

Lecture IV: Elements of Bayesian Inference

Charles B. Moss

May 2, 2013

- I. Geweke, J. 2005. *Contemporary Bayesian Econometrics and Statistics*. Hoboken, New Jersey: John Wiley & Sons – Chapter 2.

II. Basics

- A. Geweke develops the Bayesian system as a learning process – $Y_t = \{y_s\}_{s=1}^t$ is a sequence of possibly vector valued variables $y_t \in \psi_t$, $Y_t \in \Psi_t$, and $\psi_0 = \Psi_0 = \emptyset$. The probability structure for this sequence can then be presented as

$$p(y_t | Y_{t-1}, \theta_A, A). \quad (1)$$

1. For example, consider a sequence of corn prices $\{p_{Cs}\}_{s=1}^t$, input prices such as the price of fuel $\{p_{Fs}\}_{s=1}^t$, and a weather index such as the Palmer Drought Severity Index $\{I_{Ws}\}_{s=1}^t$

$$y_s = \begin{bmatrix} p_{Cs} \\ p_{Fs} \\ I_{Ws} \end{bmatrix} \quad (2)$$

The data for this formulation are presented in Table 1.

2. The sequence of observed outcomes could then be written as

$$Y_t = \left\{ \begin{bmatrix} p_{Cs} \\ p_{Fs} \\ I_{Ws} \end{bmatrix} \right\}_{s=1}^t \quad (3)$$

3. We could define the Model A as a finite ordered Vector Autoregression (VAR)

Table 1: Corn Production, Price, and Weather Data for Arkansas

Year	Corn Yield	Nominal		Real		Weather Index	
		Corn Price	Fuel Price	Corn Price	Fuel Price	Mean	Std Dev
2012	178	6.95	3.70	6.95	3.70	-2.183	1.406
2011	142	6.27	3.50	6.38	3.56	-0.944	1.674
2010	150	4.55	2.48	4.74	2.58	1.088	1.330
2009	148	3.79	1.60	4.03	1.70	1.531	1.106
2008	155	4.42	3.54	4.70	3.76	1.455	0.603
2007	169	3.80	2.26	4.17	2.48	-1.377	0.750
2006	146	2.73	2.21	3.08	2.49	-2.822	0.175
2005	131	2.15	1.92	2.49	2.22	-1.050	0.854
2004	140	2.39	1.38	2.85	1.65	0.263	0.710
2003	140	2.37	1.15	2.90	1.40	-0.381	1.028
2002	134	2.43	0.95	3.03	1.19	1.404	0.848
2001	145	2.02	1.01	2.56	1.28	-0.334	0.737
2000	130	1.75	1.08	2.26	1.39	-1.396	1.381
1999	130	1.74	0.75	2.30	0.98	-0.407	0.310
1998	100	1.85	0.65	2.48	0.88	-0.538	1.439

$$y_t = \beta_0 + \sum_{k=1}^K B_K y_{t-k} + \epsilon_t \quad (4)$$

the unobservable values are the vector β_0 , the K autoregression matrices $\{B_k\}_{k=1}^K$, and the parameters of the distribution for ϵ_t – say its variance matrix Ω . With each observation, it is possible to "learn" some more about the unknown parameters. In general the probability density function can be written as

$$p(Y_T | \theta_A, A) = \prod_{t=1}^T p(y_t | Y_{t-1}, \theta_A, A). \quad (5)$$

4. If the outcomes are independent – if y_t is not dependent on y_{t-1} either based on B_k or because $E[\epsilon_t \epsilon'_{t-s}] = 0$

$$p(y_t | Y_{t-1}, \theta_A, A) = p(y_t | \theta_A, A) \quad (6)$$

or

$$p(Y_T | \theta_A, A) = \prod_{t=1}^T p(y_t | \theta_A, A). \quad (7)$$

As one example of independent outcomes, consider the Translog cost function

$$\begin{aligned} \ln(C) = & \alpha_0 + \alpha_1 \ln(w_1) + \alpha_2 \ln(w_2) + \alpha_3 \ln(w_3) + A_{11} \ln(w_1) \ln(w_1) / 2 + \\ & A_{12} \ln(w_1) \ln(w_2) + A_{22} \ln(w_2) \ln(w_2) / 2 + A_{23} \ln(w_2) \ln(w_3) + \\ & A_{33} \ln(w_3) \ln(w_3) / 2 + \beta \ln(q) + B_{11} \ln(q) \ln(q) / 2 + \\ & \Gamma_{11} \ln(w_1) \ln(q) + \Gamma_{21} \ln(w_2) \ln(q) + \Gamma_{31} \ln(w_3) \ln(q) + \epsilon_1 \\ s_1 = & \alpha_1 + A_{11} \ln(w_1) + A_{12} \ln(w_2) + A_{13} \ln(w_3) + \Gamma_{11} \ln(q) + \epsilon_2 \\ s_2 = & \alpha_2 + A_{12} \ln(w_1) + A_{22} \ln(w_2) + A_{23} \ln(w_3) + \Gamma_{21} \ln(q) + \epsilon_3 \\ s_3 = & \alpha_3 + A_{13} \ln(w_1) + A_{23} \ln(w_2) + A_{33} \ln(w_3) + \Gamma_{31} \ln(q) + \epsilon_4 \end{aligned} \quad (8)$$

where w_1 , w_2 , and w_3 are input prices, q is the output level, s_1 , s_2 , and s_3 are the share of cost for each input, and ϵ_1 , ϵ_2 , ϵ_3 , and ϵ_4 are the residuals. This formulation can be used

Table 2: Jorgenson KLEM (Capital, Labor, Energy, and Material) Data for Agriculture

Year	Output	Capital		Labor		Energy		Material	
		Quantity	Price	Quantity	Price	Quantity	Price	Quantity	Price
1960	153474	35474.5	0.2022	100351.0	0.1023	5196.6	0.1676	105589	0.2701
1961	156256	35714.3	0.2199	98162.3	0.1080	5320.1	0.1687	106874	0.2724
1962	155497	36121.5	0.2142	96396.5	0.1162	5508.9	0.1673	107933	0.2791
1963	160845	36847.0	0.2169	95453.7	0.1210	5700.5	0.1668	111161	0.2792
1964	159818	37528.9	0.1840	94510.9	0.1385	5655.4	0.1630	109132	0.2796
1965	164301	38256.0	0.2211	86246.5	0.1520	5839.9	0.1666	113028	0.2884
1966	163392	39281.7	0.2423	84037.2	0.1668	6102.9	0.1697	115790	0.3057
1967	169478	40297.8	0.2218	82051.1	0.1793	6269.0	0.1716	115553	0.3013
1968	171661	41925.8	0.2135	78578.8	0.1980	6491.2	0.1705	115996	0.3082
1969	173386	42162.3	0.3002	76414.2	0.2140	6956.3	0.1727	120276	0.3276
1970	174451	42943.0	0.3021	75021.7	0.2196	7155.6	0.1782	119990	0.3408
1971	183510	43468.1	0.3598	72182.6	0.2137	7301.0	0.1902	121859	0.3453
1972	186534	43516.7	0.4927	73255.9	0.2171	8234.6	0.1972	135623	0.3827
1973	190194	44315.1	0.5714	69179.4	0.2445	10088.5	0.2300	159667	0.4699
1974	183488	45337.8	0.5905	64548.4	0.2683	9430.8	0.3314	149627	0.5149
1975	188571	45708.6	0.5761	65179.6	0.2789	8971.2	0.3831	142520	0.5305
1976	192838	46163.3	0.5503	71072.6	0.3067	9272.9	0.4134	143341	0.5469
1977	200567	47826.1	0.5369	68440.4	0.3272	8730.9	0.4606	145657	0.5580
1978	203682	48754.6	0.8061	67991.2	0.3622	9480.2	0.4930	151936	0.6235
1979	220251	49763.9	1.0361	66170.8	0.3769	10596.4	0.6173	166234	0.6800
1980	208824	50125.1	1.1602	67244.9	0.3967	10186.8	0.8561	157684	0.7430
1981	230896	50418.4	1.1889	67313.8	0.4048	8620.4	1.0072	156314	0.7484
1982	234524	50717.8	1.0247	66520.9	0.4834	9312.6	1.0123	166379	0.7671
1983	215450	49819.9	0.7640	81869.0	0.4403	8403.9	0.9764	142506	0.7816
1984	231061	48978.0	0.9717	73590.2	0.4820	9945.0	0.9752	145318	0.7999
1985	245257	48256.5	0.9075	71519.1	0.5435	8869.2	0.9627	145262	0.7809
1986	245492	47715.7	0.9587	70930.2	0.5594	7439.1	0.7894	138931	0.7738
1987	254229	47261.5	1.1383	70399.2	0.5791	8314.3	0.8084	144895	0.8029
1988	243603	46627.4	1.1706	76095.8	0.6128	8120.2	0.7993	145481	0.8691
1989	256601	46891.1	1.1241	74135.0	0.6986	8452.1	0.8589	145955	0.8873
1990	268591	47213.8	1.0449	72955.2	0.7851	8621.7	0.9602	150062	0.9027
1991	269660	47103.2	0.9386	76571.8	0.7857	8002.6	0.9314	148325	0.9048
1992	285108	47192.3	1.0462	74753.0	0.8084	8462.2	0.9260	151953	0.9052
1993	270593	48062.7	1.2479	75275.8	0.8717	8381.3	0.9170	150312	0.9435
1994	299636	48510.2	1.4118	77155.3	0.8792	8603.5	0.9042	156693	0.9367
1995	294954	49007.9	1.4445	85809.6	0.8950	8086.9	0.9289	158422	0.9741
1996	305745	49929.8	1.0000	82818.2	1.0000	8852.2	1.0000	164145	1.0000
1997	304443	51173.2	1.2752	84646.5	1.0145	8472.4	0.9930	164514	1.0159
1998	307862	52505.3	1.5733	83157.0	1.0981	8728.9	0.8713	159117	0.9975
1999	319071	53171.0	1.6789	89571.2	1.1672	9416.5	0.9253	164197	1.0068
2000	348475	53515.9	1.9269	87635.4	1.2034	9622.8	1.2126	177534	1.0226
2001	346322	53672.7	1.8653	81945.6	1.2161	9886.5	1.1738	177848	1.0360
2002	346868	53844.9	1.5460	68581.6	1.2588	9374.6	1.1287	170610	1.0289
2003	359890	54275.4	2.0205	75478.1	1.2697	9614.7	1.3013	182605	1.0814
2004	382321	55219.1	2.0883	67169.8	1.4252	10207.2	1.5077	199310	1.1401
2005	378033	56024.9	2.2702	59610.4	1.6277	9077.5	1.9199	189927	1.1630

to estimate changes in productivity (which is Jorgenson's primary use of this dataset) or possibly examine the effect of an increase in energy prices on the demand for other inputs. The KLEM (Capital, Labor, Energy, and Materials) data for agriculture in the United States provided by Dale W. Jorgenson are presented in Table 2. The transformations required to estimate the Translog cost system are presented in Table 3.

5. There is a transformation or reversal to the probability of the structure after the sample is observed.
 - a. Before a sample is drawn – those values (Y_T) are unknown. In that case we can talk about the probability of a set of potentially observable events

$$Y_T \Rightarrow p(Y_t | \theta_A, A). \quad (9)$$

- b. However, after the sample is drawn or observed $Y_T \Rightarrow Y_T^0$ it is no longer random. We can talk about *likelihood function* which says something about the probability of the parameters – θ_A is the random variable of interest

$$\theta_A \Rightarrow L(\theta_A; Y_T^0, A) \quad (10)$$

- c. Linking the two together

$$L(\theta_A; Y_T^0, A) \propto p(Y_t^0 | \theta_A, A) \quad (11)$$

Consider the first five outcomes for the first share equation in Equation 8

$$p(\{Y_t\}_{t=1960}^{1964} | \{\alpha_1, A_{11}, A_{12}, A_{13}, \Gamma_{11}, \sigma_{\epsilon_2}^2\}, A) = \left(\frac{1}{\sqrt{2\pi\sigma_{\epsilon_2}^2}} \right)^5 \times \exp \left[- \frac{\sum_{t=1960}^{1964} (s_{1t} - \alpha_1 - A_{11} \ln(w_{1t}) - A_{12} \ln(w_{2t}) - A_{13} \ln(w_{3t}) - \Gamma_{11} \ln(y_t))^2}{2\sigma_{\epsilon_2}^2} \right]. \quad (12)$$

Table 3: KLEM Data for Translog Cost Function

Year	Factor Shares			Ln(Output)	Normalized Price		
	Capital	Labor	Energy		Capital	Labor	Energy
1960	0.1532	0.2192	0.0186	11.9413	-0.2894	-0.9707	-0.4768
1961	0.1620	0.2187	0.0185	11.9593	-0.2144	-0.9253	-0.4794
1962	0.1548	0.2241	0.0184	11.9544	-0.2646	-0.8764	-0.5117
1963	0.1551	0.2242	0.0185	11.9882	-0.2523	-0.8360	-0.5148
1964	0.1343	0.2545	0.0179	11.9818	-0.4185	-0.7026	-0.5397
1965	0.1534	0.2378	0.0177	12.0095	-0.2654	-0.6403	-0.5484
1966	0.1587	0.2338	0.0173	12.0039	-0.2324	-0.6057	-0.5884
1967	0.1501	0.2471	0.0181	12.0405	-0.3064	-0.5192	-0.5632
1968	0.1459	0.2536	0.0180	12.0533	-0.3671	-0.4424	-0.5920
1969	0.1818	0.2349	0.0173	12.0633	-0.0874	-0.4258	-0.6404
1970	0.1812	0.2301	0.0178	12.0694	-0.1203	-0.4394	-0.6484
1971	0.2098	0.2070	0.0186	12.1200	0.0412	-0.4797	-0.5962
1972	0.2360	0.1750	0.0179	12.1364	0.2528	-0.5669	-0.6631
1973	0.2117	0.1414	0.0194	12.1558	0.1954	-0.6534	-0.7146
1974	0.2155	0.1394	0.0251	12.1199	0.1370	-0.6519	-0.4408
1975	0.2131	0.1471	0.0278	12.1472	0.0825	-0.6429	-0.3256
1976	0.1963	0.1684	0.0296	12.1696	0.0063	-0.5783	-0.2797
1977	0.1925	0.1679	0.0302	12.2089	-0.0387	-0.5339	-0.1919
1978	0.2406	0.1508	0.0286	12.2243	0.2568	-0.5432	-0.2349
1979	0.2630	0.1272	0.0334	12.3025	0.4212	-0.5901	-0.0967
1980	0.2760	0.1266	0.0414	12.2492	0.4456	-0.6275	0.1417
1981	0.2816	0.1280	0.0408	12.3497	0.4629	-0.6145	0.2970
1982	0.2350	0.1454	0.0426	12.3653	0.2896	-0.4618	0.2774
1983	0.1965	0.1861	0.0424	12.2805	-0.0228	-0.5739	0.2225
1984	0.2277	0.1697	0.0464	12.3504	0.1945	-0.5065	0.1982
1985	0.2140	0.1900	0.0417	12.4101	0.1502	-0.3624	0.2094
1986	0.2301	0.1996	0.0295	12.4110	0.2143	-0.3244	0.0200
1987	0.2472	0.1873	0.0309	12.4460	0.3491	-0.3267	0.0069
1988	0.2331	0.1992	0.0277	12.4033	0.2978	-0.3494	-0.0837
1989	0.2185	0.2147	0.0301	12.4553	0.2366	-0.2391	-0.0325
1990	0.1971	0.2288	0.0331	12.5009	0.1463	-0.1395	0.0618
1991	0.1797	0.2445	0.0303	12.5049	0.0367	-0.1411	0.0290
1992	0.1935	0.2368	0.0307	12.5606	0.1448	-0.1131	0.0228
1993	0.2180	0.2385	0.0279	12.5084	0.2796	-0.0791	-0.0285
1994	0.2354	0.2332	0.0267	12.6103	0.4102	-0.0634	-0.0353
1995	0.2288	0.2482	0.0243	12.5946	0.3940	-0.0847	-0.0475
1996	0.1633	0.2709	0.0290	12.6305	0.0000	0.0000	0.0000
1997	0.1998	0.2629	0.0258	12.6262	0.2273	-0.0013	-0.0228
1998	0.2428	0.2684	0.0224	12.6374	0.4556	0.0961	-0.1353
1999	0.2427	0.2842	0.0237	12.6732	0.5113	0.1478	-0.0845
2000	0.2566	0.2625	0.0290	12.7613	0.6336	0.1628	0.1704
2001	0.2531	0.2519	0.0293	12.7551	0.5881	0.1603	0.1249
2002	0.2340	0.2427	0.0297	12.7567	0.4072	0.2017	0.0926
2003	0.2639	0.2307	0.0301	12.7936	0.6251	0.1605	0.1851
2004	0.2542	0.2110	0.0339	12.8540	0.6053	0.2232	0.2795
2005	0.2750	0.2098	0.0377	12.8427	0.6688	0.3361	0.5012

After the outcomes have been observed, this function becomes

$$L(\{\alpha_1, A_{11}, A_{12}, A_{13}, \Gamma_{11}, \sigma_{\epsilon_2}^2\} ; Y_T^0, A) = \left(\frac{1}{\sqrt{2\pi\sigma_{\epsilon_2}^2}} \right)^5$$

$$\times \exp \left[\left((0.1532 - \alpha_1 + A_{11}0.2894 + A_{12}0.9707 + A_{13}0.4768)^2 + \right. \right.$$

$$(0.1620 - \alpha_1 + A_{11}0.2144 + A_{12}0.9253 + A_{13}0.4794)^2$$

$$(0.1548 - \alpha_1 + A_{11}0.2646 + A_{12}0.8764 + A_{13}0.5117)^2$$

$$(0.1551 - \alpha_1 + A_{11}0.2523 + A_{12}0.8360 + A_{13}0.5148)^2$$

$$\left. (0.1343 - \alpha_1 + A_{11}0.4185 + A_{12}0.7026 + A_{13}0.5397)^2 \right) / (2\sigma_{\epsilon_2}^2) \Big]$$
(13)

So the unknown variables in Equation 13 are the parameters $\{\alpha_1, A_{11}, A_{12}, A_{13}, \Gamma_{11}, \sigma_{\epsilon_2}^2\}$.

6. Given the independent sample distribution

$$p(Y_t | \theta_A, A) = \prod_{t=1}^T p(y_t | \theta_A, A) \quad (14)$$

and a prior density function for θ_A ($p(\theta_A | A)$) we can derive the joint distribution function for the sample and the parameters

$$p(y, \theta_A | A) = p(\theta_A | A) \times p(y | \theta_A, A). \quad (15)$$

Alternatively if we observe the distribution function for y and the conditional distribution function for θ_A based on y , we can define the joint distribution function as

$$p(y, \theta_A | A) = p(y | A) \times p(\theta_A | y, A). \quad (16)$$

Following Geweke, the marginal density function for y is

$$p(y | A) = \int_{\Theta_A} p(y, \theta_A | A) dv(\theta_A)$$

$$= \int_{\Theta_A} p(\theta_A | A) p(y | \theta_A, A) dv(\theta_A). \quad (17)$$

For example, consider the simplest form of the share model

$$s_{1t} = \alpha_1 + \epsilon_{2t}. \quad (18)$$

Assuming that the variance is a known constant k

$$p(y|\alpha_1, A) = \left(\frac{1}{\sqrt{2\pi k}}\right)^3 \exp\left[-\frac{1}{2k}((s_{11} - \alpha_1)^2 + (s_{12} - \alpha_1)^2 + (s_{13} - \alpha_1)^2)\right] \quad (19)$$

Next, assume that our prior for α_1 is normal with a mean of 0.25 and a variance of 0.125

$$p(\alpha_1|A) = \frac{1}{\sqrt{2\pi \times 0.125}} \exp\left[-\frac{(\alpha_1 - 0.25)^2}{2 \times 0.125}\right]. \quad (20)$$

The joint distribution function then becomes

$$\begin{aligned} p(\{s_{11}, s_{12}, s_{13}\}, \alpha_1|A) &= K \\ &\times \exp\left[-\frac{1}{2k}((s_{11} - \alpha_1)^2 + (s_{12} - \alpha_1)^2 + (s_{13} - \alpha_1)^2)\right] \times \exp\left[-\frac{(\alpha_1 - 0.25)^2}{2 \times 0.125}\right] \\ K &= \left(\frac{1}{\sqrt{2\pi k}}\right)^3 \frac{1}{\sqrt{2\pi \times 0.125}}. \end{aligned} \quad (21)$$

Assume that we observe the first three results from Table 3.
In that case, Equation 21 becomes

$$\begin{aligned} p(\{0.1532, 0.1620, 0.1548\}, \alpha_1|A) &= K \\ &\times \exp\left[-\frac{1}{2k}((0.1532 - \alpha_1)^2 + (0.1620 - \alpha_1)^2 + (0.1548 - \alpha_1)^2)\right] \times \exp\left[-\frac{(\alpha_1 - 0.25)^2}{2 \times 0.125}\right] \\ K &= \left(\frac{1}{\sqrt{2\pi \times 0.1125}}\right)^3 \frac{1}{\sqrt{2\pi \times 0.125}}. \end{aligned} \quad (22)$$

Given a value of k (say $k = 0.1125$) we could compute the value of $p(\{0.1532, 0.1620, 0.1548\} | A)$ by integrating out α_1

$$p(\{0.1532, 0.1620, 0.1548\} | A) = K \times \int_{-\infty}^{\infty} \exp \left[-\frac{1}{2 \times 0.1125} \left((0.1532 - \alpha_1)^2 + (0.1620 - \alpha_1)^2 + (0.1548 - \alpha_1)^2 \right) \right] \times \exp \left[-\frac{(\alpha_1 - 0.25)^2}{2 \times 0.005} \right] d\alpha_1$$

$$K = \left(\frac{1}{\sqrt{2\pi \times 0.1125}} \right)^3 \frac{1}{\sqrt{2\pi \times 0.125}}. \quad (23)$$

Using Mathematica to numerically integrate the integral in Equation 23

$$\int_{-\infty}^{\infty} \exp \left[-\frac{1}{2 \times 0.1125} \left((0.1532 - \alpha_1)^2 + (0.1620 - \alpha_1)^2 + (0.1548 - \alpha_1)^2 \right) \right] \times \exp \left[-\frac{(\alpha_1 - 0.25)^2}{2 \times 0.125} \right] d\alpha_1 = 0.414389 \quad (24)$$

$K = 1.8987$ so the overall probability is equal to 0.786801 . However, it will be useful in a couple of steps to write the result as

$$p(\{0.1532, 0.1620, 0.1548\} | A) = K \times 0.786801 \ni K = 1.8987 \quad (25)$$

B. Next for a couple of definitions:

1. **Definition 2.1.1** The *marginal likelihood* of the model A is $p(y^0 | A)$
2. The conditional density is then

$$p(\theta_A | y, A) = \frac{p(\theta_A | A) \times p(y | \theta_A, A)}{p(y | A)} \quad (26)$$

3. **Definition 2.1.2** The *posterior density* of the vector of unobservables θ_A in model A is

$$p(\theta_A | y^0, A) = \frac{p(\theta_A | A) \times p(y^0 | \theta_A, A)}{p(y^0 | A)} \quad (27)$$

4. Using the joint distribution from Equation 22 and the marginal distribution from Equation 25

$$p(\alpha_1 | \{0.1532, 0.1620, 0.1548\}, A) =$$

$$K \exp \left[-\frac{\left((0.1532 - \alpha_1)^2 + (0.1620 - \alpha_1)^2 + (0.1548 - \alpha_1)^2 \right)}{2 \times 0.1125} - \frac{(\alpha_1 - 0.25)^2}{2 \times 0.125} \right] \quad (28)$$

$$K \times 0.786801$$