PARISH, WILLIAM JAMES, Ph.D. Three Essays in Health Econometrics and Health Economics. (2015)
Directed by Dr. Christopher Swann. 124 pp.

This dissertation makes contributions to the field of health econometrics as well as the field of health economics. Chapter II presents a review of the "Cake Debates" literature. I also describe a generic econometric framework, which I use to clarify some points of controversy in the literature. Extant approaches to endogeneity in two-part models have failed to consider "separability." Chapter III presents a novel estimator that accounts for endogeneity without imposing "separability." Chapter IV presents a mediation analysis that explores whether improvements in beneficiary health can explain the negative relationship between Medicare Part D and the use of hospital care.

THREE ESSAYS IN HEALTH ECONOMETRICS AND HEALTH ECONOMICS

by

William James Parish

A Dissertation Submitted to
the Faculty of the Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro
2015

Approved by

_____
Committee Chair

To my loving wife

APPROVAL PAGE

This dissertation written by William James Parish has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair_____

Committee Members_____

_____

_____

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

ACKNOWLEDGMENTS

The research presented here benefited greatly from mentorship received from Drs. Joseph Terza, Christopher Swann, and Jeremy Bray.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

This dissertation is entitled "Three Essays in Health Econometrics and Health Economics." This reflects the notion that there are two broad focuses of the research contained herein. In particular, chapters II and III provide econometric contributions that are relevant in the estimation of health care expenditure models. Chapter IV is an applied contribution that investigates the effect of Medicare Part D, Medicare's prescription drug program, on the use of hospital services.

The second chapter provides an extensive review of the literature that has debated the relative merits of the two-part model, popularized by the RAND Health Insurance Experiment researchers, and Heckman's classic sample selection model. Upon completion of this extensive review a general econometric model is described, which we demonstrate nests both a conventional two-part model and a conventional sample-selection model. This serves to clarify some of the more nuanced points of controversy in the reviewed literature.

The third chapter further explores the two-part model, but addresses the issue of endogeneity on the right-hand side of the model. In this chapter it is demonstrated that extant approaches to endogeneity have failed to consider the implications of endogeneity on the "separability" of the two parts of the two-part model. This chapter also presents a

novel estimator that can estimate two-part model parameters without imposing "separability" and in the presence of an endogenous dummy regressor.

The fourth chapter analyzes the role that Medicare Part D plays in the health of Medicare beneficiaries and the effect that this relatively new program has had on the use of hospital services. In particular, this chapter presents a mediation analysis that explores whether improvements in beneficiary health can explain the negative relationship that has previously been documented in the literature between Medicare Part D and the use of, or expenditures on, hospital care.

The fifth chapter concludes.

CHAPTER II

TWO-PART MODELS VERSUS SAMPLE-SELECTION MODELS

The two-part model (Cragg, 1971) remains a popular method for estimating patient-level healthcare expenditures (Bell, Zimmerman, Arterburn, & Maciejewski, 2011; Finkelstein, Trogdon, Cohen, & Dietz, 2009; Trogdon, Finkelstein, Feagan, & Cohen, 2012). The popularity of the two-part model is likely due to its use in the Rand Health Insurance Study (Manning, Newhouse, Duan, Keeler, & Leibowitz, 1987). Applied researchers frequently test whether the two-part model is appropriate by applying Heckman's (1976; 1978) sample-selection model to determine if there exists correlation between the two parts of the model (Andreyeva, Sturm, & Ringel, 2004; Sullivan, Ghushchyan, Wyatt, & Hill, 2007). This practice is illegitimate and results from a misunderstanding of the "cake debate" literature, which debated the relative strengths and weaknesses of these two approaches. The objective of this chapter is to provide some technical resolutions to the "cake debates." This objective is accomplished by developing an encompassing theoretical framework that hierarchically organizes a comprehensive set of health care expenditure models, including double-hurdle models and minimally parametric versions of both the two-part and sample-selection models.

Two recent treatments of the two-part and sample-selection models have proposed a unifying classification system that in principle accomplishes these objectives. Madden (2008) presents a classification system based on the bivariate normal double-hurdle model (see also Jones, 1989; Jones & Yen 2000). Wooldridge (2010) considers normal versions of both models. Both of these analyses rely on restrictive primitive assumptions that are stronger than necessary for consistent estimation. Furthermore, both analyses overlook some of the more nuanced points raised in the literature. Therefore they fail to provide a comprehensive understanding of these two methodological approaches. A more *complete* analysis can be accomplished with weaker assumptions imposed from the outset. In particular, a more complete analysis should formally synthesize the extant knowledge in a way that does not implicitly ignore more generic cases.

A Recent Debate

A recent exchange (Salas & Raferty, 2001; Seshamini & Gray, 2004; Zweifel, Felder, & Meiers, 1999; 2001) provides an application-oriented presentation of what is currently known about two-part and sample-selection models. This exchange illustrates the significant confusion that still exists in applied health economics regarding these two models. Before reviewing the early "cake debate" literature I consider this more recent exchange.

Health economists have for some time expressed concern over the explosion of healthcare costs experienced in many developed countries. It has been proposed that a large proportion of the increase in healthcare expenditures may be due to the changing demographic composition of developed economies. In particular, it is noted that OECD

countries have experienced a radical increase in the proportion, and number, of individuals aged 65 and older. Since there exists a strong association between age and likelihood of severe and chronic illnesses, it is plausible that a more elderly population will spend more on healthcare. Zweifel et al. (1999) offer an alternative explanation. Specifically, they test the hypothesis that proximity to death is a more important causal factor in the increase in health expenditures. Under this hypothesis, a more elderly population will have no effect on healthcare expenditures unless that elderly population also has a higher proportion of individuals close to death.

Zweifel et al. (1999) test their hypothesis using panel data from two Swiss sickness funds, one representative of a more rural population and the other representative of a more urban population. They use Heckman's sample-selection model to estimate age and quarter-to-death effects on hospital expenditures. Zweifel et al. (1999) argue that a sample-selection correction is necessary because the strictly positive expenditure subsample represents a "sick" sample. Their test of the proximity-to-death hypothesis amounts to testing the significance of the coefficients on age and the coefficients on quarter-to-death in the second-stage equation. They find that quarter-to-death has a significant and increasing effect on health expenditures, but that age has an insignificant effect. This finding confirms their hypothesis.

Salas and Raftery (2001) critique Zweifel et al.'s (1999) implementation of the sample-selection model. In particular, they note the lack of exclusion restrictions in their specification. They argue that without exclusion restrictions the inverse Mills ratio is approximately a linear function of the same covariates used in the linear index portion of

the second-stage equation. Assuming that Zweifel et al. (1999) included only age and age squared in their first-stage probit equation, the inverse Mills ratio and age will be highly collinear in the second-stage equation.[1] Therefore, the lack of significance of the age coefficient might be explained by a standard error problem resulting from multi-collinearity. Zweifel et al. (2001) note that they included much more than age and age squared in their first-stage probit specification, and thus a smaller degree of multi-collinearity is implied.

Seshamani and Gray (2004) replicate the study conducted by Zweifel et al. (1999), and explore further the issue of multi-collinearity raised by Salas and Raftery (2001). They use data from the Oxford Record Linkage study (ORLS), a data set similar to the Swiss data used by Zweifel et al. In a comparable specification to Zweifel et al.'s (1999) specification, they find significant evidence of multi-collinearity between the inverse Mills ratio and the age variable. Furthermore, they find quantitatively and qualitatively similar results regarding the effect of age and proximity to death. However, in a two-part specification they find a significant age effect. They conclude that the distinction between the two-part and sample-selection models is practically important, because the two methods can radically differ.

In addition to revisiting the multi-collinearity issue, Seshamani and Gray (2004) also argue that the sample-selection model is inappropriate on conceptual grounds. In particular, they claim that "a sample selection model is really only necessary when the selection is unobserved… This is not the case with the [Zweifel et al., 1999] and ORLS

[1] Zweifel et al. (1999) do not present results from their first-stage probit, but discuss the results on age and age squared in the first-stage.

data sets, where all zero cost observations were observed" (p. 306).[2] In the case of "observed selection," the sample-selection model estimates parameters consistent with the concept of *potential* health expenditures; that is, the expenditures that non-hospitalized individuals would have had if they had been hospitalized. Seshamani and Gray (2004) propose that potential expenditures do not inform actual budgetary considerations.

The "Cake Debates"

The review of the "Cake Debates" provided here is organized around the following four points:

1. Two-part models estimate *actual* outcomes, while sample-selection models estimate *potential* outcomes;
2. The two-part model is more flexible than Tobit and sample-selection models;
3. The two-part model is not technically nested in the sample-selection model; and
4. Even when the sample-selection model is the conceptually "correct" approach the two-part model may have better statistical properties.

The first point motivates the need for a comprehensive understanding of two-part and sample selection models. Specifically, these methodologies are categorically distinct, and the distinctions have implications with regard to the interpretation of the obtained results. The second and third points indicate a significant gap in our knowledge since consensus remains illusive regarding these points in the literature. Finally, the fourth point implies that there are important statistical considerations that applied researchers should be aware of regardless of conceptual considerations.

[2] The statement "selection is unobserved" doesn't really make sense here. It is likely that the authors' mean that selection is based on unobservables.

The distinction between an actual versus a potential outcome is a common distinction made in the literature regarding the two-part and sample-selection models. It provides a good way of categorizing the theoretical distinctions between the two competing methods. It also represents an important distinction that must be understood by the applied researcher if the researcher wants to impart the correct interpretation on the obtained results.

Duan, Manning, Morris, and Newhouse (1983) compared several possible estimators for estimating health care expenditure equations. They rejected the sample-selection model from this comparison on the basis of theoretical considerations alone. In particular, they observed that the sample-selection model estimates an "unconditional" expenditure equation. They argued that this is an inappropriate estimation objective because they were interested in modeling "conditional" expenditures. In other words, zero expenditure is an *actual* observation, not a conveniently coded missing value for some unknown *potential* observation.

Hay and Olsen (1984) argued that potential outcomes *are* interesting in the context of health expenditures. In particular, they proposed that one can consider the dichotomous decision to seek medical care or not as a "latent propensity to access the health care system" (p. 281). Under this interpretation, one may be interested in the potential expenditures of individuals who failed to have a large enough "propensity to access the health care system." Furthermore, they argued that if one is interested in potential outcomes one should not ignore information obtained from the first-part

dichotomous participation decision equation in the analysis of "unconditional" expenditures.

Maddala (1985a) noted that one might conceptualize all individuals as needing some sort of medical intervention. Individuals only decide, however, to seek medical care when their perception of need for medical intervention becomes sufficiently high. Even when the perception of need is low, there may be known or unknown health issues that would benefit from medical treatment. Unmet medical need translates conceptually into what Duan et al. (1983) and Duan, Manning, Morris, and Newhouse (1984) call "potential" expenditures.

Dow and Norton (2003) attempted to clarify the distinction between actual and potential outcomes by classifying an actual outcome as a "corner solution outcome" (Wooldridge, 2010 also uses this language). They also put forth the argument that potential outcomes are uninformative in terms of actual private and/or public budgetary considerations.[3] Finally, they claimed that if the outcome is fully observed (i.e., the outcome is an actual outcome), then there is no need to correct for selection bias. This latter point is of particular relevance to the choice between a two-part and sample-selection model.

It is worth bearing in mind that the two-part model was initially researched as a flexible extension to the Tobit model (Tobin, 1958). It is well known that two restrictive features hamper the Tobit model (Wooldridge, 2010; Greene, 2012). First, the effects of

[3] It is important to make the distinction here between a potential expenditure that would occur generally if non-spenders became spenders, and a potential expenditure that might occur given a change in an exogenous regressor. These are distinct concepts, and treatment of the two concepts should be treated distinctly.

a regressor on the probability that the outcome is positive and on the conditional mean of the outcome must have the same sign since a single coefficient vector is estimated. This may be particularly restrictive in a variety of contexts. For instance, Wooldridge (2010) considered an example of estimating the extensiveness of life insurance coverage. It is reasonable to assume that individuals have an increasing probability of obtaining life insurance policies as they age, but that the extent of coverage will decline with age. This assumption cannot be incorporated in the Tobit estimation framework. Second, the proportional effect of two continuous regressors on both the probability that the outcome is positive and on the conditional mean of the outcome must necessarily be equal. Again, this is because there is only one estimated coefficient vector. Since the two-part model estimates two sets of parameters it is a more flexible approach than the Tobit model.

Furthermore, Pohlmeier and Ulrich (1995) argued that separating the decision to seek medical treatment from the decision regarding how much medical treatment to consume is natural. From a theoretical perspective they propose that the decision to seek medical treatment is most naturally governed by the theory developed by Grossman (1972), wherein individuals choose to consume medical services to increase their "health stock." They argue that Grossman's theory is applicable here because individuals choose whether or not to see a physician on their own, and Grossman's model is an individual-level decision-making model. On the other hand, the decision regarding how much medical treatment to consume is most naturally governed by principal-agent theory, wherein individuals do not make decisions per se but contract physicians to act as agents

on their behalf. Accordingly, Pohlmeier and Ulrich suggest that a two-part model is more theoretically congruent than a one-part model such as the Tobit model.

Duan et al. (1983; 1984) argued that the two-part model represents a more flexible specification than Heckman's sample-selection model. They pointed out that the sample-selection model relies on restrictive assumptions regarding the joint distribution of the error terms in the first and second equations. They reasoned that almost any parametric assumption could be used in the two-part model specification, whereas the sample-selection model relies on the bivariate normality assumption. This is a notion that is echoed in Angrist (2001). The implication of Angrist's argument suggests that the flexibility of the two-part model is more apparent than real, however.

Hay and Olsen (1984) argued that the two-part model is not the general model, but is a special case of the more general sample-selection model. Namely, they argue that the two-part model is essentially the same model, but it is assumed from the outset that there is independence between the first and second equation. On the face, it appears straightforward that if one begins with the sample-selection model's second-stage equation, then the two-part model obtains when zero correlation is imposed. Hay and Olsen (1984) made this argument with a bit more sophistication. In particular, they added "structure" to the two-part model by assuming the existence of a potential outcome, which is observed only when the hurdle component of the two-part model is exceeded. Of course, the two-part model makes no such assumption since the two-part model describes an actual outcome. Nonetheless, Hay and Olsen attempted to

demonstrate that with a latent level assumed at zero it must be the case that independence is assumed. They were unable to definitively demonstrate this, however.

Duan et al. (1984) made two arguments to the contrary. First, they noted that Duan et al. (1983) presents a "normal-theory" version of the two-part model. The normal-theory version of the two-part model assumes that the error terms in both the first- and second-parts of the model are normally distributed. Nothing is specified with regard to whether the error terms are jointly distributed, and if so, whether they are correlated. Importantly, they propose that the two-part model is consistent with a wide variety of alternative parametric assumptions regarding the second- as well as first-part error terms. In this sense, any two-part model that does not assume second-part normality is not strictly nested in Heckman's sample-selection model.

Duan et al. (1984) also debunked the independence argument that Hay and Olsen made. They accomplish this by constructing an example wherein two jointly distributed error terms with nonzero correlation can be transformed to obtain error term distributions that are consistent with the two-part model. The correlation between the two error terms plays no role in the likelihood function for that example, and therefore they claim that the two-part model avoids estimation of a "nuisance parameter." Maddala (1985a) noted that this example is "purely semantic," because the two-part model cannot identify the correlation between the two error terms anyway. He further commented that it might be an ill-advisable practice to construct models wherein the correlation cannot ever be estimated.

Maddala also made an informal argument that the two models are non-nested. He distinguished between the two methods by the way in which each method defines error terms. Specifically, he notes that the sample-selection model makes use of an error term in the second-stage that is defined over the entire population, while the two-part model makes use of an error term in the second-part that is defined only over the subpopulation of participants. This distinction has practical importance in the sense that the former error term is useful for modeling a "joint" decision process (i.e., a process wherein the two levels of decision-making are dependent on each other), while the latter error term is useful for modeling a "sequential" decision process (i.e., a process wherein the two levels of decision-making are independent of each other). Maddala concluded that the issue of nesting is a second-order concern since the two models are clearly different at a fundamental level; the first-order concern should be whether the concept of sequential decision-making is more appropriate than joint decision making.

A more recent literature (Jones, 1989; Jones & Yen, 2000; Madden, 2008) has attempted to provide clarification with regard to the classification, and nesting status, of the two-part and sample-selection models by nesting both models in the bivariate normal double-hurdle model with dependence. The double hurdle model specifies two latent variable processes: 1) the first-hurdle equation, and 2) the extent of participation equation. The error terms are assumed to be bivariate normally distributed in all of the works cited above. Any transformation of the observed dependent variable can be

applied.[4] The first-hurdle equation generates either zero or one values. The extent of participation equation generates zero outcomes whenever a Tobit-type corner solution occurs, or a continuous and positive outcome otherwise. The observed outcome is the product of the first-hurdle outcome, and the Tobit outcome that derives from the extent of participation equation.

Importantly, this literature demonstrates that first-hurdle dominance implies a single-hurdle model. First-hurdle dominance assumes that the dichotomous decision of zero versus nonzero is the only threshold that must be overcome in order for the observed continuous outcome to be equal to the latent outcome defined in the second-hurdle equation. In economic terms, Madden (2008) describes first-hurdle dominance as assuming that "zero consumption does not arise from a standard corner solution but instead represents a separate discrete choice" (p. 301). In the case of first-hurdle dominance, the second hurdle does not matter, but dependence is still a possibility. Since they assume dependence and bivariate normality, the single-hurdle model that obtains is the sample-selection model. The likelihood function for the normal-theory version of the two-part model is obtained if the correlation parameter in the sample-selection model is set equal to zero.

Another strand of the literature (Dow & Norton, 2003; Hay, Leu, & Rohrer, 1987; Leung & Yu, 1996; Manning, Duan, & Rogers, 1987; Norton, Dow, & Do, 2008; Puhan, 2000) has conducted a series of Monte Carlo studies regarding the relative performances of the two-part and sample-selection models under one of the following three data

[4] Jones and Yen (2000) consider the Box-Cox transformation since it relaxes the normality assumption.

generating assumptions: 1) the two-part model is correct, 2) the sample-selection model is correct, and 3) neither model is correct. The general consensus in this literature seems to be that even when the sample-selection model is the "correct" model, the two-part model is likely to outperform the sample-selection model in terms of parameter estimation and predictive power. The reason the literature has provided for this strange conclusion is that the two-part model excels in its small sample properties, whereas the sample-selection model suffers from rather important issues regarding "practical identification." In particular, the two-part model is biased whenever the sample-selection model correctly describes the data generating process, but the sample-selection model's use of the inverse Mills ratio correction frequently induces a severe multi-collinearity problem that causes incredible imprecision. Of course, valid exclusion restrictions can in principle solve the multi-collinearity problem. In the absence of exclusion restrictions, though, one can test between the methods by assuming that the sample-selection model is consistent (i.e., "correct"), and computing what Dow and Norton (2003) call an "empirical" mean-squared error. The mean squared error is unknowable in practice since one does not have access to the amount of bias induced by applying the two-part model. The "empirical" version of the mean squared error criterion uses the statistical distance between the estimates to approximate the bias assuming that the sample-selection estimates are true.

A General Model With Dependence

In this section I develop a fully general modeling framework with which I can organize the two-part and sample-selection models. With that in mind I use as few

distributional assumptions from the outset to avoid defining a hierarchical classification that is the result of the distributional class of models under consideration. Dependence between the first- and second-equations is a more general concept than correlation in a bivariate normal framework. Accordingly, from the outset I let the dependence between the two equations follow a much less restricted definition. I specify the first-equation as a linear-index probit/logit process, as this is typical in all the models discussed above and is not the point of contention between any of the models.

Let H^* denote the (first-)hurdle latent outcome, which is generated as follows:

$$H^* = X_p\beta_{p1} + X_{o1}\beta_{o1} + X_u, \tag{1}$$

where X_p denotes a *policy variable*, informally defined as the variable of interest, X_{o1} denotes the (first-)hurdle vector of observable confounders, X_u denotes an unobserved stochastic determinant and will elsewhere be referred to as an unobserved confounder to be distinguished from the observed confounders, and $\beta_1' = \left[\beta_{p1} \quad \beta_{o1}'\right]$ denotes the parameter vector. The conditional random variable $(X_u \mid X_p, X_{o1})$ must have a known, or specified, distribution. This is typically normal or logistic. In double-hurdle modeling, the first-hurdle defines whether or not the individual is a participant (e.g., in health expenditure applications, the first-hurdle determines whether the individual is in the market for healthcare at all). Let H denote the observed dichotomous participation decision. Equation (1) relates to the observed dichotomous participation decision as follows:

$$H = \begin{cases} 1 & \text{if } H^* > 0 \\ 0 & \text{if } H^* \leq 0. \end{cases} \tag{2}$$

Now let Y^* denote a latent extent of participation outcome. It is generically generated as follows:

$$Y^* = q\left(X_p, X_{o2}, X_u; \gamma_p, \gamma_o, \gamma_u\right) \times \kappa_2, \tag{3}$$

where q( ) denotes an arbitrary functional form, X_{o2} is a subset, though not necessarily strict, of X_{o1}, $\gamma' = \begin{bmatrix} \gamma_p & \gamma_o' & \gamma_u \end{bmatrix}$ denotes the parameter vector, and κ_2 denotes an error term. It must be minimally maintained that

$$E\left[\kappa_2 \mid X_p, X_{o1}, X_u, H\right] = 1. \tag{4}$$

Equation (4) is a minimal parametric assumption and could be replaced with a full distributional assumption regarding the conditional random variable $(\kappa_2 \mid X_p, X_{o1}, X_u, H)$. Inclusion of the (first-)hurdle error term in (3) imposes dependence between the (first-)hurdle and the extent of participation outcomes. A second-hurdle is said to exist if the latent outcome, Y^*, can be overridden by a corner solution outcome. Define a new latent variable that represents the utility-maximizing choice regarding extent of participation, and denote this variable by C^*. If one cannot assume first-hurdle dominance (Jones, 1989; Madden, 2008), then the latent extent of participation is given by:

$$C^* = \max\left(0, Y^*\right), \tag{5}$$

and the observed outcome, Y, is given by:

$$Y = H \times C^*. \tag{6}$$

The structure embodied in (5) suggests the use of a Tobit-type specification for the conditional on participation portion of the estimation problem. Equation (6) implies that there are two sources of the zero outcomes: non-participants have zero outcomes, and participants who choose a corner solution have zero outcomes. Equation (6) also implies that there is a potential outcome, since C^* could be positive when $H = 0$ resulting in a zero outcome for Y.

If first-hurdle dominance is assumed, then the double-hurdle model reduces to a single-hurdle model, and the set of models available reduces to either a two-part or sample-selection model. In this case one does not need to define a latent extent of participation variable, C^*, since the zero outcomes are all generated by the first-hurdle. The concept of a potential outcome is now solely captured by Y^*. If one wishes to model potential outcomes, then the observed outcome would be given by:

$$Y = H \times Y^*, \tag{7}$$

and it would become necessary to derive a sample-selection model to estimate parameters that relate X_p to the potential outcome Y^*. If instead one is interested in modeling an

actual outcome, then one replaces equation (3) with a conditional-on-positive data generating process where the outcome, denoted by Y^1, is a fully observed outcome (i.e., there are no latent levels generated). Y^1 is generically generated as follows:

$$Y^1 = q\left(X_p, X_{o2}, \left(X_u \mid X_p, X_{o1}, H_1 = 1\right); \gamma_p^a, \gamma_o^a, \gamma_u^a\right) \times \kappa_2^a. \tag{3a}$$

The parameter vector, $\gamma'^a = \left[\gamma_p^a \quad \gamma_o'^a \quad \gamma_u^a\right]$, clearly has a different interpretation in the actual outcome case versus the potential outcome case. Furthermore, the stochastic component in (3a), κ_2^a, must have a restricted support since the outcome Y^1 must be strictly positive. In this case (7) is no longer adequate for describing the observed outcome since (3a) is only defined whenever $H = 1$ and therefore an interaction between H and Y^1 is not mathematically sensible. The observed outcome must instead be given by:

$$Y = \begin{cases} 0 & \text{if and only if } H = 0 \\ Y^1 & \text{if and only if } H = 1. \end{cases} \tag{8}$$

We continue with the actual outcome version of the first-hurdle dominated model [defined by (1), (3a), and (8)]. One can write the probability density function for the conditional random variable $(X_u \mid X_p, X_{o1}, H)$ as follows (Terza, 1998; 2009):

$$f(X_u \mid X_p, X_{o1}, H) = H \times \left(\frac{g_{[-X_{po1}\beta_1 < X_u \leq \infty]}(X_u)}{1 - G(-X_{po1}\beta_1)} \right) + (1 - H) \times \left(\frac{g_{[-\infty < X_u \leq -X_{po1}\beta_1]}(X_u)}{G(-X_{po1}\beta_1)} \right), \tag{9}$$

where $X_{po1} = \left[X_p \quad X_{o1} \right]$, $g_{[I]}(\ )$ denotes the known (or specified) probability density function for $(X_u \mid X_p, X_{o1})$ evaluated over the interval I, and $G_{[I]}(\ )$ denotes the known (or specified) cumulative distribution function for $(X_u \mid X_p, X_{o1})$ likewise evaluated over the interval I. Using (9) we can integrate out the unobserved confounder in (3a), leaving a feasible conditional mean expression suitable for subsample regression-based estimation. In particular, we can derive a function (left unspecified for now), denoted by k( ), as follows:

$$
\begin{aligned}
k(X_p, X_{o1}; \beta_2) &= E\left[Y \mid X_p, X_{o1}, H = 1 \right] \\
&= \int_{-X_{po1}\beta_1}^{\infty} q(X_p, X_{o2}, (X_u \mid X_{po1}, H = 1); \gamma_2^a) f(X_u \mid X_{po1}, H = 1) d(X_u \mid X_{po1}, H = 1) \\
&= \frac{\int_{-X_{po1}\beta_1}^{\infty} q(X_{po1}, (X_u \mid X_{po1}, H = 1); \gamma_2^a) g(X_u \mid X_{po1}) d(X_u \mid X_{po1})}{1 - G(-X_{po1}\beta_1)},
\end{aligned}
\tag{10}
$$

where β_2 denotes a possibly (and appropriately) transformed version of the parameter vector $\left[\beta_1' \quad \gamma_2'^a \right]$. Once the calculation in (10) has been conducted, the actual outcomes model with first-hurdle dominance can be written in terms of k( ). In particular, one retains (1) and uses (10) to form the following regression-type equation:

$$
Y^1 = k(X_p, X_{o1}; \beta_2)\zeta_2, \tag{11}
$$

where it must be minimally maintained that $E\left[\zeta_2 \mid X_p, X_{o1} \right] = 1$.

The two-part model is a reduced-form approach to estimation in the sense that it does not specify a primitive behavioral model regarding the relationship between all confounders (observable and unobservable alike) and the conditional-on-positive outcome (i.e., the two-part model does not specify, or concern itself with, the q( ) function). Rather, the two-part model directly assumes the functional form of k( ). That is, the two-part model assumes the solution to the integration problem of (10) is in an easy-to-estimate equation that is separable from the first-part equation. Much of the "cake debate" literature has focused on whether or not the two-part model imposes independence between its two-parts. Here it is proposed that given an assumption regarding the functional form of k( ), one can recover a corresponding q( ) function with dependence from which one could derive, via the integration problem described in (10), the exact expression assumed by the chosen k( ) function. I consider two typical examples below.

The original two-part model (developed by Cragg, 1971 and researched extensively by Duan et al., 1983 and 1984) assumes, in our notation, that

$$k(X_p, X_{o1}, \beta_2) = \exp(X_p\beta_{p2} + X_{o1}\beta_{o2}). \tag{12}$$

Since the original version of the two-part model uses the logarithm of Y^1 the second-part equation in regression form is written as

$$\ln\left(Y^1\right) = X_p\beta_{p2} + X_{o1}\beta_{o2} + \ln(\zeta_2), \tag{13}$$

where it must be minimally maintained that $\mathrm{E}\left[\ln(\zeta_2) \mid \mathrm{X_p}, \mathrm{X_{o1}}\right] = 0$. Additionally, the original two-part model assumes that $(\mathrm{X_u} \mid \mathrm{X_p}, \mathrm{X_{o1}}) \sim \mathrm{N}(0,1)$. Under these assumptions, (12) can be obtained if one assumes that

$$\mathrm{q}(\mathrm{X_p}, \mathrm{X_{o1}}, \mathrm{X_u}; \gamma_2) = \frac{\Phi(\mathrm{X_{po1}}\beta_1)\exp(\mathrm{X_p}\beta_{p2} + \mathrm{X_{o1}}\beta_{o2} + \mathrm{X_u}\beta_u)}{\Phi(\mathrm{X_{po1}}\beta_1 + \beta_u)\exp(0.5\beta_u^2)}. \tag{14}$$

We show in Appendix A that the integration problem described in (10) using (14) does in fact result in the expression given in (12). Furthermore, it is clear that (14) does not impose independence between the hurdle and extent of participation decisions.

To explore the issue of nesting we consider the comparable log-linear version of the sample-selection model. In our notation, the log-linear version of the sample-selection model assumes that

$$\ln\left(\mathrm{q}(\mathrm{X_p}, \mathrm{X_{o1}}, \mathrm{X_u}; \gamma_2)\right) = \mathrm{X_p}\beta_{p2} + \mathrm{X_{o2}}\beta_{o2} + \mathrm{X_u}\beta_u. \tag{15}$$

In this version of the sample-selection model, the conditional mean of the logarithm of Y^1 is obtained rather than the conditional mean of Y^1. As with the log-linear version of the two-part model, the log-linear version of the sample-selection model assumes that $(\mathrm{X_u} \mid \mathrm{X_p}, \mathrm{X_{o1}}) \sim \mathrm{N}(0,1)$. Since (15) is additively separable in the observable components and the unobserved confounder, the integration problem in (10) can be written as

$$k(X_p, X_{o1}; \beta_2) = X_p\beta_{p2} + X_o\beta_{o2} + \beta_u \frac{\int_{-X_{po1}\beta_1}^{\infty} X_u \varphi(X_u) dX_u}{\Phi(X_{po1}\beta_1)}$$

$$= X_p\beta_{p2} + X_{o2}\beta_{o2} + \beta_u \frac{\varphi(X_{po1}\beta_1)}{\Phi(X_{po1}\beta_1)}. \quad (16)$$

Equation (16) is equivalent to Heckman's approach, though its derivation does not require bivariate normality.

A branch of the two-part model literature not discussed above concerns itself with the problem of retransforming results obtained from a log- or otherwise transformed outcome model such as the log-linear models discussed above (Ai & Norton, 2000; Ai & Norton, 2008; Abrevaya, 2002; Basu & Manning, 2009; Blough, Madden, & Hornbrook, 1999; Don, 1984; Duan, 1983; Manning, 1998; Manning, 2006; Manning & Mullahy, 2001; Mullahy, 2001; Taylor, 1986; Zhou, Gao, & Hui, 1997). For this reason it has been advocated that one use a nonlinear conditional mean model in the second part of the two-part model rather than a log-linear regression model (Mullahy, 1998). It is clear from (12) that an exponential version of the two-part model obtains from the same primitive model used to derive (12). One simply doesn't transform the outcome as it is generated by the reduced-form equation k( ).

A comparable exponential conditional mean version of the sample-selection model can be derived from an alternative primitive behavioral model. Furthermore, the same qualitative features of the previous example will hold here as well. In particular, assume that $(X_u \mid X_p, X_{o1}) \sim N(0, 1)$ and that

$$q(X_p, X_{o1}, X_u; \gamma_2) = \exp(X_p\beta_{p2} + X_{o2}\beta_{o2} + X_u\beta_u). \quad (17)$$

Equation (17) is multiplicatively separable in the observable components and the unobserved confounder so that the integration problem of (10) can be written as

$$k(X_p, X_{o1}; \beta_2) = \exp(X_p\beta_{p2} + X_{o2}\beta_{o2}) \times \frac{\int_{-X_{po1}\beta_1}^{\infty} \exp(X_u\beta_u)\varphi(X_u)dX_u}{\Phi(X_{po1}\beta_1)}$$

$$= \exp(X_p\beta_{p2} + X_{o2}\beta_{o2}) \times \exp(0.5\beta_u^2) \times \frac{\Phi(X_{po1}\beta_1 + \beta_u)}{\Phi(X_{po1}\beta_1)}. \quad (18)$$

Equation (18) suggests a two-stage estimator. In stage one, the hurdle parameters, β_1 can be estimated via conventional probit. In stage two, the extent of participation parameters, $\beta_2' = \begin{bmatrix} \beta_{p2} & \beta_{o2}' & \beta_u \end{bmatrix}$ can be estimated by applying conventional nonlinear least-squares to the following regression version of (18):

$$Y^1 = \exp(X_p\beta_{p2} + X_{o2}\beta_{o2}^+) \times \frac{\Phi(X_{po1}\hat{\beta}_1 + \beta_u)}{\Phi(X_{po1}\hat{\beta}_1)} + \omega, \quad (19)$$

where β_{o2}^+ is equivalent to β_{o2} except that the constant term has been shifted by $+0.5\beta_u^2$, $\hat{\beta}_1$ denote the first-stage estimates, and ω denotes a regression residual.

Discussion

The examples above illustrate the fact that the two-part and sample-selection approaches are fundamentally different. Even though (13) and (16) [as well as (12) and

(18)] appear to be quite similar, they are derived from quite different primitive behavioral models [i.e., (14) and (15), respectively; or (14) and (17), respectively]. This strongly indicates that the two approaches are not nested in one another, and therefore testing a zero restriction on β_u is not an adequate test to determine whether the two-part approach is appropriate or not. Testing a zero restriction on β_u does however represent a test of independence, and in the case of independence the restricted model will in principle yield more precise results. Furthermore, the restricted model is in computational terms equivalent to the two-part model. This suggests a concept of practical nesting could be used if one is willing to specify the primitive behavioral model.

These examples also illustrate that the two approaches are apparently equally flexible in terms of the assumptions necessary to conduct estimation. The only meaningful difference, in terms of the statistical assumptions used, is the specification of the primitive behavioral function. The two-part model takes a more reduced-form approach to estimation, while the sample-selection model takes a more structural approach to estimation. The sample-selection model, though, imposes a very important restriction on the researcher. In particular, to identify the parameters in (16) [or (18)], one must have at least one valid exclusion restriction. In the context of health expenditure analyses exclusion restrictions are quite difficult to conceptualize, let alone obtain in data.

Seshamani and Gray (2004) make a big deal about the fact that the implications of the two-part and sample-selection models can be (and are in their application) radically different. One thing to bear in mind, though, is that the fundamental difference between

the two models—in terms of relative performance—comes down to a multi-collinearity issue that occurs as a result of the selectivity correction in the sample-selection model, but is absent from the potentially biased two-part model. The obvious criticism with regard to Zweifel et al. (1999) is to point out their reliance on statistical *in*significance of age in touting their hypothesis as "proven" is misguided, if the insignificance is due to a flaw in their singularly chosen econometric model. In the reverse case one wonders whether a significant result produced by the sample-selection model is particularly impressive, since it is presumably (large sample) unbiased and likely suffers from the multi-collinearity problem. Similarly, an insignificant effect, when significance is the (soundly reasoned) theorized result, may not be the end of the story if the sample-selection model is misapplied. It is in this sense that one's best bet may be to let the data decide through statistical testing of the fit, or performance, of the two competing methods.

CHAPTER III

THE TWO-PART MODEL WITH AN ENDOGENOUS REGRESSOR

Analyzing the causal factors contributing to health care expenditures or utilizations is an important endeavor in the field of health economics, replete with many policy implications, and associated with a rich economic theory and a number of testable hypotheses. The two-part model, developed by Cragg (1971) as a flexible extension to the Tobit model (Tobin, 1958), has achieved considerable and persistent popularity in the estimation of health care expenditure or utilization models. In this chapter, we cast the two-part model in the broader context of a treatment effect model (Rubin, 1974; 1977). We then develop an important methodological extension that facilitates estimation of treatment effects when treatment is endogenous due to omitted variables. Empirical applications that would require such a methodology abound. For example, if one is interested in the effect of health insurance on health care expenditures, then one is likely concerned that some unobserved factors may contribute simultaneously to the decision to enroll in health insurance and to the probability, and/or extent, of health care expenditures.

Despite the popularity of the two-part model, and the obvious need for a robust methodology adept at handling endogenous treatment effects, surprisingly few methodological studies have developed a two-part model extension along this line. Our literature search produced only a handful of possible two-part model approaches that

account for endogenous treatments (Welch, Frank, & Diehr, 1984; Parente & Evans, 1998; Deb, Munkin, & Trivedi, 2006; Zheng & Zimmer, 2008; Cawley & Meyerhoefer, 2012; Biro, 2013). In our review we noted that none of these approaches handled the issue of "separability" with any level of sophistication, where by "separability" we mean whether the two-part model can continue to be estimated sequentially with none of the first-part parameters contributing to the second-part estimator, or vice versa, *once endogeneity has been introduced*. In fact, insofar as separability was discussed in these studies, it was imposed as an assumption of the endogenous two-part model. Thus our first contribution is to develop a new estimation protocol that extends the two-part model in such a way that separability can be avoided, and empirically tested for, within the context of endogenous treatment effects.

The extant literature is also limited in two other respects. First, Welch et al. (1984), Parente and Evans (1998), and Deb et al. (2006) all assume that there exists some latent positive level of the outcome for individuals with observed zero outcomes. This assumption does not comport with the conventional two-part model, and implies the possibility that there exists some correlation or dependence between the two parts of the model. This possibility is not accounted for in the modeling strategy developed by any of these authors. In contrast, we develop our two-part model extension by explicitly assuming no latent levels at zero at the outset. Second, the Deb et al. (2006) two-part model is based on full distributional assumptions in both parts. While this approach may yield efficiency gains, it introduces the possibility of bias if the distributional assumptions do not hold. This is a significant drawback of their approach. Therefore, we

develop a model that can be consistently estimated with fewer distributional assumptions, but could be straightforwardly extended to a full-information maximum likelihood framework if desired.

The chapter proceeds as follows. In the next section, our treatment effect model is rigorously described. We then develop the two-part model extension and discuss estimating the average treatment effect within the context of the resultant model. Next, we present a simulation study that provides compelling evidence regarding the asymptotic properties of our new approach and demonstrates the potential bias that could result from ignoring the separability issue as previous approaches have. The final section concludes.

An Average Treatment Effect Model

The average treatment effect parameter described here is based on the generic parameter described in Terza (2014b). The average treatment effect of a policy relevant dichotomous variable (e.g., whether or not a Medicare beneficiary enrolls in a supplementary health insurance plan) on a quantitative outcome is intuitively defined as the average amount by which the quantitative outcome is changed when an individual is treated (i.e., when an individual has supplementary health insurance). In the canonical case, established by Rubin (1974, 1977), one makes use of linear regression-based estimation to predict two (potentially) counterfactual outcomes: the outcome that would occur if an individual were treated, and the outcome that would occur if an individual were not treated. Rubin's (1974, 1977) average treatment effect is defined as the difference between these two counterfactual outcomes. We need to define a more precise

statement of the average treatment effect in the context of nonlinear regression-based estimation. We can then link this precise statement of the average treatment effect with the two-part model.

With this in mind we begin with some notation and definitions. First, we have framed our estimation objective in terms of a binary, policy relevant, variable. Henceforth we let X_p denote the policy variable. We have also framed our estimation objective in terms of two counterfactual outcomes. To make the concept of a counterfactual outcome more concrete we first define an exogenously mandated version of the policy variable, and denote this exogenously mandated version of the policy variable as X_p^*. Then let Y denote the observed outcome, and let $Y_{X_p^*}$ denote the counterfactual outcome that would occur if the policy variable were exogenously mandated to be X_p^*. To describe the link between regression-based parameter estimation and estimation of the average treatment effect, we use a generic regression model for the observed outcome, Y:

$$Y = \mu(X_p, X_o, X_u; \tau) + e, \tag{20}$$

where X_o denotes a vector of observable confounders (i.e., variables that are correlated with the outcome, Y, as well as the policy variable, X_p), and where X_u denotes a scalar confounder that encompasses all sources of unobserved confounding. The function $\mu(\ )$ denotes the conditional mean of Y given all relevant information (i.e., X_p, X_o, and X_u),

and is a potentially nonlinear function of some unknown parameters denoted by τ. The residual e is tautologically defined as the difference between Y and the conditional mean of Y [i.e., $e \equiv Y - \mu(X_p, X_o, X_u; \tau)$]. By construction, the regression model specified in (20) is comprehensive in the sense that all sources of confounding are accounted for since X_u is included in the conditioning set. We use this fact, and the fact that the exogenously mandated version of the policy variable, X_p^*, is independent of the outcome and all sources of confounding, to define the counterfactual version of the outcome, $Y_{X_p^*}$, as:

$$Y_{X_p^*} = \mu(X_p^*, X_o, X_u; \tau) + e_{X_p^*}. \tag{21}$$

That is, in the presence of comprehensive conditioning, the conditional mean function provides access to the counterfactual outcome we desire.

The average treatment effect was described above as the difference between two counterfactual outcomes. In the notation developed in the previous paragraph, we define the average treatment effect parameter in the nonlinear context as:

$$\begin{aligned} ATE &= E\left[Y_1\right] - E\left[Y_0\right] \\ &= E_{X_o, X_u}\left[\mu(1, X_o, X_u; \tau)\right] - E_{X_o, X_u}\left[\mu(0, X_o, X_u; \tau)\right], \end{aligned} \tag{22}$$

where our notation intends to make clear that the expectations in (22) are with respect to the confounding data. Consistent estimation of the average treatment effect parameter

defined in (22) requires resolving two issues. First, we must specify an appropriate functional form for $\mu(\ )$ (i.e., choose a particular regression-based estimator). Second, we must determine how to deal with the presence of the unobserved confounder, X_u. The two-part model constitutes a particular assumption regarding the correct functional form for $\mu(\ )$. In the next section we describe our approach to handling the presence of X_u (or, alternatively, our approach to handling the endogeneity of X_p) within the confines of a two-part model. We then are prepared to revisit the problem of estimating equation (22) consistently.

Endogenous Treatment in the Two-Part Model Context

The conventional two-part model features two components that are typically referred to as the hurdle and levels components. The hurdle component models the dichotomous outcome: Y is either observed with a value of zero or Y is observed with a strictly positive value. The levels component models the positive observed outcomes, conditional on the outcome being positive. Let H denote the hurdle outcome and let Y^{ℓ} denote the levels outcome. We specify these components as follows:[5]

$$H = I(X_p\beta_{p1} + X_o\beta_{o1} + X_u\beta_{u1} + \varepsilon_1 > 0) \tag{23}$$

[5] Note that the levels components specified here is equivalent to that employed by Duan, Manning, Morris, and Newhouse (1983). It is presented on a raw scale rather than a log scale, though. We adopt this approach to avoid the retransformation problem. See Abrevaya (2002), Ai and Norton (2000, 2008), Barber and Thompson (2004), Basu and Manning (2009), Blough, Madden, and Hornbrook (1999), Miller (1984), Duan (1983), Manning (1998), Manning, Basu, and Mullahy (2005), Manning and Mullahy (2001), Mullahy (1998), and Taylor (1986), Zhou, Gao, and Hui (1997) for details about the retransformation problem and available solutions. It is our contention that the retransformation problem has no solution that rivals an avoidance solution.

$$Y^{\ell} = \exp(X_p\beta_{p2} + X_o\beta_{o2} + X_u\beta_{u2} + \varepsilon_2) \tag{24}$$

where I(C) denotes the indicator function, which reports one if condition C is true and zero otherwise. We assume that there are no latent levels at zero so that the levels outcome, Y^{ℓ}, is only defined if H = 1. At a minimum we impose the following two assumptions on the model as stated thus far: 1) a probit assumption for the hurdle component:

$$(\varepsilon_1 \mid X_p, X_o, X_u) \sim N(0,1), \tag{25}$$

and 2) the following conditional mean assumption:

$$E\left[\exp(\varepsilon_2) \mid X_p, X_o, X_u, H = 1\right] = 1. \tag{26}$$

A logistic assumption could easily replace the normality assumption [equation (25)]. The normality assumption is typical in the context of binary outcome models though, and turns out to be especially convenient in the present example. Assumption 2) is minimally parametric in the sense that it suggests a nonlinear least-squares estimator, which is relatively more robust than a model based on full distributional assumptions regarding ε_2.

Based on the two components of the model [i.e., on the basis of equations (23) and (24)], and based on the assumption of no latent levels at zero, the outcome, Y, can be written as follows:

$$Y = \begin{cases} 0 & \text{if } H = 0 \\ \exp(X_p\beta_{p2} + X_o\beta_{o2} + X_u\beta_{u2} + \varepsilon_2) & \text{if } H = 1 \end{cases}. \quad (27)$$

Typically when the two-part model is applied, it is posited that

$$E\left[Y \mid X_p, X_o, X_u \right] = \Phi(X_p\beta_{p1} + X_o\beta_{o1} + X_u\beta_{u1}) \exp(X_p\beta_{p2} + X_o\beta_{o2} + X_u\beta_{u2}). \quad (28)$$

However, this implicitly assumes that there are some latent levels at zero (see Terza, 2014b). We follow the practice proposed by Terza (2014b), and use the following conditional mean expression:

$$E\left[Y \mid X_p, X_o, X_u \right] = H \times \exp(X_p\beta_{p2} + X_o\beta_{o2} + X_u\beta_{u2}) \quad (29)$$

in our formulation of the average treatment effect. This approach is logically compatible with the no latent levels at zero assumption. Equation (29) leads naturally to the following average treatment effect parameter:

$$\begin{aligned} ATE = E_{X_o, X_u}\left[I(\beta_{p1} + X_o\beta_{o1} + X_u\beta_{u1} + \varepsilon_1 > 0) \exp(\beta_{p2} + X_o\beta_{o2} + X_u\beta_{u2}) \right] \\ - E_{X_o, X_u}\left[I(X_o\beta_{o1} + X_u\beta_{u1} + \varepsilon_1 > 0) \exp(X_o\beta_{o2} + X_u\beta_{u2}) \right], \end{aligned} \quad (30)$$

where we have replace H with its data generating equation. We have also tested, and prefer, an alternative that makes use of the fact that the best prediction for $Y_{X_p^*}$ when an

individual is treated is the observed outcome, Y. That is, we have tested an average treatment effect estimator based on the following:

$$\text{ATE} = \text{E}_{\text{X}_\text{o},\text{X}_\text{u}}\left[\ddot{\text{Y}}_1 - \ddot{\text{Y}}_0\right], \tag{31}$$

where

$$\ddot{\text{Y}}_{\text{X}_\text{p}^*} = \begin{cases} \text{Y} & \text{if } \text{X}_\text{p} = \text{X}_\text{p}^* \\ \text{I}(\text{X}_\text{p}^*\beta_{\text{p1}} + \text{X}_\text{o}\beta_{\text{o1}} + \text{X}_\text{u}\beta_{\text{u1}} + \varepsilon_1 > 0)\exp(\text{X}_\text{p}^*\beta_{\text{p2}} + \text{X}_\text{o}\beta_{\text{o2}} + \text{X}_\text{u}\beta_{\text{u2}}) & \text{if } \text{X}_\text{p} \neq \text{X}_\text{p}^* \end{cases} \tag{32}$$

The discussion above serves to define the average treatment effect parameter in the two-part model context. To proceed with estimating the average treatment effect we need to estimate the two-part model parameters and deal with the fact that X_u is unobserved. To derive a consistent estimator for the two-part model parameters we adopt the generic approach developed in Terza (2009). Specifically, let W_1 denote a vector of identifying instrumental variables, and let $\text{W} = \left[\text{W}_1 \quad \text{X}_\text{o}\right]$ denote the full set of exogenous variables.[6] Then assume that the policy variable is generated as follows:

$$\text{X}_\text{p} = \text{I}(\text{W}\alpha + \text{X}_\text{u} > 0), \tag{33}$$

[6] Note that the model is just-identified when $\text{W}_1 \in \mathbb{R}$ and is over-identified when $\text{W}_1 \in \mathbb{R}^\text{k}$.

where $(X_u \mid W) \sim N(0,1)$. The probit assumption for the hurdle component and the probit assumption in (33) can be combined to derive a bivariate probit model that can consistently estimate the parameters α and $\beta_1' = \left[\beta_{p1} \quad \beta_{o1}' \quad \beta_{u1}\right]$. To see this, note that by appealing to the argument in Terza and Tsai (2006), one can re-parameterize the hurdle component as follows.[7] Let

$$\gamma_{p1} = \beta_{p1}\sqrt{1-\rho^2}\,, \tag{34}$$

$$\gamma_{o1} = \beta_{o1}\sqrt{1-\rho^2}\,, \tag{35}$$

and

$$\rho = \frac{\beta_{u1}}{\sqrt{1+\beta_{u1}^2}}\,. \tag{36}$$

Then

$$H = I(X_p\gamma_{p1} + X_o\gamma_{o1} + \upsilon > 0) \tag{37}$$

can be used in place of the hurdle equation specified in (23), and it can be shown that, conditional on W, the errors X_u and υ are standard bivariate normally distributed with correlation ρ. Thus the parameters can be consistently estimated by applying

[7] Wooldridge (2010) demonstrates a similar result as well (see pages 594-95 of his text).

conventional bivariate probit analysis to equations (33) and (37). The original parameters can be obtained by re-transforming via equations (34) through (36).

Since the levels component of the two-part model is defined only when H = 1, we can formulate an estimator for the parameters of the levels equation on the basis of a conditional mean expression for Y conditional on H = 1. It can be shown under the assumptions made thus far that[8]

$$E\left[Y \mid X_p, W, H = 1 \right] = X_p \frac{\int_{-W\alpha}^{\infty} \Phi(X\beta_1)\exp(X\beta_2)\varphi(X_u)dX_u}{\int_{-W\alpha}^{\infty} \Phi(X\beta_1)\varphi(X_u)dX_u}$$

$$+(1-X_p)\frac{\int_{-\infty}^{-W\alpha} \Phi(X\beta_1)\exp(X\beta_2)\varphi(X_u)dX_u}{\int_{-\infty}^{-W\alpha} \Phi(X\beta_1)\varphi(X_u)dX_u}, \quad (38)$$

where $X = \left[X_p \quad X_o \quad X_u \right]$. The conditional mean in (22) can serve as the basis for a nonlinear least-squares estimator since the unobserved confounder, X_u, is "integrated out" so that X_u is not a conditioning variable. The integrals in (22) have no closed form solution, but can be approximated via Monte Carlo integration or Gauss-Legendre quadrature after applying an appropriate change of variables (see Appendix B). On the basis of (22) the following two-stage estimator is consistent for the full set of parameters.

Stage 1: Estimate α and β_1 via conventional bivariate probit analysis as suggested above. ***Stage 2:*** Estimate β_2 by applying nonlinear least squares to:

[8] See Appendix B.

$$Y = X_{pi}\frac{\int_{-W_i\hat{\alpha}}^{\infty}\Phi(X_i\hat{\beta}_1)\exp(X_i\beta_2)\varphi(X_u)dX_u}{\int_{-W_i\hat{\alpha}}^{\infty}\Phi(X_i\hat{\beta}_1)\varphi(X_u)dX_u}$$

$$+(1-X_{pi})\frac{\int_{-\infty}^{-W_i\hat{\alpha}}\Phi(X_i\hat{\beta}_1)\exp(X_i\beta_2)\varphi(X_u)dX_u}{\int_{-\infty}^{-W_i\hat{\alpha}}\Phi(X_i\hat{\beta}_1)\varphi(X_u)dX_u}+\omega_i, \tag{39}$$

using the subsample of observations for whom H = 1. The "$\hat{}$ s" in (39) denote first-stage estimates and ω_i is a regression error term. Appendix B provides mathematical details needed for constructing an optimization algorithm for the second-stage. The asymptotic properties of this two-stage estimator can be obtained as described in Terza (2014c). They are also detailed in Appendix B.

To complete the discussion we need to describe a consistent estimator for the average treatment effect parameter defined in (31). With consistent estimates for all of the model parameters it is fairly straightforward to replace the population moments in (31) with sample analogs. The main difficulty is that X_u remains in these expressions. Thus proceeding from (31) we can use the law of iterated expectations to write:

$$ATE = E_{X_o}\left[\int_{-\infty}^{\infty}(\ddot{Y}_1 - \ddot{Y}_0)\varphi(X_u)dX_u\right], \tag{40}$$

which can be replaced with the following consistent estimator:

$$\widehat{ATE} = \Sigma\frac{1}{n}\left[\int_{-\infty}^{\infty}\left\{\hat{\ddot{Y}}_1 - \hat{\ddot{Y}}_0\right\}\varphi(X_u)dX_u\right], \tag{41}$$

where $\hat{\hat{Y}}_{X_p^*}$ denotes the estimated version of (32), where as described by Terza (2014a), one replaces the error term ε_1 with pseudo-random variates drawn from the standard normal distribution. The integrals in (41) can also be approximated via Monte Carlo integration or Gauss-Hermite quadrature.

The discussion above makes clear that the hurdle parameters enter into the levels estimator when endogeneity is introduced in the conventional two-part model. Therefore we conclude that it is not generally true that separability is an appropriate imposition once endogeneity has been introduced. This is the inherent weakness involved with the approaches developed by Welch et al. (1984), Parente and Evans (1998), and Zheng and Zimmer (2008). It can be seen, however, that separable estimation constitutes a consistent approach if $\beta_{u1} = 0$. This is demonstrated straightforwardly by noting that if $\beta_{u1} = 0$, then the following conditional mean expression obtains:

$$E\left[Y \mid X_p, W, H = 1 \right] = \exp(X_p\beta_{p2} + X_o\beta_{o2}^{+}) \times \left(X_p \frac{\Phi(\beta_{u2} + W\alpha)}{\Phi(W\alpha)} + (1 - X_p)\frac{1 - \Phi(\beta_{u2} + W\alpha)}{1 - \Phi(W\alpha)} \right) \qquad (42)$$

where β_{o2}^{+} is equivalent to β_{o2} except that the constant term is shifted by $0.5\beta_{u2}^2$.[9] Clearly the hurdle parameters are absent from (42). On the basis of equation (42) the following three-stage alternative is consistent for the full set of parameters.

[9] See Appendix E.

Stage 1: Estimate α by applying conventional probit analysis to equation (33) using the full sample. ***Stage 2:*** Estimate β_{p1} and β_{o1} by applying conventional probit analysis to equation (23) using the full sample, and setting $\beta_{u1} = 0$. ***Stage 3:*** Estimate β_2 by applying nonlinear least-squares to the following regression version of (42):

$$Y_i = \exp(X_{pi}\beta_{p2} + X_{oi}\beta_{o2}^{+}) \times \left(X_{pi} \frac{\Phi(\beta_{u2} + W_i\hat{\alpha})}{\Phi(W_i\hat{\alpha})} + (1 - X_{pi}) \frac{1 - \Phi(\beta_{u2} + W_i\hat{\alpha})}{1 - \Phi(W_i\hat{\alpha})} \right) + \zeta_i, \qquad (43)$$

using the subsample of observations for whom H = 1, where $\hat{\alpha}$ denotes the first-stage estimate of α and ζ_i is a regression error term.

Stages 1 and 3 constitute a more general version of the two-stage estimator suggested by Terza (1994, 1998) and comprise a special case of the generic method introduced by Terza (1996, 2009). Although there are three stages to this estimator it is relatively simple compared to the two-stage estimator describe previously since the methods applied at the first two stages are conventional, widely packaged routines, and the method applied at stage three is a simpler nonlinear least-squares objective function than the one given in (39). It is worth noting that $\beta_{u1} = 0$ if and only if $\rho = 0$. Thus a test of whether the three-stage estimator is preferred over the two-stage estimator can be conducted by estimating the first-stage of our two-stage estimator and testing the null hypothesis that $\rho = 0$.

The average treatment effect parameter when $\beta_{u1} = 0$ is slightly different also. In this case we can show that the following is appropriate:

$$\text{ATE}_0 = \text{E}_{\text{X}_o}\left[\breve{\text{Y}}_1 - \breve{\text{Y}}_0\right], \tag{44}$$

where

$$\breve{\text{Y}}_{\text{X}_p^*} = \begin{cases} \text{Y} & \text{if } \text{X}_p = \text{X}_p^* \\ \text{I}(\text{X}_p^*\beta_{p1} + \text{X}_o\beta_{o1} + \varepsilon_1 > 0)\exp(\text{X}_p^*\beta_{p2} + \text{X}_o\beta_{o2}^+) & \text{if } \text{X}_p \neq \text{X}_p^* \end{cases}. \tag{45}$$

The sample analog estimator in this case is

$$\widehat{\text{ATE}_0} = \Sigma\frac{1}{\text{n}}\left[\hat{\breve{\text{Y}}}_1 - \hat{\breve{\text{Y}}}_0\right], \tag{46}$$

where $\hat{\breve{\text{Y}}}_{\text{X}_p^*}$ denotes the estimated version of (45), and where we replace the error term ε_1 with pseudo-random variates drawn from the standard normal distribution.

Simulation Analysis of the New Methodology

To test the performance of our new methodological approach we simulated data and applied our estimator alongside a candidate comparison estimator. The comparison estimator closely aligns with the approaches taken by Welch, Frank, and Diehr (1984), Parente and Evans (1998), Zheng and Zimmer (2008), Cawley and Meyerhoefer (2012), and Biro (2013). In particular, we apply first-, and second-, part estimators that would

adequately account for the inclusion of X_u if each of the parts were considered in isolation from the other. That is, if the two parts retained their separability after endogeneity is introduced, then this approach would be consistent. We refer to the comparison estimator as the naïve TPM henceforth. The naïve TPM comprises of the following two stages.

Stage 1: Estimate α and $\gamma' = \left[\gamma_{p1} \quad \gamma_{o1}' \right]$ (and ρ) by applying conventional bivariate probit analysis to equations (33) and (37) using the full sample. β_1 can be obtained by using (34) through (36). (Note that this is equivalent to the first-stage of our proposed estimation protocol.) ***Stage 2:*** Estimate β_2 as described with Stage 3 of the 3S2PT model above. Stage 2 is consistent if the conditional mean is (42), which we showed is not generally true. We apply the same average treatment effect estimator as described with our consistent two-stage approach, but use the inconsistent second-stage estimates.

The Monte Carlo study we conducted is based on the following data generating process. The exogenous data vector, W, was obtained as independent pseudo-random draws from uniform distributions with means set to 0.5 and 0.25 and variances set to 4 and 2, respectively. Once the exogenous data were obtained we generated the policy variable, X_p, and the hurdle outcome, H, according to the data generating equations:

$$X_p = I(W\alpha + X_u > 0), \tag{47}$$

and

$$H = I(X_p\beta_{p1} + X_o\beta_{o1} + X_u\beta_{u1} + \varepsilon_1 > 0), \tag{48}$$

where X_u and ε_1 were obtained as draws from independent standard normal distributions. Finally the levels outcome was obtained as a draw from the generalized gamma distribution with parameters $\sigma = 0.5$, $\kappa = 0.25$, and

$$\mu = \exp(X_p\beta_{p2} + X_o\beta_{o2} + X_u\beta_{u2}). \tag{49}$$

This is the parameterization of the generalized gamma distribution described by Manning, Basu, and Mullahy (2005). Manning et al. (2005) show that under this data generating process the conditional mean is

$$E\left[Y^{\ell} \mid X_p, X_o, X_u\right] = \exp(X\beta_2 + f\{\sigma,\kappa\}), \tag{50}$$

where

$$f\{\sigma,\kappa\} = \frac{\sigma}{\kappa}\ln(\kappa^2) + \ln\left(\Gamma\left\{\frac{1}{\kappa^2} + \frac{\sigma}{\kappa}\right\} - \ln\left(\Gamma\left\{\frac{1}{\kappa^2}\right\}\right)\right), \tag{51}$$

where $\Gamma\{\ \}$ denotes the gamma function. The outcome, Y, was set to equal zero whenever H = 0, and equal to Y^{ℓ} whenever H = 1. The estimator we apply in our Monte Carlo study does not make use of the knowledge that the levels outcome is distributed generalized gamma. Thus the constant term in the second-stage of our consistent

estimator is actually converging to the "true" constant term plus $f\{\sigma,\kappa\}$. We correct for this when we present our results.

We conducted two Monte Carlo studies with samples of size 100,000 and 50 replications each. The first Monte Carlo study sets $\beta_{u1} = 2.00$ so that our estimator is consistent but the naïve TPM is not consistent. In terms of the first stage estimator that is common across both approaches setting $\beta_{u1} = 2.00$ implies a correlation of 0.9. Thus this data generating process is a fairly extreme case and gives a worst-case picture of the naïve TPM. The second Monte Carlo study sets $\beta_{u1} = 0.00$. The naïve TPM should perform very similar to the 2S2PT estimator.

Table 1 presents results from these two studies. The 2S2PT estimator is consistent under both Monte Carlo scenarios. The naïve TPM is inconsistent under the first Monte Carlo study. The absolute percent bias (APB) for the average treatment effect is 50.37%, which is similar in magnitude to the biases associated with β_{p2} and β_{o2}. The naïve TPM performs very well, and similarly to the 2S2PT estimator, under the second Monte Carlo scenario.

Discussion

The conventional TPM remains a popular model in the estimation of health care expenditures. Many health economic questions require modeling health care expenditures with an endogenous right-hand side dummy variable. Despite the

Table 1

Results from a Simulation Analysis

	$\beta_{u1} \neq 0$			$\beta_{u1} = 0$		
	True Values	2S2PT[a]	Naïve TPM	True Values	2S2PT[a]	Naïve TPM
ATE:						
X_p	1.56	1.57	2.34	1.60	1.61	1.61
		(0.81)	(50.37)		(0.71)	(0.60)
		[0.03]	[0.63]		[0.04]	[0.04]
Policy Variable:						
W_1	0.50	0.50		0.50	0.50	
		(0.19)			(0.04)	
		[0.00]			[0.00]	
X_o	0.50	0.50		0.50	0.50	
		(0.09)			(0.02)	
		[0.00]			[0.00]	
Constant	0.50	0.50		0.50	0.50	
		(0.16)			(0.01)	
		[0.00]			[0.00]	
Hurdle Outcome:						
X_p	0.50	0.50		0.50	0.51	
		(0.23)			(1.26)	
		[0.00]			[0.00]	
X_o	0.50	0.50		0.50	0.50	
		(0.13)			(0.34)	
		[0.00]			[0.00]	
X_u	2.00	2.01		0.00	-0.00	
		(0.31)			(NA)	
		[0.00]			[0.00]	
Constant	0.50	0.50		0.50	0.50	
		(0.32)			(0.88)	
		[0.00]			[0.00]	
Levels Outcome:						
X_p	0.50	0.50	0.75	0.50	0.51	0.50
		(0.99)	(50.73)		(1.05)	(0.89)
		[0.00]	[0.07]		[0.01]	[0.01]
X_o	0.50	0.50	0.47	0.50	0.50	0.50
		(0.19)	(5.29)		(0.11)	(0.13)
		[0.00]	[0.00]		[0.00]	[0.00]
X_u	0.50	0.49	0.24	0.50	0.50	0.50
		(1.05)	(52.87)		(0.76)	(0.68)
		[0.00]	[0.07]		[0.00]	[0.00]
Constant	0.50	0.50	0.49	0.50	0.50	0.50
		(0.08)	(1.07)		(0.98)	(0.92)
		[0.00]	[0.00]		[0.00]	[0.00]
Observations:			100,000			
Replications:			50			

[a] 2S2PT: Our two-stage two-part treatment effect estimator.
Absolute percent bias in parentheses.
Mean squared error in square brackets.

popularity of the TPM, and the overwhelming need for appropriate methodology to handle dummy endogenous regressors, little methodological research has been conducted toward this end. We have derived a novel approach to handling dummy endogenous regressors within the context of the TPM. Our approach improves upon the extant approaches in that it comports more closely with the conventional TPM approach and because it relies on relatively few distributional assumptions thereby making it robust to distributional misspecification.

Our simulation analysis demonstrates that naïve TPM approaches to accounting for endogeneity in TPMs is not adequate. In fact significant biases exist with both the slope parameters and with a typical post-estimation treatment effect parameter. Furthermore, we demonstrated that a testable restriction can be imposed such that naïve approaches as well as our approach will be consistent. This provides applied researchers with the necessary statistical testing procedures so that one can determine the best modeling approach.

CHAPTER IV

MEDICARE PART D AND THE USE OF HOSPITAL SERVICES

Prescription medicines represent a clinically important component in the treatment of a variety of illnesses. This is especially true among the chronically ill and elderly. Furthermore, drug spending accounts for a large and expanding proportion of total medical expenditures among both of these groups. Stagnitti (2007) reports that prescription drug spending accounted for approximately 20% of total medical expenditure in 2004, which represents a marked increase from 13% in 1997. In the years preceding the implementation of Medicare Part D a substantial number of Medicare beneficiaries were without prescription drug coverage. Safran et al. (2005) estimate that approximately 27% of seniors were without medication insurance in 2003 and that many of these uninsured seniors were poor or near poor. Concerns regarding the health and financial well being of the uninsured provided impetus for the passage of the Medicare Prescription Drug, Improvement, and Modernization Act of 2003, the legislative foundation for Part D.

In the health services research literature a hypothesis important to the evaluation of Medicare Part D has emerged known as the "cost-offset" hypothesis.[10] This hypothesis

[10] See Briesacher, Stuart, Ren, and Doshi (2005), Sokol, McGuigan, Verbrugge, and Epstein (2005), Hsu et al., 2006; Shang and Goldman (2007), Deb, Trivedi, and Zimmer (2009), Stuart, Doshi, and Terza (2009), Zhang, Donohue, Lave, O'Donnell, and Newhouse (2009), Chandra, Gruber, and McKnight (2010), Afendulis, Zaslavsky, and Chernew (2011), Liu et al., 2011; McWilliams, Zaslavsky, and Huskamp (2011), Kaestner and Khan (2012), and Kaestner, Long, and Alexander (2014).

proposes that Part D may reduce the use of non-drug medical resources, and thereby mitigate the cost of the drug insurance program. The rationale behind this hypothesis is that drug insurance could positively affect medication adherence, and consequently decrease the likelihood of adverse health events that would have otherwise led to the use of hospital and/or physician services. Evidence in support of the cost-offset hypothesis has been used to claim that Medicare Part D has improved the health of beneficiaries (e.g., MAPRx Coalition, 2013). A negative relationship between Part D and the use of hospital services could also indicate a substitution across medical services without a corresponding change in health, however.

Understanding the relationship between Part D and hospital utilizations is particularly relevant in assessing the costs and benefits associated with Medicare Part D. On the cost side, if Part D insurance negatively impacts the use of other health care services, then the calculated cost of the program should be adjusted to reflect the savings associated with a diminished use of these services. On the benefit side, a negative relationship between Part D and the use of non-drug medical resources may indicate that Part D has an important effect on the health of beneficiaries. In broader terms of economic efficiency, Goldman and Philipson (2007) argue that the degree of substitutability across multiple insured medical technologies is an important factor in determining the optimal cost-sharing arrangement. Thus additional knowledge regarding the extent to which drugs are substituted for hospital resources could help policymakers increase the cost-effectiveness of Medicare.

Previous studies have attempted to link prescription drug use or coverage with a change in patient health,[11] to link prescription drug insurance to changes in drug use or expenditure,[12] and to link prescription drug use or coverage with the use of other health care services.[13] Among these latter studies the literature broadly supports the conclusion that drug coverage reduces the use of other health care services (see Congressional Budget Office, 2012). More research is necessary, however, to precisely determine the magnitude of the relationship between Part D and the use of non-drug medical care (see Lau and Stubbings, 2012). Furthermore, no study has attempted to holistically investigate the relationships between drug coverage, health, and the use of hospital services. The first contribution of this chapter is to investigate the extent to which changes in health explain the relationship between Part D and the number of hospitalizations.[14]

The second contribution is a focus on the previously uninsured. Despite the obvious policy relevance of this group, little research has directly analyzed the effect of Medicare Part D among these beneficiaries. Briesacher et al. (2005), Zhang et al. (2009), and McWilliams et al. (2011) are the only studies reviewed that provide point estimates

[11] See Lichtenberg (1996; 2001) and Heisler et al. (2004).

[12] See Lichtenberg and Sun (2007), Shea, Terza, Stuart, Briesacher (2007), Simoni-Wastila, Zuckerman, Shaffer, Blanchette, and Stuart (2008), Yin et al., 2008; Schneeweiss et al., 2009; Zhang et al., 2009; Basu, Yin, and Alexander (2010), Duggan and Morton (2010), Polinski, Kilabuk, Schneeweiss, Brennan, and Shrank (2010), Safran et al., 2010; and Briesacher et al., 2011 for studies that have looked at changes in expenditures or uses of prescription drugs. See also Piette, Wagner, Potter, and Schillinger (2004), Goldman et al. (2004), Soumerai et al. (2006), Madden et al. (2008), Madden, Graves, Ross-Degnan, Briesacher, and Soumerai (2009), Polinski, Donohue, Kilabuk, and Shrank (2011), and Williams, Steers, Ettner, Mangione, and Duru (2013) for studies that have looked for changes in medication adherence.

[13] See the references in footnote 1.

[14] In the psychometric literature investigating the extent to which a third variable explains the causal relationship between two variables is referred to as a mediation analysis. See Baron and Kenny (1986), Judd and Kenny (1981), and James and Brett (1984).

associated with the previously un-, or under-, insured. Moreover, among these studies, Briesacher et al.'s study is the only one that focuses solely on drug coverage "gainers" versus "nevers," and their study may not sufficiently reflect the experiences of Part D "gainers" as opposed to "gainers" of pre-Part D drug coverage plans. In this chapter I estimate separate parameters for respondents that reported limited or no drug coverage in all survey waves before 2006 and for respondents that reported generous coverage in at least one survey wave before 2006.

Background

Analyses of disease-specific populations indicate a strong negative association between prescription drug coverage (or use) and the use of hospital services. Sokol et al. (2005) estimated the effect that medication adherence has on the probability of being hospitalized among samples of patients with one of four health conditions.[15] They found that medication adherence, derived from data on prescription drug fills, strongly affects the probability of being hospitalized. For instance, they estimated that diabetes patients with more than 80% adherence are 17 percentage points less likely to be hospitalized than diabetes patients with less than 20% adherence. Cole et al. (2006) estimated the effect of increasing prescription drug co-payments on the risk of hospitalization for patients with chronic heart failure. They found that, depending on the medication regimen, the probability of hospitalization increases between 6 and 9% for every $10 increase in drug co-payment. Afendulis et al. (2011) analyzed state-level data on the number of

[15] Diabetes, hypertension, hypercholesterolemia, and congestive heart failure.

hospitalizations associated with eight illnesses.[16] They compared the changes in the hospitalization rates associated with these illnesses across states with "high" and "low" Part D take-up rates. They found that Part D is associated with a 4.1% reduction in the hospitalization rates associated with these illnesses. These studies do not straightforwardly generalize to the Medicare beneficiary population because of the selected nature of the study populations used and because these analyses are primarily descriptive.

Studies that use proprietary insurance data are uniquely capable of identifying the role that drug coverage versus drug use plays in the relationship between prescription drug insurance and the use of physician or hospital services. Furthermore, these studies have clear sources of identifying variation lending them increased causal interpretability. Hsu et al. (2006) analyzed the effect of capping drug benefits at $1,000 per year. They report that capping benefits increased the number of emergency department visits by 9%, and increased the number of non-elective hospitalizations by 13%. Zhang et al. (2009) compared spending outcomes before and after the implementation of Medicare Part D among a group of beneficiaries enrolled in a "large Pennsylvania insurer." They found that those individuals with limited drug coverage before Part D experienced offsetting decreases in non-drug medical spending. Chandra et al. (2010) used data on Medicare beneficiaries receiving supplementary insurance through California Public Employees Retirement System. They compared physician and hospital utilizations across

[16] Short-term complications of diabetes, uncontrolled diabetes, chronic obstructive pulmonary disorder, congestive heart failure, angina, asthma, stroke, and acute myocardial infarction.

beneficiaries that experienced drug co-payment increases and those that did not. They concluded that cost-offsets are possible.

Though the aforementioned studies suggest strong cost-offset effects associated with Part D, concerns regarding the external validity of these results remain. Thus it is important to also consider evidence produced by nationally representative studies. Shang and Goldman (2007) and Deb et al. (2009) employed robust methodology and found evidence of a statistically significant relationship between prescription drug insurance and non-drug medical expenditure. However, these studies produced radically different point estimates. Shang and Goldman estimated that for every dollar spent on drug coverage Medicare saves approximately $2.06 on Part A and B expenditures. Accordingly, their results imply that the net cost of Medicare Part D is negative. Deb et al., on the other hand, found that the cost-offsets resulting from drug insurance are less than dollar-for-dollar, implying a positive net cost of Part D. Kaestner and Khan (2012) conducted a descriptive analysis that did not indicate any significant relationship between Part D and non-drug medical spending. Liu et al. (2011) employed a differences-in-differences model with a small sample (N= 1,105), and estimated a small and statistically insignificant positive relationship between drug coverage and emergency department and hospital use. Kaestner et al. (2014) found that Part D reduced hospital admissions by 8%.

Two additional studies are very similar to mine: a study conducted by Briesacher et al. (2005) and a study conducted by McWilliams et al. (2011). Briesacher et al. (2005) provide prospective evidence of the effect of gaining Part D coverage on hospital and physician spending by comparing drug coverage "gainers" with drug coverage "nevers"

in the pre-Part D era. They estimated a fixed-effects differences-in-differences model, and found small and statistically insignificant effects. McWilliams et al. (2011) investigated whether individuals with limited or no drug coverage before 2006 had a greater reduction in health care expenditures in post-2006 quarters relative to those individuals with generous drug coverage before 2006. They found that those with limited or no drug coverage before 2006 had a relative reduction in total non-drug medical spending of $306 per quarter. The majority of this reduction was driven by changes in inpatient and skilled nursing facility expenditures.

I extend both of the latter two analyses by using an extended panel data set, and by employing a more robust econometric methodology. Furthermore, my study is the first among all of the aforementioned studies to investigate the role that health plays as a potential mechanism that could explain the negative relationship between Part D and non-drug health care utilizations/expenditures. Briesacher et al. (2005), Zhang et al. (2009), and McWilliams et al. (2011) are the only studies that provide evidence regarding the effect of Part D among beneficiaries who had limited or no drug coverage prior to the implementation of Part D. This chapter provides new evidence of the effect of Part D among the previously uninsured. I also provide evidence of the effect of Part D among beneficiaries that switch from generous pre-Part D drug coverage to Part D insurance.

Data

The Health and Retirement Study (HRS) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan. The HRS survey was first conducted in 1992, and has been conducted every

two years thereafter. The HRS is representative of the over-50 population of American residents. The Asset and Health Dynamics Among the Oldest Old Study (AHEAD) is a companion study conducted by the University of Michigan and sponsored by the same NIA grant. The AHEAD survey was administered separately from the HRS survey until 1998 when the samples from the HRS and AHEAD were merged and tracked thereafter as a part of the HRS. The majority of the analysis variables I use were constructed using raw HRS data or AHEAD data from the 1995 survey,[17] but a handful of the control variables come from a file constructed by the RAND Corporation.[18] The RAND HRS data file is an easy-to-use longitudinal data set derived from raw HRS data. It was developed at RAND with funding from the National Institute on Aging and the Social Security Administration.

The data used in this chapter consist of a panel covering the years 1996 through 2010. 28,261 respondents were interviewed over this time period. The primary sample inclusion criterion is enrollment in Medicare Parts A and B.[19] Enrollment in both parts is a prerequisite for enrollment in Medicare Part D. After omitting respondents not enrolled in both Part A and B, the sample included 21,922 individuals.

The dependent variables derive from questions that elicit the number of utilizations that occurred over the past two years or since the previous interview, whichever is longer. Therefore, I omitted all observations on respondents that ever

[17] Approximately half of the respondents included in my 1996 cross-section come from the 1995 AHEAD survey.

[18] Specifically, the following measures were derived from the RAND file: age, marital status, household income, education, gender, race, and geographic region.

[19] Alternatively, a survey respondent can be included in my sample if they are enrolled in a Medicare Advantage plan. In my analyses of Part D I include Medicare Advantage prescription drug plan holders with Part D enrollees.

skipped an interview to ensure that the dependent variables were measured over the same length of time across respondents. Under this inclusion criterion individuals who died or were permanently dropped from the HRS study are retained. Less than 1,000 respondents in my final analysis sample died. After omitting respondents that ever skipped an interview, the sample included 17,974 individuals.

Only 15,534 of the remaining respondents reported their Part D status at each of the post-Part D interviews. At each survey wave the HRS instrument asks respondents to report whether they have ever been diagnosed with several physical health conditions. Some records indicate that a respondent has been diagnosed with a particular health condition in an early wave, but at a later wave has never been diagnosed with that condition. Since a secondary focus is the role that health plays in the relationship between Part D and the use of hospital services, I omit respondents with this type of inconsistent record. After making this omission, the sample included 13,726 individuals. Finally, after omitting respondents with missing left- or right-hand side variables the sample included 10,075 individuals. To assess whether any of the sample restrictions are responsible for the results, I compared the pre-restricted sample to the final analysis sample along all observed dimensions. There were no substantial differences in the means across these two samples.

Two variables were derived to measure hospital utilizations. The first variable measures the number of inpatient stays occurring over a two-year period. The second variable measures the number of hospital nights occurring over this same length of time. The primary regression controls include age, household income, marital status, education,

Table 2

Descriptive Statistics for the Dependent and Demographic Variables

	Full Sample	FE Sample	
	Mean (Overall SD)	Mean (Overall SD)	Within SD
Inpatient Stays	0.59	0.82	1.06
	(1.28)	(1.37)	
Hospital Nights	3.15	4.20	9.94
	(11.58)	(12.35)	
Age	74.22	74.90	3.03
	(6.85)	(6.81)	
Log(Household Income)	10.27	10.25	0.47
	(0.95)	(0.92)	
Married	0.59	0.57	0.19
	(0.49)	(0.49)	
Nonwhite	0.14	0.13	-
	(0.35)	(0.33)	
Female	0.61	0.61	-
	(0.49)	(0.49)	
High School Graduate	0.38	0.38	-
	(0.48)	(0.48)	
Some College	0.20	0.19	-
	(0.40)	(0.39)	
College or Beyond	0.18	0.18	-
	(0.39)	(0.39)	
Other Drug Coverage	0.31	0.30	0.33
	(0.46)	(0.46)	
HMO	0.21	0.21	0.23
	(0.41)	(0.41)	
Medicaid	0.07	0.07	0.13
	(0.26)	(0.26)	
CHAMPUS	0.05	0.05	0.14
	(0.22)	(0.22)	
Medigap	0.63	0.65	0.28
	(0.48)	(0.48)	
Northeast	0.16	0.16	0.05
	(0.36)	(0.36)	
Midwest	0.27	0.28	0.05
	(0.44)	(0.45)	
South	0.39	0.38	0.07
	(0.49)	(0.49)	
Observations:	30650	19495	

race, gender, and a set of indicators to control for the set of health insurance contracts held. All econometric models also include region and year fixed effects. Table 2 provides descriptive statistics for the outcome measures, and the primary regression controls.

Several proxy measures were derived to characterize respondent's health: in particular, measures associated with physical health conditions, measures indicating the respondent's degree of functional limitation, and measures indicating the respondent's mental health status. The physical health conditions include hypertension, diabetes, cancer (excluding skin cancer), lung disease (excluding asthma), stroke, arthritis, and heart disease. For each of these conditions, I derived an indicator that reports whether the respondent has ever been diagnosed with the condition and it is "active." For example, the hypertension variable indicates if the respondent has ever been diagnosed with hypertension and it is not under control. This variable takes a zero value if the respondent has never been diagnosed with hypertension or if the respondent has been diagnosed with hypertension in the past but it is now under control. In all models I use a count of the number of these "active" health conditions because the majority of conditions had too little within variation to be used in the fixed effects specifications.

To measure functional limitations I use an activities of daily living (ADL) score and an instrumental activities of daily living (IADL) score. The ADL score counts how many of the following tasks the respondent has difficulty performing: walking across a room, dressing oneself, bathing oneself, eating, and getting into or out of bed. The IADL

Table 3

Descriptive Statistics for the Health Measures

	Full Sample	FE Sample	
	Mean (Overall SD)	Mean (Overall SD)	Within SD
ADL	0.27	0.30	0.47
	(0.76)	(0.79)	
IADL	0.08	0.08	0.24
	(0.34)	(0.34)	
CESD	1.41	1.50	1.11
	(1.84)	(1.88)	
Number of Active Conditions[a]	0.57	0.63	0.49
	(0.82)	(0.86)	
Hypertension (not under control)	0.02	0.02	0.11
	(0.13)	(0.13)	
Diabetes (not under control)	0.01	0.01	0.08
	(0.11)	(0.11)	
Cancer (in treatment)	0.08	0.07	0.10
	(0.27)	(0.26)	
Lung Disease (limits activity)	0.04	0.05	0.12
	(0.20)	(0.21)	
Stroke (problems persist)	0.03	0.03	0.11
	(0.17)	(0.18)	
Arthritis (limits activity)	0.26	0.29	0.28
	(0.44)	(0.46)	
Myocardial Infarction	0.03	0.03	0.14
	(0.16)	(0.18)	
Congestive Heart Failure	0.03	0.04	0.14
	(0.18)	(0.20)	
Angina (limits activity)	0.03	0.04	0.13
	(0.18)	(0.20)	
Psychiatric (in treatment)	0.04	0.04	0.11
	(0.19)	(0.19)	
Observations:	30650	19495	

[a] This variable records the number of active health conditions the respondent has; the conditions are: hypertension (not under control), diabetes (not under control), cancer (in treatment), lung disease (limits activity), stroke (problems persist), arthritis (limits activity), myocardial infarction, congestive heart failure, angina (limits activity), and psychiatric disorder (in treatment).

score counts how many of the following tasks the respondent has difficulty performing: making phone calls, handling money, and taking medication.

To measure mental health I use two measures. The first is an indicator that reports if the respondent has ever been diagnosed with a psychiatric or emotional disorder and they are currently receiving treatment for that disorder. I include this measure in the count of "active" health conditions described above. The second measure is the respondent's Center for Epidemiological Studies of Depression (CESD) score, which indicates the presence and intensity of depressive disorders. The CESD score is not, however, intended for determining whether the respondent meets clinical definitions of depressive disorders. Table 3 provides descriptive statistics for these measures.

Econometric Approach

The objective of this chapter is to estimate the role that beneficiary health plays in the relationship between Medicare Part D and hospital utilizations. To achieve this objective I conduct a set of analyses to quantify the relationships between Part D, beneficiary health, and the use of hospital services. In the psychometric literature this approach is known as "mediation analysis." Mediation analysis is used to establish that the causal relationship between two variables flows through a third, mediating, variable. There are four hypotheses involved in establishing mediation, which are depicted in Figure 1.[20] The first hypothesis tests whether Part D has a causal relationship with hospital utilizations. The second hypothesis tests whether Part D has any direct effect (i.e., not through health) on the use of hospital services. This hypothesis can be used to

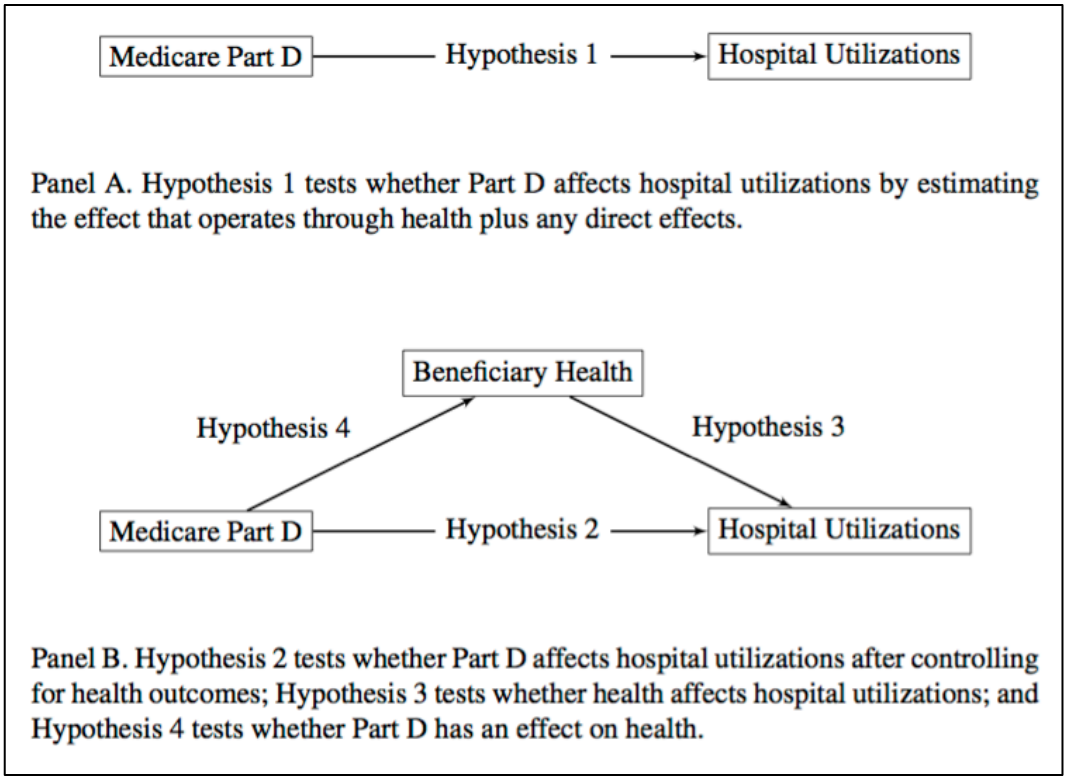
[20] See Baron and Kenny (1986), Judd and Kenny (1981), and James and Brett (1984).

determine whether health completely, or only partially, mediates the causal relationship. The third hypothesis tests whether health has an effect on the use of hospital services. The final hypothesis tests whether Part D has an effect on beneficiary health.

To test these hypotheses I estimate three distinct econometric models. The first model is used to estimate the effect of Part D on hospital utilizations. The second model is used to simultaneously estimate the direct effects of health and Part D on hospital utilizations. The third model is used to estimate the effect of Part D on beneficiary health. I describe each of these models in turn.

Figure 1

The Four Hypotheses Involved in Establishing Mediation



Panel A. Hypothesis 1 tests whether Part D affects hospital utilizations by estimating the effect that operates through health plus any direct effects.

Panel B. Hypothesis 2 tests whether Part D affects hospital utilizations after controlling for health outcomes; Hypothesis 3 tests whether health affects hospital utilizations; and Hypothesis 4 tests whether Part D has an effect on health.

To quantify the causal impact of Medicare Part D on the use of hospital services, I estimate incidence rate ratios (IRRs), which represent the percentage difference in hospital utilizations between Part D participants and abstainers. IRRs are similar to an average treatment effect (ATE), which represents the raw scale difference in hospital utilizations between Part D participants and abstainers. IRRs have the distinct advantage of omitting multiplicatively separable unobserved effects in the same way that ATEs omit additively separable unobserved effects. This is convenient since hospital utilizations are most appropriately modeled non-linearly with multiplicative unobserved effects. IRRs also have the advantage that they measure the effect of Part D in percentage terms so that the magnitude of the effect can easily be assessed.

The effect of Medicare Part D on the use of hospital services is likely different for those beneficiaries who had limited or no pre-Part D drug coverage compared to those beneficiaries who had generous pre-Part D drug coverage. Accordingly, I estimate a separate IRR for these two beneficiary types. Let y_1^a denote the number of inpatient stays (or hospital nights) that would have occurred if the respondent had enrolled in a Part D plan after having had limited or no pre-Part D drug coverage. Similarly, let y_0^a denote the number of inpatient stays (or hospital nights) that would have occurred if the respondent had not enrolled in Part D after having had limited or no pre-Part D drug coverage. Finally, let y_1^b and y_0^b be analogously defined for those respondents who had generous pre-Part D drug coverage. The IRR for those beneficiaries who had limited or no pre-Part D drug coverage can be specified as follows:

$$\text{IRR}_a = \frac{y_1^a}{y_0^a}. \tag{52}$$

If $\text{IRR}_a > 1$, then more inpatient stays (or hospital nights) would have occurred with Part D than without Part D; if $\text{IRR}_a < 1$, then fewer inpatient stays (or hospital nights) would have occurred with Part D than without Part D; and if $\text{IRR}_a = 1$, then the same number of inpatient stays (or hospital nights) would have occurred with Part D as without Part D. The following transformation of the IRR defined in (52) allows the effect of Part D to be interpreted as a percentage difference in inpatient stays (or hospital nights) that would have occurred with and without Part D:

$$\widetilde{\text{IRR}}_a = \left\{ \frac{y_1^a}{y_0^a} - 1 \right\} \cdot 100. \tag{53}$$

I present the transformed IRR defined in (53) since it is relatively easier to interpret. To quantify the impact of Part D for those beneficiaries who had generous pre-Part D drug coverage I estimate the following analog of (53):

$$\widetilde{\text{IRR}}_b = \left\{ \frac{y_1^b}{y_0^b} - 1 \right\} \cdot 100. \tag{54}$$

As described in Terza (2014b) the counterfactuals in (53) and (54) can be replaced with regression predictions when the set of confounders employed is comprehensive. Intuitively, if one can assume that the regression model does not omit

any factors that are simultaneously related to Part D and hospital utilizations, then one can assume that treatment is conditionally exogenous. Let d_{1it} denote a binary variable that indicates if respondent i was enrolled in a Medicare Part D plan at period t, let d_{2it} denote a binary variable that indicates if respondent i had limited or no drug coverage in all periods prior to period t, and let d_{3it} denote the interaction $d_{1it} \times d_{2it}$. The vector $d_{it} \equiv \left[d_{1it} \quad d_{2it} \quad d_{3it} \right]$ encompasses the four distinct "treatment" states implicit in (53) and (54).[21] With a comprehensive vector of confounders, denoted by v_{it}, the following equality holds:[22]

$$y_1^a = E\left[y_{it} \mid d_{it}, v_{it} \right]. \tag{55}$$

That is, with comprehensive confounding the counterfactual outcome, y_1^a, can be replaced with the mean of y_{it} conditional on the treatment state associated with y_1^a. Analogous equalities hold for y_0^a, y_1^b, and y_0^b. Equation (55) is useful because the right-hand side can often be specified, and consistently estimated, via standard regression-based methods. I assume that for all t= 1, 2, ..., T

$$E\left[y_{it} \mid d_{it}, v_{it} = \left[u_i \quad \lambda_t \quad x_{it} \right] \right] = u_i \cdot \exp(\lambda_t + d_{it}\alpha_d + x_{it}\alpha_x), \tag{56}$$

[21] Specifically, beneficiaries who had limited or no pre-Part D drug coverage and enrolled or did not enroll in a Medicare Part D plan, and beneficiaries who had generous pre-Part D drug coverage and enrolled or did not enroll in a Medicare Part D plan.

[22] Equation (55) is not definitional; rather, equation (55) reflects the notion that regression predictions can replace the potential outcome on the left-hand side when the set of confounders is comprehensive. See Terza (2014b) for a more complete derivation of this result.

where u_i denotes an unobserved individual-specific effect, λ_t denotes a period-specific fixed effect, and x_{it} denotes a vector of time-varying regression controls. The vector x_{it} includes age, household income, marital status, a set of variables that indicate the presence of other health insurance contracts, a set of variables that indicate the region of the country, and a set of variables that measure baseline health.

Hausman, Hall, and Griliches (1984) developed a fixed effects Poisson (FEP) estimator and a random effects Poisson (REP) estimator for model (56). The fixed effect assumption is potentially less restrictive since it allows for u_i to be correlated with d_{it} and/or x_{it}. In this case u_i accounts for any potentially omitted factors that are time-invariant, thereby strengthening the argument of comprehensive confounding. Accordingly, the fixed effects assumption lends additional credibility to the argument that the Part D estimates are causally interpretable. However, a fixed effect identification strategy is not robust to the presence of unobserved confounders that are time varying or cannot be conceptualized as a baseline initial condition. Another limitation of the FEP model is that it can only be estimated on the subsample of observations such that $\Sigma_t y_{it} \neq 0$, however. This restriction on the estimation sample could induce a selectivity bias that would otherwise not be present. Furthermore, the choice between the fixed effects version of model (56) and a random effects version is empirically testable. Therefore, I estimate model (56) using the FEP estimator and two random effects Poisson (REP) estimators. The first REP estimator is applied to the fixed effects sample to test whether the random effects assumption is appropriate. The second REP estimator is

applied to the unrestricted sample to test whether restricting the sample induces selection bias. In the event that the random effects assumption is rejected, this second test is no longer valid, but remains potentially informative. Testing is based on generalized Hausman tests of the equality of the coefficients across each specification. Additional detail regarding these testing procedures and the variance-covariance matrices required to calculate the test statistics is provided in Appendix C.

Using (55) and (56) equation (53) can be estimated as follows:

$$\widehat{\text{IRR}}_{\text{a}} = \left\{ \frac{u_i \cdot \exp(\lambda_t + \alpha_{d1} + \alpha_{d2} + \alpha_{d3} + x_{it}\alpha_x)}{u_i \cdot \exp(\lambda_t + \alpha_{d2} + x_{it}\alpha_x)} - 1 \right\} \cdot 100 . \tag{57}$$

which simplifies to

$$\widehat{\text{IRR}}_{\text{a}} = \left\{ \exp(\alpha_{d1} + \alpha_{d3}) - 1 \right\} \cdot 100 . \tag{58}$$

Likewise, after simplification, the following is a consistent estimator for (58):

$$\widehat{\text{IRR}}_{\text{b}} = \left\{ \exp(\alpha_{d1}) - 1 \right\} \cdot 100 \tag{59}$$

The asymptotic variances for the IRR estimators in (58) and (59) can be obtained via the general method developed in Terza (2014c). The generic approach laid out therein is equivalent to the delta method approximation in this case, however. I use "nlcom," a post-estimation command available in Stata 11, to calculate (58) and (59), and I conduct inference based on delta method standard errors.

Hypotheses two and three can be tested in the context of a single econometric model that simultaneously controls for Part D and improvements in health. Then one can estimate the effect that health has on the use of hospital services in such a way that health is not serving as a proxy for Part D. One can furthermore estimate the direct effect that Part D has on hospital utilizations. That is, the effect of Part D net of any effect that operates through changes in health. The model that I employ assumes, as in hypothesis one, that for all t = 1, 2, .., T

$$E\left[y_{it} \mid d_{it}, u_i, \lambda_t, x_{it}, z_{it} \right] = u_i \cdot \exp(\lambda_t + d_{it}\beta_d + x_{it}\beta_x + z_{it}\beta_z), \tag{60}$$

where z_{it} denotes a vector of health factors for respondent i measured at the end of period t (or at follow-up). End-of-period health characteristics represent a set of health outcomes, which could be influenced by Part D and which could explain the number of inpatient stays or hospital nights that occur during period t.[23] The vector z_{it} includes ADL, IADL, CESD, and a count of the number of active health conditions. See Table 3 for a list of the active health conditions included in this count. Since beginning-of-period health characteristics cannot be influenced by Part D these can be included as controls in model (56).

The direct effect of Part D can be quantified by estimating the same IRRs as defined in (58) and (59). If these effect measures are zero, then health completely mediates the relationship between Part D and the use of hospital services. If these effects

[23] The variables included in z_{it} derive from a retrospective questionnaire. Thus the end-of-period health characteristics quantify the health of respondents over the previous two-year period.

are non-zero, then health only partially mediates. Finally, the indirect effect of Part D (i.e., the effect of Part D that operates through its effect on health) can be recovered as the difference between the total and direct effects of Part D. To test whether the indirect effects are statistically different from zero I use generalized Hausman tests (see Appendix C).

The effect of health is not easily quantifiable from model (60) since health enters multi-dimensionally. Hypothesis three can be tested, however, by testing whether the set of end-of-period health measures are jointly significant. If the set of end-of-period health characteristics are jointly significant, then it can be concluded that health has an effect on the use of hospital services.

Hypothesis four involves estimating the relationship between Part D and beneficiary health. Since the primary variable that I use to control for health outcomes in hypotheses two and three is a count of the number of active health conditions, I use this variable as the outcome in a third panel Poisson analysis. Specifically, I assume that for all t= 1, 2, .., T

$$E\left[h_{it} \mid d_{it}, u_i, \lambda_t, x_{it}\right] = u_i \cdot \exp(\lambda_t + d_{it}\gamma_d + x_{it}\gamma_x), \tag{61}$$

where h_{it} denotes the count of active health conditions. To quantify the effect of Part D on health I again calculate IRRs. As before, model (61) allows the effects of Part D to differ across beneficiaries who had limited or no pre-Part D drug coverage and beneficiaries who had generous pre-Part D drug coverage. I conduct the same testing procedures across FEP and two REP specifications as with previous hypotheses.

Results

In 2006, 57% of the sample had Part D drug coverage. An additional 7 percentage of the sample gained Part D coverage in 2008. These enrollment rates are higher than those reported by Levy and Weir (2009), who reported a 24% enrollment rate in stand-alone Part D plans. This discrepancy is due to the fact that I have a different sample and because I include those beneficiaries with Medicaid or Medicare Advantage drug coverage in my Part D measure. After 2006, "dual eligibles" (i.e., Medicare and Medicaid enrollees) were automatically enrolled in Part D plans. Also, if beneficiaries are Medicare Advantage plan holders they cannot get their drug benefits through a stand-alone Part D plan, but can get comparable drug benefits through a Medicare Advantage prescription drug plan. In 2006, 60% of the sample had limited or no pre-Part D drug coverage. This is more than double the uninsured rate reported by Safran et al. (2005), but the majority of this group is composed of respondents that had limited pre-Part D drug coverage versus no coverage. In 2006, 65% of those respondents that enrolled in a Part D plan had limited or no pre-Part D drug coverage.

Table 2 (above) reports descriptive statistics for the number of inpatient stays, the number of hospital nights, and the demographic characteristics of the analysis sample. The first column presents means and standard deviations for the full sample. The second column presents means and standard deviations for the sample used in the FEP analyses. The third column presents the within standard deviations for this latter sample. Since the FEP specifications exclude respondents who never had any inpatient stays or nights, the

mean number of inpatient stays and hospital nights are larger in the FE sample. All of the other demographic characteristics are remarkably similar across the two samples.

Table 3 (above) reports descriptive statistics for the health measures. The table organization mirrors that of Table 2. The mean ADL and IADL scores are very similar across the full and FE samples. The mean CESD score and the mean number of active conditions are larger in the FE sample. This is not surprising since it is expected that health affects the use of hospital services. Thus, after omitting respondents who never were hospitalized, it is expected that the remaining respondents are the least healthy among the full sample. Also, it appears that differences in the proportion of respondents with active arthritis diagnoses explains a large portion of the difference in the mean number of active conditions. Nagamine, Jiang, and Merrill (2006) reported that osteoarthritis is included in the top ten conditions that explain 40% of the annual number of hospitalizations among the elderly. Chronic obstructive pulmonary disease, stroke, myocardial infarction, and congestive heart failure are also on that list. There are also slightly more respondents in the FE sample with lung disease, and congestive heart failure.

Table 4 presents the inpatient stay results associated with testing hypotheses one, two, and three. The table reports raw coefficients, which indicate the direction of the effect but not the magnitude. To assess the magnitude of the effect I calculate transformed IRRs (i.e., $\{\text{IRR}-1\}\cdot 100$). These transformed IRRs are interpreted as the

Table 4

The Effects of Part D on Inpatient Stays

	Total Effects of Part D[a]			Direct Effects[b]
	FEP	REP[c]	REP[d]	FEP
Limited Pre-Part D Coverage	-0.14**	-0.06	-0.05	-0.14
	(0.06)	(0.05)	(0.05)	(0.09)
	[-13.02]	[-5.53]	[-4.96]	[-12.76]
Generous Pre-Part D Coverage	0.03	0.01	0.05	0.04
	(0.07)	(0.05)	(0.06)	(0.09)
	[2.53]	[1.03]	[5.22]	[3.93]
ADL (Follow-Up)				0.05*
				(0.03)
IADL (Follow-Up)				0.14**
				(0.06)
CESD (Follow-Up)				0.02*
				(0.01)
Active Conditions (Follow-Up)				0.23***
				(0.03)
Observations:	19495	19495	30650	11689
FEP v. REP:[e]		221.59	373.10	
p value:		{0.00}	{0.00}	
REP v. REP:[f]			324.11	
p value:			{0.00}	
Limited Pre-Part D Coverage:[g]				0.00
p value:				{0.97}
Generous Pre-Part D Coverage:[h]				0.04
p value:				{0.84}
Health (Follow-Up):[i]				97.61
p value:				{0.00}

Cluster-robust standard errors are in parentheses

Transformed incidence rate ratios (i.e., {IRR-1} × 100) in square brackets; the transformed IRRs represent the percentage impact of the covariate

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[a] The total effects of Part D measure the effects of Part D that operate through changes in beneficiary health plus any direct effects.

[b] The direct effects of Part D measure the effects of Part D that do not operate through changes in beneficiary health.

[c] Column reports results from a random effects Poisson estimator applied to the fixed effects sample.

[d] Column reports results from a random effects Poisson estimator applied to the full sample.

[e] Tests the equality of the common coefficients across the FEP and REP models.

[f] Tests the equality of the coefficients across the two REP models.

[g] Tests the equality of the Part D effect among beneficiaries who had limited or no pre-Part D drug coverage.

[h] Tests the equality of the Part D effect among beneficiaries who had generous pre-Part D drug coverage.

[i] Tests the joint significance of ADL (Follow-Up), IADL (Follow-Up), CESD (Follow-Up), and Active Conditions (Follow-Up).

percentage impact of Part D on the number of inpatient stays. For the Part D effects, the table also reports these transformed IRRs in square brackets. Table C1 presents all model coefficients and can be found in Appendix C. The FEP results indicate that Part D reduced the number of inpatient stays by 13.02% [95% CI: -23.59%, -2.44%] among beneficiaries who had limited or no pre-Part D drug coverage. Among beneficiaries who had generous pre-Part D drug coverage, the FEP results indicate a small, positive, and statistically insignificant effect of 2.53% [95% CI: -11.81%, 16.87%]. The second and third columns of Table 4 present results from REP specifications applied to the FEP and full samples, respectively. The results from the REP specification applied to the FEP sample indicate that Part D reduced the number of inpatient says by 5.53% [95% CI: -14.29%, 3.22%] among beneficiaries who had limited or no pre-Part D drug coverage, and that Part D increased the number of inpatient stays by a statistically insignificant 1.03% [95% CI: -9.57%, 11.63%]. The results from the REP specification applied to the full sample indicates that Part D reduced the number of inpatients stays by 4.96% [95% CI: -14.29%, 4.37%] among beneficiaries who had limited or no pre-Part D drug coverage, and that Part D increased the number of inpatient stays by 5.22% [95% CI: -6.57%, 17.19%] among beneficiaries who had generous pre-Part D drug coverage. Thus it appears that Part D reduced the number of inpatient stays among beneficiaries who had limited or no pre-Part D drug coverage, but had little to no impact among beneficiaries who had generous pre-Part D drug coverage.

To assess the validity of the RE assumption, and to assess the extent to which the FE sample drives the results associated with the FE specification, I conduct generalized

Hausman tests of the equality of the coefficients across each of the specifications reported in columns one through three of Table 4. The results are summarized at the bottom of the table. Comparing the FE specification with the RE specification applied to the FE sample, the evidence indicates that the RE assumption is likely violated (χ^2 = 221.59, p = 0.00). Comparing the two RE specifications, the evidence indicates that the coefficients in each of the RE specifications are statistically significantly different from each other (χ^2 = 324.11, p = 0.00). This latter result lends informal evidence that the FE results are partially a reflection of the estimation sample used. It is plausible, however, that the FE specification represents a consistent approach to estimating the effect of Part D among patients that reported at least one hospitalization. For this reason, and since the generalized Hausman test indicates that the RE assumption is inappropriate, the FE results are preferred.

The final column in Table 4 presents FE results associated with testing hypotheses two and three. I only present a FE specification since the FE specification was preferred above. Hypothesis two involves testing whether Part D has a direct effect on the number of inpatient stays. The results indicate that Part D has a direct effect of -12.76% [95% CI: -27.31%, 1.78%] among beneficiaries who had limited or no pre-Part D drug coverage and a direct effect of 3.93% [95% CI: -15.19%, 23.06%] among beneficiaries who had generous pre-Part D drug coverage. This implies a very small indirect effect among both beneficiary types, where the indirect effect is given by the total effect reported in the first column minus the direct effect reported in the fourth column. This also suggests that health does not explain much of the effect.

Hypothesis three involves testing whether health outcomes have a direct effect on the number of inpatient stays. Health outcomes are measured with a set of follow-up health characteristics, which were recorded at the end of each period. These factors are jointly significant (χ^2 = 97.61, p = 0.00). Additionally, each follow-up health factor has a positive and statistically significant effect on the number of inpatient stays individually. The largest effect is associated with the number of active health conditions at follow-up. For each additional active health condition, the number of inpatient stays increases by 26.36% [95% CI: 19.49%, 33.23%]. This would translate into a very large effect at high levels of active health conditions. It is worth noting, however, that the average number of active health conditions is low (0.63) and that only one respondent ever had a maximum of 8 out of the 10 conditions.

Table 5 presents the hospital night results associated with testing hypotheses one through three. The results again only indicate the direction of the effect. To assess the magnitude of the effect I calculate transformed IRRs (i.e., $\{\text{IRR}-1\}\cdot 100$), which are interpreted as the percentage impact of enrolling in a Part D insurance plan on the number of hospital nights. All model coefficients are reported in Table C2 in Appendix C. The FEP results indicate that Part D reduced the number of hospital nights by 16.52% [95% CI: -34.06%, 1.01%] among beneficiaries who had limited or no pre-Part D drug coverage. Among beneficiaries who had generous pre-Part D drug coverage, the FEP results indicate that Part D reduced the number of hospital nights by 21.91% [95% CI: -41.28%, -2.53%]. As before, the second and third columns present results from REP specifications applied to the FE and full samples, respectively. The results from the REP

Table 5

The Effects of Part D on Hospital Nights

	Total Effects of Part D[a]			Direct Effects[b]
	FEP	REP[c]	REP[d]	FEP
Limited Pre-Part D Coverage	-0.18*	-0.16	-0.17	-0.14
	(0.11)	(0.10)	(0.10)	(0.16)
	[-16.52]	[-14.58]	[-15.31]	[-12.87]
Generous Pre-Part D Coverage	-0.25*	-0.20*	-0.21*	-0.15
	(0.13)	(0.12)	(0.12)	(0.16)
	[-21.91]	[-18.53]	[-18.84]	[-13.55]
ADL (Follow-Up)				0.14***
				(0.05)
IADL (Follow-Up)				-0.14
				(0.14)
CESD (Follow-Up)				0.06***
				(0.02)
Active Conditions (Follow-Up)				0.23***
				(0.05)
Observations:	19495	19495	30650	11679
FEP v. REP:[e]		218.42	208.83	
p value:		{0.00}	{0.00}	
REP v. REP:[f]			103.43	
p value:			{0.00}	
Limited Pre-Part D Coverage:[g]				0.11
p value:				{0.73}
Generous Pre-Part D Coverage:[h]				0.67
p value:				{0.41}
Health (Follow-Up):[i]				50.46
p value:				{0.00}

Cluster-robust standard errors are in parentheses

Transformed incidence rate ratios (i.e., {IRR-1} × 100) in square brackets; the transformed IRRs represent the percentage impact of the covariate

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[a] The total effects of Part D measure the effects of Part D that operate through changes in beneficiary health plus any direct effects.

[b] The direct effects of Part D measure the effects of Part D that do not operate through changes in beneficiary health.

[c] Column reports results from a random effects Poisson estimator applied to the fixed effects sample.

[d] Column reports results from a random effects Poisson estimator applied to the full sample.

[e] Tests the equality of the common coefficients across the FEP and REP models.

[f] Tests the equality of the coefficients across the two REP models.

[g] Tests the equality of the Part D effect among beneficiaries who had limited or no pre-Part D drug coverage.

[h] Tests the equality of the Part D effect among beneficiaries who had generous pre-Part D drug coverage.

[i] Tests the joint significance of ADL (Follow-Up), IADL (Follow-Up), CESD (Follow-Up), and Active Conditions (Follow-Up).

specification applied to the FEP sample indicates that Part D reduced the number of hospital nights by 14.58% [95% CI: -31.81%, 2.64%] among beneficiaries who had limited or no pre-Part D drug coverage and reduced the number of hospital nights by 18.53% [95% CI: -37.74%, 0.68%] among beneficiaries who had generous pre-Part D drug coverage. The results from the REP specification applied to the full sample are remarkably similar to both the FEP results and the results from the REP specification applied to the FEP sample. Accordingly, the evidence consistently indicates that Part D has had a large and negative effect on the number of hospital nights.

I conduct the same generalized Hausman tests as described above. The evidence is summarized at the bottom of Table 5 and indicates that the RE assumption may be violated, and that the FE results may be partially driven by the FE estimation sample. Therefore, by the same argument as before, the FE results are preferred. This preference is further supported by statistical testing of the equality of the Part D effects associated with each specification. In particular, test results (not reported) indicate that the estimated IRRs from the RE specifications are not statistically different from each other, while the IRRs from the FE specification is statistically different from the RE specifications.

The final column of Table 5 presents the hospital night results associated with testing hypotheses two and three. Hypothesis two involves testing whether Part D has a direct effect on the number of hospital nights. The results indicate that, after controlling for health outcomes, Part D reduced the number of hospital nights by 12.87% [95% CI: -39.54%, 13.81%] among beneficiaries who had limited or no pre-Part D drug coverage

and decreased the number of hospital nights by 13.55% [95% CI: -40.68%, 13.57%] among beneficiaries who had generous pre-Part D drug coverage. As before, this implies a small (but larger) negative indirect effect (i.e., the total effect reported in the first column minus the direct effect reported in the fourth column). The change in coefficients from the first column to the fourth column is larger among beneficiaries who had generous pre-Part D drug coverage. The indirect effect implied by the difference in effects is not statistically significant at any conventional level, however.

Hypothesis three involves testing whether health outcomes have a direct effect on the number of hospital nights. Health outcomes are measured by a set of health factors recorded at end of each period. These factors are jointly significant (χ^2 = 50.46, p = 0.00). Again the number of active conditions at follow-up has a large, positive, and statistically significant effect on the number of hospital nights. In terms of transformed IRRs, the results indicate that each additional active health condition increases the number of hospital nights by 26.12% [95% CI: 14.14%, 38.10%]. ADL and CESD also have large effects. The effect of IADL is negative and approximately counterbalances the effect of ADL on the number of hospital nights. This may reflect a high level of correlation between ADL and IADL, which prevents uniquely identifying the separate effects of each functional limitation measure.

Table 6 presents the results associated with testing hypothesis four. Table C3 included in Appendix C presents all model coefficients. Hypothesis four involves testing whether Part D has any effect on health. Since the primary health outcome used in testing hypotheses two and three is a variable that counts the number of active health

conditions, I analyze this outcome with a third set of panel Poisson specifications. After calculating transformed IRRs (i.e., $\{IRR-1\}\cdot 100$), the FEP results indicate that Part D reduced the number of active health conditions by 8.46% [95% CI: -17.74%, 0.82%] among beneficiaries who had limited or no pre-Part D drug coverage. Among beneficiaries who had generous pre-Part D drug coverage, the FEP results indicate that Part D reduced the number of active health conditions by 9.00% [95% CI: -17.42%, -0.57%]. The second and third columns present REP specifications applied to the FEP and full samples, respectively. Both REP specifications indicate a statistically insignificant and small negative effect among beneficiaries who had limited or no pre-Part D drug coverage. Among beneficiaries who had generous pre-Part D drug coverage, the REP specification applied to the FEP sample estimates a small, positive, and statistically insignificant effect. The REP specification applied to the full sample indicates that Part D increased the number of active health conditions by 10.84% [95% CI: 1.38%, 20.29%] among these beneficiaries.

As with the hospital utilization models I test the differences between the FEP and REP specifications. The generalized Hausman testing results indicate that the REP assumption may be violated (χ^2 = 497.01, p= 0.00). The REP specifications are also statistically different from each other. Since the REP assumption may be violated, and because the FEP results likely reflect consistent results for the sample of beneficiaries analyzed in hypotheses one through three, the FEP results are preferred.

Table 6

The Effect of Part D on Active Health Conditions[a]

	FEP	REP[b]	REP[c]
Limited Pre-Part D Coverage	-0.09*	-0.04	-0.01
	(0.05)	(0.04)	(0.04)
	[-8.46]	[-4.36]	[-0.71]
Generous Pre-Part D Coverage	-0.09**	0.02	0.10**
	(0.05)	(0.04)	(0.04)
	[-9.00]	[1.63]	[10.84]
Observations:	11609	11609	20961
FEP v. REP:[d]		497.01	611.33
p value:		{0.00}	{0.00}
REP v. REP:[e]			147.46
p value:			{0.00}

Cluster-robust standard errors are in parentheses

Transformed incidence rate ratios (i.e., {IRR-1} × 100) in square brackets; the transformed IRRs represent the percentage impact of the covariate

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[a] Active health conditions include: hypertension (not under control), diabetes (not under control), cancer (in treatment), lung disease (limits activity), stroke (problems persist), arthritis (limits activity), myocardial infarction, congestive heart failure, angina (limits activity), and psychiatric disorder (in treatment).

[b] Column reports results from a random effects Poisson applied to the fixed effects sample.

[c] Column reports results from a random effects Poisson applied to the full sample.

[d] Tests the equality of the coefficients in the FEP and REP models.

[e] Tests the equality of the coefficients in the two REP models.

One important limitation of this chapter is that the identification strategy is a fixed effects strategy. This means that the econometric approach is not robust to the potential that there exist some unobserved factors that vary across time and explain the beneficiary's decision to enroll in Part D plans as well as variation in their hospital utilization outcomes. For instance, risk preferences are not controlled for in any of the models presented above. If risk preferences cannot be modeled as a stationary unobserved factor, then fixed effects will fail to control for the possibility that unobserved risk preferences may influence the Part D enrollment decision and the number of hospitalizations. In this scenario an instrumental variables approach could

provide a better estimation strategy. Early in the research process I tested a number of possible instruments, but failed to find anything that passed weak instrument tests. One difficulty I encountered that hampered my ability to find strong instruments is that the public use files do not contain state identifiers. There are a number of possible instruments that could be constructed on the basis of variation in the implementation of Medicare Part D across states, as well as variation across states that derives from interactions between Medicare Part D and state-specific implementations of the Medicaid system. In the future I hope to gain access to restricted data files that will allow me to pursue identification strategies based on such instruments. Despite this limitation, this chapter provides a reasonable "first-pass" approach that models unobserved factors such as risk preferences as baseline conditions that do not change over time.

All of the models discussed above control for age and period fixed effects. It appears that these measures are highly collinear, and that the degree of collinearity makes it difficult to consistently and separately identify the effects of age and period. This is most apparent in the first column of Table 4, where the period effects indicate that hospital nights decrease over time. This is a counterintuitive result that disappears after dropping the age variable. Furthermore, the effect of age on hospital nights is larger than expected. In terms of a transformed IRR, the FEP model estimates that each additional year of age increases the number of hospital nights by 26%. This implies that hospital nights double every 3.8 years, which seems unlikely. None of the results reported above are substantively different if age is dropped from the analysis. It may be the case that age

is an unnecessary control in the FEP models since the unobserved individual fixed effect may control for the effect of time-invariant baseline age effects.

The period effects also appear to be inflated even after dropping age, however. Further investigation revealed that the FEP models estimate larger period effects when comparing more distant periods. With the base period set to 2004-2006, and after dropping age, the last two period effects are both less than or equal to the average within-person change in inpatient stays or nights. The estimated difference in inpatient stays or nights between the 1996-1998 period and the 2004-2006 period are slightly larger than the average within-person change, however, suggesting that these period effects are slightly inflated. This may reflect the fact that the panel is unbalanced and fewer respondents are available to make comparisons across more distant periods. Restricting the analysis to the last three waves of data produces slightly larger (in absolute value) Part D effects with similar post-Part D period effects as the model applied on the full panel.

In the FEP hospital night model, the region fixed effects also appear to be inflated. The FEP model estimates that respondents in the Midwest have 63% fewer hospital nights than respondents in the West. Though it is plausible that there is variability in hospital utilizations across regions of the country, this degree of variability seems unlikely. This inflation is likely due to the fact that there is almost no within variation in the region indicators. All of the results reported above remain after dropping the region variables. Furthermore, the Part D and prior drug coverage indicators have much higher within variation and so likely would not be inflated for this reason. In fact,

the within variation can be arbitrarily increased for the Part D indicator by omitting all but one of the pre-Part D waves and re-estimating the models with only data from 2004-2010. Since the results are slightly larger (in absolute value) rather than smaller, this suggests that the within variability is large enough to get consistent estimates from the FEP estimator.

In spite of these potential concerns the effects of Part D estimated in this chapter appear to be reliable. In particular, none of the robustness checks described above affected the main results. Furthermore, the effect of Part D on the number of inpatient stays is very similar to results reported in Hsu et al. (2006) and in Kaestner et al. (2014). Hsu et al. (2006) report an effect of 13%, which is identical to my estimated effect of 13% among beneficiaries who had limited or no pre-Part D drug coverage. My estimated effect is larger than the 8% effect estimated by Kaestner et al. (2014), but they do not differentiate between beneficiaries who had limited or no pre-Part D drug coverage and beneficiaries who had generous pre-Part D drug coverage.

Discussion

On the extensive margin (i.e., inpatient stays) the evidence indicates that Part D had a large negative effect among beneficiaries who had limited or no pre-Part D drug coverage and that improvements in health do not explain any of this relationship. These results could be explained by the fact that Medicare Part A covers inpatient prescription drug use. To see this, let "ex ante drug use" refer to medication use before an adverse health event occurs, and which may prevent its occurrence. Similarly, let "ex post drug use" refer to medication use after an adverse health outcome has occurred, and which

may have been avoided with ex ante drug use. It is possible that beneficiaries who had limited drug coverage prior to the implementation of Part D substituted toward relatively less costly ex post drug use, but after the implementation of Part D substituted toward relatively less costly ex ante drug use. Assuming that both types of drug use are productive in terms of improving patient health, then the health produced by ex ante drug use may only be slightly better than the health produced by ex post drug use. Thus we might observe a negative relationship between Part D and the use of hospital services that is not explained by the improvement in beneficiary health. This remains an empirical question, however. More research should be conducted to investigate this, and other hypothesized, mechanisms. Among beneficiaries who had generous pre-Part D drug coverage the evidence indicates that Part D had almost no effect on the number of hospital stays.

Among both beneficiary types the evidence indicates that Part D had a stronger impact on the intensive margin (i.e., hospital nights) than on the extensive margin. This suggests that Part D is more effective at reducing the intensity with which illnesses need to be treated than at reducing the occurrence of illness. The Part D effects slightly change after controlling for health outcomes, but are statistically imprecise. Thus there is no statistical evidence that health explains the relationship between Part D and the use of hospital nights. The relatively small indirect effects are in line with the modest effect that Part D had on the number of active health conditions.

This chapter investigated the effect of Part D on the use of inpatient hospital services. More research should be conducted to identify the relationship between Part D

and other health care services, and the extent to which changes in beneficiary health explain these relationships. For instance, it is possible that Part D has a different relationship with the use of outpatient and/or emergency department services. Furthermore, since physicians act as gatekeepers for the use of pharmaceuticals, Part D could have a positive effect on the use of physician services. On the other hand, the effect of Part D on health could imply a negative relationship between Part D and the use of physician services. A formal mediation analysis could investigate whether there is a negative indirect effect that operates through the effect of Part D on health, and whether there is a positive direct effect that operates through the use of physician's gatekeeper services.

The evidence supports the conclusion that cost-offsets are likely and fairly large. Pfuntner, Weir, and Steiner (2013) estimated that the average cost per hospital stay for an individual between the ages of 65 and 84 to be about $12,300. Using this figure the estimated effect of Part D on the number of hospital stays implies a reduction in hospital stay expenditures of about $1,600 for every $12,300 that would have otherwise been spent in the absence of Part D. The Kaiser Family Foundation (KFF) reports that the average cost of a hospital night could be as much as $1,600 in 2006 (see http://kff.org/other/state-ind1icator/expenses-per-inpatient-day-by-ownership/). Using this figure the estimated impacts of Part D on the number of hospital nights implies a reduction in hospital night expenditures between $240 and $320 for every $1,600 that would have otherwise been spent.

Cost-offset estimates may be useful in assessing the costs and benefits of Medicare Part D, as well as conducting more broadly defined economic evaluations. The evidence produced in this chapter suggests that a majority of the cost-offset effect is unrelated to health, however. Accordingly, cost-offset estimates may be useful in tabulating the costs of the Medicare Part D program, but should not be used in tabulating the monetary benefits associated with the program without further investigations into the role that health plays and proper accounting of the proportion of the cost-offset effect that can be attributed to improvements in health. This chapter does indicate a high degree of substitutability between the use of prescription medications and hospital services. Future research should attempt to determine whether these substitution effects can be leveraged to increase the cost-effectiveness of the Part D program.

CHAPTER V

CONCLUSION

This dissertation has explored three important issues related to econometrics and economics of health care. The second and third chapters explored issues that relate to appropriate methodology required for estimating health care expenditure/utilization models. The fourth chapter explored the effect of Medicare Part D on the use of hospital services. The knowledge gained in this dissertation research is both scientifically rigorous and currently relevant. The methodology contributions have many future possible applications. The applied contribution in chapter IV provides a strong foundation for a number of additional research topics.

REFERENCES

Abrevaya, J. (2002). Computing Marginal Effects in the Box-Cox Model. *Econometric Reviews, 21*(3), 383-393. doi: 10.1081/etc-120015789

Afendulis, C. C., He, Y., Zaslavsky, A. M., & Chernew, M. E. (2011). The Impact of Medicare Part D on Hospitalization Rates. *Health Services Research, 46*(4), 1022-1038. doi: 10.1111/j.1475-6773.2011.01244.x

Ai, C., & Norton, E. C. (2000). Standard errors for the retransformation problem with heteroscedasticity. *Journal of Health Economics, 19*(5), 697-718. doi: 10.1016/s0167-6296(00)00046-1

Ai, C., & Norton, E. C. (2008). A semiparametric derivative estimator in log transformation models. *Econometrics Journal, 11*(3), 538-553. doi: 10.1111/j.1368-423X.2008.00252.x

Andreyeva, T., Sturm, R., & Ringel, J. S. (2004). Moderate and Severe Obesity Have Large Differences in Health Care Costs. *Obesity Research, 12*(12), 1936-1943. doi: 10.1038/oby.2004.243

Angrist, J. D. (2001). Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice. *Journal of Business & Economic Statistics, 19*(1), 2-16.

Barber, J., & Thompson, S. (2004). Multiple regression of cost data: use of generalised linear models. *Journal of Health Services Research & Policy, 9*(4), 197-204. doi: 10.1258/1355819042250249

Baron, R. M., & Kenny, D. A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.

Basu, A., & Manning, W. G. (2009). Issues for the Next Generation of Health Care Cost Analyses. *Medical Care, 47*(7_Supplement_1), S109-S114 110.1097/MLR.1090b1013e31819c31894a31811.

Basu, A., Yin, W., & Alexander, G. C. (2010). Impact of Medicare Part D on Medicare–Medicaid Dual-Eligible Beneficiaries' Prescription Utilization and Expenditures. *Health Services Research, 45*(1), 133-151. doi: 10.1111/j.1475-6773.2009.01065.x

Bell, J. F., Zimmerman, F. J., Arterburn, D. E., & Maciejewski, M. L. (2011). Health-Care Expenditures of Overweight and Obese Males and Females in the Medical Expenditures Panel Survey by Age Cohort. *Obesity, 19*(1), 228-232.

Biro, A. (2013). Supplementary private health insurance and health care utilization of people aged 50+. *Empirical Economics*.

Blough, D. K., Madden, C. W., & Hornbrook, M. C. (1999). Modeling risk using generalized linear models. *Journal of Health Economics, 18*(2), 153-171. doi: 10.1016/s0167-6296(98)00032-0

Briesacher, B. A., Stuart, B., Ren, X., Doshi, J. A., & Wrobel, M. V. (2005). Medicare Beneficiaries and the Impact of Gaining Prescription Drug Coverage on Inpatient and Physician Spending. *Health Services Research, 40*(5p1), 1279-1296. doi: 10.1111/j.1475-6773.2005.00432.x

Briesacher, B. A., Zhao, Y., Madden, J. M., Zhang, F., Adams, A. S., Tjia, J., . . . Soumerai, S. B. (2011). Medicare Part D and Changes in Prescription Drug Use and Cost Burden. *Medical Care, 49*(9), 834-841.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications* (1st Ed. ed.). New York, New York: Cambridge University Press.

Cawley, J., & Meyerhoefer, C. (2012). The medical care costs of obesity: An instrumental variables approach. *Journal of Health Economics, 31*, 219-230.

CBO. (2012). Offsetting Effects of Prescription Drug Use on Medicare's Spending for Medical Services.

Chandra, A., Gruber, J., & McKnight, R. (2010). Patient Cost-Sharing and Hospitalization Offsets in the Elderly. *The American Economic Review, 100*(1), 193-213. doi: 10.2307/27804926

Cole, J. A., Norman, H., Weatherby, L. B., & Walker, A. M. (2006). Drug Copayment and Adherence in Chronic Heart Failure: Effect on Cost and Outcomes. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy, 26*(8), 1157-1164.

Cragg, J. G. (1971). Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. *Econometrica, 39*(5), 829-844.

Deb, P., Munkin, M. K., & Trivedi, P. K. (2006). Bayesian analysis of the two-part model with endogeneity: application to health care expenditure. *Journal of Applied Econometrics, 21*(7), 1081-1099. doi: 10.1002/jae.891

Deb, P., Trivedi, P. K., & Zimmer, D. M. (2009). *Dynamic Cost-Offsets of Prescription Drug Expenditures: Panel Data Analysis Using a Copula-Based Hurdle Model*. NBER Working Paper Series, (Working Paper 15191). NBER.

Dow, W., & Norton, E. (2003). Choosing Between and Interpreting the Heckit and Two-Part Models for Corner Solutions. *Health Services and Outcomes Research Methodology, 4*(1), 5-18. doi: 10.1023/a:1025827426320

Duan, N. (1983). Smearing Estimate: A Nonparametric Retransformation Method. *Journal of the American Statistical Association, 78*(383), 605-610.

Duan, N., Manning, W. G., Morris, C. N., & Newhouse, J. P. (1983). A Comparison of Alternative Models for the Demand for Medical Care. *Journal of Business & Economic Statistics, 1*(2), 115-126.

Duan, N., Manning, W. G., Morris, C. N., & Newhouse, J. P. (1984). Choosing between the Sample-Selection Model and the Multi-Part Model. *Journal of Business & Economic Statistics, 2*(3), 283-289.

Duan, N., Manning, W. G., Morris, C. N., & Newhouse, J. P. (1985). Comments on Selectivity Bias. *Advances in Health Economics and Health Services Research, 6*, 19-24.

Duggan, M., & Morton, F. S. (2010). The Effect of Medicare Part D on Pharmaceutical Prices and Utilization. *The American Economic Review, 100*(1), 590-607. doi: 10.2307/27804943

Finkelstein, E. A., Trogdon, J. G., Cohen, J. W., & Dietz, W. (2009). Annual Medical Spending Attributable to Obesity: Payer-and-Service-Specific Estimates. *Health Affairs, 28*(5), w822-w831.

Goldman, D., & Philipson, T. J. (2007). Integrated Insurance Design in the Presence of Multiple Medical Technologies. *The American Economic Review, 97*(2), 427-432.

Goldman, D. P., Joyce, G. F., Escarce, J. J., Pace, J. E., Solomon, M. D., Laouri, M., . . . Teutsch, S. M. (2004). Pharmacy benefits and the use of drugs by the chronically ill. *Journal of the American Medical Association, 291*(19), 2344-2350. doi: 10.1001/jama.291.19.2344

Greene, W. H. (2012). *Econometric Analysis* (Seventh Edition ed.). Upper Saddle, NJ: Prentice Hall.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy, 80*(2), 223-255. doi: 10.2307/1830580

Hausman, J., Hall, B. H., & Griliches, Z. (1984). Econometric Models for Count Data with an Application to the Patents-R & D Relationship. *Econometrica, 52*(4), 909-938.

Hay, J. W., Leu, R., & Rohrer, P. (1987). Ordinary Least Squares and Sample-Selection Models of Health-Care Demand: Monte Carlo Comparison. *Journal of Business & Economic Statistics, 5*(4), 499-506.

Hay, J. W., & Olsen, R. J. (1984). Let Them Eat Cake: A Note on Comparing Alternative Models of the Demand for Medical Care. *Journal of Business & Economic Statistics, 2*(3), 279-282.

Health and Retirement Study (HRS), AHEAD 1995 public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI, (1995).

Health and Retirement Study (HRS), HRS 1996-2010 public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI, (1996-2010).

Heckman, J. J. (1976). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Annals of Economic and Social Measurement, 5*(4), 120-137.

Heckman, J. J. (1978). Dummy Endogenous Variables in a Simultaneous Equation System. *Econometrica, 46*(4), 931-959.

Heisler, M., Langa, K. M., Eby, E. L., Fendrick, A. M., Kabeto, M. U., & Piette, J. D. (2004). The Health Effects of Restricting Prescription Medication Use Because of Cost. *Medical Care, 42*(7), 626-634.

Hsu, J., Price, M., Huang, J., Brand, R., Fung, V., Hui, R., . . . Selby, J. V. (2006). Unintended Consequences of Caps on Medicare Drug Benefits. *The New England Journal of Medicine, 354*(22), 2349-2359.

James, L. R., & Brett, J. M. (1984). Mediators, Moderators, and Tests for Mediation. *Journal of Applied Psychology, 69*(2), 307-321.

Jones, A. M. (1989). A double-hurdle model of cigarette consumption. *Journal of Applied Econometrics, 4*(1), 23-39. doi: 10.1002/jae.3950040103

Jones, A. M., & Yen, S. T. (2000). A Box–Cox Double-hurdle Model. *The Manchester School, 68*(2), 203-221. doi: 10.1111/1467-9957.00190

Judd, C. M., & Kenny, D. A. (1981). Process Analysis: Estimating Mediation in Treatment Evaluations. *Evaluation Review, 5*(5), 602-619.

Kaestner, R., & Khan, N. (2012). Medicare Part D and Its Effect on the Use of Prescription Drugs and Use of Other Health Care Services of the Elderly. *Journal of Policy Analysis and Management, 31*(2), 253-279. doi: 10.1002/pam.21625

Kaestner, R., Long, C., & Alexander, G. C. (2014). Effects of Prescription Drug Insurance on Hospitalization and Mortality: Evidence from Medicare Part D. *National Bureau of Economic Research Working Paper Series, No. 19948*.

Lau, D. T., & Stubbings, J. (2012). Medicare Part D Research and Policy Highlights, 2012: Impact and Insights. *Clinical Therapeutics, 34*(4), 904-914. doi: http://dx.doi.org/10.1016/j.clinthera.2012.02.012

Leung, S. F., & Yu, S. (1996). On the choice between sample selection and two-part models. *Journal of Econometrics, 72*(1-2), 197-229. doi: 10.1016/0304-4076(94)01720-4

Levy, H., & Weir, D. R. (2010). Take-up of Medicare Part D: Results From the Health and Retirement Study. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 65B*(4), 492-501. doi: 10.1093/geronb/gbp107

Lichtenberg, F. R. (1996). Do (More and Better) Drugs Keep People Out of Hospitals? *The American Economic Review, 86*(2), 384-388. doi: 10.2307/2118157

Lichtenberg, F. R. (2001). Are the benefits of newer drugs worth their cost? Evidence from the 1996 MEPS. *Health Affairs, 20*(5), 241-251.

Lichtenberg, F. R., & Sun, S. X. (2007). The Impact of Medicare Part D on Prescription Drug Use by the Elderly. *Health Affairs, 26*(6), 1735-1744.

Liu, F. X., Alexander, G. C., Crawford, S. Y., Pickard, A. S., Hedeker, D., & Walton, S. M. (2011). The Impact of Medicare Part D on Out-of-Pocket Costs for Prescription Drugs, Medication Utilization, Health Resource Utilization, and Preference-Based Health Utility. *Health Services Research, 46*(4), 1104-1123. doi: 10.1111/j.1475-6773.2011.01273.x

Maddala, G. S. (1985a). A Survey of the Literature on Selectivity Bias as it Pertains to Health Care Markets. *Advances in Health Economics and Health Services Research, 6*, 3-18.

Maddala, G. S. (1985b). Further Comments on Selectivity Bias. *Advances in Health Economics and Health Services Research, 6*, 25-26.

Madden, D. (2008). Sample selection versus two-part models revisited: The case of female smoking and drinking. *Journal of Health Economics, 27*, 300-307.

Madden, J. M., Graves, A. J., Ross-Degnan, D., Briesacher, B. A., & Soumerai, S. B. (2009). Cost-Related Medication Nonadherence After Implementation of Medicare Part D, 2006-2007. *Journal of American Medical Association, 302*(16), 1755-1756.

Madden, J. M., Graves, A. J., Zhang, F., Adams, A. S., Briesacher, B. A., Ross-Degnan, D., . . . Soumerai, S. B. (2008). Cost-Related Medication Nonadherence and Spending on Basic Needs Following Implementation of Medicare Part D. *Journal of American Medical Association, 299*(16), 1922-1928.

Manning, W. G. (1998). The logged dependent variable, heteroscedasticity, and the retransformation problem. *Journal of Health Economics, 17*(3), 283-295.

Manning, W. G. (2006). Dealing with skewed data on costs and expenditures. In A. M. Jones (Ed.), *The Elgar companion to health economics* (pp. 439-446): Edward Elgar Publishing Limited.

Manning, W. G., Duan, N., & Rogers, W. H. (1987). Monte Carlo evidence on the choice between sample selection and two-part models. *Journal of Econometrics, 35*(1), 59-82. doi: 10.1016/0304-4076(87)90081-9

Manning, W. G., & Mullahy, J. (2001). Estimating log models: to transform or not to transform? *Journal of Health Economics, 20*(4), 461-494. doi: 10.1016/s0167-6296(01)00086-8

Manning, W. G., Newhouse, J. P., Duan, N., Keeler, E. B., & Leibowitz, A. (1987). Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment. *The American Economic Review, 77*(3), 251-277.

MAPRx Coalition. (December 2013). Medicare Part D: Saving Money and Improving Health. *Medicare Access for Patients Rx.*, 2014, from http://maprx.info/wp-content/uploads/2013/12/MAPRx-Ten-Year-Anniversary_White-Paper_FINAL_December-2013.pdf

McWilliams, J. M., Zaslavsky, A. M., & Huskamp, H. A. (2011). Implementation of medicare part d and nondrug medical spending for elderly adults with limited prior drug coverage. *Journal of the American Medical Association, 306*(4), 402-409. doi: 10.1001/jama.2011.1026

Miller, D. M. (1984). Reducing Transformation Bias in Curve Fitting. *The American Statistician, 38*(2), 124-126.

Mullahy, J. (2001). [Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice]: Comment. *Journal of Business & Economic Statistics, 19*(1), 23-25.

Nagamine, M., Jiang, H. J., and Merrill, C. T. Trends in Elderly Hospitalizations, 1997-2004. HCUP Statis- tical Brief #14. October 2006. Agency for Healthcare Research and Quality, Rockville, MD. Available at http://www.hcup-us.ahrq.gov/reports/statbriefs/sb14.pdf.

Norton, E., Dow, W., & Do, Y. (2008). Specification tests for the sample selection and two-part models. *Health Services and Outcomes Research Methodology, 8*(4), 201-208. doi: 10.1007/s10742-008-0037-8

Parente, S. T., & Evans, W. N. (1998). Effect of Low-Income Elderly Insurance Copayment Subsidies. *Health Care Financing Review, 20*(2), 19.

Pfuntner, A (Truven Health Analytics), Wier, LM (Truven Health Analytics), Steiner, C (AHRQ). Costs for Hospital Stays in the United States, 2010. HCUP Statistical Brief #146. January 2013. Agency for Healthcare Research and Quality, Rockville, MD. Available at http://www.hcup-us.ahrq.gov/reports/statbriefs/sb146.pdf.

Piette, J. D., Wagner, T. H., Potter, M. B., & Schillinger, D. (2004). Health Insurance Status, Cost-Related Medication Underuse, and Outcomes Among Diabetes Patients in Three Systems of Care. *Medical Care, 42*(2), 102-109.

Pohlmeier, W., & Ulrich, V. (1995). An Econometric Model of the Two-Part Decisionmaking Process in the Demand for Health Care. *The Journal of Human Resources, 30*(2), 339-361.

Polinski, J. M., Donohue, J. M., Kilabuk, E., & Shrank, W. H. (2011). Medicare Part D's Effect on the Under- and Overuse of Medications: A Systematic Review. *Journal of the American Geriatrics Society, 59*(10), 1922-1933. doi: 10.1111/j.1532-5415.2011.03537.x

Polinski, J. M., Kilabuk, E., Schneeweiss, S., Brennan, T., & Shrank, W. H. (2010). Changes in Drug Use and Out-of-Pocket Costs Associated with Medicare Part D Implementation: A Systematic Review. *Journal of the American Geriatrics Society, 58*(9), 1764-1779. doi: 10.1111/j.1532-5415.2010.03025.x

Puhani, P. (2000). The Heckman Correction for Sample Selection and Its Critique. *Journal of Economic Surveys, 14*(1), 53-68. doi: 10.1111/1467-6419.00104

RAND HRS Data, Version M. Produced by the RAND Center for the Study of Aging, with funding from the National Institute on Aging and the Social Security Administration. Santa Monica, CA (September 2013).

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*(5), 688-701. doi: 10.1037/h0037350

Rubin, D. B. (1977). Assignment to Treatment Group on the Basis of a Covariate. *Journal of Educational Statistics, 2*(1), 1-26. doi: 10.2307/1164933

Safran, D. G., Neuman, P., Schoen, C., Kitchman, M. S., Wilson, I. B., Cooper, B., . . . Rogers, W. H. (2005). Prescription Drug Coverage And Seniors: Findings From A 2003 National Survey. *Health Affairs*, w5-152.

Safran, D. G., Strollo, M. K., Guterman, S., Li, A., Rogers, W. H., & Neuman, P. (2010). Prescription Coverage, Use and Spending Before and After Part D Implementation: A National Longitudinal Panel Study. *Journal of General Internal Medicine, 25*(1), 10-17. doi: 10.1007/s11606-009-1134-2

Salas, C., & Raftery, J. P. (2001). Econometric issues in testing the age neutrality of health care expenditure. *Health Economics, 10*(7), 669-671. doi: 10.1002/hec.638

Schneeweiss, S., Patrick, A. R., Pedan, A., Varasteh, L., Levin, R., Liu, N., & Shrank, W. H. (2009). The Effect Of Medicare Part D Coverage On Drug Use And Cost Sharing Among Seniors Without Prior Drug Benefits. *Health Affairs, 28*(2), w305-w316. doi: 10.1377/hlthaff.28.2.w305

Seshamani, M., & Gray, A. (2004). Ageing and health-care expenditure: the red herring argument revisited. *Health Economics, 13*(4), 303-314. doi: 10.1002/hec.826

Shang, B., & Goldman, D. P. (2007). *Prescription Drug Coverage and Elderly Medicare Spending*. NBER Working Paper Series. NBER.

Shea, D. G., Terza, J. V., Stuart, B. C., & Briesacher, B. (2007). Estimating the Effects of Prescription Drug Coverage for Medicare Beneficiaries. *Health Services Research, 42*(3p1), 933-949. doi: 10.1111/j.1475-6773.2006.00659.x

Simoni-Wastila, L., Zuckerman, I. H., Shaffer, T., Blanchette, C. M., & Stuart, B. (2008). Drug Use Patterns in Severely Mentally Ill Medicare Beneficiaries: Impact of Discontinuities in Drug Coverage. *Health Services Research, 43*(2), 496-514. doi: 10.1111/j.1475-6773.2007.00779.x

Sokol, M. C., McGuigan, K. A., Verbrugge, R. R., & Epstein, R. S. (2005). Impact of Medication Adherence on Hospitalization Risk and Healthcare Cost. *Medical Care, 43*(6), 521-530. doi: 10.2307/3768169

Soumerai, S. B., Pierre-Jacques, M., Zhang, F., Ross-Degnan, D., Adams, A. S., Gurwitz, J., . . . Safran, D. G. (2006). Cost-related medication nonadherence among elderly and disabled medicare beneficiaries: A national survey 1 year before the medicare drug benefit. *Archives of Internal Medicine, 166*(17), 1829-1835. doi: 10.1001/archinte.166.17.1829

Stagnitti, M. N. *Trends in Outpatient Prescription Drug Utilization and Expenditures: 1997 and 2004*. Statistical Brief # 168. April 2007. Agency for Healthcare Research and Quality, Rockville, Md. http://www.meps.ahrq.gov/mepsweb/data files/publications/st168/stat168.pdf

StataCorp. 2009. Stata Statistical Software: Release 11. College Station, TX: StataCorp LP.

Stuart, B. C., Doshi, J. A., & Terza, J. V. (2009). Assessing the Impact of Drug Use on Hospital Costs. *Health Services Research, 44*(1), 128-144.

Sullivan, P. W., Ghushchyan, V., Wyatt, H. R., & Hill, J. O. (2007). The Medical Cost of Cardiometabolic Risk Factor Clusters in the United States. *Obesity, 15*(12), 3150-3158. doi: 10.1038/oby.2007.375

Taylor, J. M. G. (1986). The Retransformed Mean After a Fitted Power Transformation. *Journal of the American Statistical Association, 81*(393), 114-118.

Terza, J. V. (1994). *Dummy Endogenous Variables and Endogenous Switching in Exponential Conditional Mean Models*. Department of Economics, Penn State University. Unpublished Manuscript.

Terza, J. V. (1996). *FIML, Method of Moments and Two Stage Method of Moments Estimators for Nonlinear Regression Models with Endogenous Switching and Sample Selection*. Department of Economics, Penn State University. Unpublished Manuscript.

Terza, J. V. (1998). Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects. *Journal of Econometrics, 84*(1), 129-154. doi: 10.1016/s0304-4076(97)00082-1

Terza, J. V. (2009). Parametric Nonlinear Regression with Endogenous Switching. *Econometric Reviews, 28*(6), 555 - 580.

Terza, J. V., Basu, A., & Rathouz, P. J. (2008). Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *Journal of Health Economics, 27*(3), 531-543. doi: http://dx.doi.org/10.1016/j.jhealeco.2007.09.009

Terza, J. V., Bradford, W. D., & Dismuke, C. E. (2008). The Use of Linear Instrumental Variables Methods in Health Services Research and Health Economics: A Cautionary Note. *Health Services Research, 43*(3), 1102-1120. doi: 10.1111/j.1475-6773.2007.00807.x

Terza, J. V., & Tsai, W.-D. (2006). Censored Probit Estimation with Correlation Near the Boundary: A Useful Reparameterization. *Review of Applied Economics, 2*(1), 1-12.

Terza, J. V. (2014a). Empirical health policy analysis in the context of the two-part model. Unpublished manuscript, Department of Economics, Indiana University Purdue University Indianapolis.

Terza, J. V. (2014b). Health policy analysis from a potential outcomes perspective: Smoking during pregnancy and birth weight. Unpublished manuscript, Department of Economics, Indiana University Purdue University Indianapolis.

Terza, J. V. (2014c). Simpler standard errors for multi-stage regression based estimators: Illustratioins in health economics. Unpublished manuscript, Department of Economics, Indiana University Purdue University Indianapolis.

The Kaiser Family Foundation State Health Facts. Data Source: 2006 Annual AHA Survey.

Tierney, J. A. (1968). *Calculus and Analytic Geometry*. Boston: Allyn and Bacon.

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica, 26*(1), 24-36. doi: 10.2307/1907382

Trogdon, J. G., Finkelstein, E. A., Feagan, C. W., & Cohen, J. W. (2012). State- and Payer-Specific Estimates of Annual Medical Expenditures Attributable to Obesity. *Obesity, 20*, 214-220.

Welch, W. P., Frank, R. G., & Diehr, P. (1984). Health Care Costs in Health Maintenance Organizations: Correcting For Self-Selection. *Advances in Health Economics and Health Services Research, 5*, 95-128.

White, H. L. (1982). Maximum likelihood estimation of misspecified models. *Econometrica, 50*, 1-25.

Williams, J., Steers, W. N., Ettner, S. L., Mangione, C. M., & Duru, O. K. (2013). Cost-related Nonadherence by Medication Type Among Medicare Part D Beneficiaries With Diabetes. *Medical Care, 51*(2), 193-198 110.1097/MLR.1090b1013e318270dc318252.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (Second Edition ed.). Cambridge, Massachusetts; London, England: The MIT Press.

Yin, W., Basu, A., Zhang, J. X., Rabbani, A., Meltzer, D. O., & Alexander, G. C. (2008). The Effect of the Medicare Part D Prescription Benefit on Drug Utilization and Expenditures. *Annals of Internal Medicine, 148*(3), 169-177.

Zhang, Y., Donohue, J. M., Lave, J. R., O'Donnell, G., & Newhouse, J. P. (2009). The Effect of Medicare Part D on Drug and Medical Spending. *New England Journal of Medicine, 361*(1), 52-61.

Zheng, X., & Zimmer, D. M. (2008). Farmers' Health Insurance and Access to Health Care. *American Journal of Agricultural Economics, 90*, 267-279.

Zhou, X.-H., Gao, S., & Hui, S. L. (1997). Methods for Comparing the Means of Two Independent Log-Normal Samples. *Biometrics, 53*(3), 1129-1135.

Zweifel, P., Felder, S., & Meiers, M. (1999). Ageing of population and health care expenditure: a red herring? *Health Economics, 8*(6), 485-496. doi: 10.1002/(sici)1099-1050(199909)8:6<485::aid-hec461>3.0.co;2-4

Zweifel, P., Felder, S., & Meiers, M. (2001). Reply to: econometric issues in testing the age neutrality of health care expenditure. *Health Economics, 10*(7), 673-674. doi: 10.1002/hec.637

APPENDIX A

TECHNICAL APPENDIX FOR CHAPTER II

Derivation of Equation (12) from (14)

Let

$$D = \exp(0.5\beta_u^2) \times \frac{\Phi(X_{po1}\beta_1 + \beta_u)}{\Phi(X_{po1}\beta_1)}. \tag{A.1}$$

Then after inserting (14) into (10) and using the assumption that $(X_u \mid X_p, X_{o1}) \sim N(0,1)$, we have

$$k(X_p, X_{o1}; \beta_2) = \exp(X_p\beta_{p2} + X_{o1}\beta_{o2}) \times D^{-1} \times \frac{\int_{-X_{po1}\beta_1}^{\infty} \exp(X_u\beta_u)\varphi(X_u)dX_u}{\Phi(X_{po1}\beta_1)}$$

$$= \exp(X_p\beta_{p2} + X_{o1}\beta_{o2}) \times D^{-1} \times \frac{1}{\sqrt{2\pi}} \times \frac{\int_{-X_{po1}\beta_1}^{\infty} \exp(X_u\beta_u - 0.5\beta_u^2)dX_u}{\Phi(X_{po1}\beta_1)}$$

$$= \exp(X_p\beta_{p2} + X_{o1}\beta_{o2}) \times D^{-1} \times \frac{1}{\sqrt{2\pi}} \times \frac{\sqrt{2\pi}\exp(0.5\beta_u^2)\Phi(X_{po1}\beta_1 + \beta_u)}{\Phi(X_{po1}\beta_1)}$$

$$= \exp(X_p\beta_{p2} + X_{o1}\beta_{o2}) \times D^{-1} \times D$$

$$= \exp(X_p\beta_{p2} + X_{o1}\beta_{o2}), \tag{A.2}$$

which is equivalent to (12). Equation (18) is easily obtained using the above calculation except that D^{-1} does not appear from the outset, and one must allow the extent of participation vector of observable confounders differ from the first-hurdle resulting in X_{o2} being used rather than X_{o1} in the exponentiated linear index.

APPENDIX B

TECHNICAL APPENDIX FOR CHAPTER III

Derivation of Equation (38)

Observe:

$$f(X_u \mid X_p, W, H = 1) = \frac{f(X_u, H \mid X_p, W)}{f(H \mid X_p, W)} \tag{B.1}$$

Under all modeling assumptions one can write[24]

$$\begin{aligned} f(X_u, H \mid X_p, W) = {} & X_p H \frac{\Phi(X\beta_1)\varphi_{[X_u > -W\alpha]}(X_u)}{\Phi(W\alpha)} \\ & + X_p(1-H)\frac{\left[1-\Phi(X\beta_1)\right]\varphi_{[X_u > -W\alpha]}(X_u)}{\Phi(W\alpha)} \\ & + (1-X_p)H\frac{\Phi(X\beta_1)\varphi_{[X_u \le -W\alpha]}(X_u)}{1-\Phi(W\alpha)} \\ & + (1-X_p)(1-H)\frac{\left[1-\Phi(X\beta_1)\right]\varphi_{[X_u \le -W\alpha]}(X_u)}{1-\Phi(W\alpha)} \end{aligned} . \tag{B.2}$$

Using (B.2), we can obtain the marginal density of H as follows:

[24] This follows from the normality assumptions, the fact that conditional on W these random variables are independent, and because X_u can be composed in terms of the truncated densities implicit in (B.2).

$$
\begin{aligned}
f(H \mid X_p, W) &= \int_{-\infty}^{\infty} f(X_u, H \mid X_p, W) dX_u \\
&= X_p H \frac{\int_{-W\alpha}^{\infty} \Phi(X\beta_1)\varphi(X_u)dX_u}{\Phi(W\alpha)} \\
&\quad + X_p(1-H)\frac{\int_{-W\alpha}^{\infty}\left[1-\Phi(X\beta_1)\right]\varphi(X_u)dX_u}{\Phi(W\alpha)} \\
&\quad + (1-X_p)H\frac{\int_{-\infty}^{-W\alpha}\Phi(X\beta_1)\varphi(X_u)dX_u}{1-\Phi(W\alpha)} \\
&\quad + (1-X_p)(1-H)\frac{\int_{-\infty}^{-W\alpha}\left[1-\Phi(X\beta_1)\right]\varphi(X_u)dX_u}{1-\Phi(W\alpha)}
\end{aligned}
\qquad . \qquad \text{(B.3)}
$$

Combining (B.1), (B.2), and (B.3), and after omitting components involving H=0, the following obtains:

$$
\begin{aligned}
f(X_u \mid X_p, W, H=1) &= X_p \frac{\Phi(X\beta_1)\varphi_{[X_u > -W\alpha]}(X_u)}{\int_{-W\alpha}^{\infty}\Phi(X\beta_1)\varphi(X_u)dX_u} \\
&\quad + (1-X_p)\frac{\Phi(X\beta_1)\varphi_{[X_u \le -W\alpha]}(X_u)}{\int_{-\infty}^{-W\alpha}\Phi(X\beta_1)\varphi(X_u)dX_u}
\end{aligned}
\qquad . \qquad \text{(B.4)}
$$

On the basis of (B.4) we can write:

$$
\begin{aligned}
E_{X_u}\left[Y \mid X_p, W, H=1\right] &= \int_{-\infty}^{\infty}\exp(X\beta_2) f(X_u \mid X_p, W, H=1)dX_u \\
&= X_p \frac{\int_{-W\alpha}^{\infty}\exp(X\beta_2)\Phi(X\beta_1)\varphi(X_u)dX_u}{\int_{-W\alpha}^{\infty}\Phi(X\beta_1)\varphi(X_u)dX_u} \\
&\quad + (1-X_p)\frac{\int_{-\infty}^{-W\alpha}\exp(X\beta_2)\Phi(X\beta_1)\varphi(X_u)dX_u}{\int_{-\infty}^{-W\alpha}\Phi(X\beta_1)\varphi(X_u)dX_u}
\end{aligned}
\qquad , \qquad \text{(B.5)}
$$

which is equivalent to equation (38).

Gauss-Legendre Quadrature Routine

Note that the range of integration in (38) depends on the value of X_p. In particular, the range of integration is $B = (-\infty, -W\alpha]$ when $X_p = 1$ and the range is $B = [-W\alpha, \infty)$ when $X_p = 0$. The following change of variables is immensely useful for conducting Gauss-Legendre quadrature with the integration problem inherent in (38):

$$\int_B f(x)dx = 2\int_{-1}^{1} f\left(-W\alpha + (2X_p - 1)\frac{1-t}{1+t}\right)\frac{dt}{(1+t)^2}. \tag{B.6}$$

Using this, the integration problem in (38) can be re-written as:

$$2\int_{-1}^{1} \exp(X_{po}\beta_{po2} + \psi)\Phi(X_{po}\beta_{po1} + \psi)\varphi(\psi)\frac{dt}{(1+t)^2}, \tag{B.7}$$

where

$$\psi = -W\alpha + (2X_p - 1)\frac{1-t}{1+t}. \tag{B.8}$$

Equation (B.7) is in the format needed for Gauss-Legendre quadrature. Therefore, all that one needs to do is obtain the appropriate nodes and weights and then calculate the weighted sum required under the Gauss-Legendre rule.

Mathematical Details Needed to Construct a Newton-Raphson Algorithm

The second-stage estimator described in (39) can be written as

$$\hat{\beta}_2 = \underset{\beta_2}{\operatorname{argmax}}\, Q_n\left(\beta_2; \hat{\alpha}, \hat{\beta}_1\right), \tag{B.9}$$

where

$$Q_n = \frac{1}{n}\sum_{i=1}^{n} q_i(\beta_2), \tag{B.10}$$

$$q_i(\beta_2) = -\left(Y_i - J_i(\beta_2)\right)^2, \tag{B.11}$$

and

$$\begin{aligned} J_i(\beta_2) = X_{pi}\frac{\int_{-W_i\hat{\alpha}}^{\infty}\exp(X_i\beta_2)\Phi(X_i\hat{\beta}_1)\varphi(X_u)dX_u}{\int_{-W_i\hat{\alpha}}^{\infty}\Phi(X_i\hat{\beta}_1)\varphi(X_u)dX_u} \\ +(1-X_{pi})\frac{\int_{-\infty}^{-W_i\hat{\alpha}}\exp(X_i\beta_2)\Phi(X_i\hat{\beta}_1)\varphi(X_u)dX_u}{\int_{-\infty}^{-W_i\hat{\alpha}}\Phi(X_i\hat{\beta}_1)\varphi(X_u)dX_u} \end{aligned}. \tag{B.12}$$

We need a gradient and a Hessian (or Hessian approximation). The gradient is

$$\nabla_{\beta_2} Q_n = \left[\nabla_{\beta_{po2}} Q_n \quad \nabla_{\beta_{u2}} Q_n\right] \tag{B.13}$$

where $\beta'_{po2} = \left[\beta_{p2} \quad \beta'_{o2}\right]$. The first component of (B.13) is

$$\nabla_{\beta_{po2}} Q_n = 2\sum_{i=1}^{n} e_i \nabla_{\beta_{po2}} J_i(\beta_2), \tag{B.14}$$

where

$$e_i = Y_i - J_i(\beta_2)\,, \tag{B.15}$$

$$\nabla_{\beta_{po2}} J_i(\beta_2) = J_i X_{poi}\,, \tag{B.16}$$

and

$$X_{poi} = \begin{bmatrix} X_{pi} & X_{oi} \end{bmatrix}. \tag{B.17}$$

The second component of (B.13) is

$$\nabla_{\beta_{u2}} Q_n = 2\sum_{i=1}^{n} e_i \nabla_{\beta_{u2}} J_i(\beta_2)\,, \tag{B.18}$$

where

$$\begin{aligned} \nabla_{\beta_{u2}} J_i(\beta_2) = X_{pi} & \frac{\int_{-W_i\hat{\alpha}}^{\infty} \exp(X_i\beta_2)\Phi(X_i\hat{\beta}_1)X_u\varphi(X_u)dX_u}{\int_{-W_i\hat{\alpha}}^{\infty} \Phi(X_i\hat{\beta}_1)\varphi(X_u)dX_u} \\ & + (1 - X_{pi})\frac{\int_{-\infty}^{-W_i\hat{\alpha}} \exp(X_i\beta_2)\Phi(X_i\hat{\beta}_1)X_u\varphi(X_u)dX_u}{\int_{-\infty}^{-W_i\hat{\alpha}} \Phi(X_i\hat{\beta}_1)\varphi(X_u)dX_u} \end{aligned}. \tag{B.19}$$

We found it more efficient to use the sample analog of the expected value of the Hessian matrix in our optimization algorithm. This matrix is given by

$$\hat{E}\left[\nabla_{\beta_2\beta_2} Q_n\right] = \begin{bmatrix} \hat{E}\left[\nabla_{\beta_{po2}\beta_{po2}} Q_n\right] & \hat{E}\left[\nabla_{\beta_{po2}\beta_{u2}} Q_n\right] \\ \hat{E}\left[\nabla_{\beta_{po2}\beta_{u2}} Q_n\right] & \hat{E}\left[\nabla_{\beta_{u2}\beta_{u2}} Q_n\right] \end{bmatrix}, \tag{B.20}$$

where

$$\hat{E}\left[\nabla_{\beta_{po2}\beta_{po2}} Q_n\right] = -2\sum_{i=1}^{n} \nabla_{\beta_{po2}} J_i(\beta_2)' \nabla_{\beta_{po2}} J_i(\beta_2), \tag{B.21}$$

$$\hat{E}\left[\nabla_{\beta_{po2}\beta_{u2}} Q_n\right] = -2\sum_{i=1}^{n} \nabla_{\beta_{po2}} J_i(\beta_2) \nabla_{\beta_{u2}} J_i(\beta_2), \tag{B.22}$$

and

$$\hat{E}\left[\nabla_{\beta_{u2}\beta_{u2}} Q_n\right] = -2\sum_{i=1}^{n} \nabla_{\beta_{u2}} J_i(\beta_2)^2 \tag{B.23}$$

Asymptotic Details for the Two-Stage Estimator

This derivation follows the approach detailed in Terza (2014c). Let D denote the variance covariance matrix for the full two-stage estimator. Consider the following decomposition of D:

$$D = \begin{bmatrix} D_{11} & D_{12} \\ D_{12}' & D_{22} \end{bmatrix}, \tag{B.24}$$

where D_{11} denotes the variance matrix for the first-stage estimator for $\delta' = \left[\alpha' \quad \beta_1'\right]$, D_{12} is the covariance matrix between the first and second-stage estimators, and D_{22} is

the variance matrix for the second-stage estimator for β_2. D_{11} is readily available since it is the standard variance matrix, which can be obtained as post-estimation output in Stata. Terza (2014c) has shown that when the second stage is nonlinear least squares, then we can write

$$D_{12} = -\text{AVAR}(\hat{\delta})\, E\left[\nabla_{\beta_2\delta} q\right]' E\left[\nabla_{\beta_2\beta_2} q\right]^{-1}, \tag{B.25}$$

and

$$D_{22} = E\left[\nabla_{\beta_2\beta_2} q\right]^{-1} E\left[\nabla_{\beta_2\delta} q\right] \text{AVAR}(\hat{\delta})\, E\left[\nabla_{\beta_2\delta} q\right]' E\left[\nabla_{\beta_2\beta_2} q\right]^{-1} + \text{AVAR}^*(\hat{\beta}_2), \tag{B.26}$$

where AVAR(A) denotes the asymptotic variance of A, AVAR* denotes the variance matrix that would be correct if the second-stage estimator did not take first-stage inputs, and

$$q = -(Y - J(\beta_2))^2, \tag{B.27}$$

and J( ) is defined in equation (B.12).

We start with the gradient of q with respect to the second-stage estimator; the gradient is:

$$\nabla_{\beta_2} q = \left[\nabla_{\beta_{po2}} q \quad \nabla_{\beta_{u2}} q\right], \tag{B.28}$$

where

$$\nabla_{\beta_{po2}} q = 2\, e\, \nabla_{\beta_{po2}} J(\beta_2)\,, \tag{B.29}$$

and

$$\nabla_{\beta_{u2}} q = 2\, e\, \nabla_{\beta_{u2}} J(\beta_2)\,. \tag{B.30}$$

Note that the gradients in (B.29) and (B.30) are given in (B.14) and (B.19). It is relatively easy to obtain the expected value of the second derivative matrix:

$$E\left[\nabla_{\beta_2\beta_2} q\right] = \begin{bmatrix} E\left[\nabla_{\beta_{po2}\beta_{po2}} q\right] & E\left[\nabla_{\beta_{po2}\beta_{u2}} q\right] \\ E\left[\nabla_{\beta_{po2}\beta_{u2}} q\right] & E\left[\nabla_{\beta_{u2}\beta_{u2}} q\right] \end{bmatrix}, \tag{B.31}$$

where

$$E\left[\nabla_{\beta_{po2}\beta_{po2}} q\right] = 2\, e\, \nabla_{\beta_{po2}} J(\beta_2)' \nabla_{\beta_{po2}} J(\beta_2)\,, \tag{B.32}$$

$$E\left[\nabla_{\beta_{po2}\beta_{u2}} q\right] = 2\, e\, \nabla_{\beta_{po2}} J(\beta_2) \nabla_{\beta_{u2}} J(\beta_2)\,, \tag{B.33}$$

and

$$E\left[\nabla_{\beta_{u2}\beta_{u2}} q\right] = 2\, e\, \nabla_{\beta_{u2}} J(\beta_2)^2\,. \tag{B.34}$$

The expected value of the cross-partial derivative matrix is also relatively easy to obtain:

$$E\left[\nabla_{\beta_2\delta} q\right] = \begin{bmatrix} E\left[\nabla_{\beta_{po2}\alpha} q\right] & E\left[\nabla_{\beta_{po2}\beta_{po1}} q\right] & E\left[\nabla_{\beta_{po2}\beta_{u1}} q\right] \\ E\left[\nabla_{\beta_{u2}\alpha} q\right] & E\left[\nabla_{\beta_{u2}\beta_{po1}} q\right] & E\left[\nabla_{\beta_{u2}\beta_{u1}} q\right] \end{bmatrix}, \tag{B.35}$$

where

$$E\left[\nabla_{\beta_{po2}\alpha} q\right] = -2E\left[\nabla_{\beta_{po2}} J(\beta_2)' \nabla_{\alpha} J(\beta_2)\right], \tag{B.36}$$

$$E\left[\nabla_{\beta_{po2}\beta_{po1}} q\right] = -2E\left[\nabla_{\beta_{po2}} J(\beta_2)' \nabla_{\beta_{po1}} J(\beta_2)\right], \tag{B.37}$$

$$E\left[\nabla_{\beta_{po2}\beta_{u1}} q\right] = -2E\left[\nabla_{\beta_{po2}} J(\beta_2) \nabla_{\beta_{u1}} J(\beta_2)\right], \tag{B.38}$$

$$E\left[\nabla_{\beta_{u2}\alpha} q\right] = -2E\left[\nabla_{\beta_{u2}} J(\beta_2) \nabla_{\alpha} J(\beta_2)\right], \tag{B.39}$$

$$E\left[\nabla_{\beta_{u2}\beta_{po1}} q\right] = -2E\left[\nabla_{\beta_{u2}} J(\beta_2) \nabla_{\beta_{po1}} J(\beta_2)\right], \tag{B.40}$$

and

$$E\left[\nabla_{\beta_{u2}\beta_{u1}} q\right] = -2E\left[\nabla_{\beta_{u2}} J(\beta_2) \nabla_{\beta_{u1}} J(\beta_2)\right]. \tag{B.41}$$

Thus all that is needed are analytical expressions for the gradient of J with respect to the first-stage estimator. This gradient is given by:

$$\nabla_{\delta} J(\beta_2) = \left[\nabla_{\alpha} J(\beta_2) \quad \nabla_{\beta_{po1}} J(\beta_2) \quad \nabla_{\beta_{u1}} J(\beta_2)\right]. \tag{B.42}$$

It will be helpful to have a couple of results and some new notation before proceeding. First, note that[25]

$$\nabla_{\alpha}\int_{-\infty}^{-W\alpha} g(X_u)dX_u = -g(-W\alpha)W\ , \tag{B.43}$$

and

$$\nabla_{\alpha}\int_{-W\alpha}^{\infty} = g(-W\alpha)W\ . \tag{B.44}$$

Also, let

$$\psi_1 = \int_{-W\alpha}^{\infty} \exp(X\beta_2)\Phi(X\beta_1)\varphi(X_u)dX_u\ , \tag{B.45}$$

$$\psi_0 = \int_{-\infty}^{-W\alpha} \exp(X\beta_2)\Phi(X\beta_1)\varphi(X_u)dX_u\ , \tag{B.46}$$

$$\theta_1 = \int_{-W\alpha}^{\infty}\Phi(X\beta_1)\varphi(X_u)dX_u)\ , \tag{B.47}$$

and

$$\theta_0 = \int_{-\infty}^{-W\alpha} \Phi(X\beta_1)\varphi(X_u)dX_u\ . \tag{B.48}$$

Now we have

$$\nabla_{\alpha}J(\beta_2) = X_p\frac{\nabla_{\alpha}\psi_1\theta_1 - \psi_1\nabla_{\alpha}\theta_1}{\theta_1^2} + (1-X_p)\frac{\nabla_{\alpha}\psi_0\theta_0 - \psi_0\nabla_{\alpha}\theta_0}{\theta_0^2}\ , \tag{B.49}$$

[25] Tierney (1968).

where

$$\nabla_{\alpha}\psi_1 = \exp(X_{po}\beta_{po2} + (-W\alpha)\beta_{u2})\Phi(X_{po}\beta_{po1} + (-W\alpha))\varphi(-W\alpha)W\,, \tag{B.50}$$

$$\nabla_{\alpha}\theta_1 = \Phi(X_{po}\beta_{po1} + (-W\alpha))\varphi(-W\alpha)W\,, \tag{B.51}$$

$$\nabla_{\alpha}\psi_0 = -\left[\exp(X_{po}\beta_{po2} + (-W\alpha))\Phi(X_{po}\beta_{po1} + (-W\alpha))\varphi(-W\alpha)W\right], \tag{B.52}$$

and

$$\nabla_{\alpha}\theta_0 = -\left[\Phi(X_{po}\beta_{po1} + (-W\alpha))\varphi(-W\alpha)W\right]. \tag{B.53}$$

Similarly, we have

$$\nabla_{\beta_{po1}} J(\beta_2) = X_p \frac{\nabla_{\beta_{po1}}\psi_1\theta_1 - \psi_1\nabla_{\beta_{po1}}\theta_1}{\theta_1^2} + (1-X_p)\frac{\nabla_{\beta_{po1}}\psi_0\theta_0 - \psi_0\nabla_{\beta_{po1}}\theta_0}{\theta_0^2}, \tag{B.54}$$

where

$$\nabla_{\beta_{po1}}\psi_1 = \int_{-W\alpha}^{\infty}\exp(X\beta_2)\Phi(X\beta_1)\varphi(X_u)dX_u X_{po}\,, \tag{B.55}$$

$$\nabla_{\beta_{po1}}\theta_1 = \int_{-W\alpha}^{\infty}\Phi(X\beta_1)\varphi(X_u)dX_u X_{po}\,, \text{(D.33)} \tag{B.56}$$

$$\nabla_{\beta_{po1}}\psi_0 = \int_{-\infty}^{-W\alpha}\exp(X\beta_2)\Phi(X\beta_1)\varphi(X_u)dX_u X_{po}\,, \tag{B.57}$$

and

$$\nabla_{\beta_{po1}}\theta_0 = \int_{-\infty}^{-W\alpha} \Phi(X\beta_1)\varphi(X_u)dX_u X_{po}\,. \tag{B.58}$$

Finally, we have

$$\nabla_{\beta_{u1}} J(\beta_2) = X_p \frac{\nabla_{\beta_{u1}}\psi_1\theta_1 - \psi_1\nabla_{\beta_{u1}}\theta_1}{\theta_1^2} + (1-X_p)\frac{\nabla_{\beta_{u1}}\psi_0\theta_0 - \psi_0\nabla_{\beta_{u1}}\theta_0}{\theta_0^2}, \tag{B.59}$$

where

$$\nabla_{\beta_{u1}}\psi_1 = \int_{-W\alpha}^{\infty} \exp(X\beta_2)\Phi(X\beta_1)X_u\varphi(X_u)dX_u\,, \tag{B.60}$$

$$\nabla_{\beta_{u1}}\theta_1 = \int_{-W\alpha}^{\infty} \Phi(X\beta_1)X_u\varphi(X_u)dX_u\,, \tag{B.61}$$

$$\nabla_{\beta_{u1}}\psi_0 = \int_{-\infty}^{-W\alpha} \exp(X\beta_2)\Phi(X\beta_1)X_u\varphi(X_u)dX_u\,, \tag{B.62}$$

and

$$\nabla_{\beta_{u1}}\theta_0 = \int_{-\infty}^{-W\alpha} \Phi(X\beta_1)X_u\varphi(X_u)dX_u\,. \tag{B.63}$$

Derivation of Equation (42)

Observe:

$$f(X_u \mid X_p, W, H=1) = \frac{f(X_u, H \mid X_p, W)}{f(H \mid X_p, W)} \tag{B.64}$$

Under all modeling assumptions one can write[26]

$$
\begin{aligned}
f(X_u, H \mid X_p, W) = {} & X_p H \frac{\Phi(X_{po}\beta_{po1})\varphi_{[X_u > -W\alpha]}(X_u)}{\Phi(W\alpha)} \\
& + X_p(1-H)\frac{\left[1-\Phi(X_{po}\beta_{po1})\right]\varphi_{[X_u > -W\alpha]}(X_u)}{\Phi(W\alpha)} \\
& + (1-X_p)H\frac{\Phi(X_{po}\beta_{po1})\varphi_{[X_u \le -W\alpha]}(X_u)}{1-\Phi(W\alpha)} \\
& + (1-X_p)(1-H)\frac{\left[1-\Phi(X_{po}\beta_{po1})\right]\varphi_{[X_u \le -W\alpha]}(X_u)}{1-\Phi(W\alpha)}
\end{aligned}
\tag{B.65}
$$

Using (E.2), we can obtain the marginal density of H as follows:

$$
\begin{aligned}
f(H \mid X_p, W) &= \int_{-\infty}^{\infty} f(X_u, H \mid X_p, W)\,dX_u \\
&= X_p H\Phi(X_{po}\beta_{po1}) + X_p(1-H)\left[1-\Phi(X_{po}\beta_{po1}\right] \\
&\quad + (1-X_p)H\Phi(X_{po}\beta_{po1}) + (1-X_p)(1-H)\left[1-\Phi(X_{po}\beta_{po1})\right]
\end{aligned}
\tag{B.66}
$$

Combining (E.1), (E.2), and (E.3), and after omitting components involving H=0, the following obtains:

$$
f(X_u \mid X_p, W, H=1) = X_p \frac{\varphi_{[X_u > -W\alpha]}(X_u)}{\Phi(W\alpha)} + (1-X_p)\frac{\varphi_{[X_u \le -W\alpha]}(X_u)}{1-\Phi(W\alpha)}. \tag{B.67}
$$

[26] This follows from the normality assumptions, the fact that conditional on W these random variables are independent, and because X_u can be composed in terms of the truncated densities implicit in (A.2).

On the basis of (E.4) we can write:

$$
\begin{aligned}
E_{X_u}\left[Y \mid X_p, W, H = 1 \right] &= \int_{-\infty}^{\infty} \exp(X\beta_2) f(X_u \mid X_p, W, H = 1) dX_u \\
&= X_p \frac{\int_{-W\alpha}^{\infty} \exp(X\beta_2)\varphi(X_u) dX_u}{\Phi(W\alpha)} \\
&\quad + (1 - X_p) \frac{\int_{-\infty}^{-W\alpha} \exp(X\beta_2)\varphi(X_u) dX_u}{1 - \Phi(W\alpha)} \qquad , \qquad \text{(B.68)} \\
&= \exp(X_{po}\beta_{po2}) \left[X_p \frac{\int_{-W\alpha}^{\infty} \exp(X_u\beta_{u2})\varphi(X_u) dX_u}{\Phi(W\alpha)} \right. \\
&\quad \left. + (1 - X_p) \frac{\int_{-\infty}^{-W\alpha} \exp(X_u\beta_{u2})\varphi(X_u) dX_u}{1 - \Phi(W\alpha)} \right]
\end{aligned}
$$

which is sufficient to demonstrate that the first and second part parameters are "separable." To obtain equation (26) we prove the following statement:

If z is normal with mean a and variance b, then

$$E\left[\exp(cz) \right] = \exp(ca + 0.5c^2 b) . \qquad \text{(B.69)}$$

Proof:

$$E\left[\exp(cz) \right] = \int_{-\infty}^{\infty} \exp(cz) \frac{1}{\sqrt{2\square}} \exp\left(-\frac{(z-a)^2}{2b} \right) dz \qquad \text{(B.70)}$$

Now make the substitution $q = \frac{z-a}{\sqrt{b}}$. We get

$$
\begin{aligned}
E\left[\exp(cz)\right] &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(c(\sqrt{b}q+a))\exp(-0.5q^2)dq \\
&= \exp(ca)\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(c\sqrt{b}q-0.5q^2)dq \\
&= \exp(ca)\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-0.5(q-c\sqrt{b})^2+0.5c^2b\right)dq \\
&= \exp(ca+0.5c^2b)\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(-0.5(q-c\sqrt{b})^2)dq \\
&= \exp(ca+0.5c^2b)
\end{aligned}
\tag{B.71}
$$

where the last equality follows from the fact that the integral evaluates to one because it defines a density function for a normal random variable q with mean $c\sqrt{b}$. QED

If we set a=0 and b=1, and let $c=\beta_{u2}$, then we can use the lemma above to show that

$$
\int_{-W\alpha}^{\infty}\exp(X_u\beta_{u2})\varphi(X_u)dX_u = \exp(0.5\beta_{u2}^2)\Phi(\beta_{u2}+W\alpha), \tag{B.72}
$$

and

$$
\int_{-\infty}^{-W\alpha}\exp(X_u\beta_{u2})\varphi(X_u)dX_u = \exp(0.5\beta_{u2}^2)\left[1-\Phi(\beta_{u2}+W\alpha)\right]. \tag{B.73}
$$

Combining equations (E.5), (E.9), and (E.10) yields equation (26).

APPENDIX C

TECHNICAL APPENDIX FOR CHAPTER IV

The generalized Hausman tests mentioned in Chapter IV require that I estimate joint variance-covariance (VCV) matrices for fixed effect (FE) and random effect (RE) estimators. The basic approach I take is to conceptualize the independent estimators as solving a "stacked" system of equations. With this "stacked" system of equations standard White's (1982) sandwich estimator for the joint VCV matrix can be used. The covariance blocks of the resultant joint VCV matrix account for correlation between the two models that exists because the same, or an overlapping, sample is used by both models.

Let $\Theta^T = \left[\alpha^T \quad \beta^T \right]$ denote the stacked parameter vector associated with a FEP model and a REP model. Let x_{1it} and x_{2it} generically denote the vectors of regression controls used in FEP model and the REP model, respectively. Now, define the following two functions:

$$Q_{1i}(\alpha) = \sum_{t=1}^{T_i} y_{it} \ln \left(\frac{\exp(x_{1it}\alpha)}{\sum_{r=1}^{T_i} \exp(x_{1ir}\alpha)} \right),$$

and

$$Q_{2i}(\beta) = \ln\Gamma\left(\eta + \sum_{t=1}^{T_i} y_{it}\right) - \sum_{t=1}^{T_i} \ln\Gamma\left(1 + y_{it}\right) - \ln\Gamma\left(\eta\right) + \eta \ln\left(u_i\right)$$
$$+ \ln\left(1 - u_i\right)\sum_{t=1}^{T_i} y_{it} + \sum_{t=1}^{T_i} y_{it} \exp(x_{2it}\beta) - \left(\sum_{t=1}^{T_i} y_{it}\right)\ln\left(\sum_{t=1}^{T_i} \exp(x_{2it}\beta)\right)$$

where T_i denotes the maximum number of observations for individual i, η denotes a gamma "variance" parameter, $\Gamma(\ )$ denotes the gamma function, and where

$$u_i = \frac{\eta}{\eta + \sum_{t=1}^{T_i} \exp(x_{2it}\beta)}$$

Finally, let

$$Q_i\left(\Theta\right) = Q_{1i}\left(\alpha\right) + Q_{2i}\left(\beta\right)$$

A legitimate estimator for Θ can be written as

$$\widehat{\Theta} = \underset{\Theta}{\operatorname{argmax}} \sum_{i=1}^{n} Q_i(\Theta) \qquad \text{(C.1)}$$

Problem (C.1) is separable in α and β. To see this note that

$$\nabla_{\Theta} Q_i = \left[\nabla_{\alpha} Q_{1i} \quad \nabla_{[\beta\ \eta]} Q_{2i}\right], \qquad \text{(C.2)}$$

where $\nabla_{\kappa} Q$ denotes the gradient vector of Q with respect to κ, and has row dimensions equal to one and column dimensions equal to the rows of κ. Therefore, the solution to (C.1) is given by the solutions to the two systems of equations:

$$G_1(\alpha) = \sum_{i=1}^{n} \nabla_{\alpha} Q_{1i} = 0, \tag{C.3}$$

and

$$G_2(\beta) = \sum_{i=1}^{n} \nabla_{[\beta\ \eta]} Q_{2i} = 0. \tag{C.4}$$

But (C.3) and (C.4) are the estimating equations associated with the FEP estimator for α and β conducted in isolation of one another.

This discussion serves to point out that problem (C.1) is equivalent to estimating a FEP model and REP model in isolation. Furthermore, the estimator described by (C.1) has a well-known VCV matrix estimator. Let

$$G(\Theta) = \sum_{i=1}^{n} \nabla_{\Theta} Q_i = 0$$

A robust estimator for the VCV of (C.1) is

$$\widehat{\mathrm{Var}}\left(\widehat{\Theta}\right) = \left\{ \left. \frac{\partial G(\Theta)}{\partial \Theta} \right|_{\Theta=\widehat{\Theta}} \right\}^{-1} \sum_{i=1}^{n} \nabla_{\Theta} Q_i \Big|_{\Theta=\widehat{\Theta}} \nabla_{\Theta} Q_i^{T} \Big|_{\Theta=\widehat{\Theta}} \left\{ \left. \frac{\partial G(\Theta)}{\partial \Theta} \right|_{\Theta=\widehat{\Theta}} \right\}^{-1} \tag{C.5}$$

The gradient of Q_{1i} is[27]

$$\nabla_{\alpha} Q_{1i} = \sum_{t=1}^{T_i} x_{1it} \left\{ y_{it} - \frac{\exp\left(x_{1it}\alpha \right)}{\sum_{r=1}^{T_i} \exp\left(x_{1ir}\alpha \right)} \sum_{r=1}^{T_i} y_{ir} \right\}. \tag{C.6}$$

The gradient of Q_{2i} is[28]

$$\nabla_{\beta} Q_{2i} = \sum_{t=1}^{T_i} x_{2it} \left\{ y_{it} - \exp(x_{2it}\beta) \frac{\sum_{r=1}^{T_i} y_{ir} + \eta}{\sum_{r=1}^{T_i} \exp(x_{2it}\beta) + \eta} \right\}. \tag{C.7}$$

The remaining derivative is

$$\nabla_{\eta} Q_{2i} = \psi\left(\eta + \sum_{t=1}^{T_i} y_{it} \right) - \psi\left(\eta\right) + \ln(u_i) + \frac{\sum_{t=1}^{T_i} \exp(x_{2it}\beta) - \sum_{t=1}^{T_i} y_{it}}{\eta + \sum_{t=1}^{T_i} \exp(x_{2it}\beta)}, \tag{C.8}$$

where $\psi(\ )$ denotes the digamma function. Combining (C.2), (C.6), (C.7), and (C.8) yields the gradient of Q_i. The inverse of the Hessian can be obtained as the standard variance-covariance matrix that Stata estimates for each of the models.

Under standard arguments Θ is asymptotically normal with variance approximated consistently by (C.5). Thus the Wald statistic

[27] See Cameron and Trivedi (2005) page 803.
[28] See Cameron and Trivedi (2005) page 805.

$$W = \left(R\widehat{\Theta} \right)^{T} \left[R\widehat{Var}\left(\widehat{\Theta}\right) R^{T} \right]^{-1} \left(R\widehat{\Theta} \right) \tag{C.9}$$

has a chi-squared distribution with k degrees of freedom, where k is equal to the number of rows in the matrix R, or equivalently the number of linear restrictions tested. The above procedures can be easily revised to test other hypotheses as well, such as linear restrictions across two FE models.

Table C1

Complete Inpatient Stay Results

	Total Effects of Part D[a]			Direct Effects[b]
	FEP	REP[c]	REP[d]	FEP
Limited Pre-Part D Coverage	-0.14**	-0.06	-0.05	-0.14
	(0.06)	(0.05)	(0.05)	(0.09)
	[-13.02]	[-5.53]	[-4.96]	[-12.76]
Generous Pre-Part D Coverage	0.03	0.01	0.05	0.04
	(0.07)	(0.05)	(0.06)	(0.09)
	[2.53]	[1.03]	[5.22]	[3.93]
Age	-0.02	0.01***	0.02***	-0.13
	(0.06)	(0.00)	(0.00)	(0.08)
Log(Household Income)	-0.04*	-0.07***	-0.08***	-0.02
	(0.02)	(0.02)	(0.02)	(0.03)
Married	0.01	-0.06	-0.06*	0.09
	(0.07)	(0.03)	(0.03)	(0.08)
High School Graduate		-0.08**	-0.08**	
		(0.04)	(0.04)	
Some College		-0.03	-0.06	
		(0.05)	(0.05)	
College or Beyond		-0.06	-0.08*	
		(0.05)	(0.05)	
Nonwhite		-0.06	-0.08*	
		(0.04)	(0.05)	
Female		-0.15***	-0.17***	
		(0.03)	(0.03)	
Other Drug Coverage	-0.01	-0.01	-0.03	0.03
	(0.04)	(0.03)	(0.04)	(0.05)
HMO	-0.02	-0.06*	-0.07*	-0.04
	(0.05)	(0.04)	(0.04)	(0.06)
Medicaid	-0.17**	0.01	0.04	-0.09
	(0.07)	(0.05)	(0.05)	(0.13)
CHAMPUS	-0.01	-0.05	0.09	0.04
	(0.08)	(0.05)	(0.09)	(0.09)
Medigap	-0.01	-0.03	-0.00	-0.01
	(0.04)	(0.03)	(0.03)	(0.05)
ADL (Baseline)	0.01	0.07***	0.09***	0.01
	(0.02)	(0.02)	(0.02)	(0.03)
IADL (Baseline)	0.01	0.05	0.07*	0.04
	(0.05)	(0.04)	(0.03)	(0.14)
CESD (Baseline)	0.01	0.03***	0.04***	0.02
	(0.01)	(0.01)	(0.01)	(0.01)
Active Conditions (Baseline)	-0.01	0.15***	0.20***	-0.00
	(0.03)	(0.01)	(0.02)	(0.03)
ADL (Follow-Up)				0.05*
				(0.03)
IADL (Follow-Up)				0.14**
				(0.06)
CESD (Follow-Up)				0.02*
				(0.01)

Active Conditions (Follow-Up)				0.23***
				(0.03)
Northeast	-0.03	0.09	0.11*	-0.11
	(0.30)	(0.06)	(0.06)	(0.41)
Midwest	-0.06	0.11***	0.17***	-0.14
	(0.25)	(0.04)	(0.05)	(0.37)
South	-0.14	0.07*	0.09**	-0.06
	(0.23)	(0.04)	(0.04)	(0.33)
1996-1998 (Base: 2004-2006)	-1.09**	-0.67***	-0.54***	-1.82***
	(0.49)	(0.10)	(0.11)	(0.58)
1998-2002	-0.79**	-0.45***	-0.34***	-1.32***
	(0.37)	(0.07)	(0.07)	(0.44)
2000-2002	-0.50**	-0.27***	-0.19***	-0.86***
	(0.25)	(0.04)	(0.04)	(0.30)
2002-2004	-0.24*	-0.12***	-0.07***	-0.41***
	(0.13)	(0.02)	(0.02)	(0.15)
2006-2008	0.24*	0.12***	0.07***	0.41***
	(0.13)	(0.02)	(0.02)	(0.15)
2008-2010	0.50**	0.27***	0.19***	0.86***
	(0.25)	(0.04)	(0.04)	(0.30)
Observations:	19495	19495	30650	11689
Respondents:	5042	5042	10075	3193
FEP v. REP:[e]		221.59	373.10	
p value:		{0.00}	{0.00}	
REP v. REP:[f]			324.11	
p value:			{0.00}	
Limited Pre-Part D Coverage:[g]				0.00
p value:				{0.97}
Generous Pre-Part D Coverage:[h]				0.04
p value:				{0.84}
Health (Follow-Up):[i]				97.61
p value:				{0.00}

Cluster-robust standard errors are in parentheses

Transformed incidence rate ratios (i.e., {IRR-1} × 100) in square brackets; the transformed IRRs represent the percentage impact of the covariate

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[a] The total effects of Part D measure the effects of Part D that operate through changes in beneficiary health plus any direct effects.

[b] The direct effects of Part D measure the effects of Part D that do not operate through changes in beneficiary health.

[c] Column reports results from a random effects Poisson estimator applied to the fixed effects sample.

[d] Column reports results from a random effects Poisson estimator applied to the full sample.

[e] Tests the equality of the common coefficients across the FEP and REP models.

[f] Tests the equality of the coefficients across the two REP models.

[g] Tests the equality of the Part D effect among beneficiaries who had limited or no pre-Part D drug coverage.

[h] Tests the equality of the Part D effect among beneficiaries who had generous pre-Part D drug coverage.

[i] Tests the joint significance of ADL (Follow-Up), IADL (Follow-Up), CESD (Follow-Up), and Active Conditions (Follow-Up).

Table C2

Complete Hospital Night Results

	Total Effects of Part D[a]			Direct Effects[b]
	FEP	REP[c]	REP[d]	FEP
Limited Pre-Part D Coverage	-0.18*	-0.16	-0.17	-0.14
	(0.11)	(0.10)	(0.10)	(0.16)
	[-16.52]	[-14.58]	[-15.31]	[-12.87]
Generous Pre-Part D Coverage	-0.25*	-0.20*	-0.21*	-0.15
	(0.13)	(0.12)	(0.12)	(0.16)
	[-21.91]	[-18.53]	[-18.84]	[-13.55]
Age	0.23**	0.04***	0.05***	-0.04
	(0.11)	(0.00)	(0.01)	(0.12)
Log(Household Income)	-0.05	-0.06**	-0.06**	0.02
	(0.03)	(0.03)	(0.03)	(0.04)
Married	-0.03	-0.09	-0.09	-0.24
	(0.10)	(0.08)	(0.08)	(0.15)
High School Graduate		-0.13*	-0.30***	
		(0.07)	(0.09)	
Some College		-0.07	-0.26**	
		(0.10)	(0.12)	
College or Beyond		-0.19**	-0.46***	
		(0.09)	(0.10)	
Nonwhite		0.19*	0.32**	
		(0.10)	(0.13)	
Female		-0.09	-0.17**	
		(0.06)	(0.07)	
Other Drug Coverage	-0.13*	-0.13*	-0.13*	-0.04
	(0.08)	(0.08)	(0.08)	(0.11)
HMO	-0.03	-0.05	-0.04	-0.09
	(0.09)	(0.08)	(0.08)	(0.10)
Medicaid	-0.18	-0.14	-0.15	0.03
	(0.13)	(0.12)	(0.12)	(0.17)
CHAMPUS	0.14	0.09	0.12	0.27*
	(0.14)	(0.13)	(0.13)	(0.16)
Medigap	0.05	0.02	0.02	0.03
	(0.08)	(0.08)	(0.08)	(0.11)
ADL (Baseline)	-0.03	-0.00	-0.00	0.08
	(0.04)	(0.04)	(0.04)	(0.06)
IADL (Baseline)	-0.16**	-0.13*	-0.12*	0.02
	(0.08)	(0.07)	(0.07)	(0.12)
CESD (Baseline)	0.02	0.02	0.03	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)
Active Conditions (Baseline)	0.00	0.04	0.04	0.00
	(0.04)	(0.03)	(0.04)	(0.05)
ADL (Follow-Up)				0.14***
				(0.05)
IADL (Follow-Up)				-0.14
				(0.14)
CESD (Follow-Up)				0.06***
				(0.02)

Active Conditions (Follow-Up)				0.23***
				(0.05)
Northeast	-0.48	0.10	0.09	-0.52
	(0.49)	(0.16)	(0.20)	(0.65)
Midwest	-1.01**	-0.11	-0.22	-1.23*
	(0.46)	(0.18)	(0.22)	(0.71)
South	-0.36	0.06	0.03	-0.23
	(0.41)	(0.15)	(0.18)	(0.59)
1996-1998 (Base: 2004-2006)	0.44	-0.85***	-0.74***	-1.32
	(0.84)	(0.20)	(0.20)	(0.88)
1998-2002	0.40	-0.59***	-0.51***	-0.93
	(0.64)	(0.13)	(0.13)	(0.67)
2000-2002	0.31	-0.37***	-0.30***	-0.58
	(0.44)	(0.08)	(0.08)	(0.46)
2002-2004	0.18	-0.17***	-0.14***	-0.27
	(0.22)	(0.03)	(0.03)	(0.23)
2006-2008	-0.18	0.17***	0.14***	0.27
	(0.22)	(0.03)	(0.03)	(0.23)
2008-2010	-0.31	0.37***	0.30***	0.58
	(0.44)	(0.08)	(0.08)	(0.46)
Observations:	19495	19495	30650	11679
Respondents:	5042	5042	10075	3190
FEP v. REP:[e]		218.42	208.83	
p value:		{0.00}	{0.00}	
REP v. REP:[f]			103.43	
p value:			{0.00}	
Limited Pre-Part D Coverage:[g]				0.11
p value:				{0.73}
Generous Pre-Part D Coverage:[h]				0.67
p value:				{0.41}
Health (Follow-Up):[i]				50.46
p value:				{0.00}

Cluster-robust standard errors are in parentheses

Transformed incidence rate ratios (i.e., {IRR-1} × 100) in square brackets; the transformed IRRs represent the percentage impact of the covariate

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[a] The total effects of Part D measure the effects of Part D that operate through changes in beneficiary health plus any direct effects.

[b] The direct effects of Part D measure the effects of Part D that do not operate through changes in beneficiary health.

[c] Column reports results from a random effects Poisson estimator applied to the fixed effects sample.

[d] Column reports results from a random effects Poisson estimator applied to the full sample.

[e] Tests the equality of the common coefficients across the FEP and REP models.

[f] Tests the equality of the coefficients across the two REP models.

[g] Tests the equality of the Part D effect among beneficiaries who had limited or no pre-Part D drug coverage.

[h] Tests the equality of the Part D effect among beneficiaries who had generous pre-Part D drug coverage.

[i] Tests the joint significance of ADL (Follow-Up), IADL (Follow-Up), CESD (Follow-Up), and Active Conditions (Follow-Up).

Table C3

Complete Results for Active Health Conditions[a]

	FEP	REP[b]	REP[c]
Limited Pre-Part D Coverage	-0.09*	-0.04	-0.01
	(0.05)	(0.04)	(0.04)
	[-8.46]	[-4.36]	[-0.71]
Generous Pre-Part D Coverage	-0.09**	0.02	0.10**
	(0.05)	(0.04)	(0.04)
	[-9.00]	[1.63]	[10.84]
Age	0.03	-0.00	0.01***
	(0.04)	(0.00)	(0.00)
Log(Household Income)	0.01	-0.05***	-0.08***
	(0.02)	(0.01)	(0.01)
Married	-0.02	-0.09***	-0.10***
	(0.05)	(0.03)	(0.03)
High School Graduate		-0.12***	-0.18***
		(0.03)	(0.04)
Some College		-0.19***	-0.23***
		(0.04)	(0.05)
College or Beyond		-0.18***	-0.33***
		(0.05)	(0.05)
Nonwhite		-0.11***	-0.07*
		(0.04)	(0.04)
Female		-0.02	0.06*
		(0.03)	(0.03)
Other Drug Coverage	-0.01	-0.03	-0.06**
	(0.03)	(0.03)	(0.03)
HMO	-0.01	-0.09***	-0.04
	(0.03)	(0.03)	(0.03)
Medicaid	-0.01	0.19***	0.20***
	(0.05)	(0.04)	(0.04)
CHAMPUS	0.06	0.08	0.10*
	(0.06)	(0.05)	(0.05)
Medigap	0.01	-0.08***	-0.06**
	(0.03)	(0.03)	(0.03)
Northeast	-0.36	-0.03	-0.02
	(0.26)	(0.04)	(0.05)
Midwest	-0.14	-0.01	0.03
	(0.20)	(0.04)	(0.04)
South	-0.25	0.06*	0.07*
	(0.20)	(0.04)	(0.04)
1996-1998 (Base: 2004-2006)	-0.47	-0.22***	0.02
	(0.32)	(0.08)	(0.08)
1998-2000	-0.33	-0.15***	0.01
	(0.24)	(0.05)	(0.05)
2000-2002	-0.20	-0.09***	0.01
	(0.16)	(0.03)	(0.03)
2002-2004	-0.09	-0.04***	0.01
	(0.08)	(0.01)	(0.01)
2006-2008	0.09	0.04***	-0.01

	(0.08)	(0.01)	(0.01)
2008-2010	0.20	0.09***	-0.01
	(0.16)	(0.03)	(0.03)
Observations:	11609	11609	20961
Respondents:	3329	3329	7511
FEP v. REP:[d]		497.01	611.33
p value:		{0.00}	{0.00}
REP v. REP:[e]			147.46
p value:			{0.00}

Cluster-robust standard errors are in parentheses

Transformed incidence rate ratios (i.e., {IRR-1} × 100) in square brackets; the transformed IRRs represent the percentage impact of the covariate

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[a] Active health conditions include: hypertension (not under control), diabetes (not under control), cancer (in treatment), lung disease (limits activity), stroke (problems persist), arthritis (limits activity), myocardial infarction, congestive heart failure, angina (limits activity), and psychiatric disorder (in treatment).

[b] Column reports results from a random effects Poisson applied to the fixed effects sample.

[c] Column reports results from a random effects Poisson applied to the full sample.

[d] Tests the equality of the coefficients in the FEP and REP models.

[e] Tests the equality of the coefficients in the two REP models.