BAYESIAN ANALYSIS OF CARTEL STABILITY AND REGIME SWITCHING

Eric EISENSTAT
University of California, Irvine[2]

Abstract

Empirical analysis of collusive regimes typically requires the construction of structural econometric models, with explicit ties to theoretical models of firm behavior in equilibrium. To that end, theory often elicits a wealth of important information regarding the structural parameters, information that is indispensable in accurately identifying desired phenomena, but nevertheless, is inevitably disregarded by classical techniques. Motivated by these considerations, the paper demonstrates how Bayesian methods may be used to better incorporate such structural knowledge through prior probabilistic beliefs. As a result, Bayesian posterior inference provides a clear and precise empirical interpretation of collusive behavior and cartel stability.

Key-words: *Cartel, dynamic oligopoly, collusion detection, regime switching, structural modelling, Bayesian methods*

JEL Classification: L_{13}, C_{11}, C_{32}

1. Introduction

Porter (1983) was one of the first studies to empirically examine collusive behavior under a particular equilibrium assumption of dynamic, oligopolistic behavior. Specifically, the empirical model is derived under the assumption that firms behave according to the Green and Porter (1984) model of collusive regimes, where unobserved periodic demand shocks lead to temporary, perfectly competitive regimes. This results in a simultaneous equation switching regression, which Porter (1983) estimates using a Maximum Likelihood procedure. Consequently, the results are used to argue in favor of the existence of collusion during the sample period, by rejecting the null hypothesis of “no regime switching” through a Likelihood Ratio test.

An analogous Bayesian estimation may be interesting for several reasons. First, the theoretical framework warrants significant prior information about the structural parameters that Maximum Likelihood estimation cannot incorporate. This is particularly significant in the Porter (1983) exercise because it leads to

[2] Department of Economics, University of California, Irvine, e-mail: eric.eisenstat@gmail.com.

insensible MLE estimates, such as for example market demand elasticities less than one (in absolute value). In addition, simulated draws from the joint posterior distribution allow for an extended analysis of the regime-switching model, and further facilitates straightforward model comparison through the computation of Marginal Likelihoods and Bayesian Posterior Odds Ratios.

The data used in the study describes the operations of the Joint Executive Comittee (a railroad cartel) for the period 1880-1886. Observations are market-level prices and quantities for each year in the sample period, along with covariates used to control for competition from outside markets as well as the market structure of the cartel itself. Aside from Porter (1983), this dataset has been analyzed extensively in the literature under various applications and is familiar to most industrial organization economists (e.g. Ripley, 1906; Ulen, 1979; Binder, 1988; Prager, 1989; Ellison 1994). The primary motivation, as in most utilizations of this example, is that the data characterizes an explicit account of the operations of a *known* cartel. Summary statistics are provided in Table 1; for a more extensive description of the JEC and the related data set, please refer to (Gilchrist, 1960; Ulen, 1979; Porter, 1983).

Table 1

Summary Statistics

	mean	std dev	min	max
GR	0.2465	0.0665	0.1250	0.4000
TQG	25,384	11,633	4,810	76,407
LAKES	0.5732	0.4954	0	1
PO	0.6189	0.4864	0	1

2. **Estimation**

This section derives and analyzes a Bayesian version of the regime-switching simultaneous regression. The primary interests for this arise from the fact that the specified model asserts significant prior information about the structural parameters, and that draws from a posterior distribution seem to be more accomodating in answering some of the questions in this paper, in comparison to point estimates.

As usual, to estimate a Bayesian model, we begin by constructing a likelihood and joint prior distribution. The construction of the likelihood is, in principle, identical to that of the Maximum Likelihood estimation technique presented in the paper. However, in order to facilitate the ensuing MCMC algorithm, we derive a slightly different representation. Consequently, consider the following definitions:

$$x_{1t}' = [1\, L_t\ M_t],\ x_{2t}' = [1\, S_t'\ I_t\ M_t'],\ y_{1t} = ln\ Q_t,\ y_{2t} = ln\ p_t$$

$$X = \begin{pmatrix} x_{1t} & 0 \\ 0 & x_{2t} \end{pmatrix},\ y_t = [y_{1t}\ y_{2t}],\ \Gamma = \begin{pmatrix} 1 & -\alpha_1 \\ -\beta_1 & 1 \end{pmatrix}$$

$$\delta_1 = [\alpha_0\ \alpha_2\ \alpha_3']',\ \delta_2 = [\beta_0\ \beta_2'\ \beta_3\ \beta_4']',\ B = [\delta_1'\ \delta_2']'$$

The simultaneous equations switching regression model can be written as:

$$y_{1t} - \alpha_1 y_{2t} = \alpha_0 + \alpha_2 L_t + M_t' \alpha_3 + \varepsilon_{1t} \quad (1)$$

$$y_{2t} - \beta_1 y_{1t} = \beta_0 + S_t' \beta_2 + \beta_3 I_t + M_t' \beta_4 + \varepsilon_{2t} \quad (2)$$

or equivalently,

$$y_t \Gamma' = [x_{1t}' \delta_1 \; x_{2t}' \delta_2] + \varepsilon_t \quad (3)$$

For notational convenience, let $X_{-I} = \{L_t, S_t, M_t\}_{t=1}^{T}$ (the set of independent variables). It follows, then, that the likelihood function is:

$$p(y \mid X_{-I}, B, \Sigma, \Gamma, I) = \sum_{t=1}^{T} \phi_{1\times2}(y_t \mid [x_{1t}' \delta_1 \; x_{2t}' \delta_2](\Gamma^{-1})', 1, \Gamma' \Sigma^{-1} \Gamma)$$

$$= \phi_{T\times2}(y \mid [x_1 \delta_1 \; x_2 \delta_2](\Gamma^{-1})', I_T, \Gamma' \Sigma^{-1} \Gamma) \quad (4)$$

where $\phi_{n\times m}$ denotes the *Matric-Normal* distribution. Further note that:

$$p(y \mid X_{-I}, B, \Sigma, \Gamma, I) \propto \phi_{T\times2}(y\Gamma' \mid [x_1 \delta_1 \; x_2 \delta_2], I_T, \Sigma^{-1}) \quad (5)$$

$$= \phi_{2T}((\Gamma \otimes I_T) vec(y) \mid XB, \Sigma \otimes I_T) \quad (6)$$

where ϕ_n denotes the *Multivariate Normal* distribution. We will exploit this relationship in deriving the conditional distributions for the MCMC algorithm.

The prior information in this model is described by a joint distribution over all the parameters (including hyper-parameters), which we decompose into the *hyper-prior* and *joint prior* distributions, respectively, as:

$$p(\alpha_0,\ldots,\alpha_3,\beta_0,\ldots,\beta_4,\Sigma,I_1,\ldots,I_T,\lambda) = p(\lambda)p(\alpha_0,\ldots,\alpha_3,\beta_0,\ldots,\beta_4,\Sigma,I_1,\ldots,I_T \mid \lambda) \quad (7)$$

Let $\gamma_1 = -\alpha_1, \gamma_2 = -\beta_1$. The joint prior distribution, $p(\cdot \mid \lambda)$, can be further decomposed as:

$$p(\alpha_0,\ldots,\alpha_3,\beta_0,\ldots,\beta_4,\Sigma,I_1,\ldots,I_T \mid \lambda) = p(I_1,\ldots,I_T \mid \lambda) \quad (8)$$

$$\times p(\Sigma \mid I_1,\ldots,I_T,\lambda) \quad (9)$$

$$\times p(\gamma_1,\gamma_2 \mid \Sigma, I_1,\ldots,I_T,\lambda) \quad (10)$$

$$\times p(\delta_1,\delta_2 \mid \gamma_1,\gamma_2,\Sigma,I_1,\ldots,I_T,\lambda) \quad (11)$$

The prior distribution for regime types, I_t, is assumed in Porter (1983) to be

$$I_t \mid \lambda \overset{iid}{:} Bernoulli(\lambda) \quad (12)$$

Furthermore, to construct (10) and (11), we consider the assumptions imposed by the model. First, note that α_1 and $\beta_1 + 1$ are the *constant* elasticities of demand and cost, respectively. Since the industry marginal revenue is given by $MR_{it} = p_t(1 - 1/|\alpha_1|)$ (for a downward-sloping demand curve), it must be true that $|\alpha_1| > 1$ in order for marginal revenue to be positive. This fact, which is pointed out by Porter (1983) explicitly, is contradicted by the subsequent maximum likelihood estimate, $\hat{\alpha}_1 = -0.8$. This, however, is a troublesome result since it inhibits our interpretation of other parameter estimates. That is, how do we attribute and estimated shift in quantity to a *collusive regime* if our estimate of the demand elasticity implies that optimal output quantity should always be *zero*? One way of describing this puzzling result is that MLE, in this case, puts too much of the burden of inference on the data. More specifically, consider the alternate perspective on data as a limited resource, which we *consult* in search of an answer to our research question. Even if we just eliminate implausible answers *prior* to consulting the data, we allow the data to focus on distinguishing among only plausible solutions. The latter is further extended by describing the alternative possible outcomes with a probability distribution. Therefore, prior knowledge allows us to exploit the data to a greater extent. This is especially true for the exercise at hand since a significant portion of the prior knowledge is asserted by the model, but cannot be identified by the data implicitly.

In addition to restricting the support of α_1, the model similarly requires that $\beta_1 + 1 > 1$ to ensure that marginal costs are positive and increasing, and therefore, that an equilibrium exists. Consequently, we assign the following prior distribution:

$$p(\gamma_1, \gamma_2 \mid \Sigma, I_1, \ldots, I_T, \lambda) = p(\gamma_1, \gamma_2) = MTt_5([1.05\ -0.5]', 50I_2, [1\ -\infty]', [\infty\ 0]') \quad (13)$$

where $MTt_5(m, S, a, b)$ denotes a *Multivariate Truncated t* distribution with 5 degrees of freedom, mean, m, scale, S, vector of lower bounds, a, and vector of upper bounds, b. Note that the independence of γ_1, γ_2 from $I_1, \ldots, I_T$ follows from the assumption that elasticities do not change across time (as specified in the model). Porter (1983) also notes the following relationships among the remaining parameters:

1) $\beta_0 = ln\ D = ln\ (\beta_1 + 1) - \beta_1 ln \left[\sum_{i=1}^{f} a_i^{-1/\beta_1}\right]$
2) $\beta_3 \in [0, ln\ (\alpha_1/(\alpha_1 + 1))]$
3) α_2 "should be" negative

To derive the prior distribution of β_0, we begin with the assumption, $a_i \approx a$. Intuitively, a_i represents input factor prices and technology parameters for firm i. Therefore, our assumption reflects a belief that all the firms in the industry face similar production costs and exhibit similar technologies. Furthermore, if $ln\ a \sim N(0,100)$, then

$$p(\beta_0 \mid \gamma_1, \gamma_2, \Sigma, I_1, \ldots, I_T, \lambda) = p(\beta_0 \mid \gamma_2) = N(ln\ (1-\gamma_2) - \gamma_2\ ln\ f, 100) \quad (14)$$

where f is the total number of firms. Although the number of firms in our sample varies across periods due to entry and exit, for the purpose of this exercise we simply let $f = 6.84$, the average number of firms, for all periods. Accounting for [2] and [3] above, we also specify the prior distributions:

$$p(\beta_3 \mid \beta_0, \gamma_1, \gamma_2, \Sigma, I_1, \ldots, I_T, \lambda) =$$
$$p(\beta_3 \mid \gamma_1) = U(0, ln\ (\gamma_1/(\gamma_1 - 1))) \quad (15)$$
$$p(\alpha_2 \mid \beta_0, \beta_3, \gamma_1, \gamma_2, \Sigma, I_1, \ldots, I_T, \lambda) =$$

$$p(\alpha_2) = N(-2,100) \quad (16)$$

$$p(\alpha_0, \alpha_3, \beta_2, \beta_4 \mid \alpha_2, \beta_0, \beta_3, \gamma_1, \gamma_2, \Sigma, I_1, \ldots, I_T, \lambda) =$$

$$p(\alpha_0, \alpha_3, \beta_2, \beta_4) = N(0_{29}, 100 I_{29}) \quad (17)$$

$$p(\Sigma \mid I_1, \ldots, I_T, \lambda) =$$

$$p(\Sigma) = IW_{10}(I_2) \quad (18)$$

The prior distribution (17) describes prior beliefs over shifts in the intercepts attributed to "monthly" and "structural" dummies. Our prior over these parameters is relatively flat and asserts stochastic independence. The prior distribution of Σ reflects a belief that ε_{1t} and ε_{2t} are uncorrelated. Note that (14)-(18) fully describe (11), and assuming that

$$\lambda : \; U(0,1) \quad (19)$$

completes the full prior specification.

The model is estimated with a Gibbs Sampling algorithm by iteratively sampling from conditional distributions (for further details regarding Bayesian posterior sampling techniques, see Gelman et al., 2003; Koop, 2003). In particular, we sample from five conditional distributions:

1) $B \mid \Gamma, \Sigma, I, \lambda, X_{-I}, y$
2) $\Sigma \mid B, \Gamma, I, \lambda, X_{-I}, y$
3) $\Gamma \mid B, \Sigma, I, \lambda, X_{-I}, y$
4) $I \mid B, \Gamma, \Sigma, \lambda, X_{-I}, y$
5) $\lambda \mid B, \Gamma, \Sigma, I, X_{-I}, y$

These distributions are straightforward to derive from the prior and likelihood as follows:

$$p(B \mid \Gamma, \Sigma, I, \lambda, X_{-I}, y) \propto p(\beta_0 \mid \gamma_2)\, p(\beta_3 \mid \gamma_1)\, p(\alpha_2)\, p(\alpha_0, \alpha_3, \beta_2, \beta_4)$$

$$\times \phi_{2T}((\Gamma \otimes I_T) vec(y) \mid XB, \Sigma \otimes I_T)$$

$$= MTN(D^{-1}d, D^{-1}, \beta_3 \in [0, ln(\gamma_1/(\gamma_1 - 1))]) \quad (20)$$

where:

$$D = \frac{1}{100} diag\{v_B\} + X'(\Sigma^{-1} \otimes I_T) X \quad (21)$$

$$d = \frac{1}{100} \mu_B + X'(\Sigma^{-1}\Gamma \otimes I_T) vec(y) \quad (22)$$

$$v_B = [\iota'_{20}\; 0\; \iota'_{13}]' \quad (23)$$

$$\mu_B = [0\; -2\; 0'_{13}\; ln(\gamma_1/(\gamma_1 - 1))\; 0'_{18}]'; \quad (24)$$

We sample B by first sampling β_3 from a univariate truncated normal distribution and the rest of the parameters from the appropriate multivariate normal distribution, given the draw of β_3. Subsequently, we derive the conditional distribution for Σ as:

$$p(\Sigma \mid B, \Gamma, I, \lambda, X_{-I}, y) \propto p(\Sigma)\,\phi_{2T}((\Gamma \otimes I_T)vec(y) \mid XB, \Sigma \otimes I_T)$$
$$= IW_{10+T}((I_2 + e'e)^{-1}) \quad (25)$$
$$e = y - [x_1\delta_1 \; x_2\delta_2] \quad (26)$$

where $IW_\eta(S)$ denotes an *Inverted Wishart* distribution with η degrees of freedom and scale S, which can be easily sampled form using a standard statistics software package. Sampling (γ_1, γ_2), on the other hand, is not trivial and requires a *Metropolis-Hastings* step. That is, the conditional distribution:

$$p(\gamma_1, \gamma_2 \mid B, \Sigma, I, \lambda, X_{-I}, y) \propto p(\gamma_1, \gamma_2)\; p(\beta_0 \mid \gamma_2)\; p(\beta_3 \mid \gamma_1)$$
$$\times \phi_{2T}((\Gamma \otimes I_T)vec(y) \mid XB, \Sigma \otimes I_T) \quad (27)$$

does not have a closed form that can be easily sampled from. Hence, we construct a *jumping* distribution to sample candidate draws from and employ an accept/reject decision rule. A reasonable jumping distribution for this problem is:

$$J(\gamma_1, \gamma_2) = Tt_5(\gamma_1 \mid -\hat{\alpha}_1, w_{11}, 1, \infty)\, Tt_5(\gamma_2 \mid -\hat{\beta}_1 + \frac{w_{12}}{w_{11}}(\gamma_1 + \hat{\alpha}_1), \frac{1}{w_{11}} |W|, -\infty, 0) \quad (28)$$

where $(\hat{\alpha}_1, \hat{\beta}_1)$ and W are the 3SLS estimates and asymptotic covariance matrix, respectively, from the regression:

$$e_1 = \alpha_1 y_2 + \nu_1 \quad (29)$$
$$e_2 = \beta_1 y_1 + \nu_2 \quad (30)$$

where $[e_1\, e_2] = e$ from (26). Note that $J(\gamma_1, \gamma_2)$ is *not* proportional to the mutivariate truncated t distribution, although it is quite similar. We decide whether or not to accept the new draws with the following procedure:

1) calculate $\theta = \dfrac{q(\gamma_1^{new}, \gamma_2^{new} \mid \cdot)/J(\gamma_1^{new}, \gamma_2^{new})}{q(\gamma_1^{old}, \gamma_2^{old} \mid \cdot)/J(\gamma_1^{old}, \gamma_2^{old})}$, where

$p(\gamma_1, \gamma_2 \mid \cdot) \propto q(\gamma_1, \gamma_2 \mid \cdot)$

2) sample $u \sim U(0,1)$

3) accept $(\gamma_1^{new}, \gamma_2^{new})$ *if* $u \le \theta$

Finally, we sample the remaining parameters, $I_1, \ldots, I_T, \lambda$, from the following distributions:

$$I_t \mid B, \Gamma, \Sigma, \lambda, X_{-I}, y \overset{iid}{\sim} Bernoulli(\rho_t) \quad (31)$$
$$\lambda \mid I_1, \cdots, I_T \sim Beta(T\bar{I} + 1, T(1 - \bar{I}) + 1) \quad (32)$$
$$\rho_t = \frac{\lambda\,\phi_{1\times2}(y_t\Gamma' \mid [x_{1t}'\delta_1\; x_{2t,I=1}'\delta_2], I_T, \Sigma^{-1})}{\lambda\,\phi_{1\times2}(y_t\Gamma' \mid [x_{1t}'\delta_1\; x_{2t,I=1}'\delta_2], I_T, \Sigma^{-1}) + (1-\lambda)\,\phi_{1\times2}(y_t\Gamma' \mid [x_{1t}'\delta_1\; x_{2t,I=0}'\delta_2], I_T, \Sigma^{-1})} \quad (33)$$

The above MCMC algorithm was run for 70,000 iterations, with the first 20,000 draws discarded as *burn-in*. The remaining 50,000 draws are assumed to be sampled from the joint posterior distribution and are summarized in table 2. Note that the marginal posterior distributions described by the draws are fairly close to the distributions of the ML estimates reported by Porter (1983). In fact, the only significant difference appears in the estimate of β_3, our main parameter of interest. In particular, note that the posterior distribution asserts $Pr(\beta_3 \le 0.568 \mid X_{-I}, y) = 0.25\%$, which is the region containing the MLE estimate $\hat{\beta}_3 = 0.545$. This difference can be, at least in part, explained by the parameter restrictions imposed in the Bayesian estimation, since we are explicitly forcing $\beta_3 > 0$ and $|\alpha_1| > 1$. Meaningful interpretation of the difference is difficult, however, since interpretation of $\hat{\beta}_3$ is unclear given that $\hat{\alpha}_1 = -0.800$ implies firms are *minimizing* profits, if an equilibrium in the given model exists.

Table 2

Estimation Results

	mean	std dev	Summary of Posterior Distribution $p(\alpha, \beta \mid X_{-I}, y)$ min	0.25%	20%	40%	median	60%	80%	97.5%	max
α_0	7.965	2.679	-3.698	2.696	5.699	7.296	7.973	8.629	10.213	13.211	17.380
α_1	-1.160	0.083	-1.509	-1.337	-1.232	-1.175	-1.153	-1.131	-1.085	-1.021	-1.000
α_2	-0.387	0.127	-0.867	-0.631	-0.494	-0.419	-0.387	-0.355	-0.281	-0.137	0.110
β_0	-4.019	2.670	-14.251	-9.259	-6.232	-4.683	-4.033	-3.370	-1.782	1.174	7.495
β_1	0.259	0.026	0.194	0.216	0.237	0.248	0.255	0.263	0.283	0.317	0.335
$\beta_{2,1}$	-0.198	0.044	-0.435	-0.288	-0.234	-0.208	-0.197	-0.187	-0.162	-0.116	-0.040
$\beta_{2,2}$	-0.227	0.059	-0.483	-0.345	-0.276	-0.241	-0.226	-0.211	-0.178	-0.115	0.002
$\beta_{2,3}$	-0.390	0.048	-0.610	-0.488	-0.429	-0.401	-0.389	-0.377	-0.350	-0.300	-0.223
$\beta_{2,4}$	-0.132	0.117	-0.552	-0.328	-0.221	-0.165	-0.141	-0.116	-0.055	0.098	0.731
β_3	0.624	0.030	0.495	0.568	0.599	0.616	0.624	0.631	0.649	0.684	0.758

In addition to the parameters reported in table 2 (and table 3 in Porter, 1983), it maybe of interest to examine the posterior probabilities of $I_1, \ldots, I_T$. Figure 1 depicts the predicted I_t as well as the posterior probability $Pr(I_t = 1 \mid X_{-I}, y)$ for each period in the sample. Note that our predictions are fairly consistent with the predictions derived by Porter (1983) using the Keifer Algorithm. Another point of interest evident from the figure is that in the later periods with more frequent (predicted) regime switches, the posterior uncertainty is relatively higher. Incidentally, this is also the period of highest market concentration (8 firms), which perhaps indicates that competition not only makes cooperative behavior more difficult in general, but also makes it more difficult to detect.

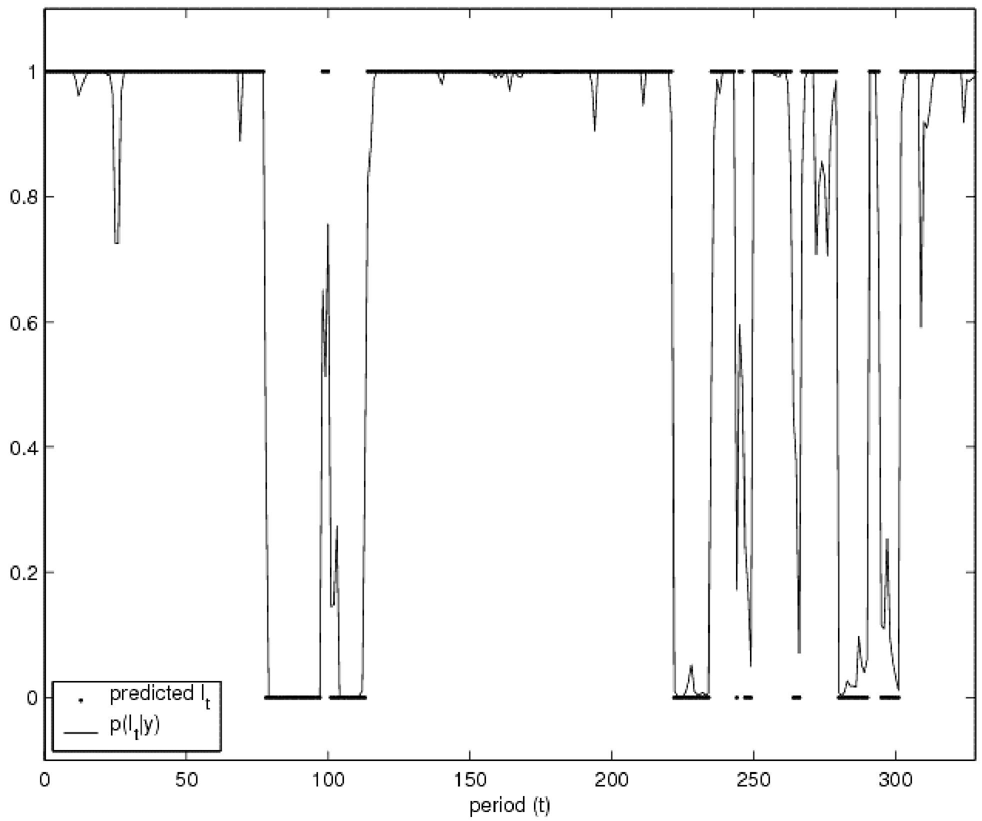

Fig. 1. *Predicted* I_t *and Posterior Probabilities* $Pr(I_t = 1 | X_{-I}, y)$

3. **Inference**

3.1. *The "collusive" regime sets static monopoly prices*

If the collusive regime sets static monopoly prices, then we should observe that $\Lambda = ln(\gamma_1/(1-\gamma_1)) - \beta_3 = 0$. While, "hypothesis testing" doesn't exactly make sense in the Bayesian context, we can perform an analogous exercise by examining the posterior distribution of Λ, which is easily approximated using the simulation draws. Specifically, we can examine the 95% *Highest Probability Density* (HPD) of Λ and check whether $\Lambda = 0$ lies in the posterior HPD. Constructing a 95% HPD given draws of Λ is straightforward using the following procedure:

1) approximate the support of $p(\Lambda | X_{-I}, y)$ with a fine grid of points Λ_i;

2) approximate the posterior density at each Λ_i using non-parametric techniques;

3) sort the grid in descending order by density values;

4) the first m sorted grid points whose densities sum to 0.95 represent the 95% posterior HPD.

Using this algorithm, we find that the 95% HPD for Λ is [0.5806, 2.8308], thus reflecting our posterior belief that it is unlikely for the collusive regime to have set static monopoly prices. This confirms the Porter (1983) prediction that "cooperative period prices exceed those implied by competitive price setting, but are less than those consistent with static joint profit maximizing." The HPD is depicted graphically with the shaded region in figure 2.

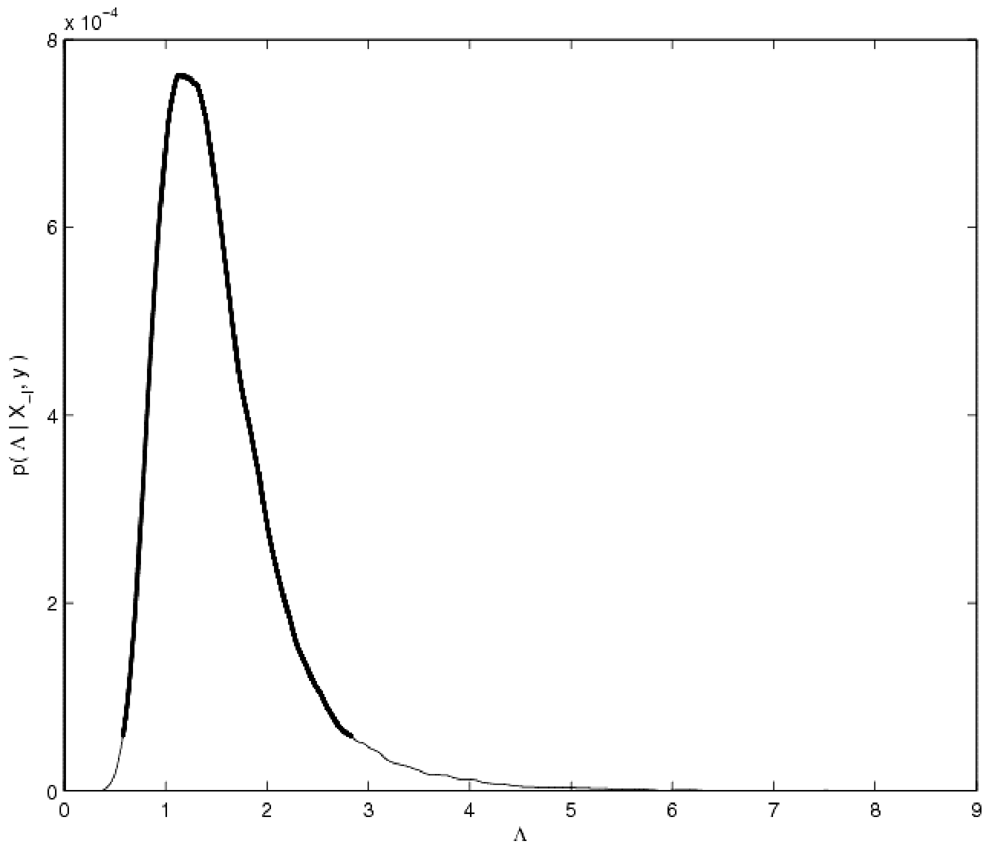

Fig. 2. *Approximated Posterior Density and* 95% *HPD Interval for* Λ

3.2. *There is only one regime in the data*

This hypothesis can be easily tested using the posterior draws of $I_1, \dots, I_T$, removing the need to examine β_3. That is, using the simulated draws we can construct the posterior distribution for the number of regime changes, $\Delta = \sum_{t=2}^{T} | I_t - I_{t-1} |$. Since Δ is discrete, we can explicitly answer the question, “What is the probability that no regime changes occured?”, by estimating the probability $Pr(\Delta = 0 \mid X_{-I}, y)$. Table 3 contains the posterior probability mass function, $Pr(\Delta = d \mid X_{-I}, y)$. Clearly, our posterior belief asserts that $Pr(\Delta = 0 \mid X_{-I}, y) = 0$. In fact, we conclude that at least 12 regime changes occured during the observed period, with the most probability around 22-24 regime switches. Furthermore, note the fact that the pmf heavily favors an even number of regime change is consistent with the posterior predictions that test period began and ended with cooperative regimes.

4. **Conclusion**

The Bayesian procedure and analysis described in the paper compliments the empirical analysis of dynamic oligopoly structure of Porter (1983) by incorporating significant information into the estimation procedure and offering a more extensive inference. In addition, the Bayesian framework may be extended to accommodate a more robust econometric model. For example, it is possible to estimate a similar model without the strict imposition of particular functional forms of supply and demand.

Consider, for example, the type of market demand assumed in the model. The constant demand elasticity reflects a CES utility function for the *representative* consumer, which may be regarded as quite restrictive in the sense that even if we accept that each consumer exhibits a CES utility, the individual demands would not aggregate to a constant elasticity market demand unless each consumer has exactly the same preferences. However, modelling a demand function without a constant elasticity (e.g. quasi-linear utility) is extremely difficult. Even if we provide a better justification for the alternative market demand, a constant elasticity of market demand is a crucial assumption in the aggregation of individual firms’ pricing decisions. Thus, changing the elasticity of the demand curve also affects the industry supply curve. It is possible, in that case, that the new demand/supply specification might entirely explain market price fluctuations in terms of demand shifts or market structure changes (e.g. entry and exit).

Even if we were able to generalize the functional form of the demand curve in a way that nested several possible demand specifications, including constant demand elasticity, estimation through Maximum Likelihood of this type of model would be difficult. Bayesian Model Averaging, on the other hand, allows for a straightforward way of incorporating different types of models, even those that cannot be nested. Such procedures, therefore, can be easily used to obtain more robust estimates.

Table 3

Posterior Probability Mass Function $Pr(\Delta = d \mid X_{-I}, y)$

value	Pr	value	Pr	value	Pr
12	0.0003	**24**	0.1896	**36**	0.0049
13	0.0000	**25**	0.0018	**37**	0.0001
14	0.0064	**26**	0.1512	**38**	0.0023
15	0.0001	**27**	0.0011	**39**	0.0001
16	0.0285	**28**	0.1055	**40**	0.0010
17	0.0004	**29**	0.0008	**41**	0.0000
18	0.0744	**30**	0.0602	**42**	0.0003
19	0.0009	**31**	0.0004	**43**	0.0000
20	0.1376	**32**	0.0309	**44**	0.0001
21	0.0012	**33**	0.0003	**45**	0.0001
22	0.1856	**34**	0.0129	**46**	0.0001
23	0.0014	**35**	0.0001	**48**	0.0001

REFERENCES

- Binder, J. J., *The Sherman Antitrust Act and the Railroad Cartels*, “Journal of Law and Economics” 31(2), 443–468, 1988.
- Ellison, G., *Theories of Cartel Stability and the Joint Executive Committee*, “The RAND Journal of Economics” *25*(1), 37–57, 1994.
- Gelman, A., J. B. Carlin, H. Stern, and D. B. Rubin, *Bayesian Data Analysis* (2 ed.). Chapman & Hall/CRC, 2003.
- Gilchrist, D. T., *Albert Fink and the Pooling System*, “The Business History Review” *34*(1), 34–49, 1960.
- Green, E. J. and R. H. Porter, *Non-cooperative Collusion under Imperfect Price Information*, “Econometrica” *52*(1), 87–100, 1984.
- Koop, G., *Bayesian Econometrics*, John Wiley & Sons, Ltd. (2003).
- Porter, R. H. A Study of Cartel Stability: The Joint Executive Committee, 1880-1886. The Bell Journal of Economics 14(2), 301–314. (1983).
- Prager, R. A., *Using Stock Price Data to Measure the Effects of Regulation: The Interstate Commerce Act and the Railroad Industry*, “The RAND Journal of Economics” *20*(2), 280–290,1989.
- Ripley, W. Z., *The Trunk Line Rate System: A Distance Tariff*, “The Quarterly Journal of Economics” *20*(2), 183–210, 1906.
- Rosen, A., *Identification and Estimation of Firms’ Marginal Cost Functions with Incomplete Knowledge of Strategic Behavior*, Working Paper, 2006.
- Ulen, T. S., *Cartels and Regulation*, Ph.D. thesis, Stanford Univer, 1979.