

# **The Effect of Extended Medicaid on the Emergency Department Visits: An analysis of the 2008 Oregon Health Program**

**Myoung-Jin Keay  
South Dakota State University**

*I analyze the effect of the extended Medicaid program on the emergency department use behavior by using the data from the Oregon Health Program conducted in 2008. Although previous literature report that the Medicaid increases the probability of ED use, they fail to properly account for the endogeneity due to an underidentification. Our contribution is twofold. First, by a copula regression, we achieved a flexibility by using different joint distributions other than multivariate normal. Our result shows that the model performs better under the nonconventional likelihood function. Second, more importantly, we achieved identification of the parameters of interest by using the copula decomposition under the conditional independence assumption. Although our result agrees with the previous research that the Medicaid indeed increases the chance of ED use in Oregon, the average partial effect is estimated to be lower than the earlier estimates.*

*Keywords: Medicaid, Endogenous Explanatory Variable, Copula*

## **INTRODUCTION**

This article explores the effects of an insurance on the emergency department (ED) visit behavior by using the data from the Oregon health experiment. Specifically we are interested in whether the extended Medicaid increased or decreased the number of ED visits during the Oregon Health Program. Those without Medicaid or an insurance tend to rely more on ED, because they will be treated even without an insurance. The medical costs can be reduced if Medicaid encourages them to use more of the preventive care and less of ED. However, it is generally difficult to confirm the effect of a medical insurance on the frequency of ED visits. The estimated effects are unclear and mixed in various randomized experiments. Anderson et al. (2012, 2014) and Newhouse (1993) find that coverage increased

ED use while Chen et al. (2011) and Miller (2012) report the opposite. By using the data from the present Oregon Health Program, Taubman et al. (2014) show that, by two-stage least squares (2SLS), the coverage increased the ED visits for the period of experiment. With the same data, Keay (2019) finds that the average treatment effect (ATE), by an exponential regime switching regression, is in effect not different from zero. There are a couple of issues in their works. The insurance variable is endogenous, because it might be correlated with other factors determining the ED visits. To solve this problem, Taubman et al. (2014) use the lottery eligibility as an instrumental variable (IV). Unfortunately, their model contains another potential endogenous variable that also requires an additional IV. Their model might suffer from an underidentification, and the coefficient estimates are dubious. We will explore a new approach to identification proposed by Keay (2019).

Another issue is that the number of ED visits are top-coded. To avoid the censoring issue, we will use the indicators of ED visits instead of using the number of visits. Thus the model that this paper considers is very similar to the trivariate probit models. In order to relax the typical assumption of the multivariate normal distribution, we will use the copula approach proposed by Winkelmann (2010) and Keay (2016).

Section 2 describes the design of the “Oregon Health Plan Standard”, an experiment conducted by the state of Oregon. Section 3 discusses how the data was collected from the experiment. We discuss the model and the assumptions used for copula regression in Section 4. The estimation results are provided in Section 5. Section 6 is a concluding remark.

## **OREGON HEALTH PLAN STANDARD**

In 2008, the state of Oregon ran an experiment called “Oregon Health Plan Standard”, which extended the Medicaid for the low-income adults. About 30,000 out of 90,000 on the Medicaid waiting list were provided with the eligibility to apply for the Medicaid. An applicant is given the Medicaid if she satisfies the criteria: she must be a citizen or a legal immigrant aged between 19-64 without any medical insurance for the previous six months, and is not otherwise eligible for Medicaid. The income must be below the federal poverty level (which is \$10,400 for an individual and \$21,200 for a family of four in 2008), and the asset less than \$2,000. The OHP covers the prescription drugs and the monthly premium is \$0-20. The lottery drawings were conducted eight times from March through September 2008. Each winner is required to submit an application to obtain the Medicaid. The Medicaid was finally given to the applicants satisfying all the above criteria.

## **DATA**

We use the same data set that Taubman et al. (2014) use. For the analysis of the OHP, they combined four data sets by matching the individuals from each data set. They are ED records for 2007-2008 from the 12 hospitals in the Portland area, the data from the OHP Standard, prerandomization demographic information provided by the participants upon signing up the lottery, and the survey data from the participants. Although the experiment was performed throughout the entire state, only the 12 hospitals in the Portland area kept the ED visit data. Therefore they reduce the sample by using the observations for the area served by the 12 hospitals. The reduced data set has 24,646 observations, roughly 1/3 of the original data. This reduced data set covers the period of Jan 2007-Sep 2009. The ED records in Portland start from Jan 2007. The lottery was conducted from Mar 2008 until Sep 2008. The survey was completed in Sep 2009. Thus the so-called prerandomization period is Jan 2007-Mar 2008, and the study period is Mar 2008-Sep 2009. The ED visits for the study period refer to the ones during this period.

## **MODEL**

We analyze the effect of the Medicaid on the ED use behavior. Let *ED*, *PED* and *MEDICAID* denote the indicators of ED visits in the study period, ED visits in the prerandomization and Medicaid. *LOTTERY* is the indicator of whether an individual is a lottery winner. It is one if she is a winner and zero, otherwise. For ED visit data, Taubman et al. (2014) use both the numerical variables that show the number of visits along with the indicators that only capture whether an individual ever visited an ED. We will use the binary indicators only, since the numerical variable is censored and can cause an inconsistency.

They also include the family size as another covariate, because the every family member is regarded as a lottery winner if one of them is. Thus a member in a larger family is more likely to be a winner. Family size is included in order to control for this. Taubman et al. (2014) use the 2SLS to find the causal effect of Medicaid with *LOTTERY* as an IV for Medicaid. They implicitly assume that there exists only one endogenous variable. However, if *PED* is also endogenous, their regression suffers from the underidentification. I will explicitly consider a triangular model as below.

$$ED_{ih} = 1[\beta_0 + \beta_1 MEDICAID_{ih} + \beta_2 PED_{ih} + X_{ih}\beta_3 - u_{ih} > 0] \quad (1)$$

$$MEDICAID_{ih} = 1[\delta_{10} + \delta_{11} LOTTERY_h + X_{ih}\delta_{12} - \epsilon_{ih} > 0] \quad (2)$$

$$PED_{ih} = 1[\delta_{20} + X_{ih}\delta_{21} - e_{ih} > 0] \quad (3)$$

where the subscript  $i$  and  $h$  denote an individual and a household which the individual belongs to. The first equation is the outcome equation that we are mainly interested in. This equation is a binary choice model, where an individual visits ED if the linear index is greater than  $u_{ih}$ . We view that *MEDICAID* and *PED* are potentially endogenous, and each of them has its own reduced form equation as (2) and (3). *FAMSIZE* is assumed to be exogenous and enters on all the equations. *LOTTERY* must be highly correlated with the actual receipt of Medicaid, but is independent of all the errors. This justifies its use as an IV for Medicaid. However, it cannot be an IV for *PED*, and hence does not enter the equation (3). The endogeneity of *MEDICAID* and *PED* is implied by  $\text{Cov}(u, \epsilon) \neq 0$  and  $\text{Cov}(u, e) \neq 0$ . The identification conditions for such a model is provided by Keay (2019). We will use them as assumptions. Let  $Y_0, Y_1$  and  $Y_2$  denote *ED*, *MEDICAID* and *PED*, respectively. Also  $X$  denotes the vector of all covariates.

**Assumption 1**  $Y_1 \perp Y_2 \mid Y_0, X$ , where  $\perp$  denotes statistical independence.

**Assumption 2**  $C_{u\epsilon}(F_u(u), F_\epsilon(\epsilon))$  and  $C_{ue}(F_u(u), F_e(e))$  are strictly increasing in joint distribution, where  $C_{u\epsilon}$ ,  $C_{ue}$ , and  $F$  are the copula functions of  $u$  and  $\epsilon$ , of  $u$  and  $e$ , and the marginal cumulative distribution function (CDF) of the specified random variables.

Assumption 1 implies that the dependence among the EEVs disappears conditional on the dependent variable and other covariates, equivalently conditional on the error. Each EEV is correlated with  $u$  by definition, and thus they might be correlated with one another. Consequently a large part of their dependence might vanish once conditioned on the common error that feeds the dependence among them. If there is any remaining dependence, they can be further eliminated by adding additional covariates. So  $X$  includes the possible additional covariates as well. This assumption can be easily tested. Keay (2019) suggests an OLS regression of one of the EEVs on the other EEV and the dependent variable along with the covariates. If the coefficient of the other EEV is significant, the independence assumption is rejected. In that case, putting additional covariates that can make it insignificant is required. This test has a limitation, though. A failure of rejection does not necessarily imply independence. The rejection simply says that they are uncorrelated. The test does not fully eliminate the need to assume the independence, since they can be uncorrelated and dependent at the same time. In that case, the power of the test might be small. However, in practice, it might be a useful indication of independence if the test is not rejected.

Assumption 2 is originally from Han and Vytlačil (2017). They show that the bivariate probit model with a dummy endogeneous regressor can be identified if the errors have a copula that is stochastically increasing in joint distribution and an IV enters the reduced form equation. Keay (2019) shows that by using the copula decomposition the model with two endogenous regressors can be identified if each bivariate copula follows the same assumption as above, and there is only one valid IV. There are some well known copulae that satisfy the property: Gaussian, Plackett, Clayton and Frank. If the population distribution follows the multivariate distribution, it becomes a trivariate probit model. Also Keay (2019) shows that the model is identified under the same assumptions even if there are two EEVs and one valid IV. The situation is similar with the Heckman correction model, where the model is identified by the nonlinearity of the inverse Mill's ratio even without an IV. Although the number of IVs is smaller than the EEVs, the lack of information is supplemented by the additional information on the joint distribution provided by the copulae. Using copula functions is not so much restrictive because all the different copulae in the class of what the Assumption 2 designates can be used in estimation.

## ESTIMATION

We run regressions by using OLS, 2SLS and copula regressions with Gaussian, Clayton and Frank copulae. Taubman et al. (2014) include *PED* on the right hand side variable. They claim that the inclusion of *PED* serves to improve the precision of estimates without changing the estimation results. Is this true? *PED* has a strongly positive correlation with *MEDICAID*, because those with previous ED experience might have a stronger incentive to have Medicaid. The correlation between *PED* and *MEDICAID* is about .085 and this implies that having *PED* in the equation essentially defines a smaller parameter. We are more interested in the model with *PED*. The inclusion of additional covariates allows us to approach the true effect of the variable of the main interest as long as holding them does not preclude the latter from varying. However, controlling for *PED* creates another issue; *PED* might be endogenous because it might be correlated with what causes an individual to visit ED in the study period. In this case, we are short of at least one IV since the experiment provided *LOTTERY* as the only valid IV. In order to solve these issues, I will use the copula decomposition proposed by Keay (2019).

We need to check if the Assumption 1 is satisfied before applying the copula decomposition. Regression of *MEDICAID* on *PED*, *ED* and *FAMSIZE* gives 5.33 as the t-statistic on *PED*. Conditioning on the errors does not fully eliminate the dependence between *MEDICAID* and *PED*. Now we use *SNAP* as another covariate. This is an indicator of *SNAP* receipt during the study period. An OLS regression of *MEDICAID* on *PED* along with *SNAP* gives .80 as the t-statistic for the coefficient on *PED*. The p-value is .42. This indicates that *SNAP* indeed eliminates a large part of the correlation between *MEDICAID* and *PED*. We will analyze from now on with *SNAP* on all the equations. The inclusion of *SNAP* allows us to identify the partial effect with *PED*, *FAMSIZE* and *SNAP* controlled for.

Table 1 and 2 display the descriptive statistics and the regression results. The OLS coefficient on *MEDICAID* implies that the insurance increases the chance of ED visit by 11.33%, but this is not a consistent estimate because the endogeneity is not properly taken care of. One can see that the 2SLS estimate is smaller than the OLS. Without *SNAP*, Taubman et al. (2014) report that the Medicaid increases the probability of ED use by 7.0%. The current estimate of 5.6% is substantially smaller due to the inclusion of *SNAP*. Nonetheless, the 2SLS estimation might suffer from underidentification if *PED* is also endogenous.

As the main results of this article, Table 2 also reports the copula regression results. By the copula decomposition, the copulae of  $(u, \epsilon)$  and  $(u, e)$  are separately modelled. Han and Vytlačil (2017) show that the bivariate probit is identified if a valid IV enters the reduced form equation and the copula is stochastically increasing in joint distribution, which Gaussian, Plackett, Clayton and Frank copulae satisfy. In estimation, I use all the copulae but Plackett that is not supported by R. In the header, the first and second letters denote the copulae used for  $(u, \epsilon)$  and  $(u, e)$ , respectively. For instance, FC indicate that the Frank and Clayton copulae are used. The results show that the coefficient estimates on *MEDICAID* are all positive around .14-.17. This is the coefficient inside the indicator function. Although the actual partial effect might be different, their signs are the same. The fact that the Medicaid increases the chance of ED visits is not rejected even by the copula regression. One can see that the APE estimates are around 4.7-5.6%. The estimates vary up to 1% as the used distributions change. This gives some idea of the location of the identified set.

The performances of the nine models can be compared by using the information criteria. Since we are not choosing over different sets of covariates, but over the copulae, the criteria simply requires us to compare the likelihood values. The Table 2 lists the results by the order of the likelihood values. It shows that the Clayton copula for  $(u, \epsilon)$  performs the best, and Gaussian does the worst.

There are two dependence parameters.  $r_1$  is the one between  $(u, \epsilon)$  and  $r_2$  is between  $(u, e)$ . The dependence parameter for Gaussian is same as the correlation coefficient. The dependence parameter of Frank copula can take any value but zero. The two random variables are independent as it approaches zero. The dependence parameter of Clayton copula can be on  $[0, \infty)$  and the independence is implied at zero. The Clayton copula cannot account for a negative dependence. Table 2 shows that the  $r_1$  estimates

are all significant, and confirms the endogeneity of *MEDICAID*. However, in all the used copulae  $r_2=0$  is not rejected. The copula regressions show that there is not a strong evidence that *PED* is endogenous.

## CONCLUSION

This article finds the effect of extended Medicaid in Oregon on the probability of ED visits. We use a new approach in this article: it is more flexible due to various copulae used in estimation procedure, and it solves the underidentification by using the copula decomposition. We agree that the extended Medicaid indeed increases the probability of ED visits in Oregon health insurance, but not as much as previously reported.

**TABLE 1  
DESCRIPTIVE STATISTICS**

Variable	mean	Std. dev.	min	max
<i>ED</i>	.343	.474	0	1
<i>PED</i>	.313	.463	0	1
<i>FAMSIZE</i>	1.210	.409	1	3
<i>SNAP</i>	.569	.495	0	1
<i>MEDICAID</i>	.241	.427	0	1
<i>LOTTERY</i>	.391	.488	0	1

**TABLE 2  
REGRESSION RESULTS**

	OLS	2SLS	CF	CG	FG	CC	FF	FC	GF	GG	GC
<i>MEDICAID</i>	.113*** (.006)	.056** (.023)	.152** (.061)	.151** (.061)	.151** (.061)	.165*** (.060)	.148** (.069)	.149** (.069)	.177*** (.069)	.178*** (.069)	.177*** (.068)
<i>PED</i>	.327*** (.006)	.330*** (.006)	.628*** (.181)	.699*** (.178)	.699*** (.178)	.808*** (.128)	.733*** (.259)	.852*** (.141)	.779** (.315)	.925*** (.254)	.870*** (.149)
<i>FAMSIZE</i>	-.072*** (.006)	-.069*** (.006)	-.265*** (.029)	-.257*** (.029)	-.257*** (.029)	-.244*** (.025)	-.254*** (.036)	-.240*** (.026)	-.251*** (.042)	-.234*** (.036)	-.240*** (.026)
<i>SNAP</i>	.119*** (.005)	.128*** (.006)	.451*** (.037)	.436*** (.038)	.436*** (.038)	.411*** (.028)	.435*** (.051)	.409*** (.029)	.421*** (.062)	.392*** (.054)	.402*** (.030)
cons	.235*** (.009)	.240*** (.009)	-.611*** (.071)	-.634*** (.071)	-.634*** (.071)	-.679*** (.05)	-.646*** (.094)	-.689*** (.051)	-.664*** (.111)	-.714*** (.093)	-.697*** (.053)
r1			.226*** (.079)	.225*** (.080)	.225*** (.080)	.207*** (.078)	.674*** (.245)	.675*** (.247)	.099** (.041)	.099** (.041)	.099** (.041)
r2			.859 (.585)	.115 (.104)	.115 (.104)	.085 (.133)	.523 (.829)	.042 (.139)	.378 (1.013)	-.017 (.153)	.024 (.145)
APE			.050** (.020)	.049*** (.019)	.049*** (.019)	.053*** (.019)	.048** (.022)	.047** (.022)	.057*** (.022)	.056*** (.021)	.056*** (.021)
Likelihood			40529.49	40529.93	40529.93	40530.25	40530.68	40530.82	40531.42	40531.46	40531.46

Note: \$MEDICAID\$ shows the coefficient estimates on Medicaid. APE is the average partial effects from the copula regressions. r1 and r2 are the dependence parameters of  $C_{u\epsilon}(F_u(u), F_\epsilon(\epsilon))$  and  $C_{ue}(F_u(u), F_e(e))$ . The standard errors are in parenthesis. The significance at 10%, 5% and 1% levels are indicated by \*, \*\* and \*\*\*, respectively. The negative signs on the likelihood are suppressed to save the space. The copula regression results are in the order of the likelihood.

## REFERENCES

- Anderson, M., Dobkin, C., & Gross, T. (2012). The Effect of Health Insurance Coverage on the Use of Medical Services. *American Economic Journal: Economic Policy*, 4(1), 1-27.
- Anderson, M., Dobkin, C., & Gross T. (2014). The Effect of Health Insurance on Emergency Department Visits: Evidence from an Age-Based Eligibility Threshold. *Review of Economics and Statistics*, 96(1), 189-195.
- Chen, C., Scheffler, G., & Chandra, A. (2011). Massachusetts' Health Care Reform and Emergency Department Utilization. *New England Journal of Medicine*, 365(12), e25.
- Han, S., & Vytlacil, E.J. (2017). Identification in a Generalization of Bivariate Probit Models with Dummy Endogenous Regressors. *Journal of Econometrics*, 41, 63-73.
- Keay, M.J. (2016). Partial Copula Methods for Models with Multiple Discrete Endogenous Explanatory Variables and Sample Selection. *Economics Letters*, 144, 85-87.
- Keay, M.J. (2019). *Identification and Estimation by a Flexible Parametric Method for Models with Discrete Endogenous Explanatory Variables*. Ness School of Management and Economics, South Dakota State University.
- Miller, S. (2012). The effect of insurance on emergency room visits: An analysis of the 2006 Massachusetts health reform. *Journal of Public Economics*, 96, 893-908.
- Newhouse, J.P. & the Insurance Experiment Group. (1993). *Free for All? Lessons from the RAND Health Insurance Experiment*. Harvard Univ. Press, Cambridge, MA
- Taubman, S.L., Allen, H.L., Wright, B.J., Baicker, K., & Finkelstein, A.N. (2014). Medicaid Increases Emergency Department Use: Evidence from Oregon's Health Insurance Experiment. *Science*, 343(6168), 263-268.
- Winkelmann, R. (2012). Copula Bivariate Probit Models: with an Application to Medical Expenditures. *Health Economics*, 21, 1444-1455.