} + \boldsymbol{\gamma}\mathbf{x}_{it} + \mu_i)]$$

where the expectation is taken with respect to the distribution of μ_i. Wooldridge (2005) uses convenient distributional assumptions (including one for the unobserved individual effects, μ_i) to show that the likelihood function for the outcome of interest has exactly the same structure as in the standard random effects probit model, except that the explanatory variables at time *t* are $\left(1, \mathbf{x}_{it}, y_{i,t-1}, y_{i0}\right)$. Then, we can consistently estimate predicted probabilities with the following estimator:

(8) $$\hat{\Upsilon} = \frac{1}{n}\sum_{i=1}^{n} \Phi\left(\hat{\alpha} + \hat{\beta} y_{i0} + \hat{\boldsymbol{\gamma}}\mathbf{x}_{it} + \hat{\rho} y_{i,t-1}\right)$$

which can be evaluated at different values of $y_{i,t-1}$ and $\boldsymbol{x}_{it}$. We can also compute changes or derivatives of this expression with respect to $\mathbf{x}_{it}$ or $y_{i,t-1}$ to obtain the main effect of interest. Finally, we use the delta method to obtain the standard errors.

Tables 8 and 9 present the results for this model. First, we estimate dynamic probit models as in Wooldridge (2005), including all the explanatory variables listed in the tables. Then, we compute the estimated probability of being in the top 10 per cent of spenders in year *t* conditional on each variable indicated in the rows, and on having been in the top 10 per cent in year *t*–1 (Table 8) or on not having been in the top 10 per cent in year *t*–1 (Table 9). It is important to note that by controlling for the initial condition y_{i0} (the value of the indicator variable in 2003), these results are not directly comparable to the transition matrices reported in Table 4.

[Table 8 about here]

[Table 9 about here]

We find some interesting general patterns. First, persistence by demographic characteristics is generally lower than persistence by co-morbidities. Because co-morbidities are harder to assess, particularly for new enrollees, than demographics, this highlights the need for robust risk prediction models. Second, people with a co-morbid condition relative to those without the condition are considerably more likely to be in the top 10 per cent of spenders in year *t* regardless of whether they were in the top 10 per cent in year *t*–1. However, people with a co-morbid condition are even more likely to be in the top 10 per cent in year *t* if they were also in the top 10 per cent in year *t*–1. For example, when congestive heart failure is present, the probability of being in the top 10 per cent of spenders is 0.27 conditional on having been in the top 10 per cent in the prior year (Table 8) versus 0.17 conditional on not having been in the top 10 per cent in the prior year (Table 9). For most conditions, the differential between those with a given co-morbidity who were and were not in the top 10 per cent in the previous year indicates that being in the top 10 per cent increases the probability of remaining in the top 10 per cent by 5–11 percentage points. These findings imply that knowledge of either co-morbidity status or prior spending will not predict subsequent spending nearly as well as knowledge of both factors. Third, those most likely to be in and remain in the top 10 per cent are those with myocardial infarction, congestive heart failure, peptic ulcer disease and in several psychiatric diagnostic groupings, which indicates that these conditions might be appropriate targets for longer-term disease management programmes. Fourth, although most conditions are less common at younger ages, when they do occur they are more predictive of persistently high spending at younger ages, as almost all conditions have the highest predicted probabilities on being in the top 10 per cent of spenders in the following year when they occur at ages under 25 and the lowest predicted probabilities when they occur in the 65-and-over population. Essentially, the presence of a condition at a younger age more clearly differentiates a person's health care trajectory from that of their peers.

VI. Conclusions

This study adds to our knowledge of spending persistence in the US in several ways. First, it studies a large population of privately-insured individuals under the age of 65. Relatively little recent and broadly representative information is available for this population. Second, we also include an important subgroup of the Medicare-eligible population aged 65 and over – those holding both Medicare and privately-provided supplementary coverage. In both of these age groups, considerable persistence is evident at both ends of the spending distribution using a number of metrics and modelling approaches. The error components models supplement the descriptive data by demonstrating the relative magnitudes of the transitory and permanent components of shocks to spending. The dynamic probit models quantify the independent contributions of co-morbidities and prior history of high spending to the likelihood of future high spending.

Greater knowledge of long-term persistence and its predictors can be useful to policymakers and insurers both for predictive purposes and to help guide the design of interventions. An example of the predictive value of such information is being able to better anticipate the self-selection incentives of potential enrollees. Given the lower-than-anticipated enrolment in health plans offered through the ACA's health insurance exchanges, premiums in some states have risen sharply to reflect the relatively sicker-than-anticipated set of enrollees,[6] and insurers are still trying to improve their predictive capabilities as those markets stabilise over time. In terms of intervention design, the varying magnitudes of persistence seen across conditions can inform the targeting of cost-control strategies. Patients likely to face greater persistence of high spending might be better targeted by chronic disease management programmes or enhanced benefits (for example, Medicare's coverage for self-management training and medical nutrition services for diabetes patients) that focus on controlling progression and complications of underlying conditions. Conversely, those whose high spending is more likely to be transient might be more appropriately targeted by high-cost case management interventions that focus on managing short-term utilisation of expensive interventions and coordination across care settings. Persistence also has important implications for the evaluation of interventions targeting patients with high levels of spending. For example, our results suggest that about half of very high spenders (those in the top quintile) would not be in the top quintile the following year even in the absence of any intervention to reduce spending. This underscores the importance of a control group in any evaluation of programmes targeting high utilisers.

Another key finding is that drug spending is substantially more persistent than medical spending, suggesting that drug coverage might be particularly vulnerable to adverse selection issues in voluntary systems of coverage. In fact, Medicare's Part D drug coverage has several design features (higher premiums for those who delay enrolment, relatively generous coverage for initial expenditures rather than having a large deductible) specifically intended to address this issue by making coverage more attractive to relatively healthy individuals.

[6]Blue Cross Blue Shield, 2016.

Finally, the high degree of spending persistence seen in these data, particularly among patients with certain conditions, suggests that the increasingly-common high-deductible health plans seen in the US may leave many individuals facing persistently high out-of-pocket costs over time. Insurance coverage and benefit design changes (for example, level of deductible and other cost-sharing requirements, use of value-based insurance design (V-BID) features to reduce cost sharing for evidence-based services used by chronically ill people) could provide greater protection from financial risk while limiting adverse incentives to skimp on potentially valuable care.

References

Blue Cross Blue Shield (2016), 'The Health of America Report: newly enrolled members in the individual health insurance market after health care reform – the experience from 2014 and 2015', http://www.bcbs.com/healthofamerica/newly_enrolled_individuals_after_aca.pdf.

Chapman, J. D. (1997), 'Biased enrollment and risk adjustment for health plans', unpublished doctoral dissertation, Harvard University.

Cohen, S. B. and Yu, W. (2012), 'The concentration and persistence in the level of health expenditures over time: estimates for the U.S. population, 2009–2010', Agency for Health Care Research and Quality, Statistical Brief no. 392.

Doris, A., O'Neill, D. and Sweetman, O. (2010), 'Identification of the covariance structure of earnings using the GMM estimator', Institute for the Study of Labor (IZA), Discussion Paper no. 4952.

Eichner, M. J., McClellan, M. B. and Wise, D. A. (1997), 'Health expenditure persistence and the feasibility of medical savings accounts', in J. M. Poterba (ed.), *Tax Policy and the Economy*, vol. 11, pp. 91–128.

Feenberg, D. and Skinner, J. (1994), 'The risk and duration of catastrophic health care expenditures', *Review of Economics and Statistics*, vol. 76, pp. 633–47.

French, E. and Jones, J. (2004), 'On the distribution and dynamics of health care costs', *Journal of Applied Econometrics*, vol. 19, pp. 705–21.

Garber, A. M., MaCurdy, T. E. and McClellan, M. B. (1998), 'Persistence of Medicare expenditures among elderly Medicare beneficiaries', in A. M. Garber (ed.), *Frontiers in Health Policy Research*, Cambridge, MA: MIT Press.

Hirth, R. A., Gibson, T. B., Levy, H. G., Smith, J. A., Calónico, S. and Anup, D. (2015), 'New evidence on the persistence of health spending', *Medical Care Research and Review*, vol. 72, pp. 277–97.

Honoré, B. E. and Kyriazidou, E. (2000), 'Panel data discrete choice models with lagged dependent variables', *Econometrica*, vol. 68, pp. 839–74.

Kohn, J. L. and Liu, J. S. (2013), 'The dynamics of medical care use in the British Household Panel Survey', *Health Economics*, vol. 22, pp. 687–710.

McKellar, M., Landrum, M. B., Gibson, T. B., Landon, B., Naimer, S. and Chernew, M. E. (2012), 'Geographic variation in health care spending, utilization, and quality among the privately insured', Institute of Medicine.

Mihaylova, B., Briggs, A., O'Hagan, A. and Thompson, S. G. (2011), 'Review of statistical methods for analyzing healthcare resources and costs', *Health Economics*, vol. 20, pp. 897–916.

Monheit, A. C. (2003), 'Persistence in health expenditures in the short run: prevalence and consequences', *Medical Care*, vol. 41, pp. III53–64.

Pauly, M. V. and Zeng, Y. (2004), 'Adverse selection and the challenges to stand-alone prescription drug insurance', *Forum for Health Economics and Policy, Volume 7*, Berkeley, CA: Berkeley Electronic Press.

Rettenmaier, A. J. and Wang, Z. (2006), 'Persistence in Medicare reimbursements and personal medical accounts', *Journal of Health Economics*, vol. 25, pp. 39–57.

Riley, G. F. (2007), 'Long-term trends in the concentration of Medicare spending', *Health Affairs*, vol. 26, pp. 808–16.

Rust, J. and Phelan, C. (1997), 'How Social Security and Medicare affect retirement behavior in a world of incomplete markets', *Econometrica*, vol. 65, pp. 781–831.

Thomson Reuters (2007), 'MarketScan Claims Supplemental Slide Library'.

van Vliet, R. C. J. A. (1992), 'Predictability of individual health care expenditures', *Journal of Risk and Insurance*, vol. 59, pp. 443–60.

Wooldridge, J. M. (2005), 'Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity', *Journal of Applied Econometrics*, vol. 20, pp. 39–54.

TABLE 1

Summary statistics

		Whole sample	*Male*	*Female*	*Aged 0–24*	*Aged 25–64*	*Aged 65 and over*
Age (years)		42.59	41.70	43.37	9.52	45.31	73.25
Male		0.47	1.00	0.00	0.51	0.46	0.45
Urban area indicator		0.79	0.79	0.78	0.80	0.78	0.79
Median household income at zip code ($1,000)[a]		48.63	48.99	48.32	50.99	48.63	45.89
Union worker indicator[b]		0.24	0.25	0.24	0.16	0.21	0.45
Region	Northeast	0.10	0.10	0.09	0.09	0.09	0.12
	North Central	0.32	0.32	0.32	0.28	0.29	0.46
	South	0.33	0.32	0.34	0.32	0.35	0.28
	West	0.24	0.24	0.24	0.29	0.26	0.13
Employment status	Active full-time	0.70	0.72	0.69	0.95	0.83	0.01
	Early retiree	0.07	0.06	0.08	0.02	0.11	0.01
	Other	0.23	0.22	0.24	0.04	0.06	0.98
Benefit plan[c]	PPO	0.35	0.35	0.34	0.35	0.38	0.24
	HMO	0.23	0.23	0.23	0.33	0.26	0.01
	POS	0.13	0.12	0.13	0.15	0.16	0.00
	Other	0.30	0.30	0.30	0.17	0.20	0.75
Medical spending ($)	2003	2,966	2,693	3,205	1,191	3,056	4,782
	2004	3,349	3,062	3,600	1,242	3,455	5,508
	2005	3,654	3,373	3,902	1,318	3,753	6,107
	2006	3,848	3,580	4,083	1,406	4,003	6,254
	2007	4,307	4,042	4,539	1,543	4,395	7,295
	2008	5,038	4,813	5,235	1,654	4,981	9,205
Prescription drug spending ($)	2003	1,256	1,138	1,359	298	1,170	2,648
	2004	1,401	1,277	1,510	331	1,302	2,967
	2005	1,471	1,356	1,572	368	1,381	3,048
	2006	1,541	1,438	1,632	398	1,491	3,046
	2007	1,594	1,502	1,676	426	1,565	3,064
	2008	1,639	1,561	1,708	439	1,633	3,074
Total spending ($)	2003	4,222	3,831	4,564	1,488	4,227	7,429
	2004	4,750	4,339	5,110	1,573	4,757	8,475
	2005	5,126	4,729	5,474	1,686	5,135	9,155
	2006	5,390	5,019	5,715	1,803	5,493	9,301
	2007	5,901	5,543	6,215	1,969	5,960	10,359
	2008	6,677	6,374	6,944	2,093	6,614	12,279
Sample size		3,177,267	1,485,211	1,692,056	713,693	1,858,557	605,017

[a]Adjusted to 2014 US dollars.

[b]Proportion that are members of a trade/professional union.

[c]HMO = health maintenance organisation; POS = point of service; PPO = preferred provider organisation.

Note: Spending is adjusted to 2014 US dollars using the personal consumption expenditure price index. The rest of the variables correspond to the year 2003.

TABLE 2

Total health spending percentiles

All

Spending percentile	*Total spending*		*Medical spending*		*Prescription drug spending*	
	Average spending ($)	% of total	Average spending ($)	% of total	Average spending ($)	% of total
All	5,344	100.0	3,860	100.0	1,484	100.0
95–100%	40,755	45.8	35,384	55.0	9,826	39.7
90–95%	14,236	13.3	10,251	13.3	4,620	15.6
70–90%	6,876	25.7	4,151	21.5	2,366	31.9
50–70%	2,747	10.3	1,344	7.0	770	10.4
0–50%	535	4.9	257	3.3	74	2.4

Aged 0–64

Spending percentile	*Total spending*		*Medical spending*		*Prescription drug spending*	
	Average spending ($)	% of total	Average spending ($)	% of total	Average spending ($)	% of total
All	4,322	100.0	3,214	100.0	1,109	100.0
95–100%	35,542	49.3	30,624	57.2	8,650	46.8
90–95%	11,338	13.1	8,277	12.9	3,414	15.4
70–90%	5,211	24.1	3,257	20.3	1,554	28.0
50–70%	1,963	9.1	1,048	6.5	442	8.0
0–50%	384	4.3	207	3.2	40	1.8

Aged 65 and over

Spending percentile	*Total spending*		*Medical spending*		*Prescription drug spending*	
	Average spending ($)	% of total	Average spending ($)	% of total	Average spending ($)	% of total
All	9,551	100.0	6,522	100.0	3,029	100.0
95–100%	55,200	34.7	49,510	45.5	12,031	23.8
90–95%	23,746	12.4	18,456	14.1	7,081	11.7
70–90%	12,798	26.8	8,141	25.0	4,767	31.5
50–70%	6,731	14.1	3,204	9.8	2,861	18.9
0–50%	2,339	12.0	735	5.5	872	14.1

Note: Adjusted to 2014 dollars.

TABLE 3

Correlation of spending (in levels) in year t with spending in years t+1 to t+5

All

	t+1	*t+2*	*t+3*	*t+4*	*t+5*
Total	0.38	0.29	0.25	0.23	0.20
Medical	0.31	0.21	0.17	0.15	0.13
Drugs	0.85	0.76	0.67	0.61	0.57

Aged 0–64

	t+1	*t+2*	*t+3*	*t+4*	*t+5*
Total	0.41	0.30	0.25	0.22	0.19
Medical	0.33	0.23	0.18	0.15	0.13
Drugs	0.85	0.75	0.66	0.59	0.56

Aged 65 and over

	t+1	*t+2*	*t+3*	*t+4*	*t+5*
Total	0.29	0.21	0.19	0.18	0.17
Medical	0.23	0.15	0.14	0.13	0.12
Drugs	0.84	0.72	0.63	0.57	0.51

TABLE 4

Transition matrices for total health expenditure

One-year transitions

Quintile in 2003	All *Quintile next year*					Aged 0–64 *Quintile next year*					Aged 6 *Quinti*	
	Bot-tom	Second	Third	Fourth	Top	Bot-tom	Second	Third	Fourth	Top	Bot-tom	Second
Bottom	59.7	26.0	8.4	3.5	2.3	56.1	26.3	10.2	4.5	2.9	65.6	20.0
Second	24.3	39.9	22.7	8.3	4.9	24.5	37.1	23.4	9.5	5.5	17.4	40.9
Third	9.0	20.6	38.0	22.6	9.8	10.6	21.3	35.0	22.7	10.4	7.2	20.3
Fourth	4.1	8.5	20.6	41.9	25.0	5.3	9.7	20.8	39.9	24.4	4.8	10.6
Top	2.9	5.0	10.4	23.7	58.0	3.6	5.6	10.7	23.4	56.8	5.0	8.3

Two-year transitions

Quintile in 2003	All *Quintile two years later*					Aged 0–64 *Quintile two years later*					Aged 6 *Quintile t*	
	Bot-tom	Second	Third	Fourth	Top	Bot-tom	Second	Third	Fourth	Top	Bot-tom	Second
Bottom	55.9	26.7	10.2	4.4	2.9	52.5	26.5	11.9	5.6	3.6	60.1	21.5
Second	24.4	36.8	23.6	9.6	5.6	24.2	34.5	23.9	11.1	6.3	18.6	36.0
Third	10.6	20.8	34.2	23.5	11.0	12.0	21.3	31.7	23.5	11.6	8.8	20.5
Fourth	5.3	9.6	20.7	38.2	26.2	6.6	10.9	20.8	36.3	25.3	6.1	12.0
Top	3.9	6.1	11.4	24.3	54.3	4.7	6.8	11.7	23.5	53.3	6.5	9.9

Five-year transitions

Quintile in 2003	All *Quintile five years later*					Aged 0–64 *Quintile five years later*					Aged 6 *Quintile f*	
	Bot-tom	Second	Third	Fourth	Top	Bot-tom	Second	Third	Fourth	Top	Bot-tom	Second
Bottom	48.4	28.1	13.5	6.0	4.0	45.4	27.1	15.1	7.6	4.8	50.6	23.4
Second	24.7	32.0	24.3	12.1	7.0	23.8	30.1	24.4	13.8	7.9	20.0	29.7
Third	13.7	20.9	28.4	23.9	13.1	14.8	21.2	26.9	23.6	13.5	11.7	20.2
Fourth	7.7	11.6	20.6	32.7	27.4	9.3	13.2	20.5	31.0	26.1	8.6	14.2
Top	5.6	7.5	13.2	25.3	48.5	6.7	8.4	13.2	23.9	47.7	9.1	12.6

TABLE 5

Measures of the concentration of total health spending over one to six years

	Total health spending averaged over:				
	1 year	*2 years*	*3 years*	*4 years*	*5 years*
All					
Gini coefficient for total spending	0.69	0.66	0.64	0.63	0.62
Percentage spent by top 1% of spenders	27.2%	22.9%	20.9%	19.6%	18.8%
Percentage spent by top 10% of spenders	59.1%	54.2%	51.6%	49.9%	48.7%
Aged 0–64					
Gini coefficient for total spending	0.71	0.69	0.67	0.65	0.64
Percentage spent by top 1% of spenders	30.5%	26.3%	24.1%	22.8%	21.8%
Percentage spent by top 10% of spenders	62.5%	57.6%	55.0%	53.3%	52.2%
Aged 65 and over					
Gini coefficient for total spending	0.57	0.53	0.50	0.48	0.47
Percentage spent by top 1% of spenders	18.7%	15.2%	13.6%	12.7%	12.0%
Percentage spent by top 10% of spenders	47.1%	41.6%	38.8%	37.0%	35.8%

TABLE 6

Empirical covariance matrix

	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

Note: The table shows autocovariances below the diagonal of the matrix, variances along the diagonal and autocorrelations above it.

TABLE 7

Error components model

Parameters	*All*		*Male*		*Female*		*Aged 0–24*		*Aged 25–64*		*Aged 65 and over*	
	AR(1)	ARMA(1,1)	AR(1)	ARMA(1,1)	AR(1)	ARMA(1,1)	AR(1)	ARMA(1,1)	AR(1)	ARMA(1,1)	AR(1)	ARMA(1,1)
ρ	0.164***	0.695***	0.166***	0.715***	0.162***	0.668***	0.166***	0.664***	0.148***	0.648***	0.216***	0.753***
	(0.001)	(0.006)	(0.001)	(0.007)	(0.001)	(0.006)	(0.001)	(0.006)	(0.001)	(0.005)	(0.001)	(0.007)
σ_{α}^{2}	0.321***	0.262***	0.329***	0.262***	0.312***	0.260***	0.149***	0.102***	0.376***	0.330***	0.358***	0.232***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.004)
$\sigma_{v_1}^{2}$	0.644***	0.703***	0.660***	0.727***	0.624***	0.676***	0.453***	0.500***	0.711***	0.757***	0.654***	0.780***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.004)
σ_{u}^{2}	0.318***	0.616***	0.336***	0.639***	0.298***	0.589***	0.197***	0.411***	0.352***	0.681***	0.347***	0.600***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.004)	(0.003)	(0.003)	(0.004)
θ		–0.449***		–0.461***		–0.433***		–0.431***		–0.432***		–0.447***
		(0.004)		(0.004)		(0.004)		(0.004)		(0.003)		(0.003)
λ_2	1.369***	1.018***	1.349***	1.015***	1.394***	1.022***	1.436***	1.021***	1.384***	1.022***	1.286***	1.023***
	(0.007)	(0.002)	(0.007)	(0.002)	(0.007)	(0.002)	(0.010)	(0.002)	(0.007)	(0.001)	(0.005)	(0.002)
λ_3	1.341***	1.030***	1.318***	1.025***	1.370***	1.036***	1.423***	1.046***	1.352***	1.028***	1.262***	1.044***
	(0.007)	(0.002)	(0.006)	(0.002)	(0.007)	(0.002)	(0.010)	(0.002)	(0.007)	(0.001)	(0.005)	(0.002)
λ_4	1.344***	1.041***	1.317***	1.034***	1.378***	1.050***	1.458***	1.081***	1.350***	1.033***	1.261***	1.055***
	(0.007)	(0.001)	(0.006)	(0.001)	(0.007)	(0.002)	(0.010)	(0.002)	(0.007)	(0.001)	(0.005)	(0.002)
λ_5	1.404***	1.049***	1.374***	1.039***	1.441***	1.062***	1.542***	1.102***	1.393***	1.034***	1.355***	1.080***
	(0.007)	(0.001)	(0.007)	(0.001)	(0.007)	(0.002)	(0.010)	(0.002)	(0.007)	(0.001)	(0.006)	(0.002)
λ_6	1.475***	1.075***	1.442***	1.062***	1.516***	1.091***	1.633***	1.143***	1.450***	1.053***	1.457***	1.127***
	(0.008)	(0.002)	(0.007)	(0.002)	(0.008)	(0.002)	(0.011)	(0.003)	(0.007)	(0.002)	(0.006)	(0.002)
p_2	1.077***	1.037***	1.077***	1.030***	1.078***	1.045***	1.093***	1.075***	1.082***	1.054***	1.051***	0.994***
	(0.001)	(0.004)	(0.001)	(0.005)	(0.001)	(0.003)	(0.003)	(0.006)	(0.001)	(0.002)	(0.001)	(0.009)
p_3	1.148***	1.034***	1.144***	1.018***	1.152***	1.054***	1.222***	1.133***	1.151***	1.069***	1.094***	0.931***
	(0.002)	(0.007)	(0.002)	(0.009)	(0.002)	(0.006)	(0.004)	(0.010)	(0.001)	(0.004)	(0.002)	(0.017)
p_4	1.186***	1.067***	1.183***	1.049***	1.190***	1.087***	1.278***	1.194***	1.184***	1.099***	1.136***	0.983***
	(0.002)	(0.009)	(0.002)	(0.011)	(0.002)	(0.007)	(0.004)	(0.013)	(0.002)	(0.005)	(0.002)	(0.022)
p_5	1.140***	1.076***	1.138***	1.062***	1.142***	1.092***	1.216***	1.223***	1.144***	1.105***	1.072***	0.982***
	(0.002)	(0.010)	(0.002)	(0.012)	(0.002)	(0.008)	(0.004)	(0.014)	(0.002)	(0.005)	(0.002)	(0.024)

p_6	1.107***	1.086***	1.108***	1.078***	1.106***	1.095***	1.141***	1.188***	1.120***	1.116***	1.037***	0.989***
	(0.002)	(0.009)	(0.002)	(0.012)	(0.002)	(0.007)	(0.004)	(0.013)	(0.001)	(0.005)	(0.002)	(0.025)
F-test for p	10,886	382.8	10,924	445.2	10,866	378.7	5,702	268.2	15,110	1,045	7,576	602.5
Prob > F	0	0	0	0	0	0	0	0	0	0	0	0

Note: Standard errors are shown in parentheses. *** indicates coefficients that are significant at the 1 per cent level.

TABLE 8

Dynamic random effects probit regressions:
estimated probability of being in the top 10 per cent of spenders in year t conditional on being in the top 10 per cent in the previous year

Dependent variable: indicator for being in the top 10 per cent of total medical spending in year *t*

	All	*Male*	*Female*	*Aged 0–24*	*Aged 25–64*	*Aged 65 and over*
Male	0.16***			0.16***	0.13***	0.15***
	(0.001)			(0.003)	(0.002)	(0.003)
Female	0.13***			0.18***	0.16***	0.16***
	(0.001)			(0.003)	(0.002)	(0.003)
Urban area	0.14***	0.14***	0.15***	0.17***	0.14***	0.15***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
Rural area	0.15***	0.15***	0.16***	0.19***	0.16***	0.16***
	(0.001)	(0.002)	(0.002)	(0.004)	(0.002)	(0.003)
Region						
Northeast	0.14***	0.13***	0.14***	0.17***	0.14***	0.16***
	(0.002)	(0.002)	(0.002)	(0.004)	(0.002)	(0.004)
North Central	0.14***	0.14***	0.15***	0.17***	0.15***	0.14***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
South	0.15***	0.15***	0.15***	0.18***	0.15***	0.16***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
West	0.14***	0.14***	0.15***	0.16***	0.14***	0.17***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	(0.004)
Employee classification						
Union worker indicator	0.14***	0.14***	0.14***	0.17***	0.15***	0.14***
	(0.001)	(0.002)	(0.002)	(0.004)	(0.002)	(0.003)
Non-union worker indicator	0.15***	0.14***	0.15***	0.17***	0.15***	0.16***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
Employment status						
Active full-time	0.17***	0.15***	0.14***	0.17***	0.14***	0.12***
	(0.002)	(0.002)	(0.002)	(0.005)	(0.002)	(0.006)
Early retiree	0.12***	0.15***	0.15***	0.17***	0.15***	0.15***
	(0.001)	(0.002)	(0.002)	(0.005)	(0.002)	(0.003)
Other	0.15***	0.14***	0.15***	0.17***	0.15***	0.20***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	(0.017)
Plan type						
PPO	0.15***	0.14***	0.15***	0.18***	0.15***	0.15***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
HMO	0.15***	0.14***	0.14***	0.17***	0.15***	0.18***
	(0.001)	(0.002)	(0.002)	(0.004)	(0.002)	(0.025)
POS	0.13***	0.13***	0.13***	0.16***	0.13***	0.19***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	(0.009)
Other	0.14***	0.14***	0.15***	0.18***	0.15***	0.15***
	(0.001)	(0.002)	(0.002)	(0.004)	(0.002)	(0.003)
Trauma	0.25***	0.26***	0.25***	0.39***	0.25***	0.26***
	(0.004)	(0.006)	(0.006)	(0.008)	(0.006)	(0.009)

No trauma	0.14***	0.14***	0.15***	0.17***	0.15***	0.15***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
Co-morbidities						
Myocardial infarction	0.40***	0.43***	0.43***	0.65***	0.60***	0.29***
	(0.006)	(0.008)	(0.01)	(0.131)	(0.01)	(0.008)
Congestive heart failure	0.27***	0.28***	0.30***	0.67***	0.37***	0.22***
	(0.003)	(0.005)	(0.005)	(0.065)	(0.008)	(0.005)
Peripheral vascular disease	0.19***	0.19***	0.18***	0.50***	0.27***	0.14***
	(0.003)	(0.004)	(0.004)	(0.069)	(0.007)	(0.004)
Dementia	0.08***	0.07***	0.08***	0.04***	0.04***	0.07***
	(0.004)	(0.005)	(0.005)	(0.010)	(0.005)	(0.005)
Cerebrovascular disease	0.21***	0.22***	0.23***	0.58***	0.30***	0.16***
	(0.003)	(0.004)	(0.004)	(0.031)	(0.005)	(0.004)
Chronic pulmonary disease	0.18***	0.18***	0.19***	0.31***	0.20***	0.14***
	(0.002)	(0.003)	(0.003)	(0.005)	(0.003)	(0.004)
Rheumatologic disease or con-nective tissue	0.25***	0.22***	0.25***	0.51***	0.29***	0.17***
disease	(0.004)	(0.007)	(0.005)	(0.035)	(0.006)	(0.006)
Peptic ulcer disease	0.29***	0.28***	0.29***	0.47***	0.33***	0.16***
	(0.007)	(0.010)	(0.009)	(0.046)	(0.010)	(0.008)
Mild liver disease	0.25***	0.23***	0.25***	0.60***	0.29***	0.15***
	(0.009)	(0.014)	(0.013)	(0.092)	(0.012)	(0.016)
Hemiplegia or paraplegia	0.20***	0.19***	0.21***	0.41***	0.23***	0.13***
	(0.005)	(0.007)	(0.007)	(0.027)	(0.009)	(0.006)
Moderate or severe renal dis-ease	0.16***	0.14***	0.16***	0.27***	0.18***	0.12***
	(0.002)	(0.003)	(0.003)	(0.017)	(0.003)	(0.003)
Diabetes	0.18***	0.16***	0.17***	0.55***	0.19***	0.11***
	(0.002)	(0.003)	(0.003)	(0.015)	(0.003)	(0.003)
Non-metastatic cancer	0.18***	0.16***	0.18***	0.32***	0.20***	0.12***
	(0.002)	(0.002)	(0.002)	(0.010)	(0.002)	(0.003)
Moderate or severe liver disease	0.14***	0.14***	0.14***	0.23***	0.16***	0.09***
	(0.004)	(0.006)	(0.006)	(0.028)	(0.006)	(0.006)
Metastatic solid tumour	0.14***	0.12***	0.14***	0.24***	0.15***	0.10***
	(0.001)	(0.002)	(0.002)	(0.017)	(0.002)	(0.003)
Diabetes + complications	0.13***	0.12***	0.13***	0.25***	0.15***	0.09***
	(0.002)	(0.002)	(0.002)	(0.032)	(0.003)	(0.003)
AIDS	0.15***	0.14***	0.15***	0.19***	0.16***	0.11***
	(0.002)	(0.003)	(0.004)	(0.009)	(0.003)	(0.007)
Psychiatric diagnostic groupings category						
Organic mental disorders	0.26***	0.27***	0.26***	0.45***	0.29***	0.25***
	(0.005)	(0.007)	(0.006)	(0.022)	(0.009)	(0.007)
Alcohol use disorders	0.25***	0.25***	0.22***	0.34***	0.27***	0.24***
	(0.007)	(0.008)	(0.011)	(0.019)	(0.009)	(0.022)
Opioid and other substance use disorders	0.28***	0.27***	0.28***	0.38***	0.29***	0.22***
	(0.004)	(0.006)	(0.006)	(0.014)	(0.005)	(0.013)
Schizophrenia disorders	0.23***	0.22***	0.24***	0.48***	0.25***	0.20***
	(0.013)	(0.019)	(0.020)	(0.065)	(0.018)	(0.030)
Other psychotic disorders (NEC, NOS)	0.26***	0.28***	0.26***	0.39***	0.25***	0.24***

	(0.006)	(0.010)	(0.009)	(0.020)	(0.011)	(0.009)
Bipolar disorders	0.26***	0.26***	0.25***	0.44***	0.25***	0.23***
	(0.006)	(0.009)	(0.007)	(0.017)	(0.007)	(0.016)
Major depressions	0.21***	0.20***	0.22***	0.33***	0.22***	0.22***
	(0.003)	(0.005)	(0.004)	(0.010)	(0.004)	(0.008)
Other specific & atypical affective disorders	0.19***	0.18***	0.18***	0.31***	0.19***	0.18***
	(0.003)	(0.006)	(0.004)	(0.010)	(0.004)	(0.011)
Post traumatic stress disorders	0.18***	0.16***	0.18***	0.34***	0.18***	0.19***
	(0.008)	(0.012)	(0.010)	(0.026)	(0.009)	(0.036)
Anxiety disorders (NOS)	0.18***	0.19***	0.18***	0.27***	0.19***	0.20***
	(0.002)	(0.004)	(0.003)	(0.007)	(0.003)	(0.007)
Personality disorders	0.18***	0.18***	0.18***	0.33***	0.20***	0.22***
	(0.014)	(0.023)	(0.018)	(0.046)	(0.018)	(0.042)
Impulse control, adjustment disorders and	0.19***	0.19***	0.20***	0.29***	0.19***	0.22***
other mental disorders	(0.002)	(0.003)	(0.003)	(0.005)	(0.003)	(0.006)

Note: Predicted probabilities are constructed by setting the corresponding row indicator variable and the lagged indicator for being in the top 10 per cent of spenders equal to 1, then taking the average over all the other variables included in the model. Standard errors are shown in parentheses. *** indicates coefficients that are significant at the 1 per cent level.

TABLE 9

Dynamic random effects probit regressions:
estimated probability of being in the top 10 per cent of spenders in year t conditional on not being in the top 10 per cent in the previous year

Dependent variable: indicator for being in the top 10 per cent of total medical spending in year *t*

	All	*Male*	*Female*	*Aged 0–24*	*Aged 25–64*	*Aged 65 and over*
Male	0.10***			0.08***	0.08***	0.08***
	(0.000)			(0.001)	(0.001)	(0.001)
Female	0.08***			0.09***	0.10***	0.09***
	(0.000)			(0.001)	(0.001)	(0.001)
Urban area	0.09***	0.09***	0.09***	0.08***	0.09***	0.09***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)
Rural area	0.10***	0.09***	0.09***	0.09***	0.09***	0.09***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Region						
Northeast	0.09***	0.08***	0.08***	0.08***	0.08***	0.09***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
North Central	0.09***	0.09***	0.09***	0.09***	0.09***	0.08***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
South	0.09***	0.09***	0.09***	0.09***	0.09***	0.09***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
West	0.09***	0.09***	0.09***	0.08***	0.08***	0.10***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Employee classification						
Union worker indicator	0.08***	0.09***	0.08***	0.08***	0.09***	0.08***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Non-union worker indicator	0.09***	0.09***	0.09***	0.08***	0.09***	0.10***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)
Employment status						
Active full-time	0.10***	0.09***	0.08***	0.08***	0.09***	0.06***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.003)
Early retiree	0.08***	0.09***	0.09***	0.08***	0.09***	0.09***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Other	0.09***	0.08***	0.09***	0.08***	0.09***	0.12***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.012)
Plan type						
PPO	0.09***	0.09***	0.09***	0.09***	0.09***	0.09***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
HMO	0.09***	0.09***	0.08***	0.08***	0.09***	0.11***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.017)
POS	0.08***	0.08***	0.08***	0.08***	0.08***	0.12***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.006)
Other	0.09***	0.09***	0.09***	0.09***	0.09***	0.09***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Trauma	0.17***	0.17***	0.16***	0.23***	0.16***	0.16***
	(0.003)	(0.004)	(0.004)	(0.005)	(0.004)	(0.006)
No trauma	0.09***	0.09***	0.09***	0.08***	0.09***	0.09***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)
Co-morbidities						
Myocardial infarction	0.29***	0.31***	0.29***	0.46***	0.46***	0.17***

	(0.005)	(0.006)	(0.008)	(0.138)	(0.009)	(0.005)
Congestive heart failure	0.17***	0.18***	0.19***	0.49***	0.25***	0.12***
	(0.002)	(0.003)	(0.004)	(0.071)	(0.006)	(0.002)
Peripheral vascular disease	0.12***	0.11***	0.10***	0.32***	0.17***	0.07***
	(0.002)	(0.002)	(0.003)	(0.060)	(0.005)	(0.002)
Dementia	0.04***	0.03***	0.04***	0.02***	0.02***	0.03***
	(0.002)	(0.003)	(0.003)	(0.004)	(0.003)	(0.002)
Cerebrovascular disease	0.14***	0.13***	0.14***	0.39***	0.19***	0.08***
	(0.002)	(0.002)	(0.002)	(0.029)	(0.003)	(0.002)
Chronic pulmonary disease	0.11***	0.10***	0.11***	0.17***	0.12***	0.07***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)
Rheumatologic disease or con-nective tissue disease	0.16***	0.14***	0.15***	0.33***	0.18***	0.09***
	(0.003)	(0.005)	(0.003)	(0.031)	(0.004)	(0.003)
Peptic ulcer disease	0.19***	0.18***	0.18***	0.30***	0.21***	0.09***
	(0.005)	(0.008)	(0.007)	(0.039)	(0.008)	(0.005)
Mild liver disease	0.16***	0.14***	0.15***	0.41***	0.18***	0.07***
	(0.007)	(0.010)	(0.010)	(0.092)	(0.009)	(0.009)
Hemiplegia or paraplegia	0.12***	0.11***	0.12***	0.25***	0.14***	0.07***
	(0.003)	(0.005)	(0.005)	(0.022)	(0.006)	(0.003)
Moderate or severe renal dis-ease	0.09***	0.08***	0.09***	0.14***	0.11***	0.06***
	(0.001)	(0.001)	(0.002)	(0.011)	(0.002)	(0.001)
Diabetes	0.11***	0.09***	0.10***	0.37***	0.11***	0.05***
	(0.001)	(0.001)	(0.001)	(0.013)	(0.001)	(0.001)
Non-metastatic cancer	0.11***	0.09***	0.10***	0.17***	0.12***	0.06***
	(0.001)	(0.001)	(0.001)	(0.007)	(0.001)	(0.001)
Moderate or severe liver disease	0.08***	0.08***	0.08***	0.11***	0.09***	0.04***
	(0.003)	(0.004)	(0.004)	(0.017)	(0.004)	(0.003)
Metastatic solid tumour	0.08***	0.07***	0.08***	0.13***	0.08***	0.05***
	(0.001)	(0.001)	(0.001)	(0.010)	(0.001)	(0.001)
Diabetes + complications	0.07***	0.06***	0.07***	0.13***	0.08***	0.04***
	(0.001)	(0.001)	(0.001)	(0.020)	(0.001)	(0.001)
AIDS	0.09***	0.08***	0.08***	0.09***	0.09***	0.05***
	(0.001)	(0.001)	(0.002)	(0.005)	(0.001)	(0.004)
Psychiatric diagnostic groupings category						
Organic mental disorders	0.17***	0.18***	0.16***	0.28***	0.19***	0.15***
	(0.004)	(0.005)	(0.005)	(0.018)	(0.007)	(0.004)
Alcohol use disorders	0.16***	0.16***	0.14***	0.19***	0.17***	0.15***
	(0.005)	(0.006)	(0.008)	(0.013)	(0.006)	(0.016)
Opioid and other substance use disorders	0.19***	0.18***	0.18***	0.22***	0.19***	0.13***
	(0.003)	(0.004)	(0.005)	(0.011)	(0.004)	(0.009)
Schizophrenia disorders	0.15***	0.14***	0.15***	0.30***	0.16***	0.12***
	(0.010)	(0.014)	(0.014)	(0.056)	(0.014)	(0.020)
Other psychotic disorders (NEC, NOS)	0.17***	0.19***	0.16***	0.23***	0.16***	0.15***
	(0.005)	(0.007)	(0.006)	(0.015)	(0.008)	(0.006)

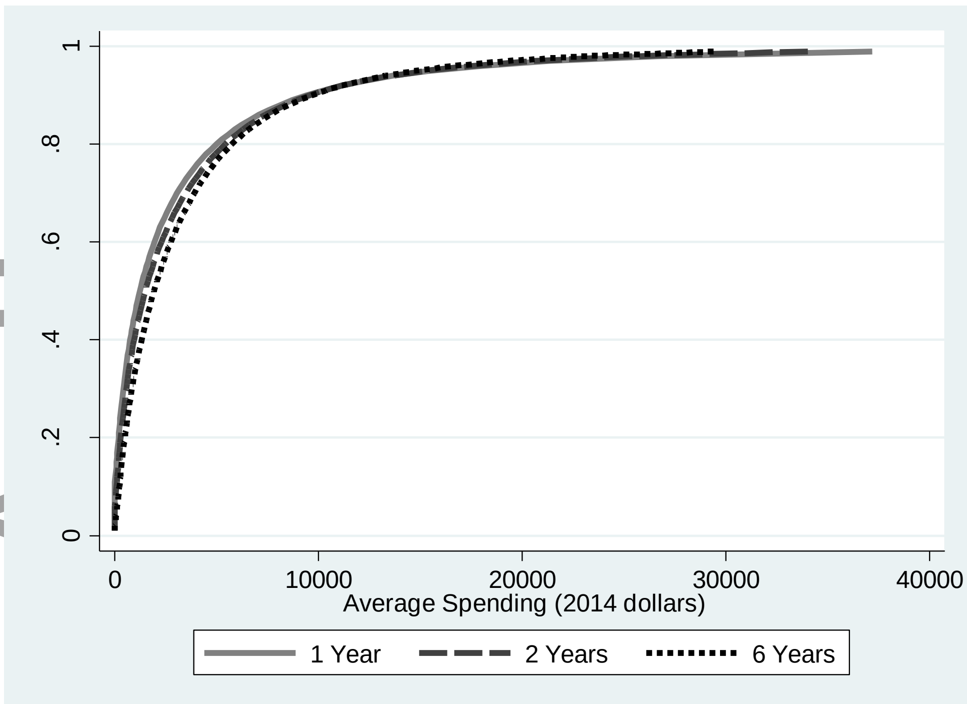
Bipolar disorders	0.17***	0.17***	0.16***	0.27***	0.16***	0.14***
	(0.004)	(0.007)	(0.005)	(0.014)	(0.005)	(0.011)
Major depressions	0.14***	0.13***	0.13***	0.18***	0.14***	0.13***
	(0.002)	(0.003)	(0.002)	(0.006)	(0.002)	(0.005)
Other specific & atypical affective disorders	0.12***	0.12***	0.11***	0.17***	0.11***	0.11***
	(0.002)	(0.004)	(0.003)	(0.007)	(0.003)	(0.007)
Post traumatic stress disorders	0.12***	0.10***	0.11***	0.19***	0.11***	0.11***
	(0.006)	(0.008)	(0.007)	(0.018)	(0.006)	(0.024)
Anxiety disorders (NOS)	0.12***	0.12***	0.11***	0.14***	0.11***	0.12***
	(0.002)	(0.003)	(0.002)	(0.004)	(0.002)	(0.004)
Personality disorders	0.12***	0.12***	0.11***	0.19***	0.12***	0.13***
	(0.010)	(0.016)	(0.012)	(0.033)	(0.012)	(0.030)
Impulse control, adjustment disorders and	0.12***	0.12***	0.12***	0.16***	0.12***	0.13***
other mental disorders	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	(0.004)

Note: Predicted probabilities are constructed by setting the corresponding row indicator variable equal to 1 and the lagged indicator for being in the top 10 per cent of spenders equal to 0, then taking the average over all the other variables included in the model. Standard errors are shown in parentheses. *** indicates coefficients that are significant at the 1 per cent level.

FIGURE 1

Empirical CDF of health expenditures, averaged over one, two and six years

Aged 0–64



Aged 65 and over

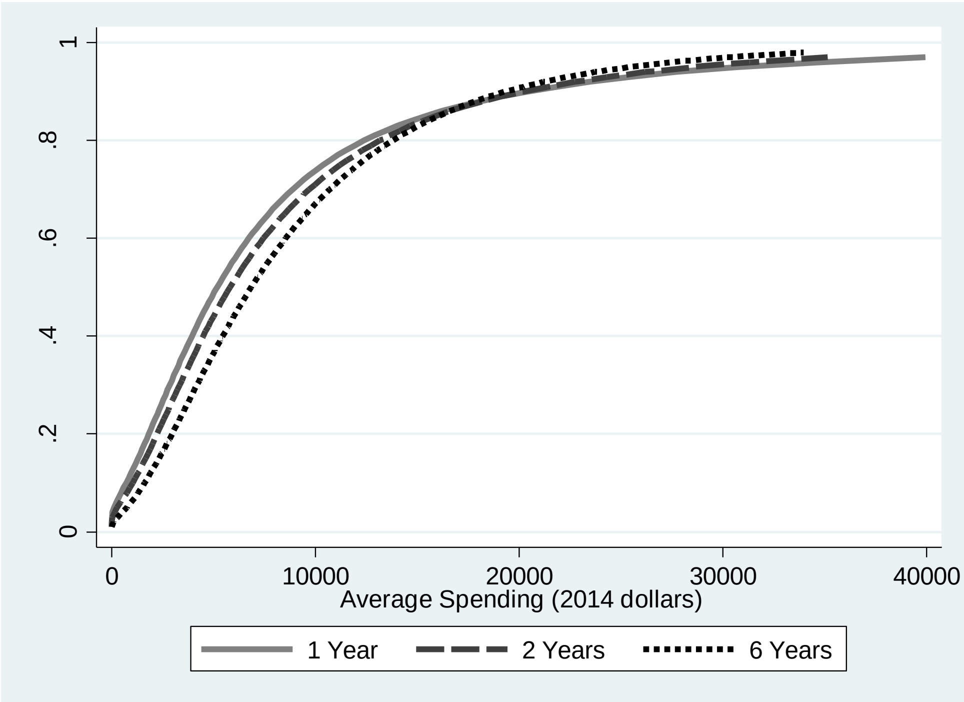


FIGURE 2

Average total health expenditures, by age