

2012

Formal and Informal Care: An Empirical Bayesian Analysis Using the Two-Part Model

Juan Du
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/economics_facpubs



Part of the [Health Economics Commons](#)

Repository Citation

Du, Juan, "Formal and Informal Care: An Empirical Bayesian Analysis Using the Two-Part Model" (2012). *Economics Faculty Publications*. 34.
https://digitalcommons.odu.edu/economics_facpubs/34

Original Publication Citation

Du, J. (2012). Formal and informal care: An empirical Bayesian analysis using the two-part model. *Forum for Health Economics & Policy*, 15(2), Article 8. doi:10.1515/1558-9544.1253

Forum for Health Economics & Policy

Volume 15, Issue 2

2012

Article 8

(AGING AND MEDICAL CARE COSTS)

Formal and Informal Care: An Empirical Bayesian Analysis Using the Two-part Model

Juan Du*

*Old Dominion University, jdu@odu.edu

DOI: 10.1515/1558-9544.1253

Copyright ©2012 De Gruyter. All rights reserved.

Formal and Informal Care: An Empirical Bayesian Analysis Using the Two-part Model*

Juan Du

Abstract

Informal care provided to the elderly by their children is proposed as a less expensive alternative to institutional long-term care. This paper explores how the elderly's consumption of medical care changes in response to changes in the informal care they receive from their children. Many earlier studies have ignored both the endogeneity of informal care and the complicated nature of health care utilization data. This paper develops a two-part model with informal care treated as an endogenous regressor and imposes exclusion restrictions on the selection process. The model is fitted using the Bayesian Markov Chain Monte Carlo (MCMC) methods, in particular the Gibbs sampler and the Metropolis-Hasting algorithm. The average treatment effects and the distributions of the treatment effects are obtained via posterior simulation. The results indicate that informal care provides a substitute for nursing home care and hospital inpatient care, but it does not affect paid home health care on average. The treatment effects are heterogeneous. The largest substitution effects occur for nursing home and hospital inpatient care at the intensive margin. The policy analysis suggests that informal care policies targeting the group that incurs the largest substitution effect may help to reduce government spending on Medicaid and Medicare.

KEYWORDS: informal care, nursing home, hospital inpatient care, Bayesian, data augmentation, treatment effect

*Direct correspondence to Juan Du at jdu@odu.edu; Address: Department of Economics, College of Business and Public Administration, Old Dominion University. Tel: 757-683-3543. I wish to thank Colin Cameron, Hilary Hoynes, Douglas Miller, and Paul Leigh for their great help with this paper. Special thanks to Oscar Jorda, Guido Kuersteiner, Mitali Das, Ann Stevens for their valuable comments and discussions. I am also grateful to all the participants at the UC Davis Economics Brownbag series. All errors are my own.

I Introduction

Meeting the demand for long-term care has become a crucial issue in the United States as the baby-boomer generation grows older and the share of elderly aged over 80 in the population is projected to double between 2010 and 2050, reaching 7.4% in 2050 (OECD, 2011). Kemper et al. (2006) project that 69% of all people turning 65 in the United States will need some form of long-term care services while over 3 out of 10 will need family care for more than 2 years and 1 out of 20 will spend more than \$100,000 on long-term care. This suggests a sizable right tail in the distribution of long-term care expenditure.

In 2010, nursing home care (6%) and paid home health care (3%) together accounted for 9% of total health care expenditure,¹ approximately 1.4% of the US GDP, or \$203 billion. Long-term care spending grew faster than GDP growth, and it is projected to grow at an average of 6.7% from 2012 - 2018,² outpacing overall health care spending in the next 40 years (OECD, 2011). Increased longevity has imposed increasing pressure on family members, traditionally the main source of long-term care. In addition, decreased fertility and changes in the family structure, such as family size becoming smaller, more daughters joining the labor force and delaying child-bearing, have reduced the availability of family caregivers and put additional strain on long-term care resources.³

The burgeoning demand for long-term care of the elderly has also introduced tremendous pressure on public programs, which cover approximately 60% of long-term care expenditures (Congressional Budget Office, 2004). Medicaid, a means-tested program, is the dominant payer for nursing home care in the United States, but only those who demonstrate a low enough income are eligible. Medicare is designed mainly to cover costs associated with short-term rehabilitation care due to acute illness rather than long-term care. When broken down by payer type, in 2008, 41% of nursing home expenditures were financed by Medicaid, 27% out-of-pocket, 19% by Medicare (19%), and the remaining 7% by private long-term care insurance. For paid home health care expenditure, Medicare paid 41%, and the rest was paid by Medicaid (35%), out-of-pocket (10%), and private insurance (9%).⁴ As more and more baby boomers become eligible for Medicare, it is projected that the growth of the

¹Center for Medicare and Medicaid Services, 2010.

²Center for Medicare and Medicaid Services, National Health Expenditure Projections 2008-2018.

³There may be a decreasing trend in the number of disabilities and other limitations of the elderly, which could offset the potential gap between the supply and demand of long-term care, but this trend has not been shown to be persistent (Pezzini and Spillman, 2000).

⁴Center for Medicare and Medicaid, National Health Expenditure Data, Table 2, 2008.

public share will outpace that of the private share.⁵

However, both Medicaid and Medicare have experienced rapid cost growth in the recent decade, and concerns about the sustainability of Medicare and Medicaid have been growing. Many states struggle to contain Medicaid costs when expanding it to institutional, home and community levels. Subsidies for purchasing private long-term care insurance have been proposed and occasionally implemented over the last 15 years; however, perceptions of future uncertainty and affordability, as well as the availability of public options through Medicaid, have limited the growth of the private insurance market. Feder et al. (2000) concluded that ‘A complex system of public and private funding often leaves elderly persons at risk of financial catastrophe and inadequate care’.

This paper examines a less costly alternative to elderly’s long-term care: informal care provided by family members. The percentage of elderly aged 70 and older receiving informal care increased from 8.1% in 1998 to 9.9% in 2010, and the market value of informal care makes up approximately 60% of market spending (Brown and Finkelstein, 2011). First, it is crucial to identify the relationship between formal and informal care. This relationship has been examined in several studies that use data from various countries in different time periods with different models, but no definite conclusions have been reached. Second, it is important to understand the policy implications of informal care and its effects on Medicaid and Medicare spending, but almost no prior studies have conducted such a policy analysis.

This paper uses the Health and Retirement Study (HRS) to study the relationship between informal care and formal medical care utilization. The goal of this paper is to understand how medical care consumption, especially long-term care utilization by the elderly, changes with the intensity of the informal care they receive from their children. Whether informal care and formal medical care are substitutes or complements may depend on the type of care. Four major types of formal medical care are examined in this paper: nursing home care, paid home health care, physician visits, and hospital care.

There are several challenges in estimating health care outcomes. A large fraction of the elderly population does not use medical care during the period of observation, leading to a mass of zeroes (the extensive margin), which must be modeled separately from positive outcomes (the intensive margin). In addition, utilization conditional on positive outcomes is highly skewed. For example, 86% of the sample did not stay in nursing homes in 2004, and the distribution of utilization has a sizable right tail (see Figure 1). There are also significant challenges in identification because formal and informal care decisions are often made jointly by parents and children, and unobservables at the family and

⁵Center for Medicare and Medicaid Services, National Health Expenditure Projections 2008-2018.

individual level could add to the complicated nature of this relationship.

This paper contributes to the body of research on informal and formal care in several ways. First, informal care is treated as an endogenous variable. Some earlier studies (Hanley, 1990; Jette et al., 1995; Freedman, 1996) treated informal care as exogenous. However, endogeneity may arise from unobservables, and treating informal care as exogenous may lead to misleading inferences. This paper uses children's characteristics as exclusion restrictions and allows formal and informal care to be correlated, thus reducing bias from unobservables. Second, this paper develops an endogenous two-part model for medical care utilization. This model has the advantage of modeling zero and positive outcomes separately while controlling for endogeneity. The proposed model is subject to Bayesian analysis based on the Markov Chain Monte Carlo (MCMC) methods, building on earlier works by Albert and Chib (1993), Koop and Poirier (1997), Chib and Hamilton (2000), Geweke et al. (2003), and Deb et al. (2006a, 2006b). In particular, latent variables generated through the data augmentation method along with the Gibbs sampler and the Metropolis-Hastings algorithms (Albert and Chib, 1993) are implemented. This paper shows the advantages of Bayesian methods, including computational convenience and the ability to determine distributions of the treatment effects. Bayesian methods also offer a unique way of testing endogeneity. Third, our study relies on data that is more recent than the data used in earlier studies.⁶ This is important as several factors related to informal care may change over-time, such as the gender of caregivers and the technology used in caregiving. Fourth, this paper contributes new insights to the literature by examining the relationship by length of stay. In particular, long-term and short-term nursing home care as well as care that follows hospital inpatient care are distinguished. Lastly, a policy analysis by payer type is performed based on simulations for two potential policies.

Evidence on whether informal care can reduce formal care utilization has important policy implications. Many policies have been designed to give incentives for family members to provide informal care, for example, proposals that provide tax credit (or exemption) for people who care for impaired relatives living in their homes, proposals on offering caregiver training, education, direct financial assistance, or funding for employment breaks, etc. There are also policy interests in increasing the effectiveness of informal care and in integrating formal and informal care in a system that is more efficient for allocating resources while reducing real care needs (Kodner and Kyriacou, 2000). To evaluate these policy proposals, the relationship between formal and informal

⁶For example, Van Houtven and Norton (2004, 2008) and Lo Sasso and Johnson (2002) used data from 1993 and 1998. Hanley (1990), Newman and Struyk (1990) and Liu et al. (2000) used data from 1982 and 1984.

care needs to be carefully examined.

This paper finds that the relationship between formal and informal care differs considerably by the type of care as well as by length of stay. Informal care is a substitute for nursing home care and hospital inpatient care. It does not affect the probability of physician visits, but it does reduce the number of visits. Departing somewhat from the literature, no statistically significant relationship is found between informal care and paid home care. With respect to length of stay, I find that informal care has a larger effect for the elderly who stayed in nursing homes for more than a hundred days than those who did not. A hundred days is the maximum length that Medicare pays for skilled nursing facilities. In addition, the substitution effect is very large for those who had on average less than 3 days of hospital inpatient care, which does not qualify for Medicare reimbursement. The treatment effects differ across different types of care at both the intensive and extensive margins. The largest substitution effects occur for nursing home and hospital care at the intensive margin. Finally, the policy analysis shows that informal care policies that target the group that incurs the largest substitution effect may help reduce government spending on Medicaid and Medicare.

This paper is organized as follows. Section II provides a review of the literature, and Section III introduces the conceptual and modeling framework. Section IV describes the main variables and the exclusion restrictions used in the estimation. Section V reports the main results as well as results by the type of care and length of stay. Section VI illustrates treatment effects at the intensive and extensive margins. Section VII provides a policy analysis focusing on potential cost reductions for Medicare and Medicaid. Section VIII tests the endogeneity of informal care. The final section concludes. Implementation of the Bayesian algorithms, computation of treatment effects, and steps for the MCMC procedure can be found in the appendices.

II Literature

Several medical studies (Jette, et al., 1995; Freedman, 1996) find that living with one's spouse and children significantly reduces the probability of nursing home stays whereas other medical studies (Hanley, 1990; Newman and Struyk, 1990) do not find statistically significant effects. More recent work in economics (Lo sasso and Johnson, 2002; Charles and Sevak, 2005) use the Asset and Health Dynamics of the Oldest Old (AHEAD 1993, 1995) and find that regular help from children reduces the likelihood of nursing home entry for disabled Americans over 70, but Lo sasso and Johnson (2002) find no significant effect when informal care is measured more broadly. The zero mass and skewed distributions of positive outcomes are largely ignored in these studies,

and the data they use oversampled people with disability.

Empirical work on the causal effects governing the relationship between paid home care and informal care is very limited. Two studies (Langa et al., 2001; Liu et al., 2000) published in medical journals find that the elderly who live with children received more paid home health care than did the elderly who live alone. This suggests a somewhat complementary relationship between paid home care and informal care. By contrast, Bosang (2009), using European data and the instrumental variables approach, shows that informal care replaced paid domestic help and that informal care is endogenous to the decision of getting paid care.

Research studies on other types of health care utilization are few. Notable exceptions are the study by Van Houtven and Norton (2004) and a recent European study (Bolin et al., 2007). Van Houtven and Norton (2004) use data from the 1995 AHEAD and the 1998 HRS and find that informal care is a substitute for all types of formal care. However, Bolin et al. (2007) and Bonsang (2009) apply similar models with European data and reach somewhat different conclusions. For example, Bolin et al. (2007) find that informal care and paid home care are substitutes, but informal care and hospitals inpatient care and physician visits are complements.

The broader literature (Christianson, 1988; Weissert, 1988; Kemper, 1992) on long-term care for the elderly includes various studies on the relationships between community care, nursing homes, and informal caregiving based on the national channeling experiment - a U.S. national intervention in the 1980s that allocated resources toward community care as a substitute for nursing home care as well as informal care. The general conclusion is that community-based health care was not a cost-effective substitute for nursing home care, and the substitution effects between several forms of long-term care were very small. These results could be due to the fact that people at higher risk of institutionalization were not well targeted in the experiment (Weissert, 1998).

As there is still no clear consensus in the literature, it is crucial to re-examine this topic at a time when aging is affecting our lives more than ever.

III Conceptual and modeling framework

The conceptual framework of this paper is based on two decision-making processes. First, children maximize utility by deciding whether to provide care and how much care to provide. This can be thought as a selection process in which children select themselves to become caregivers. Secondly, elderly parents maximize utility by choosing whether to use formal care and how much care to use, which also depends on the informal care decisions that their children make. In reality the two decisions are not made independently and are

often determined by characteristics of parents and children, and could be affected by factors that are not observed by researchers. Therefore, it is important to treat informal care as endogenous when estimating the relationship between formal and informal care.

According to Andersen (1968), health care utilization can be viewed as a function of predisposing factors, enabling factors, and need factors. Predisposing factors, which predicts the likelihood of using care, include demographic variables (such as age and gender) and social structure variables (such as race and education). Wealth, income, health insurance, and long-term care insurance are regarded as enabling factors that must be present for care to take place. Need factors are the primary determinants of utilization, such as self-perceived health or actual illnesses (ADL, IADL, and chronic conditions). Although the observed factors may play different roles depending on the type of care,⁷ they are all included as control variables in the equations that estimate formal and informal care.

Under this framework, there is a potential for endogeneity stemming from unobserved factors at both the family and individual level that affect formal and informal care decisions. These factors may include values and the knowledge people have about their health, their risk preferences, social networks and interactions, the availability of medical facilities, knowledge about health providers, varying levels of insurance coverage, unrevealed or undiagnosed health information, children and parent relationships, etc. These unobserved factors are relevant in determining the relationship between formal and informal care. For example, if an elderly person values her health greatly and is very risk averse to health fluctuations, she may desire more formal and informal care. On the other hand, when the elderly person is socially active and maintain a certain social status, she may perceive entering nursing homes as a last resort, thus requiring child's care as a substitute for institutional care. This implies that the provision of informal care is likely to be endogenous because unobservables determining informal care are likely to be correlated with unobservables determining formal care. This paper reduces the potential omitted variable bias by using the instrumental variables approach that assumes that certain children's characteristics will only affect informal care decisions, and formal care decisions are not affected directly. The residuals from the selection equation are included in the formal care equation to control for potential omitted variables.

Formal care utilization is modeled using the two-part model, following Duan et al. (1982, 1983). The advantage of the two-part model as opposed to the Tobit model is in its ability to avoid having a single distribution for both zero

⁷For example, hospital inpatient care that often occurs in response to more serious illnesses, may be better explained by need factors.

and positive outcomes. If the underlying distributions are different for zeros and positive outcomes, the Tobit model would generate biased estimates. In addition, hospital care and physician visits are measured as counts. This requires estimation using the Poisson distribution (Cameron and Trivedi, 1998). For nursing homes, log-normal distribution is used instead.

There are several hypotheses being tested in this paper. First, following previous studies in the literature, I test whether informal care is a substitute for or a complement to formal care. Second, I hypothesize that the effect of informal care on formal care differs by length of stay and by Medicare eligibility, in particular, for nursing home and hospital inpatient care. The reason is that a shorter length of stay is typically covered by Medicare, and informal care may not result in large substitution effects. Lastly, I hypothesize that informal care is endogenous; that is, unobservables at the family or individual levels should be accounted for.

Informal care decision

I begin the modeling framework by describing children's decision-making process when deciding to become caregivers. There are several factors that affect both the supply and demand of informal care. First, on the supply side, with more and more daughters joining the labor force and contributing to their family income, the opportunity cost of caregiving is higher than before. Based on Bureau of Labor Statistics, labor force participation rates for women aged 55 to 64 increased from 44% in 1988 to 59% in 2008. For women aged 45 to 54, the labor force participation rate increased from 69% in 1988 to 76% in 2008.⁸ Carmichael, Charles, and Hulme (2010) show that children with high earnings and full-employment face a higher opportunity cost than those who are not in the paid employment, and the former are thus less likely to become caregivers. This means that daughters are less available to provide the type of care that involves substantial hours, including co-residential caring, and the burden of informal care could fall on sons as primary or secondary caregivers. It is therefore important to include the employment status of both daughters and sons when estimating the intensity of informal care.

On the demand side, parents' socioeconomic status and their need for care play important roles. Parents with severe illnesses or multiple chronic conditions have higher needs for informal care. Availability of long-term care insurance is expected to make long-term care more affordable and to make parents less dependent on children. Parents' wealth may become a factor as an elderly person's bequest motives may induce children to provide more care.

⁸Labor force participation rate for men aged 55 to 64 increased from 67% in 1998 to 70% in 2008 whereas men aged 45 to 54 experienced a slight decrease from 90% to 88%.

Let z^* denote the latent utility associated with the desired hours of informal care.

$$z^* = w\alpha + u \quad (1)$$

where $u \sim N(0, \sigma_u^2)$. w includes parents' demographic and socioeconomic characteristics, health status, wealth, and insurance status. w also includes children's employment status as well as factors that affect the availability of children as caregivers. As these variables are used as exclusion restrictions, they do not appear in the formal care equation. Let z be the hours of care actually received by the elderly. $z > 0$ if the utility from desired hours of care is greater than zero, and the elderly person receives informal care from children, and $z = 0$ if the utility is less than or equal to zero, and the elderly person does not receive informal care. In estimation, the logarithm of z is used.

Formal care decision

The focus of this paper is the second decision process – how do parents' formal care utilization changes in response to changes in the informal care they receive from children.

Traditionally we distinguish between two types of elderly care: health-related care and long-term social care. Health-related care addresses mental health and physical health conditions, such as diabetes and heart diseases. Long-term social care can be thought of as services that use non-medical inputs to support the daily functioning and broader well-being of the individual. Informal care is more likely to serve as a substitute for long-term social care, which requires less medical knowledge. However, informal care may have an effect on health-related care as well. For example, following a surgery or an acute illness, the availability of informal caregivers could shorten hospital inpatient stays or skilled nursing facilities utilization. On the one hand, lack of informal care can lead to avoidable admissions into and unnecessarily long stays in hospitals. On the other hand, the availability of children can make health-related services easier to access and potentially enhance their effectiveness. For example, children can help their elderly parents better understand their treatment options, modify behaviors, and follow treatment schedules more carefully, which may increase or decrease subsequent care. Children can also visit their parents in nursing homes and improve their quality of life on weekends, and they may facilitate transportation to hospitals should such a need arise. Johri et al. (2003) find benefits in an integrated care model that combines health-related services and social care for the elderly, resulting in reduced use of acute hospitalization and institutionalization as well as community-based services. It is possible that formal and informal care are substitutes in some cases but

complements in others, with a considerable amount of heterogeneity across individuals.

The standard two-part model consists of two parts. Part 1, the hurdle part, is a binary probit model that specifies $(\Pr(Y > 0 | x) = \Phi(x'\beta_1))$ for the full sample and part 2 is a log-linear model $(E[\ln(Y) | Y > 0, x] = x'\beta_2)$ conditional on positive outcomes. The endogenous two-part model differs from the standard two-part model in two respects: 1) endogeneity is controlled for by allowing correlation between the error terms; 2) a count model is applied in the second part for physician visits and hospital care whereas in the standard two-part model, the second part is linear. This adds significant nonlinearity in the model, thus complicating computation. The details of the model are described below.

Part 1: predicting the probability of positive formal care utilization. Define latent variable Y_1^* as the difference in utility between formal care utilization and no utilization, and let Y_1 be the binary variable taking values of 1 and 0, indicating users and non-users. An individual becomes an user of formal care if $Y_1^* > 0$. Y_1^* is determined by predisposing factors, enabling factors, and need factors of the elderly (denoted by x) as well as the intensity of the informal care they receive from children (denoted by z). Any unobserved factors are included in the error term ϵ_1 . Mathematically, Y_1 and Y_1^* can be written as,

$$Y_1^* = x\beta_1 + z\tau_1 + \epsilon_1 \quad (2)$$

$$Y_1 = \begin{cases} 1 & \text{if } Y_1^* > 0 \\ 0 & \text{if } Y_1^* \leq 0 \end{cases} \quad (3)$$

Part 2: predicting the level of formal care utilization conditional on users. There are several formal care variables in this paper, thus requires different modeling assumptions. For physician visits and hospital inpatient care, I adopt the count data model by assuming potential utilization (Y_2^*) follows the Poisson distribution with mean $exp(\mu)$, defined as

$$\mu = x\beta_2 + z\tau_2 + \epsilon_2. \quad (4)$$

Utilization is determined by the same set of control variables as in part 1. The actual utilization (Y_2) is observed only if potential utilization is greater than zero. The two-part modeling structure is particularly attractive if there are many zeroes in the outcome.

For nursing homes, I use the log-normal distribution because nursing home utilization has a much longer right tail (see Figure 1) than hospital utilization and physician visits. Log-normal distribution is known to fit the right skewed distribution relatively better than other distributions. Potential nursing home utilization is defined as Y_{nh}^* , and can be written as

$$Y_{nh}^* = x\beta_2 + z\tau_2 + \epsilon_2 \quad (5)$$

and $Y_{nh} = \exp(Y_{nh}^*)$ if $Y_1 = 1$; $Y_{nh} = 0$ if $Y_1 = 0$. Again positive outcomes are observed for users.

A restrictive assumption of the standard two-part model is that the errors (ϵ_1, ϵ_2) are independent. Following Deb et al. (2006a), correlation of the error terms ϵ_1 and ϵ_2 are allowed by setting them as a function of the error term (u) in the selection equation in the following fashion,

$$\epsilon_1 = u\pi_1 + v_1 = (z^* - w\alpha)\pi_1 + v_1 \tag{6}$$

$$\epsilon_2 = u\pi_2 + v_2 = (z^* - w\alpha)\pi_2 + v_2 \tag{7}$$

where $cov(v_1, v_2) = 0$, $cov(\epsilon_1, u) = \pi_1$, $cov(\epsilon_2, u) = \pi_2$ and $v_1 \sim N(0, 1)$, $v_2 \sim N(0, \sigma^2)$. The unobservables could affect both selection and outcome and it is reasonable to believe that the unobservables that determine selection are a major source of endogeneity.

Therefore, plugging eqs. (6) and (7) into eqs. (2) and (4) yields,

$$Y_1^* = x\beta_1 + z\tau_1 + (z^* - w\alpha)\pi_1 + v_1 \tag{8}$$

$$\mu = x\beta_2 + z\tau_2 + (z^* - w\alpha)\pi_2 + v_2 \tag{9}$$

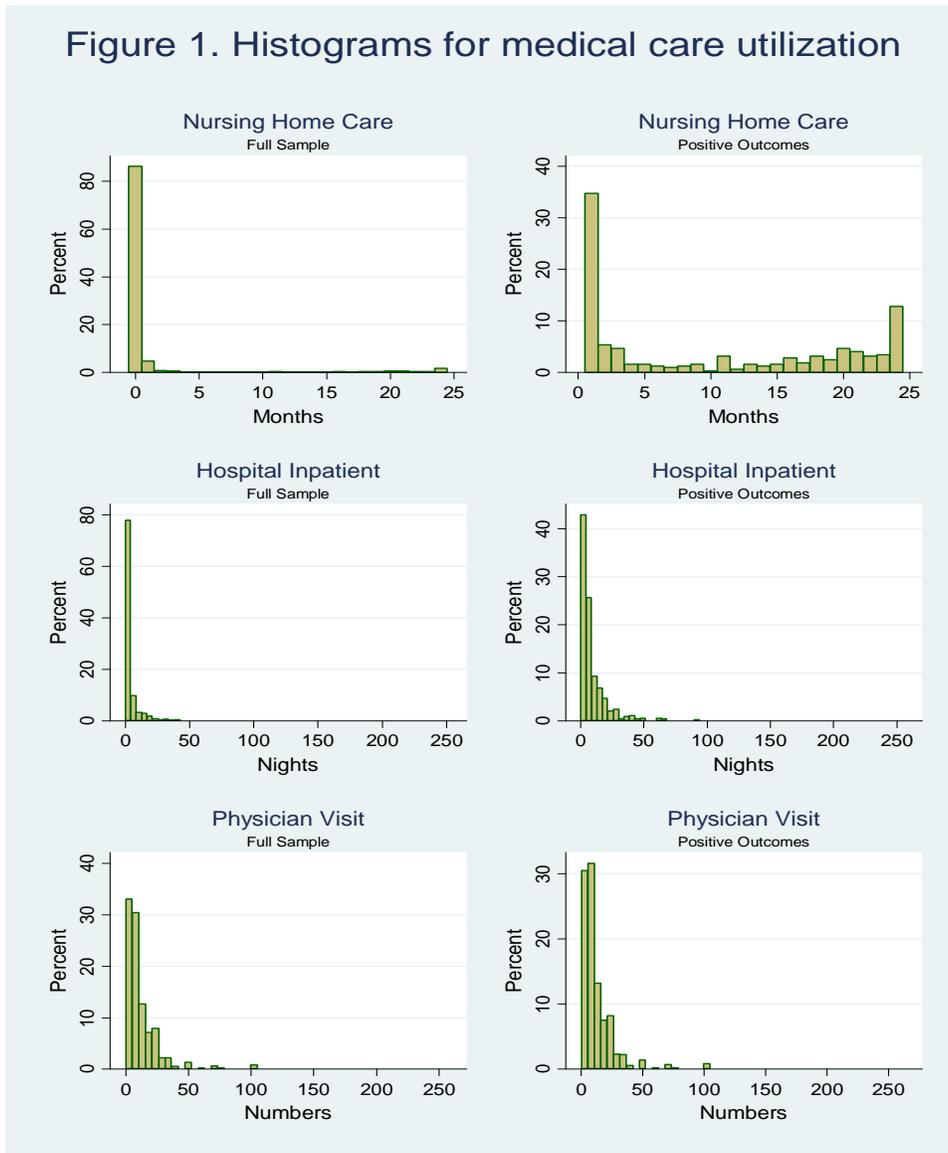
with error variance-covariance matrix,

$$cov \left(\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ u \end{bmatrix} \right) = \begin{bmatrix} \sigma_u^2\pi_1^2 + 1 & \sigma_u^2\pi_1\pi_2 & \pi_1 \\ \sigma_u^2\pi_1\pi_2 & \sigma_u^2\pi_2^2 + \sigma^2 & \pi_2 \\ \pi_1 & \pi_2 & \sigma_u^2 \end{bmatrix} \tag{10}$$

The observed dependent variables consist of (Y_1, Y_2, Y_{nh}, z) with corresponding latent variables $(Y_1^*, Y_2^*, Y_{nh}^*, z^*)$. The endogeneity of informal care can be tested by a joint hypothesis test of $(\pi_1, \pi_2) = (0, 0)$, which will be illustrated in detail in section VIII.

This model is estimated using the Bayesian Markov Chain Monte Carlo (MCMC) procedure because the likelihood function is intractable and does not have analytical solutions. In particular, data augmentation for latent variables, Gibbs sampling, and Metropolis Hasting are applied. The advantages of the Bayesian method are demonstrated by the heterogeneous treatment effects and the test of endogeneity discussed in subsequent sections. A full description of the Bayesian Algorithm can be found in Appendix A. The steps for the MCMC procedure are shown in Appendix B.

Figure 1. Histograms for medical care utilization



IV Data

This paper uses the 2002 and 2004 wave of the Health and Retirement Survey (HRS). The HRS covers a wide range of information including family structure, income, health care, as well as informal care provision. The HRS started as a nationally representative sample at the baseline that consists of non-institutionalized elderly Americans. The first wave began in 1992 and took place every two years thereafter. The data used in this paper is comparatively

more recent than the data used in earlier studies. This is important since the relationship between formal and informal care is likely to change overtime. With modern technology enhancing the communication of patients and doctors and more sons joining daughters in becoming caregivers, the skills required for caregivers may be different from those of the past. This changing nature of caregiving could lead the relationship between formal and informal care to differ.

TABLE 1A Descriptive Statistics

	Full Sample		Treated		Non-treated	
Number of observations	2317		1878		439	
	mean	s.d.	mean	s.d.	mean	s.d.
Independent Variables						
Age	82.60	5.93	85.42	6.08	81.94	5.69
Sex (female=1)	0.79	0.41	0.85	0.36	0.78	0.42
Black(black=1)	0.12	0.33	0.17	0.38	0.11	0.31
Hispanic(Hispanic=1)	0.06	0.25	0.10	0.29	0.06	0.23
Years of education	11.26	3.51	9.77	3.91	11.60	3.31
Ln(total wealth)	9.95	4.05	8.38	4.55	10.31	3.83
Ln(total income)	9.69	0.93	9.41	0.99	9.76	0.90
Medicaid coverage(0/1)	0.16	0.36	0.32	0.46	0.12	0.32
Long-term care insurance(0/1)	0.10	0.29	0.04	0.20	0.11	0.31
Proxy respondent(0/1)	0.16	0.36	0.47	0.50	0.08	0.28
Number of IADL limits	0.90	1.56	2.67	1.85	0.50	1.15
Number of ADL limits	0.80	1.42	2.15	1.82	0.49	1.09
Self-rated health (1:excellent,5:poor)	3.18	1.08	3.73	1.00	3.05	1.06
Number of chronic conditions	2.70	1.40	3.36	1.44	2.55	1.35
Instrumental Variables						
Number of adult children	3.34	2.23	3.84	2.67	3.23	2.21
Whether a child lives within 10 miles	0.62	0.48	0.71	0.45	0.59	0.49
Whether there is a resident child	0.19	0.39	0.34	0.47	0.15	0.36
Number of full-time working children	3.11	2.70	3.08	2.81	3.18	2.67

Note: the sample consists of single elderly aged 70 and older with at least one living adult child. The treated group consists of elderly who received informal care.

Because the main focus of this paper is the demand for long-term care, the sample is restricted to single individuals over 70 years old with at least one living adult child. All variables are summarized in TABLE 1A (independent variables and instruments) and 1B (dependent variables). The Rand HRS data file is a cleaned version of the original data that cover all cohorts and contain consistently derived key variables. All variables are obtained from the Rand

HRS data file except for informal care and children-related variables, which were separately obtained from the original HRS survey. The final data contain 2,317 observations.

Descriptive statistics

The full sample consists of 82% females, in part because women on average outlive men (the sample average age is 82.6 years). The sample is mostly white (12% black and 6% Hispanic) with an average education level of just under a high-school degree and an average wealth of \$252,747. 16% of the sample is eligible for Medicaid and 9% has long-term care insurance⁹.

The survey includes several measures of health status: help with a series of activities of daily living (ADLs) and instrumental activities of daily living (IADLs). The five categories of ADL limitations are walking, dressing, bathing, eating and getting in and out of bed. The five categories of IADL limitations are making phone calls, managing money, taking medication, shopping for groceries and preparing hot meals. The average numbers of ADL and IADL limitations are, respectively, 0.80 and 0.90. Self-rated health status is ranked from 1 to 5 (1 for excellent and 5 for poor). The number of chronic conditions is the sum of doctor diagnosed diseases that include histories of heart attacks, strokes, high blood pressure, diabetes, cancer, lung disease, psychological problems, and arthritis. The average elderly person in the sample has 2.7 chronic conditions. A variable that did not receive attention from prior studies is the proxy variable. A proxy respondent was asked to complete the questionnaire if the respondent herself was not available at the time of the interview or unable to give an interview. Most proxy respondents had physical or cognitive limitations that prevented them from being interviewed and some of them lived in nursing homes. The HRS conducts a test on cognitive abilities and a substantial number of proxy respondents achieved low scores on the test. Therefore, the proxy variable is included in the estimating equation to indicate cognitive functions and reduce omitted variable bias. We observe that 47% of the elderly who had informal care are proxy respondents whereas 8% of those who did not have informal care are proxy respondents.

Informal care is defined as hours of care received from a child, the child's spouse, stepchild, stepchild's spouse and grandchild in the last month. A timing problem occurs because informal care refers to care received in the past month whereas formal care refers to care received over the past two years. I therefore define informal care by using data from the 2002 wave, and I define formal care by using data from the 2004 wave. 428 individuals in the 2004

⁹Based on the HRS, in 2008, 13.8% of individuals aged 60 and older had long-term care insurance. In the late 1990s, only 10% had long-term care insurance. The number of long-term care insurance holders appears to be increasing.

wave were not interviewed in 2002. For these individuals, the informal care variable from the 2004 wave is used instead.

I define receiving informal care as being treated. Descriptive statistics for an average treated individual and a non-treated individual can be found in TABLE 1A. Compared with the average elderly person who did not receive informal care, the average treated elderly was 3.5 years older, more likely to be a female, and had fewer years of education. The average treated elderly person was more likely to be enrolled in Medicaid and less likely to have long-term care insurance. They also had a poorer health status. For example, the average treated elderly person had approximately two more IADL/ADL limitations and one additional chronic condition.

The dependent variables are formal care utilizations that include nursing home care, paid home health care, physician visits, and hospital inpatient care. Descriptive statistics can be found in TABLE 1B. Respondents were asked whether they had been a patient overnight in a nursing home, convalescent home, or other long-term health care facility and, if yes, how long they stayed there in the last two years. Nursing home, hospital care, and physician visits are examined in two ways: whether the service is used (part 1) and intensity of utilization (part 2). Due to a lack of data, paid home health care is only studied as a binary variable (part 1). 14% of respondents used nursing home care in 2004 with an average length of stay of 1.43 months and 10.37 months for users. For the elderly who used nursing home care, the average length of stay was 12.5 months for those receiving informal care and 8.5 months for those not. This shows a positive correlation between informal care and nursing home utilization. Over one-third of nursing home stays were for one month or less, reflecting post-hospital recovery in a skilled nursing home facility. Approximately 9% of the sample were in nursing homes at the time of the interview.

For hospital inpatient care, respondents were asked whether they had been a patient in a hospital overnight and if yes, how many nights they had stayed over the last two years. For the 38% of the sample that used hospital care, the average length of stay was 13 nights for those receiving informal care and 10 for those not, with considerable amount of variation in the sample. With regard to physician visits, respondents were asked how many times they had seen or talked to a medical doctor about their health, including emergency room or clinic visits over the past two years. 96% of the sample had seen a doctor, 13 times on average, 15 times for those receiving informal care and 12 times for those not. With regard to paid home health care, 14% of respondents used paid home health care in the past two years, 22% for those receiving informal care and 11% for those not. In summary, for all types of utilization, the treated group had consistently higher levels of utilization than the non-treated group.

Du: Formal and Informal Care: An Empirical Bayesian Analysis

TABLE 1B Descriptive Statistics Dependent Variables

	Full Sample		Treated		Non-treated	
Number of observations	2317		1878		439	
	mean	s.d.	mean	s.d.	mean	s.d.
Dependent Variables						
Any Informal care	0.19	0.39	1	0	0	0
Informal care hours	22.76	100.81	120.12	204.98	0	0
Any home health care	0.13	0.34	0.22	0.41	0.11	0.32
Any nursing home stay	0.14	0.35	0.35	0.47	0.09	0.28
Months in nursing homes	1.42	4.98	4.29	8.10	0.76	3.58
Months in nursing homes (+)	10.37	9.35	12.47	9.43	8.51	8.90
Any physician visit	0.96	0.19	0.97	0.17	0.96	0.19
Number of physician visits	12.52	15.77	15.32	18.27	11.86	15.06
Number of physician visits (+)	12.99	15.96	15.71	18.40	12.36	15.20
Any hospital care	0.38	0.48	0.51	0.50	0.35	0.48
Nights in hospital	4.21	13.33	6.67	15.98	3.63	12.58
Nights in hospital (+)	10.91	19.70	12.64	20.27	10.36	19.47

Note: our sample consists of single elderly aged 70 and older with at least one living living adult child. The treated group consists of elderly who received informal care. (+) indicates positive outcomes.

Exclusion restrictions

As in the case of the linear simultaneous equations model, the regressors in the selection equation include one or more variables that can be excluded from the outcome equation. In this paper, variables related to children's characteristics are considered potential instruments, and two sets of instruments are explored. The variables under consideration are whether there is a child who lives within 10 miles, whether there is a resident child, the number of full-time working children, the number of adult children, the number of adult daughters, and the number of unmarried adult children.¹⁰ All variables are from the 2002 wave, as in the case of the informal care variable. Instruments are selected based on the F-statistics that indicate the strength of the instruments, the R^2 of the regression, and the statistical significance of each variable. As shown in TABLE 2, after controlling for geographical location, employment status, and the number of children, the gender and marital status of children are

¹⁰I also considered the number of frequent contact, children's income, the number of children's offsprings, and whether there is over \$500 financial support from children as exclusive restrictions, however the F statistics for these variables is 0.39, indicating weak instruments thus they are not included in this paper.

not statistically significant.¹¹ These variables are therefore not included as exclusion restrictions.

The exclusion restrictions used in this study are whether a child lives within 10 miles, whether there is a resident child, the number of full-time working children, and the number of adult children (TABLE 2 column 2). These four instruments are applied in all regressions; together, they explain 36% of the variation in informal care.

The instrumental variable approach identifies the causal effect of informal care through marginal individuals for whom variations in informal care are induced by instruments measuring the availability of children. As most single elderly persons are cared for by their children if they are cared for informally, the marginal elderly person represents a typical senior who has children. They are more likely to be individuals who have needs for informal care such as those with chronic conditions and disabilities and those who do not favor long-term care in an institutional setting (such as females). They are also likely to be middle-income seniors instead of the very rich or the very poor. The very poor would qualify for Medicaid and the very rich would not hesitate to purchase high-quality formal care for themselves.

TALBE 2 Selection of Instruments

Instrument	Set 1	Set 2
Whether a child lives within 10 miles	0.123* (0.057)	0.124* (0.057)
Whether there is a resident child	0.854* (0.073)	0.842* (0.071)
# of working children	-0.043* (0.017)	-0.038* (0.015)
# of children	0.074* (0.028)	0.086* (0.019)
# of daughters	0.046 (0.029)	
# of unmarried children	-0.024 (0.032)	
R-squared	0.363	0.363
F statistics for coefficients = 0	32.83	48.49

Note: the dependent variable is the natural log of informal care hours. Age, gender, education, assets and income, insurance, proxy variable. Health conditions of the elderly are included as control variables. * indicates statistically significant at 5%. Standard errors are reported in parentheses.

The average number of children for the full sample is 3.34. The number of children and full-time working children for the elderly who received informal care is 3.87 and 3.35, respectively, slightly higher than the number of children for those who did not receive informal care (3.24 and 3.13). We also observe

¹¹Some studies find that daughters have a more significant role than sons do in providing informal care (e.g. Lo Sasso and Johnson, 2002), which is not supported in our sample.

that 62% of the elderly in the sample have at least one child living within 10 miles and 19% have at least one resident child. The instruments are highly predictive with an F statistic of 48.49.

The number, gender and geographic location of children have been considered by previous studies (Chevkovich et al., 2002; Stern, 1995; Charles and Sevak, 2005; Bolin et al., 2007; Bonsang, 2008), but the working conditions of children have not. This paper argues that the number of full-time working children is one of the most important factors determining the availability of caregivers. Work status represents the opportunity cost of caregiving and this opportunity cost is expected to increase as more daughters join the labor force. As shown in TABLE 2, additional full-time working children is associated with a decrease in informal care hours by 3.8%. Given that the mean number of hours of informal care is 23 for the full sample and 120 for the treated, 3.8% translates to a 0.9 hour reduction in informal care during one month for the full sample and a 4.5 hour reduction for the treated.

Some of the instrumental variables deserve further discussion. On the one hand, children may choose to move closer or live with parents if their parents have long-term care needs. On the other hand, parents may move in and live with children due to deteriorating health conditions. The same logic applies to the number of full-time working children, as a child may quit his/her job to take care of their elderly parents if they cannot afford living independently. Children's characteristics as instruments for informal care have been used and tested in the literature (Charles and Sevak, 2005; Bolin et al., 2007), and no evidence of endogeneity has been found with regards to the number of children and the geographic location of children.

Two additional methods are used to check the validity of the exogeneity of the instruments. First, I regress each instrument on elderly persons' health conditions and a set of control variables (x in eq. (2)). There are four health indicators included in the regression: the number of ADL, the number of IADL, self-rated health, and the number of chronic conditions. If any health indicators are related to the children's characteristics, the exogeneity assumption is likely to be violated. I find that almost none of the health indicators are statistically significant in predicting location and work status.¹²¹³ Second, a refutability test is performed. A subsample of very healthy elderly (no ADL, IADL, or chronic conditions) is chosen, that I do not expect their formal care be affected by informal care. The exogeneity assumption is likely to be

¹²The only exception is that the number of chronic conditions is statistically significant at 5% when used to predict whether there is a resident child. For robustness check, I excluded this variable, the results for all regressions did not vary much and therefore it is not discussed further.

¹³Charles and Sevak (2005) find that health conditions of the elderly have no prediction power for the variable that indicates children's location.

violated if children’s characteristics are associated with formal care for reasons other than informal care. However, no statistical associations are found between the selected children’s characteristics and formal care for the very healthy subsample, and this supports the exogeneity assumption.

V Results

The results for nursing home care are presented in detail. The results for other types of formal care are presented in the second subsection. Short-term and long-term care and other subsample analyses are presented in the third subsection. A sensitivity test is performed at the end of this section.

TABLE 3 Effect of Informal Care on Nursing Home Utilization

	Part1 (hurdle)	Part2 (conditional)	Informal care
Informal care	-0.181 (0.024)	-0.389 (0.104)	
Age	0.003 (0.002)	0.023 (0.009)	0.010 (0.004)
Sex (female=1)	0.012 (0.024)	0.036 (0.147)	0.181 (0.054)
Black	-0.174 (0.043)	0.215 (0.194)	0.025 (0.060)
Hispanic	-0.090 (0.032)	-0.036 (0.281)	-0.233 (0.082)
Years of education	0.003 (0.003)	0.027 (0.018)	-0.048 (0.006)
Ln (total wealth)	-0.007 (0.003)	-0.031 (0.013)	0.022 (0.006)
Ln (total income)	-0.009 (0.011)	0.064 (0.050)	-0.121 (0.023)
Medicaid coverage	0.126 (0.042)	0.576 (0.142)	-0.131 (0.059)
Long-term care insurance	-0.016 (0.033)	0.146 (0.263)	-0.115 (0.087)
Proxy respondent	0.174 (0.049)	0.283 (0.148)	0.369 (0.058)
# of IADL limits	0.089 (0.026)	0.209 (0.062)	0.507 (0.017)
# of ADL limits	0.054 (0.016)	0.129 (0.041)	0.083 (0.018)
Self-rated health	0.021 (0.012)	0.088 (0.060)	0.165 (0.022)
# of chronic conditions	0.022 (0.009)	-0.074 (0.043)	0.164 (0.015)
# of adult children			0.137 (0.012)
A child lives within 10 miles			0.344 (0.043)
Resident child			1.269 (0.045)
# of working children			-0.057 (0.011)
Variance		0.867 (0.070)	2.095 (0.064)
π_1, π_2	0.468 (0.083)	0.404 (0.103)	

Notes: posterior means are reported and standard deviations are in parentheses.

Part 1 is a probit model estimating nursing home utilization. The marginal effects are presented. Part 2 is a log-linear model for intensity of use.

Results for nursing home care

Parameter estimates for nursing home care are reported in TABLE 3. Part 1 is a *probit* model estimating the probability of entering nursing homes, and marginal effects are reported. The coefficients in the nursing home utilization equation are interpreted as semi-elasticities because the dependent variable is transformed using logarithm.

After controlling for endogeneity, I find that informal care is a substitute for nursing home care. A 10% increase in informal care hours leads to a 1.8% point decrease in the probability of entering nursing homes and a 3.9% decrease in the length of stay conditional on nursing home users¹⁴. This translates to a 12-day reduction in a two-year period. Being Black or Hispanic decreases the probability of entering nursing homes by 17% point and 9% point, respectively. Medicaid eligibility increases the probability of entering a nursing home by 13% point and increases the length of stay by 57%. Long-term private insurance does not have an effect on the probability of entering nursing homes or on the length of stay. A small percent (9%) of the sample had long-term private insurance, and it is plausible that those with long-term care insurance are more risk averse in health fluctuations and are thus less likely to have chronic conditions. One additional IADL and ADL limitation increases the probability of entering a nursing home by 9% point and 5% point, respectively, and increases the length of stay by 42% and 25%, respectively.

The coefficients for the error term (u) in the selection equation are reported in the last row of TABLE 3. These coefficients represent the effect of the unobservables on formal care utilization. The results show that the unobservables increase the probability of entering a nursing home as well as the length of stay. For example, there may be certain health conditions known to the elderly but unknown to the researchers that increase both formal and informal care. Endogeneity can be tested using the Bayesian method by comparing the likelihood of the model that restricts $\pi_1 = \pi_2 = 0$ with that of the model that allows them to vary. Details for the test are presented in section VIII and appendix A.

The determinants of informal care are presented in the last column of TABLE 3. Having an additional child and having a child who lives within 10 miles increases the hours of informal care by 14% and 34%, respectively,¹⁵

¹⁴The marginal effects of this paper are generally larger compared to studies that also use the U.S. data. For example, Van Houtven and Norton (2004) show that a 10% increase in informal care hours decreases the probability of nursing home use by 0.77% point and decreases the length of stay by 1.8%. Other studies (Charles and Sevak, 2005; Lo Sasso and Johnson, 2002) that use informal care as a binary variable find a 40–60% point decrease in the probability of entering nursing homes with frequent children's help.

¹⁵Charles and Sevak (2005) find that having a child nearby is associated with a 12-16% point higher probability of receiving informal care.

which translates to an increase of 3 and 11 hours of informal care per month. Among the four instruments, the presence of a resident child has the largest coefficient and is also the strongest indicator for informal care among all the variables. The elderly who live with a child had 29 more hours of informal care per month than those who did not have a resident child. In the sample 33% of elderly persons with a resident child received informal care compared to 15% for those without a resident child. All instruments are precisely estimated with small standard errors. Other factors that have positive effects on informal care are being female, being a proxy respondent and having poor health conditions. Being Hispanic, having more years of education, a higher income, Medicaid eligibility and having private long-term care insurance decrease informal care hours.

Results for other types of formal care

We next examine the estimates for other types of care shown in TABLE 4. Informal care is found to be a substitute for nursing home and hospital inpatient care. A 10% increase in informal care hours reduces the probability of hospital inpatient stays by 0.31% point and reduces the length of stay by 1.21 nights. The availability of informal care from children may lead to a quicker discharge from nursing homes or hospitals.

TABLE 4A Effect of Informal Care for All Types of Formal Care

	Probability (part1)	Intensity (part2)
Nursing home (months)	-0.181 (0.024)	-0.389 (0.104)
Hospital care (nights)	-0.031 (0.016)	-0.121 (0.062)
Physician visits (numbers)	0.039 (0.043)	-0.118 (0.035)
Home health care (prob.)	0.057 (0.068)	-

Note: posterior means are reported and standard deviations are included in parentheses.

I also find that informal care does not affect the probability of physician visits but does decrease the number of visits. This is consistent with Van Houtven and Norton (2004). Based on the distribution of the treatment effect (shown in Figure 2), large positive and negative effects are found; however, they cancel out after being averaged. As discussed above, children can facilitate transportation to clinics, which increases the probability of physician visits. The availability of children can also make one more visit to the doctor less attractive because it is less costly.

For paid home care, informal care is shown not to have a significant effect. This result departs somewhat from those of Van Houtven and Norton (2004),

who find them to be substitutes. A further investigation shows that informal care only serves as a substitute when the elderly is relatively healthy and has no need for full-time informal care. To demonstrate this result, I decompose sample into frail and non-frail. Frail is defined as having at least one IADL or ADL limitation. The coefficient of informal care is positive (0.102 (0.063)) for the frail sample and negative (-0.378 (0.665)) for the non-frail sample, which is consistent with the findings in Spillman and Pezzin (2000).

Subsample analysis

Several subsample analyses would be useful in examining the heterogeneous effects of informal care. First, I distinguish short-term and long-term nursing home stay. Short-term users stay in nursing homes primarily for purposes of rehabilitation or restoration of health from a sudden negative health shock, such as a knee replacement surgery or a cardiovascular procedure. Long-term users are individuals with severe disabilities or chronic conditions, such as dementia or Alzheimer's disease, and who have difficulty living independently. The relationship between formal and informal care may be different for the two types of users due to the nature of the utilization. As Medicare pays for skilled nursing facilities for no more than a hundred days, and only if the utilization follows a consecutive hospital stay of at least three days, I separate the sample based on length of nursing home stay as well as hospital stay. Long-term users are defined as those who stay in nursing facilities for more than a hundred days.¹⁶ There are 176 long-term users (approximately 8% of the sample) and 143 short-term users (approximately 6% of the sample).

Second, the sample is separated based on the length of hospital stay. Individuals with at least three consecutive days of inpatient hospital stay qualify for Medicare reimbursement for the subsequent nursing facility utilization. Qualification for Medicare reimbursement could potentially make informal care less attractive. However, the HRS only contains information on the total number of nights in a hospital during the two years before the interview, and the length of each stay is therefore unknown. Because 58% of the hospital users had only one hospital stay in the past two years, I define the average length of stay as total nights divided by number of stays. There are 164 elderly persons who averaged at least three nights in a hospital and they had also entered nursing homes.

The results by length of stay are presented in TABLE 4B. Only intensity of utilization is examined in this analysis. I find that informal care has a larger effect for long-term nursing home users (-0.162 (0.075)) than for short-term users

¹⁶85% of the nursing home users stayed in nursing homes for only once during the two years before the interview.

(-0.012 (0.06)). For those who had on average less than three days of hospital inpatient care, the effect of informal care is the largest across all subsamples. This confirms the expectation that the substitution effect of informal care is larger when nursing home utilization does not meet Medicare reimbursement criterion. This result may have implications for long-term care policies. For example, policy makers have discussed removing the three-day hospital stay requirement. Based on our analysis, this proposal would potentially increase Medicare spending by making family members less likely to provide informal care.

TALBE 4B Effect of Informal Care on Nursing Home Utilization by Length of Stay

Subsample definition	Informal care coefficient	Sample size
Long-term NH (> 100 days)	-0.162 (0.075)	176
Short-term NH (between 0 and 100)	-0.012 (0.060)	143
Hospital care (≥ 3 nights)	-0.206 (0.147)	164
Hospital care (between 0 and 3)	-0.852 (0.313)	155
Short-term NH + hospital ≥ 3 nights	-0.055 (0.048)	104
Long-term NH + hospital ≥ 3 nights	-0.335 (0.123)	60

Note: Part 2 of the two-part model and the equation for informal care are used in this analysis. Standard deviations are in parentheses.

To check the robustness of this result, I examine two additional subsamples by length of stay: short-term nursing home users with at least three nights in a hospital and long-term nursing home users with at least three nights in a hospital. I expect to find little effect of informal care for the first subsample because payment is covered by Medicare. For the second subsample, informal care may potentially reduce long-term care costs for the elderly because the cost above a hundred days cannot be reimbursed unless the person is eligible for Medicaid. The results are consistent with the expectation that informal care acts as a substitute for nursing home utilization for long-term users but does not do so for short-term users.

Sensitivity test

The insurance variables are potentially endogenous. Medicaid eligibility can come from spending-down due to nursing home utilization, and long-term care insurance purchase decisions may be made with expected utilization in mind.

According to America's Health Insurance Plan, 2007, the average age of long-term care insurance buyers was 61 years old in 2005, and 84% of buyers were younger than 70 years old. 71% of buyers had an income of more than \$50,000, and 61% were college graduates. It can be argued that because in-

insurance choices were typically made at an earlier age and by individuals in high income categories, the endogeneity problem of long-term care insurance may be mitigated because our sample consists of those above 70 years old. Nevertheless a sensitivity analysis is performed. I focus on two subsamples that are likely to cause endogeneity: those who had long-term care insurance and those who newly purchased long-term care insurance. Approximately 9% of the sample had insurance in 2004. 2.4% of the sample had long-term care insurance in 2004, but did not have it in 2002, and we would expect that new buyers would be more likely to use formal care. The two subsamples are separately excluded from the full sample and the results for nursing home utilization are presented in TABLE 4C. No significant changes in estimates are detected. The only noticeable change is that the effects of informal care are slightly smaller for the subsamples.

TABLE 4C Sensitivity Analysis

	Informal Care	Medicaid	LTC insurance
(1) Full sample	-0.389 (0.104)	0.576 (0.142)	0.146 (0.263)
(2) Exclude LTC insured	-0.295 (0.088)	0.571 (0.136)	NA
(3) Exclude newly LTC	-0.298 (0.089)	0.600 (0.136)	0.329 (0.270)
(4) Exclude newly Medicaid	-0.356 (0.100)	0.658 (0.149)	0.177 (0.262)

Note: the coefficients shown are for Part 2 of the two-part model. The coefficients for Part 1 are largely unchanged. Standard deviations are in parentheses.

For Medicaid, endogeneity arises due to individuals' spending-down their assets. I therefore exclude the subsample that was not eligible for Medicaid in 2002 but became eligible in 2004. 5.19% of the sample belongs to this category. Again I do not find the coefficient of informal care to be significantly different from that of the full sample. The coefficient for Medicaid is largely unchanged as well. Additional estimations are performed by excluding the Medicaid and the long-term care insurance variables, and the informal care coefficient is found to be robust.

VI Treatment effect and its distribution

The distribution of treatment effects is generally ignored in a frequentist non-Bayesian approach, but it can be conveniently obtained from the posterior distributions of the Bayesian MCMC procedure. The nature of health care utilization especially long-term care warrants further examination of the distribution of treatment effects because a simple mean treatment effect would not be enough to reveal the entire picture. In this section I focus on the distribution of treatment effects that consist of individual treatment effects.

Paid home care is omitted from this exercise because data are not available for intensity of utilization.

For a randomly chosen individual, the expected difference in outcome between the treated (defined as those who had informal care) and the non-treated is the average treatment effect (ATE) defined as

$$E(y_1 - y_0 | \Delta) = E(y_1 | \Delta) - E(y_0 | \Delta) \tag{11}$$

where $\Delta = (\beta_1, \beta_2, \tau_1, \tau_2, \alpha, \pi_1, \pi_2, \sigma^2, z^*)$. The expected difference conditional on being treated is the average treatment effect on the treated (ATET) defined as

$$E(y_1 - y_0 | \Delta, d = 1) = E(y_1 | \Delta, d = 1) - E(y_0 | \Delta, d = 1) \tag{12}$$

where y_1 is the outcome for the treated, and y_0 is the counterfactual (the outcome if the elderly were not treated), and $d = 1$ indicates treatment.

The ATE and the ATET are presented in TABLE 5. The mean treatment effects for both ATE and ATET are negative for nursing home care, hospital care and physician visits. I find that an elderly person who had informal care would have stayed in nursing homes for 2.3 months longer if she had not had informal care. All results have large standard deviations, indicating heterogeneous individual treatment effects.

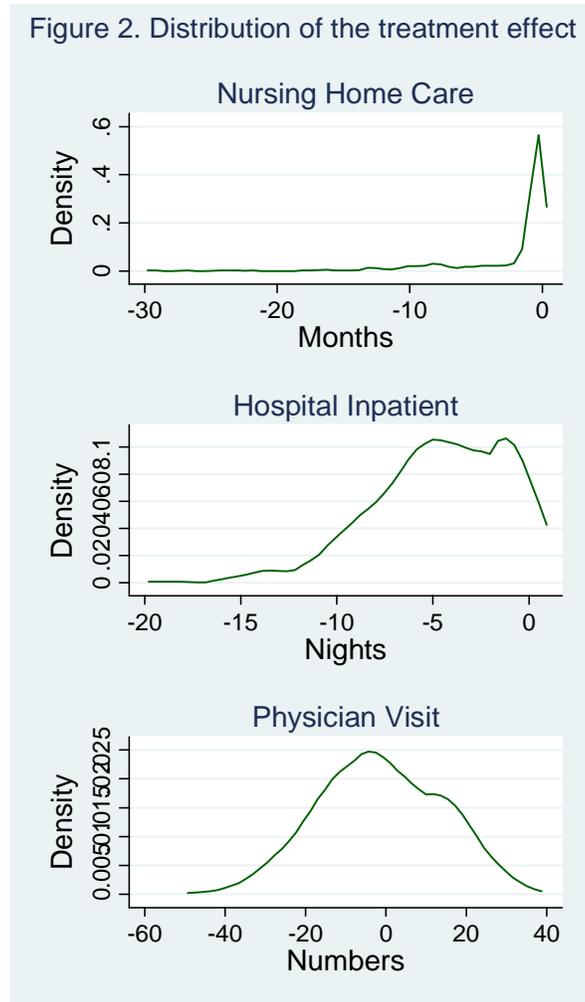
TABLE 5 Average Treatment effect and Average Treatment Effect on the Treated

	ATE	ATET
	mean (st.dev)	mean (st.dev)
Nursing home (months)	-0.44 (2.06)	-2.31 (4.27)
Hospital inpatient (nights)	-0.86 (2.33)	-4.56 (3.45)
Physician visits (numbers)	-0.25 (6.67)	-1.30 (15.30)

The distributions of ATET are shown in Figure 2. There are large variations depending on the types of care. Compared with hospital inpatient care and physician visits, the distribution for nursing home care has a sizable left tail, which reveals large substitution effects for long-term nursing home users. According to Brown and Finkelstein (2009), more than 10% of elderly persons live in nursing homes for more than 5 years, and our results indicate that these extremely long stays can be substituted by informal care. Sorting the ATET by its absolute value from largest to smallest, I find that the top 10% would have stayed in nursing homes for 8 months longer in a two-year span if they had not received informal care, which is considerably larger than the mean treatment effect. Compared with an average treated person, the top 10% are more likely to be older (average 86 years old), Black (28%), rather than Hispanic (2%) and to have more children (average 4.22). This subpopulation should be the focus of long-term care policies.

The distribution of treatment effects for hospital is also skewed to the left, though less so than nursing homes, with a median of -4.6 and minimum of -19. For physician visits, the distribution is symmetric around the mean. The range of the distribution is very large (max = 35, min = -45). This suggests that informal care and physician visits are substitutes for approximately 50% of the sample, and complements for 48% of the sample, and neither serves as a substitute nor a complement for the remaining 2%. Again, if we sort the ATET into its absolute value from largest to smallest, the top 10% would have visited doctors 21 more times in the two-year span if they had not received informal care. A careful examination of this subsample reveals that they are better educated compared to the average treated person.

Figure 2. Distribution of the treatment effect



Treatment effect at the intensive and extensive margins

In this subsection, I simulate the changes of informal care at the intensive and extensive margins and examine its impact on formal care utilization. The results of this exercise can be used to design policies that may reduce costs for Medicare and Medicaid. On the intensive margin, I let informal care hours increase from 34 hours (median level) to 160 hours (full-time care) for elderly who had informal care because most of them received part-time care. On the extensive and intensive margins, I let informal care increase from 0 to 34 hours and from 0 to 160 hours, respectively, for those who did not receive informal care.

TABLE 6 presents the results. On the intensive margin, increasing informal care from median to full-time leads to a 6.5 months reduction in nursing home care, a 2.6 nights reduction in hospital stay and a 3 times reduction in physician visits in a two-year period. Increasing informal care from 0 to 34 hours leads to a 0.34 months (≈ 10 days) decrease in nursing home care, a 1.3 nights decrease in hospital stays, and a 1.7 times decrease in physician visits in a two-year period. When informal care hours are increased from 0 to full-time, the mean treatment effects are a 1-month decrease in nursing home care, a 2.4 nights decrease in hospital stay and a 5.4 times decrease in physician visits. In summary, informal care has the largest substitution effects for nursing home and hospital inpatient care at the intensive margin.

TABLE 6 Treatment Effects for Formal Care Utilization

	Intensive margin	Intensive and extensive margins	
	Median to full	Zero to median	Zero to full
Nursing home (months)	-6.48 (7.94)	-0.34 (1.05)	-1.04 (3.33)
Hospital inpatient (nights)	-2.55 (1.28)	-1.30 (0.74)	-2.36 (1.51)
Physician visits (numbers)	-3.14 (1.45)	-1.70 (0.74)	-5.43 (2.16)

Note: treatment effects at the intensive margin are calculated for the treated group.

Treatment effects at the intensive and extensive margins are for the non-treated.

Median level of care is 34 hours and full-time of care is 160 hours per month.

VII Policy analysis

Government cost-saving analysis is performed in this section. I focus on nursing home and hospital care because large substitution effects are found for them. In 2009, long-term care accounted for 32% of total Medicaid expenses and institutional care accounted for 45% of overall long-term care expenditure (Eiken et al., 2010). Medicare does not cover long-term care per se, but it covers skilled nursing facilities following hospitalization, and it does not cover any stay of more than a hundred days. Because Medicaid and Medicare cover

different types of care, in the following analysis I examine policy implications by payer type.

Because most tax-credit programs require the caregiver to provide full-time care, two potential policies are studied based on the simulation in the previous section. The estimates in TABLE 6 are used as references in this section. The first is to let an average non-treated elderly person receive full-time care from children. Because the per diem cost of staying in a nursing home was \$181 in 2003 (p.x, CBO, 2004), total saving by informal care would be \$5,647 ($\$181 \times 30 \times 1.04$). According to the Center for Medicare & Medicaid Services national health expenditure data, in 2004, Medicaid paid 44.7% of nursing home expenditure and Medicare paid 14.7%. The cost saving for Medicaid would be \$2,524 ($0.447 \times 5,647$) and that for Medicare would be \$830 ($0.147 \times 5,647$) in a two-year period. The second policy intervention is to have an average median-level cared-for elderly person receive full-time care. The total savings that would result from informal care would equal \$35,186 ($\$181 \times 30 \times 6.48$). The cost saving in a two-year period for Medicaid and Medicare would be \$15,728 and \$5,172, respectively. The reduction in spending is larger for the second policy because informal care has a larger substitution effect.

For hospital inpatient care, the treatment effects for the two policy interventions are -2.36 and -2.55, respectively. Medicare pays 29.2% of hospital spending, and Medicaid pays approximately 17.1%.¹⁷ Based on the estimates of Agency for Healthcare Research and Quality (AHRQ), the average charge for one-day of hospital inpatient care for a Medicare patient is \$4,455 and \$3,456 for a Medicaid patient. Under the first policy, Medicare would save \$3,070 ($4,455 \times 2.36 \times 0.29$), and Medicaid would save \$1,394 ($3,456 \times 2.36 \times 0.17$) in a two-year period. Under the second policy, the cost saving for Medicare and Medicaid would be \$3,317 ($4,455 \times 2.55 \times 0.29$) and \$1,507 ($3,456 \times 2.55 \times 0.17$), respectively. In the case of hospital care, the two policies yield similar amount of savings, but yield a greater amount for Medicare than for Medicaid.

TABLE 7 summarizes the simulation results. The first policy provides similar cost saving for Medicare and Medicaid whereas the second policy provides more cost saving for Medicaid. If both policies are implemented, the total savings would reach \$12,389 for Medicare and \$21,153 for Medicaid (both federal and state) in a two-year period for an average elderly person. The cost saving for Medicaid is much larger than for Medicare because Medicaid is the main payer for long-term care.

To provide care to the elderly, caregivers may incur negative labor market outcomes, therefore we need to take into account the loss in tax revenue due to caregivers who quit their jobs or shorten working hours. Carmichael et

¹⁷The Center for Medicare & Medicaid Services national health expenditure data.

al. (2010) report that people who are not employed or who are in low-paid jobs are more likely to provide care. Based on our sample, the median annual income for an adult child caregiver is approximately \$24,000, which is subject to a 15.3% payroll tax (half paid by employer and half paid by employee) and a federal income tax that I assume to be 15%. For a full-time worker who quits her job, the tax revenue loss in a two-year span is around \$14,000 ($2 \times 24,000 \times (15\% + 15.3\%)$). For a part-time employee who quits her job or a full-time employee who switches to a part-time job, the loss in tax revenue would be smaller. This suggests a net gain for the government when informal care hours are increased from medium to full time (the second policy). Under the first policy, the government could incur either a net gain or a net loss depending on how large the reduction in tax revenue is.¹⁸

TABLE 7 Cost Savings for Medicare and Medicaid

		Nursing home	Hospital Inpatient	Total savings
Medicare	Simulation 1	830(\$)	3,070(\$)	3,900(\$)
	Simulation 2	5,172	3,317	8,489
	Subtotal	6,002	6,387	12,389
Medicaid	Simulation 1	2,524	1,394	3,918
	Simulation 2	15,728	1,507	17,235
	Subtotal	18,252	2,901	21,153

Note: Simulation 1 targets the non-treated group and let informal care hours increase from 0 to full-time. Simulation 2 targets the treated group and let informal care hours increase from median to full-time.

To conclude, the cost saving for Medicaid and Medicare largely derive from the second policy that targets elderly persons who had received some informal care from children. The cost saving for Medicaid arises mainly from substituting informal care for nursing home utilization whereas the saving for Medicare is largely from reductions in the utilization of hospital inpatient care. Because there exists large heterogeneity in the utilization patterns (as shown in Figure 2), informal care policies should focus more on the subgroup that incurs the largest substitution effects in the distribution.

VIII Test of endogeneity

Bayesian estimation provides a convenient tool for testing the endogeneity of

¹⁸In order to obtain more precise and aggregate estimates of total savings, we will need information on public insurance savings and costs, state tax revenue losses, costs or savings in other types of formal care, indirect cost of informal care, and the number of people who are eligible for the tax credit.

informal care by comparing the ratio of marginal likelihoods for two models. The restricted model imposes $(\pi_1, \pi_2) = (0, 0)$, whereas the unrestricted model allows π_1 and π_2 to vary. The Bayes factor is used to test endogeneity and the results are reported in TABLE 8. For nursing home care, the Bayes factors calculated using the Savage Dickey density ratio are very close to zero for all three priors ($\gamma = \frac{1}{4}, \gamma = \frac{1}{8}, \gamma = \frac{1}{20}$) (see Appendix A.), which supports the unrestricted model that allows endogeneity. The Bayes factors hospital inpatient care and paid home health care are greater than unity, which supports the restricted model, indicating no endogeneity. For physician visits, the Bayes factors suggest the restricted model at ($\gamma = \frac{1}{4}, \gamma = \frac{1}{8}$) but indicate the existence of endogeneity when $\gamma = \frac{1}{20}$.

These results show that informal care is endogenous for nursing home care, but I do not find informal care to be endogenous for hospital and paid home health care. For physician visits, the evidence is mixed. These results suggest that caution should be exercised when we estimate the relationship between informal care and nursing home utilization.

TABLE 8 Is Informal Care Endogenous?

	$\gamma = \frac{1}{4}$	$\gamma = \frac{1}{8}$	$\gamma = \frac{1}{20}$	Endogeneity?
Nursing home care	0.000	0.000	0.000	yes
Hospital inpatient	3.413	1.581	1.083	no
Physician visit	3.322	1.975	0.828	mixed
Home health care	7.000	5.044	3.268	no

Note: Bayes factor >1 indicates no endogeneity.

IX Conclusion

This paper estimates the effects of informal care on formal care utilizations using an endogenous two-part model that treats informal care as endogenous. The results are based on the Bayesian MCMC procedure and the data augmentation method that generates latent variables. The main finding of this paper is that informal care serves as a substitute for nursing home and hospital inpatient care. Elderly persons who received informal care were less likely to enter nursing homes and hospitals. In addition, the length of stay was shortened if they were admitted to nursing homes or hospitals. In studies using a European dataset, Bonsang (2009) find that the substitution effect disappears for the more disabled elderly, and Bolin et al. (2007) find hospital visits and informal care to be complements rather than substitutes, which is clearly a departure from the results of this paper and other papers that use the U.S.

data.

This study also finds that informal care does not affect the probability of physician visits, but it does reduce the number of visits. For paid home health care, the effects of informal care are statistically insignificant, which is somewhat unexpected. Further analysis shows that informal care is a complement to paid home health care for disabled or more impaired elderly, but acts as a substitute for the relatively healthy elderly. Compared with previous studies in the literature, larger substitution effects of informal care are found. The marginal effects almost double those of Van Houtven and Norton (2004), who applied a different modelling strategy and used data from the 1990s. This could be due to several factors. First, in more recent years, the elderly have been less disabled and are thus more able to fulfill their needs through informal care utilization. Second, recent development in tele-care that facilitates monitoring health conditions at home through technological advances may have enhanced the effectiveness of informal care. Third, the changing demographics of caregivers may play a role.

Adding to the literature on informal care, this paper explores the heterogeneous treatment effects. I find that the substitution effect of informal care is larger for long-term nursing home users than for short-term users. I find that the substitution effect is much larger for those who had less than three days of hospital stay and who are thus unlikely to qualify for Medicare skilled nursing facility reimbursement. These results have implications for policies that call for expanding Medicare coverage (such as, dropping the requirement of a three-day prior hospitalization and increase the one-hundred-day limit) to cover more long-term care related services.

Average treatment effects are calculated and the distributions of the treatment effects show great heterogeneity in the elderly population. In particular, large substitution effects are found for older African American women who have multiple children. Treatment effects are calculated at both the intensive and extensive margins. The largest substitution effects occur for nursing home and hospital care at the intensive margin. The policy analysis shows that an appropriate amount of subsidy can induce children to provide full-time care, thus potentially reducing government spending on Medicaid and Medicare. The cost saving for Medicaid is larger than that for Medicare and depends on the work status of the caregivers.

Future research may focus on other types of long-term or mixed types of care. For example, there is an increase in the use of assisted living facilities in the United States, which cost about half as much as nursing homes (MetLife Mature Market Institute, 2009). Assisted living facilities tend to attract middle-income seniors with low or moderate levels of disability. As this is a different demographic and socioeconomic group, the relationship between

informal care and the use of assisted living facilities may not be the same. Other types of care, such as intensive service days in nursing homes in an attempt to avoid hospitalization are shown to reduce admissions as well as length of stay if admitted (Kane et al., 2002). Future studies may also focus on the trade-off between outcome and cost. In some cases, better outcomes may be achieved for institutional care when safety becomes an issue. However, the extent to which outcomes are compromised by using informal care or other non-institutionalized care is still unknown.

Appendix A: Bayesian algorithm

The Bayesian estimation algorithm is presented in the following order: (1) the posterior distributions of the parameters, priors and the likelihood function; (2) the treatment effects of informal care on formal care utilization; and (3) the test for the endogeneity of informal care.

Prior, likelihood and posteriors

A normal prior with large variance is used for the coefficients $\beta_1, \beta_2, \tau_1, \tau_2, \alpha$ (Chib, 1992) and a prior with relatively smaller variance is used for the coefficients on informal care (π_1, π_2) ,¹⁹ and it follows that

$$\begin{aligned} \beta_1 &\sim N(0_k, 10I_k) & \beta_2 &\sim N(0_k, 10I_k) \\ \tau_1 &\sim N(0_k, 10I_k) & \tau_2 &\sim N(0_k, 10I_k) \\ \alpha &\sim N(0_k, 10I_k) \end{aligned} \quad (13)$$

$$\pi_1 \sim N(0_k, \gamma I_k) \quad \pi_2 \sim N(0_k, \gamma I_k) \quad (14)$$

where three sets of informative priors ($\gamma = \frac{1}{4}, \gamma = \frac{1}{8}, \gamma = \frac{1}{20}$) are used to test for sensitivity and endogeneity. The natural conjugate prior for the error variances σ^2 and σ_u^2 is the inverse gamma density, which can be expressed as

$$\sigma^2, \sigma_u^2 \sim invG(c_0/2, (d_0/2)^{-1}) \quad (15)$$

where $c_0 = 5$ and $d_0 = 10$.

The likelihood for this model is the joint density of $z^*, z, Y_1^*, Y_1, Y_2^*$, and Y_2 , in which both latent variables and observed variables are included.

Let $\phi = (w, x, \beta_1, \beta_2, \tau_1, \tau_2, \alpha, \pi_1, \pi_2, \sigma^2, \sigma_u^2)$, one can write the joint density

¹⁹A proper informative prior for π_1 and π_2 is needed for the test of endogeneity.

as

$$\begin{aligned}
 p(z^*, z, Y_1^*, Y_1, Y_2^*, Y_2 \mid \phi) &= p(z^* \mid \phi) \times p(z \mid z^*, \phi) \\
 &\times p(Y_1^* \mid z, z^*, \phi) \times p(Y_1 \mid Y_1^*, z, z^*, \phi) \\
 &\times p(Y_2^* \mid Y_1, Y_1^*, z, z^*, \phi) \\
 &\times p(Y_2 \mid Y_2^*, Y_1, Y_1^*, z, z^*, \phi)
 \end{aligned} \tag{16}$$

for nursing home care, Y_{nh}^* is used instead Y_2^* .

The posterior density is proportional to the product of the prior given in (13), (14), and (15), and the joint density is given in (16). Since there are no analytical solutions for the joint density, the Gibbs sampler (a MCMC procedure) is used (Koop, 2003). This entails making sequential draws from each of the conditional densities. These draws include draws for the latent variables as well. The data augmentation approach makes the Bayesian MCMC method especially convenient for draws of latent variables (Tanner and Wong, 1987).

For count data such as hospital inpatient care and physician visits, the Bayesian algorithm additionally involves the Metropolis-Hastings algorithm (this is also a MCMC procedure) for Y_2^* since it is hard to take draws from the conditional posterior density. Instead, draws will be taken from a candidate density and they are accepted with an acceptance ratio. The MCMC procedure is presented in Appendix B in detail.

Average treatment effects

The goal of this study is to estimate the effect of informal care on formal care utilization.

Treatment is defined as the receipt of informal care ($z > 0$). The observed outcome for formal care utilization can be written as

$$y = dy_1 + (1 - d)y_0 \tag{17}$$

$$d = 1 \text{ if } z > 0 \text{ and } d = 0 \text{ if } z = 0$$

where y_1 is the outcome for the treated and y_0 is the counterfactual (the outcome if the elderly person were not treated), and y corresponds to Y_2 in Section III.

For a randomly chosen individual, the expected difference in the outcome between the treated and the non-treated is the average treatment effect (ATE)

$$E(y_1 - y_0 \mid \Delta) = E(y_1 \mid \Delta) - E(y_0 \mid \Delta) \tag{18}$$

where $\Delta = (\beta_1, \beta_2, \tau_1, \tau_2, \alpha, \pi_1, \pi_2, \sigma^2, z^*, z_0, z_1)$. The expected difference conditional on being treated is the average treatment effect on the treated (ATET).

$$E(y_1 - y_0 \mid \Delta, d = 1) = E(y_1 \mid \Delta, d = 1) - E(y_0 \mid \Delta, d = 1) \tag{19}$$

For the ATE, the conditional expected outcomes and the counterfactual for the Poisson distribution can be written as, respectively,

$$E(y_1 \mid \Delta) = \Phi(x\beta_1 + z_1\tau_1 + (z^* - w\alpha)\pi_1) \times \frac{\exp(x\beta_2 + z_1\tau_2 + (z^* - w\alpha)\pi_2 + 0.5\sigma^2)}{1 - \exp(-\exp(x\beta_2 + z_1\tau_2 + (z^* - w\alpha)\pi_2))}, \quad (20)$$

$$E(y_0 \mid \Delta) = \Phi(x\beta_1 + z_0\tau_1 + (z^* - w\alpha)\pi_1) \times \frac{\exp(x\beta_2 + z_0\tau_2 + (z^* - w\alpha)\pi_2 + 0.5\sigma^2)}{1 - \exp(-\exp(x\beta_2 + z_0\tau_2 + (z^* - w\alpha)\pi_2))}. \quad (21)$$

For log-normal distribution (for nursing homes), the ATE can be written as

$$E(y_1 \mid \Delta) = \Phi(x\beta_1 + z_1\tau_1 + (z^* - w\alpha)\pi_1) \times \exp(x\beta_2 + z_1\tau_2 + (z^* - w\alpha)\pi_2 + 0.5\sigma^2), \quad (22)$$

$$E(y_0 \mid \Delta) = \Phi(x\beta_1 + z_0\tau_1 + (z^* - w\alpha)\pi_1) \times \exp(x\beta_2 + z_0\tau_2 + (z^* - w\alpha)\pi_2 + 0.5\sigma^2). \quad (23)$$

The ATE are obtained at both the extensive and intensive margins, in particular for $(z_0, z_1) = (0, 3.5), (0, 5), (3.5, 5)$. Those values correspond to an imposed increase in hours of informal care from zero to median level care, zero to full-time care, and median level to full-time care, respectively. The last step is to integrate out Δ by averaging over the sample from the posterior distribution. That is,

$$\overline{ATE} = \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{S} \sum_{s=1}^S \hat{E}(y_{1i} - y_{0i} \mid \Delta_i) \right) \quad (24)$$

where N is the sample size, $s = 1, \dots, S$ are the posterior draws after burn-in, and $\hat{E}(y_{1i} - y_{0i} \mid \Delta_i)$ is an estimate of $E(y_{1i} - y_{0i} \mid \Delta_i)$.

Testing endogeneity

The endogeneity of informal care is tested by comparing the likelihood of the restricted model (M_1), where $(\pi_1, \pi_2) = (0, 0)$, with the likelihood of the unrestricted model (M_2) that allows the correlation parameters to vary. The Bayesian method does not yield numerical values for these likelihoods, but a simplified method referred to as the Savage-Dickey density ratio can be applied when the two models are nested (Koop, 2003). This ratio avoids calculating the marginal likelihood directly. It does require that the priors for the restricted

and the unrestricted model to be the same at the point $(\pi_1, \pi_2) = (0, 0)$, an assumption that is reasonable in this case.

Using the Savage-Dickey density ratio, the Bayes factor can be written as the ratio of the posterior and the prior evaluated at $(\pi_1, \pi_2) = (0, 0)$ (Chib, Greenberg and Winkelmann, 1998).

$$BF_{1,2} = \frac{p(\pi_1 = 0, \pi_2 = 0 \mid data, M_2)}{p(\pi_1 = 0, \pi_2 = 0 \mid M_2)} \quad (25)$$

The Savage-Dickey density ratio only involves the unrestricted model (M_2) and the calculation of the prior and posterior densities. The prior density can be evaluated directly, whereas the posterior density can be calculated using the output from the Gibbs sampler or the Metropolis Hasting procedure because the posterior density depends on all parameters in the model. Because the less informative priors tend to favor the null hypothesis that $(\pi_1, \pi_2) = (0, 0)$, informative priors are used. $BF_{1,2}$ larger than 1 indicates no endogeneity, and its being less than 1 indicates endogeneity.

Appendix B: Bayesian markov chain monte carlo procedure

The MCMC procedure consists of the following steps.

steps (1)-(3): treat the latent variables z^, Y_1^*, Y_2^* as additional parameters in the model and data-augmenting them by taking draws from the appropriate full conditional density.*

steps (4)-(6): take draws from the appropriate full conditional density for all other parameters $(\beta_1, \beta_2, \tau_1, \tau_2, \alpha, \pi_1, \pi_2)$.

step (7): take draws from the full conditional density for σ^2 and σ_u^2 .

These steps are presented in the following order.

(1) When $z = 0$, the latent variable z^* can be drawn from the truncated normal distribution $N(\bar{z}, \bar{v})$, where

$$\bar{v} = (1 + \pi_1^2 + \pi_2^2 \sigma^{-2})^{-1} \quad (26)$$

and

$$\bar{z} = w\alpha + \bar{v}[\pi_1(Y_1^* - x_1\beta_1 - z\tau_1) + \pi_2\sigma^{-2}(Y_2^* - x_2\beta_2 - z\tau_2)] \quad (27)$$

Truncation is below zero. When $z > 0$, we have $z^* = z$.

(2) The latent variable Y_1^* can be drawn from a truncated normal distribution $N[x\beta_1 + z\tau_1 + (z^* - w\alpha)\pi_1, 1]$. The distribution is truncated below zero if $Y_1 = 1$.

(3) For count outcomes (hospital inpatient care and physician visits), μ is drawn from a normal distribution with a mean of $x\beta_2 + z\tau_2 + (z^* - w\alpha)\pi_2$ and a variance of σ^2 when $Y_1 = 0$. When $Y_1 = 1$, the full conditional density for μ is proportional to

$$p(\mu \mid \phi) = \frac{\exp(-\exp(\mu)) \exp(\mu Y_2)}{Y_2!} \times \frac{1}{1 - \exp(-\exp(\mu))} \times \exp\left(-\frac{1}{2}\sigma^{-2}(\mu - x\beta_2 - z\tau_2 - (z^* - w\alpha)\pi_2)^2\right). \quad (28)$$

The Metropolis-Hastings Method makes candidate draw μ^* from a proposal distribution $q(\mu^*)$ (Chib and Greenberg, 1995). This draw is accepted with a probability of

$$\alpha(\mu, \mu^*) = \min\left\{\frac{p(\mu^* \mid \phi, z^*)q(\mu)}{p(\mu \mid \phi, z^*)q(\mu^*)}, 1\right\} \quad (29)$$

where μ is the current draw and μ^* is the candidate draw.

For continuous variable (nursing home care), Y_{nh}^* is instead drawn from a normal distribution with a mean of $x\beta_2 + z\tau_2 + (z^* - w\alpha)\pi_2$ and a variance of σ^2 when $Y_1 = 0$. When $Y_1 = 1$, we have $Y_{nh}^* = \exp(Y_{nh})$.

(4) Given that the prior of α is $N(\underline{\alpha}, \underline{v}_\alpha)$, the posterior distribution of α is $N(\bar{\alpha}, \bar{v}_\alpha)$, where

$$\bar{v}_\alpha = (\underline{v}_\alpha + w'(1 + \pi_1^2 + \pi_2^2\sigma^{-2})w)^{-1} \quad (30)$$

$$\bar{\alpha} = \bar{v}_\alpha \left[\frac{\underline{\alpha}\underline{v}_\alpha + w'(1 + \pi_1^2 + \pi_2^2\sigma^{-2})z^* - w'\pi_1(Y_1^* - x_1\beta_1 - z\tau_1)}{-w'\pi_2\sigma^{-2}(Y_2^* - x_2\beta_2 - z\tau_2)} \right]. \quad (31)$$

(5) I specify the prior distribution for (β_1, τ_1) to be $N(\underline{\eta}_1, \underline{v}_{\eta_1}^{-1})$ and the prior distribution for π_1 to be $N(\underline{\pi}_1, \underline{v}_{\pi_1}^{-1})$. Define $\lambda_1 = (\underline{\beta}_1, \tau_1, \pi_1)$, $C = (x, z, (z^* - w\alpha))$, then the conditional distribution of λ_1 is $N(\bar{\lambda}_1, \bar{v}_{\lambda_1}^{-1})$ where

$$\bar{v}_{\lambda_1} = \begin{bmatrix} \underline{v}_{\eta_1} & 0 \\ 0 & \underline{v}_{\pi_1} \end{bmatrix} + C'C \quad (32)$$

and

$$\bar{\lambda}_1 = \bar{v}_{\lambda_1}^{-1} \left[\begin{pmatrix} \underline{\eta}_1 \underline{v}_{\eta_1} \\ \underline{\pi}_1 \underline{v}_{\pi_1} \end{pmatrix} + C'Y_1^* \right] \quad (33)$$

(6) Similarly to step (5), I specify the prior for (β_2, τ_2) to be $N(\underline{\eta}_2, \underline{v}_{\eta_2}^{-1})$ and the prior distribution for π_2 to be $N(\underline{\pi}_2, \underline{v}_{\pi_2}^{-1})$. Define $\lambda_2 = (\underline{\beta}_2, \tau_2, \pi_2)$, then the conditional distribution of λ_2 is $N(\bar{\lambda}_2, \bar{v}_{\lambda_2}^{-1})$ where

$$\bar{v}_{\lambda_2} = \begin{bmatrix} \underline{v}_{\eta_2} & 0 \\ 0 & \underline{v}_{\pi_2} \end{bmatrix} + C'\sigma^{-2}C \quad (34)$$

and

$$\bar{\lambda}_2 = \overline{v_{\lambda_2}}^{-1} \left[\left(\frac{\eta_2 v_{\eta_2}}{\pi_2 v_{\pi_2}} \right) + C' \sigma^{-2} Y_2^* \right] \quad (35)$$

(7) Finally, let the prior of σ^2 and σ_u^2 be $invG(c_0/2, (d_0/2)^{-1})$. The posterior conditional distribution can be written as,

$$G \left(\begin{array}{c} \frac{c_0 + N}{2}, \\ \left[\frac{d_0}{2} + \frac{(Y_2^* - x\beta_2 - z\tau_2 - (z^* - w\alpha)\pi_2)'(Y_2^* - x\beta_2 - z\tau_2 - (z^* - w\alpha)\pi_2)}{2} \right]^{-1} \end{array} \right) \quad (36)$$

(8) Repeat the above steps for a burn-in of 5,000 draws that are discarded, and then for another $S = 10,000$ draws that are kept.

In summary, the Gibbs sampler procedure entails draws from the normal, the truncated normal, and the gamma distributions. If the health outcome is a count, the Metropolis-Hastings algorithm is additionally applied as shown in step (3).

References

- [1] Andersen, R. 1968. Behavioral model of families' use of health services. Research Series No. 25. Chicago, IL: Center for Health Administration Studies, University of Chicago.
- [2] Angrist, J. D. 2001. 'Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice', *Journal of Business & Economic Statistics*, Vol. 19, pp. 2-16.
- [3] Albert, J. H., Chib S. 1993. 'Bayesian Analysis of Binary and Polychotomous Response Data', *Journal of the American Statistical Association*, Vol. 88, pp. 669-679.
- [4] Bolin, K., Lindgren, B., and Lundborg, P. 2007. 'Informal and Formal Care among Single-living Elderly in Europe', *Health Economics*, Vol. 17, pp. 393-409.
- [5] Bonsang, E. 2009. 'Does informal care from children to their elderly parents substitute for formal care in Europe?', *Journal of Health Economics*, Vol. 28, pp. 143-154.
- [6] Brown, J. R. and Finkelstein, A. 2011. 'Insuring long-term care in the United States', *Journal of Economic Perspectives*, Vol. 25, pp. 119-142.

- [7] Brown, J. R. and Finkelstein, A. 2009. 'The private market for long-term care insurance in the U.S.: A review of the evidence', *Journal of Risk and Insurance*, Vol. 76, pp. 5-29.
- [8] Cameron, C. and Trivedi, P. K. 1998. *Regression Analysis of Count Data*. Cambridge University.
- [9] Charles, K. K. and Sevak, P. 2005. 'Can Family Caregiving Substitute for Nursing Home Care?', *Journal of Health Economics*, Vol. 24, pp. 1174-1190.
- [10] Checkovich, J. J. and Stern, S. 2002. 'Shared Caregiving Responsibilities of Adult Siblings with Elderly Parents', *Journal of Human Resources*, Vol. 37, pp. 441-478.
- [11] Center for Medicare and Medicaid Services National Health Expenditure Data, available at https://www.cms.gov/NationalHealthExpendData/03_NationalHealthAccountsProjected.asp.
- [12] Chib, S. and Greenberg, E. 1995. 'Understanding the Metropolis-Hastings Algorithm', *The American Statistician*, Vol. 49, pp. 327-335.
- [13] Chib, S. 1992. 'Bayesian Inference in the Tobit Censored Regression Model', *Journal of Econometrics*, Vol. 51, pp. 79-99.
- [14] Chib, S., Greenberg, E., and Winkelmann, R. 1998. 'Posterior Simulation and Bayes Factor in panel Count Data Models', *Journal of Econometrics*, Vol. 86, pp. 33-54.
- [15] Chib, S., and Hamilton, B. H. 2000. 'Bayesian Analysis of Cross-section and Clustered Data Treatment Models', *Journal of Econometrics*, Vol. 97, pp. 25-50.
- [16] Christianson, J. B. 1988. 'The evaluation of the national long-term care demonstration: the effect of channeling on informal caregiving', *Health Services Research*, Vol. 23, pp. 99-117.
- [17] Congressional Budget Office (CBO), 2004. 'Financing long-term care for the elderly', available at <http://www.cbo.gov/ftpdocs/54xx/doc5400/04-26-LongTermCare.pdf>.
- [18] Deb, P., Munkin, M. K., and Trivedi, P. K. 2006a. 'Bayesian Analysis of the Two-Part Model With Endogeneity: Application to Health Care Expenditure'. *Journal of Applied Econometrics*, Vol. 21, pp. 1081-1099.

- [19] Deb, P., Munkin, M. K., and Trivedi, P. K. 2006b. 'Private Insurance, Selection, and Health Care Use: A Bayesian Analysis of a Roy-Type Model', *Journal of Business & Economic Statistics*, Vol. 24, pp. 403-415.
- [20] Duan, N., Manning, W. G., Morris, C. N., Newhouse, J. P. 1984. 'Choosing between the Sample-selection and Multi-part model', *Journal of Business and Economic Statistics*, Vol. 2, pp. 283-289.
- [21] Feder, J., Komisar, H. L., and Niefeld, M. 2000. 'Long-term Care in the United States: An Overview', *Health Affairs*, Vol. 19, pp. 40-56.
- [22] Freedman, V. (1996). 'Family Structure and the Risk of Nursing Home Admission', *The Gerontologist*, Vol. 51B, pp. S61-S69.
- [23] Geweke, J., Gowrisankaran, G., and Town, R. J. 2003. 'Bayesian Inference for Hospital Quality in a Selection Model', *Econometrica*, Vol. 71, pp. 1215-1238.
- [24] Hanley, R. J., Alecxih, L. M. B., Wiener, J. M., and Kennell, D. L. 1990. 'Predicting Elderly Nursing home Admissions: Results from the 1982-1984 National Long-Term Care Survey', *Research on Aging*, Vol. 12, pp. 199-228.
- [25] Jette, A. M., Tennstedt, S., and Crawford, S. 1995. 'How does Formal and Informal Community Care Affect Nursing Home Use?', *Journals of Gerontology*, Vol. 50, pp. S4-S12.
- [26] Johri, M., Beland, F., and Bergman, H. 2003. 'International experiments in integrated care for the elderly: a synthesis of the evidence', *Journal of Geriatric Psychiatry*, Vol. 18, pp. 222-235.
- [27] Kane, R., Keckhafer, G., and Robst, J. 2002. 'Evaluation of the Evercare demonstration program: final report', available at http://www.cms.gov/DemoProjectsEvalRpts/downloads/Evercare_Final_Report.pdf.
- [28] Kemper, P. 1992. 'The use of formal and informal home care by the disabled elderly', *Health Services Research*, Vol. 27, pp. 421-451.
- [29] Kemper, P., Komisar, H.L. and Alecxih, L. 2006. 'Long-term care over an uncertain future: what can current retirees expect?' *Inquiry*, Vol. 42, pp. 335-350.
- [30] Kodner, D., and Kyriacou, C. 2000. 'Fully integrated care for the frail elderly: two American models', *International Journal of Integrated Care*, Vol. 1, pp. 1-24.

- [31] Koop, G. 2003. *Bayesian Econometrics*. John Wiley & Sons, Ltd publishing, England.
- [32] Koop, G., and Poirer, D. J. 1997. 'Learning about the Across-Regime Correlation in Switching Regression Models', *Journal of Econometrics*, Vol. 78, pp. 217-227.
- [33] Langa, K. M., Chernew, E., Kabeto, M. U., and Katz, S. J. 2002. 'The Explosion in Paid Home Care in the 1990s: Who received the Additional Services?', *Medical Care*, Vol. 39, pp. 147-157.
- [34] Lo Sasso, A. T. and Johnson, R. W. 2002. 'Does Informal Care From Adult Children Reduce Nursing Home Admissions for the Elderly?', *Inquiry*, Vol. 39, pp. 279-297.
- [35] MetLife Mature Market Institute, 2009. 'The 2009 MetLife market survey of nursing home, assisted living, adult day services, and home care costs', October, available at <http://www.metlife.com/assets/cao/mmi/publications/studies/mmi-market-survey-nursing-home-assisted-living.pdf>.
- [36] National Health Expenditures Accounts: Definitions, Sources, and Methods, 2007. Center for Medicare and Medicaid Services.
- [37] Newman, S. J. and Struyk, R. 1990. 'Overwhelming Odds: Caregiving and the Risk of Institutionalization', *Journals of Gerontology*, Vol. 45, pp. S173-S183.
- [38] OECD, 2011. 'Help wanted? Providing and paying for long-term care', available at www.oecd.org/dataoecd/30/24/47836116.pdf.
- [39] Spillman, B. C., and Pezzin, L. E. 2000. 'Potential and Active Family Caregivers: Changing Networks and the "Sandwich Generation"', *Milbank Quarterly*, Vol. 78, pp. 339-347.
- [40] Stern, S. 1995. 'Estimating family long-term care decisions in the presence of endogenous child characteristics', *Journal of Human Resources*, Vol. 30, pp. 551-580.
- [41] Rivers, D. and Vuong, Q. H. 1988. 'Limited information estimators and exogeneity tests for simultaneous Probit models', *Journal of Econometrics*, Vol. 39, pp. 347-366.
- [42] Tanner, M. A. and Wong, W. 1987. 'The calculation of posterior distributions by data augmentation (with discussion)', *Journal of the American Statistical Association*, Vol. 82, pp. 528-550.

- [43] Van Houten, C. H. and Norton, E. C. 2004. 'Informal Care and Health Care Use of Older Adults', *Journal of Health Economics*, Vol. 23, pp. 1159-1180.
- [44] Van Houten, C. H. and Norton, E. C. 2008. 'Informal Care and Medicare Expenditure: Testing for Heterogeneous Treatment Effects', *Journal of Health Economics*, Vol. 27, pp. 134-156.
- [45] Weissert, G. W. 1988. 'The national channeling demonstration: what we knew, know now, and still need to know', *Health Services Research*, Vol. 23, pp. 175-187.