

Métodos bayesianos para modelos autorregresivos y de corrección de equilibrio

TRABAJO FIN DE GRADO



Grado en Estadística y Empresa

GONZALO ESPAÑA-HEREDIA
LLANZA

TUTOR: JUAN MIGUEL MARÍN
DIAZARAQUE

Getafe, Junio 2022

Índice general

Prólogo	III
Resumen	V
Abstract	VI
Índice de Figuras	X
Índice de Tablas	XII
1. Introducción	1
1.1. Objetivos generales y específicos	3
2. Metodología y análisis de datos	5
2.1. Metodología	5
2.1.1. Modelo VAR - Frecuentista	5
2.1.2. Modelo VAR - Bayesiano (BVAR)	7
2.1.3. Modelo VECM - Frecuentista	8
2.1.4. Modelo VECM - Bayesiano (BVECM)	11
2.1.5. Análisis de la capacidad predictiva de los modelos	14
2.2. Análisis de datos	16
2.2.1. Simulación de los datos	16
2.2.2. Datos reales	36
3. Resultados	49
3.1. Estimación de los modelos VAR y BVAR	49
3.2. Estimación de los modelos VECM y BVECM	70
3.3. Análisis de la capacidad predictiva	80
4. Conclusiones	83
4.1. Motivación de la inferencia bayesiana	83
4.2. Problemas en la inferencia bayesiana	83
4.3. Estimación de los coeficientes	84
4.4. Capacidad predictiva	85

A. Apéndice:	87
A.1. Primera sección	87
B. Apéndice: Título del Apéndice	89
B.1. Primera sección	89
Bibliografía	91

Prólogo

Este trabajo de fin de grado es el punto y final a una etapa que ha durado 4 años y de la que estoy tremadamente orgulloso, me llevo grandes recuerdos y multitud de conocimientos y lecciones aprendidas.

Aquí termina mi etapa de *undergraduate*, y, aunque no voy a continuar con estudios superiores por el momento, me voy de la universidad con una curiosidad y hambre por seguir aprendiendo enorme. Esto, en parte, se ve reflejado en este trabajo de investigación que no es más que el fruto de mi interés por investigar sobre la intersección de dos campos de estudio a los que se me ha introducido este último año de carrera y que me han fascinado.

Este trabajo es una parte ínfima de lo que representa mi paso por el Grado en Estadística y Empresa en la Universidad Carlos III, pero refleja algunas de las lecciones aprendidas durante mi formación:

- Rigor. Ha sido un tópico durante los 4 años y nos obliga a no buscar la respuesta por tener una respuesta, si no que se llegue a ella de manera fundada y objetiva.
- El cuidado por el detalle es otra parte importante de la forma en la que nos han enseñado a trabajar, tu trabajo refleja parte de cómo eres.
- La profundidad de conocimiento se ve reflejada en las extensas explicaciones sobre la teoría que hay detrás de cada modelo que se ha implementado. Una frase que recoge muy bien este concepto es “No sirve con saber hacer algo, hay que saber por qué lo haces”.
- Por último, pero el más importante, la honestidad científica. En estos 4 años nos han enseñado que no siempre se consiguen resultados concluyentes, que “no model is better than a bad model” y que la forma en la que se lleva a cabo una investigación o análisis de datos puede influir en los resultados, por esto es importante ser objetivo y honesto a la hora de sacar conclusiones. Este aspecto se ve reflejado en el capítulo 4 del trabajo.

Todo esto, en parte, ha hecho que el trabajo sea más extenso de lo que me habría gustado.

Con todo esto lo que quiero expresar es que aunque esta sea la última huella que dejo en mi paso por la universidad, no habría sido posible llegar hasta este lugar sin todas las personas que me he cruzado durante estos cuatro años. A los profesores, compañeros de clase, mi familia y amigos, gracias por todo.

Por último, quiero agradecer a Juan Miguel Marín Diazaraque su labor como tutor de este trabajo de fin de grado, en muchos momentos he estado perdido y no sabía muy bien por donde avanzar pero gracias a su ayuda y consejo creo que he tomado buenas decisiones.

Resumen

Este trabajo de fin de grado es un trabajo de investigación sobre métodos bayesianos para modelos autorregresivos y de corrección de equilibrio. El objetivo principal del trabajo es entender las motivaciones para aplicar métodos bayesianos para la estimación de este tipo de modelos, sus ventajas y desventajas frente a los métodos frecuentistas tradicionales y analizar las diferencias en la estimación de los parámetros y la capacidad predictiva de cada uno.

El principal escenario en el que se eligiría un método bayesiano de estimación frente a uno frecuentista sería el caso en el que hay numerosos predictores con varios retardos a incluir en el modelo lo que genera una sobreparametrización que resulta en estimaciones inestables y desviaciones estándar muy altas mediante el método frecuentista. Otra ventaja es que se pueden obtener distribuciones de probabilidad para cada parámetros estimado independientemente del tamaño muestral que se tenga.

Aún así, los métodos bayesiano tienen sus inconvenientes, especialmente la elección de la información a priori. Si no se elige bien esta información y se tienen pocos datos las estimaciones pueden ser muy malas, y, además, según la a priori que se elija se puede llegar a obtener una a posteriori impropia. Estos problemas, además del problema de estimación de la relación de largo plazo en los modelos de corrección de equilibrio se abordan en el trabajo, y se proponen soluciones basadas en la literatura que generan estimaciones más robustas y computacionalmente eficientes.

Para comparar las estimaciones y la capacidad predictiva de cada método, se analizan 6 datasets distintos. 3 de ellos contienen variables no cointegradas y los otros 3 variables cointegradas. En cada grupo, hay un dataset real y dos dataset simulados siendo uno de 1000 observaciones y otro de 100. De esta forma, con los datos simulados se pueden comparar las estimaciones de cada método con los valores reales de los parámetros. Los dataset reales se usan para comparar estimaciones pero, además, se usan para comparar la capacidad predictiva de los modelos mediante un algoritmo walk-forward validation modificado para que sea computacionalmente asequible.

En cuanto a la inferencia, las estimaciones de forma frecuentista y bayesiana casi idénticas para datasets con muchas observaciones (debido al peso que tienen en la a posteriori) y en los data sets simulados con pocas observaciones también son muy similares aunque hay una ligera mejora en el método bayesiano para modelos autorregresivos. De todas formas, esta ligera mejora no llega a ser concluyente y significativa ya que se produce para unos pocos parámetros y en un modelo.

La capacidad predictiva tampoco muestra diferencias entre los métodos, pero en gran parte está condicionada por la moficiación realizada al algoritmo walk-forward validation para que fuese computacionalmente asequible.

Abstract

This final thesis is a research work on Bayesian methods for autoregressive and equilibrium correction models. The main objective of the work is to understand the motivations for applying Bayesian methods for the estimation of this type of models, their advantages and disadvantages versus traditional frequentist methods and to analyze the differences in the estimation of the parameters and the predictive capacity of each one.

The main scenario in which a Bayesian estimation method would be chosen over a frequentist method would be the case in which there are numerous predictors with several lags to be included in the model, which generates an over-parameterization resulting in unstable estimates and very high standard deviations using the frequentist method. Another advantage is that probability distributions can be obtained for each estimated parameter regardless of the sample size.

Even so, Bayesian methods have their drawbacks, especially the choice of a priori information. If this information is not well chosen and there is little data, the estimates can be very poor, and, in addition, depending on the a priori chosen, an improper a posteriori can be obtained. These problems, in addition to the problem of estimating the long-run relationship in equilibrium correction models are addressed in the paper, and literature-based solutions that generate more robust and computationally efficient estimates are proposed.

To compare the estimates and predictive ability of each method, 6 different datasets are analyzed. Three of them contain non-cointegrated variables and the other three contain cointegrated variables. In each group, there is a real dataset and two simulated datasets, one of 1000 observations and the other of 100. In this way, the simulated data can be used to compare the estimates of each method with the real values of the parameters. The real datasets are used to compare estimates but, in addition, they are used to compare the predictive capacity of the models by means of a walk-forward validation algorithm modified to be computationally affordable.

In terms of inference, the almost identical frequentist and Bayesian estimates for datasets with many observations (due to the weight they have in the a posteriori) and in simulated data sets with few observations are also very similar although there is a slight improvement in the Bayesian method for autoregressive models. However, this slight improvement is not conclusive and significant since it occurs for only a few parameters and in one model.

The predictive capacity does not show differences between the methods either, but it is largely conditioned by the modification made to the walk-forward validation algorithm to make it computationally affordable.

Índice de figuras

2.1. Histograma de los valores del estadístico t para la relación lineal entre dos paseos aleatorios independientes	10
2.2. Datos simulados para VAR.	17
2.3. Variable X1 simulada para VAR.	17
2.4. Autocorrelación simple de X1 para VAR.	18
2.5. Serie temporal X2 datos simulados.	19
2.6. Autocorrelación simple de X2 simulada para VAR.	19
2.7. Datos simulados para VAR.	20
2.8. Variable X1 simulada para VAR.	20
2.9. Autocorrelación simple de X1 para VAR.	21
2.10. Serie temporal X2 datos simulados.	22
2.11. Autocorrelación simple de X2 simulada para VAR.	22
2.12. Datos simulados para VECM.	23
2.13. Variable X1 para VECM.	24
2.14. Autocorrelación simple de X1 para VECM.	24
2.15. Variable X1 en diferencias para VECM.	25
2.16. Autocorrelación simple de X1 en diferencias para VECM.	25
2.17. Variable X2 para VECM.	26
2.18. Autocorrelación simple de X2 para VECM.	26
2.19. Variable X2 en diferencias para VECM.	27
2.20. Autocorrelación simple de X2 en diferencias para VECM.	27
2.21. Relación lineal entre X1 y X2 simuladas para VECM.	28
2.22. Residuos de la relación lineal entre variables simuladas para VECM.	29
2.23. ACF de los residuos de la relación lineal entre variables simuladas para VECM.	29
2.24. Datos simulados para VECM.	30
2.25. Variable X1 para VECM.	30

2.26. Autocorrelación simple de X1 para VECM.	31
2.27. Variable X1 en diferencias para VECM.	31
2.28. Autocorrelación simple de X1 en diferencias para VECM.	32
2.29. Variable X2 para VECM.	32
2.30. Autocorrelación simple de X2 para VECM.	33
2.31. Variable X2 en diferencias para VECM.	33
2.32. Autocorrelación simple de X2 en diferencias para VECM.	34
2.33. Relación lineal entre X1 y X2 simuladas para VECM 100 observaciones.	34
2.34. Residuos de la relación lineal entre variables simuladas para VECM. . .	35
2.35. ACF de los residuos de la relación lineal entre variables simuladas para VECM.	35
2.36. Datos macro USA 1959-2007.	36
2.37. Inflación USA 1959-2007.	37
2.38. ACF inflación.	38
2.39. Desempleo USA 1959-2007.	38
2.40. ACF desempleo	39
2.41. Primera diferencia de la Inflación en USA 1959-2007.	39
2.42. ACF de la primera diferencia Inflación.	40
2.43. Primera diferencia del desempleo USA 1959-2007	40
2.44. ACF de la primera diferencia desempleo	41
2.45. Relación lineal entre desempleo e inflación.	41
2.46. Residuos de la relación lineal entre inflación y desempleo.	42
2.47. ACF de los residuos de la relación lineal entre inflación y desempleo. . .	42
2.48. Serie de LIBOR y OVERNIGHT 1999-2007.	43
2.49. ACF de LIBOR.	44
2.50. ACF de OVERNIGHT	44
2.51. Serie de LIBOR con primera diferencia.	45
2.52. ACF de LIBOR con primera diferencia.	45
2.53. Serie de OVERNIGHT con primera diferencia.	46
2.54. ACF de OVERNIGHT con primera diferencia.	46
2.55. Relación lineal entre LIBOR y OVERNIGHT.	47
2.56. Residuos de la relación lineal entre LIBOR y OVERNIGHT.	47
2.57. ACF de residuos de la relación lineal entre LIBOR y OVERNIGHT. . .	48
3.1. Residuos modelo VAR para X1 y X2.	51

3.2. Convergencia de los Hiperparámetros BVAR.	53
3.3. Convergencia de los parámetros BVAR para X1.	54
3.4. Densidad de los parámetros BVAR para X1.	54
3.5. Convergencia de los parámetros BVAR para X2.	55
3.6. Densidad de los parámetros BVAR para X2.	55
3.7. Convergencia matriz covarianza modelo BVAR.	56
3.8. Densidad de los parámetros de covarianza muestra BVAR.	56
3.9. Residuos modelo VAR para X1 y X2.	58
3.10. Convergencia de los Hiperparámetros BVAR.	59
3.11. Convergencia de los parámetros BVAR para X1.	60
3.12. Densidad de los parámetros BVAR para X1.	60
3.13. Convergencia de los parámetros BVAR para X2.	61
3.14. Densidad de los parámetros BVAR para X2.	61
3.15. Convergencia matriz covarianza modelo BVAR.	62
3.16. Densidad de los parámetros de covarianza muestra BVAR.	62
3.17. Residuos modelo VAR para Inflación y Desempleo.	64
3.18. Draws y distribuciones a posteriori de los hiperparámetros para los datos US.	65
3.19. Convergencia de los parámetros BVAR para Inflación	66
3.20. Densidad de los parámetros BVAR para Inflación	66
3.21. Convergencia de los parámetros BVAR para Desempleo	67
3.22. Densidad de los parámetros BVAR para Desempleo	68
3.23. Convergencia matriz covarianza modelo BVAR datos macro EE.UU.	69
3.24. Densidad de los parámetros de covarianza muestral BVAR datos macro EE.UU.	69
3.25. Residuos X1 modelo VECM para datos simulados de 1000 observaciones.	71
3.26. Residuos X2 modelo VECM para datos simulados de 1000 observaciones.	71
3.27. Muestreo de beta modelo BVECM para datos simulados de 1000 observaciones.	73
3.28. Residuos X1 modelo VECM para datos simulados de 100 observaciones.	74
3.29. Residuos X2 modelo VECM para datos simulados de 100 observaciones.	75
3.30. Muestreo de beta modelo BVECM para datos simulados de 100 observaciones.	76
3.31. Residuos de LIBOR modelo VECM.	78
3.32. Residuos OVERNIGHT modelo VECM.	78

3.33. Muestreo de beta modelo BVECM para los datos de LIBOR y OVER-NIGHT.	80
---	----

Índice de tablas

3.1. Estimación modelo VAR(3) para X1 con 1000 observaciones.	50
3.2. Estimación modelo VAR(3) para X2 con 1000 observaciones.	50
3.3. Estimación matriz var-cov modelo VAR(3) datos simulados 1000 observaciones.	50
3.4. Estimación modelo BVAR(3) para datos simulados con 1000 observaciones.	52
3.5. Estimación matriz var-cov para modelo BVAR(3) para datos simulados con 1000 observaciones.	52
3.6. Estimación modelo VAR(3) para X1 con 100 observaciones.	57
3.7. Estimación modelo VAR(3) para X2 con 100 observaciones.	57
3.8. Estimación matriz var-cov modelo VAR(3) con 100 observaciones.	57
3.9. Estimación modelo BVAR(3) para datos simulados con 100 observaciones.	58
3.10. Estimación matriz var-cov para modelo BVAR(3) para datos simulados con 100 observaciones.	59
3.11. Estimación modelo VAR(2) para Inflación datos macro EE.UU. 1959-2007.	63
3.12. Estimación modelo VAR(2) para Desempleo datos macro EE.UU. 1959-2007.	63
3.13. Estimación matriz var-cov modelo VAR(2) para datos macro EE.UU. 1959-2007..	63
3.14. Estimación modelo BVAR(2) para datos macro EE.UU. 1959-2007.	64
3.15. Estimación matriz var-cov modelo BVAR(2) para datos macro EE.UU. 1959-2007.	64
3.16. Estimación relación largo plazo modelo VECM(2) para datos simulados con 1000 observaciones.	70
3.17. Estimación modelo VECM(2) para datos simulados con 1000 observaciones.	70
3.18. Estimación matriz var-cov modelo VECM(2) para datos simulados con 1000 observaciones.	70
3.19. Estimación relación de largo plazo modelo BVECM(2) para datos simulados con 1000 observaciones.	72

3.20. Estimación modelo BVECM(2) para datos simulados con 1000 observaciones.	72
3.21. Estimación matriz var-cov modelo BVECM(2) para datos simulados con 1000 observaciones.	72
3.22. Estimación relación largo plazo modelo VECM(2) para datos simulados con 100 observaciones.	73
3.23. Estimación modelo VECM(2) para datos simulados con 100 observaciones.	73
3.24. Estimación matriz var-cov modelo VECM(2) para datos simulados con 100 observaciones.	74
3.25. Estimación relación de largo plazo modelo BVECM(2) para datos simulados con 100 observaciones.	75
3.26. Estimación modelo BVECM(2) para datos simulados con 100 observaciones.	75
3.27. Estimación matriz var-cov modelo BVECM(2) para datos simulados con 100 observaciones.	75
3.28. Estimación relación largo plazo modelo VECM(5) para datos de LIBOR y OVERNIGHT.	76
3.29. Estimación del modelo VECM(5) para datos de LIBOR y OVERNIGHT.	77
3.30. Estimación matriz var-cov modelo VECM(5) para datos de LIBOR y OVERNIGHT.	77
3.31. Estimación relación de largo plazo modelo BVECM(5) para datos de LIBOR y OVERNIGHT.	79
3.32. Estimación modelo BVECM(5) para datos de LIBOR y OVERNIGHT	79
3.33. Estimación matriz var-cov modelo BVECM(5) para datos de LIBOR y OVERNIGHT.	79
3.34. Media de MSRE en 30 iteraciones con datos simulados.	81
3.35. Media de MSRE en 30 iteraciones para datos macroeconómicos USA 1959-2007.	82
3.36. Media de MSRE en 30 iteraciones para datos simulados.	82
3.37. Media de MSRE en 30 iteraciones para datos de LIBOR y OVERNIGHT.	82

Capítulo 1

Introducción

Este trabajo de fin de grado trata de analizar los beneficios que puede aportar estimar modelos vectoriales autorregresivos y de corrección de equilibrio de manera bayesiana frente a métodos frecuentistas tradicionales.

Los modelos vectoriales autorregresivos (VAR) fueron introducidos por Christopher Sims (1980) y son una extensión de los modelos autorregresivos univariantes hacia modelos dinámicos multivariantes. Son uno de los modelos más populares, flexibles y fáciles de implementar para el análisis de series temporales multivariantes. Los modelos VAR han destacado principalmente en economía y finanzas para entender la relación entre las diferentes variables, y, también, para hacer predicciones (Stock y Watson 2001). Estos modelos son muy flexibles porque se pueden construir en función de otras series temporales. Esto significa que no es necesario tener un modelo estructural adecuado para una serie de interés, sino que se crea un modelo en el que una serie es explicada por el comportamiento de otras series y su propio pasado. Esto hace que lo más importante a la hora de crear modelos de este tipo es elegir bien las series que se van a relacionar y los retardos de cada una.

A parte de economía y finanzas, los modelos VAR son muy utilizados en el ámbito político para analizar las causas de ciertos escenarios macroeconómicos, así como para medir los impactos de las decisiones políticas. Estas son algunas de las aplicaciones más comunes de los modelos VAR, pero en la realidad, se aplican en múltiples campos con fines muy diferentes. Algunos ejemplos son:

- Medicina: Para entender cómo están relacionados los signos vitales de pacientes con problemas cardiorespiratorios (Bose, Eliezer; Hravnak, Marilyn; Sereika, Susan M 2017).
- Epidemiología: Para entender cómo los riesgos de infección del COVID-19 interactúan entre diferentes grupos de edad (Yu, Xinhua 2020).
- Biología: Cómo modelar las redes de expresiones genéticas (Fujita, A., Sato, J.R., Garay-Malpartida, H.M. et al. 2007).

En definitiva, estos modelos son muy comunes y se pueden aplicar a múltiples tipos de series temporales donde el objetivo sea explicar el comportamiento de una a través de su pasado y del de otras variables.

En cuanto a los modelos de corrección de equilibrio, estos modelos son una extensión de los modelos autorregresivos. La diferencia es que estos modelos recogen la relación de largo plazo entre dos variables además de las relaciones contemporáneas. Primero, Sargan (1964) desarrolló la metodología para los modelos uniecuacionales, y, más adelante, casi a la vez que los VAR, se desarrolló la metodología para los VECM.

Los modelos VECM suelen ser muy útiles ya que es muy común que al menos un par de series estén relacionadas en el largo plazo (Koop, Strachan, van Dijk y Villani 2005). Si no se recoge esta relación, se pierde información.

Los ejemplos en macroeconomía y finanzas abundan, como el consumo y la renta permanente, los precios y los dividendos en el mercado de valores y las series de tipos de interés a corto y largo plazo (Koop, Strachan, van Dijk y Villani 2005). La globalización ha causado que muchas series macroeconómicas de diferentes países estén relacionadas y se puedan aplicar estos modelos. También, como los índices macroeconómicos tienen gran influencia en los mercados bursátiles, hay series macroeconómicas relacionadas en el largo plazo con series de mercados financieros. Algunas aplicaciones más de actualidad son:

- Predicción del retorno de Bitcoin (Liang 2021)
- Análisis de la relación entre el índice de precios percibidos por el agricultor (PIR), Índice de precios pagados por los agricultores (PIP) y la relación de intercambio de los agricultores (FTT) para tratar de conocer el comportamiento de (FTT) si hay un shock en las variables PIR y PIP (Mustofa Usman; Dhia Fadhilah Fatin; M. Yusuf S. Barusman; Faiz A. M. Elfaki; Widiarti 2017).

Tanto los modelos VAR como los modelos VECM se pueden estimar también mediante métodos bayesianos. En general se usan métodos de simulación basados en cadenas de Markov por medio de la metodología MCMC. Desde finales de los 90 hasta mitad de los 2000, se empezaron a aplicar métodos bayesianos a los modelos autorregresivos y de corrección de equilibrio. Trabajos como los de Bauwens y Lubrano (1996), Bauwens y Giot (1998), Strachan (2003), Strachan y Inder (2004) son algunos de los ejemplos.

Los modelos VAR y VECM pueden llegar a tener muchos parámetros, esto hace que los errores estándar de las estimaciones sean grandes. Uno de los beneficios de estimar estos modelos con técnicas bayesianas es que se pueden mejorar las estimaciones si se incluye información a priori. Otra ventaja, es que se obtienen distribuciones de probabilidad para cada uno de los parámetros, independientemente del tamaño de la muestra. Estas propiedades de los métodos bayesianos son muy buenas pero depende de la información a priori. Se han propuesto varias distribuciones a priori a lo largo de los años que serán discutidas más adelante.

En este trabajo de fin de grado, lo que se pretende es entender cuáles son las diferencias a la hora de estimar estos modelos de manera frequentista y de manera bayesiana, así como analizar las ventajas e inconvenientes de cada método para la inferencia y predicción. Para hacer esto, se van a aplicar los diferentes modelos primero a unos datos simulados para comparar las estimaciones de los parámetros con los *reales*, y después a dos conjuntos de series temporales reales diferentes. Uno de los conjuntos contiene la tasa de inflación para Estados Unidos y el desempleo para cada trimestre desde 1959 hasta 2007. Los datos están dentro del paquete **bvar tools**. El otro conjunto de datos con el que se va a trabajar recoge

los tipos de cambio (swaps) a 9 meses, denominados OVERNIGHT, y el LIBOR a 12 en la eurozona. En la simulación de datos se va a simular un data set grande y otro pequeño para cada tipo de modelo. Se va a hacer esto para comparar las estimaciones cuando hay pocos datos (los modelos bayesianos suelen tener mejor resultado) y cuando hay muchos datos (los datos suelen tener más peso que las a priori por lo que las estimaciones deberían ser bastante similares en teoría).

Para la estimación de los modelos con métodos bayesianos se van a utilizar las librerías `bvartools` y `BVAR`. Para la estimación de los modelos frecuentistas se usarán las librerías `vars`, `tsDyn`, `tseries` y `urca` principalmente.

1.1. Objetivos generales y específicos

Este trabajo pretende exponer las ventajas e inconvenientes de estimar modelos vectoriales autorregresivos y de corrección de equilibrio de manera bayesiana frente a un enfoque frecuentista.

Los objetivos generales son:

- Explorar y entender las diferencias entre la estimación bayesiana y frecuentista para modelos autorregresivos y de corrección de equilibrio.
- Comprender las limitaciones que tienen los enfoques bayesianos en los modelos de corrección de equilibrio.
- Comparar las estimaciones de los parámetros entre el método frecuentista y bayesiano.
- Comparar la capacidad predictiva de estos modelos estimados por métodos bayesianos frente a la forma frecuentista

En cuanto a los objetivos específicos:

- Entender las motivaciones por las que una estimación bayesiana de estos modelos puede ser beneficiosa.
- Exponer los problemas que pueden surgir a la hora de hacer la estimación de forma bayesiana.
- Determinar unas distribuciones a priori adecuadas para los modelos.
- Comparar la inferencia de los parámetros de los modelos BVECM frente a los VECM y de los VAR frente a los BVAR.
- Analizar la capacidad predictiva de los diferentes modelos.

Capítulo 2

Metodología y análisis de datos

2.1. Metodología

El objetivo principal de este trabajo es entender las diferencias entre estimar los modelos VAR y VECM de forma bayesiana y de forma frecuentista, tanto para la inferencia como para la predicción de nuevas observaciones. Para esto vamos a utilizar dos conjuntos de datos simulados grandes, dos conjuntos de datos simulados pequeños y dos conjuntos de datos reales. En cada par de conjunto de datos habrá un dataset con series cointegradas y otro dataset con series no cointegradas. De esta forma podemos estimar los modelos autorregresivos y de corrección de equilibrio en datos simulados para comparar los parámetros estimados con los reales, y, también, sobre unos datos reales sin saber la verdadera relación entre las variables ni el valor de los coeficientes para comparar las estimaciones entre los modelos así como su capacidad predictiva.

Para poder comparar el método bayesiano frente al frecuentista para modelos VECM, primero es necesario que las series sean cointegradas ya que si no, carece de sentido estimar este tipo de modelo. Por esto, lo primero que se va a hacer es un análisis de las series temporales para determinar su orden de integración y si están cointegradas.

En cuanto al análisis de las series y la identificación del orden autorregresivo de los modelos, para los conjuntos de datos simulados sabemos la verdadera relación entre las variables por lo que no habrá que probar diferentes modelos. Se estimarán según cada método especificando la estructura según la estructura real para ver si hay diferencias en la estimación. Para las series de datos reales, se mostrará solamente los modelos definitivos que serán elegidos según la metodología *Box-Jenkins*.

2.1.1. Modelo VAR - Frecuentista

Los modelos VAR son una extensión multivariante dinámica de los modelos autorregresivos univariantes. Es preferible usar los modelos VAR para relacionar diferentes series, puesto que ayuda a explicar mejor el comportamiento de cada una incorporando información de la otra. Además, como también se van a aplicar los modelos de corrección de equilibrio y estos utilizan al menos 2 series temporales, al usar los modelos VAR podemos compararlos frente a los modelos de corrección de equilibrio ya que son una extensión de los modelos autorregresivos multivariantes.

Los modelos VAR tienen n ecuaciones, en este caso $n = 2$, y cada ecuación incorpora la información dinámica de cada una de las variables, por esto, se obtienen matrices de orden $n \times n$ para cada retardo. La formulación para un modelo $VAR(1)$ con $n = 2$ es la siguiente:

$$\begin{aligned} X_{1t} &= \phi_{11}X_{1t-1} + \phi_{12}X_{2t-1} + a_{1t} \\ X_{2t} &= \phi_{21}X_{1t-1} + \phi_{22}X_{2t-1} + a_{2t} \end{aligned} \tag{2.1}$$

Ambas variables X_{1t}, X_{2t} deben ser estacionarias, tanto X_{1t} como X_{2t} dependen de su pasado y del pasado de la otra, a_{1t} es la innovación de X_{1t} en el momento t y a_{2t} es la innovación de X_{2t} en el momento t . Tanto a_{1t} como a_{2t} son ruido blanco y pueden estar relacionadas entre sí, y en este caso se dice que hay relación *contemporánea*. La relación entre las innovaciones está recogida por $Cov(a_{1t}, a_{2t}) = \sigma_{12}$.

El caso del modelo expresado en 2.1, refleja un caso donde hay *realimentación*, esto es que X_{1t} depende de X_{2t} y también X_{2t} depende de X_{1t} . La realimentación está controlada por los parámetros ϕ_{12} y ϕ_{21} .

Si $\sigma_{12} = 0$ no hay relación contemporánea, y, se dice que hay *exogeneidad débil*. Si $\sigma_{12} = 0$ y $\phi_{12} = 0$ o $\phi_{21} = 0$, se dice que hay exogeneidad fuerte.

El modelo expresado en 2.1, se puede expresar también de la siguiente forma:

$$\mathbf{X}_t = \Phi_1 \mathbf{X}_{t-1} + \mathbf{a}_t \tag{2.2}$$

Donde \mathbf{X}_t es el vector columna de variables en t , Φ_1 es una matriz $n \times n$ con los parámetros autorregresivos, \mathbf{X}_{t-1} es el vector columna de variables en $t - 1$, y \mathbf{a}_t es el vector columna de innovaciones.

El caso anterior en 2.2, puede incorporar un número mayor de retardos donde recoger la dependencia temporal entre las variables y ellas mismas. Los p retardos que se introduzcan definen el orden del modelo VAR, siendo el caso de 2.2 un $VAR(1)$ y pudiendo ser del tipo $VAR(p)$ según el número de retardos, p , que se incorporen. La formulación de un modelo $VAR(p)$ es la siguiente:

$$\mathbf{X}_t = \Phi_1 \mathbf{X}_{t-1} + \Phi_2 \mathbf{X}_{t-2} + \cdots + \Phi_p \mathbf{X}_{t-p} + \mathbf{a}_t \tag{2.3}$$

Al igual que en 2.2, las matrices Φ_k serán $n \times n$ para el caso de dos variables. Si al modelo anterior en 2.3 se le incluye un modelo de medias móviles, se obtendrá un modelo $VARMA(p, q)$.

Un aspecto muy importante de los modelos VAR es que han de cumplir la *condición de estacionariedad*. Esta, en un modelo multiecuacional, viene recogida por los coeficientes ϕ_{ij} . En paralelo con el caso univariante, para que se cumpla la condición, las raíces del determinante de cada matriz polinomial, Φ_k , han de ser mayores que la unidad.

Por todo esto, para estimar un modelo $VAR(p)$ tendremos que seguir los siguientes pasos:

1. Comprobar que las variables son estacionarias, y, en caso contrario, aplicar diferencias y transformaciones hasta que lo sean. En cuanto a aplicar diferencias, se realizará el test de Dickey-Fuller aumentado junto el correlograma para determinar si es necesario aplicar diferencias o no.
2. Estimar el modelo con las series estacionarias.
3. Comprobar la condición de estacionariedad del modelo *VAR* estimado.
4. Por último, comprobar que los residuos de los modelos son ruido blanco, y, en caso contrario, modificar el modelo con la información del correlograma y análisis de los residuos.

2.1.2. Modelo VAR - Bayesiano (BVAR)

Los modelos autorregresivos vectoriales (VAR) son modelos flexibles de series temporales que pueden capturar relaciones dinámicas complejas entre variables. Sin embargo, sufren de sobreparametrización, lo que conduce a una inferencia inestable y estimaciones imprecisas de nuevas observaciones, sobre todo en el caso de modelos con muchas variables. Una solución a este problema es incluir información a priori con lo que se reduce el modelo no restringido hacia uno *naive* más parsimonioso, y, así, se reduce la incertidumbre de la estimación y se mejoran las predicciones de observaciones fuera de la muestra¹.

Los métodos bayesianos que se han usado para modelos de este tipo se denominan BVAR, difieren en que tratan a los parámetros como variables aleatorias con probabilidades a priori. Una implementación estándar de los modelos VAR bayesianos, es mediante el llamado muestreador de Gibbs, que pertenece a la familia de los métodos MCMC. Se puede encontrar un tratamiento más detallado de la diferencia entre la inferencia frecuentista y la bayesiana de modelos VAR en Kennedy (2008, ch. 14).

A medida que aumenta la relación entre las variables y las observaciones, el papel de las probabilidades a priori es cada vez más importante². La elección de las distribuciones a priori es tema de discusión desde que se empezaron a implementar estos modelos. Se han propuesto varios enfoques a la selección desde las *flat priors*, que se demostró que no funcionaban bien dado que se necesita aportar algún tipo de información (Sims 1980), pasando por hacer uso de la teoría económica (Villani 2009) u otros enfoques como maximizar el ajuste a las predicciones *out-of-sample* (e.g. Litterman 1980) o un ajuste dentro de la muestra para el control del overfitting (Báñbara et al. 2010).

No hay un consenso en que exista un tipo de distribuciones a priori que funcionen mejor para todos los casos, por esto, un método muy flexible y que elimina esta incertidumbre es el de los *modelos bayesianos jerárquicos* (Giannone et al. 2015). Este método conduce a una inferencia robusta, sustentado teóricamente y puede ser implementado de manera eficiente. Estos hiperparámetros tienen sus propias probabilidades llamadas *hyperpriors* que, extendiendo el teorema de Bayes, siguen la forma:

$$p(\boldsymbol{\gamma}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\gamma})p(\boldsymbol{\theta}|\boldsymbol{\gamma})p(\boldsymbol{\gamma}) \quad (2.4)$$

¹Karlsson, Sune (2012). Forecasting with Bayesian Vector Autoregression. Handbook of Economic Forecasting. Vol. 2 B. pp. 791–897.

²Koop, G.; Korobilis, D. (2010). “Bayesian multivariate time series methods for empirical macroeconomics” (PDF). Foundations and Trends in Econometrics. 3 (4): 267–358.

$$p(\mathbf{y}|\boldsymbol{\gamma}) = \int p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\gamma})p(\boldsymbol{\theta}|\boldsymbol{\gamma})d\boldsymbol{\theta} \quad (2.5)$$

donde \mathbf{y} son los datos, $\boldsymbol{\theta}$ los parámetros del modelo VAR, $\boldsymbol{\gamma}$ los hiperparámetros de los a priori.

Para implementar este método, existe una librería en R llamada **BVAR** que aplica este método de estimación jerárquico. Por eficiencia computacional, la librería utiliza una a priori conjugada, la Normal-inversa-Wishart. La distribución de los parámetros del modelo VAR según esta a priori conjugada queda de la siguiente forma:

$$\begin{aligned} \boldsymbol{\beta}|\Sigma &\sim \mathcal{N}(\mathbf{b}, \Sigma \otimes \boldsymbol{\Omega}), \\ \Sigma &\sim \mathcal{IW}(\boldsymbol{\Psi}, \mathbf{d}) \end{aligned} \quad (2.6)$$

donde $\mathbf{b}, \boldsymbol{\Omega}, \boldsymbol{\Psi}, \mathbf{d}$ son funciones de un vector de hyperparámetros $\boldsymbol{\gamma}$.

En el artículo de Giannone et al. (2015) sobre el que se basa esta librería, consideran tres distribuciones a priori específicas: *Minnesota* (es el benchmark), *SOC*, y *SUR*.

La a priori Minnesota (Litterman 1980), impone la hipótesis de que todas las variables siguen procesos de paseo aleatorio. Esta especificación parsimoniosa suele dar buenos resultados en previsiones de series temporales macroeconómicas (Kilian y Lütkepohl 2017) y suele utilizarse como referencia para evaluar la precisión de un modelo. Sus parámetros son λ , α y ψ_j . λ controla anchura, α controla el nivel de reducción de observaciones más lejanas, y ψ_j reduce los retardos de las variables distintas de la dependiente.

Esta librería utiliza el muestreador de Gibbs, de la familia de los métodos MCMC, para obtener las probabilidades a posteriori.

2.1.3. Modelo VECM - Frecuentista

Los modelos de corrección de equilibrio son unos modelos clásicos de la econometría de series temporales que surgieron con la idea de no desperdiciar las relaciones a largo plazo existentes entre las series. Forman parte de los modelos *no estacionarios* en cuanto a que se trabaja con series no estacionarias.

Partimos de dos series temporales que han de ser integradas de orden 1. Una serie temporal integrada de orden 1 significa que necesita una diferencia para ser estacionaria. Lo siguiente es crear un modelo de regresión simple del estilo $Y_t = c + \beta X_t + u_t$ que recoge la relación de largo plazo entre las variables. u_t constituye las desviaciones/perturbaciones de la variable con respecto a ese nivel de equilibrio de largo plazo, el que determina el modelo. Si los residuos de este modelo son estacionarios (se utilizará el test de Dickey-Fuller para comprobarlo), se dice que las variables están cointegradas, y, entonces, se trabaja con las series sin ser estacionarias para conservar la relación de largo plazo.

Si las series son cointegradas, $Y_t = c + \beta X_t$ se denomina la relación de equilibrio, y, $m_t = Y_t - c - \beta X_t$ son las desviaciones respecto al equilibrio.

Cuando las series están cointegradas, existe una relación de largo plazo, pero también puede existir una relación contemporánea o de corto plazo. Por esto, si las series están cointegradas el modelo adecuado para estimar las relaciones entre las variables es:

$$\nabla Y_t = b \nabla X_t + \alpha(Y_{t-1} - c - \beta X_{t-1}) + \epsilon_t \quad (2.7)$$

donde ϵ_t es ruido blanco, α es la tasa de ajuste sobre los desequilibrios de largo plazo, β recoge la relación de largo plazo (en caso de ser variables expresadas en logaritmos expresa la elasticidad de Y_t con respecto de X_t), y b recoge la relación contemporánea entre las variables.

Para que se cumpla la igualdad en 2.7, $(Y_{t-1} - c - \beta X_{t-1})$ tiene que ser estacionario ya que $\nabla Y_t, \nabla X_t$ son estacionarias porque Y_t, X_t son $I(1)$ y ϵ_t es ruido blanco.

En el modelo expresado en 2.7, se pueden incluir más retardos tanto para la variable Y_t como para X_t . Quedaría así, un modelo de retardos distribuidos para la relación estacionaria, y la parte de corrección de equilibrio para la relación a largo plazo no estacionaria. El modelo presenta la siguiente forma:

$$\begin{aligned} \nabla Y_t = & b \nabla X_t + b_1 \nabla X_{t-1} + b_2 \nabla X_{t-2} + \cdots + b_s \nabla X_{t-s} + \\ & \alpha(Y_{t-1} - c - \beta X_{t-1}) + \\ & a_1 Y_{t-1} + a_2 Y_{t-2} + \cdots + a_r Y_{t-r} + \epsilon_t \end{aligned} \quad (2.8)$$

Al relacionar dos series $I(1)$ con un modelo de largo plazo, existe el peligro de que la relación entre las variables sea significativa cuando en verdad no existe relación alguna entre ellas, esto se denomina *relaciones espurias*. Cuando hay relaciones espurias entre dos series temporales, el modelo de relación a largo plazo presenta un R^2 muy elevado y coeficientes significativos indicando que existe una relación y que el modelo se ajusta muy bien, pero, cuando analizamos los residuos, tienen comportamiento no estacionario lo que sugiere que el modelo está mal formulado. Lo que está pasando es que la supuesta relación entre las variables está causada por otra variable, normalmente el *tiempo*.

Si los residuos no son estacionarios las variables no están cointegradas, y, si volvemos a estimar un modelo para la relación de largo plazo pero aplicando una primera diferencia a cada serie, todas las buenas propiedades del modelo anterior desaparecen.

Las relaciones espurias son muy comunes cuando se relacionan dos series temporales independientes no estacionarias. Daniel Peña en su libro *Análisis de series temporales* muestra este problema en la Figura 1.1

La figura presenta el histograma del estadístico $t = \frac{\hat{\beta}}{\text{se}(\hat{\beta})}$ al estimar regresiones entre dos paseos aleatorios independientes. Como indica Daniel Peña en su libro “el 75 % de las veces encontraremos un valor de este estadístico mayor que dos, con lo que concluiremos que la pendiente de la ecuación de regresión simple que relaciona ambas variables es distinta de cero. Por tanto, utilizando este estadístico encontraremos con alta probabilidad una relación entre las dos variables que no existe”.

A la hora de relacionar dos series temporales, es importante comprobar si están cointegradas o no ya que, si dos series están cointegradas y las relacionamos según la ecuación 2.7 o 2.8, la estimación del parámetro β es mejor que si se estimara con las series estacionarias, y, la varianza del estimador de β tiende a cero. Esto es lo que se denomina *superconsistencia* del estimador $\hat{\beta}$ cuando se estima un modelo como el de la ecuación 2.7 con las series no estacionarias si las variables están cointegradas (véase Daniel Peña *Análisis de series temporales* páginas 544-545).

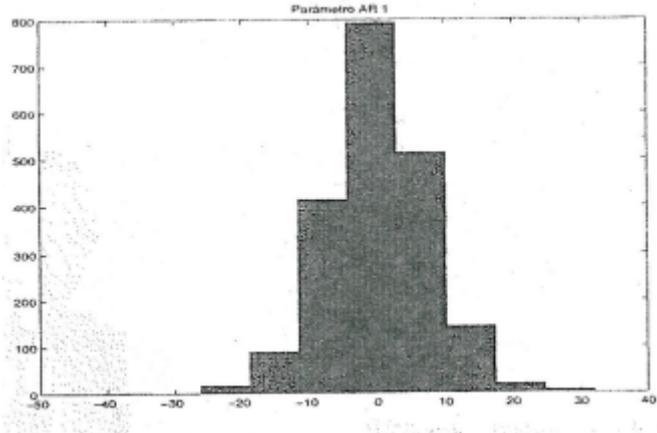


Figura 2.1: Histograma de los valores del estadístico t para la relación lineal entre dos paseos aleatorios independientes

Hasta ahora, se han presentado un tipo de modelos de corrección de equilibrio en el que sólo interactúan dos variables y de manera unidireccional. Estos modelos se pueden aplicar de manera bidireccional, $X_t \rightarrow Y_t$ y de $Y_t \rightarrow X_t$. Estos son los denominados modelos VEQCM (modelos vectoriales con mecanismos de corrección de equilibrio). Estas relaciones a largo plazo se pueden estimar también para más de dos variables.

En caso de que haya cointegración entre las variables, un modelo VEQCM quedaría de la siguiente forma:

$$\begin{aligned}\nabla X_t &= \alpha_1(X_{t-1} - c - \beta Y_{t-1}) + \epsilon_{1t} \\ \nabla Y_t &= \alpha_2(X_{t-1} - c - \beta Y_{t-1}) + \epsilon_{2t}\end{aligned}\tag{2.9}$$

El modelo anterior puede hacerse más complejo e incorporar relaciones a corto plazo como en 2.8 con el número de retardos que se desee.

Si partimos de un modelo $VAR(1)$ como el de 2.2, podemos comprobar si las series son estacionarias o no y qué tipo de modelo hay que considerar. Reformulamos el modelo en 2.2 de la siguiente forma:

$$\nabla \mathbf{X}_t = \Pi \mathbf{X}_{t-1} + \mathbf{a}_t\tag{2.10}$$

donde

$$\Pi = \begin{pmatrix} \phi_{11} - 1 & \phi_{12} \\ \phi_{21} & \phi_{22} - 1 \end{pmatrix} = -(I - \Phi_1)\tag{2.11}$$

Si Π tiene rango 2, las dos series son independientes y estacionarias, no hay relación de largo plazo por lo que se estimaría un modelo con las variables estacionarias. Si el rango es 1, hay una fila linealmente dependiente de la otra por lo que existe una relación de largo plazo y habría que estimar un modelo de corrección de equilibrio. En caso de que el rango fuese 0, tendríamos dos paseos aleatorios y se haría un modelo con las variables en diferencias.

En definitiva, estos modelos incorporan parte autorregresiva para las relaciones a corto plazo, y la relación a largo plazo estimada para no desperdiciar esta información. Estos modelos son una extensión de los modelos VAR.

Por todo esto, el procedimiento para crear un modelo de corrección de equilibrio a partir de dos series X_t e Y_t es el siguiente:

1. Comprobar que ambas series sean no estacionarias $I(1)$, si son integradas de otros órdenes todo se complica. Se utilizará el test de Dickey-Fuller para comprobar el grado de integración de cada serie junto con el correlograma. Un correlograma típico de variables no estacionarias suele presentar una estructura marcadamente decreciente en su función de autocorrelación simple.
2. Estimar una relación lineal como la de 2.7 por mínimos cuadrados.
3. Obtener los residuos del modelo estimado en el paso 2. y comprobar si son estacionarios, también se utilizará el test de Dickey-Fuller para contrastarlo.
4. Si las series son cointegradas, estimar un modelo de corrección de equilibrio como el de la ecuación 2.9.

2.1.4. Modelo VECM - Bayesiano (BVECM)

Los modelos bayesianos aportan mejoras a la estimación de los modelos VAR cuando se quiere estimar muchos parámetros, y, al ser los modelos de corrección de equilibrio una extensión de estos modelos, es posible que un enfoque bayesiano mejore también los resultados tanto para la inferencia de los parámetros, como para la capacidad predictiva.

Al igual que con los modelos BVAR, el enfoque bayesiano a los modelos de corrección de equilibrio trata los parámetros como variables con probabilidades a priori. La diferencia es que ahora, aparte de los términos autorregresivos, tenemos que estimar también la parte que recoge la relación de largo plazo y esto genera problemas en la estimación. Si miramos el modelo representado por 2.9, vemos que la estimación bayesiana se puede hacer directamente ya que esta ecuación junto con sus supuestos definen la función de verosimilitud y añadiendo una a priori se puede obtener la estimación bayesiana a posteriori. Pero, surgen problemas de *identificación* por el hecho de que la matriz Π en 2.10 suele ser de rango reducido. Existe un problema de *identificación global* ya que $\Pi = \alpha\beta'$ y $\Pi = \alpha AA^{-1}\beta'$ son idénticas para cada A no singular siendo α y β matrices $p \times r$ de rango completo donde $0 \leq r \leq p$ es el número de relaciones de cointegración. Si $r = p$, todas las variables tienen tendencia estacionaria. Este problema suele ser superado mediante la *normalización lineal*, donde $\beta = [I_r B']'$.

También puede haber un problema de *identificación local* cuando $\alpha = 0$ (en este caso β no entra en el modelo) pues la estimación bayesiana con una a priori no informativa puede resultar en una a posteriori impropia, esto es, una distribución a posteriori que no integra a 1. Este problema fue identificado por Kliebergen y Van Dijk (1994, 1998).

Para evitar estos problemas, Koop et al (2010) proponen una a priori para una “estimación eficiente de las a posteriori”. Su metodología se puede implementar junto con la librería `bvarools` en R por lo que se usará este paquete para la estimación de los modelos BVEC.

El procedimiento que proponen evita el problema que surge con las distribuciones a priori linealmente normalizadas que son muy utilizadas para el problema de la identificación global y que generan una a priori poco eficiente computacionalmente para trabajar en el espacio de cointegración y que además, impone restricciones en el espacio de cointegración y según la distribución a priori que se elija, puede estimar valores en el espacio de cointegración donde la normalización lineal es inválida, violando el supuesto de que esta transformación es válida. Además, evitan también el problema de no convergencia en el muestreador de Gibbs causado por la identificación local.

La especificación de las a priori y el algoritmo para obtener las a posteriori implementado en **bvartools** está basado en el trabajo de Koop & Strachan (2010). A continuación, se detalla la especificación del modelo.

Tenemos que $\mathcal{P} = sp(\beta)$ denota el espacio de cointegración que es un hiperplano r -dimensional en un espacio n -dimensional. Deseamos realizar una inferencia bayesiana sobre este espacio sin imponer la identificación de manera que restrinja este espacio. Además, queremos desarrollar unas a priori sensibles.

Como a priori para β , y, por ello, de $\mathcal{P} = sp(\beta)$, usamos una matriz angular central gaussiana (MACG(P_t), Chikuse, 1990):

$$p(\beta) \propto |P_t|^{-r/2} |\beta'(P_t)^{-1}\beta|^{-n/2} \quad (2.12)$$

P_t es $n \times n$ y determina la localización central de $\mathcal{P} = sp(\beta)$ y también la dispersión alrededor de la localización central.

P_t se define de la siguiente forma:

$$P_t = HH' + \tau H_{\perp}H'_{\perp} \quad (2.13)$$

donde τ es un escalar entre 0 y 1 que controla la dispersión alrededor del centro, y H , es una matriz semiortogonal fruto de la transformación de una matrix H^g que contiene los valores de los coeficientes a priori. La relación entre las matrices es $H = H^g(H^g'H^g)^{-1/2}$.

Cabe destacar que si $\tau = 1$, entonces, $P_t = I_n$ y tendremos una a priori uniforme no informativa para β , y si $\tau \approx 0$ estaríamos introduciendo mucha confianza en que la localización de la a priori es cierta.

En cuanto a α , se elige una a priori con media cero y de dimensión reducida

$$\alpha|\beta, \tau, \Sigma, \nu \sim MN(0, \nu(\beta'P_{1/\tau}\beta)^{-1} \otimes G) \quad (2.14)$$

donde, MN es una distribución matriz-variante-Normal (ver Bauwens et al., 1999, Appendix A), $P_{1/\tau} = HH' + \tau^{-1}H_{\perp}H'_{\perp}$, G es una matrix $n \times n$ y ν es un escalar que controla el nivel de reducción de dimensión. La matriz G puede ser igual a Σ (Strachan & Inder, 2004), pero puede tomar otros valores para más flexibilidad. ν puede ser fijo o se hace un modelo jerárquico donde tomará diferentes valores.

Por último, se usa la priori no informativa estándar para Σ

$$p(\Sigma) \propto |\Sigma|^{-(n+1)/2} \quad (2.15)$$

pero también se puede usar una Wishart invertida.

Las distribuciones a priori reflejadas desde 2.12 hasta 2.15, son sensibles. Además, Strachan y Inder (2004) dan más información y motivación para usar estas a prioris para el caso que $G = \Sigma$ ya que además de las buenas propiedades con las que cuenta, es computacionalmente más simple. Otra ventaja de esta formulación, es que la a priori estándar no informativa para α se consigue simplemente fijando $\frac{1}{\nu} = 0$ y permite obtener las estimaciones a posteriori también con esta a priori.

En cuanto a la estimación a posteriori de los parámetros, Koop y Strachan utilizan muestreador de Gibbs colapsado que junto con las a prioris anteriores ofrece una computación eficiente y simple.

Los muestras de Σ mediante MCMC se pueden obtener de una Wishart invertida

$$\Sigma | \alpha, \beta, Data \sim IW([y - X\beta\alpha']'[y - X\beta\alpha'], T), \quad (2.16)$$

pero, si G es fija o igual a Σ , la distribución apropiada es

$$\Sigma | \alpha, \beta, Data \sim IW([y - X\beta\alpha']'[y - X\beta\alpha'] + \nu^{-1}\alpha(\beta'P_{1/\tau}\beta)\alpha', T + r), \quad (2.17)$$

τ y ν pueden tratarse como parámetros desconocidos, y, en ese caso, se ha de elegir una a priori para ellos y un paso extra en el MCMC que saque valores de τ y ν . En este algoritmo implementado por Koop y Strachan, no los tratan como parámetros desconocidos y se centran en los problemas relacionados a obtener α y β .

En el algoritmo MCMC que desarrollan, tratan de evitar el problema que surge al imponer la semiortogonalidad de β . Esta restricción hace que la distribución a posteriori condicional de β es no-estándar. Para solucionar esto aplican la siguiente transformación:

$$\beta\alpha' = (\beta k)(\alpha k^{-1})' = [\beta(\alpha'\alpha)^{1/2}] [\alpha(\alpha'\alpha)^{1/2}] = BA' \quad (2.18)$$

siendo k una matriz definida positiva y $A = \alpha k^{-1}$ semiortogonal. Cabe destacar las siguientes relaciones entre los parámetros en 2.18:

$$\begin{aligned} k &= (\alpha'\alpha)^{1/2} \\ \beta &= B(B'B)^{-1/2} \\ B'B &= \alpha'\alpha \end{aligned} \quad (2.19)$$

Estas últimas relaciones son importantes ya que en 2.18 β es semiortogonal y α no está restringida, y en la que relaciona A y B resulta que B no está restringida y A es semiortogonal, por lo que el muestreador colapsado de Gibbs que desarrollan va alternando entre estas dos ya que se demuestra que es útil.

Teniendo en cuenta que la a priori para (α, β) de 2.12 y 2.14 implica que la a priori para (A, B) es

$$B|A \sim MN(0, (A'G^{-1}A)^{-1} \otimes P_\tau), \quad (2.20)$$

$$p(A) \propto |G|^{-r/2} |A'G^{-1}A|^{-n/2} \quad (2.21)$$

y combinando las a priori de 2.20 y 2.21 con la verosimilitud, se obtiene una a distribución a posteriori condicionada que es Normal. El algoritmo que han desarrollado, después de elegir un valor inicial para $\beta^{(0)}$ repite los siguientes pasos:

1. Sacar $\alpha^{(*)}$ de $P(\alpha|\beta, Data)$ y transformar esto para obtener $A^{(*)} = \alpha^{(*)}(\alpha^{(*)}'\alpha^{(*)})^{-1/2}$
2. Sacar $B^{(*)}$ de $p(B|A^{(*)}, Data)$ y transformar esto para obtener $\beta^{(s)} = B^{(s)}(B^{(s)'}B^{(s)})^{-1/2}$ y $\alpha^{(s)} = A^{(*)}(B^{(s)'}B^{(s)})^{1/2}$

Para más detalles sobre por qué esto forma un muestreador colapsado de Gibbs y más detalles sobre el algoritmo, ver Koop y Strachan (2010).

Una ventaja de este algoritmo es que sólo saca muestras de una distribución normal.

Cabe destacar que los modelos BVEC rara vez se tienen en cuenta para predecir las variables macroeconómicas. Particularmente en el caso de las aplicaciones en dimensiones superiores. El contraste de cointegración con dimensiones crecientes se convierte en una carga y no hay una forma automática de estimar los modelos BVEC (e.g. Jan Prüser (2021)). Se han hecho avances en este aspecto incorporando reducciones jerárquicas de la matriz de relación de largo plazo para los modelos BVEC (e.g. Jan Prüser (2021)).

Aún así, un estudio de simulación revela que los modelos BVEC equipados con ciertas distribuciones a priori funcionan bien en una serie de escenarios. En presencia de cointegración, los modelos BVEC pueden mejorar la precisión de la estimación respecto a los BVAR en primeras diferencias. En ausencia de cointegración, estas distribuciones a priori específicas son capaces de reducir la matriz de relación de largo hacia cero, de modo que la precisión de la estimación del modelo BVEC es similar a la del modelo BVAR estimado en primeras diferencias (e.g. Jan Prüser (2021)).

Si observamos los modelos en 2.7, 2.8 y 2.9, vemos que una vez estimados los modelos, sus parámetros permanecen constantes en el tiempo. Estaríamos asumiendo que esta relación estimada se mantiene en el tiempo, tanto la relación a corto como a largo plazo. Esto plantea diferentes problemas ya que las series temporales por su naturaleza evolucionan, fluctúan, y, algunas, tienen ciclos. Es de este problema de donde surge la idea para este trabajo de fin de grado. Además, como se ha comentado, la estimación frecuentista de estos modelos suele tener malas capacidades de predicción para grandes conjuntos de datos, y, sobre todo, cuando hay muchas variables.

2.1.5. Análisis de la capacidad predictiva de los modelos

A parte de analizar las diferencias en la inferencia de los parámetros según cada método de estimación, es interesante estudiar la capacidad predictiva de los modelos. Las series temporales tienen la característica de que las observaciones que están muy cercanas en el tiempo están muy correlacionadas, por lo que separar los datos en muestra de train y

test de forma aleatoria no es lo más apropiado ya que se produce un tipo de *information leakage*. Predecir una observación en el tiempo $j + s$ cuando $s \rightarrow 1$ habiendo utilizado la observación j en los datos de entrenamiento, hace que el error de test resultante esté sesgado y no sea representativo de la capacidad predictiva del modelo ya que a medida que $s \rightarrow 1$ la correlación entre las observaciones suele ser más fuerte. Además, no sólo es cuestión de la correlación entre observaciones cercanas, sino que una característica importante de los modelos para series temporales es si son capaces de captar el comportamiento que tendrá la serie en el futuro, tanto en tendencia como encuanto a los ciclos y evolución de la variabilidad. Al muestrear de forma aleatoria, no es posible medir esto.

Por todo lo comentado, el procedimiento habitual a la hora de estimar la capacidad predictiva de modelos para series temporales se basa en conservar la estructura temporal de los datos. Algunos de los procedimientos habituales son: usar un conjunto de train y otro de test, hacer varias divisiones de train y test a lo largo de la serie que conserven la estructura temporal, y walk-forward validation. Este último procedimiento es el que se va a emplear para medir la capacidad predictiva de los modelos.

El método de walk-forward validation funciona de la siguiente forma:

1. Se elige un número de observaciones t , normalmente pequeño, con el que se estimará el modelo inicial.
2. Con el modelo inicial estimado con las t primeras observaciones, se predice la observación en el momento $t + 1$.
3. Se mide el error entre la predicción y el valor real, normalmente mediante con el cuadrado del error.
4. La observación en el periodo $t + 1$ se añade al conjunto de entrenamiento, por lo que ahora $t = t + 1$ para entrenar el modelo.
5. Se repite 1-4 hasta que se haya recorrido toda la serie. Al terminar, se calcula el error cuadrático medio (RMSE).

Para este trabajo de fin de grado, el algoritmo de walk-forward validation se va a modificar por motivos computacionales. Debido a que los modelos bayesianos utilizan un tipo de muestreador de Gibbs, tardan más en estimarse que los modelos frecuentistas. Lo mismo pasa con el cálculo de las predicciones. Por esto, el algoritmo se va a modificar de la siguiente forma:

0. Para las series diarias, el horizonte temporal de predicción en vez de ser $t + 1$ será $t + 30$, es decir, se predecirán con un horizonte temporal de un mes. Para las series trimestrales, el horizonte temporal será $t + 4$, también un horizonte temporal de un año.
1. Se elige un número de observaciones t tal que el número de iteraciones del algoritmo sea igual a 30, teniendo en cuenta el horizonte temporal de predicción especificado en 0.
2. Con el modelo inicial estimado con las t primeras observaciones, se predicen las observaciones hasta el momento $t + 30$ para series diarias y $t + 4$ para series trimestrales.
3. Se mide el error entre las predicciones y los valores reales y se calcula el RMSE para esa iteración del algoritmo.

4. Las observaciones que se han predecido se añaden al conjunto de entrenamiento, por lo que ahora $t = t + 30$ para series diarias y $t = t + 4$ para series trimestrales a la hora de volver a estimar el modelo.
5. Se repite 1-4 hasta que se haya recorrido toda la serie. Al terminar, se calcula la media de todos los 30 RMSE.

De esta forma se consigue que sea computacionalmente viable realizar este procedimiento, y se consigue una muestra de 30 RMSE para cada serie que se puede utilizar para calcular intervalos de confianza.

2.2. Análisis de datos

Como se indicó en la introducción al trabajo, para comparar el enfoque bayesiano frente al frecuentista se van a aplicar los diferentes modelos a una serie de conjuntos de datos. Primero, vamos a simular dos conjuntos de series temporales: uno para los modelos autorregresivos y otro para los modelos de corrección de equilibrio. De esta forma, sabiendo los parámetros “reales” se compararán las estimaciones entre el modelo bayesiano y el frecuentista. Por otro lado, se usarán dos conjuntos de datos reales que recogen por un lado los swaps a 9 meses y el LIBOR a 12 meses en la eurozona, y, por otro lado, datos macroeconómicos de los estados Unidos desde mediados del siglo 20 hasta principios del 21.

Primero se simularán los datos, y después se empezará por un análisis descriptivo de los mismos (gráfico de las series simuladas y correlogramas) seguido de los respectivos contrastes para determinar el orden de integración con el contraste de Dickey-Fuller. Después, se comentarán las posibles transformaciones necesarias para trabajar con series estacionarias, y, para los datos que se usarán en la estimación de los modelos de corrección de equilibrio, se analizará si las series son cointegradas o no. Para los conjuntos de datos reales se seguirán los mismos pasos.

2.2.1. Simulación de los datos

Primero vamos a simular unas series temporales que se usarán para comparar la estimación frecuentista y bayesiana en modelos autorregresivos. Para la simulación, vamos a generar dos conjuntos de series temporales que sigan el siguiente modelo VAR:

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} 0.5 & 0.0 \\ -0.3 & 0.4 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} -0.1 & 0.1 \\ 0.2 & -0.3 \end{bmatrix} \begin{bmatrix} y_{t-2} \\ x_{t-2} \end{bmatrix} + \begin{bmatrix} 0.2 & 0.2 \\ 0.3 & -0.1 \end{bmatrix} \begin{bmatrix} y_{t-3} \\ x_{t-3} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \quad (2.22)$$

con matriz de varianza-covarianzas

$$\Omega = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix} \quad (2.23)$$

Se simulará un conjunto de datos con dos series de 1000 observaciones cada una, y otro conjunto de dos series con 100 observaciones cada una. Ambas siguiendo la misma

relación entre las variables. Para esto, se va a hacer uso de la función `VAR.sim` dentro de la librería `tsDyn`.

- Análisis del conjunto de datos simulado para modelos autorregresivos con 1000 observaciones por cada serie:

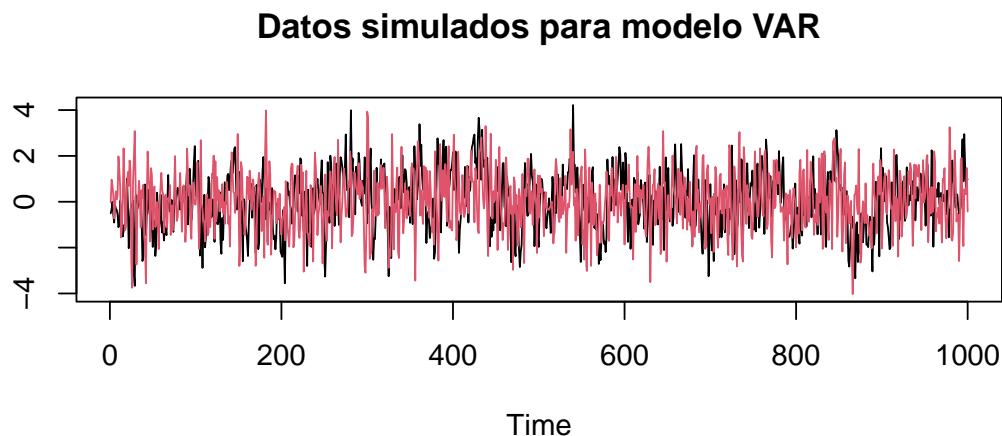


Figura 2.2: Datos simulados para VAR.

El modelo VAR que se ha usado para simular los datos ha de cumplir la condición de estacionariedad, lo que significa que las raíces del determinante de cada matriz polinomial, Φ_k , han de ser mayores que la unidad. Se puede comprobar que todas las matrices polinomiales en 2.22 cumplen esta condición. Ahora, vamos a analizar cada serie por separado.

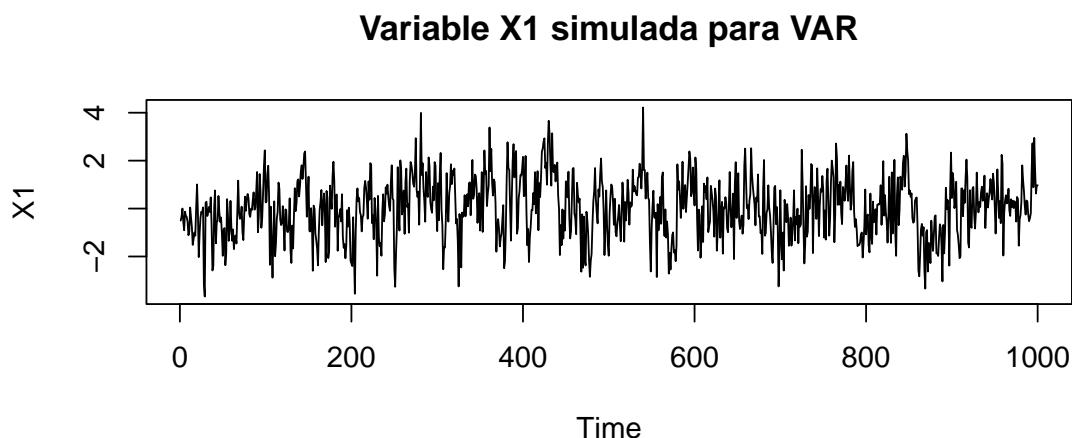


Figura 2.3: Variable X1 simulada para VAR.

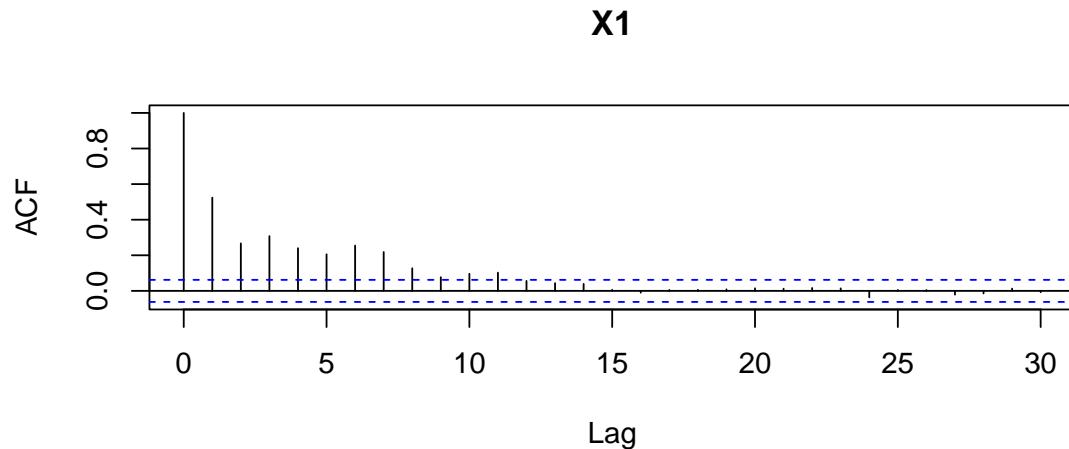


Figura 2.4: Autocorrelación simple de X1 para VAR.

En cuanto a la estacionariedad en varianza, la serie presenta una varianza muy constante a lo largo del tiempo, no se aprecian factores estacionales, y, la serie fluctúa en torno a una media alrededor de 0 sin presentar tendencias, crecimiento sistemático o patrones no estacionarios. Esto se debe a que en la simulación de los datos no se ha introducido ni estacionalidad, tendencia o crecimiento sistemático.

En cuanto a su autocorrelación simple y parcial, hay algunos retardos que se salen de las bandas de confianza pero en general la mayoría están dentro de ellas y no hay una estructura marcada decreciente característica de series no estacionarias. Para comprobar la hipótesis de estacionariedad que se ha planteado con el análisis descriptivo, vamos a realizar un test de Dickey-Fuller aumentado para contrastar si la serie tiene raíces unitarias o no:

	Augmented DF	P-Value
X1_var	-7.1873	< 0.01

El test indica que la serie es estacionaria y que no presenta raíces unitarias al 5 % de significación. Recordemos que la hipótesis alternativa es que la serie es estacionaria.

Vamos a repetir este proceso para la otra serie simulada para el modelo VAR.

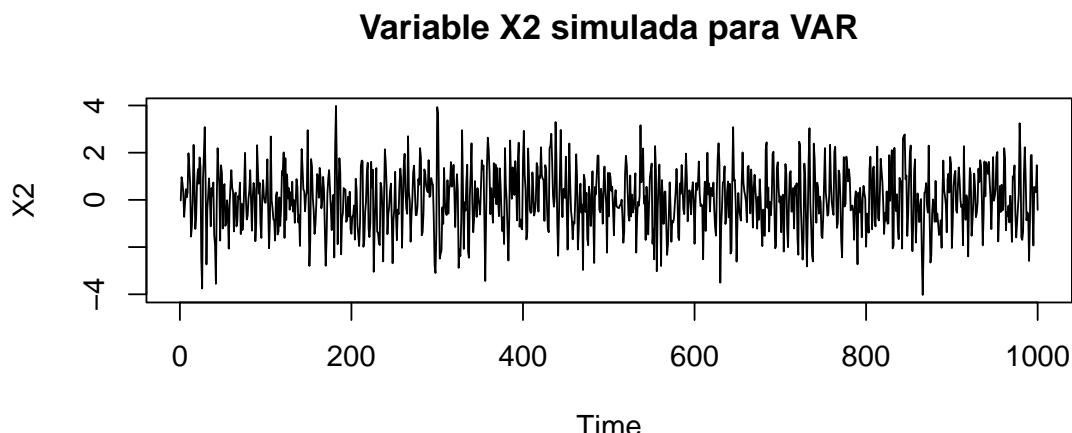


Figura 2.5: Serie temporal X2 datos simulados.

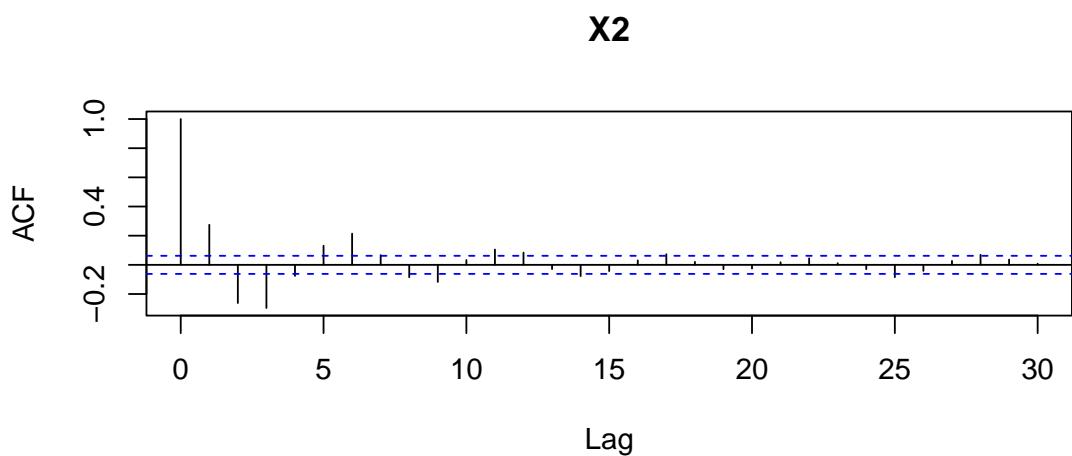


Figura 2.6: Autocorrelación simple de X2 simulada para VAR.

Igual que con la serie anterior, la varianza parece estacionaria, la serie fluctúa en torno a una media que es más o menos el 0, no hay comportamientos estacionales ni crecimientos sistemáticos o tendencias.

En cuanto a su autocorrelación simple y parcial, hay algunos retardos que se salen de las bandas de confianza pero no existe ningún tipo de estructura marcada que indique no estacionariedad. Por último, realizamos el test de Dickey-Fuller aumentado para contrastar la hipótesis de no estacionariedad:

	Augmented DF	P-Value
X2_var	-9.1044	< 0.01

Esta serie también es estacionaria según el test de Dickey-Fuller aumentado tanto al 5 % como al 1 %.

Estos resultados no son sorprendentes dado que la función utilizada para simular los datos recibe un modelo VAR con parámetros específicos y devuelve series estacionarias sacadas de ese modelo VAR que se introduce. Además, no se ha metido ningún tipo de tendencia, constante o factor estacional a la función para simular los datos.

- Análisis del conjunto de datos simulado para modelos autorregresivos con 100 observaciones por cada serie:

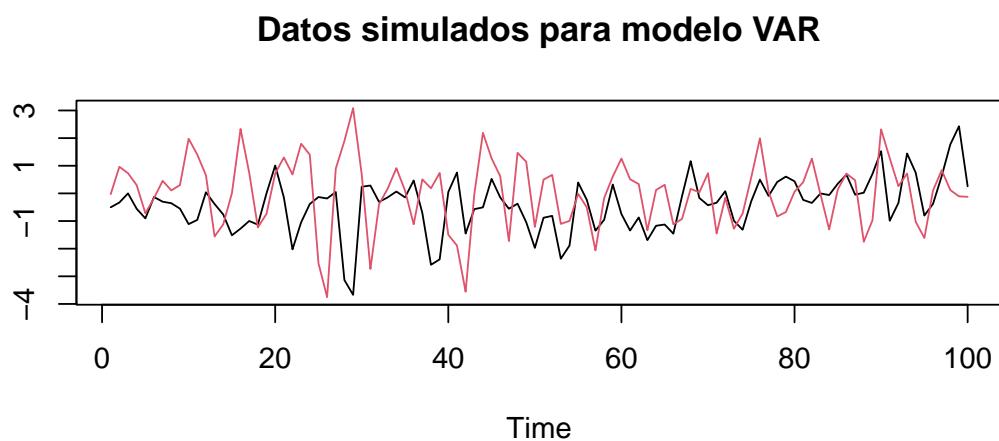


Figura 2.7: Datos simulados para VAR.

Ahora, vamos a analizar cada serie por separado.

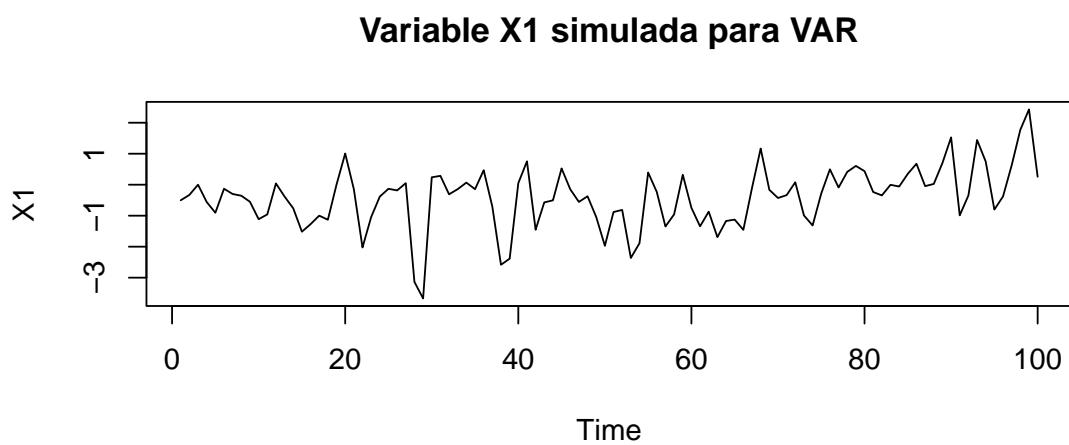


Figura 2.8: Variable X1 simulada para VAR.

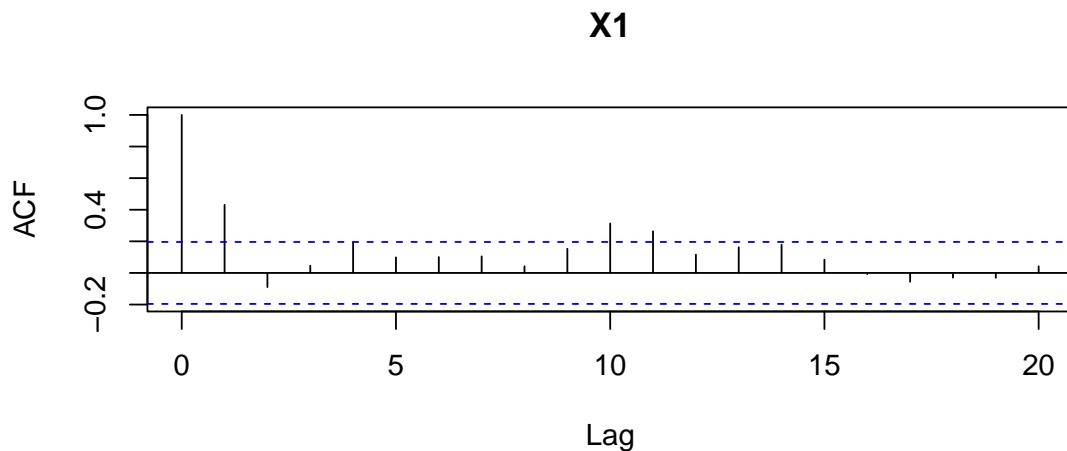


Figura 2.9: Autocorrelación simple de X1 para VAR.

En cuanto a la estacionariedad en varianza, la serie presenta una varianza constante a lo largo del tiempo, no se aprecian factores estacionales, y, la serie fluctúa en torno a una media alrededor de 0 con algunas rachas por encima y por debajo de este valor pero sin llegar a ser tendencial, crecimiento sistemático o patrones no estacionarios.

En cuanto a su autocorrelación simple y parcial, hay algunos retardos que se salen de las bandas de confianza pero en general la mayoría están dentro de ellas y no hay una estructura marcada decreciente característica de series no estacionarias. Para comprobar la hipótesis de estacionariedad que se ha planteado con el análisis descriptivo, vamos a realizar un test de Dickey-Fuller aumentado para contrastar si la serie tiene raíces unitarias o no:

	Augmented DF	P-Value
X1_var100	-3.9391	0.01498

El test indica que la serie es estacionaria y que no presenta raíces unitarias al 5 % de significación.

Vamos a repetir este proceso para la otra serie simulada para el modelo VAR.

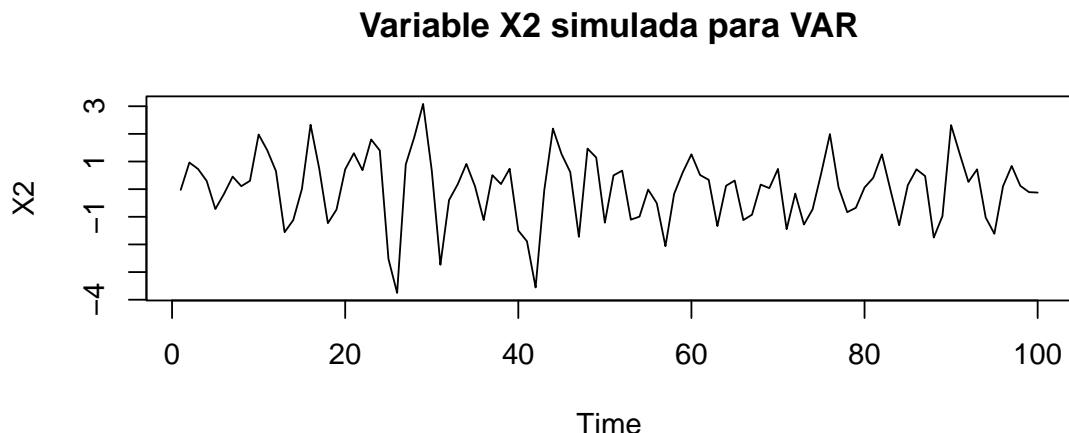


Figura 2.10: Serie temporal X2 datos simulados.

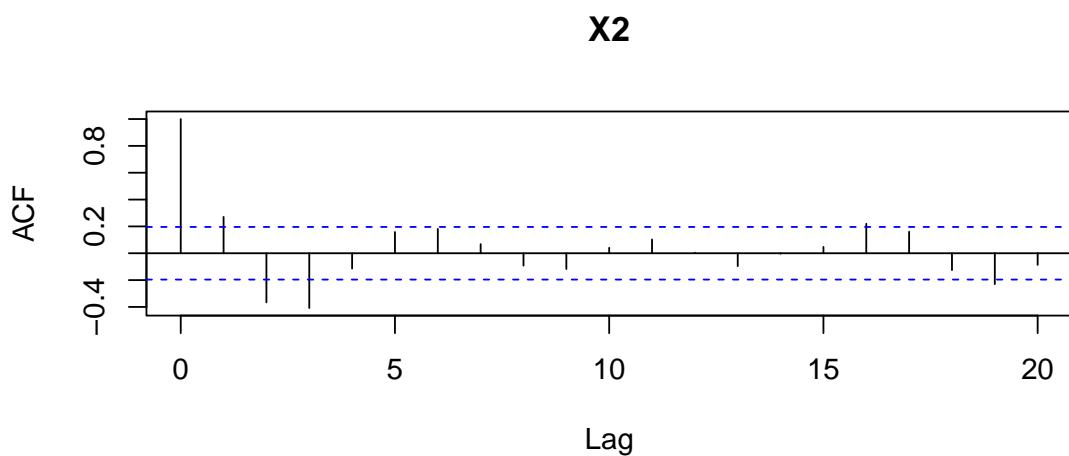


Figura 2.11: Autocorrelación simple de X2 simulada para VAR.

Igual que con la serie anterior, la varianza parece estacionaria, la serie fluctúa en torno a una media que es más o menos el 0, no hay comportamientos estacionales ni crecimientos sistemáticos o tendencias.

En cuanto a su autocorrelación simple y parcial, hay algunos retardos que se salen de las bandas de confianza pero no existe ningún tipo de estructura marcada que indique no estacionariedad. Por último, realizamos el test de Dickey-Fuller aumentado para contrastar la hipótesis de no estacionariedad:

	Augmented DF	P-Value
X2_var100	-5.8493	< 0.01

Esta serie también es estacionaria según el test de Dickey-Fuller aumentado tanto al 5 % como al 1 %.

Estos resultados no son sorprendentes dado que la función utilizada para simular los datos recibe un modelo VAR con parámetros específicos y devuelve series estacionarias sacadas de ese modelo VAR que se introduce. Además, no se ha metido ningún tipo de tendencia, constante o factor estacional a la función para simular los datos.

El siguiente paso es generar los datos para estimar los modelos de corrección de equilibrio. Estos modelos reciben series cointegradas, por lo que para simular datos que luego se puedan usar para estimar un modelo VEC, aparte de relaciones contemporáneas o de corto plazo como las que se han simulado para los datos del modelo VAR, habrá que incorporar una relación de largo plazo entre las variables.

Para conseguir esto, existe una función en R llamada `VECM.sim` que funciona de la misma forma que la función utilizada para simular series para un modelo VAR. El modelo VEC a través del cual vamos a simular datos es el siguiente:

$$\begin{bmatrix} \nabla X_{1t} \\ \nabla X_{2t} \end{bmatrix} = \begin{bmatrix} -0.3 \\ 0.2 \end{bmatrix} (X_{1t-1} - 1.5X_{2t-1}) + \begin{bmatrix} 0.2 & -0.3 \\ 0.4 & -0.2 \end{bmatrix} \begin{bmatrix} \nabla X_{1t-1} \\ \nabla X_{2t-1} \end{bmatrix} + \begin{bmatrix} 0.2 & 0.1 \\ 0.1 & -0.25 \end{bmatrix} \begin{bmatrix} \nabla X_{1t-2} \\ \nabla X_{2t-2} \end{bmatrix} \quad (2.24)$$

- Análisis del conjunto de datos simulado para modelos de corrección de equilibrio con 1000 observaciones:

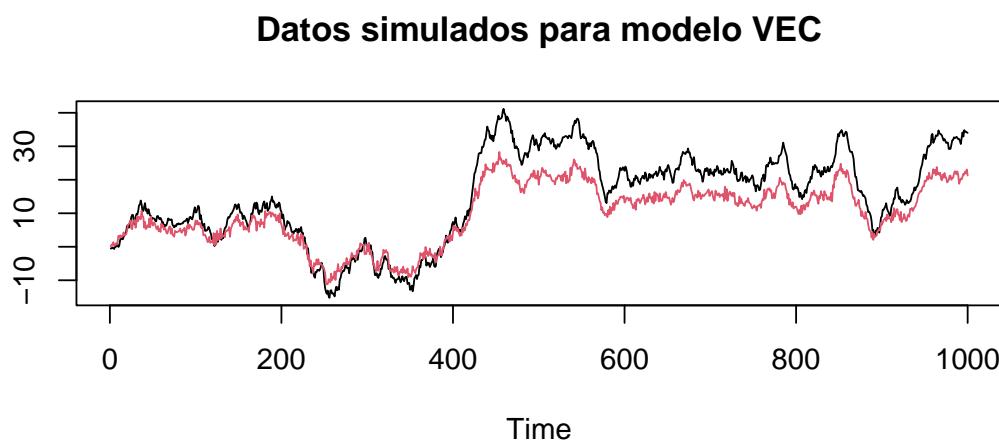


Figura 2.12: Datos simulados para VECM.

Como hemos visto en la metodología de los modelos VECM, para poder estimar este tipo de modelos las series han de ser ambas $I(1)$ y estar cointegradas, por lo tanto en el análisis de los datos tendremos que contrastar que las series sólo necesitan una diferencia regular, que están relacionadas por un modelo lineal, y, que los residuos de este modelo son estacionarios. Lo primero que vamos a hacer es comprobar la condición de que ambas series han de ser $I(1)$.

La serie X_1 presenta las siguientes características recogidas en su evolución temporal y el correlograma:

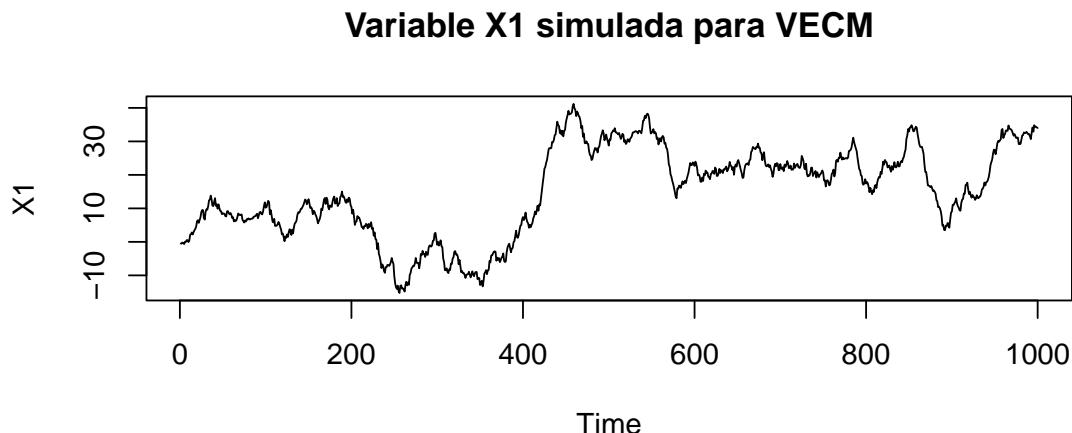


Figura 2.13: Variable X1 para VECM.

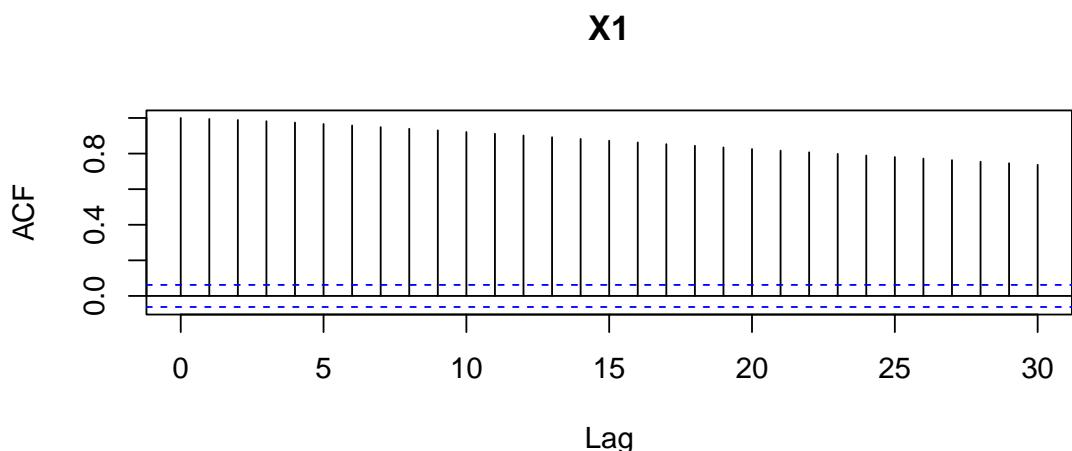


Figura 2.14: Autocorrelación simple de X1 para VECM.

Observando la serie, vemos que la varianza es constante a lo largo de la misma, pero no hay estacionariedad en la parte regular pues la serie presenta una tendencia positiva marcada. Además, si miramos la autocorrelación simple, se observa la típica estructura decreciente característica de series no estacionarias. En cuanto a estacionalidad, la serie no es estacionaria en la parte estacional como se puede observar en el gráfico de la serie, esto se debe a que a la hora de simular los datos no se han introducido patrones estacionales en el modelo.

Vamos a contrastar la hipótesis de no estacionariedad, y, por lo tanto de que la serie es al menos $I(1)$ mediante el test de Dickey-Fuller aumentado.

	Augmented DF	P-Value
X1_vecm	-2.5244	0.3563

Al no rechazar la hipótesis nula, el test nos indica que la serie presenta raíces unitarias y que es necesario aplicar diferencias. Vamos a aplicar una primera diferencia a la serie X_1 generada para el modelo VECM, y, volveremos a contrastar si tiene raíces unitarias. Como la serie ha de ser $I(1)$ para entrar en el modelo VECM, se tiene que rechazar el contraste de Dickey-Fuller aumentado.

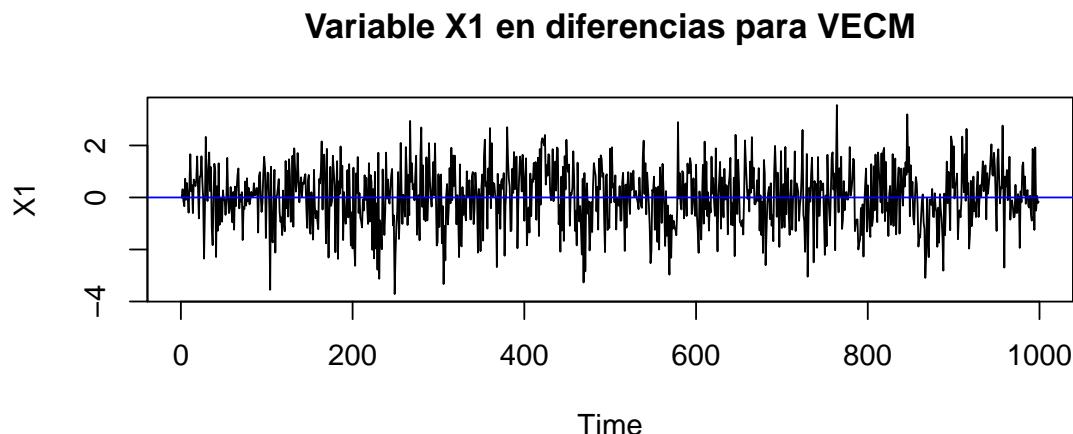


Figura 2.15: Variable X1 en diferencias para VECM.

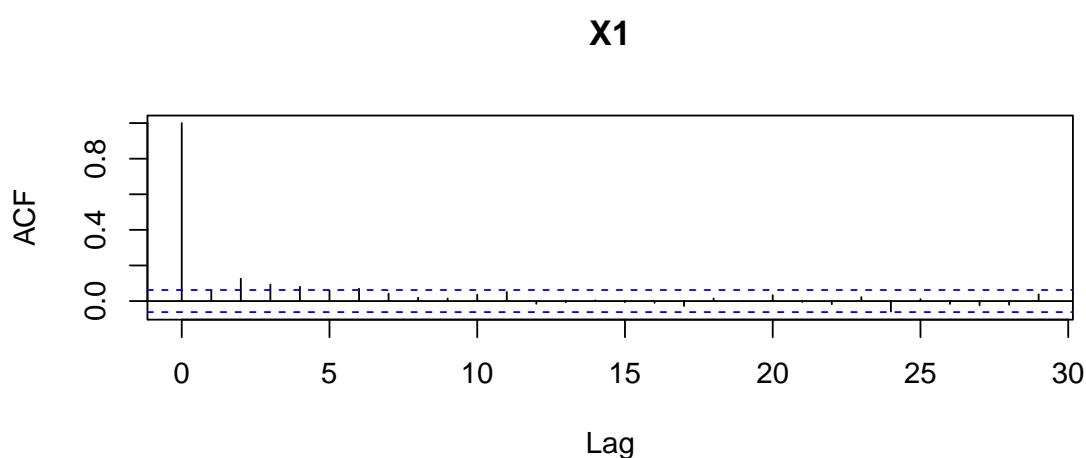


Figura 2.16: Autocorrelación simple de X1 en diferencias para VECM.

Vemos que al aplicar una primera diferencia regular la serie toma valores en torno a una media (0) y no presenta ni tendencia, crecimiento sistemático, o comportamientos de paseo aleatorio. Además, la estructura de la autocorrelación simple se ha eliminado y ahora no hay retardos que sobrepasen las bandas de confianza. Por último, vamos a constatar que la serie no necesita otra diferencia regular mediante el test de Dickey-Fuller aumentado.

	Augmented DF	P-Value
diff_X1_vecm	-8.381	< 0.01

Vemos que se rechaza la hipótesis nula de que la serie no es estacionaria por lo que podemos concluir que la serie original X_1 es $I(1)$.

Ahora, vamos a repetir este proceso con la serie X_2 . Comenzaremos con el análisis descriptivo de la serie original.

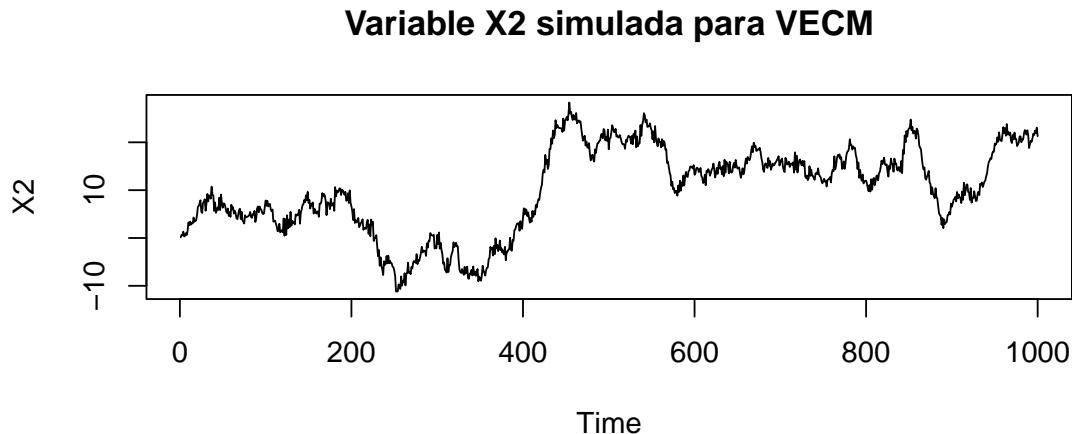


Figura 2.17: Variable X2 para VECM.

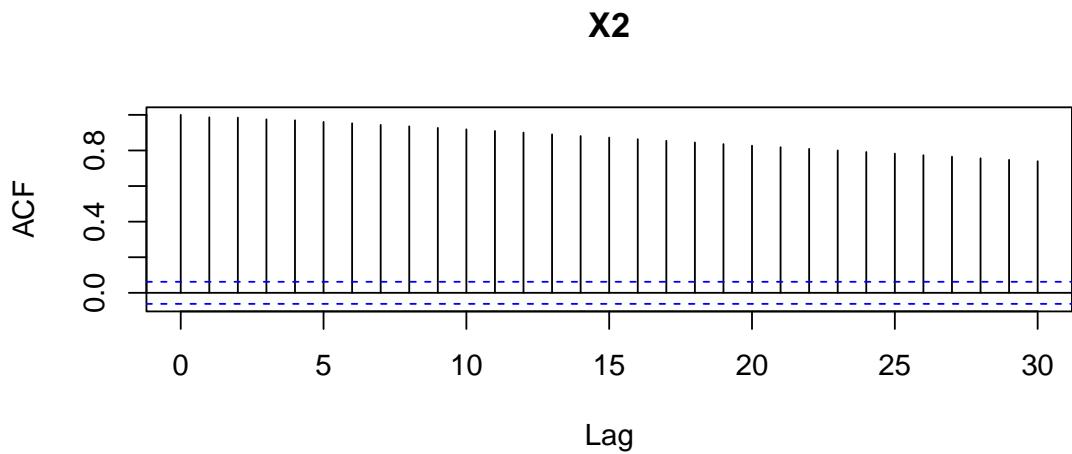


Figura 2.18: Autocorrelación simple de X2 para VECM.

Igual que con X_1 , la serie presenta una tendencia positiva, no oscila en torno a una media, y su autocorrelación simple presenta una estructura decreciente muy marcada con varios retardos significativos. En cuanto a la estacionalidad, la serie no presenta comportamiento estacional. A continuación, usamos otra vez el test de Dickey-Fuller aumentado para contrastar que la serie necesita una diferencia regular.

	Augmented DF	P-Value
X2_vecm	-2.4161	0.4022

Se acepta la hipótesis de que la serie no es estacionaria por lo que necesitamos aplicar diferencias. Vamos a diferenciar la serie y ver si necesita otra diferencia o si con una sirve.

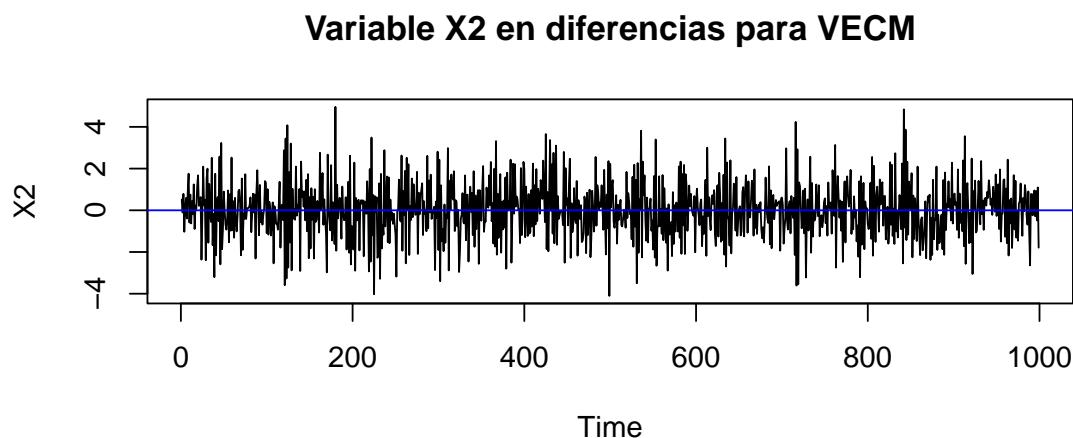


Figura 2.19: Variable X2 en diferencias para VECM.

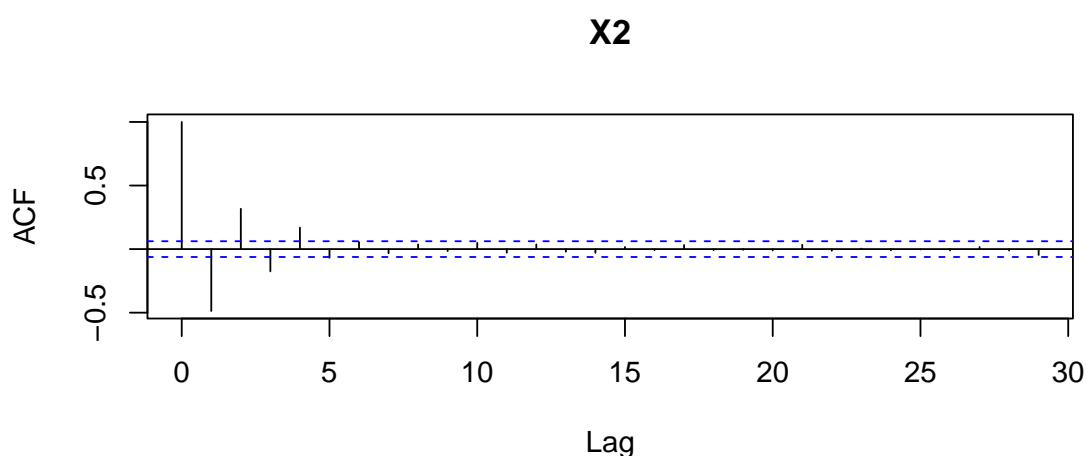


Figura 2.20: Autocorrelación simple de X2 en diferencias para VECM.

Tanto el gráfico de la serie diferenciada como la autocorrelación simple indican que la serie es estacionaria y que no hace falta otra diferencia en la parte regular. La serie oscila en torno al 0, y se ha eliminado la estructura de la autocorrelación simple. Vamos a contrastarlo con el test de Dickey-Fuller aumentado.

	Augmented DF	P-Value
diff_X2_vecm	-8.7532	< 0.01

Rechazamos H_0 con lo que tenemos que $X2$ es $I(1)$ igual que $X1$. El siguiente paso es crear un modelo que relacione las variables en el largo plazo del tipo $Y_t = c + \beta X_t + u_t$ por mínimos cuadrados ordinarios.

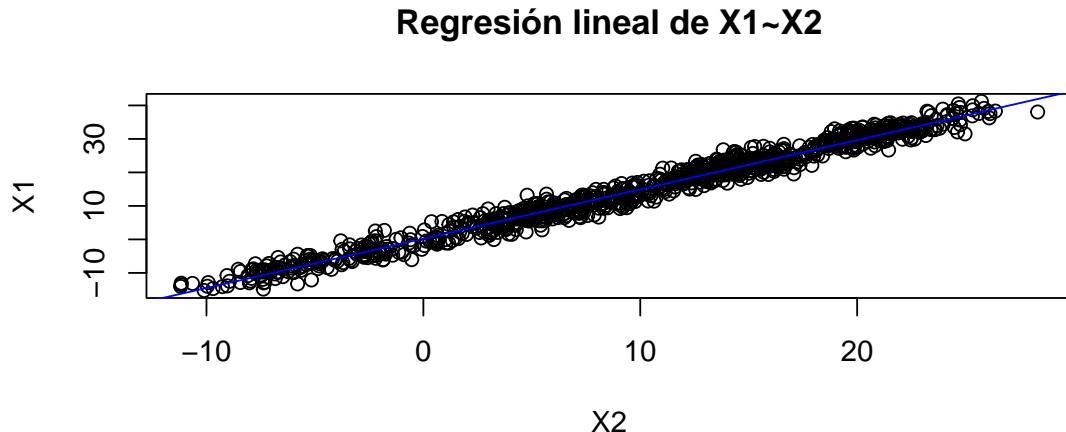


Figura 2.21: Relación lineal entre X1 y X2 simuladas para VECM.

La relación de largo plazo es positiva y los valores de los estadísticos e indicadores de ajuste se recogen en el siguiente summary:

	Estimate	Std. Error	t.value	p-value
Intercept	0.2066	0.0954	2.164	0.0307
X2	1.4723	0.0070	207.80	<2.2e-16

El estimador de β_1 es significativo con valor 1.472, muy cercano al 1.5 real, el R^2 es de 0.9774, y el estadístico F es significativo por lo que el modelo es globalmente significativo. Para contrastar si existe cointegración entre las series hace falta que los residuos de este modelo sean estacionarios. La serie temporal de los series presenta las siguientes características.

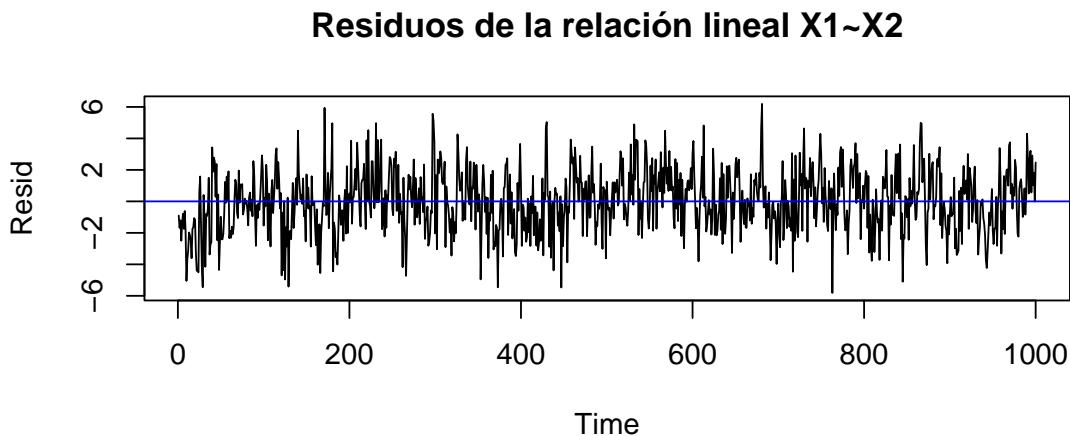


Figura 2.22: Residuos de la relación lineal entre variables simuladas para VECM.

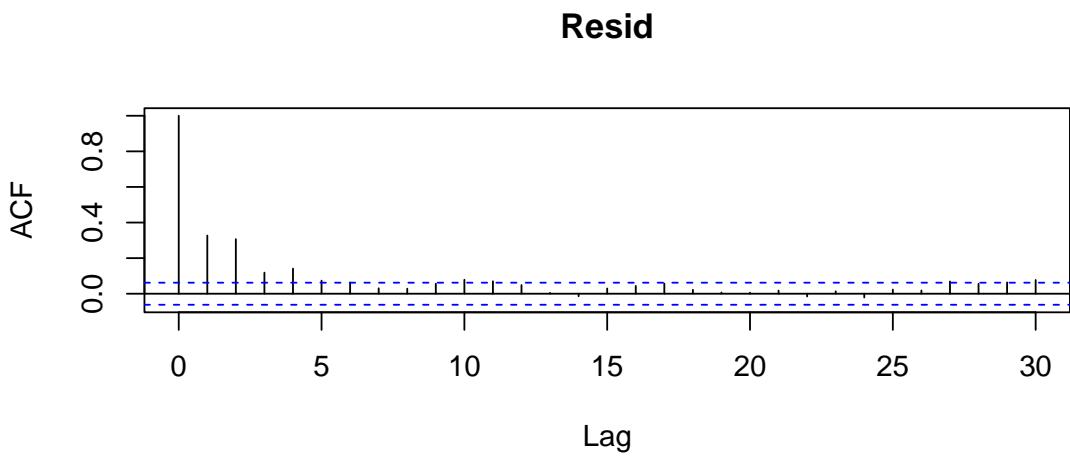


Figura 2.23: ACF de los residuos de la relación lineal entre variables simuladas para VECM.

La serie temporal de los residuos oscila en torno al cero, y tiene un comportamiento que parece estacionario. La autocorrelación simple muestra que el retardo 1 y 2 sobrepasan las bandas de confianza teniendo el retardo 1 una autocorrelación simple muestral cercana a 0.4. También algunos retardos en torno al retardo 10 sobrepasan las bandas de confianza, aún así no hay una estructura marcada y excepto el retardo 1 y 2, el resto de retardos que sobrepasan las bandas de confianza lo hacen por poco. Vamos a realizar el test de Dickey-Fuller aumentado para salir de dudas y ver si existe cointegración o no.

	Augmented DF	P-Value
coint_resid	-4.914	< 0.01

El test rechaza la hipótesis nula de que los residuos no sean estacionarios, por lo que al ser el modelo lineal globalmente significativo y los residuos estacionarios, concluimos que

las series están cointegradas y que se puede estimar un modelo de corrección de equilibrio entre las series.

- Análisis del conjunto de datos simulado para modelos de corrección de equilibrio con 100 observaciones:

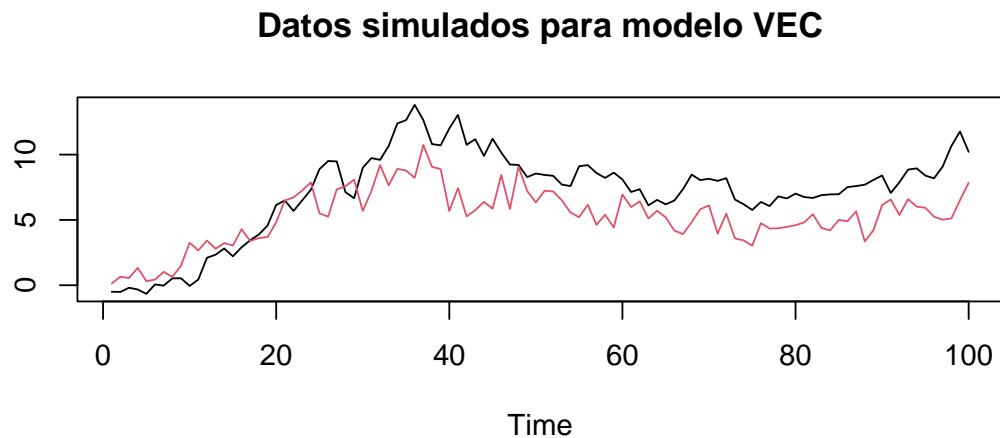


Figura 2.24: Datos simulados para VECM.

Vamos a comprobar que ambas series son $I(1)$.

La serie X_1 presenta las siguientes características recogidas en su evolución temporal y el correlograma:

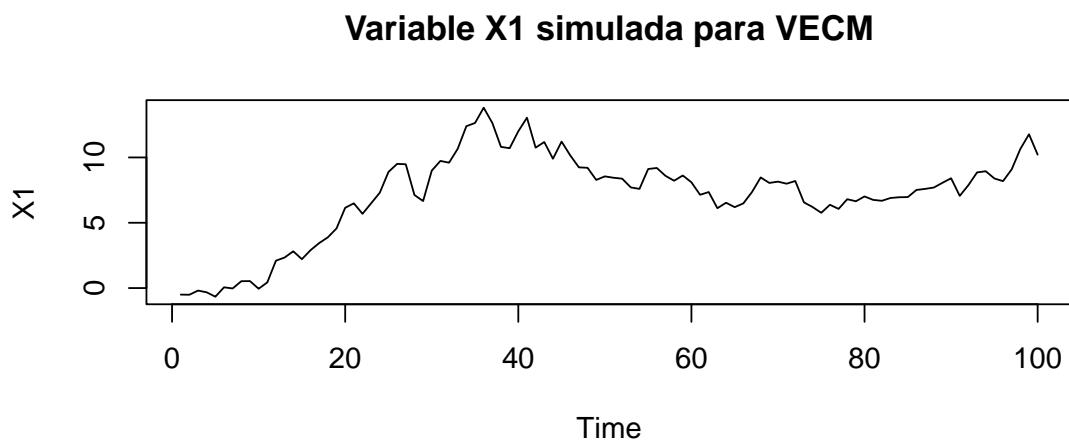


Figura 2.25: Variable X1 para VECM.

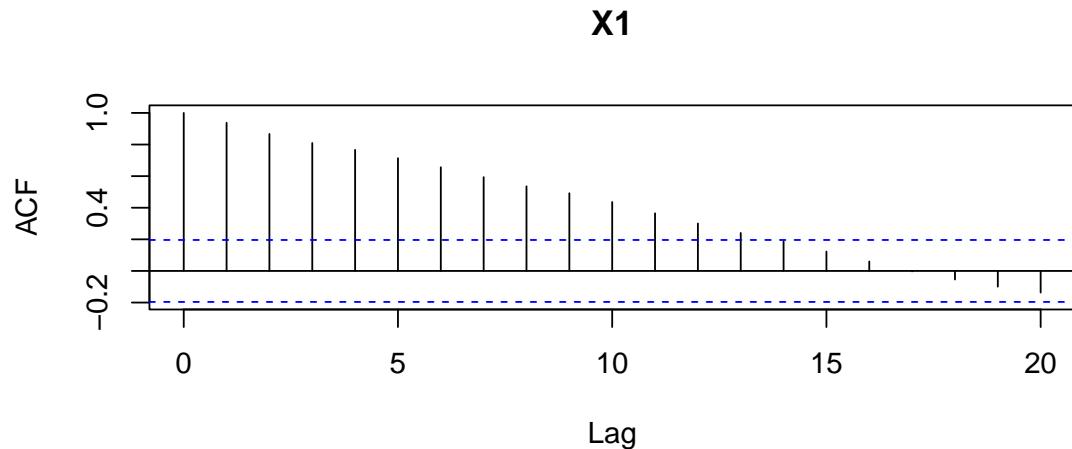


Figura 2.26: Autocorrelación simple de X1 para VECM.

Vemos que la varianza es constante a lo largo de la misma, pero no hay estacionariedad en la parte regular pues la serie presenta una tendencia positiva marcada. Además, si miramos la autocorrelación simple, se observa la típica estructura decreciente característica de series no estacionarias. En cuanto a estacionalidad, la serie no es estacionaria en la parte estacional como se puede observar en el gráfico de la serie.

Vamos a contrastar la hipótesis de no estacionariedad.

	Augmented DF	P-Value
X1_vecm100	-2.1256	0.5246

El test nos indica que la serie presenta raíces unitarias y que es necesario aplicar diferencias. Vamos a aplicar una primera diferencia a la serie X_1 simulada, y, volveremos a contrastar si tiene raíces unitarias.

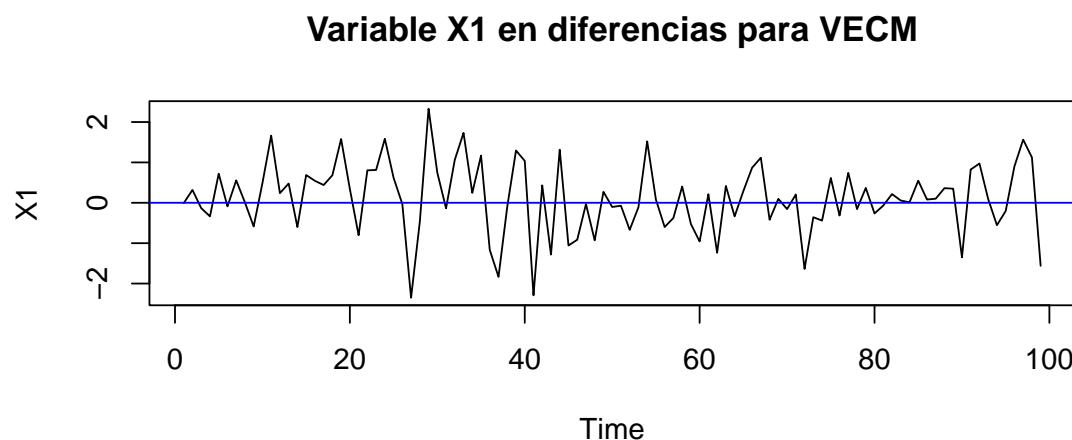


Figura 2.27: Variable X1 en diferencias para VECM.

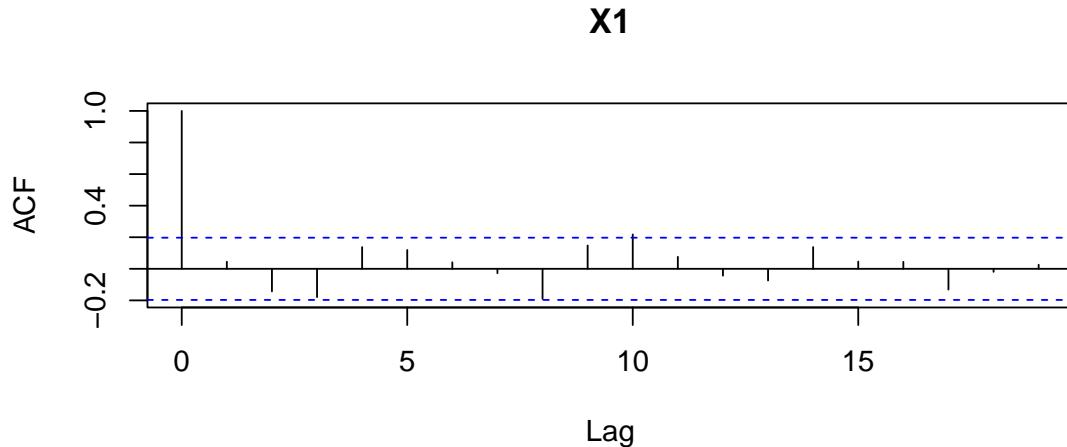


Figura 2.28: Autocorrelación simple de X_1 en diferencias para VECM.

Vemos que al aplicar una primera diferencia regular la serie toma valores en torno a una media (0) y no presenta ni tendencia, crecimiento sistemático, o comportamientos de paseo aleatorio. Además, la estructura de la autocorrelación simple se ha eliminado y ahora no hay retardos que sobrepasen las bandas de confianza. Por último, vamos a constatar que la serie no necesita otra diferencia regular mediante el test de Dickey-Fuller aumentado.

	Augmented DF	P-Value
diff_X1_vecm100	-4.0675	< 0.01

Se rechaza la hipótesis nula de que la serie no es estacionaria por lo que podemos concluir que la serie original X_1 es $I(1)$.

Ahora, vamos a repetir este proceso con la serie X_2 . Comenzaremos con el análisis descriptivo de la serie original.

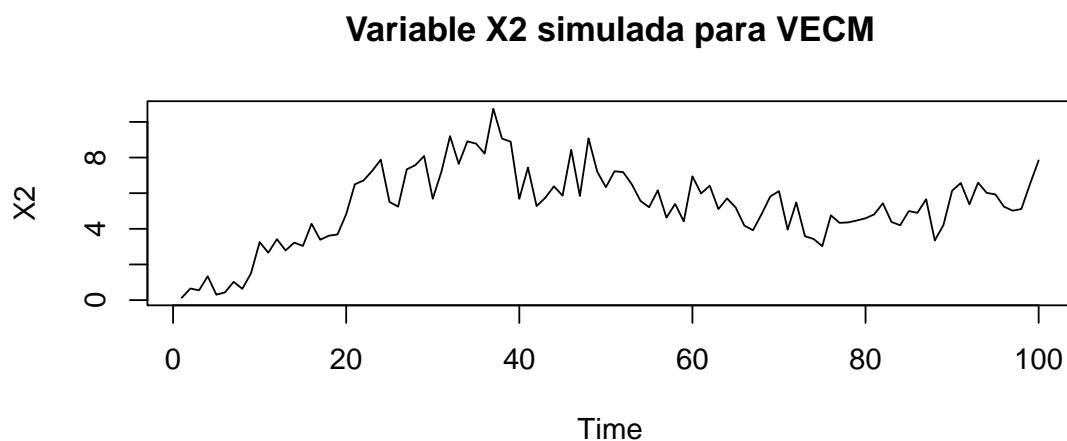


Figura 2.29: Variable X_2 para VECM.

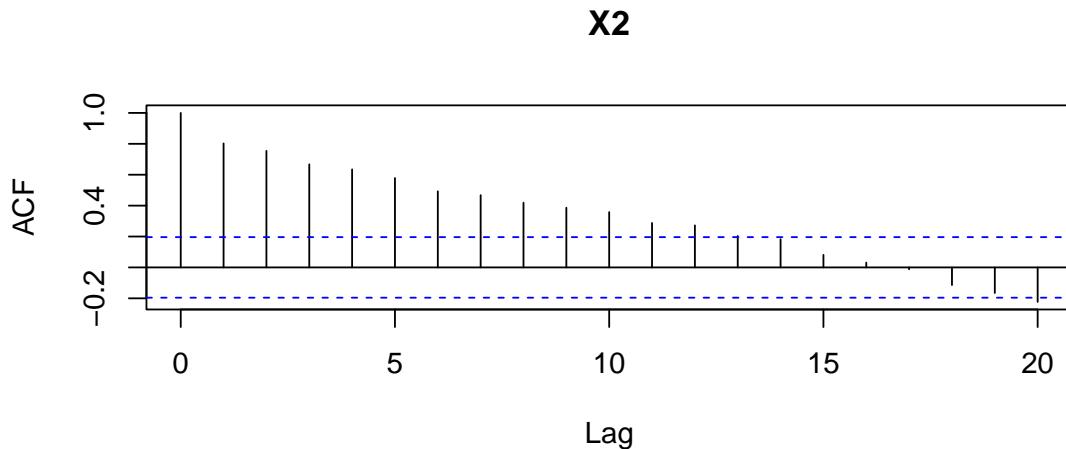


Figura 2.30: Autocorrelación simple de X2 para VECM.

Igual que con X_1 , la serie presenta una tendencia positiva, no oscila en torno a una media, y su autocorrelación simple presenta una estructura decreciente muy marcada con varios retardos significativos. En cuanto a la estacionalidad, la serie no presenta comportamiento estacional. A continuación, usamos otra vez el test de Dickey-Fuller aumentado para contrastar que la serie necesita una diferencia regular.

	Augmented DF	P-Value
X2_vecm100	-2.3672	0.4245

Se acepta la hipótesis de que la serie no es estacionaria por lo que necesitamos aplicar diferencias. Vamos a diferenciar la serie y ver si necesita otra diferencia o si con una sirve.

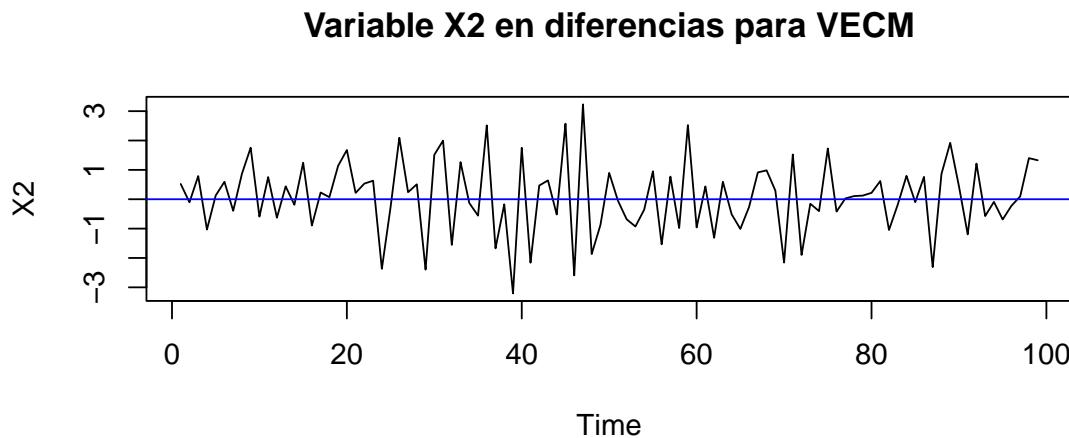


Figura 2.31: Variable X2 en diferencias para VECM.

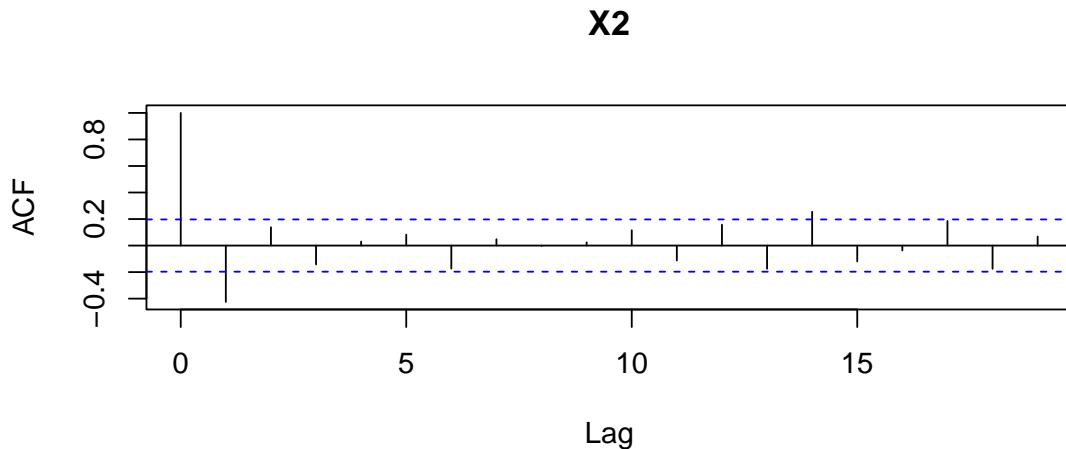


Figura 2.32: Autocorrelación simple de X2 en diferencias para VECM.

Tanto el gráfico de la serie diferenciada como la autocorrelación simple indican que la serie es estacionaria y que no hace falta otra diferencia en la parte regular. La serie oscila en torno al 0, y se ha eliminado la estructura de la autocorrelación simple. Vamos a contrastarlo con el test de Dickey-Fuller aumentado.

	Augmented DF	P-Value
diff_X2_vecm100	-4.7642	< 0.01

Rechazamos H_0 con lo que tenemos que $X2$ es $I(1)$ igual que $X1$. El siguiente paso es crear un modelo que relacione las variables en el largo plazo del tipo $Y_t = c + \beta X_t + u_t$ por mínimos cuadrados ordinarios.

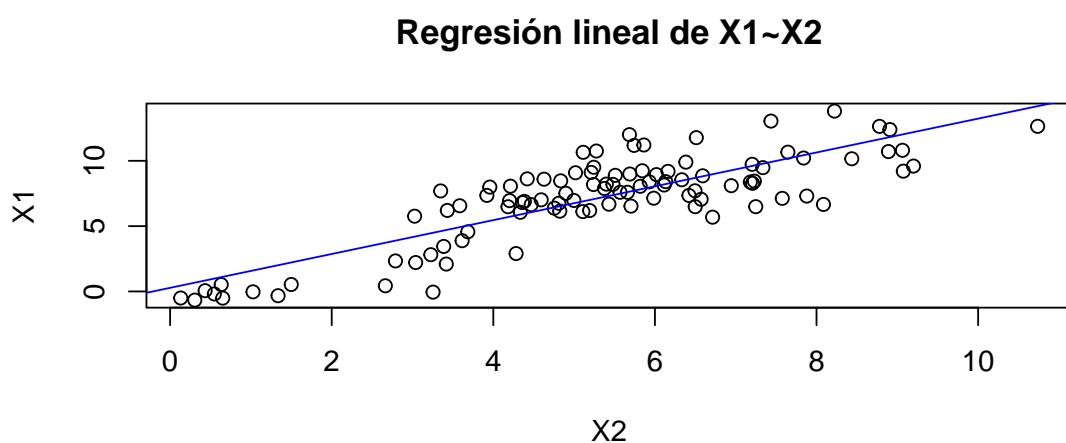


Figura 2.33: Relación lineal entre X1 y X2 simuladas para VECM 100 observaciones.

La relación de largo plazo es positiva y los valores de los estadísticos e indicadores de ajuste se recogen en el siguiente summary:

	Estimate	Std. Error	t.value	p-value
Intercept	0.27714	0.50245	0.552	0.582
X2	1.29552	0.08815	14.696	<2e-16

El estimador de β_1 es significativo con valor 1.472, muy cercano al 1.5 real, el R^2 es de 0.9774, y el estadístico F es significativo por lo que el modelo es globalmente significativo. Para contrastar si existe cointegración entre las series hace falta que los residuos de este modelo sean estacionarios. La serie temporal de los series presenta las siguientes características.

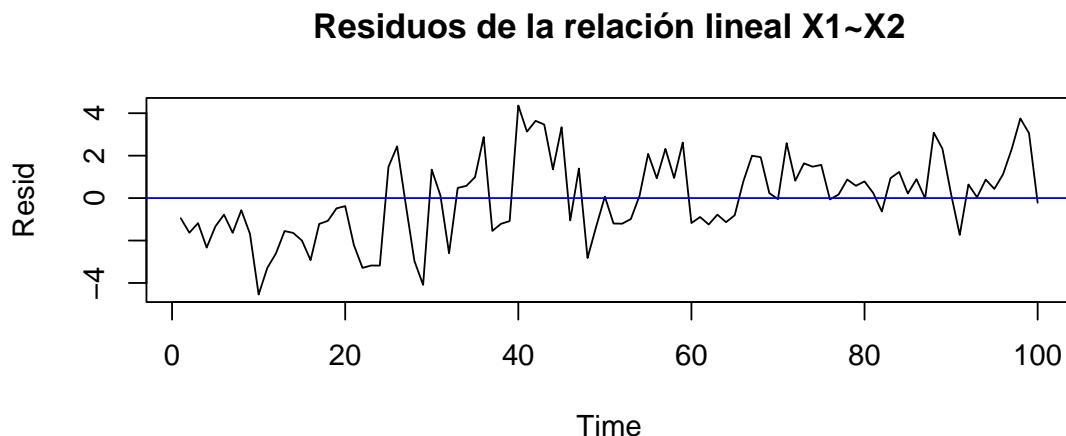


Figura 2.34: Residuos de la relación lineal entre variables simuladas para VECM.

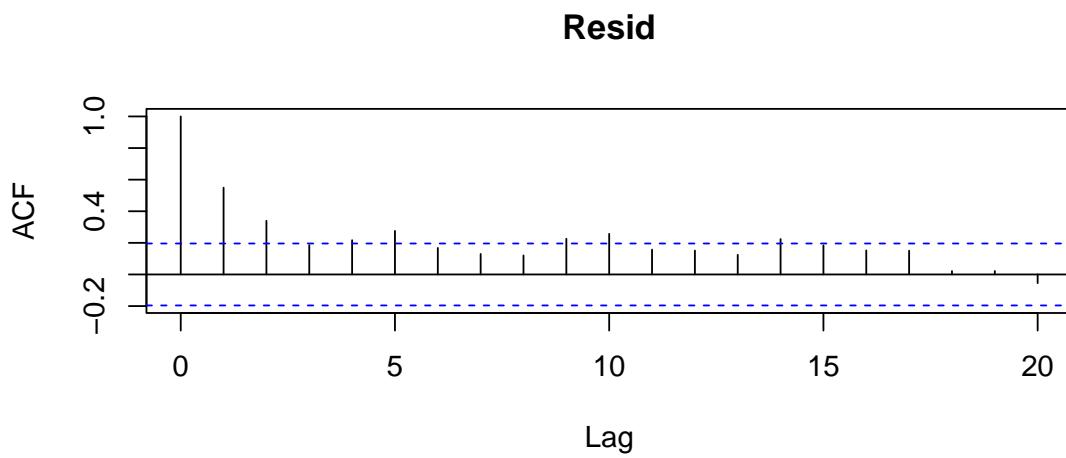


Figura 2.35: ACF de los residuos de la relación lineal entre variables simuladas para VECM.

La serie temporal de los residuos oscila en torno al cero, y tiene un comportamiento que parece estacionario. La autocorrelación simple muestra que el retardo 1 y 2 sobrepasan las

bandas de confianza teniendo el retardo 1 una autocorrelación simple muestral cercana a 0.5. También algunos retardos en torno al retardo 10 sobrepasan las bandas de confianza, aún así no hay una estructura marcada y excepto el retardo 1 y 2, el resto de retardos que sobrepasan las bandas de confianza lo hacen por poco. Vamos a realizar el test de Dickey-Fuller aumentado para salir de dudas y ver si existe cointegración o no.

	Augmented DF	P-Value
coint_resid100	-3.3996	0.05855

El test rechaza la hipótesis nula de que los residuos no sean estacionarios, por lo que al ser el modelo lineal globalmente significativo y los residuos estacionarios, concluimos que las series están cointegradas y que se puede estimar un modelo de corrección de equilibrio entre las series.

2.2.2. Datos reales

A parte de los datos simulados, se van a utilizar dos datasets que contienen datos reales. El primero, recoge variables macroeconómicas de Estados Unidos desde mitad del siglo XX hasta principios del siglo XXI. El segundo, recoge datos sobre los swaps a 9 meses y el LIBOR a 12 meses en la eurozona.

Dentro de la librería `bvarools` hay un dataset llamado `us_macrodata` que recoge datos cuatrimestrales desde 1959 hasta 2007 sobre la tasa de inflación, la tasa de desempleo, y la tasa de fondos de la reserva federal estadounidense que viene a ser como la tasa de interés a la cual el BCE presta dinero pero para la FED. El dataset original se puede encontrar en https://web.ics.purdue.edu/~jltobias/second_edition/Chapter20/code_for_exercise_1/US_macrodata.csv.

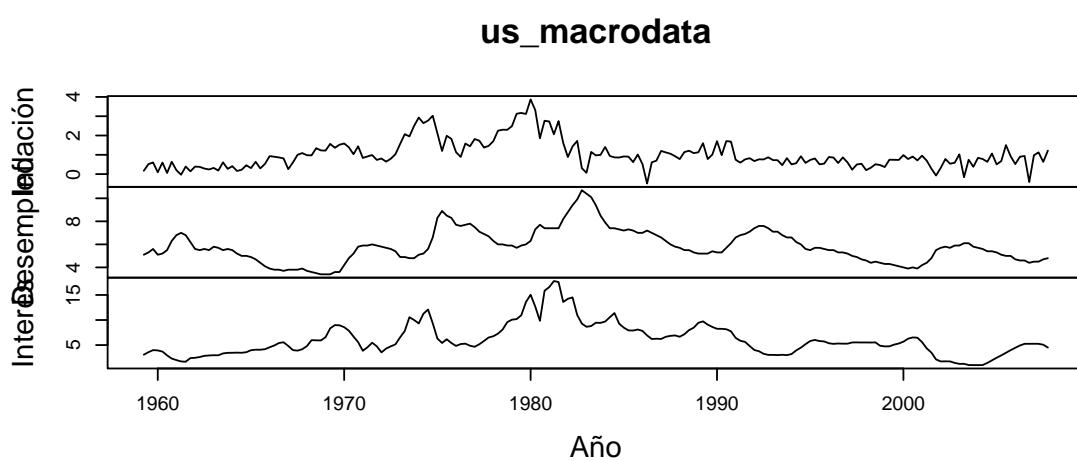


Figura 2.36: Datos macro USA 1959-2007.

Las variables macroeconómicas están muy relacionadas y destaca la relación entre inflación y desempleo. En el corto plazo, cuando el desempleo baja, la inflación sube y

viceversa, esta relación se recoge en la *Curva de Phillips*, aunque, cabe destacar que esta relación está condicionada por el ciclo económico y en el largo plazo no suele mantenerse. Si miramos el gráfico anterior que muestra las tres series temporales que recoge el conjunto de datos, vemos que en la mayoría de los casos, cuando la serie temporal de inflación sube, el desempleo se reduce, y al revés cuando la inflación baja.

En cambio, en cuanto a la tasa de interés y la inflación, la tasa de interés es un instrumento de política monetaria que se utiliza para controlar la inflación. Los bancos centrales tienen unas tasas de inflación que estiman como *ideales* en torno a las cuales tienen que fluctuar éstas. Cuando la inflación es elevada, un mecanismo de corrección suele ser subir los tipos de interés, ya que si los tipos de interés son más caros, cuesta más dinero financiarse por lo que tanto las empresas como los particulares pedirán menos dinero prestado, y, por lo tanto, consumirán menos reduciendo la demanda de los productos y consigo su precio. Por esto, la inflación y la tasa de interés están muy ligadas pero existe un pequeño desfase temporal entre ellas debido a la naturaleza reactiva de las tasas de interés, es decir, estas suelen moverse según los movimientos de las otras. Mirando las series de la gráfica anterior, se puede apreciar cómo el comportamiento de estas series es muy parecido y la serie de tasas de interés está algo desplazada a la derecha respecto de la inflación.

Por otro lado, las series de interés y desempleo parecen seguir un comportamiento parecido. En general, cuando las tasas de desempleo han sido bajas, los tipos de interés se han mantenido altos ya que una tasa de desempleo baja significa una economía sana por lo que bajar los tipos de interés sólo haría que subiese la inflación. Por esto, si observamos las series de desempleo e interés desde el inicio hasta los años 80, cuando bajaba el desempleo se aumentaban los tipos, y viceversa. Entre el año 80 y el 87 aproximadamente se rompe esta relación y las series se comportan de manera similar disminuyendo en este periodo. Ya a finales de los 80 hasta el final de la serie, la relación habitual en la que una bajada del desempleo supone un aumento de los tipos y viceversa parece que se restaura.

De las tres series, se van a utilizar para los modelos la serie de desempleo y la inflación que serán utilizadas para modelos VAR ya que sabemos que por teoría estas dos variables no presentan relación en el largo plazo. De todas formas, se realizarán los contrastes pertinentes para contrastar esto.

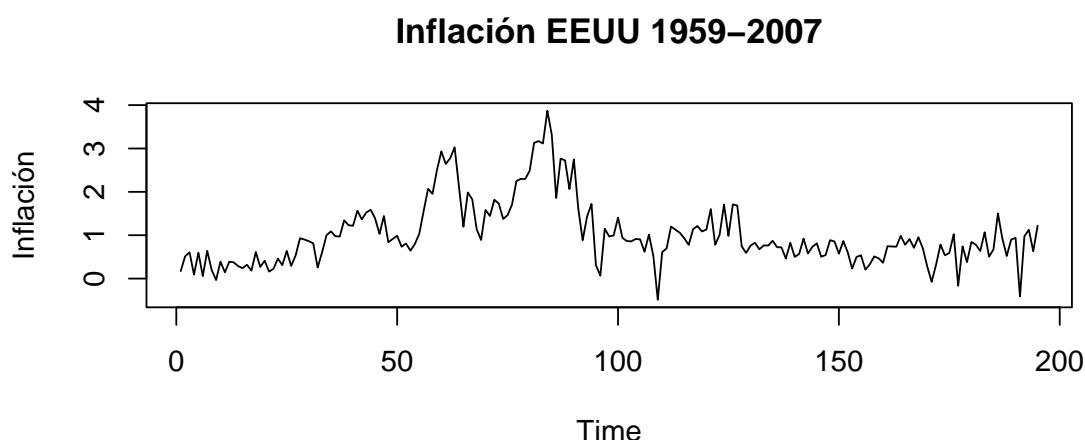


Figura 2.37: Inflación USA 1959-2007.

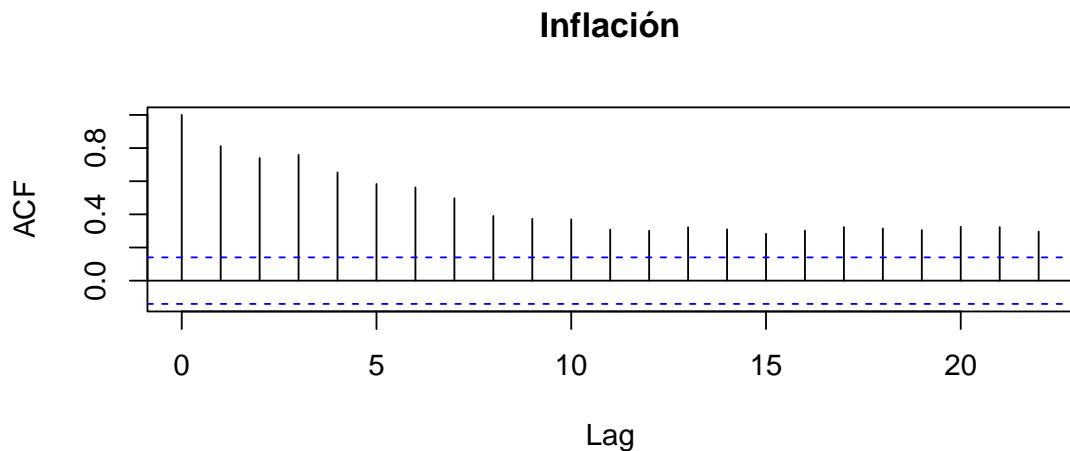


Figura 2.38: ACF inflación.

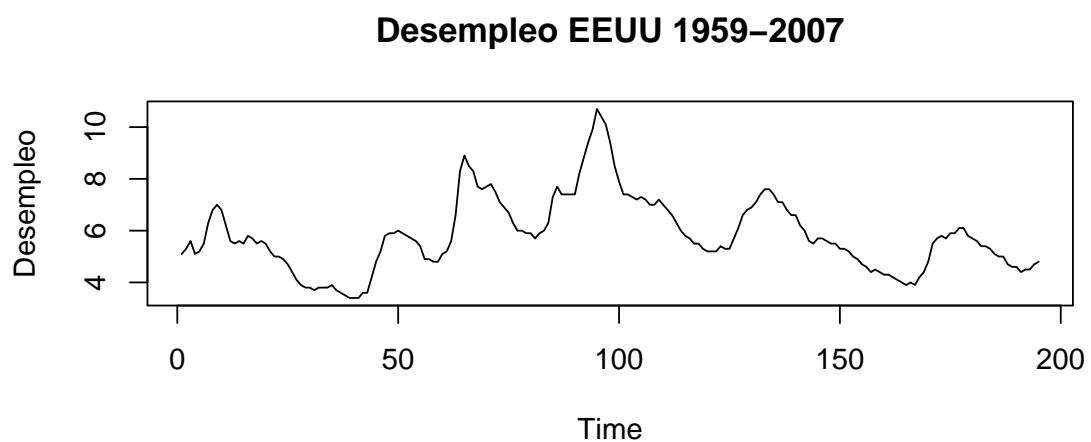


Figura 2.39: Desempleo USA 1959-2007.

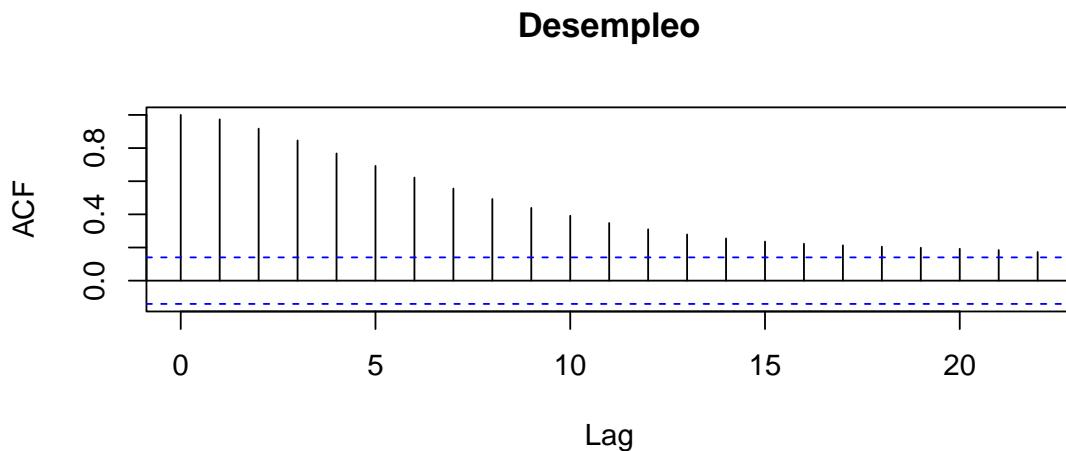


Figura 2.40: ACF desempleo

	Augmented DF	P-Value
Inflacion	-2.8292	0.229

	Augmented DF	P-Value
Desempleo	-2.6612	0.2994

La estructura de los correlogramas indica que las series no son estacionarias y el test de Dickey-Fuller sugiere que al 5 % de significación, ambas series presentan raíces unitarias. Aplicamos una primera diferencia para ver si las series son $I(1)$.

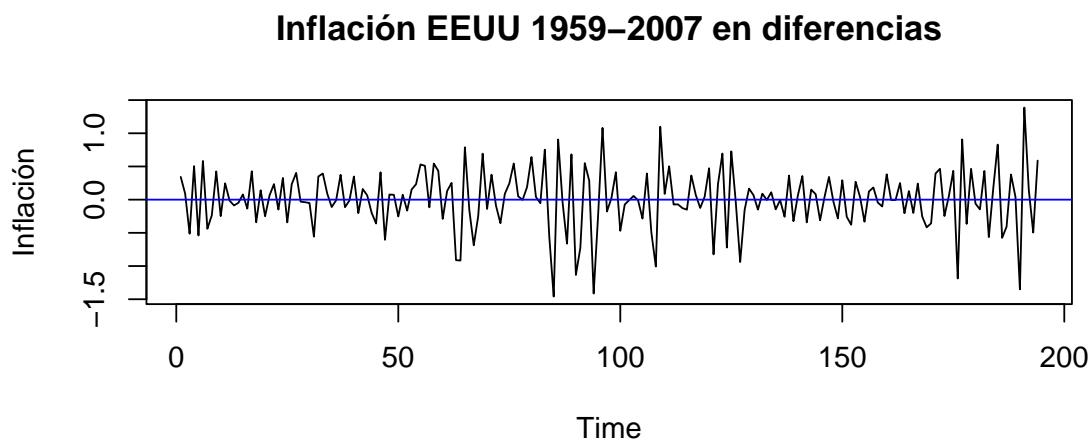


Figura 2.41: Primera diferencia de la Inflación en USA 1959-2007.

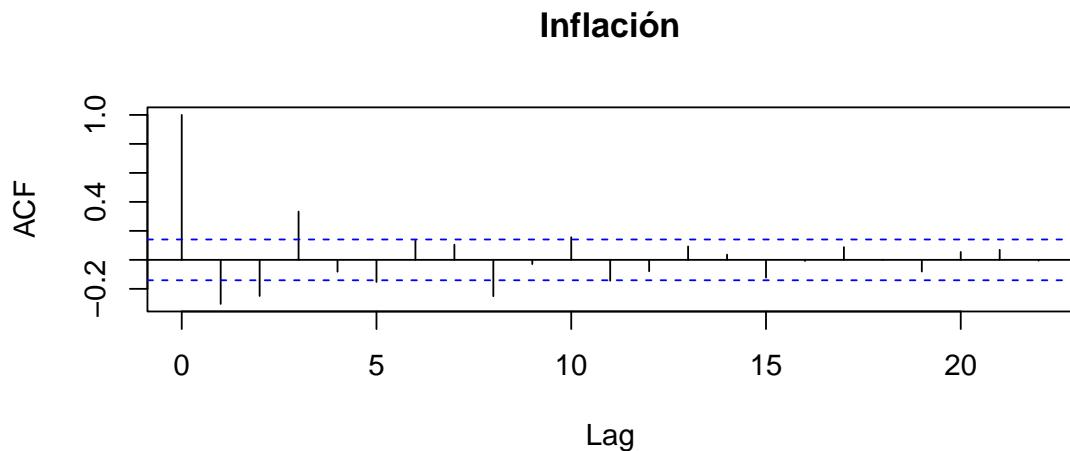


Figura 2.42: ACF de la primera diferencia Inflación.

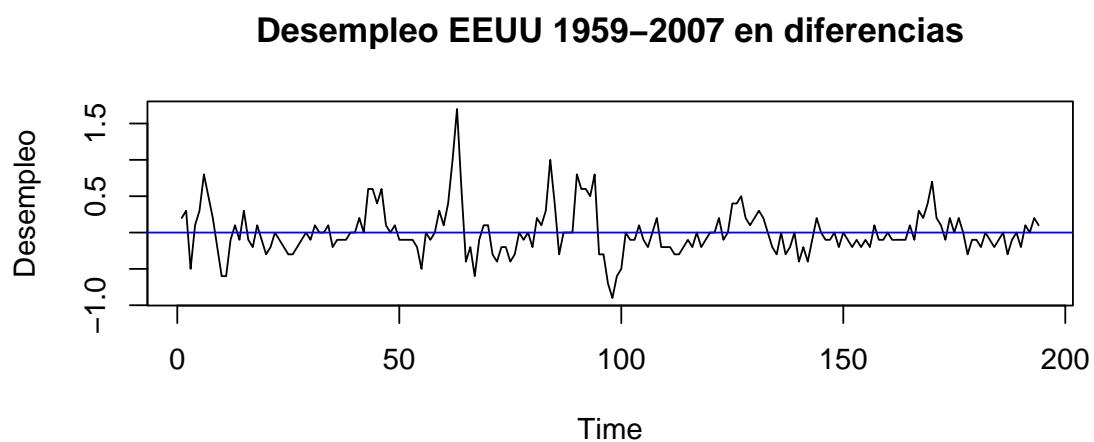


Figura 2.43: Primera diferencia del desempleo USA 1959-2007

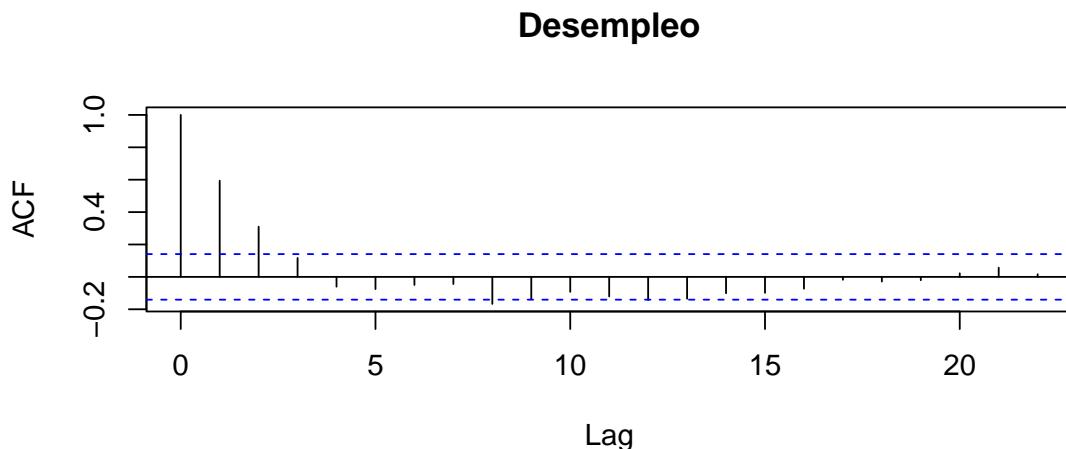


Figura 2.44: ACF de la primera diferencia desempleo

	Augmented DF	P-Value
diff_Inflacion	-6.0731	< 0.01

	Augmented DF	P-Value
diff_Desempleo	-5.1733	< 0.01

Todos los datos anteriores sugieren que las series son $I(1)$ por lo que se cumple la primera condición para determinar si están cointegradas o no. Ahora vamos a relacionar las series por un modelo lineal mediante mínimos cuadrados ordinarios, y, analizaremos los residuos para ver si son estacionarios. En caso afirmativo, las series estarán cointegradas.

Regresión lineal de Inflación~Desempleo

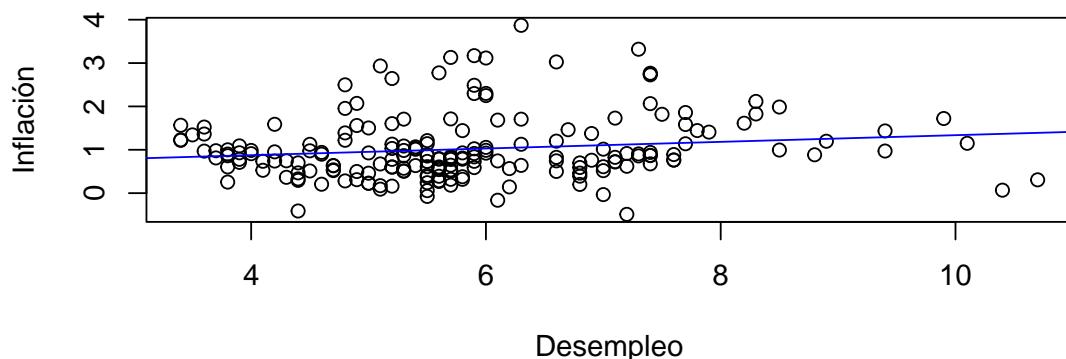


Figura 2.45: Relación lineal entre desempleo e inflación.

	Estimate	Std. Error	t.value	p-value
Intercept	0.5683	0.2212	2.570	0.0109
Desempleo	0.0768	0.0368	2.087	0.0382

El coeficiente $\beta_{desempleo}$ es significativo, positivo igual a 0.0768018. El modelo es globalmente significativo aunque el R^2 es prácticamente 0. Vamos a ver si los residuos son estacionarios.

Residuos de la relación lineal Inflación~Desempleo

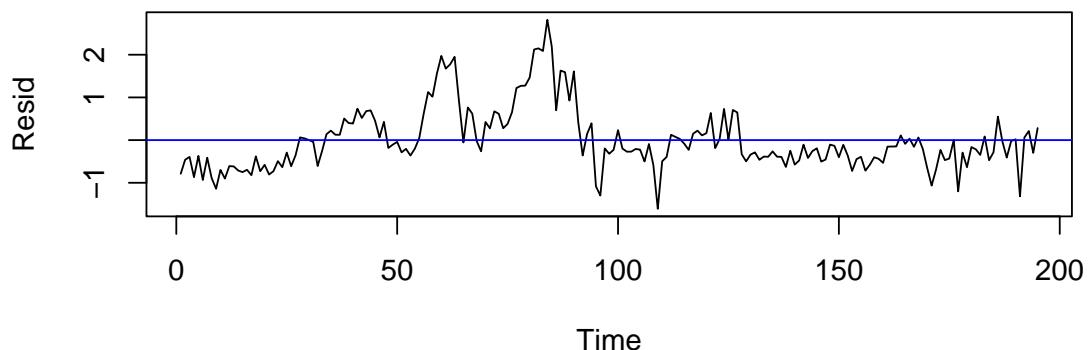


Figura 2.46: Residuos de la relación lineal entre inflación y desempleo.

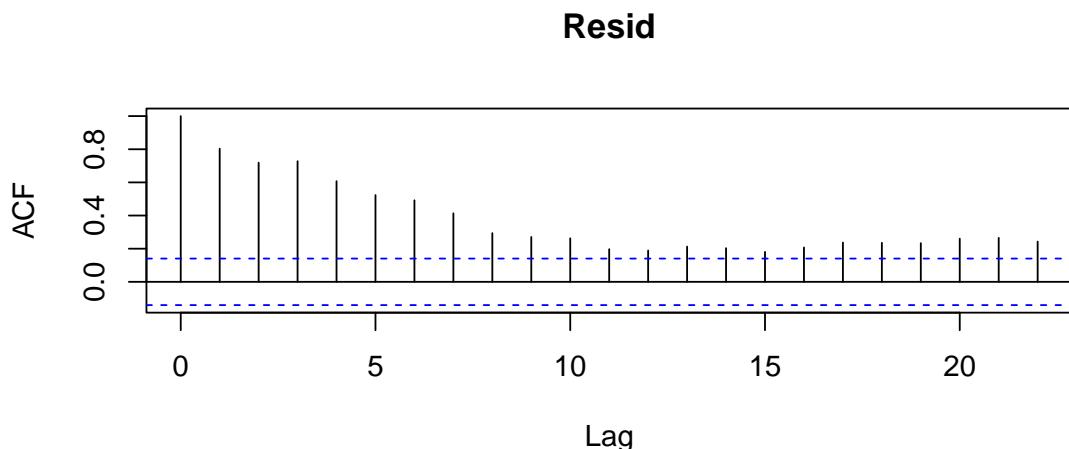


Figura 2.47: ACF de los residuos de la relación lineal entre inflación y desempleo.

	Augmented DF	P-Value
coint_Inflacion	-2.4655	0.3813

El test de Dickey-Fuller indica que los residuos no son estacionarios, esto sugiere que no se puede aplicar un modelo de corrección de equilibrio a las series ya que no hay una

relación de largo plazo subyacente. Por esto, estas series se usarán para los modelos VAR frecuentista y bayesiano.

El otro conjunto de datos que se va a analizar recoge tipo de interés OVERNIGHT (Swaps) a 9 meses y el LIBOR (interbancario) a 12 meses en la euro zona, observados diariamente desde 10 de marzo de 1999 hasta el 21 de mayo de 2007. La variable *OVERNIGHT* recoge los swaps a 9 meses sobre los tipos de interés en Inglaterra. Esto es, unos contratos entre dos partes en los que se acuerda intercambiar un activo; en este caso dinero invertido en unos depósitos a tipo de interés fijo, en una fecha concreta. Por otro lado, la variable *LIBOR* recoge los tipos de interés interbancarios en el mercado de Londres (London InterBank Offered Rate) a 12 meses. Estas variables se sabe que están muy relacionadas ya que los swaps actúan como futuros, o, especulaciones de los valores que tomará el LIBOR.

Vamos a analizar si las series están cointegradas, y, en caso afirmativo, se usarán para los modelos de corrección de equilibrio.

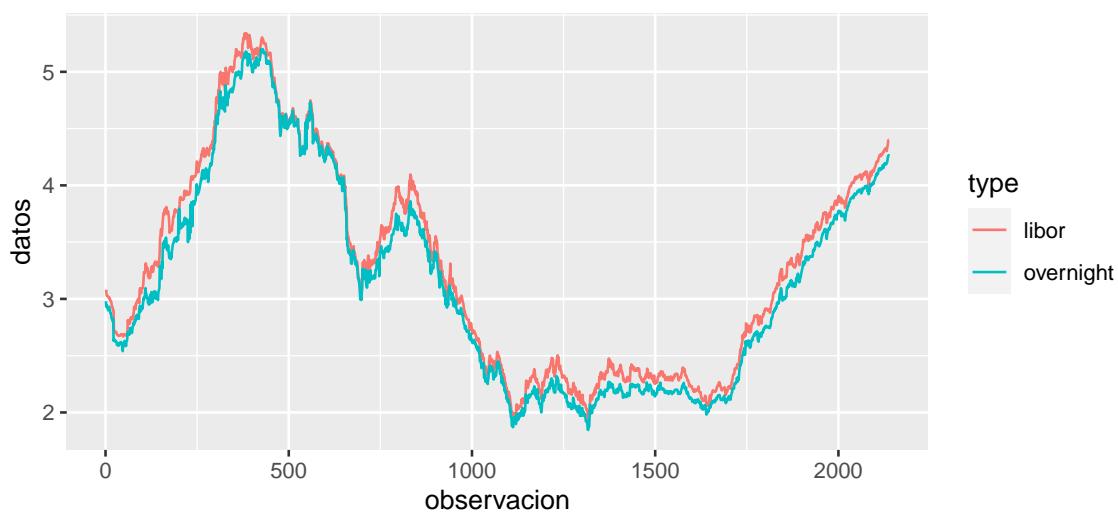


Figura 2.48: Serie de LIBOR y OVERNIGHT 1999-2007.

Se puede apreciar que ambas series no son estacionarias en la parte regular ya que no presentan una media constante a lo largo del tiempo. También, destaca la ausencia de estacionalidad tanto trimestral como mensual. Otro aspecto que salta a la vista es el comportamiento tan similar que presentan las series; esto podría indicar que están cointegradas, pero para esto hay que estimar un modelo de largo plazo entre las dos series y ver si los residuos de dicho modelo son estacionarios.

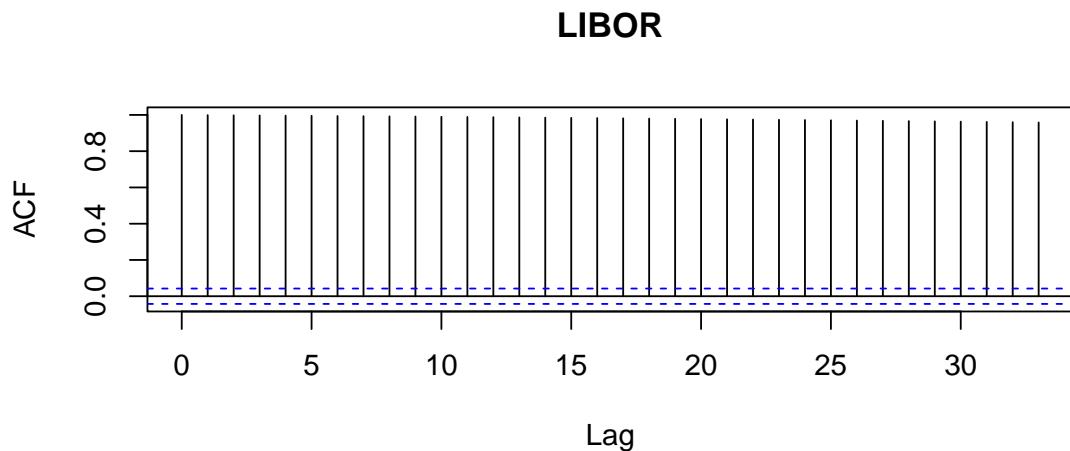


Figura 2.49: ACF de LIBOR.

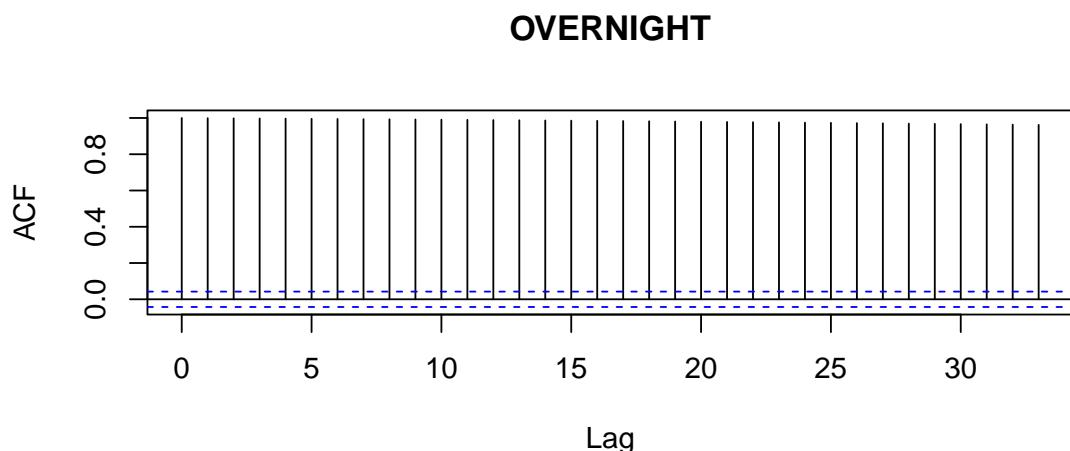


Figura 2.50: ACF de OVERNIGHT

	Augmented DF	P-Value
LIBOR	-0.40051	0.9863

	Augmented DF	P-Value
OVERNIGHT	-0.24908	0.99

Las series diferenciadas presentan las siguientes características:

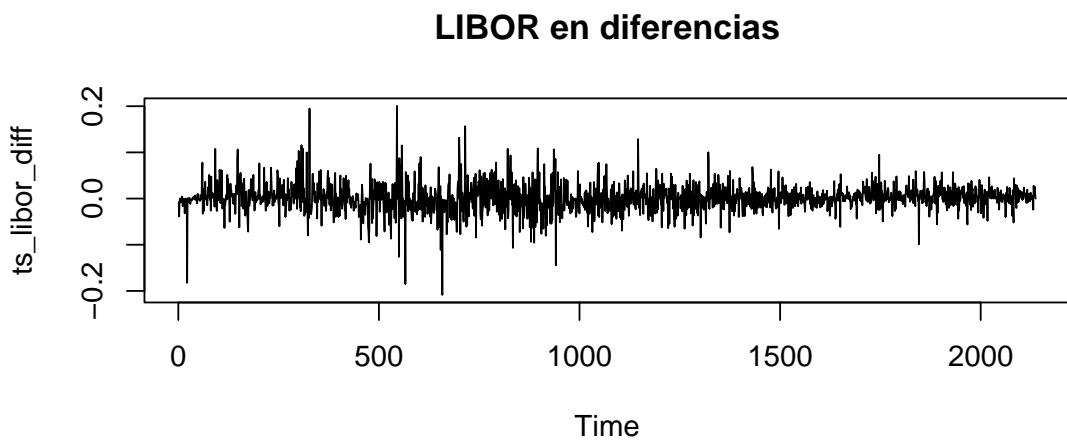


Figura 2.51: Serie de LIBOR con primera diferencia.

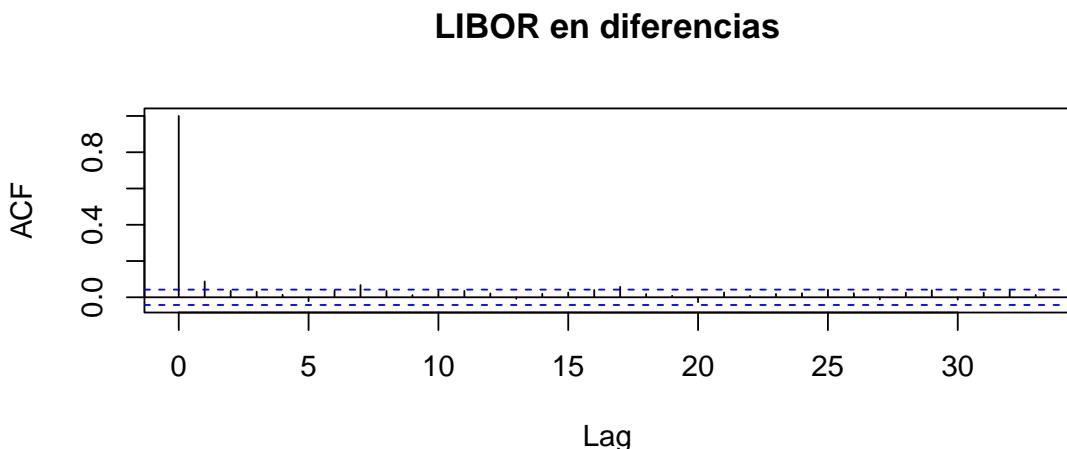


Figura 2.52: ACF de LIBOR con primera diferencia.

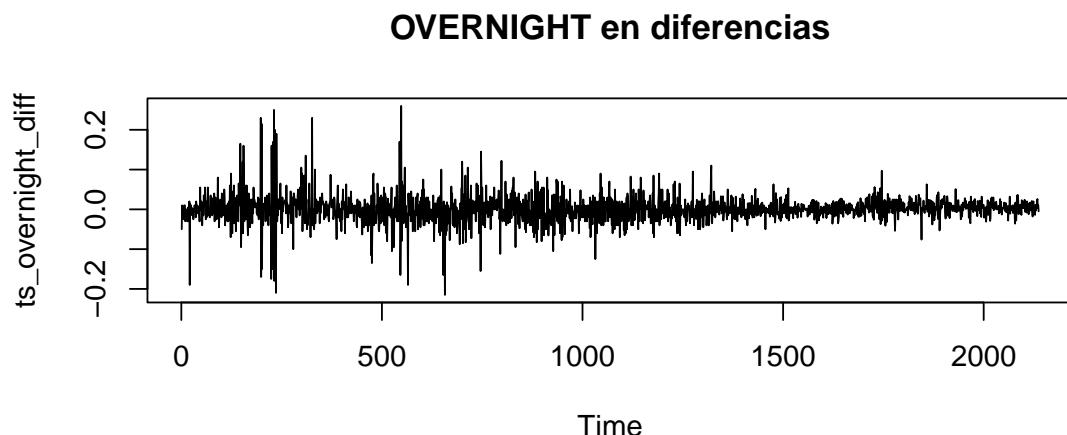


Figura 2.53: Serie de OVERNIGHT con primera diferencia.

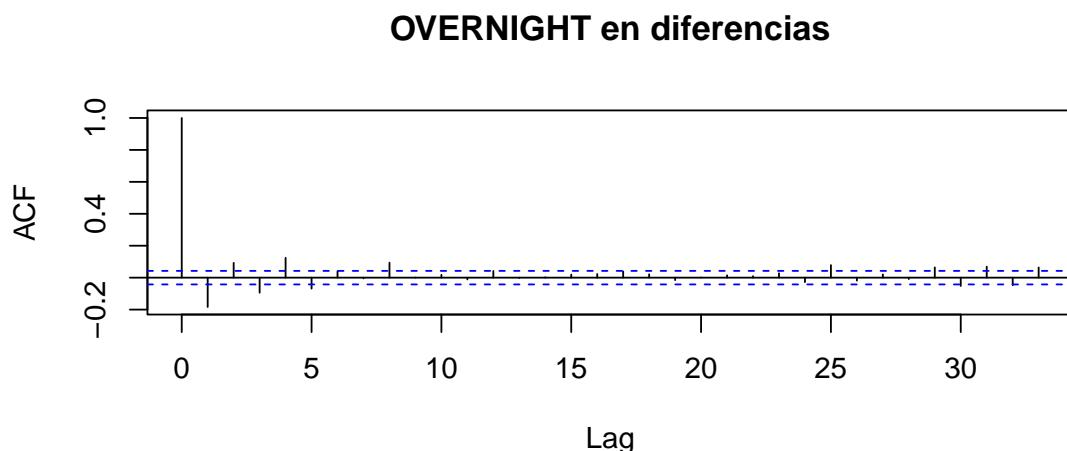


Figura 2.54: ACF de OVERNIGHT con primera diferencia.

	Augmented DF	P-Value
diff_LIBOR	-11.236	< 0.01

	Augmented DF	P-Value
diff_OVERNIGHT	-11.008	< 0.01

El test de Dickey-Fuller para ambas series indica que existen raíces unitarias, y el test para las series diferenciadas nos dice que ambas son $I(1)$ que es la primera condición que se tiene que cumplir para llegar a la conclusión de que están cointegradas. Falta ver si los residuos de la relación de largo plazo son estacionarios.

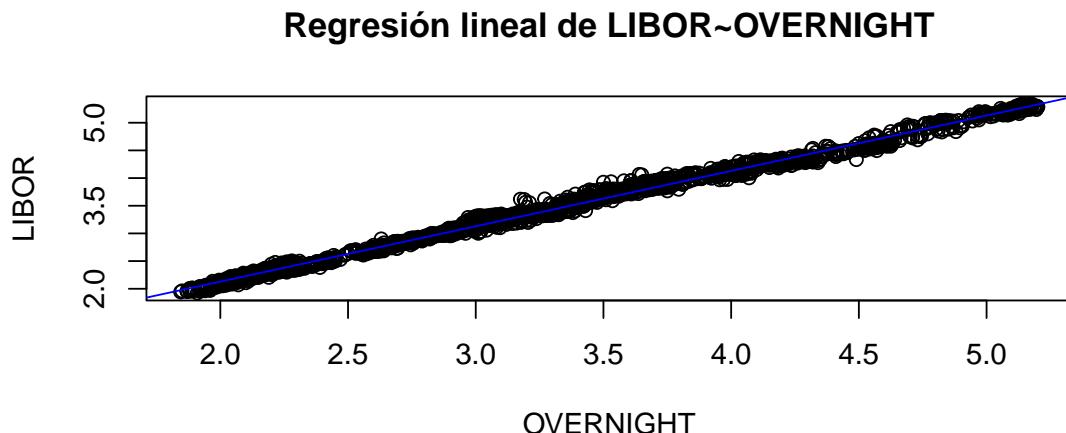


Figura 2.55: Relación lineal entre LIBOR y OVERNIGHT.

	Estimate	Std. Error	t.value	p-value
Intercept	0.128911	0.004961	25.98	<2.e-16
OVERNIGHT	1.000089	0.001508	663.19	<2.e-16

La tabla anterior recoge los datos de la relación estimada y muestra que tanto el contraste de significación global y el individual son significativos siendo la estimación de $\beta_{overnight}$ 1.0000892. La relación tiene un R^2 altísimo del 99.52 %. Cuando hemos introducido el concepto de relaciones espúrias comentábamos que cuando se daba esta situación los estadísticos y métricas de ajuste eran muy buenas pero que luego a la hora de analizar la estacionariedad de los residuos estos no resultaban ser estacionarios. Por ello, el último paso para determinar la cointegración es analizar si los residuos son estacionarios.

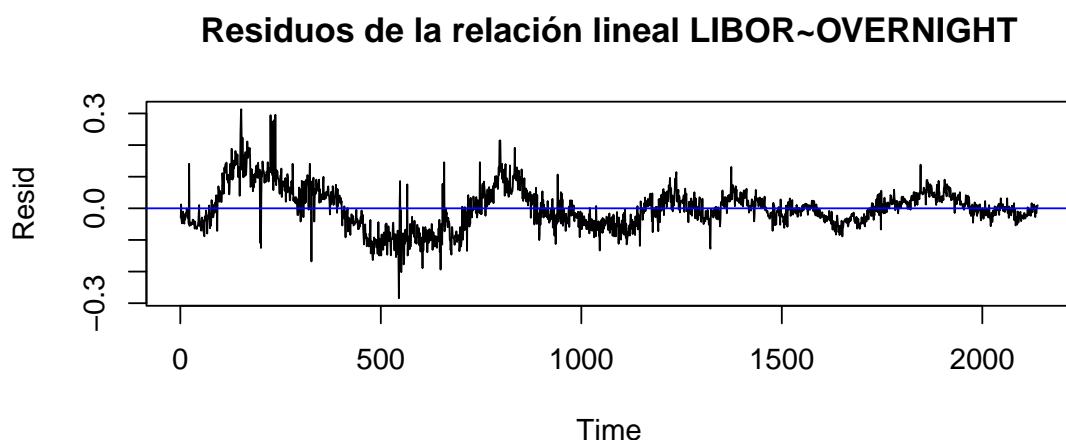


Figura 2.56: Residuos de la relación lineal entre LIBOR y OVERNIGHT.

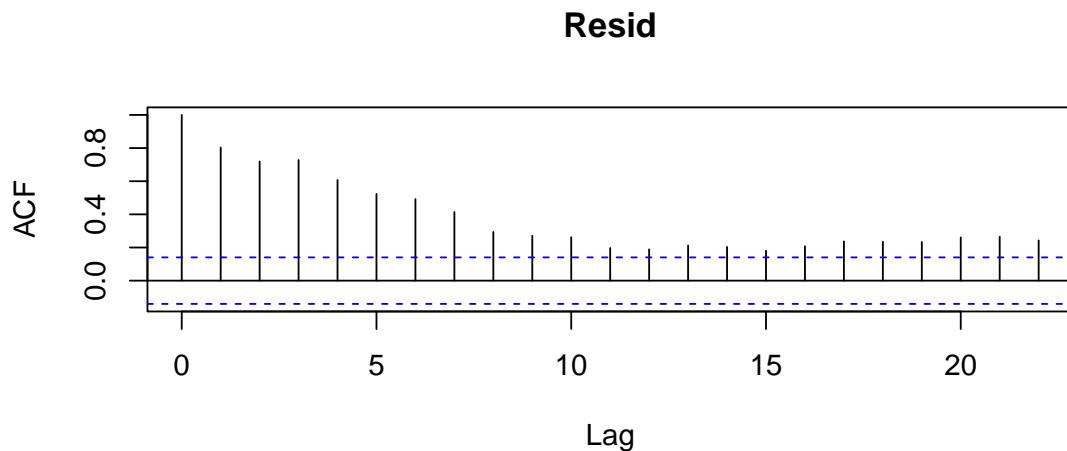


Figura 2.57: ACF de residuos de la relación lineal entre LIBOR y OVERNIGHT.

	Augmented DF	P-Value
coint_LIBOROVER	-3.5066	0.04165

La serie temporal de los residuos muestra que oscilan alrededor de un valor constante con una estructura de dependencia sin comportamiento aleatorio. Esto es irrelevante ya que no es necesario que los residuos sean ruido blanco sino que sean estacionarios. El test de Dickey-Fuller aumentado da la información necesaria para concluir que las series están cointegradas por lo que existe una relación de largo plazo entre ellas y se pueden estimar modelos de corrección de equilibrio.

Capítulo 3

Resultados

3.1. Estimación de los modelos VAR y BVAR

Para estudiar las diferencias en la estimación de modelos autorregresivos de manera frecuentista y bayesiana, se van a utilizar tres conjuntos de datos: El primero, recoge datos simulados con la estructura definida en 2.22 en el capítulo 2 con 1000 observaciones en cada serie, el segundo conjunto de datos también son simulados y siguen la misma estructura que el conjunto anterior solo que con 100 observaciones, y el tercero, es el conjunto de datos macroeconómicos de Estados Unidos entre 1959 y 2007 que recoge la inflación y el desempleo durante estos años para cada cuatrimestre.

Con los datos simulados sabemos los parámetros *reales* por lo que podemos comparar directamente las estimaciones de cada método con estos valores. Con los datos reales macroeconómicos de Estados Unidos, no sabemos el valor real de los parámetros, ni siquiera se sabe si la relación real puede seguir un modelo autorregresivo ni de qué orden exactamente. Por esto, para la estimación de los modelos en estos datos, se va a utilizar la metodología Box-Jenkins partiendo de modelos sencillos e incorporando diferentes retardos según lo que indiquen los residuos del modelo, ya que como sabemos, han de tener estructura de ruido blanco para que se acepte el modelo. Además, se usará el AIC para comparar los modelos. Se compararán las estimaciones obtenidas con el método frecuentista y el bayesiano para ver las diferencias en los valores de los parámetros estimados y sus desviaciones estándar.

Primero, se van a estimar los modelos sobre el conjunto de datos simulados. Empezaremos estimando el modelo VAR y después se estimará el modelo BVAR para el conjunto con 1000 observaciones.

Como se ha visto en el análisis de los datos en el capítulo 2, las series simuladas son estacionarias tanto en varianza como en la parte regular por lo que no hace falta aplicar ninguna transformación a los datos para trabajar con ellos. Para estimar el modelo se va a usar la función **VAR()** de la librería **vars**. Como sabemos el número *real* de retardos de la relación entre las variables, vamos a especificar este valor a la hora de estimar la relación. La función **VAR()** estima mediante mínimos cuadrados ordinarios para cada ecuación, con el orden de retardos igual a 1 por defecto, y se fijará en 3 que es el orden estipulado en el modelo con el que se han simulado los datos, y se pueden incluir parámetros que

Tabla 3.1: Estimación modelo VAR(3) para X1 con 1000 observaciones.

	Estimate	Std. Error	t value	Pr(> t)
y1.l1	0.4818838	0.0310126	15.5383065	0.0000000
y2.l1	0.0055399	0.0299340	0.1850709	0.8532113
y1.l2	-0.1449881	0.0366042	-3.9609747	0.0000800
y2.l2	0.1331521	0.0291826	4.5627275	0.0000057
y1.l3	0.2498043	0.0318613	7.8403662	0.0000000
y2.l3	0.2097965	0.0288529	7.2712409	0.0000000

Tabla 3.2: Estimación modelo VAR(3) para X2 con 1000 observaciones.

	Estimate	Std. Error	t value	Pr(> t)
y1.l1	-0.2711104	0.0331114	-8.187826	0.000000
y2.l1	0.3351355	0.0319598	10.486160	0.000000
y1.l2	0.1428288	0.0390813	3.654654	0.000271
y2.l2	-0.2788044	0.0311575	-8.948228	0.000000
y1.l3	0.3207577	0.0340175	9.429194	0.000000
y2.l3	-0.1227051	0.0308055	-3.983215	0.000073

recojan regresores deterministas y estacionalidad, aunque en el caso que nos concierne, los datos no presentan ninguna de estas características.

Las tablas anteriores muestran que las estimaciones de los parámetros para el modelo de cada variable son significativas y se acercan mucho a los valores reales. Algunos valores de los coeficientes autorregresivos no son completamente precisos, pero teniendo en cuenta el valor de su desviación típica, todos incorporan el valor real del coeficiente en el intervalo al 95 %. Lo mismo pasa con la estimación de la matriz de varianza-covarianzas que es casi igual a la matriz real. Este resultado era de esperar ya que los datos se han simulado con estructura autorregresiva y se sabe exactamente el número de retardos en las relaciones dinámicas.

Las siguientes figuras recogen la estructura de los residuos del modelo estimado para cada serie:

Tabla 3.3: Estimación matriz var-cov modelo VAR(3) datos simulados 1000 observaciones.

x1	x2
0.9426939	0.2870155
0.2870155	1.0742672

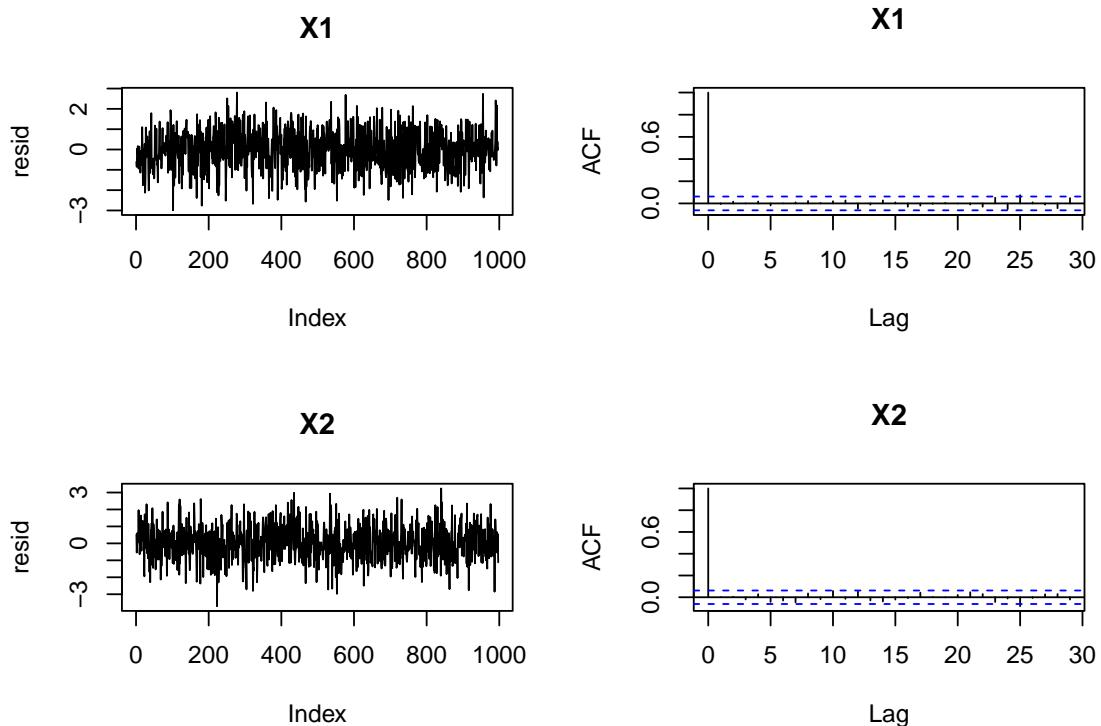


Figura 3.1: Residuos modelo VAR para X1 y X2.

Los residuos muestran comportamiento aleatorio en su serie sin presencia de dependencia. La autocorrelación muestral muestra que no hay retardos con autocorrelación significativa, lo cual indica que los residuos tienen estructura de ruido blanco y el modelo está bien estipulado. Tiene sentido ya que sabemos el número de retardos con el que se han generado las series, por lo que si el algoritmo de estimación está bien elaborado los residuos deben tener estructura de ruido blanco.

Ahora, vamos a estimar el modelo BVAR. Para esto, vamos a utilizar el paquete BVAR que permite estimar modelos VAR bayesianos jerárquicos. La función que realizar la estimación se llama **bvar** pero antes de estimar el modelo hay que fijar las distribuciones a priori de los hiperparámetros y sus parámetros. Para la a priori *minnesota*, usamos la función **bv_mn()** de modo que los valores a priori de sus parámetros son: 0.2 de moda con desviación estándar 0.4 para λ , 2 de moda y 0.25 de desviación estándar para α , y 0.004 de escala y 0.004 de forma para ψ . Estos son los valores por defecto implementados en el paquete y sugeridos por los autores.

En cuanto a la a priori *SOC* y *SUR*, se pueden incorporar con las funciones **bv_soc()** y **bv_sur()**. Se usarán también los parámetros por defecto propuestos por los autores. Por último, hay que usar Metropolis-Hasting para fijar el ratio de convergencia que asegure que se logre la convergencia. También se usarán los valores propuestos por los autores para la convergencia que en este caso son 0.05, 0.0001 y 0.0001 respectivamente.

Lo primero que cabe destacar es el tiempo que tarda en estimar el modelo de forma bayesiana con el muestreador de Gibbs. Este modelo ha tardado algo más de 1 minuto en estimarse con un número de muestras igual a 50000 y *burning* de 25000 (han sido los valores necesarios para que las estimaciones convergieran). El modelo estimado sugiere

Tabla 3.4: Estimación modelo BVAR(3) para datos simulados con 1000 observaciones.

	var1	var2
constant	0.0009997	0.0169813
var1-lag1	0.4834029	-0.2705301
var2-lag1	0.0060564	0.3394940
var1-lag2	-0.1391966	0.1464528
var2-lag2	0.1313580	-0.2800944
var1-lag3	0.2419005	0.3116188
var2-lag3	0.2062712	-0.1182739

Tabla 3.5: Estimación matriz var-cov para modelo BVAR(3) para datos simulados con 1000 observaciones.

	var1	var2
var1	0.9383818	0.2851532
var2	0.2851532	1.0709325

que el hiperparámetro λ asociado a la a priori *minnesota* es 0.51907, para la a priori *SOC* el hiperparámetro toma valor 0.95185, y para la *SUR* 0.214. Estas dos últimas a prioris suelen introducirse para contrarrestar el problema de la *minnesota* que tiene un componente determinista que prima el modelo con los valores iniciales.

En cuanto a la estimación de los parámetros, se ajustan muy bien a los reales, al igual que para la matriz de varianza-covarianzas. Las estimaciones de los coeficientes autorregresivos y de la matriz de varianza-covarianzas son muy similares a las estimaciones por el método frecuentista. Esto, en parte, se debe a que las series contienen 1000 observaciones y aunque existe información a priori, los datos tienen mucho peso en la estimación de la a posteriori. Estos resultados indican que el algoritmo es preciso pero este no es el contexto en el que saca su potencial.

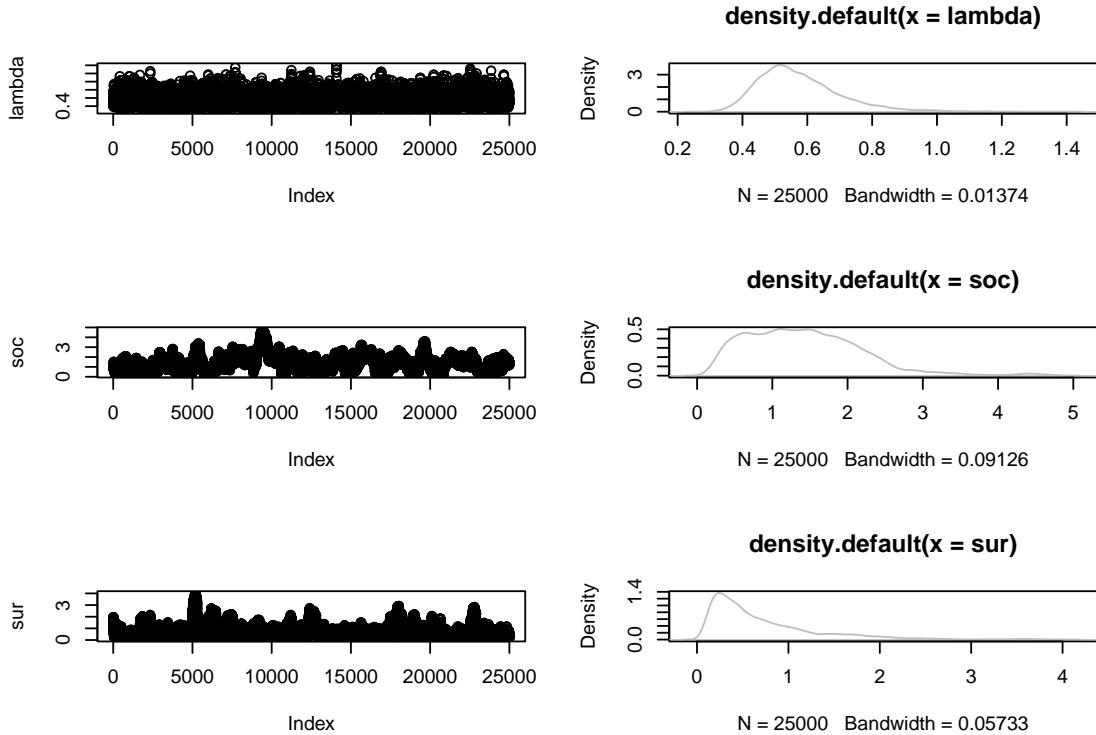


Figura 3.2: Convergencia de los Hiperparámetros BVAR.

La figura 3.2 muestra que el hiperparámetro para la a priori *minnesota* ha convergido, y lo mismo para la *SOC* y *SUR* aunque presentan desviaciones muy marcadas cada cierto tiempo.

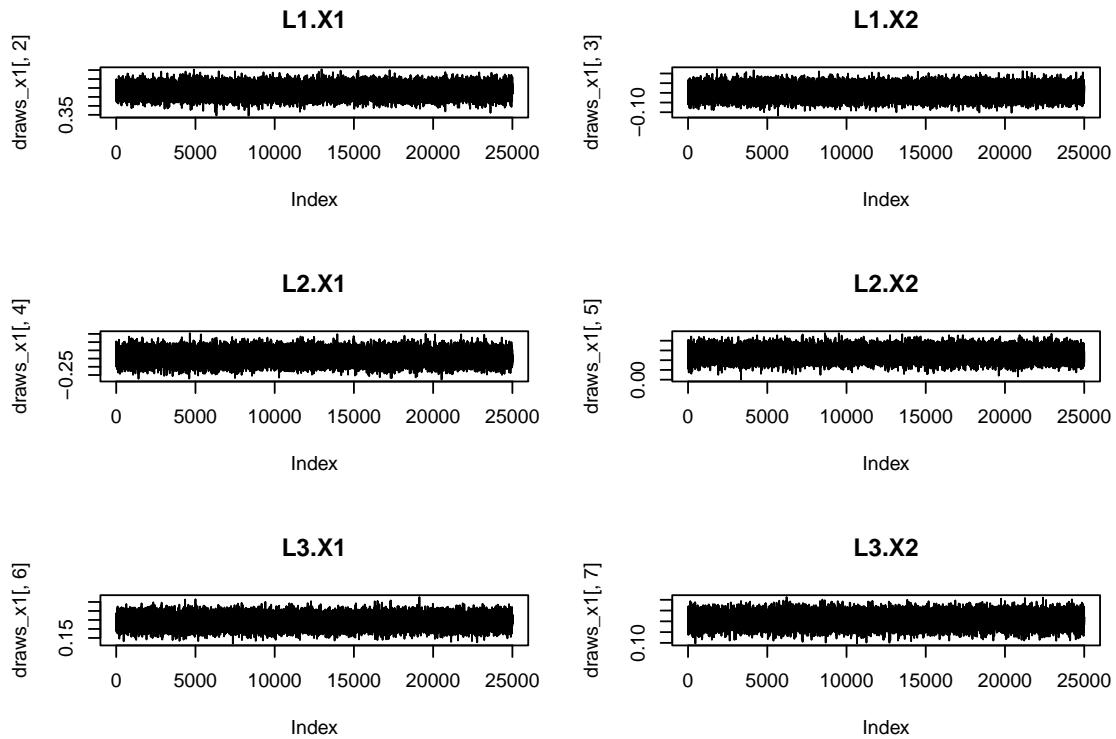


Figura 3.3: Convergencia de los parámetros BVAR para X1.

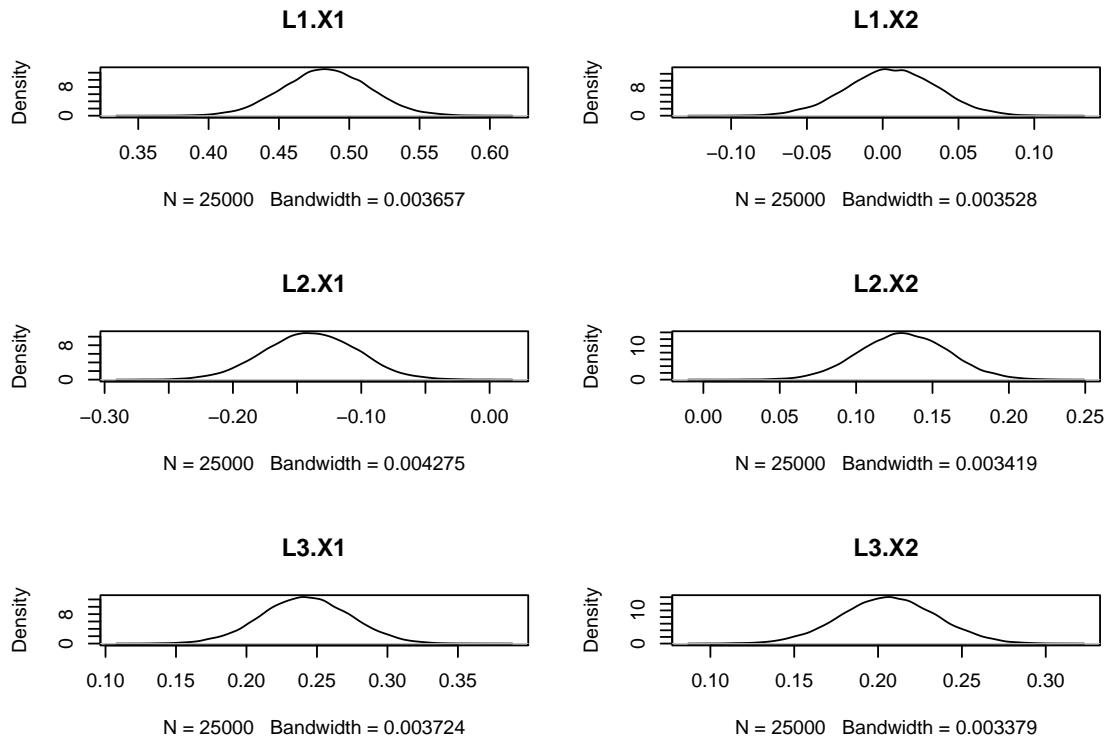


Figura 3.4: Densidad de los parámetros BVAR para X1.

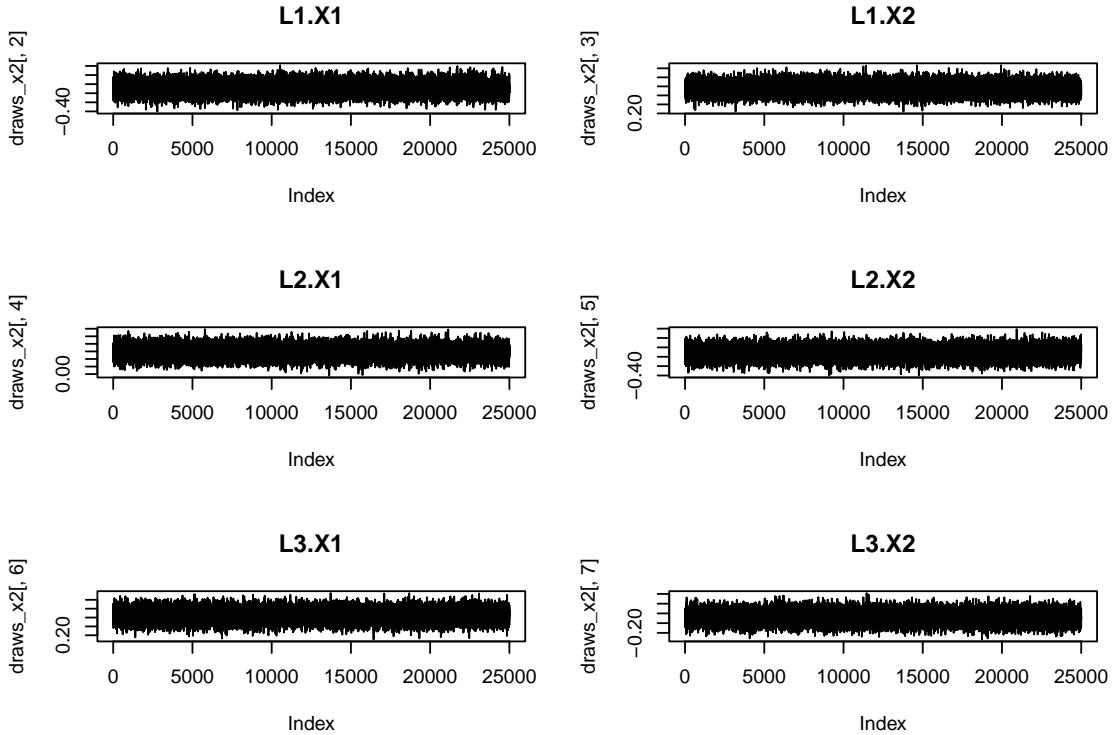


Figura 3.5: Convergencia de los parámetros BVAR para X2.

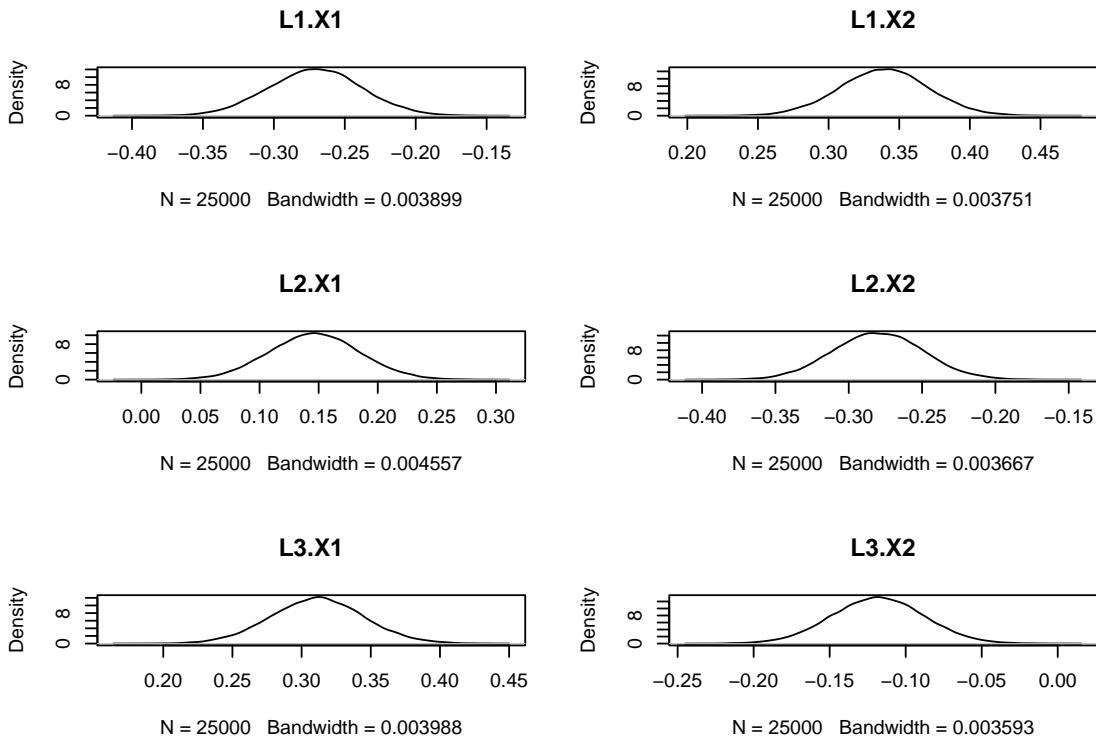


Figura 3.6: Densidad de los parámetros BVAR para X2.

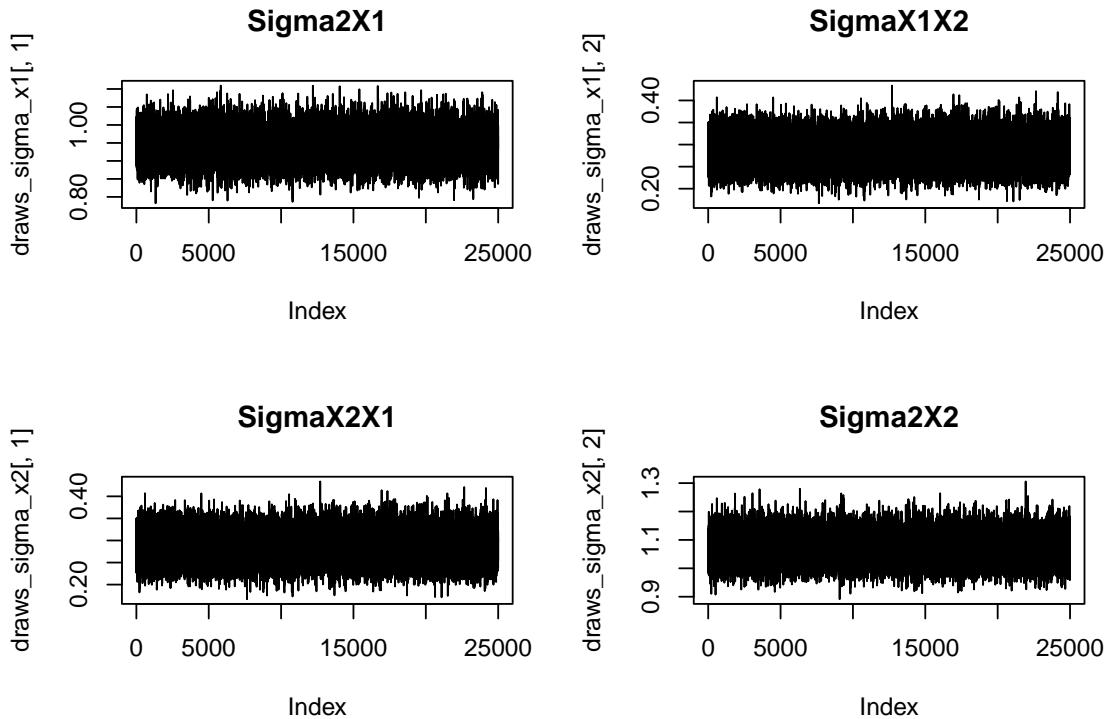


Figura 3.7: Convergencia matriz covarianza modelo BVAR.

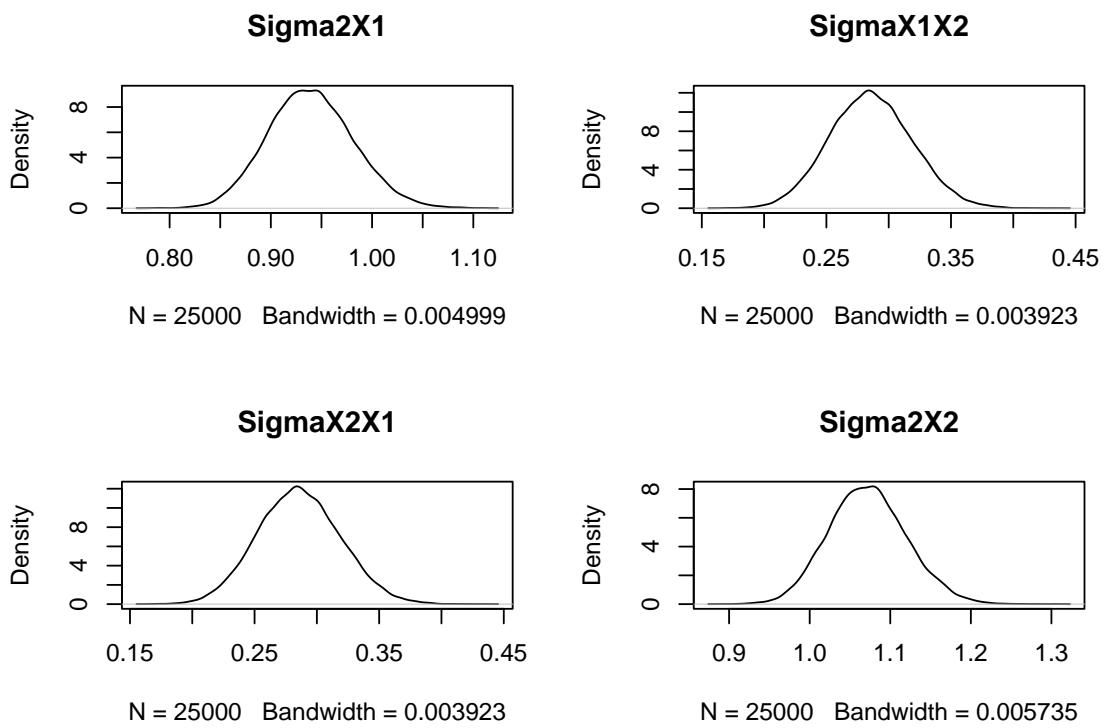


Figura 3.8: Densidad de los parámetros de covarianza muestra BVAR.

Las figuras 3.3, 3.4, 3.5 y 3.6 muestran que los parámetros estimados para las relaciones

Tabla 3.6: Estimación modelo VAR(3) para X1 con 100 observaciones.

	Estimate	Std. Error	t value	Pr(> t)
y1.l1	0.8230128	0.0927669	8.8718378	0.0000000
y2.l1	-0.2254239	0.0791106	-2.8494784	0.0054155
y1.l2	-0.6475107	0.1157129	-5.5958362	0.0000002
y2.l2	0.2196381	0.0753636	2.9143811	0.0044852
y1.l3	0.5416297	0.0959147	5.6469951	0.0000002
y2.l3	0.0008944	0.0784249	0.0114048	0.9909255

Tabla 3.7: Estimación modelo VAR(3) para X2 con 100 observaciones.

	Estimate	Std. Error	t value	Pr(> t)
y1.l1	-0.2051464	0.1195771	-1.7156002	0.0896382
y2.l1	0.2692510	0.1019740	2.6403891	0.0097450
y1.l2	-0.0405570	0.1491547	-0.2719123	0.7863054
y2.l2	-0.3649489	0.0971441	-3.7567808	0.0003034
y1.l3	0.3868822	0.1236346	3.1292396	0.0023553
y2.l3	-0.1866598	0.1010902	-1.8464679	0.0680748

autorregresivas han convergido y que su distribución a posteriori es normal. En cuanto a la estimación de la matriz de varianza-covarianzas, las figuras 3.7 y 3.8 también indican que las estimaciones han convergido.

Después de estimar con los datos simulados de 1000 observaciones, vamos a estimar los modelos para el conjunto de datos simulados con 100 observaciones. En este caso, el conjunto de datos es significativamente más pequeño. Los métodos bayesianos con información a priori suelen funcionar mejor que los frecuentistas cuando tenemos pocos datos. Vamos a comprobar si esto se cumple para nuestros datos.

La tabla anterior tiene resultados interesantes. Aún especificando la estructura real que existe entre las variables, al utilizar 100 datos las estimaciones de los coeficientes no son buenas. Hay un coeficiente, el del primer retardo de X para la estimación de Y , que lo estima como negativo y muy significativo cuando en realidad es positivo. También hay un coeficiente que lo estima como no significativo cuando en realidad toma el valor 0.2, este es el coeficiente del tercer retardo de X para la estimación de Y . Aparte de esos dos casos tan claros, el resto de estimaciones están muy lejos de sus verdaderos valores.

En cuanto a las estimaciones de los coeficientes de la segunda ecuación, pasa igual que en la primera, hay retardos que los estima como negativos y muy significativos cuando en verdad son positivos, hay un retardo que según el modelo no es significativo cuando en verdad sí lo es, y la mayoría de las estimaciones están muy lejos del valor real.

En cuanto a la estimación de la matriz de varianza-covarianzas, las estimaciones de las

Tabla 3.8: Estimación matriz var-cov modelo VAR(3) con 100 observaciones.

	y1	y2
y1	0.5577058	0.0940530
y2	0.0940530	0.9388269

Tabla 3.9: Estimación modelo BVAR(3) para datos simulados con 100 observaciones.

	var1	var2
constant	-0.1994538	0.0927026
var1-lag1	0.6964892	-0.1904004
var2-lag1	-0.1703243	0.3036261
var1-lag2	-0.5109228	0.0264880
var2-lag2	0.1567760	-0.3611953
var1-lag3	0.3353957	0.3492451
var2-lag3	0.0437455	-0.1520908

covarianzas son casi 0 cuando en verdad son 0.3 y en cuanto a la varianza de las variables, estima muy bien la varianza de la segunda variable pero mal la de la primera (recordemos que ambas variables tienen varianza 1).

Las siguientes figuras recogen la estructura de los residuos del modelo estimado para cada serie:

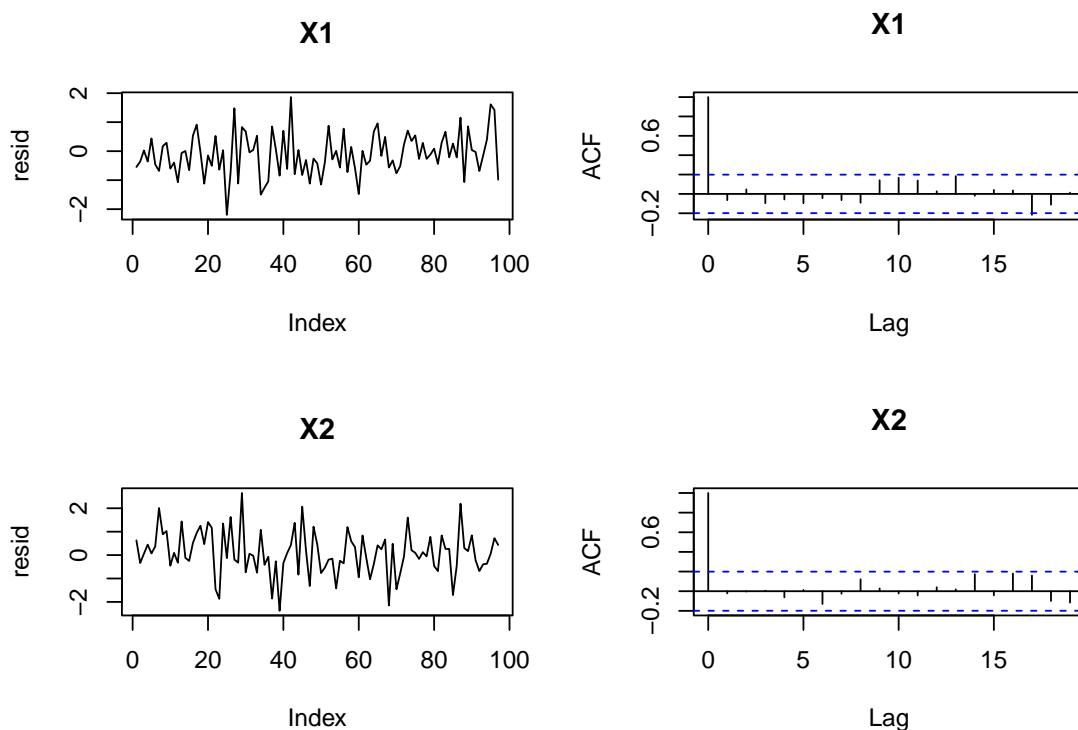


Figura 3.9: Residuos modelo VAR para X1 y X2.

Los residuos muestran comportamiento aleatorio en su serie sin presencia de dependencia. La autocorrelación muestral afirma que no hay retardos con autocorrelación significativa lo cual indica que los residuos tienen estructura de ruido blanco y el modelo está bien estipulado.

Ahora, vamos a estimar el modelo BVAR. Los valores de las a priori serán los mismos que los utilizados en el conjunto con 1000 observaciones y que proponen los autores.

Tabla 3.10: Estimación matriz var-cov para modelo BVAR(3) para datos simulados con 100 observaciones.

	var1	var2
var1	0.5475024	0.1093523
var2	0.1093523	0.8936469

En cuanto a la estimación de los coeficientes, para la primera variable, en general las estimaciones no son muy precisas aunque hay algunos coeficientes que sí se aproximan bastante a su valor real como el segundo retardo de la *var2* para la primera variable. Por otro lado, las estimaciones para el modelo de la segunda variables son bastante buena, muy cercanas a los valores reales. Esto es una mejora notable con respecto al modelo frecuentista donde en general todos los coeficientes estimaciones estaban muy lejos de sus valores reales.

En cuanto a la matriz de varianza-covarianzas, en general no estima muy bien los coeficientes aunque para las covarianzas las estimaciones se acercan más al valor real que con el método frecuentista, pero las varianzas se alejan en comparación.

Vamos a sacar los gráficos de convergencia y densidad para la estimación bayesiana.

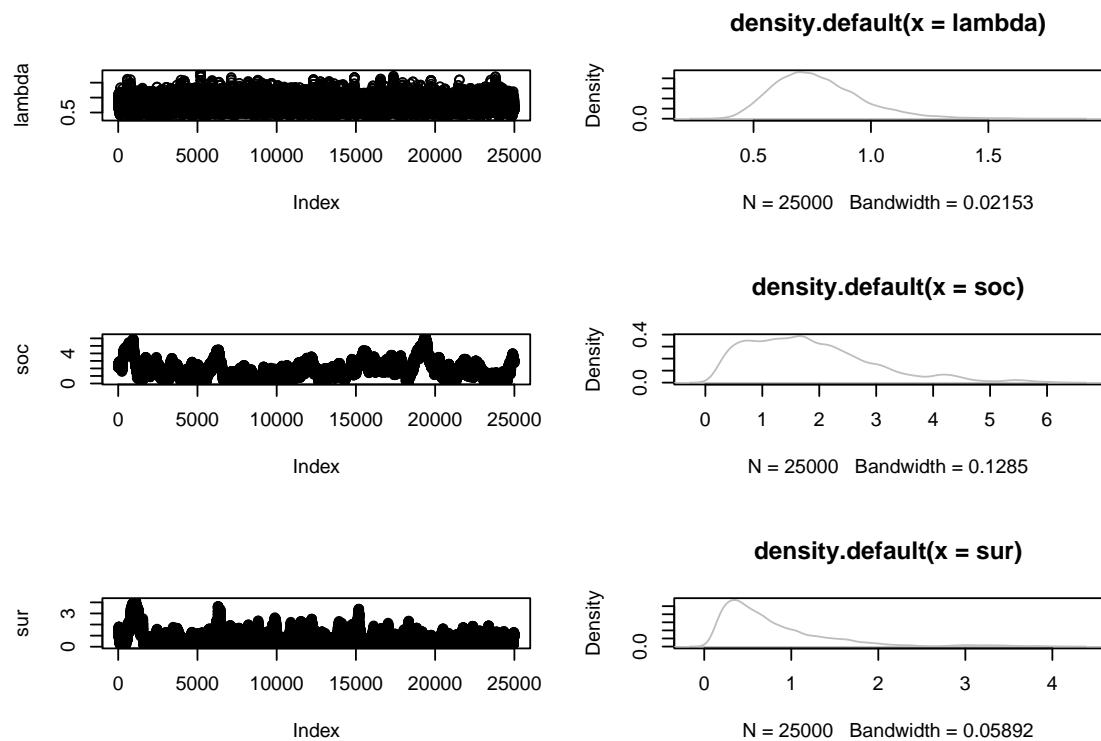


Figura 3.10: Convergencia de los Hiperparámetros BVAR.

La figura 3.10 muestra que el hiperparámetro para la a priori *minnesota* ha convergido, igual para la *SUR* y en cuanto a la *SOC*, parece que ha llegado a converger pero presenta algunas rachas pronunciadas.

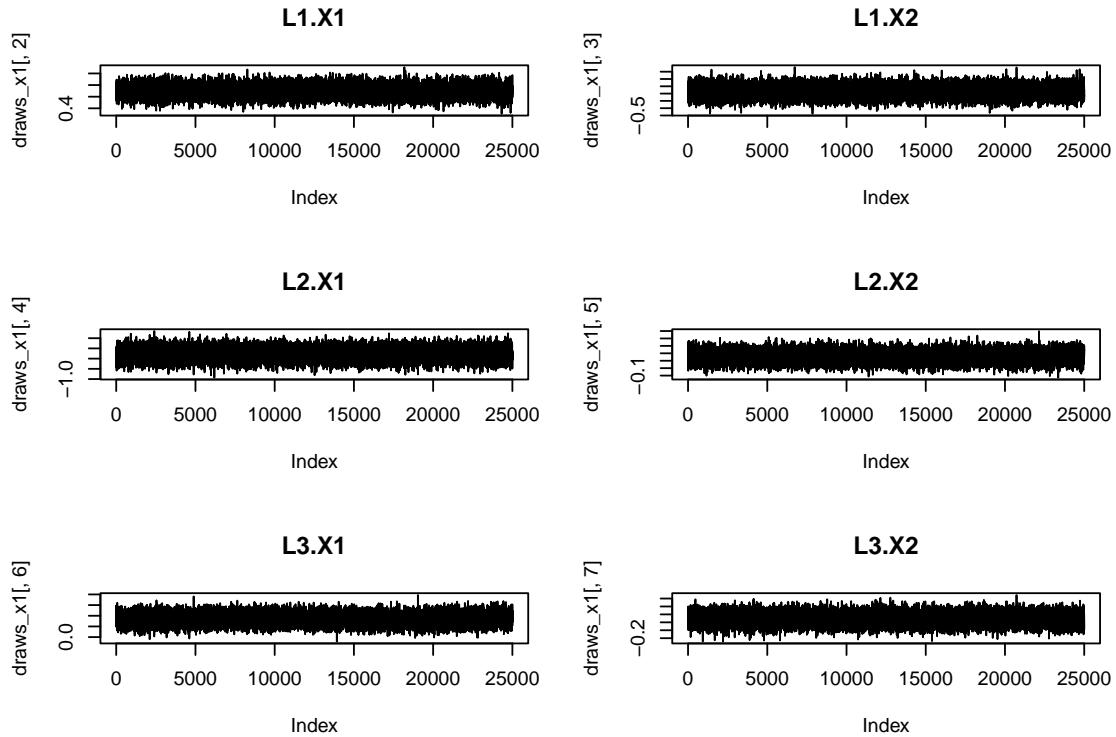


Figura 3.11: Convergencia de los parámetros BVAR para X1.

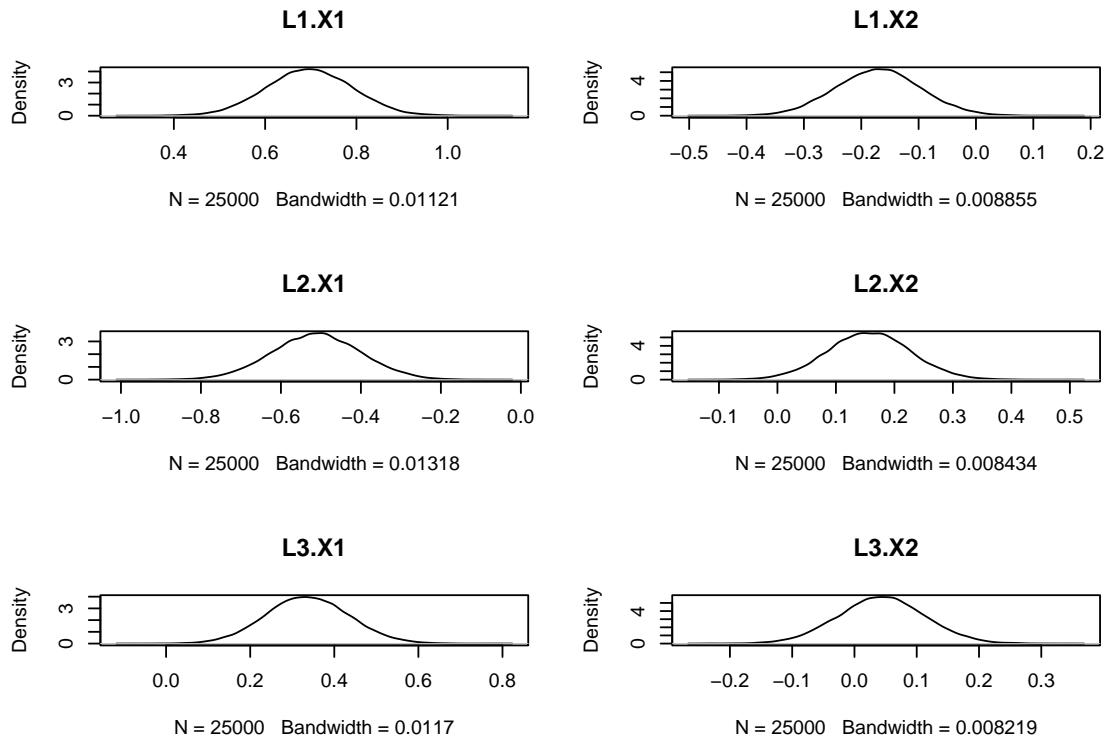


Figura 3.12: Densidad de los parámetros BVAR para X1.

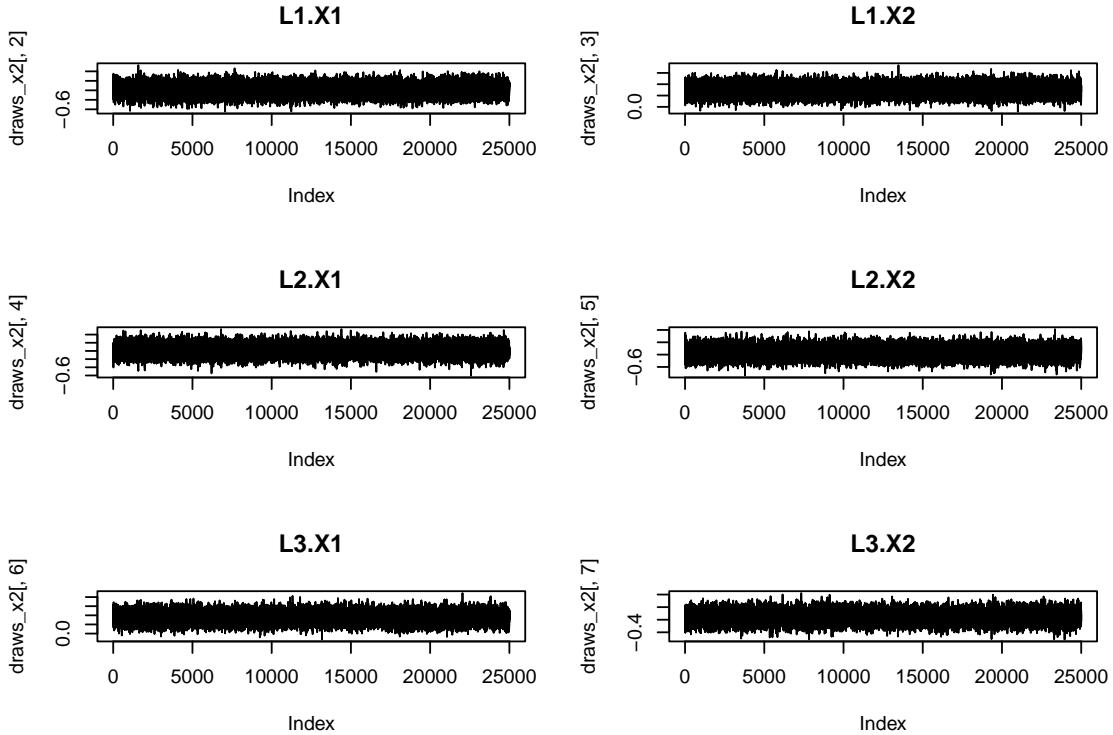


Figura 3.13: Convergencia de los parámetros BVAR para X2.

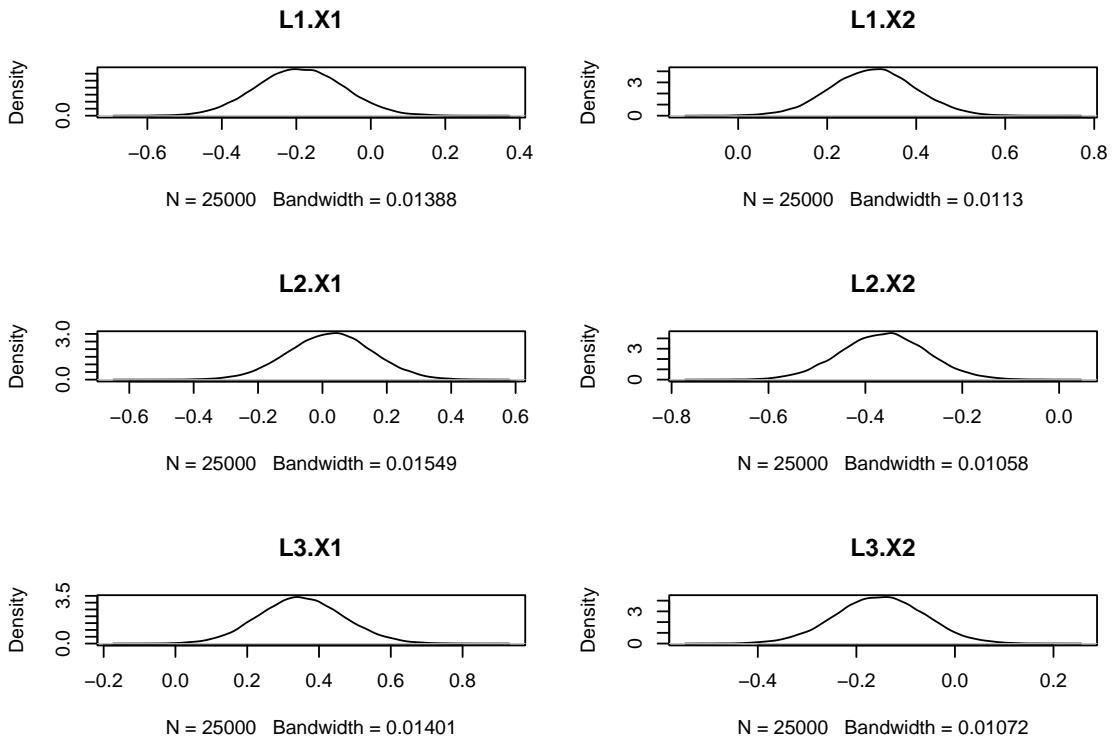


Figura 3.14: Densidad de los parámetros BVAR para X2.

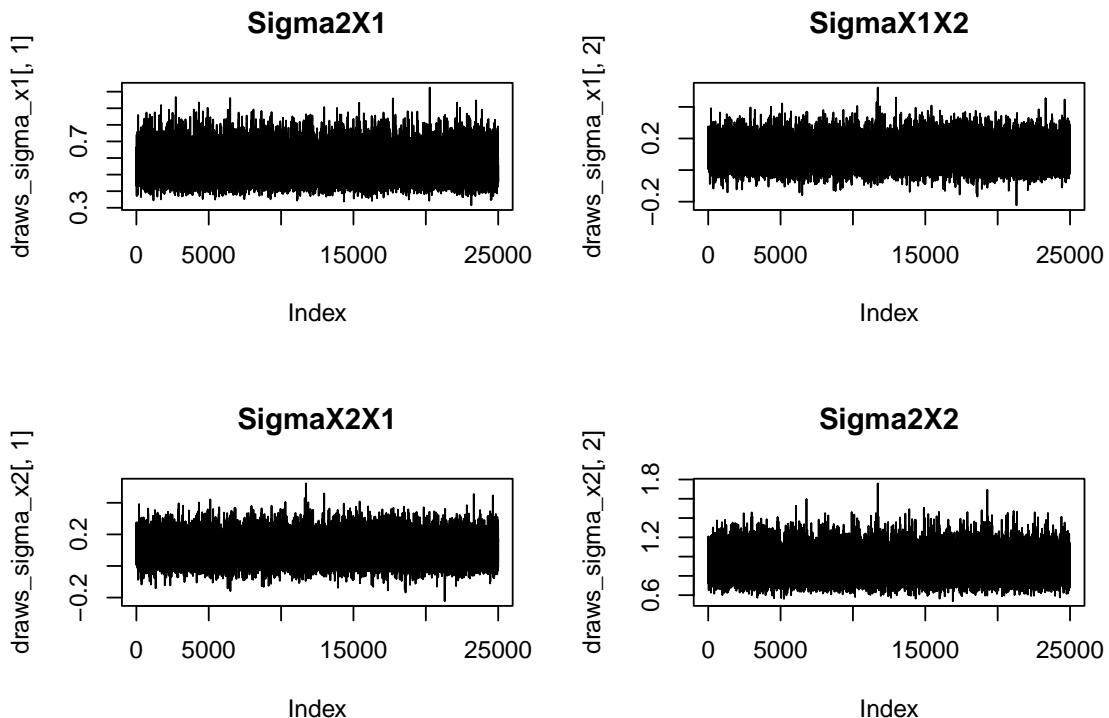


Figura 3.15: Convergencia matriz covarianza modelo BVAR.

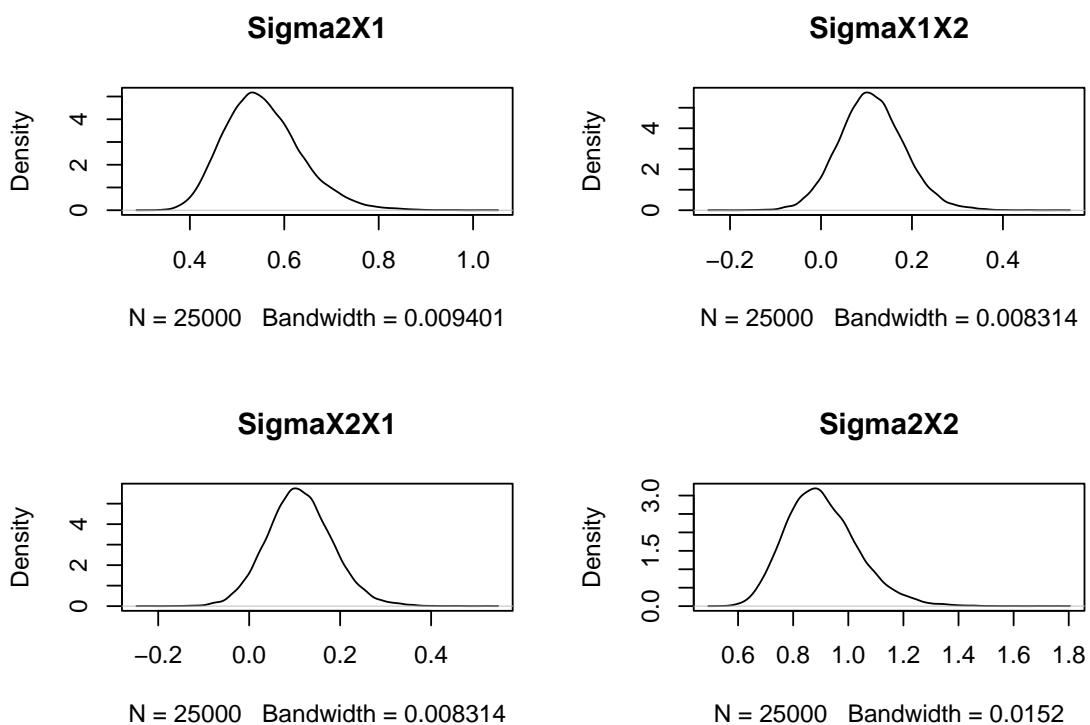


Figura 3.16: Densidad de los parámetros de covarianza muestra BVAR.

Las figuras 3.3, 3.4, 3.5 y 3.6 muestran que los parámetros estimados para las relaciones

Tabla 3.11: Estimación modelo VAR(2) para Inflación datos macro EE.UU. 1959-2007.

	Estimate	Std. Error	t value	Pr(> t)
Inflación.l1	-0.4957409	0.0656421	-7.552179	0.0000000
Desempleo.l1	-0.5858517	0.1036993	-5.649526	0.0000001
Inflación.l2	-0.3887190	0.0655551	-5.929654	0.0000000
Desempleo.l2	0.1184310	0.1099568	1.077069	0.2828298

Tabla 3.12: Estimación modelo VAR(2) para Desempleo datos macro EE.UU. 1959-2007.

	Estimate	Std. Error	t value	Pr(> t)
Inflación.l1	0.0853327	0.0461121	1.8505478	0.0658037
Desempleo.l1	0.6403199	0.0728464	8.7900000	0.0000000
Inflación.l2	0.0096612	0.0460510	0.2097938	0.8340558
Desempleo.l2	-0.0359724	0.0772422	-0.4657087	0.6419631

autorregresivas han convergido y que su distribución a posteriori es normal. En cuanto a la estimación de la matriz de varianza-covarianzas, las figuras 3.7 y 3.8 también indican que las estimaciones han convergido.

En general, cabe destacar que aunque las estimaciones no hayan sido demasiado precisas, el intervalo de confianza al 95 % para la estimación bayesiana contiene a los verdaderos valores de los coeficientes en casi todas las estimaciones. Esto es una de las ventajas de la estimación bayesiana, aunque pueda tener un sesgo, se dispone de distribuciones de probabilidad.

Las densidades de las covarianzas no llegan a recoger el verdadero valor, y para las varianzas solo se captura el verdadero valor dentro del intervalo de confianza al 95 % para σ_{var2}

Ahora vamos a estimar estos modelos para el conjunto de datos macroeconómicos de desempleo e inflación en Estados Unidos durante 1959-2007.

En el apartado de análisis de datos se ha detectado que las series originales no son estacionarias en la parte regular y que necesitan una primera diferencia. Comenzamos estimando el modelo VAR frecuentista. Como no sabemos cuál es el número de retardos óptimo para la relación entre las variables, se empezará estimando un modelo con 3 retardos y después de analizar los residuos, el correlograma de éstos, y el nivel de significación de los parámetros estimados, se decidirá si el modelo está bien estipulado o si hace falta cambiar el número de retardos. En cuanto a la estacionalidad y a incluir constantes en el modelo, los datos no presentan estacionalidad de ningún tipo, ni regresores deterministas que modelar a través de constantes.

Tabla 3.13: Estimación matriz var-cov modelo VAR(2) para datos macro EE.UU. 1959-2007..

	Inflación	Desempleo
Inflación	0.1332610	-0.0023836
Desempleo	-0.0023836	0.0657631

Tabla 3.14: Estimación modelo BVAR(2) para datos macro EE.UU. 1959-2007.

	Inflación	Desempleo
constant	0.0016338	-0.0040510
Inflación-lag1	-0.4694629	0.0865896
Desempleo-lag1	-0.5603795	0.6505770
Inflación-lag2	-0.3653001	0.0098056
Desempleo-lag2	0.1189121	-0.0339052

Tabla 3.15: Estimación matriz var-cov modelo BVAR(2) para datos macro EE.UU. 1959-2007.

	Inflación	Desempleo
Inflación	0.1369881	-0.0019946
Desempleo	-0.0019946	0.0656465

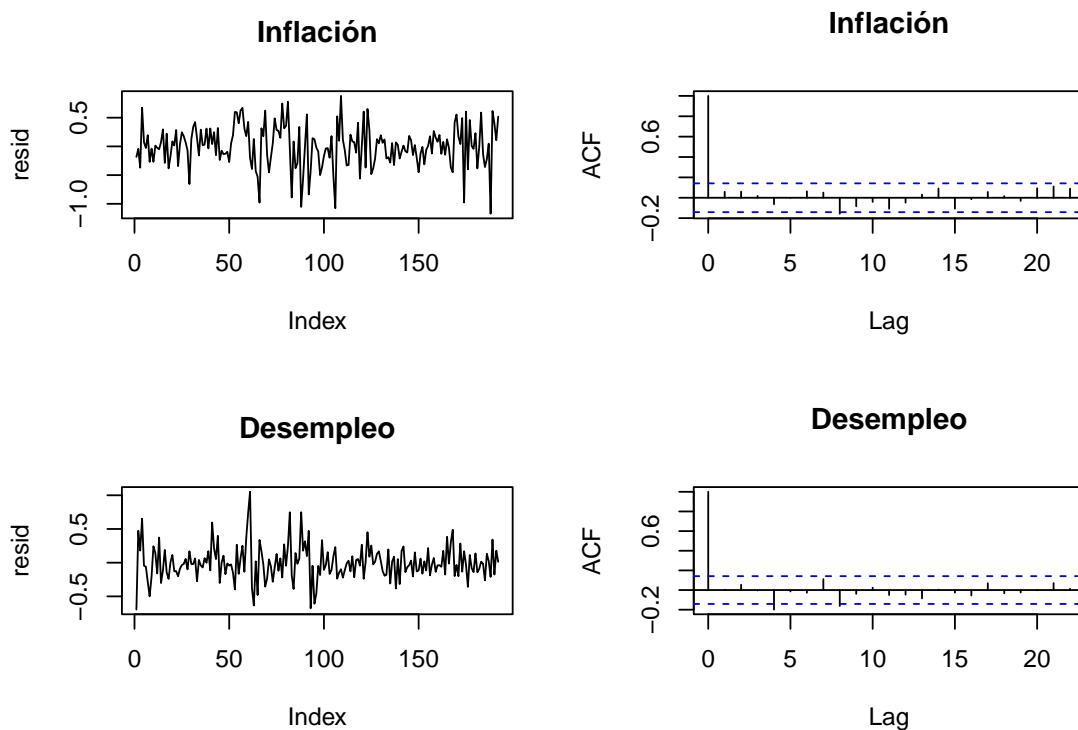


Figura 3.17: Residuos modelo VAR para Inflación y Desempleo.

Después estimar varios modelos VAR, el modelo que parece ser el más indicado para recoger la relación entre las variables es un *VAR(2)* ya que si incorporamos más retardos, las estimaciones no son significativas para todos los retardos superiores a 2, y, como podemos ver en 3.17, los residuos de los modelos para cada variable presentan estructura de ruido blanco. Cabe resaltar que en el modelo estimado para el Desempleo, sólo el primer retardo es significativo, pero al ser un modelo VAR, hay que dejar los dos retardos ya que para el modelo de Inflación sí es significativo el segundo retardo.

Ahora vamos a estimar este modelo pero de manera bayesiana.

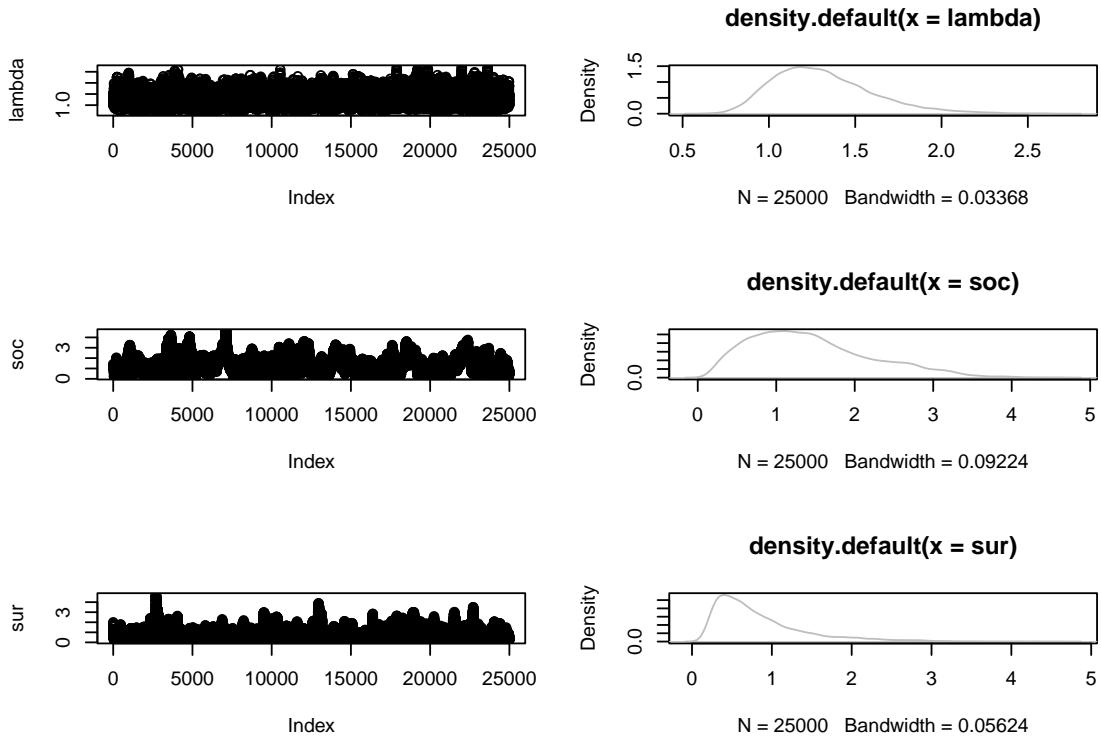


Figura 3.18: Draws y distribuciones a posteriori de los hiperparámetros para los datos US.

Después de estimar varios modelos, al igual que con la inferencia frecuentista, el modelo que parece ser más adecuado es el que contiene dos retardos. Los resultados del modelo recogidos en 3.14 y 3.15 son bastante parecidos a los resultados frecuentistas tanto en los coeficientes autorregresivos como en la matriz de varianza-covarianzas. Los hiperparámetros han convergido y el valor estimado de la log-verosimilitud es prácticamente idéntico.

En cuanto a los coeficientes autorregresivos del modelo estimado para la inflación,

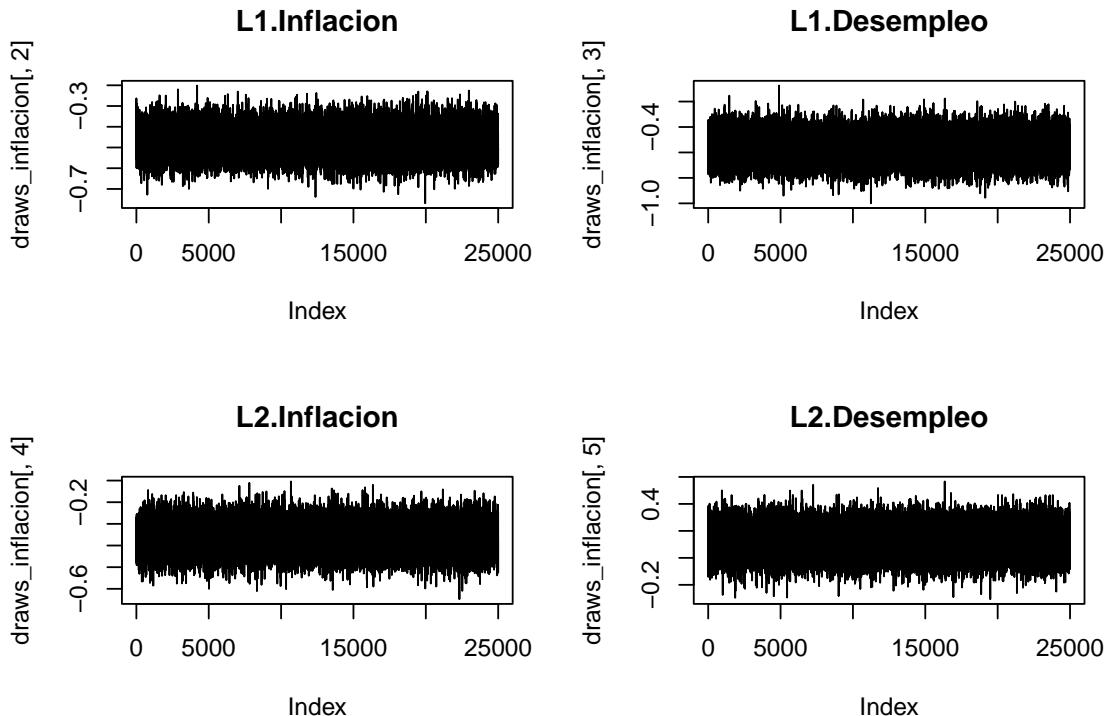


Figura 3.19: Convergencia de los parámetros BVAR para Inflación

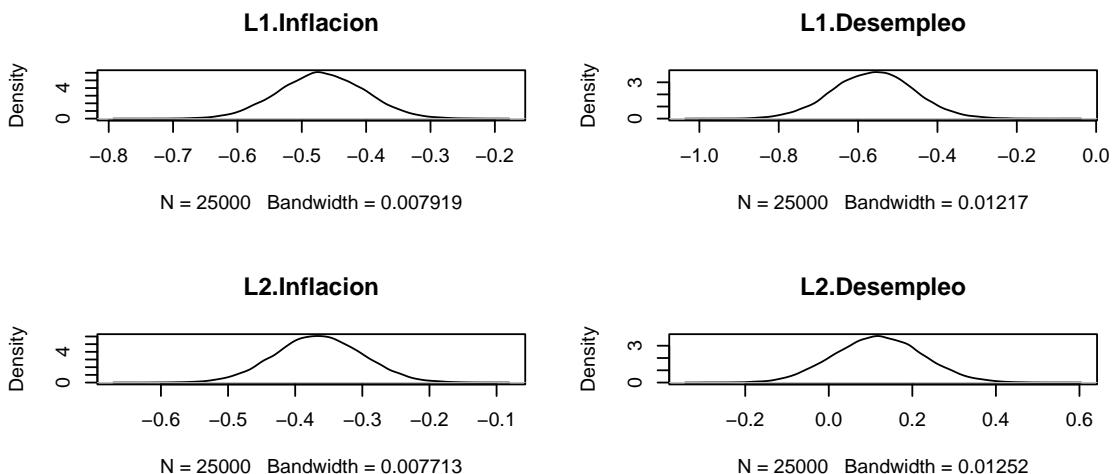


Figura 3.20: Densidad de los parámetros BVAR para Inflación

Las estimaciones de la distribución a posteriori convergen y las densidades de cada

parámetro muestran que todos son significativos excepto para el retardo de orden 2 de desempleo, en línea con los resultados obtenidos con el método frecuentista.

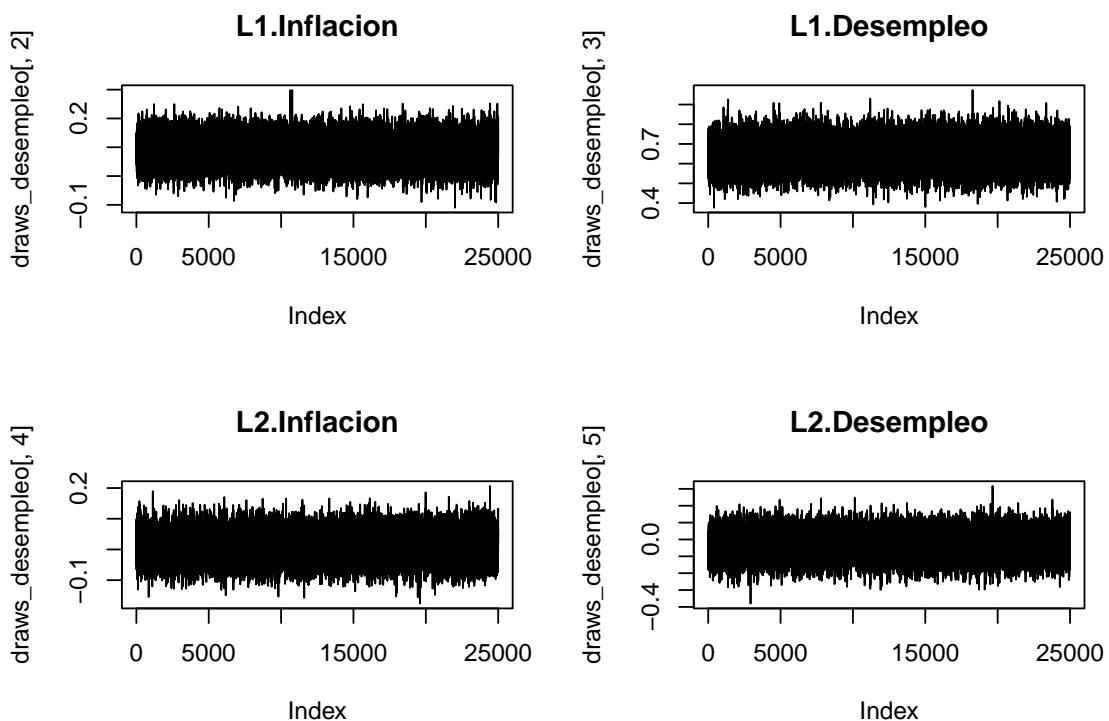


Figura 3.21: Convergencia de los parámetros BVAR para Desempleo

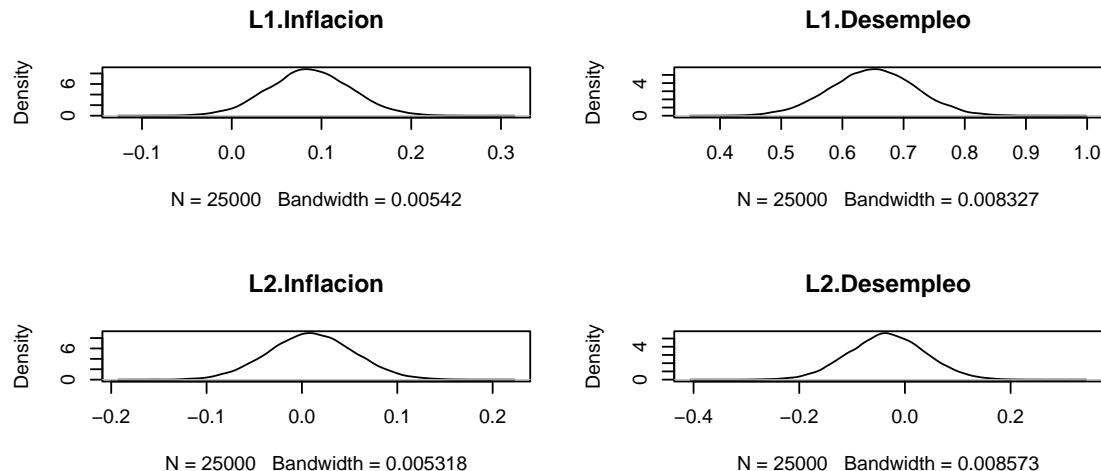


Figura 3.22: Densidad de los parámetros BVAR para Desempleo

Las figuras 3.21 y 3.22 muestran que las estimaciones de las distribuciones a posteriori para los coeficientes del modelo para el desempleo también han convergido y que solo es significativo el primer retardo para esta variable.

Por último, las figuras 3.23 y 3.24 indican que la estimación de la matriz de varianza-covarianzas ha convergido y que al 95 %, si miramos la densidad de $\sigma_{Desem,Infl}$, la estimación a posteriori sugiere que las series tienen covarianza 0 por lo que no habría relación contemporánea.

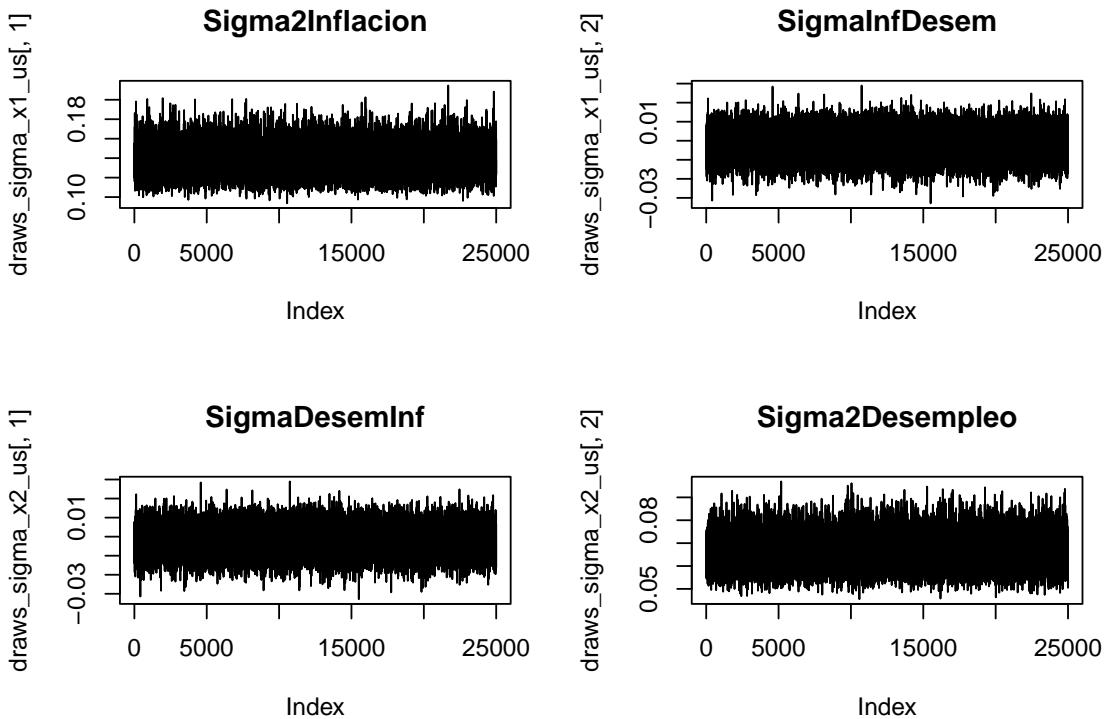


Figura 3.23: Convergencia matriz covarianza modelo BVAR datos macro EE.UU.

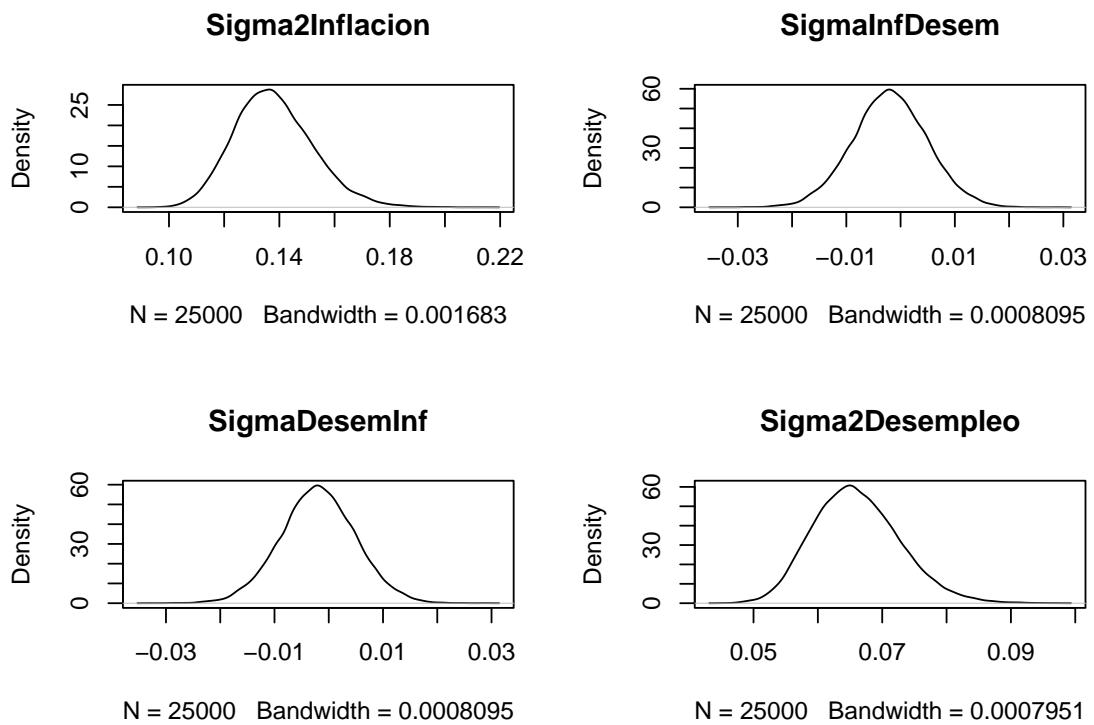


Figura 3.24: Densidad de los parámetros de covarianza muestral BVAR datos macro EE.UU.

Tabla 3.16: Estimación relación largo plazo modelo VECM(2) para datos simulados con 1000 observaciones.

	r1
Var1	1.000000
Var2	-1.496641

Tabla 3.17: Estimación modelo VECM(2) para datos simulados con 1000 observaciones.

	ECT	Var1 -1	Var2 -1	Var1 -2	Var2 -2
Equation Var1	-0.3087045	0.1849178	-0.3122232	0.1747201	0.1110475
Equation Var2	0.2213124	0.3761328	-0.2335169	0.1127229	0.2544009

3.2. Estimación de los modelos VECM y BVECM

En este apartado vamos a analizar las series simuladas que son cointegradas, y los datos que recogen el LIBOR a 12 meses y los swaps OVERNIGHT a 9 meses. Primero, vamos a estimar el modeo VECM frecuentista a los datos simulados y después el modelo bayesiano para comparar los parámetros estimados con los *reales*. Hay que destacar que los parámetros reales con los que se han simulado los datos siguen la forma de 2.24. Se va a usar la librería `tsDyn`, y, en específico la función `VECM` para estimar el modelo frecuentista. Como método de estimación se usará la máxima verosimilitud de Johansen, ya que generalmente, produce mejores estimaciones que mínimos cuadrados ordinarios en dos fases. Como sabemos que los datos simulados se han creado con una estructura autorregresiva de 2 retardos, lo especificaremos en los parámetros del modelo.

Primero, estimamos para el conjunto de datos simulados con 1000 observaciones:

Las estimaciones son muy precisas, sobre todo para el parámetro β de la relación de largo plazo. En general todos los coeficientes estimados son muy cercanos a los reales reflejados en 2.24 y son significativos al 5% como era de esperar. Los residuos de las figuras 3.25 y 3.26 indican que el modelo está bien especificado ya que los residuos tienen estructura aleatoria de ruido blanco.

Tabla 3.18: Estimación matriz var-cov modelo VECM(2) para datos simulados con 1000 observaciones.

0.9444194	0.0013178
0.0013178	1.0843747

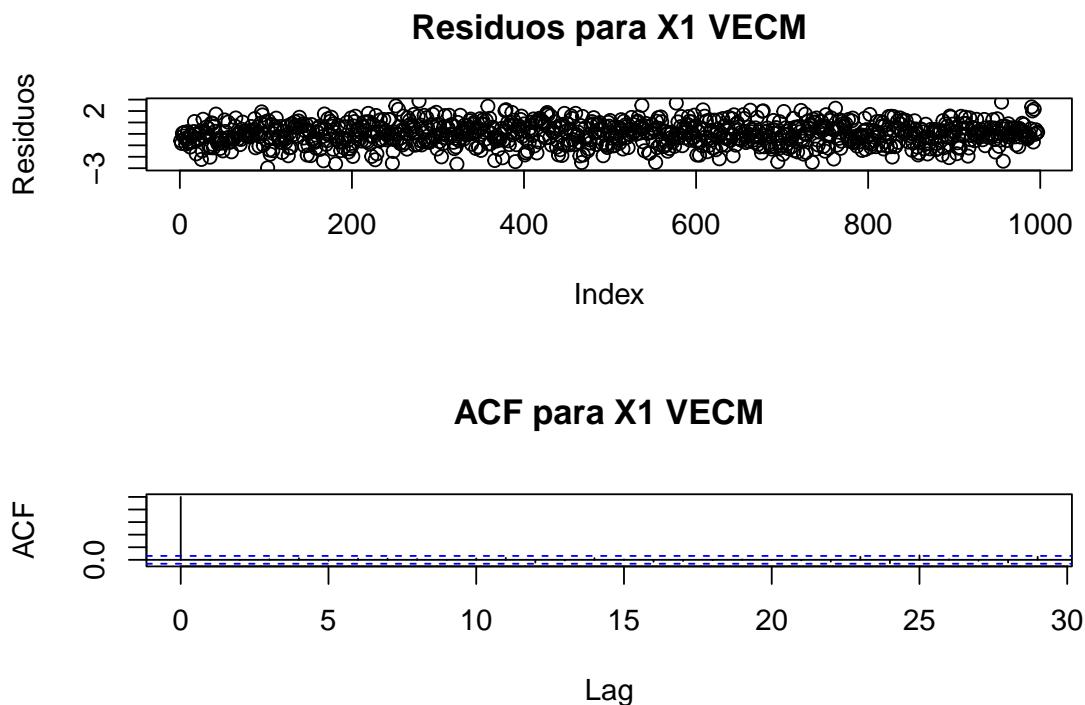


Figura 3.25: Residuos X1 modelo VECM para datos simulados de 1000 observaciones.

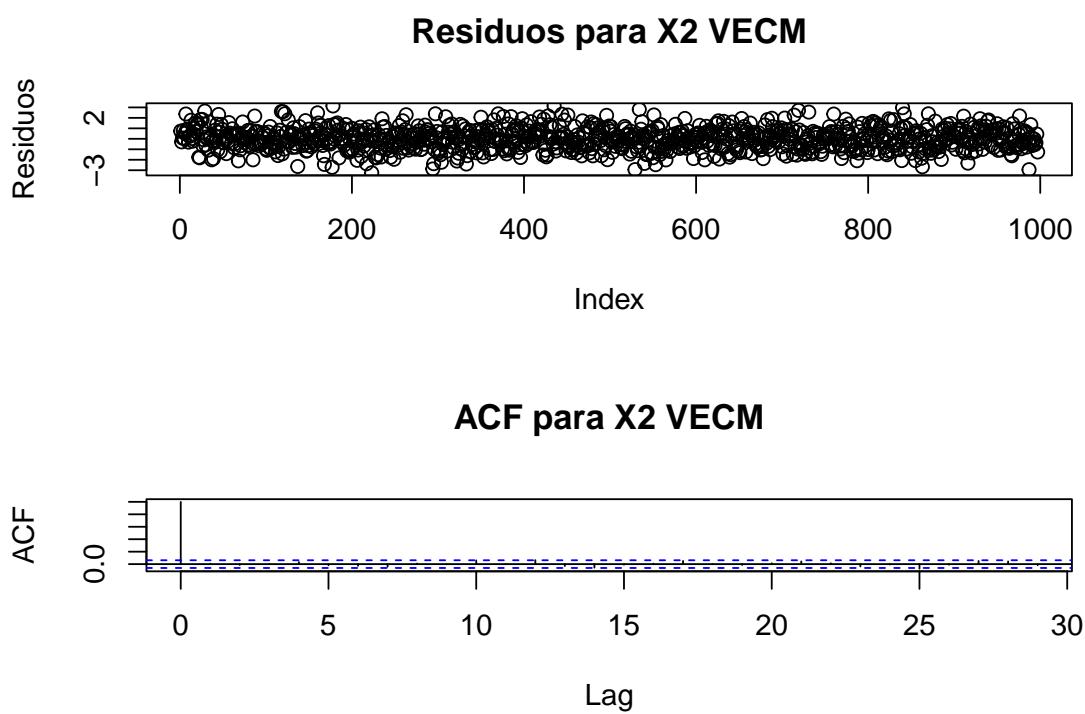


Figura 3.26: Residuos X2 modelo VECM para datos simulados de 1000 observaciones.

Ahora, vamos a estimar estos coeficientes mediante el método bayesiano. La librería

Tabla 3.19: Estimación relación de largo plazo modelo BVECM(2) para datos simulados con 1000 observaciones.

l.Series 1	1.000
l.Series 2	-1.497

Tabla 3.20: Estimación modelo BVECM(2) para datos simulados con 1000 observaciones.

	l.Series 1	d.Series 1.l01	d.Series 2.l01	d.Series 1.l02	d.Series 2.l02
d.Series 1	-0.3072476	0.1837231	-0.3100575	0.1742681	0.1122703
d.Series 2	0.2208009	0.3760533	-0.2339719	0.1125441	0.2543304

`bvarools` permite estimar modelos BVECM. Con la función `gen_vec()` se crea el objeto de clase *BVEC* donde se introduce la estructura del modelo que se quiere estimar (número de retardos, parte determinista, iteraciones del muestreador y número de burnin). Después, se introduce la función `add_priors()` donde se añaden las a prioris al objeto de tipo *BVEC*. En este caso, se van a utilizar distribuciones a priori no informativas para ver qué tal resultados se obtienen. Después de crear el objeto *BVEC* con las a prioris definidas, se puede utilizar la función `draw_posterior()` para obtener las estimaciones a posteriori. En este caso, se ha desarrollado un código específico para la estimación a partir un ejemplo que hay dentro de los ejemplos que contiene la librería. Se ha hecho esto porque la estimación a posteriori con la función `draw_posterior()` tardaba demasiado, y, teniendo en cuenta que más adelante se va a realizar un walk forward validation, esto hacía que fuese computacionalmente muy costoso de hacer con la función `draw-posterior()`.

La estimación de la relación de largo plazo es muy buena (-1.497 frente a -1.5 real) igual que con el método frecuentista. En cuanto a los términos de corrección de equilibrio, los IC al 95 % indican que son distintos de cero y toman valores muy cercanos a los reales. Todos los coeficientes autorregresivos estimados son significativos según indica el IC al 95 % y se ajustan muy bien a la relación verdadera entre las variables que se ha usado para simular los datos.

En cuanto a la estimación de la matriz de varianza-covarianzas, las estimaciones para $\sigma_{x_1}^2$ y $\sigma_{x_2}^2$ contienen al 1 en su IC al 95 % y su valor estimado es muy cercano. En cuanto a las covarianzas, los intervalos de confianza incluyen al 0 al 95 % por lo que recogen el verdadero valor.

Aún habiendo incorporado distribuciones a prioris no informativas, los resultados son muy acertados. Esto se debe a la cantidad de datos que se han usado para estimar el modelo. Como se han usado 1000 datos, éstos tienen mucho peso en la estimación de la a posteriori y por ello aún incorporando distribuciones a priori no informativas los resultados son igual de buenos que con el método frecuentista. Cabe destacar que el

Tabla 3.21: Estimación matriz var-cov modelo BVECM(2) para datos simulados con 1000 observaciones.

	d.Series 1	d.Series 2
d.Series 1	0.9468743	0.0006633
d.Series 2	0.0006633	1.0874113

Tabla 3.22: Estimación relación largo plazo modelo VECM(2) para datos simulados con 100 observaciones.

	r1
Var1	1.000000
Var2	-1.425288

Tabla 3.23: Estimación modelo VECM(2) para datos simulados con 100 observaciones.

	ECT	Var1 -1	Var2 -1	Var1 -2	Var2 -2
Equation Var1	-0.2095335	0.2098882	-0.4255205	0.0249050	-0.0012693
Equation Var2	0.2281950	0.3913241	-0.2011688	0.0802597	0.2299260

tiempo de computación es significativamente mayor.

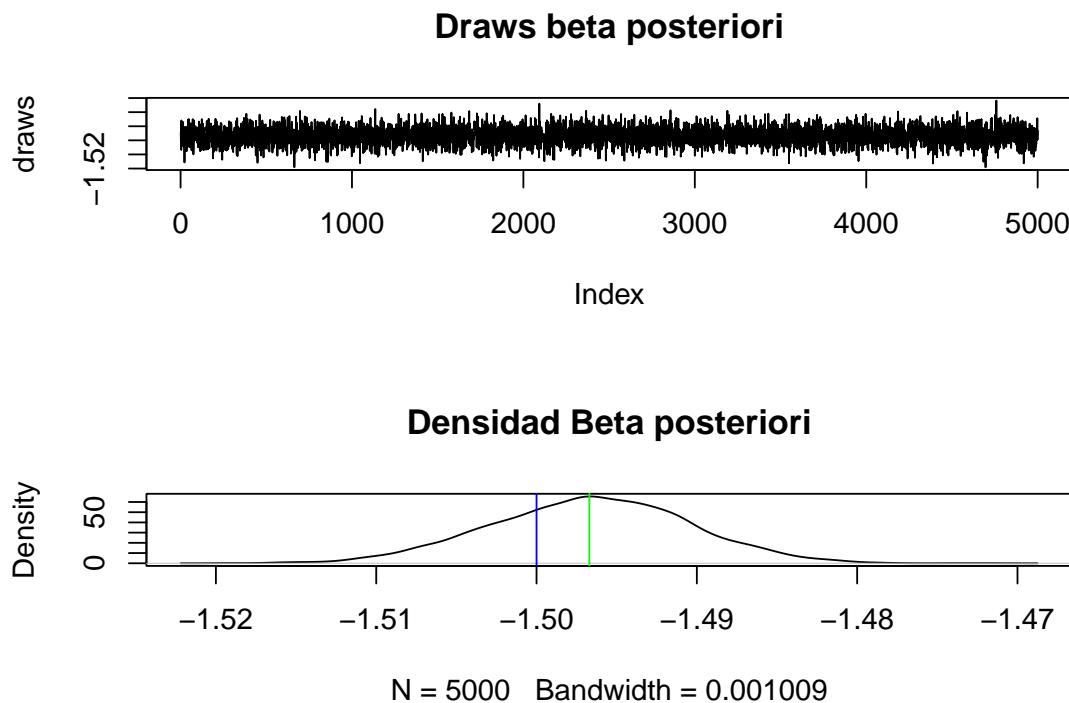


Figura 3.27: Muestreo de beta modelo BVECM para datos simulados de 1000 observaciones.

Ahora, vamos a estimar estos modelos para el conjunto de datos simulado con 100 observaciones. Hay que destacar que este conjunto de datos también sigue la estructura definida en 2.24.

La estimación de la relación de largo plazo es bastante acertada al igual que los términos de corrección de equilibrio. Para la relación autorregresiva, las estimaciones del primer retardo son bastante buenas pero para el segundo retardo fallan bastante.

Tabla 3.24: Estimación matriz var-cov modelo VECM(2) para datos simulados con 100 observaciones.

0.5526159	-0.0675504
-0.0675504	1.0575092

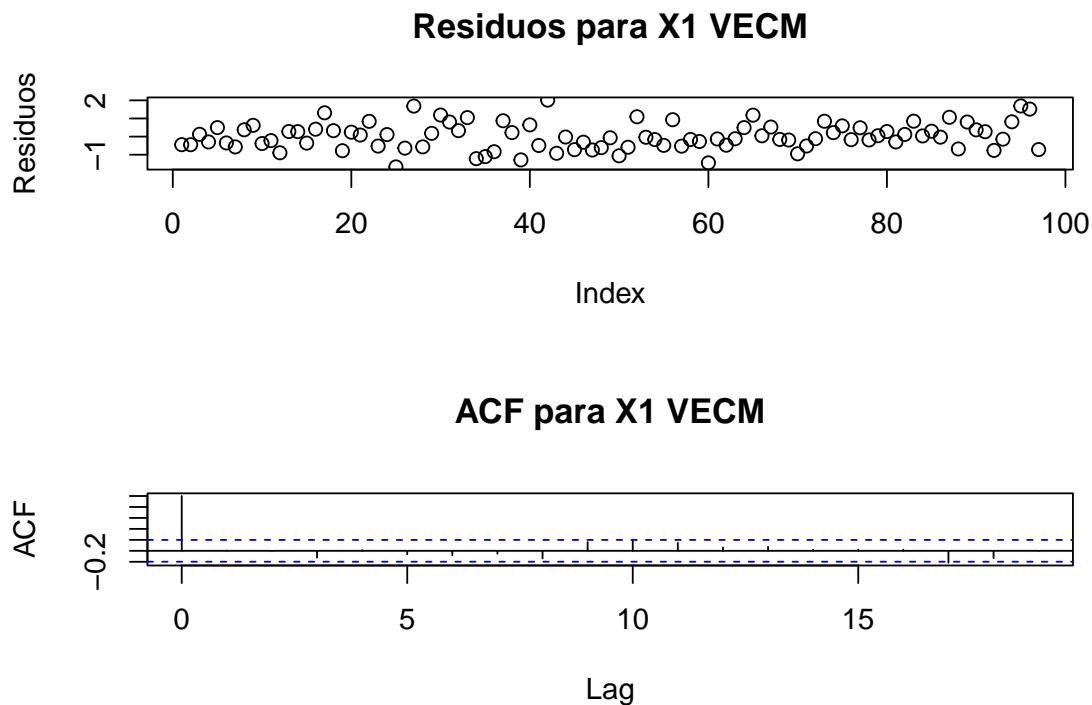


Figura 3.28: Residuos X1 modelo VECM para datos simulados de 100 observaciones.

Tabla 3.25: Estimación relación de largo plazo modelo BVECM(2) para datos simulados con 100 observaciones.

l.Series 1	1.000
l.Series 2	-1.428

Tabla 3.26: Estimación modelo BVECM(2) para datos simulados con 100 observaciones.

	l.Series 1	d.Series 1.l01	d.Series 2.l01	d.Series 1.l02	d.Series 2.l02
d.Series 1	-0.2004860	0.2053551	-0.4146520	0.0197387	0.0056306
d.Series 2	0.2192007	0.3927578	-0.2099023	0.0821863	0.2236160

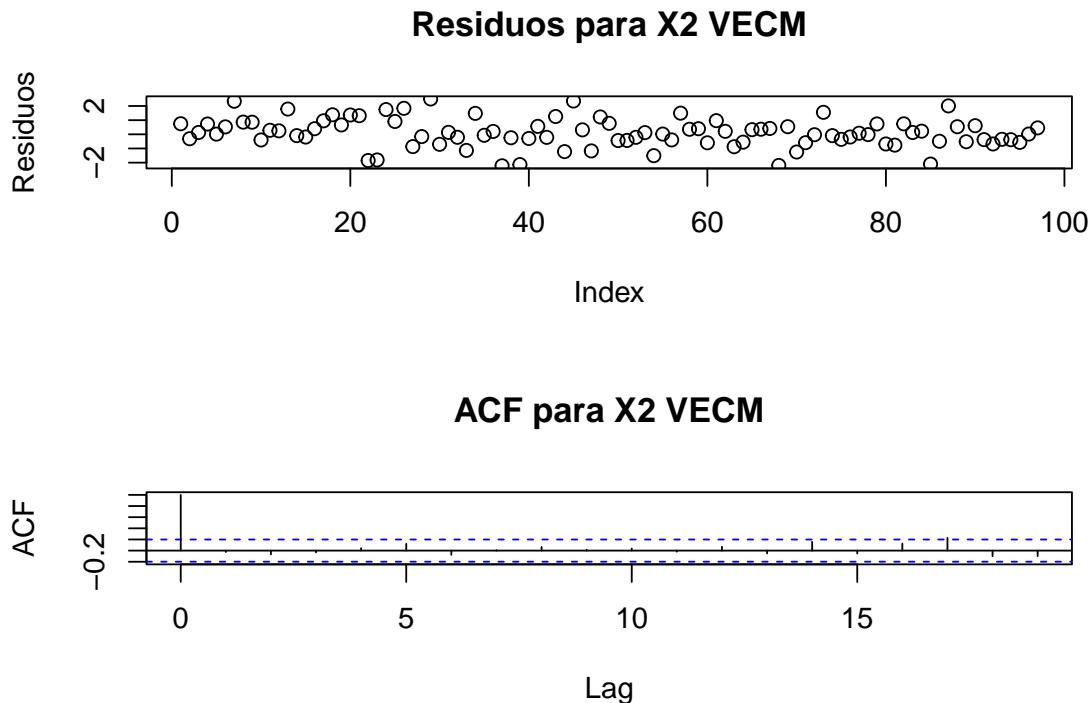


Figura 3.29: Residuos X2 modelo VECM para datos simulados de 100 observaciones.

Los residuos muestran que el modelo está bien estimado. Las discrepancias entre los coeficientes estimados y los valores reales aparecen principalmente porque n no es suficientemente grande como para que mediante un método frecuentista las estimaciones sean precisas.

Vamos a ver cómo funciona el método bayesiano.

Tabla 3.27: Estimación matriz var-cov modelo BVECM(2) para datos simulados con 100 observaciones.

	d.Series 1	d.Series 2
d.Series 1	0.5685477	-0.0743035
d.Series 2	-0.0743035	1.0886847

Tabla 3.28: Estimación relación largo plazo modelo VECM(5) para datos de LIBOR y OVERNIGHT.

	r1
libor	1.0000000
overnight	-0.9943074
const	-0.1405477

La relación a largo plazo se estima con bastante precisión, igual que el método frecuentista, y lo mismo pasa con los términos de corrección de equilibrio. En cuanto a los coeficientes autorregresivos, pasa lo mismo que con el método frecuentista: para el primer retardo las estimaciones son bastante buenas, pero para el segundo retardo están lejos del valor real.

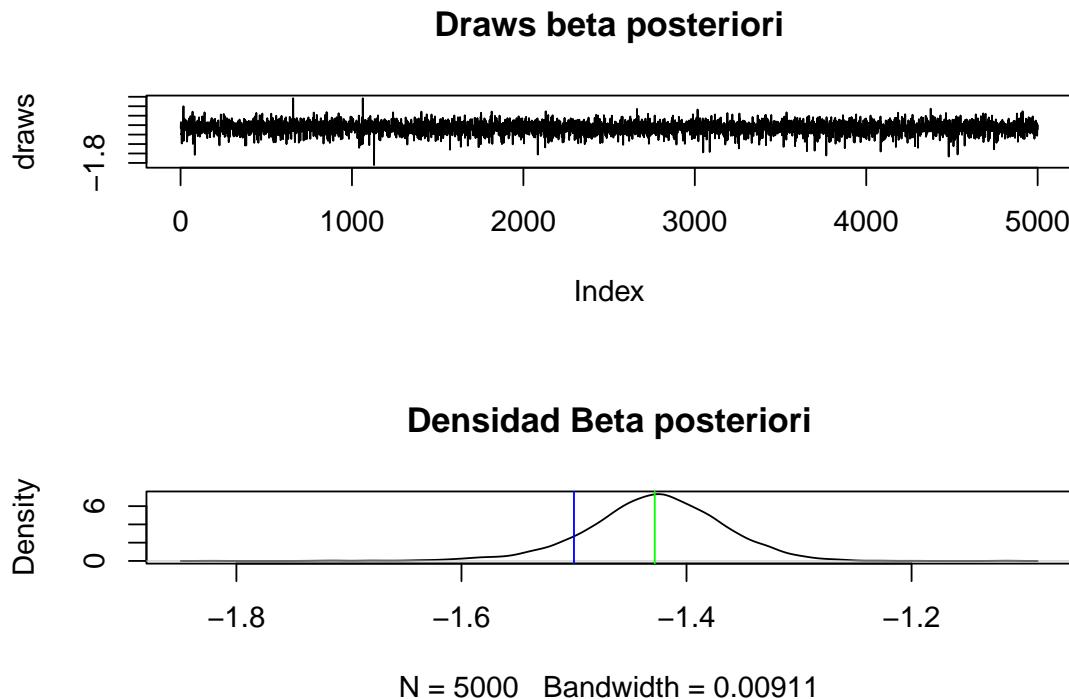


Figura 3.30: Muestreo de beta modelo BVECM para datos simulados de 100 observaciones.

Por último, vamos a estimar estos modelos para los datos macroeconómicos de LIBOR y OVERNIGHT. Después de estimar varios modelos con diferente números de retardos, el modelo que menor AIC tiene incorpora 5 retardos, una constante en la relación de largo plazo y sin otros términos deterministas en la parte de largo plazo o en la de corto plazo.

La estimación de los coeficientes sugiere que la relación a largo plazo para el modelo de LIBOR no es significativa ya que su término de corrección de equilibrio es cero. En cambio, la relación a largo plazo para OVERNIGHT sí es significativa, tanto para su término de corrección de equilibrio como para β . El modelo estimado sugiere que hay

Tabla 3.29: Estimación del modelo VECM(5) para datos de LIBOR y OVERNIGHT.

	Equation libor	Equation overnight
ECT	-0.0073746	0.0659235
libor -1	-0.4315390	0.0176173
overnight -1	0.6463308	-0.1356712
libor -2	-0.2585935	0.0277717
overnight -2	0.4563416	0.0144257
libor -3	-0.1773171	0.0017362
overnight -3	0.2770738	-0.0862848
libor -4	-0.1266505	-0.0253111
overnight -4	0.2468885	0.0751321
libor -5	-0.0435998	0.0000740
overnight -5	0.1091049	-0.0286700

Tabla 3.30: Estimación matriz var-cov modelo VECM(5) para datos de LIBOR y OVERNIGHT.

0.0005011	0.0004054
0.0004054	0.0011451

una relación a largo plazo positiva entre las variables. En cuanto a la relación de corto plazo, recogida en los términos autorregresivos, solo son significativos el primer, tercer y cuarto retardo de OVERNIGHT por lo que el modelo sugiere que LIBOR está relacionado a largo plazo con OVERNIGHT pero que no le influye en el corto plazo.

En cambio, la relación estimada para LIBOR es totalmente diferente. Según los coeficientes que se han estimado, el modelo sugiere que no existe relación a largo plazo pero sí hay a corto plazo con retardos significativos 5 períodos.

Los residuos recogidos en 3.31 y 3.32 sugieren que el modelo está bien estipulado.

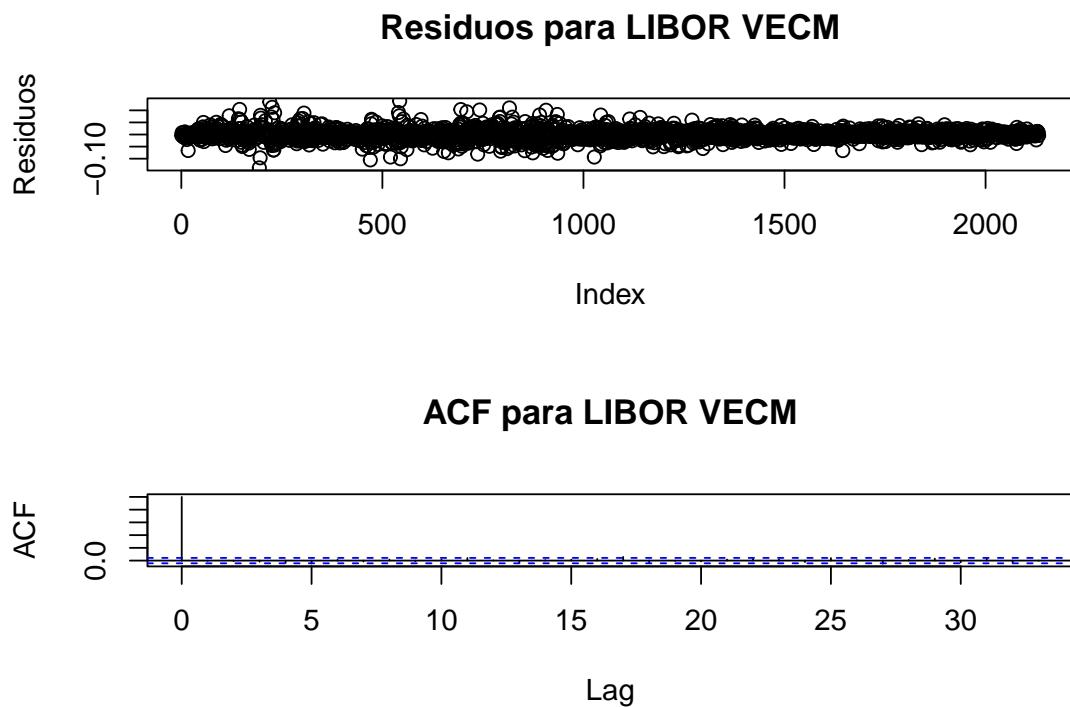


Figura 3.31: Residuos de LIBOR modelo VECM.

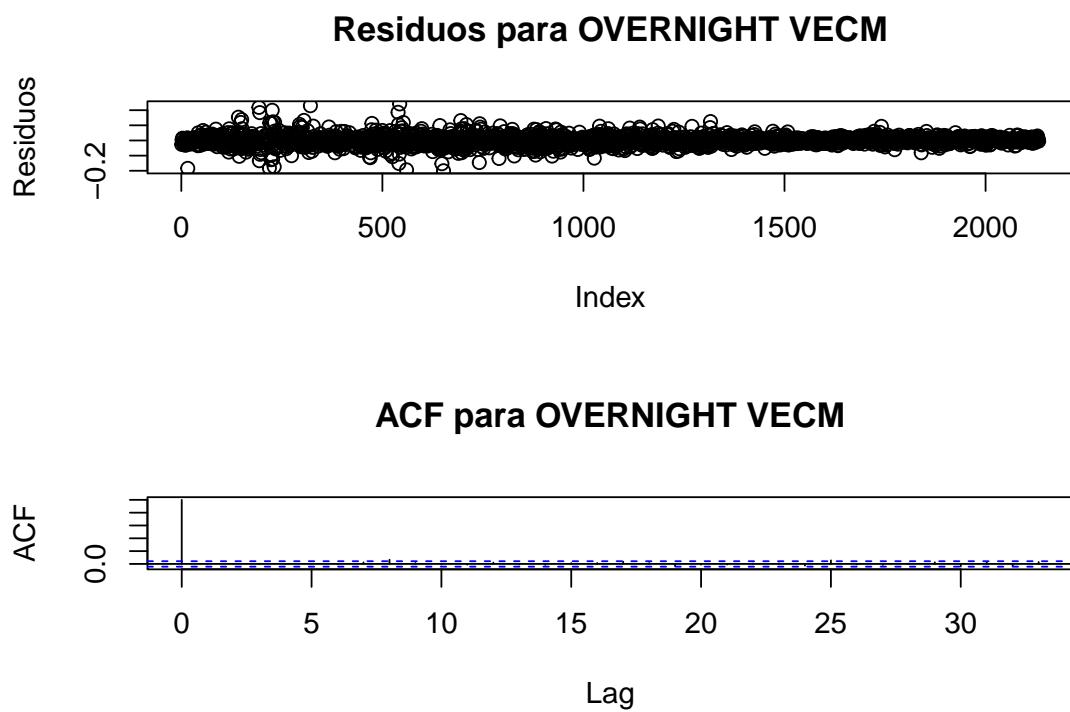


Figura 3.32: Residuos OVERNIGHT modelo VECM.

Por otro lado, la estimación de manera bayesiana con 5 retardos, constante en la relación

Tabla 3.31: Estimación relación de largo plazo modelo BVECM(5) para datos de LIBOR y OVERNIGHT.

l.libor	1.000
l.overnight	-0.994
const	-0.141

Tabla 3.32: Estimación modelo BVECM(5) para datos de LIBOR y OVERNIGHT

	d.libor	d.overnight
l.libor	-0.0072606	0.0639565
d.libor.l01	-0.4317697	0.0197076
d.overnight.l01	0.6465151	-0.1378444
d.libor.l02	-0.2589432	0.0287119
d.overnight.l02	0.4564390	0.0132412
d.libor.l03	-0.1778088	0.0023591
d.overnight.l03	0.2773270	-0.0866676
d.libor.l04	-0.1262717	-0.0239807
d.overnight.l04	0.2471502	0.0752002
d.libor.l05	-0.0438039	0.0007513
d.overnight.l05	0.1085678	-0.0292495

a largo plazo y sin otros términos deterministas es la siguiente:

Tabla 3.33: Estimación matriz var-cov modelo BVECM(5) para datos de LIBOR y OVERNIGHT.

	d.libor	d.overnight
d.libor	0.0005012	0.0004053
d.overnight	0.0004053	0.0011465

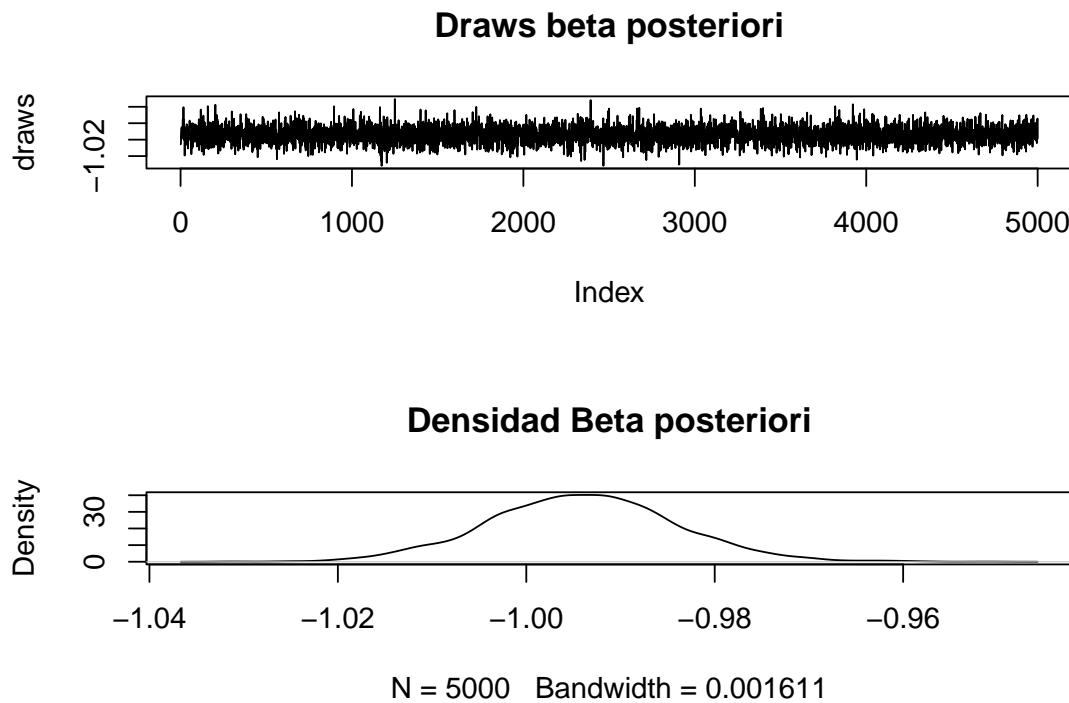


Figura 3.33: Muestreo de beta modelo BVECM para los datos de LIBOR y OVERNIGHT.

La figura 3.33 contiene los draws y densidad a posteriori de β para los datos de LIBOR y OVERNIGHT.

Se puede ver como las estimaciones con el método bayesiano son muy parecidas a las estimaciones frecuentistas, y sugiere también que OVERNIGHT solo depende de su pasado en el corto plazo, pero se ve influenciada por LIBOR en el largo plazo.

3.3. Análisis de la capacidad predictiva

Por último, se presentan los resultados de la capacidad predictiva de cada método. Para analizar la capacidad predictiva, se va a realizar un *walk-forward validation*. Se parte de un modelo inicial estimado con un número (normalmente pequeño) de observaciones t y se predice el valor en $t + 1$. Una vez estimado el valor, se calcula el error cuadrático de la estimación y esa observación pasa a formar parte de los datos con los que se estimará otra vez el modelo. Se repite este proceso hasta haber recorrido todos los datos que se tengan. Al final se obtiene el error de predicción como la media de los errores cuadráticos de cada observación predicha.

Como hemos visto en el apartado anterior, estimar los modelos de forma bayesiana mediante MCMC es computacionalmente mucho más costoso que de forma frecuentista (más de 100 veces más costoso). Por esto, ha sido necesario modificar un poco el algoritmo de *walk-forward validation* para que sea viable con los recursos informáticos que dispongo.

Tabla 3.34: Media de MSRE en 30 iteraciones con datos simulados.

	Variable	
	X1	X2
VAR	1.22794	1.23896
BVAR	1.24198	1.24056

Tanto para los modelos BVAR como para los BVECM, la estimación del modelo se ha de hacer con 50000 iteraciones y 25000 *burning* para asegurar la convergencia. Cada modelo que se estima tarda entre 45 segundos y 70 segundos. Además, la predicción de observaciones también es más lenta que para los modelos frequentistas por lo que en general, estamos hablando que estimar un modelo y hacer una predicción tarda más o menos 1 minuto y medio. Hacer esto 1000 veces para 4 conjuntos de datos diferentes supondría 100h de cómputo.

La estrategia que se ha decidido tomar para reducir el tiempo de cómputo ha sido la siguiente: El modelo inicial contendrá un número de observaciones iniciales tal que el número de iteraciones del *walk-forward validation* sea de 30 bajo la condición de que si las series son diarias se predirá a un horizonte temporal de $t + 30$, y si son trimestrales de $t + 4$ en vez de ser $t + 1$ el horizonte temporal.

Esta decisión se ha tomado para que haya 30 estimaciones de *MRSE* y se pueda aproximar a una distribución normal para sacar intervalos de confianza, y para que el tiempo de cómputo por cada conjunto de datos no sea superior a 1 hora (4 horas en total). Además, tiene sentido usar un horizonte temporal de $t + 30$ para series diarias y $t + 4$ para las trimestrales por que lo que se está haciendo es estimar a un mes vista para las series diarias, y a un año vista en las series trimestrales.

La tabla 3.34 muestra la media de *MSRE* para las 30 iteraciones del algoritmo *walk-forward validation* para obtener el error de predicción de cada modelo para cada variable. Se puede apreciar que la capacidad predictiva del modelo bayesiano y el frequentista es igual de buena. Esto se debe principalmente a que el *walk-forward* se ha iniciado con 100 observaciones y hay un total de 1000 datos en el dataset por lo que los datos tienen mucho peso en la distribución a posteriori dado la gran cantidad de datos.

En la tabla 3.35, tenemos los resultados del *walk-forward validation* para los datos macroeconómicos de inflación y desempleo en Estados Unidos desde 1959-2007. Las series de este conjunto de datos son trimestrales por lo que el algoritmo se ha configurado de tal forma que en cada iteración prediga un periodo temporal de $t + 4$ que viene a ser un año. El algoritmo se ha iniciado con 75 observaciones de 195 en total. De nuevo, como el tamaño de los datos es elevado, estos tienen mayor peso que las distribuciones a priori y los resultados sobre la capacidad predictiva indican que los modelos predicen igual de bien los valores futuros en ambas series.

En cuanto a los conjuntos de datos con series cointegradas, para el conjunto de datos que ha sido simulado, la tabla 3.36 refleja la capacidad predictiva del modelo bayesiano y frequentista al realizar un *walk-forward validation* con predicciones a $t + 30$. El conjunto de datos tiene 1000 observaciones y se ha iniciado el algoritmo con 100. Una vez más, los resultados son prácticamente idénticos debido a que el peso de los datos es muy alto a la hora de sacar las estimaciones a posteriori.

Tabla 3.35: Media de MSRE en 30 iteraciones para datos macroeconómicos USA 1959-2007.

Variable		
	Inflación	Desempleo
VAR	0.39388	0.21507
BVAR	0.39598	0.21350

Tabla 3.36: Media de MSRE en 30 iteraciones para datos simulados.

Variable		
	X1	X2
VECM	5.408236	3.818258
BVECM	5.407065	3.815903

Por último, para las series cointegradas de LIBOR y OVERNIGHT, la tabla 3.37 recoge el error de predicción para cada serie según cada tipo de modelo a través de un *walk-forward validation*, con periodo temporal de predicción de $t + 30$ en cada iteración. Aún hay pequeñas diferencias, pero no llegan a ser significativas y los modelos presentan una capacidad predictiva similar para este conjunto de datos reales.

Tabla 3.37: Media de MSRE en 30 iteraciones para datos de LIBOR y OVERNIGHT.

Variable		
	LIBOR	OVERNIGHT
VECM	0.0739661	0.0638624
BVECM	0.0784036	0.0712384

Capítulo 4

Conclusiones

4.1. Motivación de la inferencia bayesiana

Los modelos VAR y VECM son modelos muy populares y muy usados desde finales del siglo XX. Su principal ventaja es la facilidad de implementación. Sólo hace falta relacionar una serie de variables que expliquen el comportamiento de las otras y determinar el número de retardos para incorporar en el modelo.

Aún así, es normal incluir un número elevado de variables en estos modelos, y, también, un número elevado de retardos. Esto hace que solamente el número de coeficientes autorregresivos que se tengan que estimar en un modelo VAR sea del orden $p \times l$ siendo p el número de variables y l el número de retardos que se incorporan en el modelo. Este es el problema de *sobreparametrización* que genera unas estimaciones inestables con errores estándar elevados.

Los métodos bayesianos ofrecen una solución a este problema ya que al incorporar información a priori, las estimaciones dejan de ser inestables y los errores estándar se reducen. Además, otra característica positiva es que se obtienen distribuciones de probabilidad para los parámetros estimados independientemente del tamaño muestral.

4.2. Problemas en la inferencia bayesiana

Si la información a priori que se incorpora en los modelos bayesianos es acertada, las estimaciones frente a métodos frequentistas son más robustas. Además, si el tamaño muestral es pequeño, aparte de ser más robustas pueden ser más precisas ya que la estimación a posteriori no se ve tan influenciada por los datos extremos.

Esto significa que la calidad de las estimaciones con métodos bayesianos depende fuertemente de la información a priori que se incorpore, sobre todo con pocos datos. Hemos visto los modelos BVAR donde no hay un consenso sobre la distribución a priori que se debe utilizar, y por ello surgen los métodos jerárquicos que tratan a las a priori como hyperparámetros con sus respectivas hyperpriors. Este enfoque obtiene estimaciones más robustas pero aún así, dependen de la información a priori que se introduzca en los hiperparámetros; aunque, en general, no hace falta tener información muy precisa para que funcionen bien y converjan.

En cuanto a los modelos de corrección de equilibrio, también es muy importante la información a priori, además, existe un problema a la hora de estimar el modelo ya que al incorporar la relación de largo plazo, la matriz que recoge esta relación suele ser de rango reducido y causa problemas de identificación. Por un lado, hay problemas de identificación global que se han ido solucionando con la normalización lineal, y, por otro, hay problemas de identificación local donde una distribución a priori no informativa genera una distribución a posteriori impropia.

Koop et al. (2010) proponen una solución a estos problemas asociados a la estimación bayesiana de modelos de corrección de equilibrio, que resulta en una estimación que no impone restricciones en el espacio de cointegración, y que asegura la convergencia del muestreador de Gibbs empleado para la obtención de las distribuciones a posteriori. Véase 2.12 a 2.21.

4.3. Estimación de los coeficientes

La estimación de estos modelos se ha aplicado a dos conjuntos de datos simulados con 1000 observaciones, dos conjuntos de datos simulados con 100 observaciones y dos conjuntos de datos reales. En cada par de conjuntos, uno contiene variables cointegradas y otro variables no cointegradas.

Para los datos simulados con 1000 observaciones, las estimaciones bayesiana y frecuentista son muy similares y cercanas a los valores reales de los parámetros, tanto para la parte autorregresiva, como para la parte de largo plazo en el modelo de corrección de equilibrio. La semejanza en las estimaciones, se debe en gran parte a la cantidad de los datos que tienen mucho peso en la estimación a posteriori de los modelos bayesianos.

El conjunto de datos simulados con 100 observaciones de variables no cointegradas, ha demostrado que las estimaciones bayesianas pueden mejorar las estimaciones frecuentistas cuando el tamaño muestral es reducido. El método frecuentista en general estima bastante mal los coeficientes autorregresivos y el método bayesiano, aunque tampoco es muy preciso, mejora las estimaciones para el modelo de la primera variable. La estimación de la matriz de varianzas-covarianzas en general es bastante mala para ambos métodos.

Las estimaciones para los datos simulados de 100 observaciones con variables cointegradas han sido muy similares. La relación a largo plazo está bastante cerca de la real (-1.4 estimado y -1.5 real) en ambos métodos, la parte autorregresiva para el primer retardo está bien estimada en ambos métodos y para el segundo retardo ambos métodos fallan bastante. El método frecuentista es normal que falle con pocos datos, y el método bayesiano en general lo hace bien porque acierta con la estimación de largo plazo y la parte autorregresiva del primer retardo, pero falla para el segundo retardo, seguramente porque las distribuciones a priori debían haberse especificado mejor para incorporar más información al modelo.

En cuanto a los datasets reales, en general las estimaciones han sido muy similares y en gran parte se debe a que los datasets contienen bastantes datos lo que hace que en la estimación a posteriori pese mucho esta parte.

4.4. Capacidad predictiva

Por último, para la capacidad predictiva se ha aplicado un algoritmo *walk-forward validation*, para los datasets reales. En general, la capacidad predictiva medida con la media del MRSE de las iteraciones del algoritmo es muy similar para ambos métodos. Las causas de esto son principalmente:

1. Debido a limitaciones computacionales, el número de observaciones iniciales con la que se ha empezado el algoritmo ha sido grande en general lo que hace que los datos pesen bastante en las estimaciones a posteriori de los parámetros y por ende en las predicciones.
2. Por limitaciones computacionales, se ha modificado el algoritmo, y en vez de iterar para cada observación fuera del conjunto inicial, en cada iteración se han hecho predicciones a $t + 30$ y $t + 4$ según si los datos eran diarios o trimestrales. Esto hace que las predicciones lejanas, sean muy similares en ambos métodos y por ello el MRSE muy similar.

Apéndice A

Apéndice:

A.1. Primera sección

Apéndice B

Apéndice: Título del Apéndice

B.1. Primera sección

Bibliografía

- M De Arriba Cotillas. "apuntes método econométricos uc3m". Apuntes no transferibles, 2021.
- D. Giannone, M. Lenza, and G Primiceri. Prior selection for vector autoregressions. *Review of Economics and Statistics*, 97, 01 2012. doi: 10.1162/REST_a_00483.
- G. Koop, R. Strachan, H. Van Dijk, and M Villani. Bayesian Approaches to Cointegration. Discussion Papers in Economics 04/27, Division of Economics, School of Business, University of Leicester, September 2004. URL <https://ideas.repec.org/p/lec/leecon/04-27.html>.
- G. Koop, R. León-González, and R Strachan. Efficient posterior simulation for cointegrated models with priors on the cointegration space. *Econometric Reviews*, 29 (2):224–242, 2009. doi: 10.1080/07474930903382208. URL <https://doi.org/10.1080/07474930903382208>.
- P Luque-Calvo. *Escribir un Trabajo Fin de Estudios con R Markdown*, 2017.
- P Luque-Calvo. *Cómo crear Tablas de información en R Markdown*, 2019.
- D Peña. *Análisis de series temporales*. Alianza, Madrid, 2a ed. edition, 2010. ISBN 9788420669458.
- J Praser. Data-based priors for vector error correction models. *International Journal of Forecasting*, 2021. ISSN 0169-2070. doi: <https://doi.org/10.1016/j.ijforecast.2021.10.007>. URL <https://www.sciencedirect.com/science/article/pii/S0169207021001709>.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. URL <https://www.R-project.org/>.
- Techopedia. "definition - what does business intelligence (bi) mean?". Disponible en <https://www.techopedia.com/definition/345/business-intelligence-bi>.
- H. Wickham and G Grolemund. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly, first edition edition, 2017.
- Y Xinhua. Risk interactions of coronavirus infection across age groups after the peak of covid-19 epidemic. *International Journal of Environmental Research and Public Health*, 17(14), 2020. ISSN 1660-4601. doi: 10.3390/ijerph17145246. URL <https://www.mdpi.com/1660-4601/17/14/5246>.
- W. Yangru and Z Xing. *VAR Models: Estimation, Inferences, and Applications*, pages 1391–1398. Springer US, Boston, MA, 2010. ISBN 978-0-387-77117-5. doi: 10.1007/978-0-387-77117-5_93. URL https://doi.org/10.1007/978-0-387-77117-5_93.