THE RIMINI CENTRE FOR ECONOMIC ANALYSIS

WP 09-45

Markus Jochmann
University of Strathclyde, UK

WHAT BELONGS WHERE? VARIABLE SELECTION FOR ZERO-INFLATED COUNT MODELS WITH AN APPLICATION TO THE DEMAND FOR HEALTH CARE

The *Rimini Centre for Economic Analysis* (RCEA) was established in March 2007. RCEA is a private, non-profit organization dedicated to independent research in Applied and Theoretical Economics and related fields. RCEA organizes seminars and workshops, sponsors a general interest journal *The Review of Economic Analysis*, and organizes a biennial conference: *Small Open Economies in the Globalized World* (SOEGW). Scientific work contributed by the RCEA Scholars is published in the RCEA Working Papers series.

The Rimini Centre for Economic Analysis
Legal address: Via Angherà, 22 – Head office: Via Patara, 3 - 47900 Rimini (RN) – Italy
www.rcfea.org - secretary@rcfea.org

What Belongs Where? Variable Selection for Zero-Inflated Count Models with an Application to the Demand for Health Care

Markus Jochmann*

Department of Economics, University of Strathclyde

markus.jochmann@strath.ac.uk

October 2009

Abstract

This paper develops stochastic search variable selection (SSVS) for zero-inflated count models which are commonly used in health economics. This allows for either model averaging or model selection in situations with many potential regressors. The proposed techniques are applied to a data set from Germany considering the demand for health care. A package for the free statistical software environment R is provided.

JEL classifications: C11, C25, I11

Keywords: Bayesian, model selection, model averaging, count data, zero-inflation, demand for health care

*The author thanks Gary Koop for helpful discussions. The usual disclaimer applies.

1 Introduction

Empirical health economics often involves the analysis of count data variables that have distributions which exhibit a high frequency of zeros. Examples include the number of visits to the doctor or the number of work days missed by individuals with special health care needs. Zero-inflated count models (Mullahy, 1986; Lambert, 1992) account for this data characteristic and are frequently used in the literature. Among others, Grootendorst (1995) analyses prescription drug utilisation, Street et al. (1999) examine data on pharmaceutical utilisation and expenditure in Russia and Böhning et al. (1999) consider caries prevention in dental epidemiology.

The zero-inflated count model increases the probability of observing a zero outcome relative to the underlying count data distribution (this can be a Poisson, negative binomial or Poisson lognormal distribution). It does so by adding a proportion of zeros and reducing the probabilities of the other frequencies by corresponding amounts. The additional probability of observing a zero may be modelled as either a constant or a function of explanatory variables.

The zero-inflated count model can be interpreted as a finite mixture model where one distribution is degenerate with a unit point mass at zero. The model divides the population into two groups: one group for which the outcome is always zero (depending on the context, we could label this group as the *non-users* or the *low-risk group*) and one group for which the outcome is drawn from the underlying count data distribution (the *potential users* or the *normal-risk group*). Thus, each explanatory variable can have an effect on either or both (i) the probability that an individual is a non-user and (ii) the magnitude of the count outcome, given that the individual is a potential user.

In many situations, potential explanatory variables are numerous, including demographic, socio-economic, lifestyle, disease history and medical variables. Assuming k potential regressors and allowing each to be either included in or excluded from each of the two equations (the *regime selection equation* governing the probability of being a non-user and the *count equation* determining the count outcomes), it follows that there are 4^k possible models. However, most studies employing the zero-inflated count model ignore this model uncertainty and include all or a specific subset of the potential explanatory variables in both equations (sometimes the proportion of added zeros is modelled as a constant). Then they proceed as if the selected model

had generated the data, this leads to over-confident inferences and decisions that are more risky than they appear (see, for example, Madigan and Raftery, 1994).

One way around this problem is to apply *stochastic search variable selection* (SSVS), a technique first proposed by George and McCulloch (1993). SSVS is a Bayesian procedure that introduces a prior distribution on all the unknowns, which in this case are the parameters in each of the models and the models themselves. It combines the prior distribution with the data and induces a posterior distribution that accounts for model uncertainty. This posterior distribution identifies "promising" subsets of explanatory variables and can be used for model selection and inference. A big advantage of SSVS is that it avoids the effort of calculating the posterior probabilities of all 4^k models. Instead, it uses Markov chain Monte Carlo (MCMC) methods to analyse the posterior distribution.[1]

The major econometric contribution of this paper is to extend the SSVS approach, so far developed for normal linear models, to the nonlinear zero-inflated count model. In particular, we propose a novel MCMC sampling algorithm that explores the posterior distribution in an efficient manner. The major empirical contribution of this paper lies in the application of the proposed methods to a data set from Germany analysing the determinants of the demand for health care.

In order to facilitate the use of the proposed techniques and the reproduction of the results of this paper, a software package was written for the free statistical software environment R (R Development Core Team, 2009). The package is called `zic` and can be downloaded from the *Comprehensive R Archive Network* (CRAN) on `http://www.R-project.org`.

The rest of the paper is organised as follows. Section 2 shows how SSVS can be applied to a zero-inflated count data model. Section 3 briefly discusses the MCMC methods needed for the analysis with further details provided in the Appendix. Section 4 gives an application to the demand for health care employing a German data set. Section 5 finally concludes.

[1]For a general introduction to Bayesian inference, see Koop (2003). Examples of Bayesian methods applied to count data applications in health economics include Jochmann and León-González (2004) and Deb et al. (2006).

2 SSVS for a Zero-Inflated Poisson Lognormal Model

As discussed in the introduction, a zero-inflated count model mixes a degenerate distribution with a unit point mass at zero with a non-degenerate count distribution. We choose the latter to be a Poisson lognormal distribution which, unlike the Poisson distribution, allows for overdispersion (see, for example, Greene, 2005). Another common alternative to the Poisson lognormal distribution is the negative binomial distribution. However, as we will see below, the Poisson lognormal distribution can be combined with the SSVS framework in a more convenient way and leads to a simpler sampling algorithm.

We consider a latent count variable y_i^* and specify its conditional distribution to be Poisson:

$$y_i^* \sim \text{Poisson}[\exp(\eta_i^*)]. \tag{1}$$

The logarithm of the conditional mean is given as follows:

$$\eta_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i,\ \varepsilon_i \sim \text{N}(0, \sigma^2), \tag{2}$$

where $\boldsymbol{x}_i$ is a $k \times 1$ vector of regressors, $\boldsymbol{\beta}$ the corresponding $k \times 1$ parameter vector, ε_i an error term that captures unobserved heterogeneity and σ^2 the variance of the error term. We can prove that y_i^* follows a Poisson lognormal distribution by integrating out the unobserved heterogeneity. Next we consider another latent variable d_i^* that comes from

$$d_i^* = \boldsymbol{x}_i'\boldsymbol{\delta} + \nu_i,\ \nu_i \sim \text{N}(0, 1), \tag{3}$$

where $\boldsymbol{\delta}$ is a $k \times 1$ parameter vector and ν_i another error term. We do not observe the latent variables but only $y_i = 1(d_i^* > 0)y_i^*$ and $\boldsymbol{x}_i$.[2] That is, if d_i^* is bigger than zero, we observe the Poisson lognormal outcome. Otherwise the observed y_i is equal to zero regardless of the value of the latent y_i^*.

Equations (1) - (3) specify a zero-inflated Poisson lognormal model whose likelihood function is given in the Appendix. Next we choose independent prior distributions for the unknown parameters $\boldsymbol{\beta}$, $\boldsymbol{\delta}$ and σ^2. As outlined above, the prior distributions for $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ shall reflect model uncertainty, i.e. they shall account for

[2] $1(a)$ denotes the indicator function that takes the value 1 if a is true and 0 otherwise.

the fact that we do not know which explanatory variables belong into equations (2) and (3). We tackle this issue using the stochastic search variable selection (SSVS) technique of George and McCulloch (1993). The underlying principle can be explained as follows. A common Bayesian prior for some parameter α is $\alpha \sim \mathrm{N}(0, V_\alpha)$. Larger values of V_α correspond to a relatively non-informative prior, smaller values of V_α shrink α towards zero. The SSVS prior now combines these two extremes by specifying a scale mixture of two Normal distributions: $\alpha \sim (1-\theta)\mathrm{N}(0, V_{\alpha 0}) + \theta \mathrm{N}(0, \mathrm{V}_{\alpha 1})$, where $\mathrm{V}_{\alpha 0}$ denotes a very "small" variance so that α is virtually zero and $\mathrm{V}_{\alpha 1}$ corresponds to a rather "large" variance and thus leads to an uninformative prior. θ is a dummy variable which equals 0 if α is drawn from the first Normal distribution and equals 1 if α is drawn from the second. This prior is called a hierarchical prior since θ is treated as an unknown parameter and estimated in a data-based fashion.

Along these lines we specify the following prior distributions for the components of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$:

$$\beta_j \sim (1-\gamma_j)\mathrm{N}\left(0, \overline{\tau}_{0j}^2\right) + \gamma_j \mathrm{N}\left(0, \overline{\tau}_{1j}^2\right), \quad j = 1, \ldots, k, \tag{4}$$

$$\delta_j \sim (1-\kappa_j)\mathrm{N}\left(0, \overline{\omega}_{0j}^2\right) + \kappa_j \mathrm{N}\left(0, \overline{\omega}_{1j}^2\right), \quad j = 1, \ldots, k, \tag{5}$$

where the vectors $\boldsymbol{\gamma} \equiv (\gamma_1, \ldots, \gamma_k)'$ and $\boldsymbol{\kappa} \equiv (\kappa_1, \ldots, \kappa_k)'$ consist of dummy variables which take the value 1 if the respective variable is included in the model and 0 otherwise. Thus, the prior parameters $\overline{\boldsymbol{\tau}}_0^2 \equiv (\overline{\tau}_{01}^2, \ldots, \overline{\tau}_{0k}^2)'$ and $\overline{\boldsymbol{\omega}}_0^2 \equiv (\overline{\omega}_{01}^2, \ldots, \overline{\omega}_{0k}^2)'$ consist of "small" variances whereas the prior parameters $\overline{\boldsymbol{\tau}}_1^2 \equiv (\overline{\tau}_{11}^2, \ldots, \overline{\tau}_{1k}^2)'$ and $\overline{\boldsymbol{\omega}}_1^2 \equiv (\overline{\omega}_{11}^2, \ldots, \overline{\omega}_{1k}^2)'$ consist of "large" variances. The inclusion probabilities are given by

$$\Pr(\gamma_j = 1) = 1 - \Pr(\gamma_j = 0) = \overline{p}_j, \quad j = 1, \ldots, k, \tag{6}$$

$$\Pr(\kappa_j = 1) = 1 - \Pr(\kappa_j = 0) = \overline{q}_j, \quad j = 1, \ldots, k, \tag{7}$$

where $\overline{\boldsymbol{p}} \equiv (\overline{p}_1, \ldots, \overline{p}_k)'$ and $\overline{\boldsymbol{q}} \equiv (\overline{q}_1, \ldots, \overline{q}_k)'$ are prior parameter vectors. Finally, the prior for the variance σ^2 is an inverse gamma distribution:

$$\sigma^2 \sim \text{Inv-Gamma}\left(\overline{e}, \overline{f}\right). \tag{8}$$

Several approaches to selecting the prior parameters $\overline{\boldsymbol{\tau}}_0^2$, $\overline{\boldsymbol{\tau}}_1^2$, $\overline{\boldsymbol{\omega}}_0^2$ and $\overline{\boldsymbol{\omega}}_1^2$ can be chosen, depending on the actual application. Basically, the "small" variance

prior parameters $\overline{\tau}_{0j}^2$ ($\overline{\omega}_{0j}^2$) should lead to β_j (δ_j) that are essentially zero and the "large" variance prior parameters $\overline{\tau}_{1j}^2$ ($\overline{\omega}_{1j}^2$) should be selected so that β_j (δ_j) are empirically substantive. We use a semi-automatic approach and set $\overline{\tau}_{0j}^2 = c_0\widehat{\text{var}}(\beta_j)$, $\overline{\tau}_{1j}^2 = c_1\widehat{\text{var}}(\beta_j)$, $\overline{\omega}_{0j}^2 = d_0\widehat{\text{var}}(\delta_j)$ and $\overline{\omega}_{1j}^2 = d_1\widehat{\text{var}}(\delta_j)$, where $\widehat{\text{var}}(\beta_j)$ and $\widehat{\text{var}}(\delta_j)$ are estimates of the parameter variances from a preliminary analysis based on a non-informative prior and $c_0 \ll c_1$ and $d_0 \ll d_1$. This choice follows the considerations of George and McCulloch (1993) and George and McCulloch (1997) and we refer to these papers for further justification.

It finally should be noted that applying SSVS can be seen as a computationally feasible way to implement Bayesian model averaging (BMA).[3] Instead of calculating the posterior probabilities of all possible 4^k models, SSVS greatly reduces the amount of computation by stochastically exploring the model space and letting the MCMC sampler average over the models. Furthermore, SSVS can also be used to select a single best model in which the count equation includes only those explanatory variables for which $\Pr(\gamma_j = 1|\text{Data}) > c$ and the regime selection equation only those for which $\Pr(\kappa_j = 1|\text{Data}) > c$, where c is some threshold (e.g. c = 0.5). We will illustrate both uses in the application below.

3 Bayesian Computation

Model inference is based on the posterior distribution which is, according to Bayes theorem, proportional to the product of the likelihood function [equations (1) - (3)] and the prior distribution [equations (4) - (8)]. Since the posterior distribution is too complex to be analysed analytically we resort to MCMC sampling techniques (see Liu (2001) or Robert and Casella (2004) for comprehensive surveys on these methods). In particular we use the Gibbs sampler for summarising features of the posterior model space. The Gibbs sampler draws a large posterior sample by successively sampling from the conditional distributions of the model parameters. In order to keep computations simple, we apply the data augmentation technique put forward by Tanner and Wong (1987). This means that we include the latent variables $\boldsymbol{y}^* \equiv (y_1^*, \ldots, y_n^*)'$, $\boldsymbol{\eta}^* \equiv (\eta_1^*, \ldots, \eta_n)'$ and $\boldsymbol{d}^* \equiv (d_1^*, \ldots, d_n^*)'$ in the parameter space. In this way we end up with conditional distributions that take convenient functional forms.

[3] See Hoeting et al. (1999) for an introduction to BMA.

Details about the proposed algorithm can be found in the Appendix.

To facilitate the application of SSVS to zero-inflated count data applications, this paper comes with software that implements the proposed algorithm for the free statistical software environment R. The corresponding package is called `zic` and can be downloaded from the *Comprehensive R Archive Network* (CRAN) on `http://www.R-project.org`. It also includes the data set used in this paper, allowing for the reproduction of our results.[4]

4 Application

We illustrate the proposed methods by applying them to a sub-sample of a data set originally used by Riphahn et al. (2003) to analyse the demand for health care.[5] The data set stems from the German Socio-Economic Panel Study (SOEP, see Wagner et al., 2007) and is an unbalanced panel of 7,293 individual families observed from one to seven times. Like in the original study, we restrict our sample to West German men aged 25 through 65 who are German nationals. In addition, we only use the last wave from 1994 and focus on male individuals. This leaves us with 1812 observations. Variable descriptions along with summary statistics are given in Table 1.[6]

Whereas Riphahn et al. (2003) estimate a bivariate model and consider both the number of doctor visits and the number of hospital visits as dependent variables, we focus on analysing only the number of doctor visits in the last three months. Figure 1 shows a histogram of the dependent variable. We clearly see a large number of zeros which cannot be modelled adequately with a Poisson model; thus, the use of a zero-inflated count model seems warranted.

The parameters for the SSVS prior are chosen using the semi-automatic approach outlined above. Thus, we base the prior distributions for $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ on a preliminary MCMC run using a non-informative prior and set $c_0 = d_0 = 0.01$ and $c_1 = d_1 = 10$. We set each element of both $\overline{\boldsymbol{p}}$ and $\overline{\boldsymbol{q}}$ to 0.5 which correspondents to a prior belief

[4]To speed up computations, large parts of the R package (especially the MCMC samplers) are coded in C++.

[5]The original data are available for download on the *Journal of Applied Econometrics Data Archive* website (`http://econ.queensu.ca/jae/`). The version of the data set that is used in this application is also included in the R package that accompanies this article.

[6]Like Greene (2005), who uses the same data set, we changed all observations on health that were recorded between 6 and 7 to 7.

Variable	Description	Mean	Sd
docvis	number of doctor visits in last 3 months	2.958	5.224
age	age	41.653	11.583
agesq	age squared / 1000	1.869	1.013
health	health satisfaction, 0 (low) - 10 (high)	6.838	2.189
handicap	1 if handicapped, 0 otherwise	0.119	0.324
hdegree	degree of handicap in percentage points	6.165	18.491
married	1 if married, 0 otherwise	0.694	0.461
schooling	years of schooling	11.830	2.494
hhincome	household monthly net income, in German marks / 1000	4.517	2.126
children	1 if children under 16 in the household, 0 otherwise	0.388	0.487
self	1 if self employed, 0 otherwise	0.084	0.278
civil	1 if civil servant, 0 otherwise	0.109	0.312
bluec	1 if blue collar employee, 0 otherwise	0.312	0.464
employed	1 if employed, 0 otherwise	0.831	0.375
public	1 if public health insurance, 0 otherwise	0.847	0.360
addon	1 if add-on insurance, 0 otherwise	0.018	0.134

Table 1: Variable descriptions and summary statistics

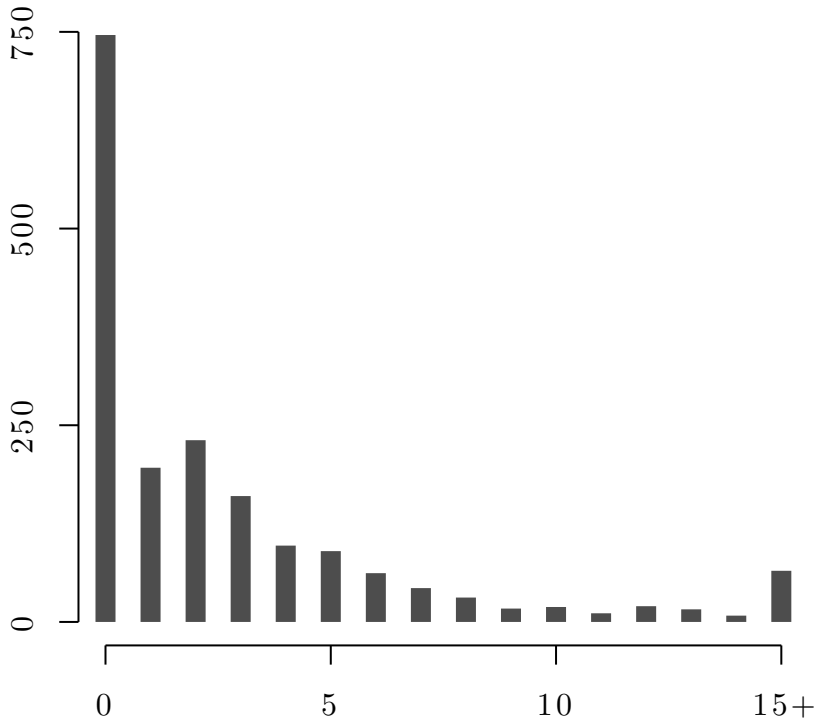

Figure 1: Histogram of number of doctor visits in the last three months (docvis)

that in both equations each explanatory variable is equally likely to be included as excluded. Finally, we set $\overline{e} = 2$ and $\overline{f} = 3$. It follows that both the prior mean and the prior variance for σ^2 are equal to 1.

Table 2 gives posterior means, posterior standard deviations and posterior inclusion probabilities for each coefficient in the model.[7] Looking at the count equation we see that only 5 out of the potential 16 coefficients have an inclusion probability which is greater than 0.5. In the regime selection equation this pattern is even more pronounced, here only the constant, the health satisfaction index (health) and the coefficient dummy variable for children in the household (children) are included in the model. Thus, SSVS chooses a very parsimonious model specification. Riphahn et al. (2003) were particularly interested in the role of the choice of private insurance on health care demand. We see that neither the dummy variable indicating public health insurance (public) nor the dummy variable for add-on insurance (addon) are included in the two equations.

Next, we look at a typical individual, which we define as follows: he is 40 years old, has a health satisfaction index of 7 and no handicap, is married, went to school for 12 years and has a monthly household income of 5,000 German marks. He has no children in the household, is not self-employed, neither a civil servant nor a blue collar worker, is employed and publicly insured and did not purchase add-on insurance. Figure 2 shows posterior predictive distributions for the number of doctor visits (docvis) based on three different variants of the model. The circles give the predictive probabilities based on the zero-inflated count model (labelled ZIC in the figures) presented in section 2 but without the SSVS prior. In this model, all coefficients are included in both equations. The triangles show the predictive probabilities for the SSVS model (labelled SSVS) which averages over the set of possible models. Finally, crosses give the predictive density derived from the "best" model (labelled BEST). Based on a previous SSVS analysis we included only the coefficients with posterior inclusion probabilities greater than 0.5 to obtain the "best" model. We see that the results of the SSVS model always lie between those of the other two models. The biggest differences are observed for the frequencies 0 and 1. Whereas the zero-inflated model predicts that our typical individual did not go to the doctor at all with a probability of 41.9 percent, the SSVS model finds a

[7]The results are based on every 10-th of 100,000 samples from the MCMC output after a burn-in period of 10,000 iterations.

	Count Equation ($\boldsymbol{\beta}$)			Regime Selection Equation ($\boldsymbol{\delta}$)		
	Mean	S.d.	P(Inc.)	Mean	S.d.	P(Inc.)
const	2.049	0.440	1.000	1.800	0.561	1.000
age	−0.004	0.020	0.164	−0.006	0.028	0.093
agesq	0.125	0.235	0.225	0.120	0.332	0.105
health	−0.181	0.015	1.000	−0.171	0.022	1.000
handicap	0.214	0.176	0.702	0.029	0.092	0.090
hdegree	0.000	0.002	0.172	0.001	0.002	0.108
married	−0.002	0.026	0.080	0.052	0.104	0.258
schooling	−0.001	0.006	0.100	−0.001	0.006	0.077
hhincome	0.001	0.005	0.096	0.002	0.009	0.105
children	−0.002	0.028	0.092	−0.263	0.133	0.888
self	−0.244	0.195	0.690	−0.008	0.067	0.081
civil	−0.188	0.180	0.588	−0.010	0.061	0.087
bluec	0.009	0.037	0.112	0.004	0.030	0.076
employed	−0.024	0.064	0.175	−0.009	0.050	0.092
public	0.078	0.128	0.341	0.008	0.044	0.071
addon	0.022	0.097	0.112	0.011	0.183	0.059
σ^2	0.623	0.051				

Table 2: Posterior moments

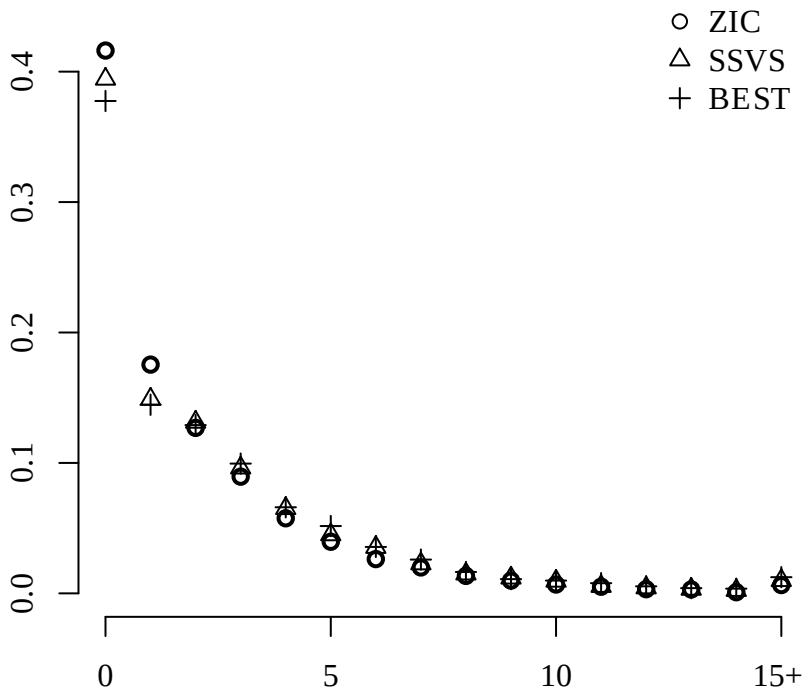

Figure 2: Predictive distributions for the number of doctor visits for a typical individual

predictive probability of 40.0 percent and the "best" model gives 37.5 percent. The predictive probabilities for docvis = 1 are 17.2 percent, 15.9 percent and 13.8 percent, respectively.

Finally, Figure 3 analyses what happens in case the typical individual buys add-on insurance (addon).[8] Posterior distributions for the effect of this on the mean of the number of doctor visits are plotted under the three models. We see that all three posterior distributions are centred around zero indicating that purchasing add-on insurance has no effect on the mean of the number of doctor visits. However, the distributions for the SSVS and the "best" model are much tighter. This result is not surprising since the inclusion probabilities for the add-on variable are only 11.2 for the count equation and 5.9 percent for the regime selection equation in the SSVS model; the "best" model does not include it at all.

[8]Note that purchasing add-on insurance might be an endogenous decision. However, accounting for this fact is outside the scope of this paper.

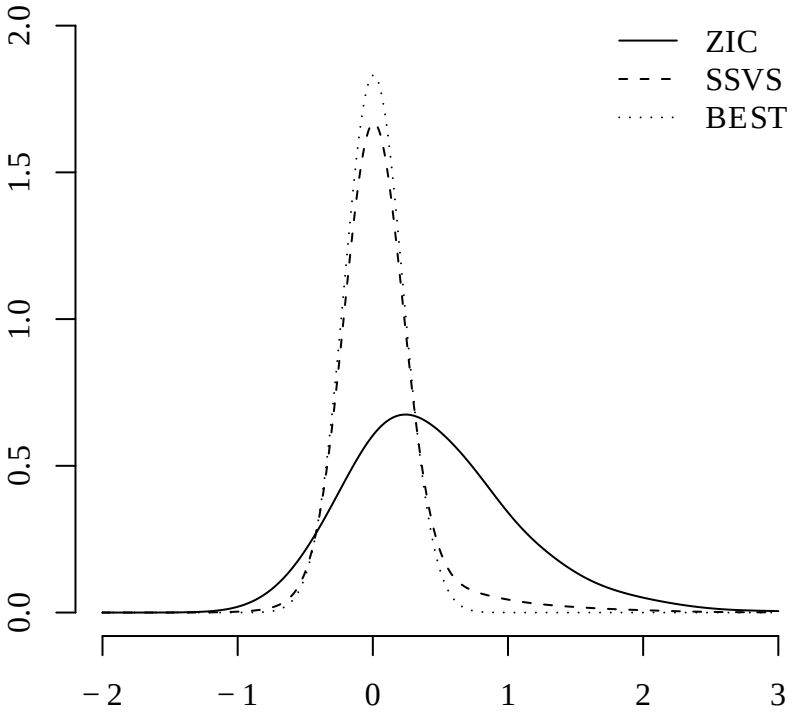

Figure 3: Posterior distributions for the effect of buying add-on insurance on the mean of the number of doctor visits for a typical individual

5 Conclusion

Stochastic search variable selection (SSVS) methods are a convenient tool for model averaging and model selection. For each parameter either a tight or a loose prior distribution is chosen in a data-based fashion. In this paper we extended this approach to zero-inflated count data models, which are commonly used in health economics. We developed an MCMC algorithm for carrying out the empirical work and provided an R package, which helps to reproduce our results and to apply the proposed techniques in further work.

Our empirical application considered the demand for health care in Germany. The proposed SSVS techniques worked well and were computationally feasible. Despite the fact that we started with a rather large model, SSVS chose a very parsimonious specification. Only one fourth of the regressors were included in the model with a probability exceeding 50 percent. We also used SSVS for model selection and obtained estimates for a "best" model. Comparing the results from the three analysed models, we found considerable differences.

References

Böhning, D., Dietz, E., Schlattmann, P., Mendonça, L., Kirchner, U., 1999. The zero-inflated poisson model and the decayed, missing and filled teeth index in dental epidemiology. Journal of the Royal Statistical Society: Series A (Statistics in Society) 162, 195–209.

Deb, P., Munkin, M. K., Trivedi, P. K., 2006. Private insurance, selection, and health care use: A Bayesian analysis of a Roy-type model. Journal of Business & Economic Statistics 24, 403–415.

George, E. I., McCulloch, R. E., 1993. Variable selection via Gibbs sampling. Journal of the American Statistical Association 88, 881–889.

George, E. I., McCulloch, R. E., 1997. Approaches to Bayesian variable selection. Statistica Sinica 7, 339–379.

Gilks, W. R., Wild, P., 1992. Adaptive rejection sampling for Gibbs sampling. Applied Statistics 41, 337–348.

Greene, W., 2005. Functional form and heterogeneity in models for count data. Foundations and Trends in Econometrics 1, 113–218.

Grootendorst, P. V., 1995. A comparison of alternative models of prescription drug utilisation. Health Economics 4, 183–198.

Hoeting, J. A., Madigan, D., Raftery, A. E., Volinsky, C. T., 1999. Bayesian model averaging: A tutorial. Statistical Science 14, 382–417.

Jochmann, M., León-González, R., 2004. Estimating the demand for health care with panel data: A semiparametric Bayesian approach. Health Economics 13, 1003–1014.

Koop, G., 2003. Bayesian Econometrics. Wiley, Chichester.

Lambert, D., 1992. Zero-inflated poisson regression, with an application to defects in manufacturing. Technometrics 34, 1–14.

Liu, J. S., 2001. Monte Carlo Strategies in Scientific Computing. Springer, New York.

Madigan, D., Raftery, A. E., 1994. Model selection and accounting for model uncertainty in graphical models using Occam's window. Journal of the American Statistical Association 89, 1535–1546.

Mullahy, J., 1986. Specification and testing of some modified count data models. Journal of Econometrics 33, 341–365.

R Development Core Team, 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Riphahn, R. T., Wambach, A., Million, A., 2003. Incentive effects in the demand for health care: A bivariate panel count data estimation. Journal of Applied Econometrics 18, 387–405.

Robert, C. P., Casella, G., 2004. Monte Carlo Statistical Methods, 2nd Edition. Springer, New York.

Street, A., Jones, A., Furuta, A., 1999. Cost-sharing and pharmaceutical utilisation and expenditure in Russia. Journal of Health Economics 18, 459–472.

Tanner, M. A., Wong, W., 1987. The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82, 528–550.

Wagner, G. G., Frick, J. R., Schupp, J., 2007. The German Socio-Economic Panel Study (SOEP) – Scope, evolution and enhancements. Schmollers Jahrbuch 127, 139–169.

Appendix: Inference via MCMC Sampling

As mentioned above, we use the data augmentation technique and treat the latent variables $\boldsymbol{y}^* \equiv (y_1^*, \ldots, y_n^*)'$, $\boldsymbol{\eta}^* \equiv (\eta_1^*, \ldots, \eta_n)'$ and $\boldsymbol{d}^* \equiv (d_1^*, \ldots, d_n^*)'$ as parameters of the model. Though the augmented model contains more parameters, inference becomes easier since MCMC techniques can be applied in a more straightforward manner.

Likelihood function

The likelihood function can be written as

$$L(\boldsymbol{\beta}, \boldsymbol{\delta}, \sigma^2, \boldsymbol{y}^*, \boldsymbol{d}^*, \boldsymbol{\eta}^* | \boldsymbol{y}) = \prod_{i=1}^{n} \frac{\exp[-\exp(\eta_i^*)] \exp(\eta_i^* y_i^*)}{y_i^*!} \times \mathrm{N}(\eta_i^* | \boldsymbol{x}_i' \boldsymbol{\beta}, \sigma^2)\, \mathrm{N}(d_i^* | \boldsymbol{x}_i' \boldsymbol{\delta}, 1)\, 1(y_i^*, d_i^*, y_i), \tag{A-1}$$

where $\mathrm{N}(\cdot | a, b)$ denotes the density of the normal distribution with mean a and variance b and

$$1(y_i^*, d_i^*, y_i) = 1(d_i^* \leq 0) 1(y_i = 0) + 1(d_i^* > 0) 1(y_i^* = y_i). \tag{A-2}$$

Prior distributions

The prior distribution for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ can be expressed as $\varphi(\boldsymbol{\beta}|\boldsymbol{\gamma})\varphi(\boldsymbol{\gamma})$ with

$$\varphi(\boldsymbol{\beta}|\boldsymbol{\gamma}) = \mathrm{N}(0, \boldsymbol{B}), \tag{A-3}$$

where $\boldsymbol{B} = \mathrm{diag}(b_1, \ldots, b_k)$,

$$b_j = \begin{cases} \overline{\tau}_{0j}^2 & \text{if } \gamma_j = 0 \\ \overline{\tau}_{1j}^2 & \text{if } \gamma_j = 1 \end{cases} \tag{A-4}$$

and

$$\varphi(\boldsymbol{\gamma}) = \prod_{j=1}^{k} \gamma_j^{\overline{p}_j} (1 - \gamma_j)^{1 - \overline{p}_j}. \tag{A-5}$$

Similarly, the prior distribution for $\boldsymbol{\delta}$ and $\boldsymbol{\kappa}$ is $\varphi(\boldsymbol{\delta}|\boldsymbol{\kappa})\varphi(\boldsymbol{\kappa})$ with

$$\varphi(\boldsymbol{\delta}|\boldsymbol{\kappa}) = \mathrm{N}(0, \boldsymbol{D}), \tag{A-6}$$

where $\boldsymbol{D} = \mathrm{diag}(d_1, \ldots, d_k)$,

$$d_j = \begin{cases} \overline{\omega}_{0j}^2 & \text{if } \kappa_j = 0 \\ \overline{\omega}_{1j}^2 & \text{if } \kappa_j = 1 \end{cases} \tag{A-7}$$

and

$$\varphi(\boldsymbol{\kappa}) = \prod_{j=1}^{k} \kappa_j{}^{\overline{q}_j}(1-\kappa_j)^{1-\overline{q}_j}. \tag{A-8}$$

The prior distribution for σ^2 is specified as

$$\varphi(\sigma^2) = \text{Inv-Gamma}\left(\overline{e}, \overline{f}\right). \tag{A-9}$$

Posterior distribution

The joint posterior distribution is thus given by

$$\begin{aligned} &\varphi(\boldsymbol{\beta}, \boldsymbol{\delta}, \sigma^2, \boldsymbol{y}^*, \boldsymbol{d}^*, \boldsymbol{\eta}^*, \boldsymbol{\gamma}, \boldsymbol{\kappa}|\boldsymbol{y}) \\ &\quad \propto \varphi(\boldsymbol{\beta}|\boldsymbol{\gamma})\varphi(\boldsymbol{\gamma})\varphi(\boldsymbol{\delta}|\boldsymbol{\kappa})\varphi(\boldsymbol{\kappa})\varphi(\sigma^2)L(\boldsymbol{\beta}, \boldsymbol{\delta}, \sigma^2, \boldsymbol{y}^*, \boldsymbol{d}^*, \boldsymbol{\eta}^*|\boldsymbol{y}) \end{aligned} \tag{A-10}$$

Gibbs sampling algorithm

The proposed Gibbs sampling algorithm consists of the following steps:

1. Sample $\boldsymbol{\beta}$ from $\varphi(\boldsymbol{\beta}|\boldsymbol{\eta}^*, \sigma^2, \boldsymbol{\gamma}) = \mathrm{N}(\boldsymbol{\beta}|\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$ with variance

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \left(\boldsymbol{B}^{-1} + \frac{1}{\sigma^2}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1} \tag{A-11}$$

and mean

$$\boldsymbol{\mu}_{\boldsymbol{\beta}} = \boldsymbol{\Sigma}_{\boldsymbol{\beta}}\left(\frac{1}{\sigma^2}\sum_{i=1}^{n} \boldsymbol{x}_i \eta_i^*\right). \tag{A-12}$$

2. Draw σ^2 from $\varphi(\sigma^2|\boldsymbol{\eta}^*, \boldsymbol{\beta}) = \text{Inv-Gamma}\left(\overline{e} + \frac{n}{2}, \overline{f} + \frac{\sum_{i=1}^{n}(\eta_i^* - \boldsymbol{x}_i'\boldsymbol{\beta})^2}{2}\right)$.

3. Sample $\boldsymbol{\delta}$ from $\varphi(\boldsymbol{\delta}|\boldsymbol{d}^*, \boldsymbol{\kappa}) = \mathrm{N}(\boldsymbol{\delta}|\boldsymbol{\mu}_{\boldsymbol{\delta}}, \boldsymbol{\Sigma}_{\boldsymbol{\delta}})$ with variance

$$\boldsymbol{\Sigma}_{\boldsymbol{\delta}} = \left(\boldsymbol{D}^{-1} + \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \right)^{-1} \tag{A-13}$$

and mean

$$\boldsymbol{\mu}_{\boldsymbol{\delta}} = \boldsymbol{\Sigma}_{\boldsymbol{\delta}} \left(\sum_{i=1}^{n} \boldsymbol{x}_i d_i^* \right). \tag{A-14}$$

4. For $i = 1, \ldots, n$ sample η_i^* from

$$\varphi(\eta_i^*|y_i^*, \boldsymbol{\beta}, \sigma^{-2}) \propto \exp\left[-\exp(\eta_i^*) + \eta_i^* y_i^* - \frac{(\eta_i^* - \boldsymbol{x}_i'\boldsymbol{\beta})^2}{2\sigma^2} \right]. \tag{A-15}$$

These density functions are log-concave and we use adaptive rejection sampling [ARS, see Gilks and Wild (1992)] to generate the η_i^*.

5. Jointly sample $(\boldsymbol{y}^*, \boldsymbol{d}^*)$: For $i = 1, \ldots, n$ sample (y_i^*, d_i^*) from $\varphi(y_i^*, d_i^*|\boldsymbol{\eta}^*, \boldsymbol{\delta})$: If $y_i > 0$, set $y_i^* = y_i$ and sample d_i^* from $\mathrm{TN}^+(\boldsymbol{x}_i'\boldsymbol{\delta}, 1, 0)$, where $\mathrm{TN}^+(\mu, \sigma^2, a)$ denotes a Normal distribution with mean μ and variance σ^2 that is truncated at the left at a. If $y_i = 0$, draw an auxiliary variable z from a standard uniform distribution $\mathrm{U}(0, 1)$. If

$$z < \frac{1 - \Phi(\boldsymbol{x}_i'\boldsymbol{\delta})}{1 - \Phi(\boldsymbol{x}_i'\boldsymbol{\delta}) + \Phi(\boldsymbol{x}_i'\boldsymbol{\delta}) \exp[-\exp(\eta_i^*)]}, \tag{A-16}$$

where $\Phi(\cdot)$ denotes the standard-normal c.d.f., sample y_i^* from Poisson$[\exp(\eta_i^*)]$ and d_i^* from $\mathrm{TN}^-(\boldsymbol{x}_i'\boldsymbol{\delta}, 1, 0)$, where $\mathrm{TN}^-(\mu, \sigma^2, a)$ denotes a Normal distribution with mean μ and variance σ^2 that is truncated at the right at a. Otherwise set $y_i^* = y_i = 0$ and sample d_i^* from $\mathrm{TN}^+(\boldsymbol{x}_i'\boldsymbol{\delta}, 1, 0)$.

6. For $j = 1, \ldots, k$ sample γ_j from $\varphi(\gamma_j|\beta_j) = \text{Bernoulli}\left(\frac{u_{1j}}{u_{1j}+u_{0j}}\right)$ with

$$u_{1j} = (\overline{\tau}_{1j}^2)^{-\frac{1}{2}} \exp\left(\frac{-\beta_j^2}{2\overline{\tau}_{1j}^2} \right) \overline{p}_j \tag{A-17}$$

and

$$u_{0j} = (\overline{\tau}_{0j}^2)^{-\frac{1}{2}} \exp\left(\frac{-\beta_j^2}{2\overline{\tau}_{0j}^2} \right) (1 - \overline{p}_j). \tag{A-18}$$

7. For $j = 1, \ldots, k$ sample κ_j from $\varphi(\kappa_j|\delta_j) = \text{Bernoulli}\left(\frac{u_{1j}}{u_{1j}+u_{0j}}\right)$ with

$$u_{1j} = (\overline{\omega}_{1j}^2)^{-\frac{1}{2}} \exp\left(\frac{-\delta_j^2}{2\overline{\omega}_{1j}^2}\right) \overline{q}_j \tag{A-19}$$

and

$$u_{0j} = (\overline{\omega}_{0j}^2)^{-\frac{1}{2}} \exp\left(\frac{-\delta_j^2}{2\overline{\omega}_{0j}^2}\right) (1 - \overline{q}_j). \tag{A-20}$$