CHAPTER 6

Analysis of ‘Number of Permits’ with the Gibbs Sampler*

6.1 Introduction

The two models in Chapter 5 did not confirm the expected positive impact of the two-rate tax on the number of building permits. One might be tempted to view this as the final result of the analysis, or at least as the final result given the available data. If one believes that the correct model ought to take into account that the data consists of nonnegative integers, then the best one can do with conventional maximum likelihood techniques and a reasonable amount of effort is to use either the Poisson or the negative binomial distribution.

However, Markov Chain Monte Carlo methods are able to overcome the computational difficulty, and the Gibbs Sampler offers a fairly straightforward possibility to examine further distributions. This chapter describes an analysis of the number of building permits with the Gibbs Sampler. Section 6.2 describes the setup of the analysis, in Section 6.3 the estimation procedure is summarized while the technical details are relegated to Appendix G, and Section 6.4 shows the results, which differ considerably from the results in the previous chapter.

6.2 Setup of the analysis

The new model incorporates several insights that were gained from the earlier work. The available data set of independent variables for the 219 municipalities seems to omit important determinants of construction, so that a variable $S_i$ that describes these determinants is estimated for each municipality. As in the model in Section 5.4, this term $S_i$ is interpreted as the economic status of municipality $i$. The only predetermined municipality-specific information used is ‘Population’, $POP_{i,t}$, ‘Months reported’, $MON_{i,t}$, and the adjusted tax differential, $ATD_{i,t}$. Additional yearly variables $Y_t$ describe exogenous macroeconomic impacts that differ from year to year.

*I thank Dr. Keying Ye, Statistics Department, Virginia Polytechnic Institute and State University, for his support and discussions regarding the work in this chapter.
As in the earlier models, the number of permits $P_{i,t}$ that are issued by municipality $i$ in year $t$ are assumed to follow a Poisson distribution with mean $\mu_{i,t}$. To assure nonnegativity, this mean is modeled as a deterministic function according to

$$\mu_{i,t} = \exp(\alpha + S_i + Y_t + \tau \cdot ATD_{i,t}) \cdot POP_{i,t} \cdot MON_{i,t}$$  \hspace{1cm} (6.1)$$

where $\alpha$ is the intercept for municipality 1 in year 1, and $\tau$ is the coefficient that measures the influence of the adjusted tax differential. So far this is in essence the same model that was used for the analysis with the Poisson distribution in Section 4.7. Yet Figure E.1 in Appendix E reveals that, because the data are overdispersed, the Poisson distribution is not the best approximation of the true distribution of building permits. To find a more suitable approximation of this distribution, $\mu_{i,t}$ can be modeled as a stochastic function with deterministic parameters and coefficients, and with a stochastic error term $\varepsilon_{i,t}$ that is assumed to be gamma distributed. The resulting distribution is negative binomial, and it is used in Chapter 5 to analyze the number of building permits.

Yet there is no guarantee that the negative binomial distribution approximates the true distribution best, and the assumption of a gamma distributed error term, made for computational ease, is rather artificial. The comparison between the results of the analysis with the Poisson and the analysis with the negative binomial distribution in Section 4.7 shows the strong impact of the distributional assumptions on the results; the Poisson distribution yielded a significantly positive effect of the two-rate tax, while the use of the negative binomial distribution did not show any significant impact. To be able to analyze additional distributions the current model assumes that the unknown expressions in equation 6.1, $\alpha$, $\tau$, $S_i$, and $Y_t$, are drawn independently from appropriate two-parametric distributions. While in the earlier analyses the coefficients of interest were included as multiplicative factors of exogenous data and dummy variables, the undetermined coefficients of this model are the parameters of the distributions that describe the unknown terms in equation 6.1.\textsuperscript{166}

Figure 6.1 summarizes these assumptions.\textsuperscript{167} Each term of the model is shown as a node in the figure; a rectangle stands for observed data, and a circle describes all unknown terms. The arrows describe the assumed connections between these terms; a dashed arrow describes a deterministic relationship, and a solid arrow stands for a probabilistic link. Note that all arrows are unidirectional: while it is for example possible to determine $\alpha$ if $\alpha_\mu$ and $\alpha_\sigma$.

\textsuperscript{166} $S_i$ and $Y_t$ cannot be called ‘dummy variables’, because their range is not restricted to {0,1} anymore but it is the range of the distribution they are assumed to follow.

\textsuperscript{167} The explanation of the figure closely follows the explanation in Spiegelhalter et al. (1996), pp. 25-27.
Lauritzen et al. (1990) show that this is an implication of any model that can be expressed in the form of eq. (6.2) are known, it is impossible to compute $\alpha_g$ if for example $\alpha$ and $\alpha_{\mu}$ are known. Repetitive structures that refer either to different municipalities or to different years are drawn as stacks of ‘sheets’. As pointed out in Spiegelhalter et al. (1996), the figure shows the properties of the model before any data is observed, and the relationships between the nodes may change after data are included. While nodes without a common ‘parent’ will initially be marginally independent (e.g. $\alpha$ and $\tau$), they might become dependent once the $P_{\mu}$ are observed. This follows from the fact that the product of all conditional distributions of each node $n_i$, given its parents, fully specifies the joint distribution of all $k$ probabilistic terms of the model, so that

$$P(n_1, n_2, \ldots, n_k) = \prod_{i=1}^{k} P(n_i | n_{-i}) , \hspace{1cm} (6.2)$$

where $n_{-i}$ stands for all nodes except node $i$. The product on the right hand side is the product of the distributions of all unknown nodes $n_i$, conditional on the distribution of all other nodes $n_j, j \neq i$. The conditional distribution of any node $n_i$ is proportional to the joint distribution of all nodes $n_l, l = 1, \ldots, k$, and therefore proportional to all terms in this joint distribution that

---

$Lauritzen et al. (1990)$ show that this is an implication of any model that can be expressed in the form of Figure 6.1.
contain \( n_i \). This means that the full conditional distribution of \( n_i \) depends only on its ‘parents’ (the distributions of the parameters of \( n_i \)), its ‘children’ (the distributions for which \( n_i \) is used to formulate a parameter), and the values of its ‘co-parents’ (the other terms that are used to formulate the parameters of the ‘children’). For example, the ‘parents’ of \( \alpha \) in Figure 6.1 are \( \alpha_{\mu_i} \) and \( \alpha_{\sigma_i} \), the only ‘child’ is \( P_{i,t} \), and the ‘co-parents’ are \( \tau, S_i \) and \( Y_i \) (the other determinants of the single parameter of \( P_{i,t} \)). The proportional relationship of the full conditional distribution of any \( n_i \) can therefore be written as

\[
P(n_i \mid n_{.-}) \propto P(n_i \mid \text{parents of } n_i) \cdot \prod_{n_{j: \text{children of } n_i}} P(n_j \mid \text{parents of } n_j)
\] (6.3)

The first term on the right hand side can be interpreted as the prior of \( n_i \), and the second term is nothing but the likelihood of \( n_i \), given its ‘children’ and its ‘co-parents’. To be able to determine the full conditional distribution for all nodes, a probability distribution needs to be assigned to each unknown quantity. At this stage of the analysis, the unknown expressions \( \alpha, \tau, S_i \) and \( Y_i \) are assumed to follow normal distributions. It is also necessary to assign prior distributions to all nodes without ancestors, that is, to the parameters of the normal distributions of \( \alpha, \tau, S_i \) and \( Y_i \). Because these priors ought to influence the analysis as little as possible, the prior distribution of the mean of every unknown term in equation 6.1 is assumed to be normal with a mean equal to 0 and a variance of 10,000, and the distribution of the inverse of the variance of each term is specified as gamma, with both parameters equal to 0.01.

Although it would be possible to assign completely flat priors, these specifications simplify the analysis without imposing to much of a bias. The full specification of the model is summarized in Table 6.1.

<table>
<thead>
<tr>
<th>Table 6.1 Model Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_{i,t} \sim \text{Poisson (} \mu_{i,t} \text{)} )</td>
</tr>
<tr>
<td>( \mu_{i,t} = \exp(\alpha + S_i + Y_i + \tau \cdot TAX_{i,t}) \cdot POP_{i,t} \cdot MON_{i,t} )</td>
</tr>
<tr>
<td>( \alpha \sim \text{N (} \alpha_{\mu}, \alpha_{\sigma}^{-1} \text{)} )</td>
</tr>
<tr>
<td>( S_i \sim \text{N (} S_{\mu_i}, S_{\sigma_i}^{-1} \text{)} )</td>
</tr>
<tr>
<td>( Y_i \sim \text{N (} Y_{\mu_i}, Y_{\sigma_i}^{-1} \text{)} )</td>
</tr>
<tr>
<td>( \tau \sim \text{N (} \tau_{\mu}, \tau_{\sigma}^{-1} \text{)} )</td>
</tr>
<tr>
<td>( \alpha_{\mu}, S_{\mu_i}, Y_{\mu_i}, \tau_{\mu} \sim \text{N (0, 10000)} )</td>
</tr>
<tr>
<td>( \alpha_{\sigma}, S_{\sigma_i}, Y_{\sigma_i}, \tau_{\sigma} \sim \text{Gamma (0.01, 0.01)} )</td>
</tr>
</tbody>
</table>
6.3 Estimation of the model

Given the distributional assumptions, it is not possible to find an analytic solution for the joint distribution in equation 6.2. Yet the determination and parameter optimization of the marginal distribution of each node requires knowledge of the joint distribution of all nodes in Figure 6.1. In addition, even if this joint distribution were known, it would still be extremely difficult to perform the integrations over all parameters that are necessary to obtain the marginal distributions. The use of maximum likelihood techniques is therefore impossible. Instead the Gibbs Sampler can be used to approach the joint distribution via a Markov Chain, and to perform a Monte Carlo integration of this distribution. In Section 3.5 it has been shown that the Gibbs Sampler consists of nothing but a (very long) set of samples from all full conditional distributions, which ultimately converges to a set of samples from all marginal distributions. From the samples from the marginal distribution of \( \tau \) it is then possible to determine the mean and the variance of \( \tau \), which yields information about the impact of the tax differential on the number of building permits.

With the distributional assumptions in Table 6.1 it is relatively straightforward to determine the full conditional distributions in equation 6.2. A (posterior) distribution that is derived from a normal prior and a normal likelihood is also normal, and a (posterior) distribution that is derived from a gamma prior and a normal likelihood is gamma. All distributions but the distributions of \( \alpha, \tau, S, \) and \( Y \) fall into either of these two categories, and do not pose any difficulty for the analysis, because very efficient sampling algorithms are available. The full conditional distributions of \( \alpha, \tau, S, \) and \( Y \), which consist of a normal prior and Poisson likelihood terms, cannot be related to known distributions, so that other than standard sampling methods need to be employed. Fortunately these full conditional distributions are log-concave, so that the very effective adaptive-rejection sampling method by Gilks and Wild (1992) can be used. All full conditional distributions for this model and their derivation are presented in Appendix G.

---

169 See for example Box and Tiao (1972), pp. 74-75, and Gilks (1996).

170 To sample from the standard normal distribution I used Box and Muller’s (not Müller’s!) (1958) algorithm (see Press et al., 1995, for a formulation of the algorithm in C). To sample from the standardized gamma distribution with single parameter \( \gamma \) I used Ahrens and Dieter’s (1974) algorithm when \( \gamma < 1 \), and Cheng and Fast’s (1979) algorithm when \( \gamma > 1 \). For \( \gamma = 1 \) the gamma distribution becomes the exponential distribution, and I used the algorithm described by Press et al. (1995). Additional algorithms are surveyed in Devroye (1986). All algorithms require the generation of pseudo-random numbers. For their generation I used L’Ecuyer’s (1986) algorithm (see Press et al., 1995, for a formulation of the algorithm in C), which has a cycle length of \( 10^{19} \) before the generator repeats itself. See Geweke (1995) and Press et al. (1995) for discussions of the difficulties and dangers that are related to the generation of pseudo-random numbers.
6.4 Results

Because the Gibbs Sampler yields only a very long chain of sampled values, it is crucial to be able to determine, from which point on these values are sampled from the correct distribution, and which values where sampled during the ‘burn-in’ phase. Although several theoretical tests of convergence are available, the most straightforward test is to plot a graph of the sampled values of the parameter in question, and to use this graph to determine when the burn-in phase was over. However, even though the samples generated by the Gibbs Sampler will ultimately converge to samples from the correct distribution, this can take a very long time, and a single chain of samples might give a wrong impression of convergence. It is safer to run several simultaneous chains with different starting values, and to make sure that all chains converge to samples from the same distribution.

For each of the four data sets I ran three simultaneous chains with different starting values of \( \tau (\tau = -5, \tau = 0, \tau = 2) \) and a length of 5,000 iterations. For both of the first two chains I set the starting values of all other variables to 0, and their variances equal to 10; for the third chain I set all starting values equal to 2, with variances equal to 10.

Graphs of the resulting samples are shown in Figures 6.2 to 6.5. Because the chains for the starting values \( \tau = -5 \) and \( \tau = 0 \) behaved very similarly, only the graphs for \( \tau = -5 \) are shown. For every data set the first two chains had converged after about 500 iterations, while the third chain took considerably longer (about 3,000 iterations for the data set of nonresidential additions and alterations) to converge. The means and standard deviations of each set of samples, which are summarized in Table 6.2, are sufficiently similar to suggest that at least convergence among the three chains has occurred in every case.

The results differ considerably from the results in Chapter 5. The impact of the two-rate tax is significantly positive for three of the four data sets, yet the results indicate that the tax has different effects in different sectors. While the impact on residential construction is significantly positive, and differs not much for the construction of whole units and additions and alterations, it does not have the same uniform effect on nonresidential construction. Surprisingly, nonresidential construction of whole units does not respond to the construction incentive of the two-rate tax, while nonresidential additions and alterations show a very strong reaction to this incentive. A likely explanation for this is provided by the fact that municipalities frequently offer generous tax breaks for new nonresidential

---

171 See Tanner (1993) and Gilks et al. (1996a) for surveys and comparisons of convergence tests.


173 On an 80 MHz Pentium processor each chain took about one hour to finish.
Figure 6.2.a Coefficient $\tau$ of the adjusted tax differential for residential whole units. Starting value of $\tau = -5$.

Figure 6.3.a Coefficient $\tau$ of the adjusted tax differential for residential additions and alterations. Starting value of $\tau = -5$.

Figure 6.2.b Coefficient $\tau$ of the adjusted tax differential for residential whole units. Starting value of $\tau = 2$.

Figure 6.3.b Coefficient $\tau$ of the adjusted tax differential for residential additions and alterations. Starting value of $\tau = 2$.

Figure 6.2 and Figure 6.3 Residential construction
Figure 6.4.a  Coefficient $\tau$ of the adjusted tax differential for nonresidential whole units. Starting value of $\tau = -5$.

Figure 6.5.a  Coefficient $\tau$ of the adjusted tax differential for nonresidential additions and alterations. Starting value of $\tau = -5$.

Figure 6.4.b  Coefficient $\tau$ of the adjusted tax differential for nonresidential whole units. Starting value of $\tau = 2$.

Figure 6.5.b  Coefficient $\tau$ of the adjusted tax differential for nonresidential additions and alterations. Starting value of $\tau = 2$.

Figure 6.4 and Figure 6.5  Nonresidential construction.
Table 6.2 Means and standard deviations for $\tau$, computed from chains with different starting points

<table>
<thead>
<tr>
<th></th>
<th>Residential Whole Units</th>
<th>Residential Additions and Alterations</th>
<th>Nonresidential Whole Units</th>
<th>Nonresidential Additions and Alterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start $\tau = -5$</td>
<td>0.1515 (0.0080)</td>
<td>0.1430 (0.0069)</td>
<td>0.0460 (0.0330)</td>
<td>0.3917 (0.0172)</td>
</tr>
<tr>
<td>Start $\tau = 0$</td>
<td>0.1508 (0.0080)</td>
<td>0.1419 (0.0066)</td>
<td>0.0481 (0.0320)</td>
<td>0.3948 (0.0170)</td>
</tr>
<tr>
<td>Start $\tau = 2$</td>
<td>0.1512 (0.0077)</td>
<td>0.1411 (0.0070)</td>
<td>0.0466 (0.0335)</td>
<td>0.3987 (0.0185)</td>
</tr>
</tbody>
</table>

Notes: Means and standard deviations for the starting values of $\tau = -5$ and $\tau = 0$ are computed from 4,000 iterations after a burn-in of 1,000 iterations; means and standard deviations for the starting value $\tau = 2$ are computed from 2,000 iterations after a burn-in of 3,000 iterations.

construction to attract new businesses, while additions to existing buildings and residential construction do not enjoy the same tax advantage.

This concludes the analysis of the number of building permits. Some additional interpretations of the results are presented together with a summary of the study in Chapter 8. The following chapter describes the analysis of the value per building permit.