ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ – ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
ФАКУЛЬТЕТ ЭКОНОМИКИ
КАФЕДРА МАТЕМАТИЧЕСКОЙ ЭКОНОМИКИ И ЭКОНОМЕТРИКИ

# *Домашнее задание №2*

# *по курсу*

# *“Байесовский подход в эконометрическом анализе”*

---

*Выполнила:*
студентка группы 71 ММАЭ
Горностаева Вероника

П*роверила:*
Демидова О.А.

Москва, 2010

1. Выберем для анализа следующие страны:

*Австралия*
*Австрия*
*Аргентина*
*Бельгия*
*Бразилия*
*Великобритания*
*Венгрия*
*Венесуэла*
*Германия*
*Гонконг*
*Греция*

В качестве зависимой переменной возьмем Y – Качество жизни населения, а в качестве независимых – $X_1$ – Загрязнение окружающей среды и $X_2$ – Расходы на личное конечное потребление на душу населения.

2. Оценим модель $Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \varepsilon$, $\varepsilon \sim N\left(0, \frac{1}{h} I_n\right)$ отдельно по данным за 2000, 2001, 2002, 2003, 2004, 2005 годы.

Данные таблицы представлены в файле ДЗ2.xls (каждый лист соответствует оцениваемому году).

3. Построим точечные и интервальные байесовский оценки для параметров $\theta_0, \theta_1, \theta_2, h$ по данным 2005 г.

В качестве $\theta_i^0 = E\theta_i, \quad i = 0, \ldots, 2; \quad h^0 = Eh$ возьмем среднее арифметическое оценок каждого из коэффициентов $\theta_0, \theta_1, \theta_2$ и параметра точности h в исходной модели, оцененной по данным 2000, 2001, 2002, 2003, 2004 гг.

Получаем следующие результаты:

$$\theta_0^0 = E\theta_0 = -1.6201$$
$$\theta_1^0 = E\theta_1 = 0.9446$$
$$\theta_2^0 = E\theta_2 = 0.00025$$
$$h^0 = Eh = 0.73$$

Стандартные отклонения $\Delta_i = \sqrt{D\theta_i}, i = 0,...,2; \Delta_h = \sqrt{Dh}$ рекомендуется вычислить по тем же оценкам с помощью стандартной формулы $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

$$\Delta_0 = \sqrt{D\theta_0} = 1.30915$$
$$\Delta_1 = \sqrt{D\theta_1} = 0.31834$$
$$\Delta_2 = \sqrt{D\theta_2} = 0.00007$$
$$\Delta_h = \sqrt{Dh} = 0.4841$$

Определим значения параметров многомерного гамма-нормального распределения α и β по следующим формулам:

$$\begin{cases} \mathbf{E}h = \dfrac{\alpha}{\beta} = h_0 \\ \mathbf{D}h = \dfrac{\alpha}{\beta^2} = \Delta_h^2. \end{cases} \Longrightarrow \quad \alpha = \frac{h_0^2}{\Delta_h^2} = 2.28 \quad \text{и} \quad \beta = \frac{h_0}{\Delta_h^2} = 3.12$$

Теперь найдем остальные параметры распределения – параметр сдвига $\Theta_0$ и элементы матрицы $\Lambda_0$. Действительно:

$$\mathbf{E}\Theta = \mathbf{E}t\left(2\alpha \mid \Theta_0; \frac{\alpha}{\beta}\Lambda_0\right) = \Theta_0$$

$$\mathbf{\Sigma}_\Theta = \mathbf{\Sigma}_{t(2\alpha|\Theta_0;\frac{\alpha}{\beta}\Lambda_0)} = \frac{2\alpha}{2\alpha-2}\left(\frac{\alpha}{\beta}\Lambda_0\right)^{-1} = \begin{pmatrix} \Delta_0^2 & & \Theta \\ & \Delta_1^2 & \\ \Theta & & \Delta_2^2 \end{pmatrix},$$

где $\Delta_j^2$ – заданные значения априорных дисперсий компонент вектора $\Theta = (\theta_0, \theta_1, \theta_2), \quad j = 0,1,2.$

Таким образом, векторный параметр сдвига в многомерном гамма-нормальном распределении определяется заданным вектором априорных средних значений $\Theta_0$, а диагональные элементы $\lambda_0^{(j)}$ $(j = 0,1,2)$ матрицы $\Lambda_0$ определяются по формулам:

$$\lambda_0^{(j)} = \frac{1}{\Delta_j^2} \cdot \frac{\beta}{\alpha - 1},$$

$$\lambda_0^{(0)} = \frac{1}{\Delta_0^2} \cdot \frac{\beta}{\alpha - 1} = 1.42,$$

$$\lambda_0^{(1)} = \frac{1}{\Delta_1^2} \cdot \frac{\beta}{\alpha - 1} = 24.01,$$

$$\lambda_0^{(2)} = \frac{1}{\Delta_2^2} \cdot \frac{\beta}{\alpha - 1} = 492020442,$$

*Пересчет значений параметров при переходе от априорного сопряженного распределения к апостериорному.*

$$\begin{cases} \tilde{\Theta}_0 = (\Lambda_0 + \mathbf{X}^T\mathbf{X})^{-1}(\Lambda_0\Theta_0 + \mathbf{X}^T Y) \text{ – параметр сдвига;} \\ \tilde{\Lambda}_0 = \Lambda_0 + \mathbf{X}^T\mathbf{X} \text{ – матрица точности;} \\ \widehat{\alpha} = \alpha + \dfrac{n}{2}; \\ \tilde{\beta} = \beta + \dfrac{1}{2}\left[(Y - \mathbf{X}\tilde{\Theta}_0)^T Y + (\Theta_0 - \tilde{\Theta}_0)^T \Lambda_0\Theta_0\right] \end{cases}$$

параметры частного апостериорного гамма-распределения параметра точности $h$

$$\begin{cases} \tilde{\Theta}_0 = \begin{pmatrix} -1.3209 \\ 0.81568 \\ 0.00023 \end{pmatrix} \\ \tilde{\Lambda}_0 = \begin{pmatrix} 12.42 & 66.78 & 134754 \\ 66.78 & 447.63 & 889964 \\ 134754 & 889964 & 2.72*10^9 \end{pmatrix}; \\ \widehat{\alpha} = \alpha + \dfrac{n}{2} = 5.28; \\ \tilde{\beta} = \beta + \dfrac{1}{2}\left[(Y - \mathbf{X}\tilde{\Theta}_0)^T Y + (\Theta_0 - \tilde{\Theta}_0)^T \Lambda_0\Theta_0\right] = 6.88 \end{cases}$$

*Точечные* байесовские оценки параметров $(\theta_0; \theta_1; \theta_2; h)$ определяются средними значениями соответствующих частных апостериорных распределений. С учетом свойств многомерного гамма-нормального распределения имеем:

$$\hat{\Theta}^{(Б)} = \mathbf{E}(\Theta \mid \mathbf{X}, Y) = \tilde{\Theta}_0 = (-1.3209; 0.81568; 0.00023)^{\mathrm{T}};$$

$$\hat{h}^{(Б)} = \mathbf{E}(h \mid \mathbf{X}, Y) = \frac{\tilde{\alpha}}{\tilde{\beta}} = 0.76742.$$

При выводе *интервальных* байесовских оценок используются те же свойства многомерного гамма-нормального распределения, а также факт $t(2\tilde{\alpha})$ – распределенности случайных величин $(\hat{\theta}_j^{(Б)} - \theta_j)\sqrt{\tilde{c}_j}$, где параметр точности $\tilde{c}_j$ вычисляется по блочным компонентам матрицы точности $\tilde{B}$ частного апостериорного обобщенного многомерного $t(2\tilde{\alpha} \mid \hat{\Theta}^{(Б)}; \tilde{B})$ – распределения по формуле

$$\tilde{c}_j = \tilde{b}_{jj} - \tilde{B}_{j.} \cdot \left(\tilde{B}(j)\right)^{-1} \cdot \tilde{B}_{.j}$$

Участвующие в этом соотношении число $\tilde{b}_{jj}$, $1 \times (k-1)$-матрица $\tilde{B}_{j.}$, $(k-1) \times 1$-матрица $\tilde{B}_{.j}$ и $(k-1) \times (k-1)$-матрица $\tilde{B}(j)$ определяются следующим блочным представлением матрицы $\tilde{B}$:

$$\tilde{B} = \begin{pmatrix} \tilde{b}_{jj} & \tilde{B}_{j.} \\ \tilde{B}_{.j} & \tilde{B}(j) \end{pmatrix}$$

В нашем случае $k = 3$, $j = 0,1,2$, матрица

$$\tilde{B} = \frac{\tilde{\alpha}}{\tilde{\beta}} \tilde{\Lambda}_0 = \begin{pmatrix} 9.531 & 51.248 & 103413.09 \\ 51.248 & 343.521 & 682977.074 \\ 103413.09 & 682977.074 & 2086270940 \end{pmatrix},$$

так что $\tilde{c}_0 = 8.149$, $\tilde{c}_1 = 51.254$, $\tilde{c}_2 = 727170357.1$.

Следовательно, с вероятностью $P_0 = 0,90$ мы можем утверждать, что $|\hat{\theta}_0^{(Б)} - \theta_0| \cdot \sqrt{8.149} < t_{0,05}(11)$, $|\hat{\theta}_1^{(Б)} - \theta_1| \cdot \sqrt{51.254} < t_{0,05}(11)$ и $|\hat{\theta}_2^{(Б)} - \theta_2| \cdot \sqrt{727170357.1} < t_{0,05}(11)$, так что (с учетом того, что $t_{0,05}(11) = 2.20$) имеем:

$\theta_0 \in [-2.092; -0.55]$ с вероятностью $P_0 = 0,95$;

$\theta_1 \in [0.508;\ 1.123]$ с вероятностью $P_0 = 0,95$;

$\theta_2 \in [0.00015;\ 0.0003]$ с вероятностью $P_0 = 0,95$.

Поскольку параметр $h$ подчиняется апостериорному гамма-распределению с параметрами $\tilde{\alpha}$ и $\tilde{\beta}$, то:

$$h \in \left[\gamma_{0,95}(\tilde{\alpha};\tilde{\beta});\gamma_{0,05}(\tilde{\alpha};\tilde{\beta})\right] \text{ с вероятностью } P_0 = 0,90.$$

Используя соотношение $\gamma_q(\tilde{\alpha};\tilde{\beta}) = \frac{1}{2\tilde{\beta}}\chi_q^2(2\tilde{\alpha})$, имеем (с учетом $\chi_{0,95}^2(11) \approx 4.57$ и $\chi_{0,05}^2(11) \approx 19.68$):

$h \in [0.332;1.429]$ с вероятностью $P_0 = 0,95.$

Сравним полученные точечные и интервальные байесовский оценки параметров с оценками МНК этих же параметров (точечными и интервальными).

$$\hat{\Theta}_{\text{мп}} = \hat{\Theta}_{\text{мнк}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}Y = (0.978;\,0.38;\,0.00026)^{\mathrm{T}}.$$

$$\hat{h}_{\text{мп}} = \left[\frac{1}{4}(Y - \mathbf{X}\hat{\Theta}_{\text{мп}})^{\mathrm{T}}(Y - \mathbf{X}\hat{\Theta}_{\text{мп}})\right]^{-1} = \ 0.8344;$$

-1.8926<$\theta_0 < 3.8488$; $-0.2075 < \theta_1 < 0.9677$; $0.000157 < \theta_2 < 0.000366$
с вероятностью $P_0 = 0,95$;

$h \in [0.1483;1.979]$ с веротяностью $P_0 = 0,95.$

При оценке модели 2005г. методом наименьших квадратов константа и коэффициент при переменной загрязнение окружающей среды оказались незначимыми, поэтому достаточно велики получились расхождения в оценках МНК и с помощью байесовского подхода для этих X.

Сравним интервальные оценки. Мы видим, что байесовский подход позволяет сузить доверительный интервал для $\theta_0$ в 3,7 раз, для $\theta_1$ — в 1,91 раз, для $\theta_2$ - в 1.28 раз и для $h$ — в 1,7 раза по сравнению с подходом, основанным на методе максимального правдоподобия.

4. Выберем две страны из определенного списка – Австралия и Австрия.

Построим:

а) точечный и интервальный байесовский прогноз для значений зависимой переменной на 2006 г. для этих стран;

$$\tilde{\mathbf{X}} = \begin{pmatrix} 1 & 7.29 & 20013 \\ 1 & 8.65 & 20633 \end{pmatrix}; \ \tilde{Y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = ?$$

Плотность условного распределения вектора $\tilde{Y}$ при заданных $\mathbf{X}, \tilde{\mathbf{X}}$ и $Y$ описывается обобщенным многомерным $t$-распределением с числом степеней свободы $v = n - k - 1 = 6 - 2 - 1 = 3$, параметром сдвига $\begin{pmatrix} 1 & 7.29 & 20013 \\ 1 & 8.65 & 20633 \end{pmatrix} \begin{pmatrix} -1.3209 \\ 0.81568 \\ 0.00023 \end{pmatrix}$ и матрицей точности $\mathbf{B}' = \frac{\tilde{\alpha}}{\tilde{\beta}} \left[ \mathbf{I}_q - \tilde{\mathbf{X}}(\Lambda_0 + \mathbf{X}^{\mathrm{T}}\mathbf{X} + \tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^{\mathrm{T}} \right]$.

Произведя необходимые вычисления по формулам и используя известные свойства обобщенного многомерного $t$-распределения, имеем:

$$\hat{y}_{n+m}(\text{прогнозное}) = (\hat{\Theta}^{(\text{Б})})^{\mathrm{T}} \cdot X_{n+m}, \quad m = 1;$$

$$\underline{\hat{y}_1(\text{прогн.}) = -1.321 + 0.816 \cdot 7.29 + 0.00023 \cdot 20013 = 9.237;}$$

$$\underline{\hat{y}_2(\text{прогн.}) = -1.321 + 0.816 \cdot 8.65 + 0.00023 \cdot 20633 = 10.49;}$$

$$B' = \begin{pmatrix} 0.679 & -0.0966 \\ -0.0966 & 0.645 \end{pmatrix};$$

$$y_{n+m} \in \left[ \hat{y}_{n+m} - t_{\frac{1-P_0}{2}}(n-k-1) \cdot \frac{1}{\sqrt{c'_m}}; \hat{y}_{n+m} + t_{\frac{1-P_0}{2}}(n-k-1) \cdot \frac{1}{\sqrt{c'_m}} \right],$$

$$\underline{y_1 \in [5.333; 13.243] \text{ с вероятностью } P_0 = 0,95;}$$

$$\underline{y_2 \in [6.484; 14.495] \text{ с вероятностью } P_0 = 0,95;}$$

б) двумерную доверительную область для этих двух стран.

$$\frac{1}{q}(\tilde{Y}-\tilde{\mathbf{X}}\hat{\Theta}^{(\text{Б})})^{\text{T}}B'(\tilde{Y}-\tilde{\mathbf{X}}\hat{\Theta}^{(\text{Б})})<F_{1-P_0}(q;\ n-k-1),$$

где $F_{0.05}(\nu_1,\nu_2)$ — 5%-ная точка $F(\nu_1,\nu_2)$- распределения, $\hat{\Theta}^{(\text{Б})}$ — байесовская точечная оценка параметров регрессии $\Theta$. Эта область имеет форму $q$-мерного эллипсоида.

$$\Delta_{\tilde{Y}}=\left\{\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}:\ \frac{1}{2}\begin{pmatrix} y_1-9.237 \\ y_2-10.49 \end{pmatrix}^{\text{T}}\begin{pmatrix} 0.679 & -0.0966 \\ -0.0966 & 0.645 \end{pmatrix}\begin{pmatrix} y_1-9.237 \\ y_2-10.49 \end{pmatrix}<9.55\right\}$$