Bayesian “VAR” Averaging

Saki Bigio
BCRP

Roddy Rivas-Llosa
APCI

“In no affairs of mere prejudice, pro or con, do we deduce inferences with entire certainty even with the most simple data.”

The Narrative of Arthur Gordon Pym of Nantucket

-Edgar Allan Poe-

Motivación

- Tenemos incertidumbre sobre los parámetros pero también tenemos incertidumbre sobre el modelo!

- Tenemos fuertes limitaciones en cuanto a grados de libertad, por lo tanto no podemos incluir todas las variables a la vez.

- Estimaciones con distintas especificaciones hacen variar los coeficientes de manera significativa. Nos gustaría conocer que tan volátiles son estos parámetros tomando en cuenta la incertidumbre del modelo.

El teorema de Bayes:

- Si le asignamos probabilides a los modelos de ser los correctos:

$$pr(\Delta / D) = \sum_{k=1}^{2^K} pr(\Delta \setminus M_k) pr(M_k \setminus D)$$

- Y gracias al teorema de Bayes sabemos que:

$$pr(M_i / D) = \frac{pr(D \setminus M_i) pr(M_i)}{\sum_{k=1}^{2^K} pr(D \setminus M_k) pr(M_k)}$$

Insumos:

- Función de Verosimilitud:

$$pr(D \setminus M_i) = \int pr(D \setminus \theta_i, M_i) pr(\theta_i \setminus M_i) d\theta_i$$

- Y un PRIOR por modelo:

$$pr(M_i)$$

Productos:

- Estimación de Parámetros o Predicciones posteriores:

$$E(\Delta / D) = \sum_{k=1}^{2^K} \hat{\Delta}_k P(M_k \setminus D)$$

- Estimación de Parámetros o Predicciones posteriores:

$$VAR(\Delta) = \sum_{k=1}^{2^k} (VAR[\Delta \setminus D, M_k] P(M_k \setminus D) + (\hat{\Delta}_k - E[\Delta \setminus D]) P(M_k \setminus D))^2$$

Un ejemplo del Producto Final: Sala-i-Martin (AER 2004)

Table 12.6
Baseline Estimation for All 67 Variables

Rank	Variable	Fraction of Regressions with $\|t$ stat$\| > 2$ (1)	Posterior Inclusion Probability (2)	Posterior Mean Conditional on Inclusion (3)	Posterior s.d. Conditional on Inclusion (4)	Posterior Unconditional Mean (3)′	Posterior Unconditional s.d. (4)′	Sign Certainty Probability (5)
1	East asian	0.99	0.823	0.021805	0.006118	0.017935	0.010010	0.999
2	Primary schooling 1960	0.96	0.796	0.026852	0.007977	0.021386	0.012945	0.999
3	Investment price	0.99	0.774	−0.000084	0.000025	−0.000065	0.000041	0.999
4	GDP 1960 (log)	0.30	0.685	−0.008538	0.002888	−0.005845	0.004631	0.999
5	Fraction of tropical area (or people)	0.59	0.563	−0.014757	0.004227	−0.008312	0.007977	0.997
6	Population density in coastal areas 1960s	0.85	0.428	0.000009	0.000003	0.000004	0.000005	0.996
7	Malaria prevalence in 1960s	0.84	0.252	−0.015702	0.006177	−0.003956	0.007489	0.990
8	Life expectancy in 1960	0.79	0.209	0.000808	0.000354	0.000168	0.000366	0.986
9	Fraction Confucian	0.97	0.206	0.054429	0.022426	0.011239	0.024275	0.988
10	African dummy	0.90	0.154	−0.014706	0.006866	−0.002260	0.005948	0.980
11	Latin American dummy	0.30	0.149	−0.012758	0.005834	−0.001905	0.005075	0.969
12	Fraction GDP in mining	0.07	0.124	0.038823	0.019255	0.004818	0.014487	0.978
13	Spanish colony	0.24	0.123	−0.010720	0.005041	−0.001320	0.003942	0.972
14	Years open	0.98	0.119	0.012209	0.006287	0.001457	0.004514	0.977
15	Fraction Muslim	0.11	0.114	0.012629	0.006257	0.001446	0.004545	0.973
16	Fraction Buddhist	0.90	0.108	0.021667	0.010722	0.002348	0.007604	0.974
17	Ethnolinguistic fractionalization	0.52	0.105	−0.011281	0.005835	−0.001181	0.003936	0.974
18	Government consumption share 1960s	0.77	0.104	−0.044171	0.025383	−0.004586	0.015761	0.975
19	Population density 1960	0.01	0.086	0.000013	0.000007	0.000001	0.000004	0.965
20	Real exchange rate distortions	0.92	0.082	−0.000079	0.000043	−0.000006	0.000025	0.966
21	Fraction speaking foreign language	0.43	0.080	0.007006	0.003960	0.000559	0.002204	0.962

Que se ha hecho hasta el momento?

- En estadística es una metodología bastante común para predicciones de corrientes marinas y de epidemias.

- En cuanto al crecimiento Ley, Fernandez y Steel (JAE 2002) y Sala-i-Martin et al. (AER 2004)

En la literatura Macro y de Series de Tiempo:

- 2003:
  - Para predecir al tipo de cambio, **Jonathan Wright** para el FED System: “Depending on the Currency-Horizon pair, the BMA forecastsdo quite a bit better than the Random Walk Benchmark while they never do much worse”
  - Para predecir la Inflación, **Jonathan Wright** para el FED System “In this paper, I consider BMA for pseudo-of-sample prediction of US inflation, and find that it gives more accurate forecasts than simple equal weighted averaging across subsamples and inflation measures…both methods substantially outperform a naive time series benchmark”
  - Sobre prediccion de Inflacion y PBI, **Koop** y **Potter** para el NY FED “ Our analysis indicates that models containing add factors do outperform autoregressive models in forecasting both GDP and inflation, but narrowly and in short horizons”
- 2004:
  - **Diebold** para la inflación “Exploring BMA, we find that combining forecasts from the 10th highest ranked indicator models (in terms of posterior probabilities) yields robust forecasts with, in general, smaller root mean squared errors than the included individual models display.
- **2005:**
  - Para predecir inflacion y Actividad, **Stock y Watson** para el World Congress del Econometric Society “…FAAR models and the principal component of BMA models with small values of g…put wheight on only few of the principal component, resulting in less sample error and more accurate forecasts.”

Procedimientos usuales:

- Para llevar a cabo las estimaciones, es indispensable utilizar tener una función para los priors de los coeficientes:

$$pr(M_i / D) = \frac{pr(D \setminus M_i)\, pr(M_i)}{\sum_{k=1}^{2^K} pr(D \setminus M_k)\, pr(M_k)}$$

- El uso de los G-Prior Zellner (1971):

$$B \sim (0, \phi\sigma^2 (X'X)^{-1})$$

  - Como quedan las estimaciones: Model Posterior ala Ley, Fernandez & Steel

$$pr(M_{ij}, / D) = \frac{pr(M_{ij}) T^{-n(M_{ij})/2} (1+\phi)^{-n(M_{ij})/2} S_{ij}^{-T/2}}{\sum_{k=1}^{2^K} pr(M_{kj}) T^{-n(M_{ij})/2} (1+\phi)^{-n(M_{kj})/2} S_{kj}^{-T/2}}$$

  - Posterior ala Sala-i-Martin:

$$pr(M_i, / D) = \frac{pr(M_i) T^{-n(M_i)/2} SSE^{-T/2}}{\sum_{k=1}^{2^K} pr(M_k) T^{(n(M_i)-n(M_j))/2} SSE^{-T/2}}$$

Prior para los modelos

- Rho es un hiperparametro [0,1]

$$P(M_k) = \rho^{n(M_k)}(1-\rho)^{n(\mathrm{M})-n(M_k)}$$

- Si tenemos rezagos tenemos Rho1 y Rho2

$$P(M_k) = \rho_1^{n(M_k)}(1-\rho_1)^{n(\mathrm{M})-n(M_k)} * \rho_2^{nL(M_k)}(1-\rho_2)^{nL(\mathrm{M})-nL(M_k)}$$

El problema de la envergadura y el algoritmo MC3

- Markov Chain Montecarlo Carlo Modelo Composition

$$Min\left\{1, \frac{p(M'\backslash D)}{p(M\backslash D)}\right\}$$

Que queremos hacer nosotros?

- Llevar esto al contexto de un VAR con el fin practico de:
  - Darnos una idea clara de que variables debemos incluir en nuestros ejercicios VAR y cuales debemos dejar fuera según sea el caso:
  - Generar Inferencia Robusta a la especificación: Impulsos Respuesta independientes del modelo:
  - Mejorar, si tenemos suerte la predicción.

Un ejercicio de Prueba

- Data Peruana desde 1997

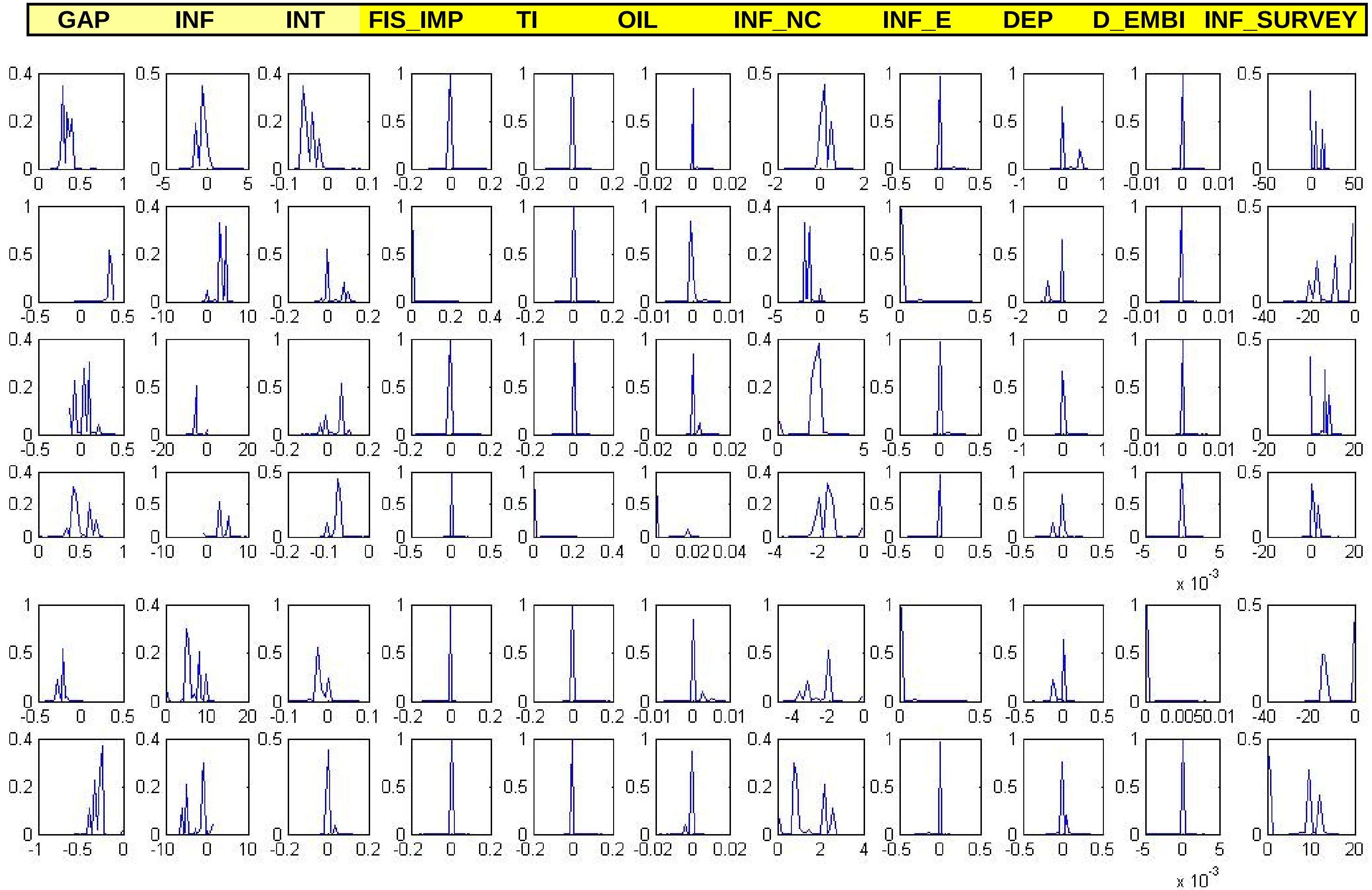
GAP	INF	INT	FIS_IMP	TI	OIL	INF_NC	INF_E	DEP	D_EMBI	INF_SURVEY

Un ejercicio de Prueba:

- Priors Utilizadas:
  - Modoel Prior= Flat Prior
  - Coef Prior = G-Prior (a la Sala-i-Martin)

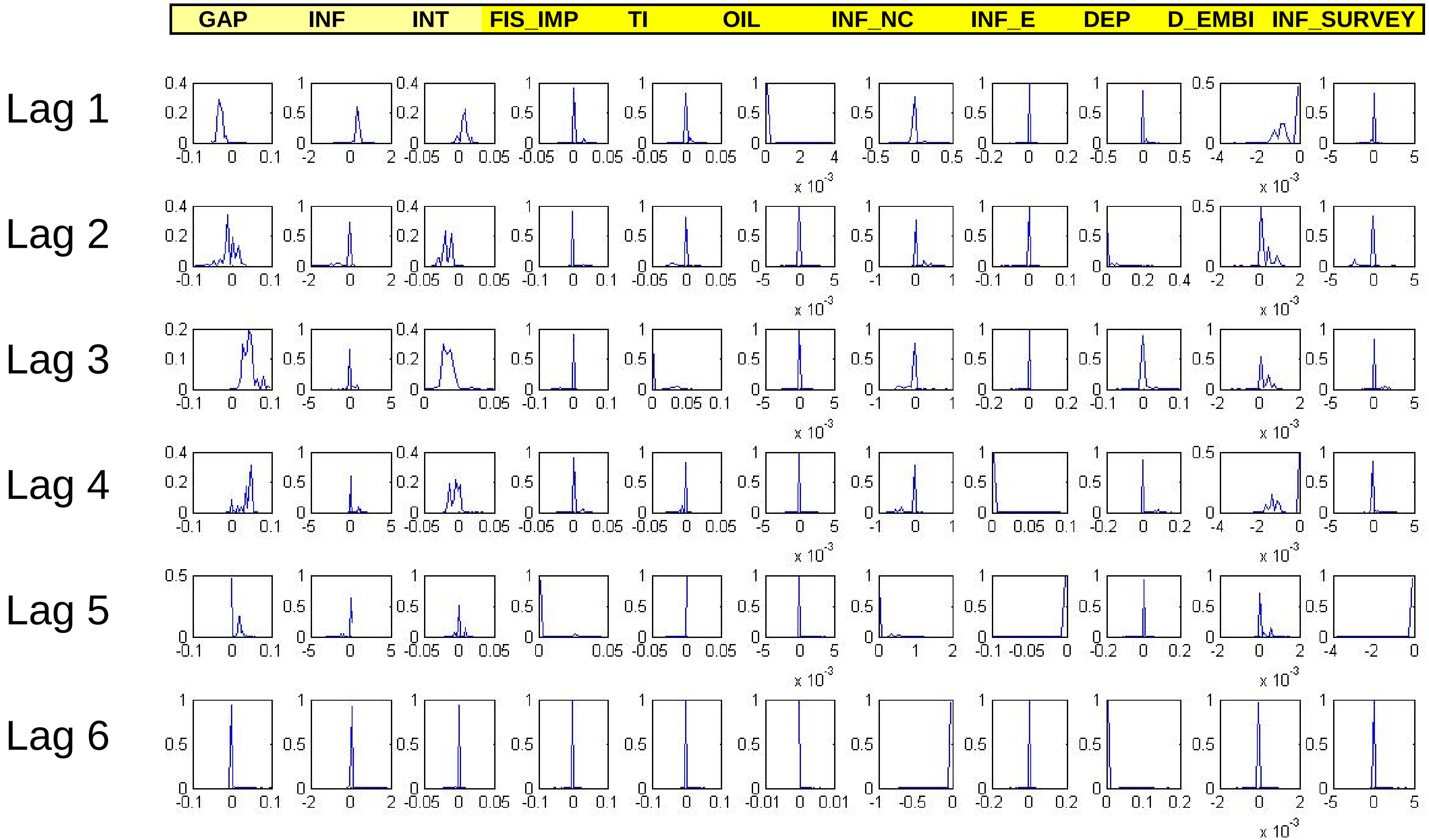
Resultados: Muestra de Resultados

- Gráfico para la brecha del Producto:



Resultados: Muestra de Resultados

- Gráfico para la brecha del Producto:



Que nos falta hacer exactamente?

- Complicaciones:

$$pr(M_i,/D) \sim= \frac{pr(M_i)T^{-n(M_j)/2}(1+\phi)^{-n(M_j)/2}\left|\Xi\right|^{-T/2}}{\sum_{k=1}^{2^K} pr(M_{kj})T^{-n(M_{ij})/2}(1+\phi)^{-n(M_{kj})/2}\left|\Xi\right|^{-T/2}}$$

- Minnesota Prior:

$$B \sim (I, \nu * \Sigma)$$

Algunas preguntas para el auditorio:

- Tomar promedios ecuación por ecuación o promedios tomando a todo el modelo en conjunto?

- Utilizar los G-Prior o tratar de implementar esto para un Minnesota Prior?