Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Univesidad de Bío-Bío
Facultad de Ciencias
Departamento de Estadística y Departamento de Matemática
Programa de Magíster en Matemática,
mención Matemática Aplicada o
mención Estadística.

Modelo Dinámico Multivariado de Localización con Intervención: Una aplicación a variables macroeconómicas

Tesis presentada por

Javier Ignacio Castillo Cruces

al Programa de Magíster en Matemática, Mención Matemática Aplicada o Mención Estadística para la obtención del grado de

Magíster en Matemática con Mención en Estadística.

Directora: Dra. Verónica Anaya Domínguez

Marzo de 2022

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Modelo Dinámico Multivariado de Localización con Intervención: Una aplicación a variables macroeconómicas

TESIS PRESENTADA POR

Javier Ignacio Castillo Cruces

Tesis presentada al Programa de Magíster en Matemática, Mención Estadística del Departamento de Estadística y Departamento de Matemática de la Universidad del Bío Bío como parte de los requisitos para la obtención del grado de Magíster en Matemática con Mención en Estadística de la Universidad del Bío-Bío.

Aprobada por:

Francisco Novoa Muñoz
Director
Departamento de estadística, Universidad del Bío-Bío

Sergio Contreras Espinoza
Co-Director
Departamento de estadística, Universidad del Bío-Bío

Christian Caamaño Carrillo
Profesor Informante
Departamento de estadística, Universidad del Bío-Bío

Bernardo Lagos Álvarez
Profesor Informante
Departamento de estadística , Universidad de Concepción

Universidad del Bío-Bío
Facultad de Ciencias
Departamento Estadística y Departamento de Matemática

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Resumen

En esta tesis estudiamos y aplicamos una clase de modelos de series de tiempo guiados por observación denominados Modelos Dinámicos Condicionales de Score (DCS). El mecanismo para actualizar los parámetros a lo largo del tiempo es la función score de la función de verosimilitud. Este enfoque proporciona un marco unificado y consistente para introducir parámetros variables en el tiempo en una amplia clase de modelos no lineales resistentes a las observaciones atipicas, lo que mejora su rendimiento estadístico en comparación con los modelos clásicos. Para el modelo Seasonal-t-QVAR el cual es un modelo dinámico multivariante de localización guiado por la score, se presenta la formulación econométrica, el cálculo de la función de respuesta al impulso (IRF), la estimación de Máxima Verosimilitud (MV) y las condiciones de las propiedades asintóticas de MV que incluyen la invertibilidad. La aplicación presentada compara los modelos Seasonal-t-QVAR con el modelo Seasonal-Gaussian-VARMA (VAR Moving Average) o modelo clásico. Además, se incorpora a estos modelos un análisis de intervención para capturar los efectos de la pandemia producidos en las variables macroeconómicas de inflación y desempleo de Estados Unidos para el período febrero de 1948 a febrero de 2022.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Abstract

In this thesis we study and apply a class of observation-driven time series models called Dynamic Conditional Score Models (DCS). The mechanism for updating the parameters over time is the score function. This approach provides a unified and consistent framework for introducing time-varying parameters into a broad class of nonlinear models resistant to outliers, improving their statistical performance compared to classical models. For the Seasonal-t-QVAR model, which is a dynamic multivariate localization model drive by the score, is presented the econometric formulation, the calculation of the impulse response function (IRF), the Maximum Likelihood estimation (ML) and the conditions of the asymptotic properties of MV that include invertibility. The presented application compares the Seasonal-t-QVAR models with the Seasonal-Gaussian-VARMA model (VAR moving average) or classic model. In addition, an intervention analysis is incorporated into these models to capture the effects of the pandemic produced on the macroeconomic variables of inflation and unemployment in the United States for the period from February 1948 to February 2022.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Índice general

Resumen **v**

Abstract **vii**

1. Introducción **1**
1.1. Componentes no obervados y filtros 2
1.2. Independencia, ruido blanco y diferencias de martingalas 5
1.2.1. La ley de esperanzas iteradas y predicciones óptimas 6
1.2.2. Definiciones y propiedades 6
1.3. Planificación 8

2. Distribuciones estadísticas y teoría asintótica **11**
2.1. Distribuciones 12
2.1.1. Distribución t-Student 12
2.1.2. Distribución General de Error 13
2.1.3. Distribución Beta 14
2.1.4. Distribución Gamma 17
2.2. Máxima Verosimilitud 18
2.2.1. Distribución t-Student 20
2.2.2. Distribución General de Error (GED) 22
2.2.3. Distribución Gamma 23
2.2.4. Consistencia y normalidad asintótica 24
2.3. Estimacion máximo verosimil para modelos Dinamicos Condicionales de Score 27
2.3.1. Lema de la matriz de información 27
2.3.2. Matriz de información para modelos de primer orden 29
2.3.3. Matriz de información con la parametrización δ 34
2.3.4. Distribución asintótica 35

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

2.3.5. Consistencia y normalidad asintótica 35
2.3.6. No estacionariedad 41
2.3.7. Varios parámetros 42
2.4. Modelos de orden superior 45
2.5. Pruebas 49
2.5.1. Correlación serial 49
2.5.2. Bondad de ajuste de las distribuciones 52
2.5.3. Residuos 53
2.5.4. Ajuste del modelo 54
2.6. Variables explicativas 54

3. Localización **57**
3.1. Modelos de Localización Dinámicos t-student 58
3.2. Propiedades básicas 59
3.2.1. Generalización y forma reducida 60
3.2.2. Momentos de las observaciones 61
3.2.3. Función de autocorrelación 61
3.3. Estimación por máxima verosimilitud 62
3.3.1. Distribución asintótica del estimador de máxima verosimilitud . 62
3.3.2. Experimento Montecarlo 66
3.3.3. Aplicación 67
3.4. Restricción de parámetros 67
3.5. Modelos de orden superior y modelos de espacio estado 69
3.5.1. Modelos lineales gaussianos y el Filtro de Kalman 69
3.5.2. Modelo DCS 72
3.6. Tendencia y estacionalidad 73
3.6.1. Modelo local level 74
3.6.2. Modelo de tendencia lineal local 75
3.6.3. Modelo estacionalidad estocastica 76
3.7. Suavizado 78
3.7.1. Pesos 78
3.8. Pronostico 79
3.8.1. Modelos QARMA 79
3.8.2. Forma de espacio estado 80

4. Modelos Multivariantes **83**
4.1. Distribución Multivariante 84
4.1.1. Estimación 84
4.1.2. Regresión 85
4.2. Modelo de localización multivariante 85

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

4.2.1. Modelos estructurales de series de tiempo 85
4.2.2. Modelo DCS para una t multivariante 87
4.2.3. Teoría asintótica 88
4.2.4. Regresión y errores en variables 89

5. Modelos de localización multivariantes para variables macroeconómicas 91
5.1. Modelo Seasonal-QVAR 91
5.1.1. Inferencia estadística del modelo Seasonal-QVAR 93
5.2. Modelo Gaussiano Seasonal-VARMA 95

6. Intervención 97
6.1. Variables impulso 97
6.2. Variable escalón 98
6.3. Relación entre impulsos y escalones 99

7. Resultados 101
7.1. Aplicación del modelo Seasonal-t-QVAR 101
7.1.1. Resultados de la estimación 102
7.2. Conclusiones 109

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Capítulo 1

Introducción

En este trabajo se presenta una teoría unificada y completa para una clase de modelos de series de tiempo no lineales que pueden manejar distribuciones dinámicas. El énfasis está en los modelos en los que la distribución condicional a observaciones pasadas puede ser de cola pesada y la localización varía con el tiempo. La característica definitoria de estos modelos es que la dinámica está impulsada por la score de la distribución condicional. Cuando se emplea una función de enlace adecuada para el parámetro variante en el tiempo, se pueden derivar expresiones analíticas para momentos incondicionales, autocorrelaciones y momentos de pronósticos de varios pasos. Además, se puede desarrollar una teoría de distribución asintótica completa para estimadores de Máxima Verosimilitud (MV), incluidas expresiones analíticas para matrices de covarianza asintótica de los estimadores.

La clase de lo que llamaremos Modelos Dinámicos Condicionales de Score (DCS) incluye modelos de series temporales lineales estándar observados con un error que puede estar sujeto a valores atípicos, modelos que capturan la varianza condicional cambiante y modelos para variables no negativas. Un principio rector que subyace a la clase de modelos propuestos es el de la extracción de señales. Cuando se combina con ideas básicas de estimación de MV, el enfoque de extracción de señales conduce a modelos que, a diferencia de muchos en la literatura, son relativamente simples y producen expresiones analíticas para sus características principales como los momentos de la dsitribución, matriz de información y matriz de varianza-covarianza de los estimadores MV. Para estimar la localización, los modelos DCS están estrechamente relacionados con los modelos de componentes no observados descritos en Harvey (1989). Dichos modelos se pueden manejar usando métodos de espacio de estado. Para estimar la escala, los modelos se acercan a los modelos de volatilidad estocástica, en los que la varianza se trata como un componente no observado.

La referencia a la localización y la escala en lugar de la media y la varianza es

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

deliberada. La localización y la escala se aplican a todas las distribuciones, mientras que la media y la varianza pueden no existir siempre, un punto que es particularmente relevante cuando se trata de colas pesadas. Además, aunque el conocimiento de la media y la varianza de una distribución gaussiana nos dice todo lo que hay que saber, este no es el caso con muchas otras distribuciones. Es imprudente centrar demasiada atención en la media y la varianza, especialmente en la econometría financiera. De manera similar, la correlación mide la fuerza de la relación entre dos variables en un mundo gaussiano, pero el concepto más general de asociación tiene una relevancia más amplia.

1.1. Componentes no obervados y filtros

Los modelos Autorregresivos Integrados de Media Móvil (ARIMA) se enfoca en pronosticar valores futuros de una serie. El paradigma de señal más ruido proporciona un marco más general. La extracción de señal es de interés en si misma, y una vez que se ha resulto el problema, sigue la solución de pronóstico.

Un modelo gaussiano simple de señal más ruido para una muestra de T observaciones, y_t, $t = 1, \ldots, T$, es

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2), \quad t = 1, \ldots, T, \tag{1.1}$$
$$\mu_{t+1} = \phi\mu_t + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2),$$

donde ϕ es el parámetro autorregresivo, las perturbaciones irregulares y de señal ε_t y η_t respectivamente, son mutuamente y serialmente independientes entre sí, y la notación $NID(0, \sigma^2)$ denota distribución normal e independiente con media cero y varianza σ^2. La relación señal/ruido, $q = \sigma_\eta^2/\sigma_\varepsilon^2$, juega un papel clave para las extracciones de señales. La forma reducida de (1.1) es un proceso $ARMA(1, 1)$,

$$y_t = \phi y_{t-1} + \xi_t - \theta\xi_{t-1}, \quad \xi_t \sim NID(0, \sigma^2) \quad t = 1, \ldots, T, \tag{1.2}$$

pero con restricciones en θ. Por ejemplo, cuando $\phi = 1$, $0 \leq \theta \leq 1$. Los pronósticos del modelo de componentes no observados y la forma reducida son los mismos. Es posible una aproximación autorregresiva a la forma reducida, pero, si q está cerca de cero, puede ser necesario un gran número de rezagos para que la aproximación produzca pronósticos aceptables.

El modelo de componentes no observados en (1.1) está efectivamente en forma de espacio de estado, y como tal, puede ser manejado por el Filtro de Kalman; véase Harvey (1989). Los parámetros ϕ y $q = \sigma_\eta^2/\sigma_\varepsilon^2$ pueden estimarse por MV, con la función de verosimilitud construida a partir de los errores de predicción de un paso adelante. El Filtro de Kalman se puede expresar como una sola ecuación que combina el

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

estimador de μ_t basado en la información en el tiempo $t-1$ con la t-ésima observación para producir el mejor estimador de μ_{t+1}. Al escribir esta ecuación junto con una ecuación que define el error de predicción de un paso adelante, v_t, se obtiene la forma de innovación del Filtro de Kalman:

$$\begin{aligned} y_t &= \mu_{t|t-1} + v_t, \quad t = 1, \ldots, T, \\ \mu_{t+1|t} &= \phi\mu_{t|t-1} + k_t v_t. \end{aligned} \tag{1.3}$$

La ganancia de Kalman, k_t, depende de ϕ y $q = \sigma_\eta^2/\sigma_\varepsilon^2$. En estado estacionario, k_t es constante. Establecerlo igual a un parámetro, κ, y reorganizar da el modelo ARMA, (1.2), con $\xi_t = v_t$ y $\phi - \kappa = \theta$. Un modelo autorregresivo puro (AR) es un caso especial en el que $\kappa = \phi$, de modo que $\mu_{t|t-1} = \phi y_{t-1}$.

Ahora suponga que el ruido ε_t en un modelo de componentes no observados proviene de una distribución de colas pesadas, como la t de Student. Tal distribución puede dar lugar a observaciones que, cuando se comparan con el criterio de una distribución gaussiana, se consideran valores atípicos. En el caso de (1.1), la forma reducida sigue siendo un proceso $ARMA(1,1)$, pero con perturbaciones ε_t que, aunque no están correlacionadas en serie, no se distribuyen de manera independiente e idéntica. Permitir que las perturbaciones tengan una distribución de colas pesadas no soluciona el problema. Un valor grande de ε_t solo afecta la observación actual, pero en la forma reducida, se incorpora al nivel y toma tiempo trabajar a través del sistema. Para ser específicos, la representación AR de un proceso $ARMA(1,1)$ es

$$y_t = (\phi - \theta)\sum_{j=1}^{\infty} \phi^{j-1} y_{t-j} + \xi_t = \mu_{t|t-1} + \xi_t.$$

Si la t-ésima observación está contaminada al agregar una cantidad arbitraria, C, entonces, después de τ períodos, la predicción de la siguiente observación todavía está contaminada por C porque contiene la cantidad $(\phi - \theta)\phi^\tau C$.

Un modelo $ARMA$ o AR en el que se permite que las perturbaciones ε_t tengan una distribución de cola pesada está diseñado para manejar los valores atípicos de innovación, a diferencia de los valores atípicos aditivos. Hay una gran cantidad de discusión sobre valores atípicos y cómo manejarlos en la literatura de robustez; ver, por ejemplo, el libro de Maronna et al. (2019) y el artículo de Muler et al. (2009) sobre estimación robusta para modelos $ARMA$. El punto de vista adoptado aquí es que un enfoque basado en modelos no solo es más simple, tanto conceptual como computacionalmente, que los métodos robustos habituales, sino que también es más adecuado para la verificación de diagnóstico y la generalización.

Los métodos de simulación, como la cadena de Markov Monte Carlo (MCMC), el muestreo de importancia y el filtrado de partículas, proporcionan la base para un

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

ataque directo a los modelos que son no lineales y/o no gaussianos. El objetivo es ampliar los algoritmos de filtrado y suavizado de Kalman que han demostrado ser tan efectivos en el manejo de modelos gaussianos lineales. Se han hecho progresos considerables en los últimos años; véanse Robert et al. (2010), Durbin and Koopman (2012) y Koopman et al. (2016). Sin embargo, el hecho es que la estimación basada en simulación puede llevar mucho tiempo y estar sujeta a cierto grado de incertidumbre. Además, las propiedades estadísticas de los estimadores no son fáciles de establecer.

El enfoque aquí comienza escribiendo la distribución de la t-ésima observación, condicionada a observaciones pasadas. Los parámetros que varían con el tiempo se actualizan luego mediante un filtro adecuadamente definido. Tal modelo es lo que Cox et al. (1981) llamó impulsado por la observación. En un modelo de componentes no observado gaussiano lineal, que está controlado por parámetros en la terminología de Cox, el Filtro de Kalman está controlado por el error de predicción de un paso adelante, como en (1.3). El ingrediente principal del filtro desarrollado aquí para distribuciones no gaussianas es el reemplazo de v_t en la ecuación del Filtro de Kalman por una variable, u_t, que es proporcional a la score de la distribución condicional, es decir, el logaritmo de la función de densidad de probabilidad en el tiempo t diferenciado con respecto a $\mu_{t|t-1}$. Así, la segunda ecuación en (1.3) se convierte en

$$\mu_{t+1|t} = \phi\mu_{t|t-1} + \kappa u_t,$$

donde κ se trata como un parámetro desconocido.

¿Por qué la score? Si la señal en (1.1) fuera fija, es decir, $\phi = 1$ y $\sigma_\eta^2 = 0$, entonces $\mu_t = \mu$, la media de la muestra, $\hat{\mu}$, cumpliría la condición

$$\sum_{t=1}^{T}(y_t - \hat{\mu}) = 0.$$

El estimador de MV se obtiene diferenciando la función logarítmica de verosimilitud con respecto a μ y estableciendo la derivada resultante, la score, igual a cero. Cuando las observaciones se distribuyen normalmente, el estimador MV es el mismo que la media muestral, el estimador de momento. Sin embargo, para una distribución no gaussiana, el estimador de momento y el estimador MV difieren. Una vez que la señal en un modelo gaussiano se vuelve dinámica, como en (1.1), su estimación se puede actualizar con cada nueva observación usando el Filtro de Kalman. Con una distribución no normal, la actualización exacta ya no es posible, pero el hecho de que la estimación MV en el caso estático establece la score en cero proporciona una razón para reemplazar el error de predicción, que tiene una media cero, por la score, que para cada observación individual también tiene media cero. El filtro resultante podría, por lo tanto, considerarse como una aproximación a la solución informática intensiva

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

para el modelo de componentes no observadas, y la evidencia presentada más adelante respalda esta noción. El apoyo teórico adicional proviene del enfoque del modo condicional para suavizar modelos no lineales.

El atractivo de tratar el filtro impulsado por la score de la distribución condicional como un modelo por derecho propio es que se hace posible derivar la distribución asintótica del estimador MV y generalizar en varias direcciones. La primera ecuación en (1.3) es entonces no lineal. La justificación de la clase de modelos de score condicional dinámico no es que se aproximen a los modelos de componentes no observados correspondientes, sino que sus propiedades estadísticas son integrales y directas.

El uso de la score de la distribución condicional para robustecer el Filtro de Kalman fue propuesto originalmente por Masreliez (1975). Sin embargo, a menudo se ha argumentado que una suposición crucial hecha por Masreliez (con respecto a la normalidad aproximada del anterior en cada paso de tiempo) es, para citar a Schick and Mitter (1994), 'insuficientemente justificada y sigue siendo controvertida'. Sin embargo, señalan que el procedimiento 'se ha encontrado que funciona bien tanto en estudios de simulación como con datos reales'. Schick and Mitter (1994) sugirieron una generalización del Filtro de Masreliez basada en fundamentos teóricos algo más sólidos. Se supone que el ruido de observación proviene de una distribución normal contaminada, y el estimador resultante emplea bancos de filtros y suavizadores de Kalman ponderados por probabilidades posteriores. Como resultado, es considerablemente más complicado que el Filtro de Masreliez. Una vez que se ha entrado en el campo de las técnicas computacionalmente intensivas, parece mejor adoptar los métodos basados en simulación a los que se aludío anteriormente.

Las situaciones abordadas por Masreliez son más complicadas que las consideradas aquí porque las matrices del sistema en el modelo de espacio de estado pueden variar en el tiempo. Los modelos de este trabajo tienen una estructura más simple y, como resultado, el uso de la partitura para impulsar la dinámica puede basarse en bases estadísticas mucho más sólidas.

1.2. Independencia, ruido blanco y diferencias de martingalas

El estudio de modelos que no son lineales y gaussianos requiere una cuidadosa distinción entre los conceptos de independencia, falta de correlación y diferencias de martingalas. Pero antes de continuar, es necesario establecer algunos resultados estadísticos básicos. Las pruebas se pueden encontrar en muchos textos introductorios de series de tiempo y econometría.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

1.2.1. La ley de esperanzas iteradas y predicciones óptimas

Un elemento clave en algunas de las derivaciones estadísticas que siguen es la ley de esperanzas iteradas. Suponga que es difícil encontrar el valor esperado de una variable aleatoria, y, pero evaluar su esperanza condicional en otra variable aleatoria, x, es sencillo. Entonces $E(y)$ se puede obtener como

$$E(y) = E_x[E(y|x)],$$

porque

$$\begin{aligned} E_x[E(y|x)] &= \int \left[\int y f(y|x) dy \right] f(x) dx \\ &= \int \int y f(y,x) dy dx = E(y). \end{aligned}$$

El proceso anterior puede generalizarse y repetirse. Por lo tanto, si $g(y_t)$ es una función de y_t, se puede encontrar un valor esperado varios pasos adelante de la secuencia de esperanzas condicionales de un paso adelante porque

$$\underset{t-j}{E}[g(y_t)] = \underset{t-j}{E} \ldots \underset{t-1}{E}[g(y_t)], \quad t = 2, 3 \ldots .$$

La esperanza incondicional se encuentra haciendo $j \to \infty$. La esperanza de una función de la observación en el tiempo $T + \ell$ basada en la información disponible en el tiempo T viene dada por establecer $t = T + \ell$ y $j = \ell$ entonces

$$\underset{T}{E}[g(y_{T+\ell})] = \underset{T}{E} \ldots \underset{T+\ell-1}{E}[g(y_{T+\ell})], \quad \ell = 2, 3, \ldots . \tag{1.4}$$

Cuando el objetivo es predecir una observación futura con base en información actual, la esperanza condicional, $E_T(y_T + \ell)$, $\ell = 1, 2, 3, \ldots$, es óptimo en el sentido de que minimiza el error cuadrático medio (MSE) del error de predicción; ver, por ejemplo, Harvey (1993). Como tal, se denomina predictor del error cuadrático medio mínimo (MMSE). Para modelos no lineales, la expresión (1.4) tiene una importancia práctica considerable para encontrar predictores de MMSE.

1.2.2. Definiciones y propiedades

Deben tenerse en cuenta las siguientes definiciones importantes.

Definición 1.1. *Las variables de ruido blanco no están serialmente correlacionadas con una media y una varianza constantes.*

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Definición 1.2. *La diferencia de martingala tiene una esperanza condicional cero (o constante), es decir,*

$$\underset{t-1}{E}(y_t) = E(y_t|Y_{t-1}) = 0.$$

También es necesario que la esperanza incondicional del valor absoluto sea finita, es decir, $E|y_t| < \infty$*; véase Davidson (2000, págs. 121–2).*

Definición 1.3. *Las variables de ruido blanco estrictas son independientes e idénticamente distribuidas (IID).*

La relación entre una diferencia de martingala y los dos tipos de ruido blanco viene dada por la siguiente proposición.

Proposición 1.4. *(a) Todas las secuencias independientes de media cero son diferencias de martingala y (b) todas las diferencias de martingala son ruido blanco, siempre que la varianza sea finita. En ninguno de los dos casos es cierto lo contrario.*

Demostración. La parte (a) no requiere prueba. Se sigue la parte (b) porque todas las diferencias de martingalas tienen una media incondicional cero y no están correlacionados en serie. Específicamente

$$E(y_t) = E[E(y_t|Y_{t-1})] = 0$$

y y_t no está correlacionada con ninguna función de observaciones pasadas porque

$$E[y_t f(Y_{t-1})] = f(Y_{t-1})E(y_t|Y_{t-1}) = 0.$$

Por lo tanto, la esperanza incondicional de $y_t f(Y_{t-1})$ es cero.

Los modelos heteroscedásticos producen observaciones que son diferencias de martingales pero no IID. Que una secuencia ruido blanco no es necesariamente una diferencia de martingala puede demostrarse mediante un ejemplo simple que muestra que puede haber un predictor no lineal no trivial. Para ser específicos, las observaciones en el modelo

$$y_t = \varepsilon_t + \beta\varepsilon_{t-1}\varepsilon_{t-2}, \quad \varepsilon_t \sim IID(0, \sigma^2), \quad t = 1, \ldots, T,$$

donde ε_0 y ε_{t-1} son fijos y conocidos, son ruido blanco, pero no una diferencia de martingala porque $E(y_{T+1}|Y_T) = \beta\varepsilon_T\varepsilon_{T-1}$. □

Observación 1.5. *Cuando una variable tiene una distribución normal, la distinción entre ruido blanco, ruido blanco estricta y diferencia de martingalas desaparece, ya que una distribución normal se describe completamente en sus dos primeros momentos. Por lo tanto, el ruido blanco gaussiano es un ruido blanco estricto.*

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Un proceso lineal generalmente se define como uno que puede escribirse como una media móvil infinita en perturbaciones $IID(0, \sigma^2)$, siendo finita la suma de los cuadrados de los coeficientes, es decir,

$$y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} \psi_j^2 < \infty, \quad \varepsilon_t \sim IID(0, \sigma^2). \tag{1.5}$$

De manera más general, un proceso lineal puede definirse como una combinación lineal de observaciones pasadas y/o perturbaciones estrictas de ruido blanco, con condiciones iniciales adecuadamente definidas [1]. Para un proceso estacionario, la representación en (1.5) significa que toda la información sobre la dinámica está en la función de la autocorrelación (ACF). Además, el predictor de error cuadrático medio mínimo de $y_{T+\ell}$ es lineal, y su MSE es $\sigma^2 \Sigma_{j=0}^{\ell-1} \psi_j^2$. Sin embargo, a menos que las perturbaciones sean gaussianas, la linealidad de (1.5) tiene un valor práctico limitado, ya que normalmente no es posible derivar la distribución predictiva de varios pasos. Por otro lado, los pronósticos óptimos en un modelo que es una función lineal de las diferencias de martingalas actuales y pasadas son los mismos que en un modelo en el que las diferencias de martingalas se reemplazan por ruidos blancos estrictos, y si las varianzas condicionales son constantes, las MSEs son las mismas.

1.3. Planificación

El orden del texto es el siguiente. El capítulo 2 proporciona algo de teoría básica, comenzando con una revisión de la t de Student y las distribuciones de errores generales. Se discuten los principios de estimación de máxima verosimilitud. Luego se desarrolla la teoría asintótica para las propiedades de los estimadores de Máxima Verosimilitud de los parámetros en la clase DCS.

El capítulo 3 expone la teoría de los modelos de localización. Inicialmente, la atención se centra en las series de tiempo estacionarias, después de lo cual se muestra cómo se pueden manejar los componentes de tendencia y estacional dibujando paralelos con el componente no observado, o modelos estructurales de series temporales que se han aplicado con éxito para modelar el nivel de series temporales gaussianas. Las manipulaciones técnicas se basan principalmente en las propiedades estándar de las distribuciones beta y gamma. Una vez que se aprecia esto, la mayoría de los resultados y fórmulas siguen de manera sencilla y elegante. Sin embargo, el atractivo de las matemáticas no debe restar valor al propósito principal de los modelos, que es tratar con distribuciones de colas pesadas de una manera que sea eficiente, tanto desde el punto de vista estadístico como desde el punto de vista práctico.

[1] Para una discusión más detallada, véase Teräsvirta (2006).

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

El Capítulo 4 proporciona una introducción a los desafíos asociados con el modelado de series temporales multivariadas. El enfoque DCS conduce a una forma atractiva de modelar la correlación variante en el tiempo, así como la escala variante en el tiempo. De hecho, abordar la construcción de modelos para relaciones dinámicas como un ejercicio para cambiar la correlación, en lugar de hacerlo a través de la ruta habitual de parámetros de regresión variables en el tiempo, parece ser la mejor manera de avanzar. Un mayor desarrollo conduce a la sustitución de la correlación por el concepto más general de asociación y las oportunidades que ofrece el modelado de cópulas dinámicas.

En el Capítulo 5 presenta un modelo multivariado de localización para series macroeconómicas

El Capítulo 6 se analiza el efecto de las intervenciones en una variable de interes. Se emplean modelos de ecuaciones en diferencias para representar las posibles características dinámicas de las intervenciones.

Finalmente el Capítulo 7 se emplean un modelo bivariado de localización a las series económicas de inflación y desempleo de Estados Unidos agregando una intervención para medir el efecto del SARS-Cov-2 en el país.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Capítulo 2

Distribuciones estadísticas y teoría asintótica

La teoría de distribución asintotica para un amplio rango de modelos DCS es de crucial importancia para mostrar su viabilidad. La matriz de informacion se puede obtener de manera explicita y las demostraciones de normalidad asintotica de los estimadores de MV son relativamente sencilla. Esto contrasta con la situación de la mayoría de las otras clases de modelos dinámicos no lineales. Por ejemplo, la matriz de información es posible encontrarla de manera explicita en la mayoria de los modelos $GARCH$, mientras que para los modelos $EGARCH$ no hay una teoría asintotica para los estimadores de MV.

La Sección 2.1 revisa las propiedades de las distribuciones que más destacan. Las interconexiones entre las diversas distribuciones proporcionan los componentes básicos de la teoría que sigue. La estimación de MV se analiza para modelos estáticos antes de pasar a los teoremas principales en la Sección 2.3. Se muestra que, para la clase de modelos en cuestión, la matriz de información se descompone en dos partes. Una parte es la matriz de información del modelo estático, mientras que la otra es una matriz vinculada a la ecuación de la dinámica. Esta segunda matriz depende de las propiedades de la score condicional. La teoría asintótica apunta a ciertas restricciones sobre los parámetros dinámicos, y la forma que toman estas restricciones se investiga en capítulos posteriores. La sección 2.4 indica cómo extender los resultados de la teoría asintótica para la ecuación dinámica de primer orden a modelos más complejos. Las pruebas del multiplicador de Lagrange contra la correlación serial se obtienen en la Sección 2.5. También se discuten los métodos para evaluar la bondad del ajuste. La última sección introduce variables explicativas en los modelos.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

2.1. Distribuciones

La distribución t-Student y la distribución general de error juegan un rol importante en el desarrollo que se plantea en esta sección. Ambas contienen a la distribución Gausiana como un caso especial. La distribución Beta y Gamma aparecen en la score y sus derivadas. Por lo tanto, sus propiedades se invocan al derivar expresiones para momentos y ACF's de la serie, así como en la obtención de distribuciones asintóticas de estimadores MV.

2.1.1. Distribución t-Student

La distribución t_ν con parametro de localización μ y parametro de escala φ tiene una función de densidad de probabilidad

$$f(y;\mu,\varphi,\nu) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\varphi\sqrt{\pi\nu}}\left(1+\frac{(y-\mu)^2}{\nu\varphi^2}\right)^{-(\nu+1)/2}, \quad \varphi,\nu>0. \tag{2.1}$$

Esta es la función de densidad de probabilidad no estandar de una t_ν, donde ν son los grados de libertad y $\Gamma(.)$ es la función gamma definida como,

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt,$$

los momentos existen solo hasta $\nu-1$; ver (2.5). Por tanto, la media es finita cuando $\nu>1$. Como la distribución es simétrica respecto al parametro de escala ν, la media es igual a la mediana (que siempre existe). Para $\nu>2$, la varianza es

$$\sigma^2 = \{\nu/(\nu-2)\}\varphi^2. \tag{2.2}$$

El exceso de curtosis o grado de apuntamiento es la cantidad en que se excede la curtosis de una distribución normal, esto es, $6/(\nu-4)$, siempre que $\nu>4$.

La distribución de Cauchy es t_1 la cual es famosa por no tener momentos. Su función de densidad de probabilidad es

$$f(y) = \frac{1}{\pi\varphi}\left(1+\frac{(y-\mu)^2}{\varphi^2}\right)^{-1}.$$

La función de densidad de probabilidad estandarizada, esto es, $\mu=0$ y $\varphi=1$, es graficada junto con la distribución normal en la Figura 2.1. Las colas de la distribución Couchy son mucho más pesadas que las de la distribución normal.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

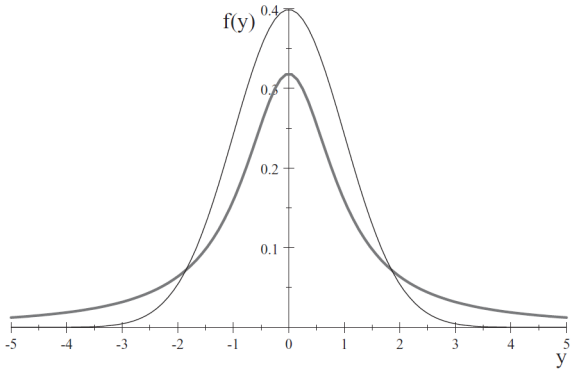


Figura 2.1: Couchy (linea gruesa) y Distribución normal.

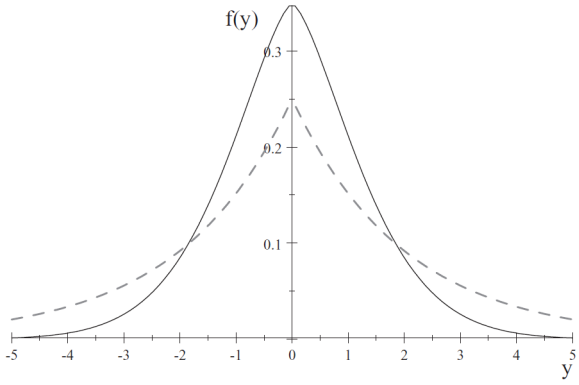


Figura 2.2: Distribución General de Error con $\upsilon = 1{,}5$ y $\upsilon = 1$ (linea segmentada).

2.1.2. Distribución General de Error

La función de densidad de probabilidad de la distribución general de error, denotada por GED(υ), es

$$f(y;\mu,\varphi,\upsilon) = [2^{1+1/\upsilon}\varphi\Gamma(1+1\upsilon)]^{-1}exp(-|(y-\mu)\varphi|^{\upsilon}/2), \quad \varphi,\upsilon > 0, \tag{2.3}$$

donde φ es el parametro de escala, relacionado con la desviación estandar por la formula

$$\sigma = 2^{1/\upsilon}(\Gamma(3/\upsilon)/\Gamma(1/\upsilon))^{1/2}\varphi,$$

y υ es el parametro que determina el peso de la cola.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

La Figura 2.2 muestra la función de densidad estandarizada ($\varphi = 1$) para $\upsilon = 1.5$ y $\upsilon = 1$ siendo la segunda de estas la distribución doble exponencial o de Laplace. La distribución normal es obtenida cuando $\upsilon = 2$, es este caso $\sigma = \varphi$. Los momentos en torno a la media son dados por φ^c multiplicado por la expresión (2.9). La kurtosis es $\Gamma(5/\upsilon)\Gamma(1/\upsilon)$, entonces para $\upsilon = 1$, el exceso de kurtosis es nueve.

2.1.3. Distribución Beta

Cuando la variable aleatoria b sigue una distribución $beta(\alpha, \beta)$, la función de densidad de probabilidad es

$$f(b) = \frac{1}{B(\alpha,\beta)} b^{\alpha-1}(1-b)^{\beta-1}, \quad 0 \geq b \geq 1, \quad \alpha, \beta > 0, \tag{2.4}$$

donde la función beta es

$$B(\alpha,\beta) = \int_0^1 b^{\alpha-1}(1-b)^{\beta-1} db,$$

o en términos de gamma $B(\alpha,\beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$.

Lema 2.1. *Cuando b tiene una distribución $beta(\alpha,\beta)$ y $w(b)$ es una función de b con esperanza finita,*

$$E[b^h(1-b)^k w(b)] = \frac{B(\alpha+h,\beta+k)}{B(\alpha,\beta)} E[w(b)], \quad h > -\alpha, k > -\beta,$$

donde la esperanza del lado derecho se entiende que es con respecto a una distribución $beta(\alpha+h,\beta+k)$

Demostración. La función de densidad de probabilidad de b es como en (2.4), entonces tomando la definición de esperanza tenemos

$$E[b^h(1-b)^k w(b)] = \frac{1}{B(\alpha,\beta)} \int b^h(1-b)^k b^{\alpha-1}(1-b)^{\beta-1} w(b) db, \quad h > -\alpha, \quad k > -\beta,$$

ahora multiplicando por $1 = \frac{B(\alpha+h,\beta+k)}{B(\alpha+h,\beta+k)}$ y sumando los exponentes de potencias de igual base tenemos

$$\begin{aligned} E[b^h(1-b)^k w(b)] &= \frac{B(\alpha+h,\beta+k)}{B(\alpha,\beta)} \frac{1}{B(\alpha+h,\beta+k)} \int b^{\alpha-1+h}(1-b)^{\beta-1+k} w(b) db \\ &= \frac{B(\alpha+h,\beta+k)}{B(\alpha,\beta)} E[w(b)]. \end{aligned}$$

□

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Notar que no es necesario que h y k sean números enteros. Al establecer $w(b) = 1$ obtenemos el siguiente corolario importante.

Corolario 2.2. *Cuando b tiene una distribución beta(α, β)*

$$E(b^h(1-b)^k) = \frac{B(\alpha+h, \beta+k)}{B(\alpha,\beta)}, \quad h > -\alpha, \quad k > -\beta, \tag{2.5}$$

y los momentos son obtenidos estableciendo $k = 0$.

La media es $\alpha/(\alpha+\beta)$. Más generalmente, la fórmula en (2.5) puede simplificarse significativamente para h y k pequeñas.

Ejemplo 2.3. *El momento de primer orden se obtiene de la siguiente manera*

$$\begin{aligned} E(b) &= \frac{B(\alpha+1,\beta)}{B(\alpha,\beta)} \\ &= \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+1+\beta)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\ &= \frac{\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\ &= \frac{\alpha}{\alpha+\beta}. \end{aligned}$$

De manera más general, la ecuación (2.5) puede simplificarse significativamente para h y k pequeñas.

Ejemplo 2.4. *Para una distribución beta$(1/2, \nu/2)$*

$$\begin{aligned} E(b(1-b)) &= \frac{B(\alpha+1,\beta+1)}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\ &= \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)} \end{aligned} \tag{2.6}$$

$$E(b(1-b)) = \frac{(1/2)(\nu/2)}{(1/2+\nu/2+1)(1/2+\nu/2)} = \frac{\nu}{(\nu+3)(\nu+1)}. \tag{2.7}$$

La función generadora de momentos de una distribución *beta* es clave en la obtención de resultados posteriores. Supongamos que b tiene una distribución *beta*(α, β) y c es un número finito. Luego,

$$M_\beta(c;\alpha;\beta) = E(e^{cb}) = 1 + \sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{\alpha+r}{\alpha+\beta+r}\right)\frac{c^k}{k!}, \quad \alpha,\beta > 0. \tag{2.8}$$

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Demostración. Por definición la función generadora de momentos de una variable b con distribución $beta(\alpha, \beta)$ es

$$\begin{aligned} E(e^{cb}) &= \int_0^1 \frac{e^{cb} b^{\alpha-1}(1-b)^{\beta-1}}{B(\alpha,\beta)} db \\ &= \frac{1}{B(\alpha,\beta)} \int_0^1 e^{cb} b^{\alpha-1}(1-b)^{\beta-1} db \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 e^{cb} b^{\alpha-1}(1-b)^{\beta-1} db, \end{aligned}$$

tomando la función confluente hypergeométrica de Kummer's; ver Slater (1965), tenemos

$$\begin{aligned} F_1(\alpha;\beta;c) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 e^{cb} b^{\alpha-1}(1-b)^{\beta-1} db \\ &= \sum_{k=1}^{\infty} \frac{\alpha^{(n)} c^n}{(\alpha+\beta)^{(n)} k!}, \end{aligned}$$

donde $a^{(0)} = 1$ y $a^{(n)} = a(a+1)(a+2)\dots(a+n-1)$ con $n \in \mathbb{N}$. Entonces

$$F_1(\alpha;\beta;c) = E(e^{cb}) = 1 + \sum_{k=1}^{\infty} \left(\prod_{r=0}^{k-1} \frac{\alpha+r}{\alpha+\beta+r} \right) \frac{c^k}{k!}.$$

□

Finalmente notemos que cuando b tiene una distribución $beta(\alpha, \beta)$ la distribución de $(1-b)$ es $beta(\beta, \alpha)$.

Demostración. Utilizando el teorema de transformación de una variable aleatoria; ver Casella and Berger (2021), con $y = 1 - b$ y $g^{-1}(y)$ la función inversa de y, tenemos

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| \\ &= \frac{1}{B(\alpha,\beta)} (1-y)^{\alpha-1} (1-(1-y))^{\beta-1} \\ &= \frac{1}{B(\alpha,\beta)} (1-y)^{\alpha-1} y^{\beta-1} \\ &= \frac{1}{B(\beta,\alpha)} y^{\beta-1} (1-y)^{\alpha-1}. \end{aligned}$$

□

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

2.1.4. Distribución Gamma

La función de densidad de probabilidad de una variable $gamma(\alpha, \gamma)$ es

$$f(g) = \frac{\alpha^{-\gamma} g^{\gamma-1} e^{-g/\alpha}}{\Gamma(\gamma)}, \quad \alpha, \gamma > 0, \tag{2.9}$$

donde α es el parametro de escala y γ es el parameto de forma. La distribución Chi-Cuadrada con ν grados de libertad es $gamma(2, \nu/2)$. Fijando $\gamma = 1$ se obtiene la distribución exponencial. Los momentos de todos los ordenes existen, dados por

$$E(g^h) = \frac{\alpha^h \Gamma(h+\gamma)}{\Gamma(\gamma)}, \quad h > 0. \tag{2.10}$$

Por tanto, la media es $\gamma\alpha$ y la varianza es $\gamma\alpha^2$. La función generadora de momentos es

$$M(c; \alpha, \gamma) = E(e^{cg}) = (1 - \alpha c)^{-\gamma}, \quad -\infty < c < 1/\alpha, \quad \gamma > 0. \tag{2.11}$$

Existen una serie de relaciones importantes ente la función gamma y otras distribuciones.

Lema 2.5. *El valor esperado del valor obsoluto del error de una variable t_ν, ε_t, elevada a la potencia c es*

$$E(|\varepsilon_t|^c) = \frac{\nu^{c/2}\Gamma(c/2 + 1/2)\Gamma(-c/2 + \nu/2)}{(\Gamma(1/2)\Gamma(\nu/2))}, \quad -1 < c < \nu. \tag{2.12}$$

Demostración. Cuando ε_t es t_ν, $|\varepsilon|^c$ puede ser expresado en terminos de una tasa de variables chi-cuadradas elevadas a la potencia $c/2$. El valor esperado de una chi-cuadrada con f grados de libertad elevada a las potencia $c/2$ es

$$\frac{2^{c/2}\Gamma(c/2 + f/2)}{\Gamma(f/2)}, \quad c > -f.$$

□

Lema 2.6. *Si z es $gamma(\theta, \alpha)$ y w es $gamma(\theta, \beta)$, entonces $x = z/(w + z)$ es $beta(\alpha, \beta)$.*

Corolario 2.7. *La variable $(t^2/\nu)/(1 + t^2/\nu)$ tiene una distribución $beta(1/2, \nu/2)$, mientras que $1/(1 + t^2\nu)$ tiene una distribución $beta(\nu/2, 1/2)$.*

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Demostración. Porque t^2/ν es la razon de una distribución normal estandar al cuadrado a una χ^2_ν, se sigue del lema 2.6 que

$$\begin{aligned}\frac{t^2/\nu}{1+t^2/\nu} &= \frac{z/w}{1+z/w} \\ &= \frac{z}{z+w}.\end{aligned}$$

Similarmente,

$$\frac{1}{1+t^2/\nu} = \frac{w}{w+z}$$

□

Lema 2.8. *Para la GED, $|(y-\mu)/\varphi|^\upsilon$ tiene una distribución gamma$(2, 1/\upsilon)$*

El resultado anterior se puede demostrar fácilmente mediante un cambio de variable. El siguiente resultado es inmediato.

Corolario 2.9. *El valor esperado del valor obsoluto de una variable GED(υ) elevada a la potencia c es*

$$E(|(y-\mu)/\varphi|^c) = \frac{2^{c/\upsilon}\Gamma((c+1)/\upsilon)}{\Gamma(1/\upsilon)}, \quad c > 1, \quad \upsilon > 0. \tag{2.13}$$

La expresión anterior entrega los momentos pares (centrales) de una variable de GED. Los momentos centrales impares son cero.

2.2. Máxima Verosimilitud

Sea $y_t, t = 1, \ldots, T$, sea un conjunto de observaciones de series temporales, cada una de las cuales se extrae de una distribución con función de densidad de probabilidad $f(y_t; \boldsymbol{\theta})$, donde $\boldsymbol{\theta}$ es un vector de parámetros. Cuando las observaciones son IID, la función de densidad conjunta es solo el producto de las funciones de densidad individuales. La función de verosimilitud tiene la misma forma que la función de densidad conjunta. Por razones prácticas y teóricas, es más conveniente trabajar con el logaritmo de la función de verosimilitud, y así

$$\ln L(\boldsymbol{\theta}; y_1, \ldots, y_T) = \sum_{t=1}^{T} \ln f(y_t; \boldsymbol{\theta}).$$

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

La función de verosimilitud difiere de la función de densidad conjunta en que las observaciones se toman como dadas, mientras que $\boldsymbol{\theta}$ es variable. El objetivo es encontrar el valor de $\boldsymbol{\theta}$ que hace que la muestra sea 'más probable'. El estimador de MV global, $\tilde{\boldsymbol{\theta}}$, maximiza $\ln L(\boldsymbol{\theta})$ en todo el espacio de parámetros. Siempre que $\ln L(\boldsymbol{\theta})$ sea diferenciable en el valor verdadero del parámetro, $\tilde{\boldsymbol{\theta}}$ se obtendrá resolviendo las ecuaciones de verosimilitud,

$$\frac{\partial \ln L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}. \tag{2.14}$$

Observación 2.10. *Es posible que el supremo de la función de verosimilitud esté en un límite, en cuyo caso la resolución de la(s) ecuación(es) de verosimilitud puede no producir el estimador MV. El modelo de paseo aleatorio más ruido, (1.1) con $\phi = 1$, proporciona un ejemplo; véase Shephard and Harvey (1990).*

Las propiedades de un estimador MV suelen ser sencillas de derivar cuando se pueden escribir en forma cerrada.

Ejemplo 2.11. *La media y varianza de la distribución poisson,*

$$f(y) = \mu^y e^{-\mu}/y!, \quad y = 1, 2, \ldots,$$

ambos son μ. La función de log-verosimilitud para T observaciones independientes es

$$\ln L(\mu) = \ln \mu \sum y_t - T\mu - \sum \ln y_t!$$

y entonces el estimador MV de μ es solo la media de la muestra, es decir, $\tilde{\mu} = \bar{y}$. Por lo tanto, las propiedades del estimador MV se encuentran fácilmente porque, a partir del teorema de Khinchine, y es consistente, ya que es el promedio de las observaciones IID con media finita, y la varianza de la media es μ/T. La normalidad asintótica se deriva de un teorema central del límite estándar.

La matriz de información para una sola observación es

$$\mathbf{I}(\boldsymbol{\theta}_0) = E_0 \left(\frac{\partial \ln f}{\partial \boldsymbol{\theta}} \frac{\partial \ln f}{\partial \boldsymbol{\theta}'} \right) = -E_0 \left(\frac{\partial^2 \ln f}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right), \quad t = 1, \ldots, T, \tag{2.15}$$

donde la esperanza se toma en el valor verdadero de $\boldsymbol{\theta}$ denotado $\boldsymbol{\theta}_0$. La matriz de información completa es $T.\mathbf{I}(\boldsymbol{\theta}_0)$. Siempre que $f(y_t; \boldsymbol{\theta})$ satisfaga ciertas condiciones de regularidad, $\tilde{\boldsymbol{\theta}}$ es un estimador consistente de $\boldsymbol{\theta}_0$, y es asintóticamente normal en el sentido de que $\sqrt{T}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converge en distribución a una normal multivariada, con vector de media cero y matriz de covarianza $\mathbf{I}^{-1}(\boldsymbol{\theta}_0)$. La matriz de información es definida positiva, siempre que el modelo sea identificable.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Ejemplo 2.12. *La cantidad de información para la distribución de Poisson es*

$$I(\mu_0) = -E_0\left(\frac{\partial^2 \ln f}{\partial \mu^2}\right) = \frac{1}{\mu_0}.$$

En este caso, $1/I(\mu_0)$ *da la varianza de muestra pequeña.*

El vector en (2.14) se denomina score de la función de verosimilitud y, como tal, es una función de $\boldsymbol{\theta}$ para un conjunto dado de observaciones. La score de una sola observación juega un papel destacado en la formulación de modelos DCS, pero las propiedades estadísticas de estos modelos se derivan con respecto al vector de score, $\partial \ln f(y_t;\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, donde las observaciones son variables aleatorias, tal y como se encuentran en las fórmulas de la matriz de información, (2.15).

Las ecuaciones normales a menudo serán no lineales, en cuyo caso deben resolverse mediante un procedimiento de optimización numérica. Puede haber más de una solución para una ecuación de probabilidad particular, pero solo el máximo global produce el estimador MV. El parámetro de localización en una distribución de Cauchy proporciona un ejemplo, como lo analizan Newey and McFadden (1994).

Observación 2.13. *Existen diferentes conjuntos de condiciones de regularidad, según la naturaleza del problema. Un tratamiento muy general se puede encontrar en Newey and McFadden (1994). A los efectos de abordar los modelos dinámicos considerados aquí, el conjunto de condiciones más práctico parece ser el adaptado de Jensen and Rahbek (2004). Estas condiciones se establecen en el apartado final.*

2.2.1. Distribución t-Student

Cuando las observaciones provienen de una distribución t_ν, el logaritmo de la funcion de densidad de probabilidad, la densidad logaritmica, es

$$\begin{aligned}\ln f(y_t,\mu,\varphi,\nu) = {} & \ln\Gamma((\nu+1)/2) - \frac{1}{2}\ln\pi - \ln\Gamma(\nu/2) \\ & - \frac{1}{2}\ln\nu - \ln\varphi - \frac{(\nu+1)}{2}\ln\left(1 - \frac{(y_t-\mu)^2}{\nu\varphi^2}\right).\end{aligned} \tag{2.16}$$

La matriz de información[1] es

$$\mathbf{I}\begin{pmatrix}\mu\\ \varphi^2\\ \nu\end{pmatrix} = \begin{bmatrix}\frac{\nu+1}{\nu+3}\varphi^{-2} & 0 & 0\\ 0 & \frac{\nu}{2\varphi^4(\nu+3)} & \frac{1}{2\varphi^2(\nu+3)(\nu+1)}\\ 0 & \frac{1}{2\varphi^2(\nu+3)(\nu+1)} & h(\nu)/2\end{bmatrix},$$

[1]La matriz de información es evaluada en el verdadero parámetro, θ_0, pero el subindice a menudo se omite

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

donde

$$h(\nu) = \frac{1}{2}\psi'(\nu/2) - \frac{1}{2}((\nu+1)/2) - \frac{\nu+5}{\nu(\nu+3)(\nu+1)}, \tag{2.17}$$

con $\psi'(.)$ siendo la función trigamma[2]; ver, por ejemplo, Taylor and Verbyla (2004) y Lin and Wang (2009). Como se verá a su debido tiempo, la derivación depende de reconocer que las primeras derivadas de lnf con respecto a μ y φ^2 son funciones simples de una variable con distribución *beta*. Si la parametrización es en términos de la varianza en lugar de la escala al cuadrado, entonces, asumiendo $\nu > 2$, el único cambio en la matriz de información es que σ^2 reemplaza a φ^2 y el primer elemento, asociado con μ, se multiplica por $(\nu-2)/\nu$. La matriz de información para la media y la varianza de una distribución normal se obtiene como un caso límite.

El parámetro λ, donde $\varphi = exp(\lambda)$, normalmente se utiliza al modelar la escala dinámica. La matriz de información es entonces

$$\mathbf{I}\begin{pmatrix}\mu\\ \lambda\\ \nu\end{pmatrix} = \begin{bmatrix} \frac{\nu+1}{\nu+3}exp(-2\lambda) & 0 & 0 \\ 0 & \frac{2\nu}{\nu+3} & \frac{1}{(\nu+3)(\nu+1)} \\ 0 & \frac{1}{(\nu+3)(\nu+1)} & h(\nu)/2 \end{bmatrix}, \tag{2.18}$$

y el parámetro de escala ya no aparece en el bloque inferior. Esta característica de la función de enlace resulta ser de crucial importancia. Las matemáticas subyacentes a la derivación (2.18) se pueden explicar de la siguiente manera. Suponga por simplicidad que se sabe que μ es cero. La score con respecto a λ para una sola observación,

$$\frac{\partial \ln f(y_t;\lambda,\nu)}{\partial\lambda} = \frac{(\nu+1)y_t^2}{\nu exp(2\lambda)+y_t^2} - 1, \quad t = 1,\ldots,T,$$

puede expresarse como

$$\frac{\partial \ln f(y_t;\lambda,\nu)}{\partial\lambda} = (\nu+1)b(y_t;\lambda,\nu) - 1, \tag{2.19}$$

donde

$$b(y_t;\lambda,\nu) = \frac{y_t^2/\nu exp(2\lambda)}{1+y_t^2/\nu exp(2\lambda)}, \quad 0 \le b \le 1, \quad 0 < \nu < \infty. \tag{2.20}$$

El resultado de $b(y_t,\lambda_0,\nu_0)$ se distribuye beta$(1/2,\nu/2)$ se sigue inmediatamente del Corolario 2.7. Porque $E(b) = 1/(\nu+1)$ y $\text{Var}(b) = 2\nu/(\nu+3)(\nu+1)^2$, la score tiene media cero y varianza $2\nu/(\nu+3)$. La varianza es el segundo elemento de la diagonal en 2.18. La matriz de información para los otros elementos puede ser obtenida de forma similar.

[2]Se definen como la segunda derivada del logaritmo de la función gamma.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

2.2.2. Distribución General de Error (GED)

La densidad logaritmica para observaciones provenientes de una GED es

$$\begin{aligned}\ln f(y_t,\mu,\varphi,\upsilon) = &-(1+\upsilon^{-1})\ln 2 - \ln\Gamma(1+\upsilon^{-1}) \\ &- \ln\varphi - \frac{1}{2\varphi^{\upsilon}}|y_t-\mu|^{\upsilon}.\end{aligned} \tag{2.21}$$

La función score de μ es

$$\frac{\partial \ln f(y_t;\mu,\lambda,\upsilon)}{\partial\mu} = \frac{\upsilon}{2\varphi^{\upsilon}} sgn(y-\mu)|y_t-\mu|^{\upsilon-1}, \quad t=1,\ldots,T,$$

pero cuando $\upsilon < 2$, la densidad logaritmica no es diferenciable con respecto a μ en $y_t = \mu$, entonces la función score no es continua. Cuando $1 < \upsilon < 2$, la teoría asintótica para μ se sigue del argumento de Zhu and Zinde-Walsh (2009). Notar que ellos transforman la escala de manera diferente, entonces las fórmulas en su matriz de información difieren ligeramente de los aquí presentados.

La matriz de información cuando la escala es parametrizada con una función de enlace exponencial, esto es $\lambda = \ln\varphi$, es

$$\mathbf{I}\begin{pmatrix}\mu\\ \lambda\\ \upsilon\end{pmatrix} = \begin{bmatrix}\upsilon/2exp(-2\lambda\upsilon) & 0 & 0\\ 0 & \upsilon & j(\upsilon)\\ 0 & j(\upsilon) & \upsilon^{-3}\ln 2 + g(\upsilon) + h(\upsilon)\end{bmatrix},$$

donde

$$\begin{aligned} j(\upsilon) &= -\upsilon^{-1}(1+0{,}5\Gamma(2/\upsilon)\psi(2/\upsilon)/\Gamma(1/\upsilon)),\\ g(\upsilon) &= 2\upsilon^{-3}\psi(1+1/\upsilon) + \upsilon^{-4}\psi'(1+1/\upsilon)\end{aligned}$$

y

$$h(\upsilon) = \frac{\Gamma(2/\upsilon)((\psi(2/\uparrow))^2) - \psi'(2/\upsilon)}{2\upsilon^2\Gamma(1/\upsilon)}$$

Como en (2.18), los elementos del bloque inferior son independientes de la escala. Cuando $\upsilon = 1$,

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

$$\frac{\partial \ln f(y_t;\mu,\lambda,\upsilon)}{\partial \mu} = \frac{1}{2\varphi} sgn(y_t-\mu)$$

y el estimador MV de μ es la mediana muestral. Una expresión general para la varianza asintótica de la mediana muestral es $(2f(0))^{-2}$, donde $f(0)$ es la función de densidad de probabilidad en μ; ver Koenker and Hallock (2001). Para la distribución de Laplace, la varianza asintótica, $4\varphi^2$, dada por la fórmula anterior es la misma que la obtenida de la matriz de información.

Los problemas de no diferenciabilidad no surgen cuando se conoce μ y se está considerando el estimador de la escala por MV. Supongamos que $\mu = 0$ en (2.21) y se conoce υ. Diferenciar la función logarítmica de verosimilitud completa con respecto a φ nos entrega la score

$$\frac{\partial L}{\partial \varphi} = \sum_{t=1}^{T} \left[-\frac{1}{\varphi} + \frac{\upsilon}{2\varphi^{\upsilon+1}} |y_t|^{\upsilon} \right].$$

Fijando la score en cero produce el estimador máximo verosimil

$$\tilde{\varphi} = \left(\frac{\upsilon}{2} \frac{\sum |y_t|^{\upsilon}}{T} \right)^{1/\upsilon}. \tag{2.22}$$

El hecho de que haya una expresión de forma cerrada para $\tilde{\varphi}$ significa que demostrar consistencia y normalidad asintótica es sencillo. Las propiedades de $\tilde{\lambda}$ siguen entonces como $\tilde{\lambda} = \ln \tilde{\varphi}$.

2.2.3. Distribución Gamma

La función densidad logaritmica para una distribución *gamma* es

$$\ln L(\alpha,\gamma) = -\gamma T \ln \alpha + (\gamma - 1) \sum \ln y_t - \frac{\sum y_t}{\alpha} - T \ln \Gamma(\gamma).$$

El estimados de MV de α es $\tilde{\gamma}$ multiplicado por la media de la muestra, pero $\tilde{\gamma}$ debe calcularse mediante optimización numérica. La matriz de información depende de α y γ, pero si la parametrización se cambia por el logaritmo de la escala, es decir $\alpha = exp\lambda$, donde $-\infty < \lambda < \infty$, depende solo de γ y está dada por

$$\mathbf{I}\begin{pmatrix} \tilde{\lambda} \\ \tilde{\gamma} \end{pmatrix} = \begin{bmatrix} \gamma & 1 \\ 1 & \psi'(\gamma) \end{bmatrix}, \tag{2.23}$$

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

donde la función trigamma $\psi^{'}(\gamma)$ aparece al diferenciar $\ln\Gamma(\gamma)$ dos veces.

El parámetro de escala puede ser reemplazado por la media, $\mu = \alpha\gamma$, en cuyo caso su estimador MV es solo la media de la muestra. Por tanto, se puede obtener un estimador eficiente sin estimar realmente γ. La matriz de información es, por supuesto, diagonal. Si μ está parametrizada en términos de su logaritmo, λ^*, entonces $\alpha = \gamma^{-1}exp\lambda^*$, y la matriz de información es

$$\mathbf{I}\begin{pmatrix}\tilde{\lambda}^* \\ \tilde{\gamma}\end{pmatrix} = \begin{bmatrix}\gamma & 0 \\ 0 & \psi^{'}(\gamma) - 1/\gamma\end{bmatrix},$$

La inversa de las matrices de información anteriores confirma que la varianza asintótica de γ es la misma en ambos casos.

2.2.4. Consistencia y normalidad asintótica

La prueba de consistencia para una amplia gama de modelos DCS se puede basar en el lema 2.1 de Jensen and Rahbek (2004). Esta subsección muestra cómo se aplican las condiciones a algunos modelos estáticos[3]. Una versión simplificada de un solo parámetro del lema es que $\tilde{\theta}$, el máximo global de la función de verosimilitud, es consistente y asintóticamente normal, en el sentido de que la distribución límite de $\sqrt{T}(\tilde{\theta} - \theta_0)$ es normal, con media cero y varianza $1/I(\theta_0)$, siempre que se cumplan las siguientes condiciones.

(i) $\ln L(\theta)$ es tres veces continuamente diferenciable para θ.

(ii) El verdadero valor del parámetro, θ_0, es un punto interior del espacio paramétrico compacto.

(iii) Como $T \to \infty$, $\frac{1}{\sqrt{T}}\frac{\partial \ln L(\theta_0)}{\partial\theta} \to N(0, I(\theta_0))$, donde $I(\theta_0)$ es positiva.(El subindice de θ_0 en $\frac{\partial \ln L(\theta_0)}{\partial\theta}$ indica que la derivada es evaluada en θ_0.)

(iv) Como $T \to \infty$, $-\frac{1}{T}\frac{\partial^2 \ln L(\theta_0)}{\partial\theta^2} \to I(\theta_0)$.

(v) Si $N(\theta_0)$ se encuentra en una vecindad de θ_0,

$$\sup_{\theta\in N(\theta)}\left|\frac{\partial^3 lnL(\theta)}{T.\partial\theta^3}\right| \leq c_T, \tag{2.24}$$

donde c_T es una variable aleatoria que converge en probabilidad una constante positiva acotada, es decir, $0 \leq c_T \xrightarrow{p} c$, $0 < c < \infty$.

[3]Para un modelo estático, las condiciones no son muy diferentes a las de los libros de texto como Greene (2003).

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Observación 2.14. *La condición (ii) descarta los valores límite, como* $\sigma_\varepsilon^2 = 0$ *en (1.1), aunque a menudo es posible desarrollar una teoría asintótica (no estándar) en tales casos. La referencia clásica es Chernoff (1954), pero véase la discusión sobre modelos de componentes no observados en Harvey (1989). La suposición de compacidad descarta valores de parámetros que no son válidos porque el modelo no es identificable. Verificar (v) no suele ser fácil, pero resulta relativamente sencillo para los modelos en los que la tercera derivada de la función de verosimilitud logarítmica está acotada porque depende de una variable, distribuida beta en los valores verdaderos de los parámetros, que siempre debe estar entre cero y uno.*

Distribución t-Student

Para la estimación de λ, el logaritmo de la escala, para una distribución t-Student, la diferenciación del logaritmo de la densidad tres veces muestra que se cumple la condición (i). La condición (ii) simplemente requiere que λ_0 sea finito porque el espacio de parámetros es $-\infty < \lambda < \infty$.

Porque la score en (2.19) es IID, con media cero y varianza finita en λ_0, la condición (iii) se sigue del Teorema Cental del Límite (Lindberg-Levy). Con respecto a (iv), diferenciando (2.19) se obtiene

$$\frac{\partial^2 \ln f(y_t;\lambda)}{\partial \lambda^2} = -2(\nu+1)b(1-b), \quad T = 1, ...T, \tag{2.25}$$

que, al igual que la score, es una variable aleatoria IID, pero con esperanza $-2\nu/(\nu+3)$; ver (2.7). Por lo tanto, su media converge en probabilidad a menos $I(\lambda_0)$ por el teorema de Khinchine.

Finalmente, la tercera derivada de la log-verosimilitud para la t-ésima observación es

$$\frac{\partial^3 \ln f(y_t;\lambda)}{\partial \lambda^3} = 4(\nu+1)b(1-b) - 8(\nu+1)b^2(1-b). \tag{2.26}$$

Para cualquier λ admisible, la estructura de b es tal que $0 \leq b \leq 1$; ver (2.20). Por lo tanto $0 \leq b(1-b) \leq 1$ y $0 \leq b^2(1-b) \leq 1$, y tambien $\partial^3 \ln f(y_t;\lambda)/\partial\lambda^3$ está acotado para cualquier λ admisible. La condición (v) se cumple sin necesidad de introducir una variable aleatoria delimitadora, c_T, en el argumento.

Cuando se desconoce ν, es sencillo adaptar los argumentos anteriores para verificar las condiciones.

Distribución General de Error

Considere el modelo GED con $\mu = 0$ y υ conocido. Aunque la existencia de una expresión de forma cerrada para el estimador de escala de Máxima Verosimilitd significa

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

que no se necesita el conjunto completo de condiciones de regularidad para establecer sus propiedades, vale la pena demostrar que sirven como preludio para probar la consistencia y la normalidad asintótica para los modelos dinámicos más complicados presentados en la siguiente sección. La score de una sola observación con respecto a λ es

$$\begin{aligned}\frac{\partial \ln f(y_t;\lambda)}{\partial \lambda} &= \frac{\upsilon}{2}|y_t exp(-\lambda)|^{\upsilon} - 1 \\ &= (\upsilon/2)g - 1, \quad t = 1, \dots T, \quad -\infty < \lambda < \infty\end{aligned}$$

donde $g = g(y_t;\lambda)$ tiene una distribución $gamma(2, 1/\upsilon)$ en $\lambda = \lambda_0$; ver Lemma 2.8. La esperanza de la score es cero, como debería ser, y la varianza es υ. Claramente, las condiciones (i) a (iii) se mantienen. la segunda derivada es

$$\frac{\partial^2 \ln f(y_t;\lambda)}{\partial \lambda^2} = -(\upsilon^2 - 2)|y_t exp(-\lambda)|^{\upsilon} = -(\upsilon^2/2)g,$$

que también tiene una distribución gamma en λ_0. Por lo tanto (iv) se cumple. En cuanto a (v),

$$\frac{\partial^3 \ln f(y_t;\lambda)}{\partial \lambda^3} = (\upsilon^3 - 2)|y_t exp(-\lambda)|^{\upsilon} = (\upsilon^3/2)g.$$

Ahora escribiendo

$$g(y_t;\lambda) = g(y_t;\lambda_0)exp(\upsilon(\lambda_0 - \lambda)).$$

Si se elige un valor λ^* tal que $\lambda_0 - \lambda^* > \lambda_0 - \lambda$, entonces $g(y_t;\lambda)exp(\upsilon(\lambda - \lambda^*)) > g(y_t;\lambda)$ y podemos construir una variable aleatoria $g(y_t;\lambda^*)$, que unirá uniformemente $g(\lambda)$ desde arriba. Como $g(y_t;\lambda_0)$ tiene primer momento finito (de hecho, todos los momentos son finitos), del teorema de Khinchine se sigue que $\sup_{\lambda \in N(\lambda_0)} |T^{-1}\partial^3 \ln L/\partial \lambda^3|$ está acotado por una variable aleatoria que converge a una cantidad finita positiva, es decir,

$$\frac{\upsilon^3}{2T}\sum_{t=1}^{T} g(y_t;\lambda^*) = \frac{\upsilon^3}{2T}exp(\upsilon(\lambda_0 - \lambda^*))\sum_{t=1}^{T} g(y_t;\lambda_0) \xrightarrow{p} \upsilon^2 exp(\upsilon(\lambda_0 - \lambda^*)) = c,$$

donde $0 < c < \infty$.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

2.3. Estimacion máximo verosimil para modelos Dinamicos Condicionales de Score

En los modelos DCS, algunos o todos los parámetros varían en el tiempo, y la distribución de y_t se define condicional a estos parámetros. Para un único parámetro variable en el tiempo, como en (2.3.4), la dinámica está impulsada por la score condicional. Un requisito crucial, aunque no el único, para establecer resultados sobre distribuciones asintóticas de estimadores MV de los parámetros que rigen los movimientos en θ_{t1t-1} es que no aparezca en la expresión de su cantidad de información. El cumplimiento de esta condición puede requerir una elección cuidadosa de la función de enlace. La primera subsección establece un lema básico sobre la matriz de información y la segunda subsección deriva su forma para el modelo de primer orden. La tercera subsección reelabora el resultado para obtener una especificación dinámica más común, pero menos elegante. La distribución asintótica del estimador de MV se establece en la Subsección 2.3.4 y se demuestra en la subsección siguiente. Luego de una breve discusión sobre la no estacionariedad, el resultado de la matriz de información se generaliza en la Subsección 2.3.7.

2.3.1. Lema de la matriz de información

Suponga inicialmente que solo hay un parámetro, θ, en un modelo estático. Definir

$$u_t = k.\partial \ln f(y_t;\theta)/\partial\theta, \quad t = 1,\ldots,T,$$

donde k es una constante finita que puede ser la cantidad de información, lo que da como resultado la score estandarizada, o alguna otra constante, incluida la unidad. La derivada $\partial \ln f(y_t;\theta)/\partial\theta$ es una variable aleatoria que tiene media cero en el verdadero valor del parámetro, θ_0. De manera similar, u_t tiene media cero en $\theta = \theta_0$, y su varianza, σ_u^2, es finita en condiciones de regularidad estándar. La cantidad de información para una sola observación es

$$\begin{aligned} I(\theta_0) &= -E(\partial^2 \ln f/\partial\theta^2) = E[(\partial \ln f/\partial\theta)^2] \\ &= E(u_t^2)/k^2 = \sigma_u^2/k^2 < \infty. \end{aligned} \tag{2.27}$$

Si, para una elección particular de función de enlace, $I(\theta_0)$ no depende de θ_0, se cumple la siguiente condición.

Condición 2.15. *La varianza de la score en el modelo estático es finita y no depende de* θ_0.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Ahora si se permite que $\theta = \theta_{t|t-1}$ evolucione con el tiempo como una función de observaciones pasadas y valores pasados de la score de la distribución condicional. Debido a que la score condicional depende de observaciones pasadas a traves de $\theta_{t|t-1}$, se puede dividir en dos partes:

$$\frac{\partial \ln f_t(y_t|Y_{t-1};\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} = \frac{\partial \ln f_t(y_t;\theta_{t|t-1})}{\partial \theta_{t|t-1}} \frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\psi}} \tag{2.28}$$

donde la notación $f_t(y_t;\theta_{t|t-1})$ indica que la distribución de y_t depende del parámetro variable en el tiempo, $\theta_{t|t-1}$, y $\boldsymbol{\psi}$ denota el vector de parámetros que gobiernan la dinámica. Debido a que $\theta_{t|t-1}$ y sus derivados dependen solo de la información pasada, la distribución de la score condicional a la información en el tiempo $t-1$ es la misma que su distribución incondicional y, por lo tanto, es invariante en el tiempo. La descomposición anterior de la score condicional conduce al siguiente resultado.

Lema 2.16. *Considere un modelo con un solo parámetro variable en el tiempo, $\theta_{t|t-1}$, que satisface una ecuación que depende de variables que están fijas en el tiempo $t-1$. El proceso está gobernado por un conjunto de parámetros fijos, $\boldsymbol{\psi}$. Si la Condición 2.15 se cumple, entonces la score condicional para la t-ésima observación, $\partial \ln f_t(y_t|Y_{t-1};\boldsymbol{\psi})/\partial\boldsymbol{\psi}$, es una diferencia de martingala en $\boldsymbol{\psi} = \boldsymbol{\psi}_0$, con matriz de covarianza condicional*

$$\begin{aligned} E_{t-1} &= \left(\frac{\partial \ln f_t(y_t|Y_{t-1};\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\right)\left(\frac{\partial \ln f_t(y_t|Y_{t-1};\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\right)' \\ &= I.\left(\frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\psi}}\frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\psi}'}\right), \quad t = 1,\ldots,T, \end{aligned} \tag{2.29}$$

donde la cantidad de información, I, es constante en el tiempo e independiente de $\boldsymbol{\psi}$.

Demostración. El hecho de que la score en (2.28) sea una diferencia de martingalas se confirma por el hecho de que la derivada del parámetro variable en el tiempo, $\partial\theta_{t|t-1}/\partial\boldsymbol{\psi}$, se fija en el tiempo $t-1$, y el valor esperado de la score en el modelo estático es cero.

La matriz de covarianza condicional de la score se encuentra escribiendo su producto exterior como

$$\left(\frac{\partial \ln f_t}{\partial \theta_{t|t-1}}\frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\psi}}\right)\left(\frac{\partial \ln f_t}{\partial \theta_{t|t-1}}\frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\psi}}\right)' = \left(\frac{\partial \ln f_t}{\partial \theta_{t|t-1}}\right)^2\left(\frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\psi}}\frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\psi}'}\right).$$

Ahora tomando las esperanzas condicionadas a la información en el tiempo $t-1$. Si $E_{t-1}(\partial \ln f_t/\partial\theta_{t|t-1})^2$ no depende de $\theta_{t|t-1}$, es fijo e igual a la esperanza incondicional

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

en el modelo estático, es decir, (2.27). Por lo tanto, debido a que $\theta_{t|t-1}$ está fijo en el tiempo $t-1$,

$$E_{t-1}\left[(\frac{\partial \ln f_t}{\partial \theta_{t|t-1}}\frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\psi}})(\frac{\partial \ln f_t}{\partial \theta_{t|t-1}}\frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\psi}})'\right] = \left[E\left(\frac{\partial \ln f_t}{\partial \theta}\right)^2\right]\frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\psi}}\frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\psi}'}.$$

□

Corolario 2.17. *La matriz de información para el modelo del Lema 2.16 es*

$$\boldsymbol{I}(\boldsymbol{\psi}) = I.\boldsymbol{D}(\boldsymbol{\psi}), \quad \textit{donde} \quad \boldsymbol{D}(\boldsymbol{\psi}) = E\left(\frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\psi}}\frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\psi}'}\right).$$

La siguiente subsección muestra cómo encontrar las esperanzas incondicionales de los elementos en la matriz del producto externo de derivadas.

2.3.2. Matriz de información para modelos de primer orden

En esta subsección, la matriz de información en el momento t se deriva para el modelo de primer orden,

$$\theta_{t+1|t} = \delta + \phi\theta_{t|t-1} + \kappa u_t, \quad \kappa \neq 0, \quad t = 1, \ldots, T, \tag{2.30}$$

y mostrando que se define positivo cuando el modelo es identificable. A menos que se sepa que ϕ es cero, la condición $\kappa \neq 0$ es necesario para la identificabilidad. La suposición de que $|\phi| < 1$ permite expresar $\theta_{t|t-1}$ como una media móvil infinita en u_t's. Debido a que las u_t's son diferencias de martingalas y, por lo tanto, ruido blanco, $\theta_{t|t-1}$ es débilmente estacionario, con una media incondicional de $\omega = \delta/(1-\phi)$ y una varianza incondicional de $\sigma_u^2/(1-\phi^2)$. Se supone que el proceso comenzó en el pasado infinito, aunque para propósitos prácticos, $\theta_{1|0}$ puede igualarse a la media incondicional.

Aunque (2.30) es la formulación convencional de un modelo dinámico de primer orden, resulta que la matriz de información toma una forma más simple si la parameterización es en términos de ω en lugar de δ. Por lo tanto

$$\theta_{t+1|t} = \omega + \theta^{\dagger}_{t+1|t}, \quad \theta^{\dagger}_{t+1|t} = \phi\theta^{\dagger}_{t|t-1} + \kappa u_t, \quad t = 1, \ldots, T \tag{2.31}$$

donde establecer $\theta^{\dagger}_{1|0} = 0$ es lo mismo que establecer $\theta_{1|0} = \omega$. Las complicaciones surgen porque u_t depende de $\theta_{t|t-1}$ y por lo tanto de los parámetros en ψ. Derivando (2.31) y observando que $\partial\theta_{t|t-1}/\partial\kappa = \partial\theta^{\dagger}_{t|t-1}/\partial\kappa$, $\partial\theta_{t|t-1}/\partial\phi = \partial\theta^{\dagger}_{t|t-1}/\partial\phi$, pero $\partial\theta_{t|t-1}/\partial\omega = 1 + \theta^{\dagger}_{t|t-1}/\partial\omega$, el vector $\partial\theta_{t+1|t}/\partial\psi$ se convierte en

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

$$\begin{aligned}
\frac{\partial \theta_{t+1|t}}{\partial \kappa} &= \phi \frac{\partial t|t-1}{\partial \kappa} + \kappa \frac{\partial u_t}{\partial \kappa} + u_t \\
\frac{\partial \theta_{t+1|t}}{\partial \phi} &= \phi \frac{\partial \theta_{t|t-1}}{\partial \phi} + \kappa \frac{\partial u_t}{\partial \phi} + \theta_{t|t-1} - \omega \\
\frac{\partial \theta_{t+1|t}}{\partial \omega} &= \phi \frac{\partial \theta_{t|t-1}}{\partial \omega} + \kappa \frac{\partial u_t}{\partial \omega} + 1 - \phi.
\end{aligned} \tag{2.32}$$

Sin embargo,

$$\frac{\partial u_t}{\partial \kappa} = \frac{\partial u_t}{\partial \theta_{t|t-1}} \frac{\partial \theta_{t|t-1}}{\partial \kappa},$$

y análogamente para las otras dos derivadas. Por lo tanto,

$$\begin{aligned}
\frac{\partial \theta_{t+1|t}}{\partial \kappa} &= x_t \frac{\partial \theta_{t|t-1}}{\partial \kappa} + u_t \\
\frac{\partial \theta_{t+1|t}}{\partial \phi} &= x_t \frac{\partial \theta_{t|t-1}}{\partial \phi} + \theta_{t|t-1} - \omega \\
\frac{\partial \theta_{t+1|t}}{\partial \omega} &= x_t \frac{\partial \theta_{t|t-1}}{\partial \omega} + 1 - \phi,
\end{aligned} \tag{2.33}$$

donde

$$x_t = \phi + \kappa \frac{\partial u_t}{\partial \theta_{t|t-1}}, \quad t = 1, \ldots, T. \tag{2.34}$$

La siguiente condición, que generaliza la Condición 2.15, es necesaria para derivar la matriz de información de $\boldsymbol{\psi}$.

Condición 2.18. *Para un modelo estatico, la score y su primera derivada o equivalentemente u_t y $u_t' = \partial u_t / \partial \theta$, tiene segundo momento finito y su covarianza es invariante en el tiempo y no depende de θ, esto es, $E(u_t^{2-k} u_t'^k) < \infty$, $k = 0, 1, 2$.*

La condición anterior establece que $E(u_t u_t^{'}) < \infty$ y $u_t^{'2} < \infty$ así como también $E(u_t^2) < \infty$. La última condición es la misma que la Condición 2.15, como lo es el hecho de que la media de $u_t^{'}$ es invariable en el tiempo.

Se necesitan las siguientes definiciones:

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

$$
\begin{aligned}
a &= E_{t-1}(x_t) = \phi + \kappa E_{t-1}\left(\frac{\partial u_t}{\partial \theta_{t|t-1}}\right) = \phi + \kappa E\left(\frac{\partial u_t}{\partial \theta}\right) \\
b &= E_{t-1}(x_t^2) = \phi^2 + 2\phi\kappa E\left(\frac{\partial u_t}{\partial \theta}\right) + \kappa^2 E\left(\frac{\partial u_t}{\partial \theta}\right)^2 \geq 0 \\
c &= E_{t-1}(u_t x_t) = \kappa E\left(u_t \frac{\partial u_t}{\partial \theta}\right).
\end{aligned} \tag{2.35}
$$

Las esperanzas en las fórmulas anteriores [4] existen en vista de la Condición 2.18. Debido a que son invariantes en el tiempo, las esperanzas incondicionales pueden reemplazar a las condicionales.

El siguiente lema es un requisito previo para el Teorema 2.20.

Lema 2.19. *Cuando el proceso para $\theta_{t|t-1}$ comienza en el pasado infinito y $|a| < 1$,*

$$
\begin{aligned}
E\left(\frac{\partial \theta_{t+1|t}}{\partial \kappa}\right) &= 0, \quad t = 1, \ldots, T, \\
E\left(\frac{\partial \theta_{t+1|t}}{\partial \phi}\right) &= 0, \\
E\left(\frac{\partial \theta_{t+1|t}}{\partial \omega}\right) &= \frac{1-\phi}{1-a}.
\end{aligned} \tag{2.36}
$$

Demostración. Tomando las esperanzas condicionales de la primera ecuación en (2.32) se obtiene

$$
E_{t-1} = \left(\frac{\partial \theta_{t+1|t}}{\partial \kappa}\right) = E_{t-1}\left(x_t \frac{\partial \theta_{t|t-1}}{\partial \kappa} + u_t\right) = a\frac{\partial \theta_{t|t-1}}{\partial \kappa} + 0
$$

y, de la ley de esperanzas iteradas, la esperanza en el tiempo $t-2$ está dada por

$$
\begin{aligned}
E_{t-2}E_{t-1}\left(\frac{\partial \theta_{t+1|t}}{\kappa}\right) &= aE_{t-2}\left(\frac{\partial \theta_{t|t-1}}{\partial \kappa}\right) \\
&= aE_{t-2}\left(x_{t-1}\frac{\partial \theta_{t-1|t-1}}{\partial \kappa} + u_{t-1}\right) = a^2\frac{\partial \theta_{t-1|t-2}}{\partial \kappa}.
\end{aligned}
$$

Por tanto, si $|a| < 1$,

[4] Tenga en cuenta que cuando b_t se use para denotar una variable beta en un modelo dinámico, tendrá un subíndice de tiempo.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

$$\lim_{n\to\infty} E_{t-n}\left(\frac{\partial\theta_{t+1|t}}{\partial\kappa}\right) = 0, \quad t = \ldots, 0, 1, \ldots, T.$$

En cuanto a ω,

$$E_{t-1}\left(\frac{\partial\theta_{t+1|t}}{\partial\omega}\right) = a\frac{\partial\theta_{t|t-1}}{\partial\omega} + 1 - \phi. \tag{2.37}$$

Podemos continuar evaluando esta expresión sustituyendo $\partial\theta_{t|t-1}/\partial\omega$, tomando esperanzas condicionales en el tiempo $t-2$ y luego repitiendo este proceso. Una vez que se ha demostrado que existe una solución, el resultado puede confirmarse tomando las esperanzas incondicionales en (2.37) para dar

$$E\left(\frac{\partial\theta_{t+1|t}}{\partial\omega}\right) = \frac{1-\phi}{1-a}.$$

De manera similar, tomando esperanzas condicionales de $\partial\theta_{t+1|t}/\partial\phi$ en el tiempo $t-1$ da

$$E_{t-1}\left(\frac{\partial\theta_{t+1|1}}{\partial\phi}\right) = a\frac{\partial\theta_{t|t-1}}{\partial\phi} + (\theta_{t|t-1} - \omega), \tag{2.38}$$

y entonces la esperanza incondicional produce

$$E\left(\frac{\partial\theta_{t+1|t}}{\partial\phi}\right) = 0, \quad t = \ldots, 0, 1, \ldots, T.$$

□

El Lema anterior requiere que $|a| < 1$. El resultado de la siguiente matriz de información requiere $b < 1$, y el cumplimiento de esta condición implica $|a| < 1$. Que este sea el caso se sigue directamente de la desigualdad de Cauchy-Schwarz $E(x_t^2) \geq [E(|x_t|)]^2$.

Teorema 2.20. *Supongamos que la condición (2.18) se cumple y que $b < 1$. Entonces la matriz de información para una sola observación es invariante en el tiempo y esta dada por*

$$\boldsymbol{I}(\boldsymbol{\psi}) = I.\boldsymbol{D}(\boldsymbol{\psi}) = (\sigma_u^2/k^2)\boldsymbol{D}(\boldsymbol{\psi}), \tag{2.39}$$

donde

$$\boldsymbol{D}(\boldsymbol{\psi}) = \boldsymbol{D}\begin{pmatrix}\kappa\\ \phi\\ \omega\end{pmatrix} = \frac{1}{1-b}\begin{bmatrix}A & D & E\\ D & B & F\\ E & F & C\end{bmatrix}, \tag{2.40}$$

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

con

$$A = \sigma_u^2, \quad B = \frac{\kappa^2\sigma_u^2(1+a\phi)}{(1-\phi^2)(1-a\phi)}, \quad C = \frac{(1-\phi)^2(1+a)}{(1-a)},$$
$$D = \frac{a\kappa\sigma_u^2}{1-a\phi}, \quad E = \frac{c(1-\phi)}{(1-a)}, \quad y \quad F = \frac{ac\kappa(1-\phi)}{(1-a)(1-a\phi)}.$$

Demostración. Para una sola observación, la matriz de información, la matriz de covarianza de la score, se obtiene tomando la esperanza incondicional de (2.29) y luego combinándola con la fórmula para $\mathbf{D}(\boldsymbol{\psi})$. La derivación de el primer término, A, se da aquí para ilustrar el método. Este término es la esperanza incondicional del cuadrado de la primera derivada en (2.33). Para evaluarlo, primero se toman las esperanzas condicionales en el tiempo $t-1$, para obtener

$$\begin{aligned} E_{t-1}\left(\frac{\partial\theta_{t+1|t}}{\partial\kappa}\right)^2 &= E_{t-1}\left(x_t\frac{\partial\theta_{t|t-1}}{\partial\kappa} + u_t\right)^2 \\ &= b\left(\frac{\partial\theta_{t|t-1}}{\partial\kappa}\right)^2 + 2c\frac{\partial\theta_{t|t-1}}{\partial\kappa} + \sigma_u^2. \end{aligned} \tag{2.41}$$

Del Lema 2.19 se sigue que la esperanza incondicional del segundo término es cero. Eliminando este término y tomando las esperanzas en $t-2$ da

$$\begin{aligned} E_{t-2}\left(\frac{\partial\theta_{t+1|t}}{\partial\kappa}\right)^2 &= bE_{t-2}\left(x_{t-1}\frac{\partial\theta_{t-1|t-2}}{\partial\kappa} + u_{t-1}\right)^2 + \sigma_u^2 \\ &= b^2\left(\frac{\partial\theta_{t-1|t-2}}{\partial\kappa}\right) + 2cb\frac{\partial\theta_{t-1|t-2}}{\partial\kappa} + b\sigma_u^2 + \sigma_u^2 \end{aligned}$$

Nuevamente, el segundo término puede ser eliminado, y es claro que

$$\lim_{n\to\infty} E_{t-n}\left(\frac{\partial\theta_{t+1|t}}{\partial\kappa}\right)^2 = \frac{\sigma_u^2}{1-b}.$$

Tomando esperanzas incondicionales en (2.41) da el mismo resultado. □

Corolario 2.21. *Si κ es cero, $B = 0$ en (2.40), por lo que $\boldsymbol{D}(\kappa, \phi, \omega)$ es singular y el modelo no es identificable. Cuando $\kappa \neq 0$, los tres parámetros se identifican localmente, siempre que $b < 1$, incluso si $\phi = 0$. Por otro lado, si se conoce ϕ, tanto κ como ω son localmente identificables, incluso si $\kappa = 0$.*

Observación 2.22. *Una expresión para la matriz de información permite calcular las estimaciones de MV mediante el método de score. Las derivadas analíticas se construyen recursivamente usando (2.32) combinada con (2.28).*

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

2.3.3. Matriz de información con la parametrización δ

Para el modelo en el que la constante está parametrizada como en (2.30),

$$\begin{aligned}\frac{\partial\theta_{t+1|t}}{\partial\kappa} &= x_t\frac{\partial\theta_{t|t-1}}{\partial\kappa} + u_t\\ \frac{\partial\theta_{t+1|t}}{\partial\phi} &= x_t\frac{\partial\theta_{t|t-1}}{\partial\phi} + \theta_{t|t-1}\\ \frac{\partial\theta_{t+1|t}}{\partial\delta} &= x_t\frac{\partial\theta_{t|t-1}}{\partial\delta} + 1,\end{aligned} \tag{2.42}$$

lo que lleva a una modificación del Lema 2.19.

Lema 2.23. *Cuando el proceso para $\theta_{t|t-1}$ comienza en el pasado infinito y $|a| < 1$,*

$$\begin{aligned}E\left(\frac{\partial\theta_{t+1|t}}{\partial\kappa}\right) &= 0, \quad t = \ldots 0, 1, \ldots, T,\\ E\left(\frac{\partial\theta_{t+1|t}}{\partial\phi}\right) &= \frac{\delta}{(1-a)(1-\phi)},\\ E\left(\frac{\partial\theta_{t+1|t}}{\partial\delta}\right) &= \frac{1}{1-a}.\end{aligned} \tag{2.43}$$

La matriz de información viene dada por la siguiente modificación del Teorema 2.20.

Teorema 2.24. *Supongamos que la condición (2.18) se cumple y que $b < 1$. Entonces la matriz de información para una sola observación es invariante en el tiempo y esta dada por (2.39), con*

$$\boldsymbol{D}(\boldsymbol{\psi}) = \boldsymbol{D}\begin{pmatrix}\kappa\\ \phi\\ \delta\end{pmatrix} = \frac{1}{1-b}\begin{bmatrix}A & D & E\\ D & B & F\\ E & F & C\end{bmatrix}, \tag{2.44}$$

donde $A = \sigma_u^2$,

$$\begin{aligned}B &= \frac{(1+a\phi)\kappa^2\sigma_u^2}{(1-a\phi)(1-\phi^2)} + \frac{(1+a\phi)\delta^2}{(1-a\phi)(1-phi)^2} + \frac{2a\delta(delta+\kappa c)}{(1-\phi)(1-a)(1-a\phi)},\\ C &= (1+a)/(1-a), \quad D = \frac{a\kappa\sigma_u^2}{1-a\phi} + \frac{c\delta}{(1-\phi)(1-a)},\\ E &= c/(1-a), \quad F = \frac{a\kappa c}{(1-a)(1-a\phi)} + \frac{\delta(1+a)}{(1-\phi)(1-a)}.\end{aligned}$$

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Observación 2.25. *La matriz de información basada en (2.40) es mucho más simple y es independiente del parámetro ω, mientras que (2.44) depende de δ. Esta diferencia puede parecer sorprendente a primera vista, pero de hecho, un contraste similar aparece en las matrices de información para un modelo Gaussiano AR(1) simple, que se puede formular como*

$$y_t = \omega(1-\phi) + \phi y_{t-1} + \varepsilon_t \quad ó \quad y_t = \delta + \phi y_{t-1} + \varepsilon_t, \quad |\phi| < 1. \tag{2.45}$$

Al invertir las matrices de información obtenidas de (2.40) y (2.44), se encuentra que las submatrices correspondientes a la matriz de covarianza de $(\tilde{\kappa}, \tilde{\phi})$ son iguales. De hecho, sería algo preocupante que no lo fueran. Los únicos términos de la inversa de (2.44) que dependen de δ son los que están conectados a $\tilde{\delta}$. Es fácil verificar que lo mismo ocurre en (2.45), donde la varianza asintótica de $\tilde{\phi}$ es siempre $(1-\phi^2)/T$.

2.3.4. Distribución asintótica

Siempre que $\tilde{\boldsymbol{\psi}}$, el estimador MV de $\boldsymbol{\psi}$, sea el máximo global de la función de verosimilitud e $\mathbf{I}^{-1}(\boldsymbol{\psi}_0)$ sea definida positiva, que la esperanza sea consistente, y la distribución límite de $\sqrt{T}(\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}_0)$ es normal multivariante con vector de media cero y matriz de covarianza

$$\text{Var}(\tilde{\boldsymbol{\psi}}) = \boldsymbol{I}^{-1}(\boldsymbol{\psi}_0) = I^{-1}\boldsymbol{D}^{-1}(\boldsymbol{\psi}_0). \tag{2.46}$$

La interpretación del resultado anterior es que $\tilde{\boldsymbol{\psi}}$ es aproximadamente normal, con media $\boldsymbol{\psi}$ y matriz de covarianza dadas por la siguiente definición.

Definición 2.26. *La matriz de covarianza aproximada o asintótica es*

$$Avar(\tilde{\boldsymbol{\psi}}) = T^{-1}Var(\tilde{\boldsymbol{\psi}}). \tag{2.47}$$

El error estándar asintótico (ASE) de un elemento de $\tilde{\boldsymbol{\psi}}$ es la raíz cuadrada del elemento diagonal correspondiente de $Avar(\tilde{\boldsymbol{\psi}})$.

Una prueba formal de consistencia y normalidad asintótica requiere la verificación de una serie de condiciones. Este asunto se aborda en la siguiente subsección.

2.3.5. Consistencia y normalidad asintótica

Una prueba de la consistencia y normalidad asintótica del estimador MV para el modelo de primer orden necesita que se cumplan una serie de condiciones de regularidad. La primera condición es simplemente una declaración formal del requisito, ya señalado, de que si se va a identificar el modelo, κ no debe ser cero o tomar un valor tal que se viole la restricción $b < 1$.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Condición 2.27. *Los elementos de $\boldsymbol{\psi}_0$ son un punto interior del espacio compacto de parámetros, que para el modelo estacionario de primer orden se tomará como $|\phi| < 1$, $|\omega| < \infty$ y $0 < \kappa < \kappa_u$, $\kappa_L < \kappa < 0$, donde κ_u y κ_L son valores determinados por la condicion $b < 1$*

Proposición 2.28. *El modelo es localmente identificable cuando se cumplen las condiciones 2.18 y 2.27 porque $\boldsymbol{D}(\boldsymbol{\psi}_0)$ y, por lo tanto, $\boldsymbol{I}(\boldsymbol{\psi}_0)$, es definida positiva.*

En la subsección 2.2.4 se proporcionó un conjunto de condiciones para la consistencia y la normalidad asintótica de los estimadores MV y se demostró que se cumple para las observaciones IID de la distribución t de Student y otras distribuciones. Los argumentos expuestos ahora deben combinarse con los resultados sobre la estacionariedad y la ergodicidad de las derivadas de $\theta_{t|t-1}$. La siguiente condición es una versión más fuerte de la Condición 2.18.

Condición 2.29. *Para el modelo estático, las distribuciones de la score y sus tres primeras derivadas, o equivalentemente u_t y $\partial^j u_t/\partial\theta^j$, $\quad j = 1, 2, 3$, son invariantes en el tiempo y no dependen de θ. Además, sus esperanzas existen, al igual que la esperanza de u_t $\partial u_t/\partial\theta$.*

Lema 2.30. *Cuando se cumple la Condición 2.29, u_t en el modelo dinámico es IID $(0, \sigma_u^2)$, por lo que el proceso $\theta_{t+1|t}$ en (2.30) es estrictamente estacionario cuando $\boldsymbol{\psi} = \boldsymbol{\psi}_0$.*

Demostración. Debido a que u_t es un diferencia de martingalas con una distribución invariable en el tiempo, es IID con media cero. □

Lema 2.31. *Las primeras derivadas de $\theta_{t+1|t}$ con respecto a κ, ϕ y ω son ecuaciones de recurrencia estocástica (SREs), y la condición $|a| < 1$ es suficiente para asegurar que son estrictamente estacionarios y ergódicos en el verdadero valor del parámetro.*

Demostración. Las primeras derivadas se dieron en (2.42). Son ecuaciones de recurrencia estocástica de la forma $y_{t+1} = x_t y_t + z_t$, donde x_t y z_t son estrictamente estacionarios. La estacionariedad estricta y la ergodicidad se derivan de los resultados estándar de los SREs; véase Brandt (1986) y Straumann y Mikosch (2006, págs. 2450–1). La condición necesaria para la estacionariedad estricta es $E(\ln|x_t|) < 0$. Esta condición se cumple en el valor verdadero del parámetro cuando $|a| = E(|x_t|) < 1$, porque de la desigualdad de Jensen, $E(\ln|x_t|) \leq \ln E(|x_t|) < 0$. □

Observación 2.32. *Recuerde que de la desigualdad de Jensen, $|a| < 1$ está implícito en $b < 1$.*

Corolario 2.33. *La condición $|a| < 1$ es suficiente para asegurar que las primeras derivadas elevadas a cualquier potencia positiva sean estrictamente estacionarias y ergódicas en el verdadero valor del parámetro.*

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Demostración. La prueba del Lema 2.31 mostró que $E(\ln|x_t|) < 0$ cuando $a < 1$ y entonces $E(\ln|x_t|^c) = cE(\ln|x_t|) < 0$ para $c > 0$ □

Corolario 2.34. *La condición $|a| < 1$ es suficiente para asegurar que las derivadas segunda y tercera de $\theta_{t+1|t}$ con respecto a κ, ϕ y ω son estrictamente estacionarias y ergódicas en el verdadero valor del parámetro.*

Demostración. Considere las segundas derivadas. Centrándose en ϕ, tenga en cuenta que debido a que

$$\frac{\partial\theta_{t+1|t}}{\partial\phi} = \phi\frac{\partial\theta_{t|t-1}}{\partial\phi} + \kappa\frac{u_t}{\partial\phi} + \theta_{t|t-1} - \omega$$

la segunda derivada de $\theta_{t+1|t}$ con respecto a ϕ es

$$\frac{\partial^2\theta_{t+1|t}}{\partial\phi^2} = \phi\frac{\partial^2\theta_{t|t-1}}{\partial\phi^2} + \frac{\partial\theta_{t|t-1}}{\partial\phi} + \kappa\frac{\partial^2 u_t}{\partial\phi^2} + \frac{\partial\theta_{t|t-1}}{\partial\phi}.$$

Ahora

$$\frac{\partial^2 u_t}{\partial\phi^2} = \frac{\partial u_t}{\partial\theta_{t|t-1}}\frac{\partial^2\theta_{t|t-1}}{\partial\phi^2} + \frac{\partial^2 u_t}{\partial\theta_{t|t-1}^2}\left(\frac{\partial\theta_{t|t-1}}{\partial\phi}\right)^2,$$

entonces

$$\begin{aligned}\frac{\partial^2\theta_{t+1|t}}{\partial\phi^2} &= \phi\frac{\partial^2\theta_{t|t-1}}{\partial\phi^2} + 2\frac{\partial\theta_{t|t-1}}{\partial\phi} + \kappa\frac{\partial u_t}{\partial\theta_{t|t-1}}\frac{\partial^2\theta_{t|t-1}}{\partial\phi^2} + \kappa\frac{\partial^2 u_t}{\partial\theta_{t|t-1}}\left(\frac{\partial\theta_{t|t-1}}{\partial\phi}\right)^2 \\ &= \left(\phi + \kappa\frac{\partial u_t}{\partial\theta_{t|t-1}}\right)\frac{\partial^2\theta_{t|t-1}}{\partial\phi^2} + 2\frac{\partial\theta_{t|t-1}}{\partial\phi} + \kappa\frac{\partial^2 u_t}{\partial\theta_{t|t-1}}\left(\frac{\partial\theta_{t|t-1}}{\partial\phi}\right)^2.\end{aligned}$$

La primera derivada, $\partial\theta_{t+1|t}/\partial\phi$, es estrictamente estacionaria cuando $|a| < 1$, como lo es su cuadrado. Así, la ecuación anterior para $\partial^2\theta_{t+1|t}/\partial\phi^2$ es un SRE con la misma forma que el SRE para $\partial\theta_{t+1|t}/\partial\phi$, es decir,

$$\frac{\partial^2\theta_{t+1|t}}{\partial\phi^2} = x_t\frac{\partial^2\theta_{t|t-1}}{\partial\phi^2} + \{\text{proceso estrictamente estacionario}\}$$

Se puede emplear un argumento similar para mostrar que las derivadas con respecto a κ y ω obedecen a SRE de la misma forma, al igual que las derivadas cruzadas, como $\partial^3\theta_{t+1|t}/\partial\phi\partial\kappa$.

La tercera derivada con respecto a ϕ es

$$\frac{\partial^3\theta_{t+1|t}}{\partial\phi^3} = \phi\frac{\partial^3\theta_{t|t-1}}{\partial\phi^3} + \frac{\partial^2\theta_{t|t-1}}{\partial\phi} + \kappa\frac{\partial^3 u_t}{\partial\phi^3} + \frac{\partial^2\theta_{t|t-1}}{\partial\phi^2},$$

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

donde

$$\frac{\partial^3 u_t}{\partial \phi^3} = \frac{\partial u_t}{\partial \theta_{t|t-1}} \frac{\partial^3 \theta_{t|t-1}}{\partial \phi^3} + \frac{\partial^2 u_t}{\partial \theta_{t|t-1}^2} \frac{\partial \theta_{t|t-1}}{\partial \phi} \frac{\partial^2 \theta_{t|t-1}}{\partial \phi^2} + \frac{\partial^3 u_t}{\partial \theta_{t|t-1}^3} \left(\frac{\partial \theta_{t|t-1}}{\partial \phi} \right)^3 .$$

El principal punto a destacar es la aparición de la tercera derivada de u_t con respecto a ϕ y el cubo de la primera derivada. Sin embargo, el argumento de que $\partial^3 \theta_{t+1|t}/\partial \phi^3$ es estrictamente estacionario como antes. □

Lema 2.35. *La condición $|a| < 1$ es necesario para que los primeros momentos de las primeras derivadas de $\theta_{t|t-1}$ sean finitos, mientras que $b < 1$ es necesario para que las segundas derivadas y los cuadrados de las primeras derivadas tengan primeros momentos finitos.*

Demostración. La existencia de primer y segundo momento para las primeras derivadas de $\theta_{t|t-1}$ se mostró al derivar la matriz de información. Como se desprende de la prueba del Corolario 2.34, la existencia de los primeros momentos de las segundas derivadas requiere la existencia del segundo momento de las primeras derivadas y $E(\partial^2 u_t/\partial \theta^2) < \infty$. □

Los resultados anteriores ahora se pueden usar para verificar las versiones multivariadas de (iii) y (iv) en la Subsección 2.2.4.

Proposición 2.36. *Sujeto a las condiciones 2.27 y 2.29,*

$$(1/\sqrt{T})\partial \ln L(\boldsymbol{\psi}_0)/\partial \boldsymbol{\psi} \to N(0, \boldsymbol{I}(\boldsymbol{\psi}_0)), \quad \textit{cuando} \quad T \to \infty,$$

donde $\boldsymbol{I}(\boldsymbol{\psi}_0)$ es definida positiva.

Demostración. Del Lema 2.16, el vector de score en $\boldsymbol{\psi}_0$ es una diferencia de martingalas con matriz de covarianza condicional, (2.29). Para un solo elemento en la score condicional, $\partial \ln f_t/\partial \psi_i, \quad i = 1, 2, 3$, donde ψ_i es el i-ésimo elemento de $\boldsymbol{\psi}$, y podemos escribir

$$E\left[\left(\frac{\partial \ln f_t}{\partial \psi_i} \right)^2 \right] = I. \left(\frac{\partial \theta_{t|t-1}}{\partial \psi_i} \right), \quad t = 1, \ldots, T.$$

La demostración del teorema central del límite para las diferencias de martingala en Brown (1971) requiere que

$$p\lim T^{-1} \sum I. \left(\frac{\partial \theta_{t|t-1}}{\partial \psi_i} \right)^2 = I. p\lim T^{-1} \sum \left(\frac{\partial \theta_{t|t-1}}{\partial \psi_i} \right)^2 < \infty, \quad i = 1, 2, 3. \quad (2.48)$$

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Dado que, como se observa en el Lema 2.42, $(\partial\theta_{t|t-1}/\partial\psi_i)^2$ es estrictamente estacionario y ergódico con media incondicional finita, del teorema ergódico se sigue que se cumple (2.48). De manera similar, la existencia de momentos de segundo orden para la score y $\partial\theta_{t|t-1}/\partial\psi_i$ es suficiente para que se cumpla la condición de Lindberg; ver Jensen and Rahbek (2004). □

Proposición 2.37. *Sujeto a las condiciones 2.27 y 2.29,*

$$(-1/T)\partial^2 \ln L(\boldsymbol{\psi}_0)/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}' \xrightarrow{P} \boldsymbol{I}(\boldsymbol{\psi}_0), \quad \textit{cuando} \quad T \to \infty.$$

Demostración. Escribiendo

$$\frac{\partial^2 \ln f_t}{\partial\psi^2} = u_t \frac{\partial^2\theta_{t|t-1}}{\partial\psi^2} + \frac{\partial u_t}{\partial\theta_{t|t-1}} \left(\frac{\partial\theta_{t|t-1}}{\partial\psi}\right)^2$$

Por lo tanto

$$\frac{1}{T}\frac{\partial^2 \ln L}{\partial\psi^2} = \frac{1}{T}\sum \frac{\partial^2 \ln f_t}{\partial\psi^2} = \frac{1}{T}\sum u_t \frac{\partial^2\theta_{t|t-1}}{\partial\psi^2} + \frac{1}{T}\sum \frac{u_t}{\partial\theta_{t|t-1}} \left(\frac{\partial\theta_{t|t-1}}{\partial\psi}\right)^2.$$

La expresión anterior es similar a la de Jensen and Rahbek (2004). Como u_t y su derivada son independientes de $\partial\theta_{t|t-1}/\partial\psi$ y su derivada, los términos del lado derecho convergen, por el teorema ergódico, a

$$E(u_t)E\left(\frac{\partial^2\theta_{t|t-1}}{\partial\psi^2}\right) + E\left(\frac{\partial u_t}{\partial\theta}\right) E\left(\frac{\partial\theta_{t|t-1}}{\partial\psi}\right)^2, \tag{2.49}$$

siempre que $E(\partial^2\theta_{t|t-1}/\partial\psi^2)$ y $E[(\partial\theta_{t|t-1}/\partial\psi)^2]$ sean finitos. El segundo término en (2.49) es el mismo que el límite en (2.48). Tomando la esperanza de $\partial^2\theta_{t|t-1}/\partial\psi^2$ como en el Corolario 2.34 – cuando ψ es ϕ – muestra que $E(\partial^2\theta_{t|t-1}/\partial\psi^2)$ es finita si $E(\partial^2 u_t/\partial\theta^2)$; véase la Condición 2.29. Siendo este el caso, el primer término en (2.49) es cero. Lo mismo es cierto para las otras derivadas y derivadas cruzadas. □

La versión general de la condición (v) en la Subsección 2.2.4 se puede establecer como

$$\max_{h,i,j=1,\ldots,n} \sup_{\boldsymbol{\psi}\in N(\boldsymbol{\psi}_0)} \left|\frac{\partial^3 \ln L(\boldsymbol{\psi})}{T.\partial\psi_h\partial\psi_i\partial\psi_j}\right| \leq c_T, \tag{2.50}$$

donde $\psi_i, \quad i = 1, \ldots, n$ denotan los parámetros en $\boldsymbol{\psi}$, $N(\boldsymbol{\psi}_0)$ es una vecindad del verdadero valor del parámetro y c_T es una variable aleatoria tal que $0 \leq c_T \xrightarrow{P} c$, $0 < c < \infty$. Encontrar las condiciones generales bajo las cuales se cumple (2.50) no es

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

fácil. En cambio, se establecen teoremas para dos clases generales de modelos a escala de la forma

$$y_t = \varepsilon_t exp(\theta_{t|t-1}), \quad t = 1, \ldots, T,$$

donde los ε_t son IID con localización cero y escala unitaria.

Teorema 2.38. *Cuando ε_t se distribuye como t_ν o beta generalizada y los verdaderos parámetros satisfacen la Condición 2.27, $\tilde{\boldsymbol{\psi}}$ es consistente y la distribución límite de $\sqrt{T}(\tilde{\boldsymbol{\psi}}-\boldsymbol{\psi}_0)$ es normal multivariante, con vector de medias cero y matriz de covarianza, $Var(\tilde{\boldsymbol{\psi}})$, como en 2.46.*

Demostración. Para las distribuciones en cuestión, las primeras tres derivadas de $\ln f_t$ consisten en combinaciones lineales de las variables $b_t^* = b_t^h(1-b_t)^k, \quad h, k = 0, 1, 2, \ldots,$ donde $b_t = b_t(y_t; \boldsymbol{\psi})$ tiene distribución beta en $\boldsymbol{\psi} = \boldsymbol{\psi}_0$. Por lo tanto, se cumplen las condiciones para que las proposiciones 2.36 y 2.37 sean verdaderas. Con respecto a las terceras derivadas de la función log-verosimilitud, considere un solo parámetro (sin subíndice) y escriba

$$\frac{\partial^3 \ln L(\psi)}{\partial \psi^3} = \sum_{t=1}^{T} z_t$$

donde

$$\begin{aligned} z_t = \frac{\partial^3 \ln f_t(y_t;\psi)}{\partial \psi^3} &= \frac{\partial \ln f_t}{\partial \theta_{t|t-1}} \frac{\partial \theta_{t|t-1}^3}{\partial \psi^3} + \frac{\partial^2 \ln f_t}{\partial \theta_{t|t-1}^2} \frac{\partial \theta_{t|t-1}}{\partial \psi} \frac{\partial \theta_{t|t-1}^2}{\partial \psi^2} \\ &+ \frac{\partial^3 \ln f_t}{\partial \theta_{t|t-1}^3} \left(\frac{\partial \theta_{t|t-1}}{\partial \psi} \right)^3 . \end{aligned} \tag{2.51}$$

La definición de z_t se extiende a cuando hay n parámetros, como en (2.50). Escribiendo

$$b_t(y_t; \boldsymbol{\psi}) = h(y_t; \boldsymbol{\psi})/(1 + h(y_t; \boldsymbol{\psi})), \quad 0 \leq h(y_t; \boldsymbol{\psi}) \leq \infty.$$

Para cualquier $\boldsymbol{\psi}$ admisible, $0 \leq b_t(y_t; \boldsymbol{\psi}) \leq 1$ y así, generalizando la discusión que sigue (2.26),$0 \leq b_t^*(y_t; \boldsymbol{\psi}) \leq 1$, donde $b_t^*(y_t; \boldsymbol{\psi})$ tiene la misma forma que b_t^* al comienzo de la prueba. Por lo tanto, $\partial^j \ln f_t / \partial \theta_{t|t-1}^j, \quad j = 1, 2, 3$ son variables acotadas. Además, las derivadas de $\theta_{t|t-1}$ deben estar acotadas en $\boldsymbol{\psi}_0$, porque como se muestra en el Lema 2.31, son SRE estables que en última instancia dependen de $b_t(y_t; \boldsymbol{\psi})$. También deben estar acotados en la vecindad de $\boldsymbol{\psi}_0$, porque la condición $b < 1$ es más que suficiente para garantizar la condición de estabilidad $E(\ln |x_t|) < 0$. Extendiendo el argumento

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

para cubrir las terceras derivadas con respecto a más de uno de los parámetros en $\boldsymbol{\psi}$ es sencillo. Por tanto, podemos concluir que todas las terceras derivadas en (2.50) están acotadas. □

Teorema 2.39. *Cuando ε_t se distribuye como GED o gamma generalizada y los parámetros verdaderos satisfacen la Condición 2.27, $\tilde{\boldsymbol{\psi}}$ es consistente, y la distribución límite de $\sqrt{T}(\tilde{\boldsymbol{\psi}} - \tilde{\boldsymbol{\psi}}_{\mathbf{0}})$ es normal multivariante con vector de media cero y matriz de covarianza, $Var(\tilde{\boldsymbol{\psi}})$, como en (2.46).*

Demostración. De la última parte de la subsección 2.2.4, sabemos que las primeras tres derivadas de la función de verosimilitud logarítmica para la GED consisten en funciones lineales de variables, $g_t = g_t(y_t; \boldsymbol{\psi})$, que tienen una distribución gamma en $\boldsymbol{\psi} = \boldsymbol{\psi}_0$. Por lo tanto, se cumplen las condiciones para que las proposiciones 2.36 y 2.37 sean verdaderas. La variable g_t toma la forma $g_t(y_t; \boldsymbol{\psi}) = |y_t exp(-\theta_{t|t-1})|^{\upsilon}$ en la GED, es $g_t(y_t; \boldsymbol{\psi}) = |y_t exp(-\theta_{t|t-1})|^{\upsilon}$, por lo que se mantienen los mismos argumentos. Como existen todos los momentos de la variable aleatoria $g_t(y_t; \boldsymbol{\psi}_0)$, todos los momentos de $g_t(y_t; \boldsymbol{\psi})$ existen en la vecindad de $\boldsymbol{\psi}_0$ porque $g_t(y_t; \boldsymbol{\psi})$ es igual a $g_t(y_t; \boldsymbol{\psi}_0)$ por un escalar que es, independiente de y_t y viene dada por $exp\theta_{t|t-1}(\boldsymbol{\psi}_0 - \boldsymbol{\psi})$, donde la notación $\theta_{t|t-1}(\boldsymbol{\psi}_0 - \boldsymbol{\psi})$ indica que $\theta_{t|t-1}$ depende de $(\boldsymbol{\psi}_0 - \boldsymbol{\psi})$. Si se elige un valor $\boldsymbol{\psi}^*$ tal que $\theta_{t|t-1}(\boldsymbol{\psi}_0 - \boldsymbol{\psi}^*) > \theta_{t|t-1}(\boldsymbol{\psi}_0 - \boldsymbol{\psi})$, para todo $t = 1, \ldots, T$, entonces $g_t(y_t; \boldsymbol{\psi}^*)$ unirá uniformemente a $g_t(y_t; \boldsymbol{\psi})$ desde arriba. Sea $z_t(\boldsymbol{\psi})$ la tercera derivada, (2.51), generalizada para ser definida con respecto a tres elementos cualquiera en $\boldsymbol{\psi}$, es decir, $\partial^3 \ln f_t(y_t; \psi)/\partial\psi_h\partial\psi_i\partial\psi_j$. Teniendo en cuenta que, como se mostró anteriormente, las derivadas de $\theta_{t|t-1}$ deben estar acotadas en la vecindad de $\boldsymbol{\psi}_0$, es posible construir una variable aleatoria $z_t(\boldsymbol{\psi}^*)$ a partir de $g_t(y_t; \boldsymbol{\psi}^*)'s$ que unirá uniformemente a $z_t(\boldsymbol{\psi})$ desde arriba. Como $g_t(y_t; \boldsymbol{\psi}^*) = g_t(y_t; \boldsymbol{\psi}_0) exp\theta_{t|t-1}(\boldsymbol{\psi}_0 - \boldsymbol{\psi}^*)$, $z_t(\boldsymbol{\psi}^*)$ es una variable aleatoria con primer momento finito (de hecho, todos los momentos son finitos), y por lo tanto, por el teorema ergódico, $\sum z_t(\boldsymbol{\psi}^*)/T$ converge a una cantidad finita, c. □

Los teoremas 2.38 y 2.39 pueden ampliarse para incluir la estimación de parámetros de forma desconocidos, incluidos los grados de libertad. El tratamiento de las terceras derivadas, incluidas las derivadas cruzadas, no implica nada nuevo en principio.

2.3.6. No estacionariedad

Cuando $\phi = 1$, la matriz $\mathbf{D}(\boldsymbol{\psi})$, y por lo tanto $\mathbf{I}(\boldsymbol{\psi})$, ya no es definida positiva. La teoría asintótica habitual no se aplica, ya que el modelo contiene una raíz unitaria. Existe una literatura considerable sobre este tema, comenzando con el trabajo de Dickey y Fuller sobre las implicaciones de ajustar un modelo autorregresivo de primer orden a una caminata aleatoria. Sin embargo, si se impone la raíz unitaria de modo que ϕ se iguale a la unidad, se aplican las asintóticas estándar. Por ejemplo, si ϕ se

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

iguala a la unidad en el modelo Gaussiano $ARMA(1,1)$, (1.2), el estimador MV de θ es asintóticamente normal, siempre que $|\theta| < 1$.

Si $\theta_{1|0}$ es fijo y conocido, el siguiente resultado es un corolario de los teoremas 2.20 a 2.39.

Corolario 2.40. *Cuando ϕ se toma como la unidad en (2.31) pero $b < 1$ y se conoce $\omega = \theta_{1|0}$, la cantidad de información para $\tilde{\kappa}$ es $\sigma_u^4/k^2(1-b)$, con*

$$b = 1 - 2\kappa\sigma_u^2/k + \kappa^2 E[(\partial u_t/\partial\theta)^2], \tag{2.52}$$

suponiendo que se cumplen las condiciones 2.27 y 2.29. Si la distribución condicional de las observaciones es t, GED, gamma generalizada o beta generalizada, el estimador MV de κ es consistente y $\sqrt{T}(\tilde{\kappa} - \kappa)$ tiene una distribución límite normal con media cero y varianza $\kappa^2(1-b)/\sigma_u^4$. La introducción de una constante β en la ecuación dinámica de (2.31) significa que tiene una pendiente (fija). La matriz de información es entonces

$$\boldsymbol{I}\begin{pmatrix}\kappa\\ \beta\end{pmatrix} = \frac{\sigma_u^2}{\kappa^2(1-b)}\begin{bmatrix}\sigma_u^2 & \frac{c}{1-a}\\ \frac{c}{1-a} & \frac{1+a}{1-a}\end{bmatrix}, \tag{2.53}$$

donde $a = 1 - \kappa\sigma_u^2/k\beta$ y los estimadores MV de κ y β son consistentes. Además, $\sqrt{T}(\tilde{\kappa}-\kappa, \tilde{\beta}-\beta)$ tiene una distribución lmite normal con vector de media cero y matriz de covarianza $\boldsymbol{I}^{-1}(\kappa,\beta)$.

Demostración. Aunque $\theta_{t|t-1}$ no es estacionario, las primeras tres derivadas de $\theta_{t|t-1}$ con respecto a κ y ω son SRE, que son estrictamente estacionarios cuando $b < 1$. Por lo tanto, se cumplen las condiciones de consistencia y normalidad asintótica. La matriz de información en (2.53) se obtiene de la misma manera que (2.44) pero reemplazando δ por β. □

Observación 2.41. *De (2.52) se puede ver que $\kappa > 0$ es una condición necesaria para $b < 1$.*

Más problemas que surgen en relación con la inicialización de un filtro no estacionario se puede encontrar en el capítulo sobre localización.

2.3.7. Varios parámetros

El lema 2.16 y el corolario subsiguiente pueden extenderse para tratar modelos en los que θ es un vector de parámetros.

Lema 2.42. *Considere un modelo estático en el que el vector de parámetros $n \times 1$ se divide como $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$, donde n_2 y n_2 son las dimensiones de $\boldsymbol{\theta}_1$ y $\boldsymbol{\theta}_2$, y la matriz de información se divide de manera similar como*

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

$$\boldsymbol{I} = \begin{bmatrix} \boldsymbol{I}_{11} & \boldsymbol{I}_{12} \\ \boldsymbol{I}_{21} & \boldsymbol{I}_{22} \end{bmatrix}.$$

Deje que $\boldsymbol{\theta}_1$ evolucione con el tiempo en un modelo dinámico que depende del vector de score y un vector de parámetros, $\boldsymbol{\psi}$. Cuando la matriz de información en el modelo estático no depende de $\boldsymbol{\theta}_1$

$$\boldsymbol{I}\begin{pmatrix} \boldsymbol{\psi} \\ \boldsymbol{\theta}_2 \end{pmatrix} = \begin{bmatrix} E\left[\frac{\partial \boldsymbol{\theta}_1'}{\partial \boldsymbol{\psi}} \boldsymbol{I}_{11} \frac{\partial \boldsymbol{\theta}_1}{\partial \boldsymbol{\psi}'}\right] & E\left(\frac{\partial \boldsymbol{\theta}_1'}{\partial \boldsymbol{\psi}}\right) \boldsymbol{I}_{12} \\ \boldsymbol{I}_{21} E\left(\frac{\partial \boldsymbol{\theta}_1}{\partial \boldsymbol{\psi}'}\right) & \boldsymbol{I}_{22} \end{bmatrix}, \tag{2.54}$$

donde $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_{1,t|t-1}$

Demostración. Generalizando la transpuesta de (2.28) se obtiene

$$\frac{\partial \ln f_t(y_t|Y_{t-1};\boldsymbol{\psi})}{\partial \boldsymbol{\psi}'} = \frac{\partial \ln f_t}{\partial \boldsymbol{\theta}_1'} \frac{\partial \boldsymbol{\theta}_1}{\partial \boldsymbol{\psi}'},$$

y entonces

$$E_{t-1} \frac{\partial \ln f_t}{\partial \boldsymbol{\psi}} \frac{\partial \ln f_t}{\partial \boldsymbol{\psi}'} = \frac{\partial \boldsymbol{\theta}_1'}{\partial \boldsymbol{\psi}} \left[E_{t-1}\left(\frac{\partial \ln f_t}{\partial \boldsymbol{\theta}_1} \frac{\partial \ln f_t}{\partial \boldsymbol{\theta}_1'}\right)\right] \frac{\partial \boldsymbol{\theta}_1}{\partial \boldsymbol{\psi}'} = \frac{\partial \boldsymbol{\theta}_1'}{\partial \boldsymbol{\psi}} \mathbf{I}_{11} \frac{\partial \boldsymbol{\theta}_1}{\partial \boldsymbol{\psi}'}.$$

□

Cuando $n = 1$, la estructura de la matriz de información se simplifica y sigue una generalización del Teorema 2.20.

Corolario 2.43. *Suponga que $\boldsymbol{\theta}_2$ consta de $n - 1 \geq 1$ parámetros fijos, mientras que θ_1 es variable en el tiempo y depende de un conjunto de parámetros, $\boldsymbol{\psi}$. Cuando los términos en la matriz de información del modelo estático que involucran a θ_1, incluidos los productos cruzados, no dependen de θ_1,*

$$\boldsymbol{I}\begin{pmatrix} \boldsymbol{\psi} \\ \boldsymbol{\theta}_2 \end{pmatrix} = \begin{bmatrix} E(\frac{\partial \ln f_t}{\partial \theta_1})^2 E(\frac{\partial \theta_1}{\partial \boldsymbol{\psi}} \frac{\partial \theta_1}{\partial \boldsymbol{\psi}'}) & E(\frac{\partial \theta_1}{\partial \boldsymbol{\psi}}) E(\frac{\partial \ln f_t}{\partial \theta_1} \frac{\partial \ln f_t}{\partial \boldsymbol{\theta}_2'}) \\ E(\frac{\partial \ln f_t}{\partial \theta_1} \frac{\partial \ln f_t}{\partial \boldsymbol{\theta}_2}) E(\frac{\partial \theta_1}{\partial \boldsymbol{\psi}'}) & E(\frac{\partial \ln f_t}{\partial \boldsymbol{\theta}_2} \frac{\partial \ln f_t}{\partial \boldsymbol{\theta}_2'}) \end{bmatrix}, \tag{2.55}$$

donde $\theta_1 = \theta_{1,t|t-1}$. Las condiciones para la normalidad asintótica son como en la Subsección 2.3.5.

La matriz de información para el modelo de primer orden es

$$\mathbf{I}\begin{pmatrix} \boldsymbol{\psi} \\ \boldsymbol{\theta}_2 \end{pmatrix} = \begin{bmatrix} E\left(\frac{\partial \ln f_t}{\partial \theta_1}\right)^2 \mathbf{D}(\boldsymbol{\psi}) & \begin{pmatrix} 0 \\ 0 \\ \frac{1-\phi}{1-a} \end{pmatrix} E\left(\frac{\partial \ln f_t}{\partial \theta_1} \frac{\partial \ln f_t}{\partial \boldsymbol{\theta}_2'}\right) \\ E\left(\frac{\partial \ln f_t}{\partial \theta_1} \frac{\partial \ln f_t}{\partial \boldsymbol{\theta}_2}\right) \begin{pmatrix} 0 & 0 & \frac{1-\phi}{1-a} \end{pmatrix} & E\left(\frac{\partial \ln f_t}{\partial \boldsymbol{\theta}_2} \frac{\partial \ln f_t}{\partial \boldsymbol{\theta}_2'}\right) \end{bmatrix}, \tag{2.56}$$

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

donde $\mathbf{D}(\boldsymbol{\psi})$ es la matriz en (2.40). Cuando las distribuciones asintóticas de los estimadores MV de θ_1 y $\boldsymbol{\theta}_2$ son independientes, la matriz de información es diagonal por bloques y el bloque superior izquierdo es $\mathbf{I}(\boldsymbol{\psi}) = I.\mathbf{D}(\boldsymbol{\psi})$, como en (2.39). La evaluación analítica del bloque superior izquierdo en (2.55) puede no ser fácil cuando $n_1 > 1$. Para un modelo estacionario de primer orden,

$$\boldsymbol{\theta}_{1,t+1|t} = (\mathbf{I} - \boldsymbol{\Phi})^{-1} + \boldsymbol{\omega} + \Phi\boldsymbol{\theta}_{1,t|t-1} + \mathbf{K}\mathbf{u}_t, \tag{2.57}$$

el vector de parámetros $\boldsymbol{\psi}$ contiene los distintos elementos del vector $\boldsymbol{\omega}$ y las matrices $n \times n$ y $\boldsymbol{K}$. En principio, se puede obtener una expresión de forma cerrada, siempre que los dos primeros momentos de $\mathbf{u}_t$ y $\partial\mathbf{u}_t/\partial\boldsymbol{\theta}_1$ y el asociado producto cruzado no dependen de $\boldsymbol{\theta}_1$. Derivando las ecuaciones dinámicas de (2.57) con respecto a $\boldsymbol{\psi}^{'}$ se obtiene una ecuación de recurrencia estocástica matricial, que será estable si las raíces de la matriz

$$\mathbf{A} = \boldsymbol{\Phi} + \mathbf{K}E(\partial\mathbf{u}_t/\partial\boldsymbol{\theta}^{'})$$

son menores que uno en valor absoluto. La existencia de la matriz de información dependerá de una generalización de la condición de que $b < 1$ en el modelo univariante. Es posible que se necesiten restricciones para la identificabilidad, por ejemplo, que se requiera que $\mathbf{K}$ sea simétrico. La cuestión de si el vector score debe o no ser estandarizado ahora se vuelve más importante. La opción más simple es $\mathbf{u}_t = \partial \ln f_t/\partial\boldsymbol{\theta}_1$, pero puede haber un buen caso para dejar que $\mathbf{u}_t = \mathbf{I}^{-1}(\boldsymbol{\psi}, \boldsymbol{\theta}_2)$. Temas de este tipo se exploran en el capítulo sobre modelos multivariados.

Es posible cierta simplificación cuando cada parámetro variable en el tiempo depende de un conjunto diferente de parámetros subyacentes. El lema a continuación es para $n = 2$, pero esto es simplemente por conveniencia de notación.

Lema 2.44. *Supongamos que hay dos parámetros en $\boldsymbol{\theta}$ pero que $\theta_{j,t|t-1} = f(\boldsymbol{\psi}_j), \quad j = 1, 2$ con los vectores $\boldsymbol{\psi}_1$ y $\boldsymbol{\psi}_2$ sin elementos en común. Cuando la matriz de información en el modelo estático no depende ni de θ_1 ni de θ_2,*

$$\boldsymbol{I}\begin{pmatrix}\boldsymbol{\psi}_1\\ \boldsymbol{\psi}_2\end{pmatrix} = \begin{bmatrix} I_{11}E\left(\frac{\partial\theta_1}{\partial\boldsymbol{\psi}_1}\frac{\partial\theta_1}{\partial\boldsymbol{\psi}_1^{'}}\right) & I_{12}E\left(\frac{\partial\theta_1}{\partial\boldsymbol{\psi}_1}\frac{\partial\theta_2}{\partial\boldsymbol{\psi}_2^{'}}\right) \\ I_{21}E\left(\frac{\partial\theta_2}{\partial\boldsymbol{\psi}_2}\frac{\partial\theta_2}{\partial\boldsymbol{\psi}_1^{'}}\right) & I_{22}E\left(\frac{\partial\theta_2}{\partial\boldsymbol{\psi}_2}\frac{\partial\theta_2}{\partial\boldsymbol{\psi}_2^{'}}\right)\end{bmatrix}. \tag{2.58}$$

Demostración. Por álgebra directa,

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

$$\mathbf{I}\begin{pmatrix}\boldsymbol{\psi}_1\\ \boldsymbol{\psi}_2\end{pmatrix} = E\left[\begin{pmatrix}\frac{\partial \ln f_t}{\partial \theta_1}\frac{\partial \theta_1}{\partial \boldsymbol{\psi}_1}\\ \frac{\partial \ln f_t}{\partial \theta_2}\frac{\partial \theta_2}{\partial \boldsymbol{\psi}_2}\end{pmatrix}\begin{pmatrix}\frac{\partial \ln f_t}{\partial \theta_1}\frac{\partial \theta_1}{\partial \boldsymbol{\psi}_1}\\ \frac{\partial \ln f_t}{\partial \theta_2}\frac{\partial \theta_2}{\partial \boldsymbol{\psi}_2}\end{pmatrix}'\right]$$
$$= \begin{bmatrix} E\left(\frac{\partial \ln f_t}{\partial \theta_1}\right)^2 E\left(\frac{\partial \theta_1}{\partial \boldsymbol{\psi}_1}\frac{\partial \theta_1}{\partial \boldsymbol{\psi}_1'}\right) & E\left(\frac{\partial \ln f_t}{\partial \theta_1}\frac{\partial \ln f_t}{\partial \theta_2}\right) E\left(\frac{\partial \theta_1}{\boldsymbol{\psi}_1}\frac{\partial \theta_2}{\partial \boldsymbol{\psi}_2'}\right) \\ E\left(\frac{\partial \ln f_t}{\partial \theta_1}\frac{\partial \ln f_t}{\partial \theta_2}\right) E\left(\frac{\partial \theta_2}{\boldsymbol{\psi}_2}\frac{\partial \theta_1}{\partial \boldsymbol{\psi}_1'}\right) & E\left(\frac{\partial \ln f_t}{\partial \theta_2}\right)^2 E\left(\frac{\partial \theta_2}{\partial \boldsymbol{\psi}_2}\frac{\partial \theta_2}{\partial \boldsymbol{\psi}_2'}\right) \end{bmatrix}.$$

□

2.4. Modelos de orden superior

Durbin and Koopman (2012) establecieron procedimientos de estimación para modelos de espacio de estados no lineales y no gaussianos en los que la ecuación de medición no lineal es

$$f(y_t|\theta_t;\boldsymbol{\psi}), \quad \theta_t = \omega + \boldsymbol{z}_t'\boldsymbol{\alpha}_t, \quad t = 1,\ldots,T, \tag{2.59}$$

donde $\mathbf{z}_t$ es un vector $m\times 1$ no estocástico, ω es una constante y el vector de estado $m \times 1$, $\boldsymbol{\alpha}_t$, se genera mediante una ecuación de transición lineal,

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{\delta} + \boldsymbol{T}\boldsymbol{\alpha}_t + \boldsymbol{\eta}_t, \quad t = 1,\ldots,T, \tag{2.60}$$

donde T es una matriz $m \times m$, $\boldsymbol{\delta}$ es un vector $m \times 1$ de constantes y $\boldsymbol{\eta}_t$ es un vector $m \times 1$ de perturbaciones no correlacionadas en serie con media cero y matriz de covarianza $\mathbf{Q}$. Por lo general, habrá restricciones en el modelo. Por ejemplo, si $\boldsymbol{\alpha}_t$ es estacionario, $\boldsymbol{\delta}$ se establece en cero si ω no es cero.

El parámetro variable en el tiempo, θ_t, es la señal. Puede depender de una serie de componentes diferentes, incluida una tendencia y estacionalidad, o puede ser que la estructura de Markov para el vector de estado sea un dispositivo para permitir que θ_t sea un modelo de orden superior, como un proceso $ARMA$. La señal está conectada a un conjunto de parámetros en la distribución, $f(y_t|\theta_t;\boldsymbol{\psi})$, por una función de enlace, un término que se origina en el trabajo sobre modelos lineales generalizados de McCullagh and Nelder (1989). Cuando θ_t es la localización y la distribución condicional es gaussiana, el modelo toma la forma de espacio de estado estándar. La actualización óptima de la distribución condicional del estado se realiza mediante el Filtro de Kalman, y la función de verosimilitud se obtiene como subproducto del Filtro de Kalman. De manera más general, la estimación se lleva a cabo mediante técnicas computacionales intensivas, como el muestreo de importancia o la cadena de Markov Monte Carlo.

El modelo DCS correspondiente es

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

$$\begin{gathered} f(y_t|\theta_{t|t-1};\boldsymbol{\psi}), \quad \theta_t = \omega + \boldsymbol{z}_t^{'}\alpha_{t|t-1}, \quad t = 1,\ldots,T, \\ \boldsymbol{\alpha}_{t+1|t} = \boldsymbol{\delta} + \boldsymbol{T}\boldsymbol{\alpha}_{t|t-1} + \boldsymbol{\kappa}u_t, \end{gathered} \tag{2.61}$$

donde u_t es la score condicional y $\boldsymbol{\psi}$ denota parámetros desconocidos en los $m \times 1$ vectores $\boldsymbol{\delta}$ y $\boldsymbol{\kappa}$ y la matriz $m \times m$ T. El filtro en (2.61) es estacionario siempre que los valores propios de la matriz T tengan un módulo menor que uno. En un modelo de localización condicionalmente gaussiano, la ecuación recursiva de (2.61) es el Filtro de Kalman de estado estacionario.

El vector $\mathbf{z}_t$ cambia con el tiempo, de forma determinista, para una cierta formulación del componente estacional, pero en otros casos de interés, es invariante en el tiempo. Para el resto de esta subsección, se supondrá que es invariable en el tiempo y se eliminará el subíndice t. Cuando los componentes de tendencia y estacionales están presentes, la matriz T contendrá raíces unitarias. Siempre que estas raíces sean fijas, en lugar de estimadas, no plantean ningún problema para la teoría asintótica.

Teorema 2.45. *Cuando $\boldsymbol{\alpha}_{t+1|t}$ es estacionario y se cumple la Condición 2.18, una condición suficiente para la existencia de la matriz de información de los parámetros desconocidos en $\boldsymbol{\psi}$ es que todas las raíces de la matriz $m^2 \times m^2$*

$$\boldsymbol{B} = \boldsymbol{T} \otimes \boldsymbol{T} + (\boldsymbol{\kappa}\boldsymbol{z}^{'} \otimes \boldsymbol{T} + \boldsymbol{T} \otimes \boldsymbol{\kappa}\boldsymbol{z}^{'})E\left(\frac{\partial u_t}{\partial \theta}\right) + \boldsymbol{\kappa}\boldsymbol{z}^{'} \otimes \boldsymbol{\kappa}\boldsymbol{z}^{'}E\left(\frac{\partial u_t}{\partial \theta}\right)^2, \tag{2.62}$$

donde $\otimes$ es el producto de Kronecker, tienen módulo menor que uno. Una condición necesaria es que las raíces de la matriz $m \times m$

$$\boldsymbol{A} = \boldsymbol{T} + \boldsymbol{\kappa}\boldsymbol{z}^{'}E(\partial u_t/\partial\theta) \tag{2.63}$$

tenga módulo menor que uno.

Demostración. Considere los parámetros en $\boldsymbol{\kappa}$. Para todo $t = 1,\ldots,T$,

$$\frac{\partial \boldsymbol{\alpha}_{t+1|t}}{\partial \kappa_i} = \boldsymbol{T}\frac{\partial \boldsymbol{\alpha}_{t|t-1}}{\partial \kappa_i} + \boldsymbol{\kappa}\frac{\partial u_i}{\partial \kappa_i} + \boldsymbol{w}_i u_t, \quad i = 1,\ldots,m,$$

donde $\boldsymbol{w}_i$ es un vector que tiene uno en la i-ésima posición y ceros en cualquier otra parte. Luego

$$\frac{\partial u_t}{\partial \kappa_i} = \frac{\partial u_t}{\partial \theta_{t|t-1}}\frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\alpha}_{t|t-1}^{'}}\frac{\partial \boldsymbol{\alpha}_{t|t-1}}{\partial \kappa_i} = \frac{\partial u_t}{\partial \theta_{t|t-1}}\boldsymbol{z}^{'}\frac{\partial \boldsymbol{\alpha}_{t|t-1}}{\partial \kappa_i}$$

entonces

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

$$\frac{\partial \boldsymbol{\alpha}_{t+1|t}}{\partial \kappa_i} = \left(\boldsymbol{T} + \boldsymbol{\kappa}\boldsymbol{z}^{'} \frac{\partial u_t}{\partial \theta_{t|t-1}} \right) \frac{\partial \boldsymbol{\alpha}_{t|t-1}}{\kappa_i} + \boldsymbol{w}_i u_t,$$

ó

$$\frac{\partial \boldsymbol{\alpha}_{t+1|t}}{\partial \kappa_i} = \boldsymbol{X}_t \frac{\partial \boldsymbol{\alpha}_{t|t-1}}{\partial \kappa_i} + \boldsymbol{w}_i u_t, \quad i = 1, \ldots, m, \tag{2.64}$$

donde

$$\boldsymbol{X}_t = \boldsymbol{T} + \frac{\partial u_t}{\partial \theta_{t|t-1}} \boldsymbol{\kappa}\boldsymbol{z}^{'} \tag{2.65}$$

Tomando esperanzas condicionales de (2.64) se obtiene

$$E_{t-1} \frac{\partial \boldsymbol{\alpha}_{t+1|t}}{\partial \kappa_i} = \boldsymbol{A} \frac{\partial \boldsymbol{\alpha}_{t|t-1}}{\partial \kappa_i}, \quad i = 1, \ldots m,$$

donde

$$\boldsymbol{A} = E_{t-1} \boldsymbol{X}_t = E_{t-1} \left(\boldsymbol{T}\boldsymbol{\kappa}\boldsymbol{z}^{'} \frac{\partial u_t}{\partial \theta_{t|t-1}} \right) = \boldsymbol{T} + E \left(\frac{\partial u_t}{\partial \theta} \right) \boldsymbol{\kappa}\boldsymbol{z}^{'}.$$

El vector de derivada incondicional es cero si las raíces de $\mathbf{A}$ son menores que uno en valor absoluto.

La matriz de información depende de la matriz $m \times m$

$$E \left(\frac{\partial \theta_{t+1|t}}{\partial \boldsymbol{\kappa}} \frac{\partial \theta_{t+1|t}}{\partial \boldsymbol{\kappa}^{'}} \right). \tag{2.66}$$

Aunque en la fórmula solo aparece la derivada de $\theta_{t+1|t}$, es necesario evaluar las matrices de covarianza de las derivadas de los elementos de $\boldsymbol{\alpha}_{t+1|t}$ con respecto a cada elemento en $\boldsymbol{\kappa}$ para encontrarla. De (2.64)

$$\begin{aligned} \frac{\partial \boldsymbol{\alpha}_{t+1|t}}{\partial \kappa_i} \frac{\partial \boldsymbol{\alpha}_{t+1|t}}{\partial \kappa_j} = {} & \boldsymbol{X}_t \frac{\partial \boldsymbol{\alpha}_{t|t-1}}{\partial \kappa_i} \frac{\partial \boldsymbol{\alpha}_{t|t-1}}{\partial \kappa_j} \boldsymbol{X}_t^{'} + \boldsymbol{X}_t \frac{\partial \boldsymbol{\alpha}_{t|t-1}}{\partial \kappa_i} \boldsymbol{w}_i^{'} u_t \\ & + u_t \boldsymbol{w}_j \frac{\partial \boldsymbol{\alpha}_{t+1|t}}{\partial \kappa_j} \boldsymbol{X}_t^{'} + u_t^2 \boldsymbol{w}_i \boldsymbol{w}_j^{'}, \end{aligned}$$

para $i, j = 1, \ldots, m$. Podemos escribir

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

$$\begin{aligned}\text{vec}\left(\frac{\partial \boldsymbol{\alpha}_{t+1|t}}{\partial \kappa_i}\frac{\partial \boldsymbol{\alpha}_{t+1|t}^{'}}{\partial \kappa_j}\right) &= \boldsymbol{X}_t \otimes \boldsymbol{X}_t.\text{vec}\left(\frac{\partial \boldsymbol{\alpha}_{t|t-1}}{\partial \kappa_i}\frac{\partial \boldsymbol{\alpha}_{t|t-1}^{'}}{\partial \kappa_j}\right)\\ &+ u_t\text{vec}\left(\boldsymbol{X}_t\frac{\partial \boldsymbol{\alpha}_{t|t-1}}{\partial \kappa_i}\boldsymbol{w}_i^{'} + \boldsymbol{w}_j\frac{\partial \boldsymbol{\alpha}_{t|t-1}^{'}}{\partial \kappa_j}\boldsymbol{X}_t^{'}\right)\\ &+ u_t^2\text{vec}(\boldsymbol{w}_i\boldsymbol{w}_j^{'}),\end{aligned} \tag{2.67}$$

donde el operador vec(.) indica que las columnas de la matriz se apilan una sobre otra. Tomando esperanzas condicionales

$$\begin{aligned}E_{t-1}\text{vec}\left(\frac{\partial \boldsymbol{\alpha}_{t+1|t}}{\partial \kappa_i}\frac{\partial \boldsymbol{\alpha}_{t+1|t}^{'}}{\partial \kappa_j}\right) &= [E_{t-1}(\boldsymbol{X}_t \otimes \boldsymbol{X}_t)]\text{vec}\left(\frac{\partial \boldsymbol{t|t-1}}{\partial \kappa_i}\frac{\partial \boldsymbol{\alpha}_{t|t-1}^{'}}{\kappa_j}\right)\\ &+ \text{vec}\left([E_{t-1}(u_t\boldsymbol{X}_t)]\frac{\partial \boldsymbol{\alpha}_{t|t-1}}{\partial \kappa_i}\boldsymbol{w}_i^{'} + \boldsymbol{w}_j\frac{\partial \boldsymbol{\alpha}_{t|t-1}^{'}}{\kappa_j}[E_{t-1}(u_t\boldsymbol{X}_t)]^{'}\right)\\ &+ [E_{t-1}(u_t^2)]\text{vec}(\boldsymbol{w}_i\boldsymbol{w}_j^{'}).\end{aligned}$$

Ahora

$$\mathbf{X}_t \otimes \mathbf{X}_t = \mathbf{T}\otimes\mathbf{T} + \boldsymbol{\kappa}\mathbf{z}^{'}\frac{\partial u_t}{\partial \theta_{t|t-1}}\otimes\mathbf{T} + \mathbf{T}\otimes\boldsymbol{\kappa}\mathbf{z}^{'}\frac{\partial u_t}{\partial \theta_{t|t-1}} + \kappa\mathbf{z}^{'}\frac{\partial u_t}{\partial \theta_{t|t-1}}\otimes\boldsymbol{\kappa}\mathbf{z}^{'}\frac{\partial u_t}{\partial \theta_{t|t-1}}$$

y su esperanza incondicional, $E(\mathbf{X}_t\otimes\mathbf{X}_t)$, es $\mathbf{B}$, como en (2.62). Así como el teorema 2.24 requiere $b<1$, aquí las raíces de $\mathbf{B}$ deben ser menores que uno en valor absoluto para que exista la esperanza incondicional de (2.67). Siendo este el caso,

$$\begin{aligned}&E\text{vec}\left(\frac{\partial \boldsymbol{\alpha}_{t+1|t}}{\partial \kappa_i}\frac{\partial \boldsymbol{\alpha}_{t+1|t}^{'}}{\partial \kappa_j}\right)\\ &= (\mathbf{I}-\mathbf{B})^{-1}\left[\text{vec}\left([E(u_t\mathbf{X}_t)]\frac{\partial \boldsymbol{\alpha}_{t|t-1}}{\partial \kappa_i}\boldsymbol{\omega}_i^{'} + \boldsymbol{\omega}_j\frac{\partial \boldsymbol{\alpha}_{t|t-1}^{'}}{\partial \kappa_j}[E(u_t\mathbf{X}_t)]^{'}\right) + [E(u_t^2)]\text{vec}(\boldsymbol{\omega}_i\boldsymbol{\omega}_j^{'})\right],\end{aligned} \tag{2.68}$$

y una vez encontradas las esperanzas, el ij -ésimo elemento, $i,j=1,\ldots,m$, en (2.66) viene dado por

$$(\text{vec}(\mathbf{z}\mathbf{z}^{'}))^{'}E\text{vec}\left(\frac{\partial \boldsymbol{\alpha}_{t+1|t}}{\partial \kappa_i}\frac{\partial \boldsymbol{\alpha}_{t+1|t}^{'}}{\partial \kappa_j}\right) = \left[E\text{vec}\left(\frac{\partial \boldsymbol{\alpha}_{t+1|t}}{\partial \kappa_i}\frac{\partial \boldsymbol{\alpha}_{t+1|t}^{'}}{\partial \kappa_j}\right)\right](\mathbf{z}\otimes\mathbf{z}). \tag{2.69}$$

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Cuando hay parámetros desconocidos en $\mathbf{T}$, las condiciones clave para la inferencia de MV son las mismas. Sea ϕ un parámetro en $\mathbf{T}$. Entonces

$$\frac{\partial \boldsymbol{\alpha}_{t+1|t}}{\partial \phi} = \mathbf{T}\frac{\partial \boldsymbol{\alpha}_{t|t-1}}{\partial \phi} + \boldsymbol{\kappa}\frac{\partial u_t}{\partial \phi} + \frac{\partial \mathbf{T}}{\partial \phi}\boldsymbol{\alpha}_{t|t-1} = \mathbf{X}_t\frac{\partial \boldsymbol{\alpha}_{t|t-1}}{\partial \phi} + \frac{\partial \mathbf{T}}{\partial \phi}\boldsymbol{\alpha}_{t|t-1}.$$

□

La evaluación de la matriz de información puede ser complicada, pero la evaluación de $\mathbf{B}$ es directa. Una vez que se ha establecido que las raíces de $\mathbf{B}$ tienen un módulo menor que uno, la matriz de información se puede calcular numéricamente.

2.5. Pruebas

Antes de ajustar un modelo DCS, es recomendable probar si la variación de tiempo está, de hecho, presente. Cuando una serie es aleatoria, un modelo DCS de primer orden no será identificable porque, como se señaló en el Corolario 2.21, la matriz de información es singular. Incluso si se induce a un procedimiento de optimización numérica a la convergencia, no se debe realizar una prueba de Wald en la que la hipótesis nula sea $\kappa = 0$.

El principio del multiplicador de Lagrange se puede utilizar para construir pruebas adecuadas contra la correlación serial en la característica de interés teniendo en cuenta la forma de la distribución condicional postulada para el modelo dinámico. Dichas pruebas deberían poder proteger contra la baja potencia como consecuencia de valores atípicos. Una vez que se ha ajustado un modelo, se pueden construir pruebas de diagnóstico basadas en principios similares. Estas pruebas pueden complementarse con procedimientos para evaluar la bondad de ajuste de la distribución condicional supuesta.

2.5.1. Correlación serial

Las pruebas estándar contra la correlación serial se basan en las autocorrelaciones de la muestra,$r(\tau) = c(\tau)/c(0)$, donde

$$c(\tau) = T^{-1}\sum_{t=\tau+1}^{T}(y_t - \bar{y})(y_{t-\tau} - \bar{y}), \quad \tau = 1, 2, 3, \ldots \tag{2.70}$$

(También se puede usar un divisor de $T - \tau$). En particular, la prueba de portmanteau utiliza la estadística

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

$$Q(P) = T\sum_{\tau=1}^{P} r^2(\tau) \tag{2.71}$$

donde P es un número entero distinto de cero; la modificación de Ljung-Box es

$$Q^*(P) = T(T+2)\sum_{\tau=1}^{T}(T-\tau)^{-1}r^2(\tau).$$

Tanto $Q(P)$ como $Q^*(P)$ son asintóticamente χ^2_P cuando las observaciones son independientes. La suposición más débil de que las observaciones no están correlacionadas en serie no es suficiente.

La prueba de Portmanteau se puede derivar como una prueba del multiplicador de Lagrange contra un proceso de media móvil gaussiano de orden P. De manera similar, se puede desarrollar una prueba DCS para correlación serial considerando el proceso dinámico $(0, P-1)$

$$\theta_{t+1|t} = \omega + \kappa_0 u_t + \cdots + \kappa_{P-1}u_{t-P+1}, \quad t = 1, \ldots, T, \tag{2.72}$$

donde u_t es la score de la t-ésima observación.

Proposición 2.46. *Considere un modelo, como en el Corolario 2.43, donde un parámetro en θ varía en el tiempo según (2.72), mientras que los otros parámetros son fijos.*

El estadístico de prueba del Multiplicador de Lagrange para probar la hipótesis nula $\kappa_1 = \kappa_1 = \cdots = \kappa_{P-1} = 0$ es el estadístico de portmanteau

$$Q_u(P) = T\sum_{\tau=1}^{P} r_u^2(\tau), \tag{2.73}$$

donde $r_u(\tau)$ es la τ-ésima autocorrelación de la muestra de u_t . La distribución asintótica de $Q_u(P)$ bajo la hipótesis nula es χ^2_P.

Demostración. Por simplicidad, suponga que $k = 1$; cuando $k \neq 1$, se cancela en la expresión final. Tenemos

$$\frac{\partial\theta_{t+1|t}}{\partial\kappa_j} = \sum_{i=0}^{P-1}\kappa_i\frac{\partial u_{t-i}}{\partial\kappa_j}, \quad j = 0, \ldots, P-1,$$

pero bajo la hipótesis nula, donde $\kappa = (\kappa_0, \ldots, \kappa_{P-1})' = \mathbf{0}$,

$$\frac{\partial\theta_{t+1|t}}{\kappa_j} = u_{t-j}, \quad j = 0, \ldots, P-1,$$

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

por lo tanto

$$\boldsymbol{D}(\kappa) = E\left(\frac{\partial \theta_{t|t-1}}{\partial \kappa}\frac{\partial \theta_{t|t-1}}{\partial \kappa'}\right) = \sigma_u^2 \boldsymbol{I}_P$$

y así, como en (2.39),

$$\boldsymbol{I}(\kappa) = \sigma_u^2 \boldsymbol{D}(\kappa) = \sigma_u^4 \boldsymbol{I}_P.$$

El resultado se sigue de la fórmula estándar del Multiplicador de Lagrange (LM), es decir

$$LM = \frac{\partial \ln L}{\partial \kappa'} \boldsymbol{I}^{-1}(\kappa) \frac{\partial \ln L}{\partial \kappa},$$

porque

$$\frac{\partial \ln f_t}{\kappa_j} = \frac{\partial \ln f_t}{\theta_{t|t-1}}\frac{\partial \theta_{t|t-1}}{\kappa_j} = u_t u_{t-1-j}, \quad j = 0, 1, \ldots, P-1. \tag{2.74}$$

La estimación de parámetros fijos, como los grados de libertad o la media incondicional, no hace ninguna diferencia en la forma del resultado, porque en (2.55) $\boldsymbol{\psi} = \kappa$ y $E(\partial \theta_{t|t-1}/\partial \kappa) = \mathbf{0}$. □

Observación 2.47. *Bajo la hipótesis nula, la condición de primer orden,* $\partial \ln L/\partial \omega = 0$*, implica que* $\bar{u} = T^{-1}\sum_{t=1}^{T} u_t = 0$ *porque*

$$\frac{\partial \ln L}{\partial \omega} = sum_{t=1}^{T} \frac{\partial \ln f_t(\theta_{t|t-1})}{\partial \theta_{t|t-1}} \frac{\partial \theta_{t|t-1}}{\partial \omega} = \sum_{t=1}^{T} u_t.$$

Observación 2.48. *La estadística de Ljung-Box*

$$Q_u^*(P) = T(T+2)\sum_{\tau=1}^{P}(T-\tau)^{-1} r_u^2(\tau),$$

también se puede utilizar. La distribución asintótica bajo la hipótesis nula es nuevamente χ_P^2*.*

Los grados de libertad en una distribución t pueden estimarse sin afectar la distribución asintótica de $Q_u(P)$. La estimación de parámetros dinámicos, por otro lado, afectará la distribución asintótica del estadístico portmanteau construido a partir de residuos. La solución, como en la prueba de Box-Pierce, es ajustar la distribución chicuadrado de referencia reduciendo los grados de libertad por el número de parámetros dinámicos estimados, es decir, $p + q$ cuando se utiliza un modelo $ARMA(p, q)$. Otra posibilidad podría ser desarrollar pruebas del Multiplicador de Lagrange.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

No es raro encontrarse con una situación en la que la correlación serial en $\theta_{t|t-1}$ es muy persistente. En el modelo de primer orden (2.30), la persistencia significa que el parámetro ϕ está cerca de uno. En estas circunstancias, una prueba de estacionariedad, como la propuesta por Nyblom and Mäkeläinen (1983) o la prueba más general de Kwiatkowski et al. (1992) tenderá a tener mayor potencia que la prueba de portmanteau. Las estadísticas de la prueba podrían construirse a partir de variables de score condicional.

2.5.2. Bondad de ajuste de las distribuciones

Evaluar qué tan bien se ajusta una distribución condicional particular a los datos es una parte importante de la metodología de los modelos DCS. La bondad de ajuste de una distribución gaussiana se prueba de forma rutinaria con el estadístico Bowman-Shenton (Jarque-Bera), que se basa en la asimetría y la curtosis de la muestra. El hecho de que la estadística de prueba consta de dos partes es útil para determinar si un rechazo se basa principalmente en la asimetría o la curtosis.

De manera más general, la bondad del ajuste se evalúa con referencia a los cuantiles de la distribución teórica supuesta. Hay dos enfoques. El primero se basa en el QQplot, en el que los T cuantiles equidistantes de la distribución de comparación se trazan en el eje vertical, y las estadísticas de orden de la muestra,$y_1 \leq y_2 \leq \cdots \leq y_T$, se trazan en el eje horizontal. Si la muestra proviene de la distribución de comparación, la gráfica se aproximará a una línea recta de 45 grados. Formalmente, la gráfica es de la función cuantil $F^{-1}(k/(T+1))$, contra y_k, $k = 1, \ldots, T$. Tenga en cuenta que, aparte de unas pocas excepciones, como la distribución exponencial, las expresiones de forma cerrada uniformes, de Weibull y logarítmicas para funciones cuantiles, $F^{-1}(y_t)$, generalmente no están disponibles, y su evaluación a menudo involucra métodos numéricos.

Una segunda gráfica, que esencialmente presenta la misma información, pero de una manera diferente, utiliza la transformada integral de probabilidad (PIT). El PIT de una variable viene dado por su CDF, es decir, $PIT(y) = F(y)$. Por construcción [5], el PIT tiene una distribución uniforme estándar, es decir, el rango es $[0, 1]$. Si los PIT no están distribuidos uniformemente, la forma del histograma puede ser informativa.

Por ejemplo, una joroba indica que los pronósticos son demasiado estrechos y que las colas no están representadas adecuadamente. La hipótesis de que un conjunto de T observaciones proviene de una distribución paramétrica particular se puede probar utilizando la estadística de Kolmogorov-Smirnov

[5] Un cambio de variable de y a F(y) entrega la función de densidad de probabilidad

$$f(F(y)) = f(y)\partial y/\partial F(y) = 1, \quad 0 \leq F(y) \leq 1$$

.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

$$KS_{PIT} = \max_j |PIT_{(j)} - j/T|,$$

donde $PIT_{(j)}$, $j = 1, \ldots, T$ denota el $PIT(y_t)$ ordenado; véase, por ejemplo, Siegal (1956, pág. 48). Cuando se han estimado los parámetros, la distribución del estadístico de Kolmogorov-Smirnov normalmente tendrá que hallarse mediante simulación.

Ejemplo 2.49. *El PIT de una variable que tiene una distribución exponencial con media θ es simplemente*

$$PIT(y_t) = F(y_t) = \int_0^{y_t} \theta^{-1} exp(-x/\theta) dx = 1 - exp(-y_y/\theta).$$

Ejemplo 2.50. *El PIT para la distribución t_ν estándar es una función beta incompleta regularizada, $beta(z_t, \nu/2, \nu/2)$, con $z_t = ((y_t^2 + \nu)^{-1/2} y_t + 1)/2$.*

Berkowitz (2001) sugirió transformar los PIT en variables normales estándar utilizando un algoritmo para calcular la inversa de la función de distribución gaussiana (la función cuantil). Esto abre una amplia gama de pruebas apropiadas para la distribución normal, incluida la prueba de Bowman-Shenton. Berkowitz argumentó que tales pruebas tienden a tener mayor potencia que las pruebas no paramétricas asociadas con la distribución uniforme.

2.5.3. Residuos

Diebold et al. (1997) discuten varias formas en que los PIT pueden usarse para evaluar los esquemas de pronóstico. En el presente contexto, esto significa verificar la especificación del modelo utilizando los PIT de los errores de predicción de un paso adelante. Por ejemplo, los gráficos de las funciones de autocorrelación de los PIT y de sus potencias pueden indicar el origen de la dependencia serial. Tales gráficos pueden ser particularmente útiles para el modelado no paramétrico.

Las pruebas basadas en residuos estandarizados, sus PIT y las variables normales creadas a partir de los PIT ofrecen tres alternativas para la verificación diagnóstica. Debido a que todas las variables son transformaciones entre sí, no está claro cuál será preferible. De hecho, es casi seguro que la respuesta depende de la distribución supuesta. El ajuste de los modelos DCS también ofrece una cuarta posibilidad, a saber, el uso de las scores condicionales. Los scores condicionales pueden ser apropiados para probar la correlación serial y, en los casos en los que tienen una distribución conocida, puede ser valioso comparar las distribuciones empírica y teórica.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

2.5.4. Ajuste del modelo

El ajuste general de un modelo se puede evaluar de varias maneras. Los diversos criterios que se analizan aquí son de aplicación general y, por lo tanto, se pueden utilizar para comparar los modelos DCS con otros modelos. Se supone que la estimación es por MV.

El logaritmo de la verosimilitud maximizado es la medida básica de la bondad del ajuste, pero para comparar diferentes modelos, a menudo se tiene en cuenta el número de parámetros, n, estimado. El criterio de información de Akaike (AIC) se define como

$$AIC = -2\ln(\boldsymbol{\psi}) + 2n, \tag{2.75}$$

mientras que el criterio de información de Bayes (BIC) tiene $2n$ reemplazado por $n \ln T$. También se pueden emplear pruebas predictivas posteriores a la muestra. La probabilidad predictiva, a veces llamada log-score, es simple y efectiva. Observar los residuos posteriores a la muestra, la score y las PIT también puede proporcionar información valiosa. Mitchell and Wallis (2011) proporcionaron una discusión reciente de los temas involucrados.

2.6. Variables explicativas

El parámetro variable en el tiempo puede depender de un conjunto de variables explicativas observables, indicadas por el vector $\mathbf{w}_t$ de dimensión $k \times 1$, así como de sus propios valores pasados y la score. El modelo se puede configurar como

$$\theta_{t|t-1} = \mathbf{w}_t^{'}\boldsymbol{\gamma} + \theta^{\dagger}_{t|t-1}, \quad t = 1, \ldots, T, \tag{2.76}$$

donde, en el caso de primer orden,

$$\theta^{\dagger}_{t+1|t} = \phi\theta^{\dagger}_{t|t-1} + \kappa u_t, \quad t = 1, \ldots, T,$$

con $\theta^{\dagger}_{1|0} = 0$. Las variables explicativas son estrictamente exógenas en el sentido de que son independientes de las variables aleatorias que entran en el resto del modelo en todos los periodos de tiempo.

El siguiente resultado es una generalización relativamente directa del Teorema 2.20. Una media constante, ω, es un caso especial, obtenido cuando $\mathbf{w}_t$ es un escalar igual a la unidad. Si, como en (2.30), la formulación del modelo es

$$\theta_{t+1|t} = \mathbf{w}_t^{'}\boldsymbol{\delta} + \phi\theta_{t|t-1} + \kappa u_t, \tag{2.77}$$

la matriz de información se construye siguiendo las líneas del Teorema 2.24.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Corolario 2.51. *Suponga que las variables explicativas son débilmente estacionarias, con media $\boldsymbol{\mu}_w$ y segundo momento Λ_w, y son estrictamente exógenas. La matriz de información para el modelo (2.76) es*

$$\boldsymbol{I}\begin{pmatrix}\kappa\\ \phi\\ \boldsymbol{\gamma}\end{pmatrix} = \frac{\sigma_u^2}{k^2(1-b)}\begin{bmatrix} A & D & E\boldsymbol{\mu}_w^{'} \\ D & B & F\boldsymbol{\mu}_w^{'} \\ E\boldsymbol{\mu}_w & F\boldsymbol{\mu}_w & \boldsymbol{C}_w \end{bmatrix}, \tag{2.78}$$

donde A, B, C, D, E y F son como en (2.40), y

$$\boldsymbol{C}_w = (1+\phi^2)\Lambda_w + 2\phi\Lambda_w(1) + \frac{2a(1-\phi)^2}{1-a}\boldsymbol{\mu}_w\boldsymbol{\mu}_w^{'},$$

con $\Lambda_w(1) = E(\boldsymbol{w}_t\boldsymbol{w}_{t-1}^{'}) = E(\boldsymbol{w}_{t-1}\boldsymbol{w}_t^{'})$.

Demostración. El modelo (2.76) se puede reescribir en la forma de (2.77), es decir,

$$\theta_{t+1|t} = (\mathbf{w}_{t+1} - \phi\mathbf{w}_t)^{'}\boldsymbol{\gamma} + \phi\theta_{t|t-1} + \kappa u_t \tag{2.79}$$

Derivando (2.79) con respecto a $\boldsymbol{\gamma}$ se obtiene

$$\frac{\partial\theta_{t+1|t}}{\partial\boldsymbol{\gamma}} = x_t\frac{\partial\theta_{t|t-1}}{\partial\boldsymbol{\gamma}} + \mathbf{w}_{t+1} - \mathbf{w}_t$$

y entonces

$$E_{t-1}\left(\frac{\partial\theta_{t+1|t}}{\partial\boldsymbol{\gamma}}\right) = a\frac{\partial\theta_{t|t-1}}{\partial\boldsymbol{\gamma}} + E_{t-1}(\mathbf{w}_{t+1}) - \phi\mathbf{w}_t.$$

Tomar esperanzas incondicionales obtenemos

$$E\left(\frac{\partial\theta_{t+1|t}}{\partial\boldsymbol{\gamma}}\right) = \frac{1-\phi}{1-a}\boldsymbol{\mu}_w.$$

Ahora

$$\begin{aligned} E_{t-1}\left(\frac{\partial\theta_{t+1|t}}{\partial\boldsymbol{\gamma}}\frac{\partial\theta_{t+1|t}}{\partial\boldsymbol{\gamma}'}\right) =& b\frac{\partial\theta_{t|t-1}}{\partial\boldsymbol{\gamma}}\frac{\partial\theta_{t|t-1}}{\partial\boldsymbol{\gamma}'} \\ &+ E_{t-1}((\mathbf{w}_{t+1} - \phi\mathbf{w}_t)(\mathbf{w}_{t+1} - \phi\mathbf{w}_t)^{'}) \\ &+ \frac{\partial\theta_{t|t-1}}{\partial\boldsymbol{\gamma}}E_{t-1}(x_t(\mathbf{w}_{t+1} - \phi\mathbf{w}_t)^{'}) \\ &+ E_{t-1}(x_t(\mathbf{w}_{t+1} - \phi\mathbf{w}_t))\frac{\partial\theta_{t|t-1}}{\partial\boldsymbol{\gamma}'}. \end{aligned} \tag{2.80}$$

Debido a que se supone que $\mathbf{w}_t$ es estrictamente exógeno,

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

$$\frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\gamma}} E_{t-1}(x_t(\mathbf{w}_{t+1} - \phi \mathbf{w}_t)') = a \frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\gamma}} E_{t-1}(\mathbf{w}_{t+1} - \phi \mathbf{w}_t)'$$

y cuando se toman esperanzas incondicionales,

$$aE\left[\frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\gamma}}(\mathbf{w}_{t+1} - \phi \mathbf{w}_t)'\right] = aE\frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\gamma}}(1-\phi)\boldsymbol{\mu}_w' = \frac{a(1-\phi)^2}{1-a}\boldsymbol{\mu}_w \boldsymbol{\mu}_w'$$

Tomando las esperanzas incondicionales en (2.80) se obtiene $(1-b)^{-1}\mathbf{C}_w$.

□

Observación 2.52. *Se obtiene un estimador de la matriz de covarianza asintótica reemplazando* Λ_w *y* $\Lambda_w(1)$ *por* $T^{-1}\sum \boldsymbol{w}_t \boldsymbol{w}_t'$ *y* $T^{-1}\sum \boldsymbol{w}_t \boldsymbol{w}_{t-1}'$*, respectivamente.*

Observación 2.53. *Las variables explicativas pueden ser fijas. Tales variables podrían incluir dummies estacionales y términos trigonométricos. La teoría asintótica del Corolario 2.51 se aplica a variables fijas si*

$$\lim T^{-1}\sum \boldsymbol{w}_t \boldsymbol{w}_t' = \Lambda_w.$$

También se pueden incluir tendencias temporales deterministas, aunque la condición anterior debe modificarse ligeramente.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Capítulo 3

Localización

Los modelos de componentes lineales gaussianos no observados, como el paseo aleatorio más el ruido de (1.1), desempeñan un papel central en el modelado de series temporales. El Filtro de Kalman y el suavizador asociado proporcionan la base para un tratamiento estadístico completo. Los estimadores filtrados y suavizados de la señal minimizan el MSE del error de estimación, la probabilidad se da como un subproducto de los errores de predicción de un paso producidos por el Filtro de Kalman y la distribución predictiva completa de varios pasos tiene una distribución gaussiana conocida.

Ahora queremos confrontar una situación en la que la localización está inserta en un ruido que no es gaussiano y que puede arrojar observaciones que, cuando se juzgan con el criterio gaussiano, son atípicas. El objetivo es desarrollar un modelo basado en la observación que sea manejable y conserve algunas de las características deseables del modelo gaussiano lineal. Dejar que la dinámica sea impulsada por la score condicional conduce a un modelo que no solo es fácil de implementar, sino que también facilita el desarrollo de una teoría integral y relativamente sencilla para la distribución asintótica del estimador MV.

Modelar el ruido aditivo con una distribución t de Student es efectivo y teóricamente sencillo. De hecho, los atractivos de usar la distribución t para protegerse contra valores atípicos en modelos estáticos están bien documentados; véase, por ejemplo,Lange et al. (1989). La distribución general de errores es menos satisfactoria por las razones que se analizan al final del capítulo. Sin embargo, apunta a algunas conexiones interesantes con los cuantiles dinámicos.

Las primeras cuatro secciones describen el modelo de localización DCS-t, con la Sección 3.2 derivando momentos y funciones de autocorrelación, y la Sección 3.3 presentando la distribución asintótica del estimador MV. La sección 3.5 muestra cómo la forma espacio estados permite generalizar el modelo de primer orden, a partir del cual

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

se introducen la tendencia y la estacionalidad. El suavizado se describe en la Sección 3.7, y la última subsección ofrece una justificación adicional para el uso de la partitura para impulsar la dinámica. La mecánica del pronóstico se establece en la Sección 3.8.

3.1. Modelos de Localización Dinámicos t-student

Cuando la localización varia en el tiempo es posible capturar este cambio por medio de un modelo en el que $y_t|Y_{t-1}$ tiene una distribución t_ν, (2.1), con media condiciona $\mu_{t|t-1}$, generada por la función lineal

$$u_t = \left(1 + \frac{(y_t - \mu_{t|t-1})^2}{\nu e^{2\lambda}}\right)^{-1} v_t \quad t = 1, \ldots, T, \tag{3.1}$$

donde $v_t = y_t - \mu_{t|t-1}$ es el error de predicción y $\varphi = exp(\lambda)$ es la escala invariante en el tiempo. Diferenciando la log-densidad, (2.16), se muestra que u_t es proporcional a la score condicional

$$\frac{\partial \ln f_t}{\partial \mu_{t|t-1}} = (\nu + 1)\nu^{-1} exp(-2\lambda) u_t.$$

No se imponen restricciones a los grados de libertad ν, además de ser estos positivos. El factor de escala, $\exp(2\lambda)$, se cancela si la score se divide por la cantidad de información para la localización, como se indica en (2.18).

El modelo puede ser escrito de manera similar a la forma de inovaciones Gausianas. Por lo tanto, en el caso del modelo de primer orden corresponde con (1.3), el cual es

$$\begin{aligned} y_t &= \mu_{t|t-1} + v_t = \mu_{t|t-1} + exp(\lambda)\varepsilon_t, \quad t = 1, \ldots, T \\ \mu_{t+1|t} &= \delta + \phi \mu_{t|t-1} + \kappa u_t, \end{aligned} \tag{3.2}$$

donde ε_t es una variable t indendiente, fijando la escala φ en uno. Para un modelo estacionario, se reparametriza en terminos de la media incondicional, ω, como en (2.31), de la siguiente manera

$$\mu_{t+1|t} = \omega + \mu^\dagger_{t+1|t}, \quad \mu^\dagger_{t+1|t} = \phi \mu^\dagger_{t|t-1} + \kappa u_t, \quad t = 1, \ldots, T, \tag{3.3}$$

donde $\mu_{1|0}$ puede fijarse en ω.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

3.2. Propiedades básicas

El error de predicción v_t, en (3.2), es por construcción un varibale t independiente e identicamente distribuida, con media cero y escala $exp(\lambda)$. Es más, las u_t son tambien IID. Las propriedades de u_t se siguen de la relación con la distribución beta.

Proposición 3.1. *La variable u_t puede ser escrita*

$$u_t = (1 - b_t)(y_t - \mu_{t|t-1}), \tag{3.4}$$

donde

$$b_t = \frac{(y_t - \mu_{t|t-1})^2/\nu exp(2\lambda)}{1 + (y_t - \mu_{t|t-1})^2/\nu exp(2\lambda)}, \quad 0 \leq b_t \leq 1, \quad 0 < \nu < \infty, \tag{3.5}$$

es distribuida beta(1/2,ν/2). Las u_t son IID(0,σ_u^2) y simetricamente distribuida. Incluso los momentos de todos los ordenes existen y son dados por

$$E(u_t^m) = \nu^{m/2} exp(m\lambda) \frac{B((1+m)/2, (\nu+m)/2)}{B(1/2, \nu/2)}, \quad m = 2, 4, \ldots \tag{3.6}$$

Observación 3.2. *La teoría tambien se puede elaborar observando que $u_t = \bar{b}_t(y_t - \mu_{t|t-1})$, donde $\bar{b}_t = (1 - b_t)$ se distribuye beta(ν/2,1/2).*

Corolario 3.3. *La varianza de u_t es*

$$Var(u_t) = \sigma_u^2 = \nu exp(2\lambda) E(b_t(1 - b_t)) = \frac{\nu^2 exp(2\lambda)}{(\nu + 3)(\nu + 1)} \tag{3.7}$$

y

$$E(u_t^4) = \nu^2 exp(4\lambda) E(b_t^2(1 - b_t^2)) = \frac{3\nu^2(\nu + 2) exp(4\lambda)}{(\nu + 1)(\nu + 3)(\nu + 5)(\nu + 7)}.$$

Por lo que la Kurtosis de u_t es

$$Kurtosis(u_t) = \frac{3(\nu + 2)(\nu + 3)(\nu + 1)}{\nu(\nu + 5)(\nu + 7)},$$

por lo que sus colas son menos pesadas que las de una distribución normal (a menos que ν <0.24). Por ejemplo, con $\upsilon = 6$, la kurtosis es 1.76.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Debido a que $u_t's$ son IID$(0, \sigma_u^2)$, $\mu_{t|t-1}$ es debil y estrictamente estacionaria mientras $|\phi| < 1$. Aunque determinar las propiedades estadisticas de $\mu_{t|t-1}$ requiere asumir que este comenzo en el paso infinito, pero en la practica necesita inicializarce, y esto se puede hacer estableciendo $\mu_{1|0} = \omega = \delta/(1-\phi)$ ó $\mu_{1|0}^{\dagger}$ en (3.3).

Observación 3.4. *La varianza del error de predicción en (3.2) es*

$$Var(v_t) = Var(y_t|Y_{t-1}) = \frac{\nu}{(\nu-2)exp(2\lambda)}. \tag{3.8}$$

mientras que

$$E(u_t v_t) = \nu exp(2\lambda)E(b_t) = \frac{\nu}{(\nu+1)} exp(2\lambda). \tag{3.9}$$

cuando $\nu = \infty$, $u_t = v_t$ y (1.3) es obtenida.

3.2.1. Generalización y forma reducida

El filtro en el modelo de primer orden de (3.2) puede generalizarse a:

$$\begin{aligned} \mu_{t+1|t} =& \delta + \phi_1 \mu_{t|t-1} + \cdots + \phi_p \mu_{t-p+1|t-p} \\ &+ \kappa_0 u_t + \kappa_1 u_{t-1} + \cdots + \kappa_r u_{t-r}. \end{aligned} \tag{3.10}$$

Dicho filtro se denota como $QARMA(p, r)$. El modelo completo se llamará DCS-t-$QARMA(p, r)$. Este modelo corresponde a uno de señal mas ruido no observada en la que la señal es $ARMA(p, r)$. En el caso Gaussiano, $u_t = v_t$. Si q se define como $max(q, r+1)$, podemos escribir

$$y_t = \delta + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + v_t - (\phi_1 - \kappa_0)v_{t-1} - \cdots - (\phi_q - \kappa_q)v_{t-q}, \tag{3.11}$$

que es un modelo $ARMA(p, q)$ con coeficientes MA $\theta_i = \phi_i - \kappa_{i-1}$, $i = 1, \ldots, q$. La condición de invertibilidad se aplican a $\theta_i = \phi_i - \kappa_{i-1}$, $i = 1, \ldots, q$ en lugar de κ_i, $i = 0, \ldots, q$. Para una distribución t_ν con $\nu < \infty$,

$$y_t = \delta + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \kappa_0 u_{t-1} + \cdots + \kappa_q u_{t-q} + v_t - \phi_1 v_{t-1} - \cdots - \phi_q v_{t-q}, \tag{3.12}$$

y las perturbaciones MA no son identicamente distribuidas, ya que cada una es una combinación diferente de variables, u_t y v_t, que tienen distribuciones diferentes (no normales). De hecho, no todos tienen las mismas varianzas. El proceso sigue siendo $ARMA(p, q)$, pero los coeficientes MA no son $\phi_i - \kappa_{i-1}$, $i = 1, \ldots, q$. (Pueden, en principio, encontrarse a partir de las autocorrelaciones).

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

3.2.2. Momentos de las observaciones

Cuando (3.10) es estacionario, la localización puede ser escrita como un MA infinito,

$$\mu_{t|t-1} = \omega + \sum_{j=1}^{\infty} \psi_j u_{t-j}, \quad \sum_{j=1}^{\infty} \psi_j^2 < \infty, \tag{3.13}$$

donde $\omega = \delta/(1 - \phi_1 - \cdots - \phi_p)$, entonces

$$y_t = \omega + \sum_{j=1}^{\infty} \psi_j u_{t-j} + v_t.$$

La existencia de momentos de y_t no se ve afectada por la dinámica. El siguiente resultado se sigue directamente de (3.7) y (3.8).

Proposición 3.5. *La varianza incondicional es*

$$E(y_t - \omega)^2 = E\left(\sum_{j=1}^{\infty} \psi_j u_{t-j} + v_t\right)^2 = \frac{exp(2\lambda)\nu}{\nu+1}\left[\frac{\nu+1}{\nu-2} + \frac{\nu}{3+\nu}\sum_{j=1}^{\infty}\psi_j^2\right], \tag{3.14}$$

siempre que $\nu > 2$.

De manera similar se puede derivar una expresión para el cuarto momento. La introducción de la dinámica en la localización significa que la curtosis es menor que la de la distribución t condicional porque el efecto de las colas pesadas en v_t se diluye con u_t.

3.2.3. Función de autocorrelación

Las autocovarianzas se pueden encontrar de la siguiente manera:

$$\begin{aligned} E((y_t - \omega)(y_{t-\tau} - \omega)) &= E(\mu_{t|t-1}(\mu_{t-\tau|t-\tau-1} + v_{t-\tau})), \quad \tau = 1, 2, 3, \ldots \\ &= E\left(\sum_{j=1}^{\infty} \psi_j u_{t-j}\right)\left(\sum_{j}^{\infty} \psi_j u_{t-\tau-j} + v_{t-\tau}\right) \\ &= \psi_\tau E(u_{t|\tau} v_{t-\tau}) + \sum_{j=1} \psi_j \psi_{j+\tau} E(u_{t-\tau}^2). \end{aligned}$$

Por lo tanto, utilizando (3.7) y (3.9),

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

$$E((y_t-\omega)(y_{t-\tau}-\omega)) = \frac{exp(2\lambda)\nu}{\nu+1}\left[\psi_\tau + \frac{\nu}{3+\nu}\sum_j^\infty \psi_j\psi_{j+\tau}\right], \quad \tau = 1,2,\ldots.$$

El modelo tiene covarianza estacionaria, siempre que $\nu > 2$, y la autocorrelación en el retardo τ venga dada por la relación entre el término entre corchetes en la ecuación anterior y el término entre corchetes en (3.14). En el modelo de primer orden, $\psi_j\psi_{j+\tau} = \kappa^2\phi^{2j+\tau-2}$ y por lo tanto $\sum_{j=1}^\infty \psi_j\psi_{j+\tau} = \kappa^2\phi^\tau/(1-\phi^2)$. En consecuencia

$$\begin{aligned}\rho_\nu(\tau) &= \frac{\psi_\tau + \frac{\nu}{3+\nu}\sum_{j=1}^\infty \psi_j\psi_{j+\tau}}{\frac{\nu+1}{\nu-2} + \frac{\nu}{3+\nu}\sum_{j=1}^\infty \psi_j^2} \\ &= \frac{\kappa\phi^{\tau-1} + \frac{\nu}{3+\nu}\frac{\kappa^2\phi^\tau}{1-\phi^2}}{\frac{\nu+1}{\nu-2} + \frac{\nu}{3+\nu}\frac{\kappa^2}{1-\phi^2}}, \quad \nu > 2, \quad \tau = 1,2,\ldots.\end{aligned}$$

La forma de la ACF es la de un modelo $ARMA(1,1)$ porque

$$\rho_\nu(1) = \left[\kappa + \frac{\nu}{3+\nu}\frac{\kappa^2\phi}{1-\phi^2}\right] / \left[\frac{\nu+1}{\nu-2} + \frac{\nu}{3+\nu}\frac{\kappa^2}{1-\phi^2}\right] \tag{3.15}$$

depende de κ y ϕ, pero apartir de entonces, $\rho_\nu(\tau) = \phi\rho_\nu(\tau-1), \quad \tau = 1,2,3,\ldots.$

3.3. Estimación por máxima verosimilitud

La función de log-verosimilitud para un modelo DCS-t es

$$\begin{aligned}\ln L(\boldsymbol{\psi},\nu) =& T\ln\Gamma((\nu+1)/2) - \frac{T}{2}\ln\pi - T\ln\Gamma(\nu/2) \\ &- \frac{T}{2}\ln\nu - T\ln\varphi - \frac{(\nu+1)}{2}\sum_{t=1}^T \ln\left(1 + \frac{(y_t-\mu_{t|t-1})^2}{\nu\varphi^2}\right).\end{aligned} \tag{3.16}$$

La maximización de la función de log-verosimilitud con respecto al parametro dinámico desconocido en el vector $\boldsymbol{\psi}$, la escala λ y el parametro de forma ν, puede ser realizado por medio de optimización numérica.

3.3.1. Distribución asintótica del estimador de máxima verosimilitud

Para aplicar la teoría asintótica de la sección 2.3, $\theta_{t|t-1}$ es reeemplazado por $\mu_{t|t-1}$ en el Teorema 2.24, y $\boldsymbol{\theta}_2$ se establece en $(\lambda,\nu)'$ en el Corolario 2.43.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Proposición 3.6. *Sea $y_t|Y_{t-1}$ observaciones con un distribución t_ν y localización, $\mu_{t|t-1}$, generada por el modelo de primer orden (3.2) o, equivalentemente , (3.3), con $|\phi| < 1$. Siempre que $\kappa \neq 0$ y $b < 1$, $(\tilde{\boldsymbol{\psi}}', \tilde{\lambda}, \tilde{\nu})'$, el estimador MV de $(\boldsymbol{\psi}', \lambda, \nu)'$ es consistente, y la distribución limite de $\sqrt{T}(\tilde{\boldsymbol{\psi}}' - \boldsymbol{\psi}', \tilde{\lambda} - \lambda, \tilde{\nu} - \nu)$ es normal multivariante, con media cero y matriz de covarianza*

$$Var\begin{pmatrix}\tilde{\boldsymbol{\psi}}\\ \tilde{\lambda}\\ \tilde{\nu}\end{pmatrix} = \begin{bmatrix}\frac{\nu+1}{\nu+3}exp(-2\lambda)\boldsymbol{D}(\boldsymbol{\psi}) & 0 & 0\\ 0 & \frac{2\nu}{\nu+3} & \frac{1}{(\nu+3)(\nu+1)}\\ 0 & \frac{1}{(\nu+3)(\nu+1)} & h(\nu/2)\end{bmatrix}^{-1}, \tag{3.17}$$

evaluado en $\boldsymbol{\psi} = \boldsymbol{\psi}_0$, $\lambda = \lambda_0$ y $\nu = \nu_0$. La matriz $\boldsymbol{D}(\boldsymbol{\psi})$ es definida como en (2.40) o (2.44), dependiendo de si $\boldsymbol{\psi} = (\phi, \kappa, \omega)'$ o $\boldsymbol{\psi} = (\phi, \kappa, \delta)'$, la varianza σ_u^2 es como en (3.7), $h(\nu)$ es definida en (2.17) y

$$\begin{aligned}
a &= \phi - \kappa\frac{\nu}{\nu+3},\\
b &= \phi^2 - 2\phi\kappa\frac{\nu}{\nu+3} + \kappa^2\frac{\nu(\nu^3+10\nu^2+35\nu+38)}{(\nu+1)(\nu+3)(\nu+5)(\nu+7)}\\
c &= 0.
\end{aligned}$$

Demostración. La matriz de información se sigue inmediatamente del Teorema 2.24 y el Corolario 2.43. Nótese que $\sigma_u^2/k^2 = [(\nu+1)/(\nu+3)]exp(-2\lambda)$ con $k = exp(2\lambda)\nu/(\nu+1)$ y $E(\partial\theta_1/\partial\boldsymbol{\psi}) = \mathbf{0}$ en (2.55) porque la distribución asintótica de μ en el modelo estático es independiente de los estimadores MV de λ y ν.

Con respecto a los valores de a, b y c en (2.35), la derivada de u_t es

$$\frac{\partial u_t}{\partial \mu} = 2(1-b_t)b_t - (1-b_t).$$

Por lo tanto $E(\partial u_t/\partial\mu) = -\nu/(\nu+3)$, que es consistente con σ_u^2 en (3.7). Ademas,

$$E\left(\frac{u_t\partial u_t}{\partial\mu}\right) = E(2(1-b_t)b_t - 1(1-b_t))(y_t - \mu_{t|t-1})(1-b_t) = 0$$

como $E((y_t - \mu_{t|t-1})|b_t) = 0$, y

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

$$
\begin{aligned}
E\left(\frac{\partial u_t}{\partial \mu}\right)^2 &= E(2(1-b_t)b_t - 1(1-b_t))^2 \\
&= 4E(1-b_t)^2 b_t^2 - 4E(1-b_t)^2 b_t + E(1-b_t)^2 \\
&= \frac{12\nu(nu+2)}{(\nu+1)(\nu+3)(\nu+5)(\nu+7)} - \frac{4\nu(\nu+2)}{(\nu+1)(\nu+3)(\nu+5)} + \frac{(\nu+2)\nu}{(\nu+3)(\nu+1)} \\
&= \frac{\nu(\nu^3+10\nu^2+35\nu+38)}{(\nu+1)(\nu+3)(\nu+5)(\nu+7)} \leq 1.
\end{aligned}
$$

La consistencia y la normalidad asintótica se siguen al hacer una corrección menor al Teorema 2.38. La primera derivada y segunda derivada de u_t son funciones lineales de $b_t^* = b_t^h(1-b_t)^k$, y entonces $0 \leq b_t^* \leq 1$ para cualquier $\boldsymbol{\psi}$ admisible. En cuanto a u_t en sí, como $u_t = (1-b_t)(y_t - \mu_{t1t-1})$, se puede ver que $u_t = 0$ cuando $y_t = 0$ y $u_t \to 0$ cuando $|y| \to \infty$. Claramente, u_t, como sus derivadas, está acotado, y se cumplen las condiciones para que se cumpla (2.50). □

Porque $c = 0$, hay alguna simplificación en $\mathbf{D}(\boldsymbol{\psi})$ como aparece en (3.17). En particular, para la paremetrización de ω en (2.40), $E = F = 0$, conduciendo al siguiente corolario.

Corolario 3.7. *Para el modelo establecido en 3.3, el estimador MV de ω se distribuye independientemente de los estimadores de κ y ϕ en muestras grandes, y su varianza asintótica es*

$$
Avar(\tilde{\omega}) = \frac{1}{T}\frac{\nu+3}{\nu+1}\frac{1-a}{1+a}\frac{1-b}{(1-\phi)^2}exp(2\lambda). \tag{3.18}
$$

Corolario 3.8. *Cuando $\mu_{t|t-1}$ depende de un conjunto de variables explicativas exógenas, la distribución asintótica del estimador MV se obtiene del Corolario 2.51.*

Observación 3.9. *La teoría asintótica es válida incluso para una distribución de Cauchy, aunque se debe tener cuidado para asegurar un máximo global; véase el Ejemplo 1 en Newey y McFadden (1994, págs. 2119, 2130, 2147).*

La figura 3.1 muestra un gráfico de b contra κ para $\phi = 0{,}9$ y $\nu = 6$. El rango adminisble es ligeramente mayor que en el caso gaussiano, donde, como se muestra en el siguiente corolario, es $-0{,}1 < \kappa < 1{,}9$.

Corolario 3.10. *Para un modelo gaussiano, $b < 1$ siempre que $\phi - 1 < \kappa < \phi + 1$.*

Demostración. Porque $u_t = v_t = y_t - \mu_{t|t-1}$, $\partial u_t / \partial \mu_{t|t-1} = -1$ y tambien $a = \phi - \kappa$, $b = \phi^2 - 2\phi\kappa + \kappa^2 = (\phi - \kappa)^2 = a^2$ y $c = 0$, todos los cuales son consistentes con las fórmulas de la Proposición 3.6. □

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

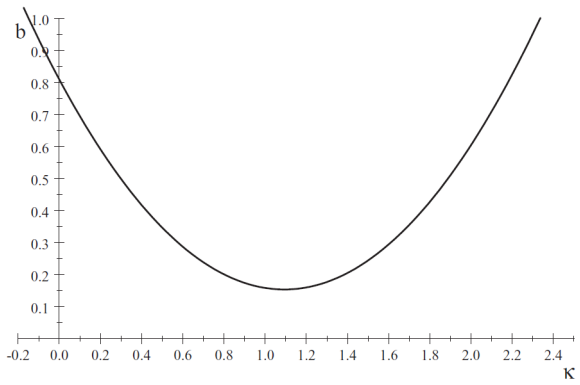


Figura 3.1: Gráfico de b contra κ para $\phi = 0{,}9$ y $\nu = 6$.

La forma reducida del modelo Gaussiano en el proceso $ARMA(1,1)$

$$y_t = \phi y_{t-1} + \upsilon_t - \theta \upsilon_{t-1}.$$

La condición para la invertibilidad estricta en el modelo $ARMA(1,1)$ es $|\theta| < 1$, y, porque $\theta = \phi - \kappa$, la invertibilidad asegura que $b < 1$. La condición $\theta \neq \phi$ es necesaria para la identificabilidad, y esta condición es equivalente a $\kappa \neq 0$.

Cuando ϕ es conocido

$$Var(\tilde{\kappa}) = 1 - b = 1 - (\phi - \kappa)^2, \tag{3.19}$$

que es consistente con el resultado estándar $MA(1)$, $Var(\tilde{\theta}) = 1 - \theta^2$. Cuando ϕ es estimado, la matriz de covarianzas asintótica corresponde con la del modelo $ARMA(1,1)$; ver Box et al. (2015) o Harvey (1993).

Con respecto al estimador MV de ω en un modelo $ARMA(1,1)$ gaussiano, se sigue de (3.18) que

$$Var(\tilde{\omega}) = \frac{1-\theta}{1+\theta}\frac{1-\theta^2}{(1-\phi)^2} exp(2\lambda) = \frac{(1-\theta)^2}{(1-\phi)^2} Var(\upsilon_t),$$

que es también la varianza de la distribución límite de la media muestral. La media de la muestra es un estimador eficiente de μ para observaciones gaussianas independientes, pero, como se mostró en la subsección 2.1.1, puede ser muy ineficiente para una distribución t con pequeños grados de libertad. Lo mismo ocurre cuando las observaciones están serialmente correlacionadas.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

3.3.2. Experimento Montecarlo

Se llevaron a cabo una serie de experimentos Montecarlo para obtener una indicador de qué tan buena es la aproximación asintótica para muestras pequeñas. Las cifras que se muestran en el cuadro (3.1) se basan en el cálculo de RMSE de 1.000 replicas para $T = 1.000$ observaciones de modelos de primer orden con $\nu = 6$ y un rango de valores (realistas) de κ y ϕ. La expresión de la matriz de información muestra que los errores estándar asintóticos (ASE) son independientes de ω y que λ solo aparece como un factor de escala. Por lo tanto fijar $\omega = \lambda = 0$ no debería implicar pérdida de generalidad (Harvey and Luati (2012)).

En la mayoría de los casos, la convergencia fue rápida, sin embargo, se encontró un problema de cálculo para λ. Los errores estándar asintóticos (ASE) se obtuvieron de las raíces cuadradas de los elementos diagonales de (3.17) dividido por el tamaño de la muestra. Los resultados muestran que los RMSE empíricos son un poco menor que los ASE pero no en gran medida. Los errores estándar numéricos se calcularon a partir de la matriz hessiana de la función de verosimilitud logarítmica. Las cifras en el Cuadro (3.1) se promedian sobre todas las replicas. En general, están razonablemente cerca de los RMSE empíricos a excepción del parámetro λ el cual presentó dificultades. La fila etiquetada como 'Estim. Par' es el promedio de la muestra y su función es dar una indicación de cualquier sesgo.

Cuadro 3.1: Resultados Montecarlo para el Modelo DCS-t de Primer Orden con $\nu = 6$ y $\omega = \lambda = 0$

Parametros			Estimaciones MV para $T = 1000$				
ϕ	κ		ϕ	κ	ω	λ	ν
0.8	0.1	Estim. Par	0.7918	0.10026	0.0014	-0.0042	6.083
		Num SE	0.1394	0.0414	0.0483	0.0379	1.1446
		RMSE	0.0755	0.0343	0.0459	0.0042	0.8415
		ASE	0.1175	0.0384	0.0478	0.0292	0.8437
0.8	0.5	Estim. Par	0.7956	0.4992	0.0022	-0.0046	6.1070
		Num SE	0.0359	0.0526	0.0933	0.0352	1.0367
		RMSE	0.0311	0.0346	0.0835	0.0046	0.7825
		ASE	0.0351	0.0433	0.0940	0.0292	0.8437
0.8	1.0	Estim. Par	0.7954	1.0009	0.0034	-0.0051	6.0620
		Num SE	0.0245	0.0660	0.1445	0.0311	0.8477
		RMSE	0.0245	0.0568	0.1070	0.0051	0.7856
		ASE	0.0240	0.0446	0.1469	0.0292	0.8437
0.95	0.1	Estim. Par	0.9374	0.0998	-0.0046	-0.0042	6.0300
		Num SE	0.0314	0.0277	0.0795	0.0376	1.1144
		RMSE	0.0329	0.0271	0.0770	0.0042	0.8426
		ASE	0.0238	0.0257	0.0835	0.0292	0.8437
0.95	0.5	Estim. Par	0.9465	0.4978	-0.0046	-0.0042	6.1470
		Num SE	0.0130	0.0472	0.2607	0.0345	1.0217
		RMSE	0.0137	0.0368	0.1247	0.0042	0.8424
		ASE	0.0120	0.0376	0.2691	0.0292	0.8437
0.95	1.0	Estim. Par	0.9476	1.0006	0.0031	-0.0047	6.0410
		Num SE	0.0110	0.0641	0.4525	0.0308	0.8229
		RMSE	0.0108	0.0535	0.1360	0.0047	0.7071
		ASE	0.0102	0.0430	0.4839	0.0292	0.8437

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

3.3.3. Aplicación

Los datos mensuales de la tasa de desempleo de Estados Unidos es utilizada desde febrero de 1948 a febrero de 2022 (fuente: Federal Reserve Economic Data). Las estadisticas descriptivas son presentadas en el Cuadro 3.2. Blazsek and Licht (2020) realizan una regresión lineal sobre las varaibles dummies $y_i = \delta_1 D_{jan,t} + \cdots + \delta_{12} D_{Dec,t} + \eta_t$ la cual sugiere que existe un posible componente de estacionalidad anual para s_t. La evolución de y_t y la estimación se presenta en la Figura (3.2).

La estimación de parámetros, criterios de selección y rendimiento del modelo son presentados en el Cuadro (3.2)

Cuadro 3.2: Estadística descriptiva, estimación de parámetros y diagnostico del modelo

Estadística descriptiva		Local Level-seasonal-t	
Fecha de inicio	Enero 1948	ω	0.7537
Fecha de término	Diciembre 2021	ϕ	1.0031
Tamaño muestra	889	κ	0.3805
Mínimo	2.50	λ	-0.3160
Máximo	14.80	ν	4.8918
Promedio	5.77	κ_1	0.2160
Mediana	5.60	κ_2	-0.0235
SD	1.70	κ_3	-0.0153
Skewness	0.8428	κ_4	-0.0167
Kurtosis	4.13	κ_5	-0.0223
		κ_6	-0.0366
		κ_7	-0.0199
		κ_8	0.0406
		κ_9	-0.0012
		κ_{10}	-0.0338
		κ_{11}	-0.0586
		κ_{12}	-0.0620
		LL	1059.8320
		AIC	42.0683
		AICC	43.9633
		BIC	176.0963
		RMSE	0.7320
		MAPE	0.0558

3.4. Restricción de parámetros

¿Qué restricciones sobre κ podrían estar implícitas si se considera que el modelo DCS de localización-t es una aproximación al modelo de componentes no observadas

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

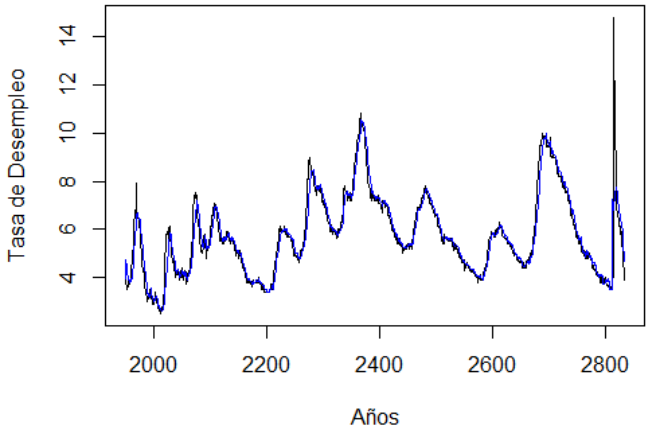


Figura 3.2: Estimación desempleo Estados Unidos.

de primer orden

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \quad t = 1, \ldots, T \\ \mu_{t+1} &= \delta + \phi\mu_t + \eta_t, \quad \eta \sim NID(0, \sigma_\eta^2), \end{aligned} \tag{3.20}$$

donde ε_t se distribuye como t_ν con media cero y escala ϕ?

En el modelo gaussiano, cuando $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$ en (3.20), la condición de invertibilidad en la forma reducida implica que el rango de κ es $\phi - 1 < \kappa < \phi + 1$. Sin embargo, en la forma reducida restringida, $\phi \leq \kappa \leq 0$ cuando $-1 \leq \phi \leq 0$ y $0 \leq \kappa \leq \phi$ para $0 \leq \phi < 1$. La expresión para κ en la forma de innovaciones de estado estacionario se obtiene igualando las autocorrelaciones en el retardo uno y es

$$\kappa = \phi - \left[\frac{q + 1 + \phi^2 - \sqrt{(q + 1 + \phi^2)^2 - 4\phi^2}}{2\phi} \right]. \tag{3.21}$$

Una vez fuera del modelo gaussiano, ya no existe una relación directa entre el modelo de componentes no observadas y el modelo DCS. Por lo tanto, las restricciones exactas correspondientes al modelo de componentes no observadas no se pueden aplicar al modelo DCS. Se puede obtener una aproximación igualando las autocorrelaciones en el retardo uno. El ACF para el modelo de componentes no observadas con ruido distribuido t, es decir, (3.20), es

$$\rho(\tau) = \frac{q\phi^\tau/(1 - \phi^2)}{\frac{\nu}{\nu-2} + \frac{q}{1-\phi^2}}, \quad \nu > 2, \quad \tau = 1, 2, \ldots$$

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Haciendo $\rho_\nu(1) = \rho(1)$, donde $\rho_\nu(1)$ se definió en (3.15), se obtiene

$$\frac{\kappa + \frac{\nu}{3+\nu}\frac{\kappa^2\phi}{1-\phi^2}}{\frac{\nu+1}{\nu-2} + \frac{\nu}{3+\nu}\frac{\kappa^2}{1-\phi^2}} = \frac{q\phi(1-\phi^2)}{\frac{\nu}{\nu-2}\frac{q}{1-\phi^2}}$$

que conduce a la ecuación cuadrática

$$\left(\frac{\nu}{\nu-2}\frac{\nu}{3+\nu}\frac{\phi}{1-\phi^2}\right)\kappa^2 + \left(\frac{\nu}{\nu-2} + \frac{q}{1-\phi^2}\right)\kappa - \frac{\nu+1}{\nu-2}\frac{q\phi}{1-\phi^2} = 0.$$

La solución para el modelo gaussiano se da en (3.21). Para grados finitos de libertad, es más complicado. Sin embargo, no es difícil ver que el rango de κ es más amplio que en el caso gaussiano, porque va de $\kappa = 0$ cuando $q = 0$ a

$$\kappa = \frac{\nu+1}{\nu-2}\phi \tag{3.22}$$

cuando $q = \infty$. Cuando $\nu = 6$, el límite superior anterior es $\kappa = 1.75\phi$. Por lo tanto, para $\nu = 6$ y $\phi = 0.9$, el rango implícito para κ está dentro del rango admisible, según lo determinado por la condición $b < 1$ requerida para la distribución asintótica del estimador MV; véase la figura 3.1. Nótese, sin embargo, que las autocorrelaciones no existen cuando $\nu = 2$, y como $\nu \to 2$ (desde arriba), el límite superior en (3.22) tiende a infinito.

3.5. Modelos de orden superior y modelos de espacio estado

El tratamiento estadístico general de los modelos de componentes no observados se basa en la forma de espacio de estado. El forma de innovaciones correspondiente facilita el manejo de modelos DCS de orden superior.

3.5.1. Modelos lineales gaussianos y el Filtro de Kalman

Para simplificar, suponga un modelo de serie de tiempo univariante invariante en el tiempo y excluya cualquier componente determinista. El caso general se expone en Harvey (1989). La observación en la forma de espacio de estado está relacionada con un vector de estado $m \times 1$, $\boldsymbol{\alpha}_t$, a través de una ecuación de medición,

$$y_t = \mathbf{Z}'\boldsymbol{\alpha}_t + \varepsilon_t, \quad t = 1, \ldots, T, \tag{3.23}$$

donde $\mathbf{z}$ es un vector $m \times 1$ y ε_t es una perturbación no correlacionada en serie con $E(\varepsilon_t) = \mathbf{0}$ y $Var(\varepsilon_t) = \sigma_\varepsilon^2$. Los elementos de $\boldsymbol{\alpha}_t$ por lo general no son observables,

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

pero se sabe que son generados por una ecuación de transición, una forma especial invariante en el tiempo de la cual se dio anteriormente en 2.61. Eliminando el vector $\boldsymbol{\delta}$ da

$$\boldsymbol{\alpha}_{t+1} = \mathbf{T}\boldsymbol{\alpha}_t + \boldsymbol{\eta}_t, \quad t = 1, \ldots, T. \tag{3.24}$$

La especificación se completa suponiendo que $E(\boldsymbol{\alpha}) = \boldsymbol{\alpha}_{1|0}$ y $Var(\boldsymbol{\alpha}_1) = \mathbf{P}_{1|0}$, donde $\mathbf{P}_{1|0}$ es una matriz semidefinida positiva, y que $E(\varepsilon_t \boldsymbol{\alpha}_0^{'}) = \mathbf{0}$ y $E(\boldsymbol{\eta}_t \boldsymbol{\alpha}_0^{'}) = \mathbf{0}$ para $t = 1, \ldots, t$. Por lo general, se supone que las perturbaciones no están correlacionadas entre sí en todos los períodos de tiempo, es decir, $E(\varepsilon_t \boldsymbol{\eta}_s^{'}) = \mathbf{0}$ para todo $s, t = 1, \ldots, T$, aunque esta suposición puede relajarse.

Cuando las perturbaciones y el estado inicial se distribuyen normalmente, las estimaciones del error cuadrático medio mínimo del estado y la observación en el tiempo t, basadas en la información en el tiempo $t-1$, son sus esperanzas condicionales. El Filtro de Kalman es un procedimiento recursivo para calcular estas estimaciones, dados $\mathbf{z}, \sigma_\varepsilon^2, \mathbf{T}$ y $\mathbf{Q}$ junto con las condiciones iniciales, $\boldsymbol{\alpha}_{1|0}$ y $\mathbf{P}_{1|0}$.

El Filtro de Kalman se puede escribir como un único conjunto de recurrencias que van directamente de $\boldsymbol{\alpha}_{t|t-1}$ a $\boldsymbol{\alpha}_{t+1|t}$, es decir,

$$\boldsymbol{\alpha}_{t+1|t} = \mathbf{T}\boldsymbol{\alpha}_{t|t-1} + \mathbf{k}_t v_t, \quad t = 1, \ldots, T \tag{3.25}$$

donde $v_t = y_t - \mathbf{z}_t^{'}\boldsymbol{t}|\boldsymbol{t-1}$ es la innovación y $f_t = \mathbf{z}^{'}\mathbf{P}_{t|t-1}\mathbf{z} + \sigma_\varepsilon^2$ es su varianza. El vector de ganancia, $\mathbf{k}_t$, es

$$\mathbf{k}_t = (1/f_t)\mathbf{T}\mathbf{P}_{t|t-1}\mathbf{z}, \quad t = 1, \ldots, T, \tag{3.26}$$

y la recursión de la matriz de covarianza es

$$\boldsymbol{P}_{t+1|t} = \mathbf{T}(\mathbf{P}_{t|t-1} - (1/f_t)\mathbf{P}_{t|t-1}\mathbf{z}\mathbf{z}^{'}\mathbf{P}_{t|t-1})\mathbf{T}^{'} + \mathbf{Q}. \tag{3.27}$$

Reorganizar las ecuaciones anteriores da la representación de la forma de innovaciones

$$\begin{aligned} y_t &= \mathbf{z}^{'}\boldsymbol{\alpha}_{t|t-1} + v_t, \quad t = 1, \ldots, T, \\ \boldsymbol{\alpha}_{t+1|t} &= \mathbf{T}\boldsymbol{\alpha}_{t|t-1} + \mathbf{k}_t v_t. \end{aligned} \tag{3.28}$$

Como en (1.3), la forma de las innovaciones refleja el estado original, la forma espacial, excepto que $\boldsymbol{\alpha}_{t|t-1}$ aparece en lugar del estado y las perturbaciones en las ecuaciones de medición y transición están perfectamente correlacionadas. Debido a que el modelo contiene solo un término de perturbación, puede considerarse como una forma reducida, con $\mathbf{k}_t$ sujeto a restricciones provenientes de la forma estructural original.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

En un modelo invariante en el tiempo, la matriz de covarianza $\mathbf{P}_{t+1|t}$ generalmente convergerá a un estado estacionario, $\mathbf{P}$, dado por la solución a la ecuación de Riccati (3.27); véase Harvey (2006). En la forma de innovaciones de estado estacionario, $\mathbf{k}_t$ y f_t son invariantes en el tiempo.

El filtro en (3.25) se puede escribir como

$$\boldsymbol{\alpha}_{t+1|t} = \mathbf{L}_t\boldsymbol{\alpha}_{t|t+1} + \mathbf{k}_t y_t, \quad t = 1, \ldots, T, \tag{3.29}$$

donde $\mathbf{L}_t = \mathbf{T} - \mathbf{k}_t\mathbf{z}_t^{'}$, lo que demuestra que $\boldsymbol{\alpha}_{t+1|t}$ se puede expresar como una combinación lineal de observaciones pasadas. En el estado estacionario de un modelo invariante en el tiempo, $\mathbf{L}_t = \mathbf{L}$. Cuando las raíces de $\mathbf{L}$ son menores que uno en valor absoluto, $\boldsymbol{\alpha}_{t+1|t}$ puede expresarse como un promedio móvil infinito de pasado $y_t^{'}$.

Hay varias formas de inicializar el Filtro de Kalman. Cuando la ecuación de transición es estacionaria, $\boldsymbol{\alpha}_{1|0}$ y $\mathbf{P}_{1|0}$ se pueden establecer en la media incondicional y la varianza de $\boldsymbol{\alpha}_t$. Cuando el estado contiene m^* elementos no estacionarios, se puede obtener un a priori adecuado a partir de las primeras m^* observaciones, ya sea directamente o indirectamente, usando un previo difuso, como se describe en Durbin and Koopman (2012).

Ejemplo 3.11. *Sea $p_{t|t-1}$ denote $\sigma_\varepsilon^2 MSE(\mu_{t|t-1})$ en el modelo $AR(1)$ más ruido de (1.1). La ecuación de Riccati es*

$$p_{t|t-1} = \phi^2 p_{t|t-1} - [\phi^2 p_{t|t-1}^2/(1 + p_{t|t-1})] + q, \quad t = 1, \ldots, T,$$

donde $q = \sigma_\eta^2/\sigma_\varepsilon^2 \geq 0$ y $p_{1|0} = q/(1 - \phi^2)$. Haciendo $p_{t+1|t} = p_{t|t-1} = p$ nos entrega

$$p\left(q + \phi^2 - 1 + \sqrt{(1 - q - \phi^2)^2 + 4q}\right)/2, \quad q \geq 0. \tag{3.30}$$

La solución cuando $q = 0$ es $p = 0$. En este caso no hay señal. Por otro lado, cuando $p = \infty$, no hay ruido. La ganancia de Kalman en estado estacionario es $\kappa = \phi p/(1+p)$, que al reacomodar se ve que es igual al resultado de (3.21). La ecuación de actualización de estado estacionario,

$$\mu_{t+1|t} = \phi\mu_{t|t-1} + \kappa v_t$$

puede escribirse como en (3.29), es decir

$$\mu_{t+1|t} = (\phi - \kappa)\mu_{t|t-1} + \kappa y_t. \tag{3.31}$$

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

3.5.2. Modelo DCS

Un modelo DCS de localización general se puede configurar de la misma manera que la forma de innovaciones de un modelo de espacio de estado gaussiano. El modelo correspondiente al estado estacionario de (3.28), pero con constantes incluidas, es

$$\begin{aligned} y_t &= \omega + \mathbf{z}^{'}\boldsymbol{\alpha}_{t|t-1} + v_t, \quad t = 1, \ldots, T, \\ \boldsymbol{\alpha}_{t+1|t} &= \boldsymbol{\delta} + \mathbf{T}\boldsymbol{\alpha}_{t|t-1} + \boldsymbol{\kappa} u_t \end{aligned} \tag{3.32}$$

El vector $\mathbf{z}$ y la matriz $\mathbf{T}$ se pueden especificar de la misma manera que para los modelos de componentes no observados gaussianos. Dichos modelos se pueden encontrar en Harvey (1989) y Koopman et al. (2009), pero la tendencia canónica y las formulaciones estacionales se describen en la siguiente sección. Esto deja la cuestión de cómo especificar los parámetros en el vector $\boldsymbol{\kappa}$. Más específicamente, ¿qué restricciones deberían imponerse? Los problemas del modelo de primer orden se discutieron en la Sección 3.4, pero incluso en ese caso, la respuesta no es sencilla. En un modelo de componentes no observadas, la ganancia de Kalman dependerá de todos los componentes, pero el principal determinante en el comportamiento de un elemento en particular será el componente con el que se relaciona. Por lo tanto, puede ser razonable explorar los diferentes componentes de forma aislada. Este es el enfoque adoptado en la siguiente sección.

La ecuación de transición en (3.32) es estacionaria, siempre que las raíces de la matriz de transición $\mathbf{T}$ tengan módulo menor que uno. Cuando este es el caso, $\boldsymbol{\delta}$ es superfluo y la inicialización se logra al establecer $\boldsymbol{\alpha}_{1|0}$. Cuando $\boldsymbol{\alpha}_{t|t-1}$ contiene elementos no estacionarios, la mejor opción parece ser tratar sus valores iniciales como parámetros desconocidos. Se puede encontrar una discusión más detallada en la siguiente sección.

La teoría asintótica para la estimación MV desarrollada para (2.61) se puede aplicar aquí reemplazando $\theta_{t|t-1}$ por $\mu_{t|t-1}$. Sean $\mathbf{A}(\nu)$ y $\mathbf{B}(\nu)$, $\mathbf{A}$ y $\mathbf{B}$ evaluados para una distribución t_ν. Luego

$$\mathbf{A}(\nu) = \mathbf{T} - \{\nu/(\nu+3)\}\boldsymbol{\kappa}\mathbf{z}^{'}, \tag{3.33}$$

mientras que $\mathbf{B}(\nu)$ se obtiene sustituyendo las esperanzas dadas en la demostración de la Proposición 3.6 en la expresión (2.62). Para la distribución t (simétrica), $E[(\partial u_t/\partial\mu)u_t] = 0$, entonces $[E(u_t\mathbf{X}_t)] = 0$ y (2.68) se convierte en

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

$$
\begin{aligned}
E\text{vec}\left(\frac{\partial \boldsymbol{\alpha}_{t+1|t}}{\partial \kappa_i}\frac{\partial \boldsymbol{\alpha}_{t+1|t}^{'}}{\partial \kappa_j}\right) &= [\mathbf{I}-\mathbf{B}(\nu)]^{-1}[E(u_t^2)]\text{vec}(\boldsymbol{\omega}_i\boldsymbol{\omega}_j^{'}) \\
&= \frac{exp(2\lambda)\nu^2}{(\nu+3)(\nu+1)}[\mathbf{I}-\mathbf{B}(\nu)]^{-1}\text{vec}(\boldsymbol{\omega}_i\boldsymbol{\omega}_j^{'}).
\end{aligned} \tag{3.34}
$$

La matriz de información para los parámetros desconocidos se obtiene como en (2.69).

La teoría asintótica requiere que las raíces de $\mathbf{B}(\nu)$ tengan un módulo menor que uno. Para un modelo gaussiano, este será el caso si las raíces de $\mathbf{A}$ tienen un módulo menor que uno. Esta condición en $\mathbf{A}$ también se da en el artículo de De Livera et al. (2011), en el que se utiliza la forma de innovaciones del espacio de estados para modelar series de tiempo gaussianas.

Lema 3.12. *En un modelo gaussiano, $\boldsymbol{B}=\boldsymbol{A}\otimes\boldsymbol{A}$, por lo que los valores propios de $\boldsymbol{B}$ son menores que uno en valor absoluto si los valores propios de $\boldsymbol{A}$ son menores que uno en valor absoluto.*

Demostración. Para $\nu=\infty$, $E(\partial u_t/\partial\mu)=-1$, entonces $\mathbf{A}=\mathbf{T}-\kappa\mathbf{z}^{'}$. Además, como $E(\partial u_t/\partial\mu)^2=1$,

$$\mathbf{B}=\mathbf{T}\otimes\mathbf{B}-\boldsymbol{\kappa}\mathbf{z}^{'}\otimes\mathbf{T}-\mathbf{T}\otimes\boldsymbol{\kappa}\mathbf{z}^{'}+\boldsymbol{\kappa}\mathbf{z}^{'}\otimes\boldsymbol{\kappa}\mathbf{z}^{'}=\mathbf{A}\otimes\mathbf{A}.$$

Por lo tanto, los N^2 valores propios de $\mathbf{B}$ son $\xi_i\xi_j$, $i,j=1,\ldots,N$, donde ξ_i, $i=1,\ldots,N$ son los valores propios de $\mathbf{A}$; véase Hamilton (2020). □

3.6. Tendencia y estacionalidad

La tendencia estocástica y los componentes estacionales pueden introducirse en los modelos de componentes no observados para la localización. Estos modelos, denominados modelos de series temporales estructurales, se describen en Harvey (1989). Las propiedades de los estimadores MV en modelos gaussianos se exponen en (Harvey, 1989, capitulo 3 y 4).

A continuación se exponen los principales elementos en la formulación de modelos estructurales de series temporales y su tratamiento estadístico. Luego se explora la forma en que las formas innovadoras de estos modelos conducen a los modelos DCS-t correspondientes. Una característica complicada es el hecho de que los componentes de tendencia y estacionales no son estacionarios, y esto tiene implicaciones para la inicialización de los filtros DCS.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

3.6.1. Modelo local level

El paseo aleatorio gaussiano más ruido o modelo de nivel local es

$$\begin{aligned} y_t &= \mu_t + \varepsilon_e, \quad NID(0, \sigma_\varepsilon^2), \\ \mu_t &= \mu_{t-1} + \eta_t, \quad NID(0, \sigma_\eta^2), \end{aligned} \tag{3.35}$$

donde $E(\varepsilon_t \eta_s)$ para todo t y s. La relación señal/ruido es $q = \sigma_\eta^2/\sigma_\varepsilon^2$, y el parámetro θ en la representación de forma reducida $ARIMA(0,1,1)$, se encuentra en el rango $0 \leq \theta < 1$ cuando $\sigma_\varepsilon^2 > 0$. Debido a que $\theta = 1 - \kappa$, el rango de κ en la forma de innovaciones de estado estacionario es $0 < \kappa \leq 1$. En este caso, $\mu_{t+1|t}$ es un promedio móvil ponderado exponencialmente en el que los pesos en las observaciones actuales y pasadas no son negativas, porque, al hacer $\phi = 1$ en (3.31),

$$\mu_{t+1|t} = (1 - \kappa)\mu_{t|t-1} + \kappa y_t. \tag{3.36}$$

Para una muestra semi-infinita

$$\mu_{t+1|t} = \kappa \sum_{i=0}^{\infty} (1 - \kappa)^i y_{t-i} \tag{3.37}$$

y los pesos de las observaciones pasadas suman uno.

Para el filtro DCS-t, establecer $\phi = 1$ y $\delta = 0$ en (3.2) nos entrega

$$\begin{aligned} y_t &= \mu_{t|t-1} + v_t, \\ \mu_{t+1|t} &= \mu_{t|t-1} + \kappa u_t \end{aligned} \tag{3.38}$$

La inicialización del Filtro de Kalman para (3.35) se realiza mejor usando un previo difuso. Sin embargo, se puede demostrar que como $p_{1|0} \to \infty$, $\mu_{2|1} \to y_1$ y $p_{2|1} \to 1 + q$; véase (Harvey, 1989, pp. 107-8). Por lo tanto, usar un prior difuso es efectivamente lo mismo que inicializar el filtro con una estimación construida a partir de la primera observación, pero teniendo debidamente en cuenta el error de medición. Un previo difuso no es una opción para un modelo DCS, pero el filtro aún podría iniciarse con $\mu_{2|1} = y_1$. El problema de proceder de esta manera es que el filtro podría verse afectado negativamente si la primera observación es un valor atípico. Una forma alternativa de inicializar el filtro DCS es tratar $\mu_{1|0}$ como un parámetro desconocido que debe estimarse junto con κ y ν. De Livera et al. (2011) utilizan el mismo enfoque para inicializar los modelos de fuente única de error de Gauss.

Como $u_t = (1 - b_t)(y_t - \mu_{t|t-1})$, reorganizando la ecuación dinámica en (3.38) se obtiene

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

$$\mu_{t+1|t} = (1 - \kappa(1 - b_t))\mu_{t|t-1} + \kappa(1 - b_t)y_t, \quad t = 1, \ldots, T. \tag{3.39}$$

Una condición suficiente para que las ponderaciones de las observaciones actuales y pasadas no sean negativas es que $\kappa(1 - b_t) < 1$ y, debido a que $0 \leq b_t \leq 1$, esto está garantizado por $0 < \kappa \leq 1$. Sin embargo, la restricción de que $\kappa \leq 1$ es mucho más estricto de lo necesario o deseable. De hecho, el argumento basado en autocorrelaciones coincidentes sugiere un rango admisible de $0 \leq \kappa \leq (\nu + 1)/(\nu - 2)$. Como muestra el siguiente ejemplo, las estimaciones de κ mayores que uno no son inusuales y son completamente apropiadas cuando la señal es fuerte en relación con el ruido.

En cuanto a las propiedades asintóticas, el Corolario 2.40 se aplica cuando $\mu_{1|0}$ es fijo y conocido o cuando $\mu_{2|1} = y_1$ e y_1 es fijo. Por eso

$$\begin{aligned}\operatorname{var}(\tilde{\kappa}) &= (1 - b)((\nu + 3)/\nu)^2 \\ &= \left(2\kappa\frac{\nu}{\nu + 3} - \kappa^2\frac{\nu(\nu^2 + 10\nu^2 + 35\nu + 38)}{(\nu + 1)(\nu + 3)(\nu + 5)(\nu + 7)}\right)\left(\frac{\nu + 3}{\nu}\right)^2.\end{aligned}$$

El rango admisible de κ se puede medir fácilmente a partir de la gráfica de b en la figura 3.1. En contraste con el caso cuando $|\phi| < 1$, es necesario que $\kappa > 0$. Para grados de libertad finitos, el límite superior será mayor que el valor de dos que se aplica a un modelo gaussiano.

3.6.2. Modelo de tendencia lineal local

El modelo de tendencia lineal local generaliza el nivel local mediante la introducción de una pendiente estocástica, β_t, que a su vez sigue un camino aleatorio. Por lo tanto

$$\begin{aligned}\mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2), \\ \beta_t &= \beta_{t-1} + zeta_t, \quad \zeta \sim N(0, \sigma_\zeta^2),\end{aligned} \tag{3.40}$$

donde las perturbaciones irregulares, de nivel y de pendiente, ε_t, η_t y ζ_t, respectivamente, son mutuamente independientes. Si ambas varianzas σ_η^2 y σ_ζ^2 son cero, la tendencia es determinista. Cuando solo σ_ζ^2 es cero, la pendiente es fija y la tendencia se reduce a una caminata aleatoria con deriva.

Permitiendo que σ_ζ^2 sea positivo, pero estableciendo σ_η^2 en cero, se obtiene una tendencia de caminata aleatoria integrada, que cuando se estima tiende a ser relativamente suave. Puede considerarse como una aproximación al ajuste de un spline cúbico que es el estimador suavizado de la tendencia a partir de un modelo de tiempo continuo sin perturbaciones de nivel; ver Kohn et al. (1992). La matriz de covarianza de tiempo discreto implícita es

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

$$\text{Var}\begin{bmatrix}\eta_t \\ \zeta_t\end{bmatrix} = \begin{bmatrix}\frac{1}{3} & \frac{1}{2} \\ \frac{1}{2} & 1\end{bmatrix}\sigma_\zeta^2.$$

El suavizado exponencial doble, sugerido por el principio de los mínimos cuadrados descontados, se obtiene estableciendo $q_\zeta = \sigma_\zeta^2/\sigma_\varepsilon^2 = (q_\eta/2)$, donde $q_\eta = \sigma_\eta^2/\sigma_\varepsilon^2$; ver (Harvey, 1989, Capítulo 2).

El filtro DCS correspondiente a la forma de innovaciones de estado estacionario para (3.40) es

$$\begin{aligned} y_t &= \mu_{t|t-1} + v_t \\ \mu_{t+1|t} &= \mu_{t|t-1} + \beta_{t|t-1} + \kappa_1 u_t \\ \beta_{t+1|t} &= \beta_{t|t-1} + \kappa_2 u_t. \end{aligned} \tag{3.41}$$

Se puede usar la inicialización $\beta_{3|2} = y_2 - y_1$ y $\mu_{3|2} = y_2$, pero, como en el modelo de nivel local, la inicialización de esta manera es vulnerable a valores atípicos al principio. Estimar los valores iniciales fijos, $\mu_{1|0}$ y $\beta_{1|0}$, parece ser una mejor opción.

Cuando las perturbaciones en el modelo de componentes no observados gaussiano, (3.40), no están correlacionadas entre sí, las restricciones sobre los coeficientes de forma de innovaciones, que están estrechamente relacionados con las constantes de suavizado en el esquema de pronóstico de Holt-Winters, son $0 < \kappa_1, \kappa_2 < 1$; véase (Harvey, 1989, pp. 177). Al igual que con el modelo de nivel local, se puede permitir que los límites superiores de κ_1 y κ_2 aumenten más allá de uno a medida que disminuyen los grados de libertad. Los valores de κ_1 y κ_2, que son demasiado grandes, pueden poner en duda la interpretación de la señal más el ruido, pero las restricciones clave son las necesarias para la validez de la teoría distribucional asintótica. Una tendencia de caminata aleatoria integrada en el modelo Gaussiano, (3.40), implica la restricción

$$\kappa_2 = \kappa_1^2(2 - \kappa_1), \quad 0 < \kappa_1 < 1, \tag{3.42}$$

que se puede encontrar usando fórmulas en (Harvey, 1989, pp. 177). La restricción puede imponerse al modelo DCS-t tratando $\kappa_1 = \kappa$ como el parámetro desconocido, pero sin imponer la unidad como límite superior.

A veces puede haber un caso para imponer una restricción derivada del principio de mínimos cuadrados descontados. En este caso, $\kappa_1 = 1 - \omega^2$ y $\kappa_2 = (1 - \omega)^2$, donde el parámetro de descuento satisface la restricción $0 \leq \omega < 1$ en el modelo gaussiano.

3.6.3. Modelo estacionalidad estocastica

Un patrón estacional fijo se puede modelar como $\gamma_t = \sum_{j=1}^{s} \gamma_j z_{jt}$, donde s es el número de temporadas y la variable ficticia z_{jt} es uno en la temporada j y cero en

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

caso contrario. Para no confundir tendencia con estacionalidad, los coeficientes, γ_j, $j = 1, \ldots, s$, están obligados a sumar cero. Se puede permitir que el patrón estacional cambie con el tiempo dejando que los coeficientes evolucionen como paseos aleatorios. Si γ_{jt} denota el efecto de la estación j en el tiempo t y $\boldsymbol{\gamma}_t = (\gamma_{1t}, \ldots, \gamma_{st})'$, entonces

$$\boldsymbol{\gamma}_t = \boldsymbol{\gamma}_{t-1} + \boldsymbol{\omega}_t, \quad t = 1, \ldots, T, \tag{3.43}$$

donde $\boldsymbol{\omega}_t$ es un vector de perturbaciones de media cero distribuido normalmente. Aunque todos los componentes estacionales están en constante evolución, solo uno afecta las observaciones en un momento determinado, es decir, $\gamma_t = \gamma_{jt}$ cuando prevalece la estación j en el tiempo t. El requisito de que los componentes estacionales se muevan de tal manera que siempre sumen cero, es decir, $\sum_{j=1}^{s} \gamma_{jt} = 0$, se impone por la restricción de que las perturbaciones suman cero en cada momento. Esta restricción es implementada por la estructura de correlación en

$$\text{Var}(\boldsymbol{\omega}_t) = \sigma_\omega^2(\mathbf{I} - s^{-1}\mathbf{i}\mathbf{i}'), \quad \sigma_\omega^2 > 0, \tag{3.44}$$

donde $\boldsymbol{\omega}_t = (\omega_{1t}, \ldots, \omega_{st})'$ e $\mathbf{i}$ es un vector de unos, junto con las condiciones iniciales que requieren que las estaciones sumen cero en $t = 0$. Puede verse en (3.44) que $\text{Var}(\mathbf{i}'\boldsymbol{\omega}_t) = 0$.

En la forma de espacio de estado, la matriz de transición es solo la matriz de identidad, pero el vector $\mathbf{z}$ debe cambiar con el tiempo para adaptarse a la temporada actual. Además de reemplazar $\mathbf{z}$ con $\mathbf{z}_t$, la forma del Filtro de Kalman permanece sin cambios. La adaptación del formulario de innovaciones al marco basado en la observación DCS, (2.61), nos entrega

$$\begin{aligned} y_t &= \mathbf{z}_t'\boldsymbol{\alpha}_{t|t-1} + v_t \\ \boldsymbol{\alpha}_{t+1|t} &= \boldsymbol{\alpha}_{t|t-1} + \boldsymbol{\kappa}_t u_t, \end{aligned} \tag{3.45}$$

donde $\mathbf{z}_t$ selecciona la temporada actual, $\gamma_{t|t-1}$, es decir, $\gamma_{t|t-1} = \mathbf{z}_t'\boldsymbol{\alpha}_{t|t-1}$; compárese con (3.32). La única pregunta es cómo parametrizar $\boldsymbol{\kappa}_t$.

Los componentes estacionales en el modelo de componentes no observas están restringidos a sumar cero, y lo mismo ocurre con sus estimaciones filtradas. Por lo tanto, $\mathbf{i}\boldsymbol{\kappa}_t = 0$ en el Filtro de Kalman, y esta propiedad debería trasladarse al filtro DCS. Si κ_{jt}, $j = 1, \ldots, s$, denota el j -ésimo elemento de $\boldsymbol{\kappa}_t$ en (3.45), luego en la temporada j establecemos $\kappa_{jt} = \kappa_s$, donde κ_s es un parámetro desconocido no negativo, mientras que $\kappa_{it} = -\kappa_s/(s-1)$ para $i \neq j$. Por lo tanto, las cantidades en las que cambian los efectos estacionales suman cero.

Las recursiones estacionales se pueden combinar con las ecuaciones de filtrado de tendencia de (3.41) para dar una estructura similar en forma a la del Filtro de Kalman

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

para el modelo de tendencia estocástica más estacional más ruido, a veces conocido como el "modelo estructural básico". Por lo tanto

$$y_t = \mu_{t|t-1} + \gamma_{t|t-1} + v_t, \tag{3.46}$$

donde $\mu_{t|t-1}$ es como se define en (3.41). El filtro se puede inicializar mediante la regresión de las primeras $s+1$ observaciones en una constante, tendencia temporal y variables ficticias estacionales restringidas para que los coeficientes sumen cero. Alternativamente, las condiciones iniciales en el momento $t = 0$ se estiman tratándolas como parámetros; hay $s-1$ parámetros estacionales porque el estado estacional inicial restante es menos la suma de los demás.

3.7. Suavizado

El objetivo del filtrado es encontrar el valor esperado del vector de estado, $\boldsymbol{\alpha}_t$, condicionado a la información disponible en el tiempo t. El objetivo del suavizar es tener en cuenta la información disponible después del tiempo t. En un modelo gaussiano lineal, la estimación suavizada es $\boldsymbol{\alpha}_{t|T}$, la esperanza condicional de $\boldsymbol{\alpha}_t$ y, por lo tanto, su MMSE, en función de todas las observaciones de la muestra. Para un modelo DCS, el filtro de suavizado se define por un argumento de simetría en lugar de derivarse como una estimación óptima. Sin embargo, como veremos, también puede ser racionalizado por un argumento basado en el modo condicional de la distribución posterior del estado. De hecho, esta interpretación de extracción de señales probablemente proporcione una base más sólida para los modelos DCS que el caso expuesto en el Capítulo 1.

3.7.1. Pesos

Las ecuaciones de filtrado en un modelo DCS tienen la misma forma que el Filtro de Kalman en un modelo de componentes lineales no observados gaussiano correspondiente. El Filtro de Kalman define un conjunto implícito de pesos para las observaciones actuales y pasadas. De manera similar, las ecuaciones recursivas hacia atrás en un intervalo fijo más suave para un modelo de componentes no observadas gaussiano lineal definen implícitamente un conjunto de pesos para todas las observaciones. Suavizar en un modelo DCS equivale a utilizar un conjunto de pesos de suavizado que coincidan con los pesos del filtro. Las complicaciones surgen porque el filtro DCS está controlado por una función no lineal de las observaciones que, a su vez, depende de la salida del filtro.

En un modelo de componentes no observadas gaussiano lineal simple, hay fórmulas explícitas para los pesos de filtrado y suavizado; véase (Whittle, 1983, capítulos 6 y 7). Para el modelo $AR(1)$ más ruido de (1.1), los pesos son una función del parámetro, θ,

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

en la forma reducida (1.2). Para ser específicos, los pesos para el filtro en una muestra semi-infinita,

3.8. Pronostico

La distribución predictiva de un paso adelante viene dada directamente por el modelo. La preocupación aquí es con la predicción de varios pasos.

3.8.1. Modelos QARMA

Cuando un modelo DCS es estacionario y $\mu_{t|t-1}$ tiene una representación de media móvil en las variables IID, como en (3.13), la localización en el tiempo $T+\ell$ se puede escribir como

$$\mu_{T+\ell|T+\ell-1} = \omega + \sum_{j=1}^{\ell-1} \psi_j u_{T+\ell-j} + \sum_{k=0}^{\infty} \psi_{\ell+k} u_{T-k}, \quad \ell = 2, 3, \ldots.$$

La estimación del MMSE de $\mu_{T+\ell|T+\ell+1}$, basada en la información en el tiempo T, es su esperanza condicional

$$\mu_{T+\ell|T} = \omega + \sum_{k=0}^{\infty} \psi_{\ell+k} u_{T-k}, \quad \ell = 2, 3, \ldots.$$

El MSE es

$$MSE(\mu_{T+\ell|T}) = \sigma_u^2 \sum_{j=1}^{\ell-1} \psi_j^2, \quad \ell = 2, 3, \ldots.$$

Para la distribución t_ν, σ_u^2 viene dada por (3.7) para cualquier ν positivo.

Cuando $\mu_{t+1|t}$ tiene la forma $QARMA(p, r)$, (3.10), el predictor, $\mu_{T+\ell|T}$, puede calcularse recursivamente, como en un modelo ARMA. Por lo tanto

$$\begin{aligned}\mu_{T+\ell|T} =& \delta + \phi_1 \mu_{T+\ell-1|T} + \cdots + \phi_p \mu_{T+\ell-p|T} \\ &+ \kappa_0 u_{T+\ell-1|T} + \cdots + \kappa_r u_{T+\ell-r|T}, \quad \ell = 2, 3, \ldots,\end{aligned}$$

donde $\mu_{T+j|T}$ se conoce para $j \leq 1$, $\mu_{T+j|T}$ se conoce para $j \leq 0$ y $u_{T+j|T} = 0$ para $j > 0$. Se puede usar una recursión de esta forma incluso si $\mu_{t+1|t}$ es no estacionario.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Ejemplo 3.13. *Para el modelo de primer orden, (3.2),*

$$\mu_{T+\ell|T} = \delta + \phi\mu_{T+\ell-1|T}, \quad \ell = 2, 3, \ldots,$$

y la recursividad sigue siendo válida para el modelo de nivel local, (3.38). Cuando el modelo es estacionario, la recursión se puede resolver para dar $\mu_{T+\ell|T} = \omega(1-\phi^{\ell-1}) + \phi^{\ell-1}\mu_{T+1|T}$, $\ell = 1, 2, 3, \ldots$, *y cuando* $\phi = 1$, $\mu_{T+\ell|T} = \delta(\ell-1) + \mu_{T+\ell|T}$.

El predictor de la observación en el tiempo $T+\ell$, es decir, $y_{T+\ell} = \mu_{T+\ell|T+\ell-1} + v_{T+\ell}$, es

$$y_{T+\ell|T} = \mu_{T+\ell|T}, \quad \ell = 1, 2, 3, \ldots.$$

Siempre que $\nu > 1$, $y_{T+\ell|T}$ es la esperanza condicional de $y_{T+\ell}$. Cuando $\nu > 2$, $y_{T+\ell|T}$ es el error cuadrático medio mínimo $\ell-$pasos adelante de $y_{T+\ell}$ con

$$MSE(y_{T+\ell|T}) = MSE(\mu_{T+\ell|T}) + \text{Var}(v_{T+\ell}), \quad \ell = 1, 2, \ldots, \tag{3.47}$$

donde MSE $(y_{T+1|T}) = 0$.

No se puede encontrar una fórmula para la distribución predictiva de varios pasos a menos que el modelo sea gaussiano. Sin embargo, la simulación es una opción viable. El error de predicción asociado con $y_{T+\ell|t}$ es

$$\sum_{j=1}^{\ell-1} \psi_j u_{T+\ell-j} + v_{T+\ell}, \quad \ell = 2, 3, \ldots,$$

y u_{T+j}, $j = 1, \ldots, \ell-1$ y $v_{T+\ell}$ pueden generarse a partir de distribuciones t independientes.

3.8.2. Forma de espacio estado

Las predicciones de varios pasos se pueden calcular utilizando la forma de espacio de estado de (2.61). El estimador óptimo de

$$\boldsymbol{\alpha}_{T+\ell|T+\ell-1} = \sum_{j=0}^{\ell-2} \mathbf{T}^j\boldsymbol{\delta} + \mathbf{T}^{\ell-1}\boldsymbol{\alpha}_{T+1|T} + \sum_{j=0}^{\ell-2} \mathbf{T}^j\boldsymbol{\kappa}\mathbf{u}_{T+j+1}, \quad \ell = 2, 3, \ldots,$$

es

$$\boldsymbol{\alpha}_{T+\ell|T} = \left[\sum_{j=0}^{\ell-2} \mathbf{T}\right]\boldsymbol{\delta} + \mathbf{T}^{\ell-1}\boldsymbol{\alpha}_{T+1|T}, \quad \ell = 2, 3, \ldots,$$

y el predictor correspondiente de $y_{T+\ell}$ es $y_{T+\ell|T} = \mathbf{z}'_{T+\ell}\boldsymbol{\alpha}_{T+\ell|T}$ para $\ell = 1, 2, 3, \ldots,$

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

El MSE de $y_{T+\ell|T}$ es

$$MSE(y_{T+\ell|T}) = \sigma_u^2 \mathbf{z}_{T+\ell}' \left[\sum_{j=0}^{\ell-2} \mathbf{T}^j\right] \boldsymbol{\kappa}\boldsymbol{\kappa}' \left[\sum_{j=0}^{\ell-2} \mathbf{T}^j\right]' \mathbf{z}_{t+\ell} + \text{Var}(v_{T\ell}), \quad \ell = 2, 3, \ldots,$$

porque el error de predicción es $\sum_{j=0}^{\ell-2} \mathbf{T}^j \boldsymbol{\kappa}\mathbf{u}_{T+j+1} + v_{T+\ell}$ para $\ell \geq 2$ y así la distribución predictiva completa puede simularse como antes.

Las funciones de pronóstico para modelos que contienen tendencias y estacionales toman la misma forma que en los modelos de componentes no observados correspondientes.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Capítulo 4

Modelos Multivariantes

Los modelos de score condicionales dinámicos se extienden de forma bastante natural para manejar cambios de localización en series de tiempo multivariadas. Aunque un tratamiento completo requiere un uso extenso del álgebra matricial, los principales problemas prácticos se refieren a la naturaleza y el alcance de las restricciones sobre las matrices de parámetros que gobiernan la dinámica. La teoría asintótica de la estimación de MV es sencilla en principio, aunque los detalles no siempre son fáciles de resolver.

El principal desafío radica en construir y analizar modelos que incorporen relaciones dinámicas entre variables. En este caso, la teoría asintótica ya no es sencilla, ni siquiera para los modelos gaussianos. Sin embargo, incluso cuando no es posible desarrollar una teoría asintótica completa, la estructura de la matriz de información y las condiciones bajo las cuales es definida positiva, proporciona una buena indicación de qué tan bien podrían funcionar en la práctica diferentes modelos y parametrizaciones.

Los modelos multivariados juegan un papel central en la macroeconometría. Los modelos que son capaces de manejar valores extremos se han vuelto más recurrentes dadas las turbulencias económicas de los últimos años de muchos indicadores macroeconómicos haciendo que estos temas estén lejos de ser solo académicos. Además, existe la posibilidad de que las relaciones económicas puedan cambiar con el tiempo. Los modelos en los que las correlaciones son dinámicas ofrecen una nueva línea de ataque al tratamiento de los parámetros variables en el tiempo en la regresión. La primera sección de este capítulo revisa la teoría de la distribución multivariada y esboza los problemas involucrados en la construcción de modelos dinámicos de localización, finalmente los modelos de localización dinámica se analizan en la Sección 4.2 y se muestra cómo algunas de las ideas principales de los modelos de series de tiempo estructurales se trasladan a los modelos DCS multivariados.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

4.1. Distribución Multivariante

Una distribución normal multivariada para un vector y_t de tamaño $N \times 1$ se parametriza en términos de un vector $N \times 1$ de medias $\boldsymbol{\mu}$, y una Matriz de covarianza $N \times N$, $\boldsymbol{\Sigma}$. la función de densidad de probabilidad es

$$f(\mathbf{y}_t; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-N/2}|\boldsymbol{\Sigma}|^{-1/2} exp(0{,}5(\mathbf{y}_t - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\mathbf{y}_t - \boldsymbol{\mu})). \tag{4.1}$$

Las distribuciones marginales de y_i son $N(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, N$. La distribución de un subconjunto de variables condicionadas por las otras también son normal multivariante.

La distribución t multivariante más común tiene función de densidad de probabilidad

$$f(\mathbf{y}_t; \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\nu}) = \frac{\Gamma((\nu + N)/2)}{(\pi\nu)^{N/2}\Gamma\nu/2|\boldsymbol{\Omega}|^{1/2} w_t^{(\nu+N)/2}},$$

donde $w_t = 1 + (1\nu)(\mathbf{y}_t - \boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \boldsymbol{\mu})$ y $\boldsymbol{\Omega}$ es la matriz de escala; véase Lange et al. (1989) y Kotz and Nadarajah (2004). La matriz de escala puede descomponerse como $\boldsymbol{\Omega} = \mathbf{DRD}$, donde $\mathbf{R}$ es una matriz de correlación y $\mathbf{D}$ es una matriz diagonal con el i-ésimo elemento diagonal igual a un parámetro de escala, ϕ_i. La matriz de covarianza, cuando existe, es $\boldsymbol{\Sigma} = \nu(\nu - 2)^{-1}\boldsymbol{\Omega}$, $\nu > 2$. La distribución puede derivarse de una normal multivariante y una chi-cuadrada independiente con ν grados de libertad, generalizando así una propiedad fundamental de la distribución t Student univariada. Específicamente, si $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y $z \sim \chi^2_\nu$, entonces $\mathbf{x}\sqrt{\nu/z} \sim t(\boldsymbol{\mu}, \boldsymbol{\Omega})$.

4.1.1. Estimación

La estimación de los parámetros en un modelo gaussiano multivariado es sencilla, con los estimadores MV dados por el primer y segundo momento. En un modelo bivariado, el estimador de la correlación, ρ, es el coeficiente de correlación muestral, r.

La estimación MV del modelo t multivariado requiere optimización numérica. La matriz de información del vector de localización $\boldsymbol{\mu}$ es

$$\mathbf{I}_{\mu\mu} = E\left(\frac{\partial \ln f}{\partial \boldsymbol{\mu}} \frac{\partial \ln f}{\partial \boldsymbol{\mu}'}\right) = \frac{\nu + N}{\nu + N + 2}\boldsymbol{\Omega}^{-1} \tag{4.2}$$

La derivación se puede encontrar en Lange et al. (1989, apéndice); véase también Kotz and Nadarajah (2004).

El estimador MV de $\boldsymbol{\mu}$ se distribuye independientemente de los estimadores MV de elementos de la matriz de escala y del parámetro grados de libertad; compare (2.18) para el modelo univariado. Lange et al. (1989) dan expresiones analíticas para el bloque de la matriz de información para ν y los distintos elementos de $\boldsymbol{\Omega}$.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

4.1.2. Regresión

Suponga que y_{1t} y y_{2t} son conjuntamente normales con medias cero y matriz de covarianza

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

La distribución de y_{1t} condicional a y_{2t} también es normal, y la ecuación de regresión se puede escribir como

$$y_{1t}|y_{2t} = \beta y_{2t} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \quad t = 1, \ldots, T, \tag{4.3}$$

donde $\beta = \sigma_1^2/\sigma_2^2 = \rho\sigma_1/\sigma_2$ y ε_t se distribuye independientemente de y_{2t} con varianza $\sigma_\varepsilon^2 = \sigma_1^2 - \rho^2\sigma_2^2$. Cuando β es el parámetro de interés, se puede estimar eficientemente haciendo una regresión de y_{1t} sobre y_{2t}. En la terminología de Engle et al. (1983), se dice que y_{2t} es débilmente exógena.

Cuando y_{1t} y y_{2t} tienen una distribución t bivariada con escala común φ, la relación de regresión es nuevamente lineal. Siempre que $\nu > 2$, el modelo se puede escribir como en (4.3), pero con $\varepsilon_t = \varphi\sqrt{1-\rho^2}t_\nu$, donde t_ν es una variable estándar de t; véase (Johnson and Kotz, 1972, pp. 135). Sin embargo, la simple regresión de y_{1t} sobre y_{2t} ya no es eficiente en comparación con la maximización de la función de verosimilitud obtenida a partir de la densidad conjunta de las dos variables. De hecho, desde el punto de vista de la estimación, pensar en el problema en términos de una sola ecuación no es útil porque las observaciones periféricas e influyentes son una característica de la distribución t bivariada, y solo al considerar la distribución conjunta se las trata adecuadamente.

4.2. Modelo de localización multivariante

Antes de analizar los modelos DCS, se analizan brevemente la naturaleza y el alcance de los modelos de componentes no observados gaussianos multivariados.

4.2.1. Modelos estructurales de series de tiempo

Los modelos de series temporales estructurales multivariantes se describen con cierto detalle en (Harvey, 1989, Capitulo 8). El modelo prototípico es una generalización directa de (1.1), es decir

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\omega} + \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim NID(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon), \quad t = 1, \ldots, T, \\ \boldsymbol{\mu}_{t+1} &= \boldsymbol{\Phi}\boldsymbol{\mu}_t + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim NID(\mathbf{0}, \boldsymbol{\Sigma}_\eta). \end{aligned} \tag{4.4}$$

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

El tratamiento estadístico de todos estos modelos se basa en el Filtro de Kalman.

El desglose en señal y ruido proporciona la base para una rica descripción de series temporales multivariadas. Establecer $\mathbf{\Phi} = \mathbf{I}$ da el modelo de nivel local multivariante y dejar que $\boldsymbol{\Sigma}_\eta$ tenga un rango menor que N da lugar a tendencias comunes y, por lo tanto, a la cointegración. Si el rango de la matriz de covarianza $\boldsymbol{\Sigma}_\eta$ es $J < N$, un orden adecuado de la serie permite escribir el modelo con J niveles comunes, o tendencias, $\boldsymbol{\mu}_t^\dagger$. Por lo tanto

$$\begin{aligned} \mathbf{y}_{1t} &= \boldsymbol{\mu}_t^\dagger + \boldsymbol{\varepsilon}_{1t} \\ \mathbf{y}_{2t} &= \mathbf{\Pi}\boldsymbol{\mu}_t^\dagger + \bar{\boldsymbol{\mu}} + \boldsymbol{\varepsilon}_{2t}, \end{aligned} \tag{4.5}$$

donde $\mathbf{\Pi}$ es una $(N-J)\times J$ matriz de coeficientes, $\bar{\boldsymbol{\mu}}$ es un $(N-J)\times 1$ vector de constantes, y

$$\boldsymbol{\mu}_t^\dagger = \boldsymbol{\mu}_t^\dagger + \boldsymbol{\eta}_t^\dagger, \quad \boldsymbol{\eta}_t^\dagger \sim NID(\mathbf{0}, \boldsymbol{\Sigma}_\eta^\dagger),$$

donde $\boldsymbol{\Sigma}_\eta^\dagger$ es definida positiva.

La presencia de tendencias comunes implica cointegración. En el modelo de nivel local anterior, existen vectores de cointegración $R = N-J$ (no únicos), de modo que la premultiplicación de (4.5) por estos vectores produce una serie temporal estacionaria de R. Por ejemplo, si la matriz de vectores cointegrantes es $A = (-\mathbf{\Pi}, \mathbf{I}_R)$, las ecuaciones para $\mathbf{y}_{2t}$ se convierten en

$$\mathbf{y}_{2t} = \mathbf{\Pi}\mathbf{y}_{1t} + \bar{\boldsymbol{\mu}} + \boldsymbol{\varepsilon}_t, \tag{4.6}$$

donde $\boldsymbol{\varepsilon}_t = \boldsymbol{\varepsilon}_{2t} - \mathbf{\Pi}\boldsymbol{\varepsilon}_{1t}$.

Cuando solo hay una tendencia común, $\mu_t^\dagger$, la matriz $\mathbf{\Pi}$ es un vector $(N-1)\times 1$, $\boldsymbol{\pi}$. Hay $N-1$ vectores cointegrantes, pero estos se pueden elegir de diferentes maneras. Restringir $\boldsymbol{\pi}$ para que sea un vector de unos da un crecimiento equilibrado.

Ejemplo 4.1. *El ingreso nacional, el consumo y la inversión exhiben un crecimiento equilibrado cuando existe una sola tendencia común para sus logaritmos. Si la tendencia común está asociada con el ingreso y las dos ecuaciones de cointegración corresponden a las "grandes proporciones"de consumo e inversión a ingreso, entonces*

$$\boldsymbol{A} = [-\boldsymbol{\pi} \quad \boldsymbol{I}_2] = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}.$$

El ruido gaussiano en (4.4) podría reemplazarse por ruido de cola pesada con distribución t multivariada, es decir, $\boldsymbol{\varepsilon}_t \sim t_\nu(\mathbf{0}, \boldsymbol{\Omega}_\varepsilon)$. Una vez que se abandona la suposición de la Gaussianidad, se necesitan técnicas intensivas en computación. Un modelo DCS ofrece una alternativa.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

4.2.2. Modelo DCS para una t multivariante

El modelo DCS-t de localización se puede formular como

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\omega} + \boldsymbol{\mu}_{t|t-1} + \mathbf{v}_t, \quad \mathbf{v}_t \sim t_\nu(\mathbf{0}, \boldsymbol{\Omega}), \quad t = 1, \ldots, T, \\ \boldsymbol{\mu}_{t|t-1} &= \boldsymbol{\Phi}\boldsymbol{\mu}_{t|t-1} + \mathbf{K}\mathbf{u}_t, \end{aligned} \tag{4.7}$$

donde $\mathbf{K}$ es una matriz $N \times N$ y el vector $\mathbf{u}_t$ depende de la score. La densidad logarítmica para la observación t-ésima es

$$\begin{aligned} \ln f_t(\boldsymbol{\omega}, \boldsymbol{\Phi}, \mathbf{K}, \boldsymbol{\Omega}, \nu) = &\ln(\Gamma(\nu + N)/2) - \ln\Gamma(\nu/2) - \frac{N}{2}\ln \pi\nu \\ &- \frac{1}{2}\ln|\boldsymbol{\Omega}| - \frac{\nu + N}{2}\ln w_t, \end{aligned} \tag{4.8}$$

donde

$$w_t = 1 + (1/\nu)(\mathbf{y}_t - \boldsymbol{\mu}_{t|t-1})'\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \boldsymbol{\mu}_{t|t-1}), \quad t = 1, \ldots, T.$$

El vector de score con respecto a $\boldsymbol{\mu}_{t|t-1}$ es

$$\frac{\partial \ln f_t}{\partial \boldsymbol{\mu}_{t|t-1}} = \frac{1}{w_t}\frac{\nu + N}{\nu}\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \boldsymbol{\mu}_{t|t-1}). \tag{4.9}$$

Si $\mathbf{u}_t$ se iguala al vector de score en la ecuación dinámica multivariante de (4.7), un valor atípico en cualquier serie afectará a los demás porque están conectados por $\boldsymbol{\Omega}$, pero el efecto se mitigará debido a la consiguiente reducción de peso de w_t. Este peso es el mismo para todas las series, incluso si no están correlacionadas entre sí. La intuición radica en el hecho de que la t multivariante se construye a partir de variables normales con un denominador común de chi-cuadrado.

El vector $\mathbf{u}_t$ puede modificarse premultiplicando por el inverso de la matriz de información. La eliminación de los términos constantes da una generalización de la variable (3.1), utilizada en el modelo de localización univariante, a saber,

$$\mathbf{u}_t = w_t^{-1}(\mathbf{y}_t - \boldsymbol{\mu}_{t|t-1}), \quad t = 1, \ldots, T.$$

Un valor atípico en una serie ahora no afectará directamente a los otros elementos de $\mathbf{u}_t$, aunque todavía existe la reducción de peso de w_t. El resultado final de los diferentes modelos depende de las restricciones impuestas a $\mathbf{K}$. Si $\mathbf{K}$ es diagonal, la especificación de $\mathbf{u}_t$ será crucial, pero este no será el caso si $\mathbf{K}$ está simplemente restringido a ser simétrico.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Se puede especificar un modelo DCS de nivel local multivariado configurando $\mathbf{\Phi} = \mathbf{I}$. La especificación se puede adaptar para permitir tendencias comunes (niveles). Así (4.5) se convierte en

$$\begin{aligned} \mathbf{y}_{1t} &= \boldsymbol{\mu}^{\dagger}_{t|t-1} + \boldsymbol{\nu}_{1t}, \quad t = 1, \ldots, T, \\ \mathbf{y}_{2t} &= \mathbf{\Pi}\boldsymbol{\mu}^{\dagger}_{t|t-1} + \bar{\boldsymbol{\omega}} + \boldsymbol{\nu}_{2t}, \\ \boldsymbol{\mu}^{\dagger}_{t|t-1} &= \mathbf{\Phi}\boldsymbol{\mu}^{\dagger}_{t|t-1} + \mathbf{K}^{\dagger}\mathbf{u}_t, \end{aligned}$$

donde $\boldsymbol{\mu}^{\dagger}_{t|t-1}$ ahora es un vector $J \times 1$ y $\mathbf{K}^{\dagger}$ es una matriz $J \times N$. Para un solo común, $\mathbf{K}^{\dagger'} = \boldsymbol{\kappa}$ es un vector $N \times 1$. En un modelo de crecimiento equilibrado, puede ser apropiado dejar que $\boldsymbol{\kappa}$ sea proporcional a un vector de unos por el inverso de la matriz de información.

Un enfoque alternativo a las tendencias comunes sería trabajar directamente con el vector $\boldsymbol{\mu}_{t|t-1}$ de tamaño $N \times 1$, como en (4.7) y luego restringir la matriz $\mathbf{K}$ de tamaño $N \times N$ para que sea simétrica, pero con rango J. Esto se puede hacer dejando $\mathbf{K} = \mathbf{K}^{*}\mathbf{K}^{*'}$, donde $\mathbf{K}^{*}$ es una matriz $N \times J$ en la que las primeras J filas son una matriz $J \times J$ triangular inferior. Un modelo de crecimiento equilibrado tendría $\mathbf{K}^{*'} = \boldsymbol{\kappa}$.

4.2.3. Teoría asintótica

Las propiedades del vector de score y sus derivadas son una generalización de las propiedades que se encuentran en el Capítulo 3.

Proposición 4.2. *La distribución de la score no depende de* $\boldsymbol{\mu}_{t|t-1}$ *porque* $\boldsymbol{y}_y - \boldsymbol{\mu}_{t|t-1}$ *y* w_t^{-1} *se distribuyen independientemente de* $\boldsymbol{\mu}_{t|t-1}$*. Es más,*

$$b_t = \frac{(1/\nu)(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})'\mathbf{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})}{1 + (1/\nu)(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})'\mathbf{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})} = 1 - \frac{1}{w_t}$$

se distribuye como beta$(N/2, \nu/2)$ *(y* w_t^{-1} *se distribuye como beta*$(\nu/2, N/2)$*). Las scores son IID (multivariante) con vector de media cero y matriz de covarianza dada por la matriz de información,* $\boldsymbol{I}_{\mu\mu}$ *de (4.2).*

Demostración. La matriz $\mathbf{\Omega}$ se puede descomponer como $\mathbf{\Omega} = \mathbf{\Omega}^{1/2}\mathbf{\Omega}^{1/2}$, donde $\mathbf{\Omega}^{1/2}$ es definida positiva. Así $\mathbf{\Omega}^{-1/2}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1}) \sim t_\nu(\mathbf{0}, \mathbf{I})$. Como $(\boldsymbol{y}_t\boldsymbol{\mu}_{t|t-1})' - \mathbf{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})$ es la suma de N variables t independientes, es la razón de una variable χ^2_N a una variable χ^2_ν independiente dividida por ν . El hecho de que b_t sea $beta(N/2, \nu/2)$ ahora se sigue del Lema 2.6, con detalles similares a los del Corolario 2.7. □

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

La matriz de segundas derivadas de la densidad logaritmica, la hessiana, es

$$\frac{\partial^2 \ln f_t}{\partial \boldsymbol{\mu}_{t|t-1} \partial \boldsymbol{\mu}_{t|t-1}^{'}} = -(1-b_t)\frac{\nu+N}{\nu}\boldsymbol{\Omega}^{-1} + 2\frac{\nu+N}{\nu^2}(1-b_t)^2\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \boldsymbol{\mu}_{t|t-1})(\mathbf{y}_t - \boldsymbol{\mu}_{t|t-1})'\boldsymbol{\Omega}^{-1}.$$

Al igual que el vector de scores, el hessiano se distribuye independientemente de $\boldsymbol{\mu}_{t|t-1}$.

El hecho de que el vector de score y su matriz de primera derivada, la hessiana, se distribuyan de forma independiente e idéntica y dependan de variables acotadas (que son funciones de variables distribuidas beta en los valores verdaderos de los parámetros) sugiere que los resultados de las Secciones 2.3, 2.4 y 3.3 puede extenderse para encontrar las condiciones bajo las cuales los estimadores MV de los parámetros desconocidos en $\boldsymbol{\omega}$,$\boldsymbol{\Phi}$,$\mathbf{K}$, $\boldsymbol{\Omega}$ y ν serán consistentes y asintóticamente normales. El primer paso es aplicar el Lema 2.42 observando que $\boldsymbol{\mu}_{t|t-1}$ depende de un conjunto de n_1 parámetros, que son los elementos distintos de las matrices $\boldsymbol{\Phi}$ y $\mathbf{K}$. En la notación del Lema 2.42, $\boldsymbol{\mu}_{t|t-1} = \boldsymbol{\theta}_{t|t-1}$ y los parámetros están contenidos en el $n \times 1$ vector $\boldsymbol{\psi}$. Por tanto, la score condicional es

$$\frac{\partial \ln f_t}{\partial \boldsymbol{\psi}} = \frac{\partial \ln f_t}{\partial \boldsymbol{\mu}_{t|t-1}^{'}} \frac{\partial \boldsymbol{\mu}_{t|t-1}}{\partial \boldsymbol{\psi}}$$

y, usando (4.2),

$$E_{t|t-1}\left(\frac{\partial \ln f_t}{\partial \boldsymbol{\psi}}\frac{\partial \ln f_t}{\partial \boldsymbol{\psi}^{'}}\right) = \frac{\partial \boldsymbol{\mu}_{t|t-1}^{'}}{\partial \boldsymbol{\psi}}\mathbf{I}_{\mu\mu}\frac{\partial \boldsymbol{\mu}_{t|t-1}}{\partial \boldsymbol{\psi}} = \frac{\nu+N}{\nu+N+2}\frac{\partial \boldsymbol{\mu}_{t|t-1}^{'}}{\partial \boldsymbol{\psi}}\boldsymbol{\Omega}^{-1}\frac{\partial \boldsymbol{\mu}_{t|t-1}}{\partial \boldsymbol{\psi}^{'}};$$

comparar (Lange et al., 1989, pp. 894) y la matriz superior izquierda en (2.54). Los elementos de la matriz de información están entonces dados por las esperanzas incondicionales de la expresión anterior; ver la discusión en la Sub-Sección 2.3.7.

4.2.4. Regresión y errores en variables

La subsección 4.1.2 explicó por qué la regresión era inadecuada para tratar con una distribución t bivariada. El caso del modelado multivariante vuelve a ser evidente cuando los componentes no observados entran en escena. La relación subyacente entre

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

y_{1t} y y_{2t} en el modelo gaussiano bivariado de (4.4) puede estar entre las señales correlacionadas en serie, μ_{1t} y μ_{2t}, mientras que las perturbaciones ε_{1t} y ε_{2t} representan términos de ruido que esencialmente no están correlacionados entre sí, además de ser independientes en serie. En el modelo de nivel local, μ_{1t} y μ_{2t} pueden considerarse componentes permanentes, mientras que ε_{1t} y ε_{2t} son transitorios.

La relación de regresión entre los componentes de la señal es

$$\mu_{1t}|\mu_{2t} = \beta_\mu \mu_{2t} + \eta_t, \quad t = 1, \ldots, T,$$

donde $\beta_\mu = \text{Cov}(\eta_{1t}, \eta_{2t})/\text{Var}(\eta_{1t})$. No se puede obtener un estimador eficiente de β_μ haciendo una regresión de y_{1t} sobre y_{2t} a menos que las matrices de covarianzas de $\varepsilon_t^{'}s$ y $\eta_t^{'}s$ sean proporcionales. Cuando este es el caso, y_{2t} es débilmente exógeno, pero más generalmente, el estimador de Mínimos Cuadrados Ordinarios de β_μ será inconsistente en un modelo estacionario porque, en una ecuación como (4.3), y_{2t} está correlacionada con ε_t; véase (Harvey, 1989, pp. 458). La consistencia se puede lograr cuando μ_{1t} y μ_{2t} son caminatas aleatorias, pero aún queda la cuestión de la eficiencia.

En el modelo clásico de errores en las variables, ambos componentes son serialmente independientes, y no es posible obtener estimadores consistentes de los parámetros de la relación subyacente a menos que exista información previa. Cuando las propiedades dinámicas de los componentes de interés son diferentes a las de los errores, el problema se resuelve. Un ejemplo lo da el modelo bivariado del Producto Interno Bruto e inflación reportado en Harvey (2011), donde la relación de la curva de Phillips se extrae del componente cíclico conjunto. En este caso, los componentes de tendencia y estacionales, así como los irregulares, pueden quedar fuera de escena porque sus dinámicas difieren de las del ciclo.

Las colas pesadas no complican, en principio, los errores en el tema de las variables. Un modelo DCS-t multivariante proporciona una solución, y la única dificultad es la práctica de estimar un modelo con varios componentes.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Capítulo 5

Modelos de localización multivariantes para variables macroeconómicas

En el presente capitulo, presentamos el modelo Seasonal-QVAR (cuasi vector autorregresivo) que incluye componentes de estacionalidad estocástica y nivel local multivariado no lineal impulsado por la score, y se aplica este modelo para estudiar la relación entre la tasa de inflación y la tasa de desempleo para los Estados Unidos. Especificamente nos centramos en los efectos que la pandemia sanitaria a causado en estas dos variables macroeconómicas, incorporando para ello un análisis de intervención. Una propiedad importante del modelo Seasonal-QVAR, a diferencia de los modelos de series temporales multivariantes con términos de error gaussianos, es que el filtro QVAR es resistente a valores extremos en el ruido.

5.1. Modelo Seasonal-QVAR

Para el modelo Seasonal-t-QVAR(1), el vector de variables dependientes $(y_{1,t}, \ldots, y_{N,t})'$ es:

$$\mathbf{y}_t = \boldsymbol{\omega} + \boldsymbol{\mu}_t + \mathbf{s}_t + \mathbf{v}_t \tag{5.1}$$

donde $\boldsymbol{\omega}$ $(N \times 1)$ incluye parámetros constantes, $\boldsymbol{\mu}_t$ $(N \times 1)$ es el componente dinámico de localización, $\mathbf{s}_t$ $(N \times 1)$ es el componente dinámico de estacionalidad y $\mathbf{v}_t$ $(N \times 1)$ es el término de error en forma reducida. Las componentes $\boldsymbol{\mu}_t$ y $\mathbf{s}_t$ son observables, condicionadas a la información pasada $(\mathbf{y}_1, \ldots, \mathbf{y}_{t-1})$.

El término de error de forma reducida $\mathbf{v}_t$ tiene una distribución t multivariada *i.i.d.* $\mathbf{v}_t \sim t_N(\mathbf{0}, \boldsymbol{\Omega}, \nu)$, donde $\boldsymbol{\Omega}$ es definida positiva y $\nu > 2$ es el parámetro de grados de libertad (es decir, la varianza de $\mathbf{v}_t$ está bien definida). Así $E[\mathbf{v}_t] = \mathbf{0}$ y $Var[\mathbf{v}_t] = \boldsymbol{\Omega} \times \nu/(\nu - 2)$.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

La forma reducida del término de error es $\boldsymbol{\varepsilon}_t = [\nu/(\nu-2)]^{-1/2}\,\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} \times \mathbf{v}_t$ con una distribución t multivariante $i.i.d$; $E[\boldsymbol{\varepsilon}_t] = \mathbf{0}$ y $Var[\boldsymbol{\varepsilon}_t] = \mathbf{I}_2$. El logaritmo de la densidad condicional de $\mathbf{y}_t$ es como en 4.8 y la score con respecto a $\boldsymbol{\mu}_t$ es como en 4.9, donde $\mathbf{u}_t$ $(N \times 1)$ es la función vector de la score escalada. En la proposición 4.2 se mostró que $\mathbf{u}_t$ es multivariante $i.i.d$ con media cero y varianza finita. En la definición de la función score, $\mathbf{v}_t$ está multiplicada por el escalar $\left[1 + \frac{\mathbf{v}_t'\boldsymbol{\Omega}^{-1}\mathbf{v}_t}{\nu}\right]^{-1} = \frac{\nu}{\nu+\mathbf{v}_t'\boldsymbol{\Omega}^{-1}\mathbf{v}_t} \in (0,1)$. Por lo tanto, la función score está ocotada por el término de error de forma reducida: $|\mathbf{u}_t| < |\mathbf{v}_t|$.

El componente de localización (4.7), donde $\boldsymbol{\Phi}$ y $\mathbf{K}$ (ambos $N \times N$) son parámetros matriciales constantes, y $\boldsymbol{\mu}_t$ se actualiza con el primer rezago de la función score escalada $\mathbf{u}_{t-1}$. El proceso se inicia utilizando la media incondicional $\boldsymbol{\mu}_1 = E[\boldsymbol{\mu}_1] = \mathbf{0}_{2\times 1}$. También se pueden utilizar alternativas a esta inicialización. Por ejemplo, los elementos de $\boldsymbol{\mu}_1$ pueden estimarse como parámetros adicionales en la estimación conjunta de todos los parámetros de Seasonal-QVAR.

La componente estacional $\mathbf{s}_t = (s_{1,t}, \ldots, s_{N,t})'$ con $\mathbf{s}_{k,t} = D_t'\boldsymbol{\gamma}_{k,t}$ para cada $k = 1, \ldots, N$, donde $\mathbf{D}_t = (D_{1,t}, ..., D_{s,t})'$ es un vector de variables ficticias estacionales (s denota el período conocido de la estacionalidad) y $\boldsymbol{\gamma}_{k,t} = (\gamma_{k,1,t}, ..., \gamma_{k,s,t})'$ es un vector de parámetros de estacionalidad dinámica. El parámetro dinámico $\boldsymbol{\gamma}_{k,t}$ es actualizado de acuardo con $\boldsymbol{\gamma}_{k,t} = \boldsymbol{\gamma}_{k,t-1} + \boldsymbol{\kappa}_{k,t}\mathbf{u}_{k,\boldsymbol{\mu},t-1}$, donde $\boldsymbol{\kappa}_{k,t} = (\kappa_{k,1,t}, \ldots, \kappa_{k,s,t})'$ es un parámetro de escala dinámico y $u_{k,t-1}$ es el k-ésimo elemento de $\mathbf{u}_{t-1}$. Cada elemento de $\boldsymbol{\kappa}_{k,t}$ es dado por $\boldsymbol{\kappa}_{k,j,t} = \kappa_{k,j}$ para $\mathbf{D}_{j,t} = 1$ y $\boldsymbol{\kappa}_{k,j,t} = -\kappa_{k,j}/(S-1)$ para $\mathbf{D}_{j,t} = 0$, donde $\boldsymbol{\kappa}_{k,j}$ con $j = 1, \ldots, s$, son parámetros de estacionalidad a estimar. Esta parametrización asegura que $\sum_{j=1}^{s} \boldsymbol{\kappa}_{k,j,t} = 0$. Como consecuencia $\mathbf{s}_{k,t}$ está centrado en cero para cada variable dependiente.

El parámetro dinámico $\boldsymbol{\gamma}_{k,t}$ se inicializa mediante un procedimiento de Mínimos Cuadrados no Lineales, en el que $\mathbf{y}_{k,t}$ es regresada en las variables ficticias estacionales, bajo la restricción de que la suma de todos los parámetros es cero. Esta parametrización asegura que $\sum_{j=1}^{s} \boldsymbol{\gamma}_{k,j,1} = 0$. Como consecuencia, $\mathbf{s}_{k,t}$ se centra en cero para cada variable dependiente. También se pueden considerar alternativas a esta inicialización. Sin embargo, resulta que el procedimiento de Mínimos Cuadrados Ordinarios utilizado para la inicialización es muy útil para la estimación efectiva del modelo Seasonal-QVAR.

En el modelo Seasonal-QVAR, el componente de localización $\boldsymbol{\mu}_t$ mide todos los efectos de interacción dinámica entre $\mathbf{y}_t$. Por lo tanto, tenemos la siguiente representación VMA(∞) no lineal de forma estructural de $\boldsymbol{\mu}_t$ Blazsek et al. (2018a) y Blazsek et al. (2018b):

$$\boldsymbol{\mu}_t = \sum_{j=0}^{\infty} \boldsymbol{\Phi}^j \boldsymbol{\Psi} [(\nu-2)\nu]^{\frac{1}{2}} \boldsymbol{\Omega}^{-1} \frac{\boldsymbol{\varepsilon}_{t-1-j}}{\nu - 2 + \boldsymbol{\varepsilon}_{t-1-j}'\boldsymbol{\varepsilon}_{t-1-j}} \tag{5.2}$$

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Sea C_1 el máximo módulo de los valores propios de $\mathbf{\Phi}$. La serie de la ecuación (5.2) es convergente si $C_1 < 1$; la función de respuesta al impulso $IRF_{j,t} = \frac{\partial \boldsymbol{\mu}_{t+j}}{\partial \boldsymbol{\varepsilon}_{t-1}}$ es Blazsek et al. (2018a) y Blazsek et al. (2018b):

$$IRF_{j,t} = \Phi^j \Psi[(\nu-2)\nu]^{\frac{1}{2}} \Omega^{-1} Q_{t-1-j} \quad \text{for} \quad j = 0, 1, \ldots, \infty, \tag{5.3}$$

donde

$$\mathbf{Q}_t = \frac{\partial \frac{\varepsilon_t}{\nu-2+\varepsilon_t' \varepsilon_t}}{\partial \varepsilon_t} \tag{5.4}$$

La matriz $\mathbf{Q}_t = \mathbf{I}_K$ para VARMA y VAR por lo que la IRF es constante para esos modelos. Para QVAR, usamos la media incondicional de $IRF_{j,t}$ (Blazsek et al. (2017)):

$$IRF_j = E[IRF_{j,t}] = \mathbf{\Phi}^j \mathbf{\Psi}[(\nu-2)\nu]^{\frac{1}{2}} \mathbf{\Omega}^{-1} E[\mathbf{Q}_{t-1-j}] \quad \text{para} \quad j = 0, 1, 2, \ldots, \infty. \tag{5.5}$$

donde $E[\mathbf{Q}_{t-1-j}]$ se estima utilizando el promedio muestral de $\mathbf{Q}_t$, que se apoya al realizar la prueba aumentada de Dicky Fuller (ADF) con constante para cada elemento de $\mathbf{Q}_t$. Una alternativa al uso de la invariante en el tiempo $E[IRF_{j,t}]$ es la estimación y análisis período por período de $IRF_{j,t}$, en la que $IRF_{j,t}$ se promedia para diferentes subperíodos de la muestra y se comparan las estimaciones de IRF resultantes. El intervalo de confianza de la IRF se estima mediante el uso de 10.000 simulaciones de Monte Carlo a partir de las estimaciones de Máxima Verosimilitud.

5.1.1. Inferencia estadística del modelo Seasonal-QVAR

Se estima el modelo Seasonal-QVAR utilizando el método de MV (ver, por ejemplo, Davidson et al. (2004)). Relacionado con esto, también se hace referencia a los trabajos de Creal et al. (2013) y Harvey (2013). El estimador MV de parámetros es

$$\hat{\mathbf{\Theta}}_{MV} = arg \quad \underset{\mathbf{\Theta}}{max} LL(y_1, \ldots, y_T; \mathbf{\Theta}) = arg \quad \underset{\mathbf{\Theta}}{max} \sum_{t=1}^{T} \ln f(y_t | y_1, \ldots, y_{t-1}; \mathbf{\Theta}), \tag{5.6}$$

donde $\mathbf{\Theta}$ denota el vector de parámetros constantes. Usamos la inversa de la matriz de información estimada numéricamente para los errores estándar de MV (Creal et al. (2013) ; Harvey (2013)).

Se utilizan los resultados de las secciones 2.3, 2.4 y 3.3) para las condiciones de consistencia y normalidad asintótica del estimador MV (ver también Blazsek et al. (2017)). En relación con las propiedades asintóticas de MV, también estudiamos la invertibilidad de Seasonal-QVAR (ver, por ejemplo, Blasques et al. (2018)).

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Primero, la condición 1 es $C_{\mu,1} < 1$, por lo tanto, $\boldsymbol{\mu}_t$ es estacionario en covarianza. En segundo lugar, la Condición 2 se cumple si $E[\mathbf{u}_{j,t}^{2-i}(\partial \mathbf{u}_{k,t}/\partial \boldsymbol{\mu}_{l,t})^i] < \infty$, donde $i = 0, 1, 2$ y $j, k, l = 1, \ldots, N$ (utilizando el promedio de la muestra para estimar esta esperanza que respaldamos con la prueba ADF). Tercero, para la Condición 3 consideramos el elemento representativo $\Psi_{i,j}$ de la matriz $\boldsymbol{\Psi}$. A partir de la ecuación dinámica de nivel local $\boldsymbol{\mu}_t = \boldsymbol{\Phi}\boldsymbol{\mu}_{t-1} + \boldsymbol{\Psi}\mathbf{u}_{t-1}$, expresamos

$$\frac{\partial \boldsymbol{\mu}_t}{\partial \boldsymbol{\Psi}_{i,j}} = \boldsymbol{\Phi}\frac{\partial \boldsymbol{\mu}_{t-1}}{\partial \boldsymbol{\Psi}_{i,j}} + \boldsymbol{\Psi}\frac{\partial \mathbf{u}_{t-1}}{\partial \boldsymbol{\Psi}_{i,j}} + \mathbf{W}_{i,j}\mathbf{u}_{t-1} \tag{5.7}$$

para todo $t = 1, \ldots, T$, donde el elemento (i, j) de la matriz $\mathbf{W}_{i,j}$ $(N \times N)$ es uno y el resto de los elementos de $\mathbf{W}_{i,j}$ son cero. Usando la regla de la cadena, también podemos escribir la Ecuación (5.7) como

$$\frac{\partial \boldsymbol{\mu}_t}{\partial \boldsymbol{\Psi}_{i,j}} = \left(\boldsymbol{\Phi} + \boldsymbol{\Psi}\frac{\partial \mathbf{u}_{t-1}}{\partial \boldsymbol{\mu}_{t-1}^{'}}\right)\frac{\partial \boldsymbol{\mu}_{t-1}}{\partial \boldsymbol{\Psi}_{i,j}} + \mathbf{W}_{i,j}\boldsymbol{u}_{t-1} = \mathbf{X}_t\frac{\partial \boldsymbol{\mu}_{t-1}}{\partial \boldsymbol{\Psi}_{i,j}} + \mathbf{W}_{i,j}\mathbf{u}_{t-1}. \tag{5.8}$$

La condición 3 es que todos los valores propios de $E(\mathbf{X}_t)$ están dentro del círculo unitario, donde la finitud de todos los elementos de $E(\mathbf{X}_t)$ se deriva de la condición 2. Denotamos el módulo máximo de los valores propios de $E(\mathbf{X}_t)$ usando $C_{\mu,3}$. Si cada elemento de $\mathbf{X}_t$ es estacionario en covarianza, entonces $E(\mathbf{X}_t)$ se puede estimar utilizando el promedio de la muestra (apoyamos el uso de este estimador mediante la prueba ADF). Cuarto, para la Condición 4, consideramos que la matriz de información depende de:

$$\frac{\partial \boldsymbol{\mu}_t}{\partial \Psi_{i,j}}\frac{\partial \boldsymbol{\mu}_t^{'}}{\partial \Psi_{k,l}} = \mathbf{X}_t\frac{\partial \boldsymbol{\mu}_{t-1}}{\partial \Psi_{i,j}}\frac{\partial \boldsymbol{\mu}_{t-1}^{'}}{\partial \Psi_{k,l}}\mathbf{X}_t^{'} + \mathbf{X}_t\frac{\partial \boldsymbol{\mu}_{t-1}}{\partial \Psi_{i,j}}\mathbf{W}_{i,j}^{'}\mathbf{u}_{t-1} + \mathbf{u}_{t-1}^{'}\mathbf{W}_{k,l}\frac{\partial \boldsymbol{\mu}_{t-1}^{'}}{\partial \Psi_{k,l}}\mathbf{X}_t^{'} + \mathbf{W}_{i,j}\mathbf{u}_{t-1}\mathbf{u}_{t-1}^{'}\mathbf{W}_{k,l}^{'} \tag{5.9}$$

que también podemos escribir como

$$\begin{aligned}\operatorname{vec}\left(\frac{\partial \boldsymbol{\mu}_t}{\partial \boldsymbol{\Psi}_{i,j}}\frac{\partial \boldsymbol{\mu}_t^{'}}{\partial \boldsymbol{\Psi}_{k,l}}\right) =& \mathbf{X}_t \otimes \mathbf{X}_t\operatorname{vec}\left(\mathbf{X}_t\frac{\partial \boldsymbol{\mu}_{t-1}}{\partial \boldsymbol{\Psi}_{i,j}}\mathbf{W}_{i,j}^{'}\mathbf{u}_{t-1}\right) + \operatorname{vec}\left(\mathbf{u}_{t-1}^{'}\mathbf{W}_{k,l}\frac{\partial \boldsymbol{\mu}_{t-1}^{'}}{\partial \boldsymbol{\Psi}_{k,l}}\mathbf{X}_t^{'}\right) \\ &+ \operatorname{vec}\left(\mathbf{u}_{t-1}^{'}\mathbf{W}_{k,l}\frac{\partial \boldsymbol{\mu}_{t-1}^{'}}{\partial \boldsymbol{\Psi}_{k,l}}\mathbf{X}_t^{'}\right) + \operatorname{vec}(\mathbf{W}_{i,j}\mathbf{u}_{t-1}\mathbf{u}_{t-1}^{'}\mathbf{W}_{k,l}^{'}),\end{aligned} \tag{5.10}$$

La condición 4 es que todos los valores propios de $E(\mathbf{X}_t \otimes \mathbf{X}_t)$ están dentro del círculo unitario, donde la finitud de todos los elementos de $E(\mathbf{X}_t \otimes \mathbf{X}_t)$ se deriva de la condición 2. Denotamos el módulo máximo de los valores propios de $E(\mathbf{X}_t \otimes \mathbf{X}_t)$

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

usando $C_{\mu,4}$. Si cada elemento de $\mathbf{X}_t \otimes \mathbf{X}_t$ es estacionario en covarianza, entonces $E(\mathbf{X}_t \otimes \mathbf{X}_t)$ se puede estimar utilizando el promedio de la muestra (apoyamos el uso de este estimador mediante la prueba ADF). Para el cálculo de $\mathbf{X}_t = \boldsymbol{\Phi} + \boldsymbol{\Psi}(\partial \mathbf{u}_{t-1}/\partial \boldsymbol{\mu}_{t-1}^{'})$ en las Condiciones 3 y 4, usamos la fórmula para $\partial \mathbf{u}_t/\partial \boldsymbol{\mu}_t^{'}$ $(N \times N)$. Como se mencionó anteriormente, la función de score está dada por

$$\mathbf{u}_t = \left(1 + \frac{\mathbf{v}_t^{'}\boldsymbol{\Omega}^{-1}\mathbf{v}_t}{\nu}\right)^{-1} \qquad \mathbf{v}_t = \frac{\nu(\mathbf{y}_t - \boldsymbol{\omega} - \boldsymbol{\mu}_t - \mathbf{s}_t)}{\nu + (\mathbf{y}_t - \boldsymbol{\omega} - \boldsymbol{\mu}_t - \mathbf{s}_t)^{'}\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \boldsymbol{\omega} - \boldsymbol{\mu}_t - \mathbf{s}_t)} \tag{5.11}$$

y la fórmula de $\partial \mathbf{u}_t/\partial \boldsymbol{\mu}_t^{'}$ se puede obtener utilizando el cálculo matricial estándar.

Además de las condiciones anteriores, también estudiamos la invertibilidad de Seasonal-QVAR que es una condición de la consistencia y normalidad asintótica de MV. La invertibilidad se estudia en la literatura reciente sobre modelos DCS (ver por ejemplo: Blasques et al. (2018)). A partir de las ecuaciones $\mathbf{y}_t = \boldsymbol{\omega} + \boldsymbol{\mu}_t + \mathbf{s}_t + \mathbf{v}_t$ y $\boldsymbol{\mu}_t = \boldsymbol{\Phi}_{t-1} + u_{t-1}$, expresamos:

$$\mathbf{y}_t = (\mathbf{I}_N - \boldsymbol{\Phi})\boldsymbol{\omega} + \boldsymbol{\Phi}\mathbf{y}_{t-1} + (\mathbf{I}_N - \boldsymbol{\Phi}\mathbf{L})\mathbf{s}_t - \boldsymbol{\Phi}\mathbf{v}_{t-1}\boldsymbol{\Psi}\mathbf{u}_{t-1} + \mathbf{v}_t. \tag{5.12}$$

Sustituimos la función de puntuación en la ecuación anterior y obtenemos:

$$\mathbf{y}_t = (\mathbf{I}_N - \boldsymbol{\Phi})\boldsymbol{\omega} + \boldsymbol{\Phi}\mathbf{y}_{t-1} + (\mathbf{I}_N - \boldsymbol{\Phi}\mathbf{L})\mathbf{s}_t + \frac{\boldsymbol{\Psi}\nu - \boldsymbol{\Phi}}{\nu + \mathbf{v}_{t-1}^{'}\boldsymbol{\Omega}^{-1}\mathbf{v}_{t-1}}\mathbf{v}_{t-1} + \mathbf{v}_t \tag{5.13}$$

Esta ecuación es un modelo VARMA(1,1) estacional, para el cual el parámetro MA(1) está basado en la observación, y se define como: $\boldsymbol{\Psi}_t = (\boldsymbol{\Psi}\nu - \boldsymbol{\Phi})/(\nu + \mathbf{v}_{t-1}^{'}\boldsymbol{\Omega}^{-1}\mathbf{v}_{t-1})$. VARMA(1,1) con parámetros constantes es invertible si el módulo máximo de valores propios del parámetro MA(1) es menor que uno. Estudiamos esta condición para $\boldsymbol{\Psi}_t$ para $t = 1, \ldots, T$.

Por último, para el componente de estacionalidad $\mathbf{s}_t$, las condiciones de MV se satisfacen si los parámetros de $\boldsymbol{\gamma}_{k,t-1}$ se establecen en la matriz de identidad (es decir, caminata aleatoria multivariada), como en la ecuación $\boldsymbol{\gamma}_{k,t} = \boldsymbol{\gamma}_{k,t-1} + \boldsymbol{\kappa}_{k,t}\mathbf{u}_{k,t-1}$. Si la raíz unitaria se impone en lugar de estimarse en los modelos DCS de tendencia o estacionalidad, se aplican las asintóticas estándar de MV.

5.2. Modelo Gaussiano Seasonal-VARMA

Seasonal-QVAR se puede relacionar con los modelos Gaussian Seasonal-VARMA y Gaussian Seasonal-VAR. Con respecto al modelo VAR, se hace referencia a los trabajos de Sims (1980), Sims et al. (1986), Sims et al. (1982), Bernanke (1986), Blanchard and

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Watson (1986) y Stock and Watson (2001). Con respecto al modelo VARMA, se hace referencia al artículo seminal de Tiao and Tsay (1989). También se está haciendo referencia a los libros de texto relacionados de (Hamilton, 2020, capítulos 10 a 12) y Lütkepohl (2005).

si $\nu \to \infty$, luego $\mathbf{v}_t \xrightarrow{d} N_N(0, \boldsymbol{\Sigma}_v)$ y $\mathbf{u}_t \xrightarrow{p} \mathbf{v}_t$ (ver Ecuación (5.11)). A partir de las ecuaciones $\mathbf{y}_t = \boldsymbol{\omega} + \boldsymbol{\mu}_t + \mathbf{s}_t + \mathbf{v}_t$ y $\boldsymbol{\mu}_t = \boldsymbol{\Phi}_{t-1} + \boldsymbol{\Psi}\mathbf{u}_{t-1}$, obtenemos el modelo Gaussiano Estacional-QVAR:

$$\mathbf{y}_t = (\mathbf{I}_N - \boldsymbol{\Phi})\boldsymbol{\omega} + \boldsymbol{\Phi}\mathbf{y}_{t-1} + (\mathbf{I}_N - \boldsymbol{\Phi}\mathbf{L})\mathbf{s}_t + (\boldsymbol{\Psi} - \boldsymbol{\Phi})\mathbf{v}_{t-1} + \mathbf{v}_t, \tag{5.14}$$

donde $\mathbf{L}$ denota el operador de retraso. Por lo tanto, Seasonal-QVAR se convierte en Gaussian Seasonal-VARMA, en el que el término de error $\mathbf{v}_t$ tiene una distribución multivariada *i.i.d.* La distribución gaussiana y tanto el nivel local $\boldsymbol{\mu}_t$ como la estacionalidad estocástica $\mathbf{s}_t$ se actualizan mediante $\mathbf{v}_{t-1}$. Gaussian Seasonal VARMA es un caso especial de Seasonal-QVAR para ν grande. Además, Gaussian Seasonal-VAR es otro caso especial de Seasonal-QVAR para ν grandes y $\boldsymbol{\Psi} = \boldsymbol{\Phi}$. En este documento, nos enfocamos en la alternativa gaussiana estacional-VARMA.

Para Seasonal-VARMA, el nivel local $\boldsymbol{\mu}_t$ mide todos los efectos de interacción dinámica de $\mathbf{y}_t$. Por lo tanto, nos enfocamos en la representación VMA(∞)con $\boldsymbol{\mu}_t = \sum_{j=0}^{\infty} \boldsymbol{\Phi}^j \boldsymbol{\Psi}\boldsymbol{\Omega}_v^{-1}\boldsymbol{\varepsilon}_{t-1-j}$, que se obtiene usando la descomposición $\boldsymbol{\Sigma}_v = \boldsymbol{\Omega}_v^{-1}(\boldsymbol{\Omega}_v^{-1})'$. Sea $C_{\mu,1}$ el módulo máximo de valores propios de $\boldsymbol{\Phi}$. Si $C_{\mu,1} < 1$, entonces la serie en la representación VMA(∞) es finita. Para VARMA estacional, $IRF_{j,t} = \partial\boldsymbol{\mu}_{t+j}/\partial\boldsymbol{\varepsilon}_{t-1} = \boldsymbol{\Phi}^j\boldsymbol{\Psi}\boldsymbol{\Omega}_v^{-1}$ para $j = 0, 1, \ldots, \infty$.

Estimamos el Seasona-VARMA utilizando el método MV. Incluso si la suposición $\mathbf{v}_t \sim N_N(0, \boldsymbol{\Sigma}_v)$ no se cumple, el estimador MV aún proporciona estimaciones de parámetros consistentes para Seasonal-VARMA de acuerdo con los resultados cuasi-MV (QMV) de Gouri eroux, Monfort y Trognon (1984). Denotamos los módulos máximos de valores propios de $\boldsymbol{\Phi}$ y $\boldsymbol{\Psi} - \boldsymbol{\Phi}$ con $C_{\mu,1}$ y $C_{\mu,2}$, respectivamente. Para Seasonal-VARMA, $C_{\mu,1} < 1$ y $C_{\mu 2} < 1$ aseguran que MV sea consistente y asintóticamente normal.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Capítulo 6

Intervención

A menudo las series reales se ven con frecuencia sujetas a la influencia de factores externos. Por ejemplo se puede mencionar medidas económicas, cambios tecnológicos, las huelgas, cambios en la legistación Uriel and Peiro (2000) o un "shock"generalizado provocados por pandemia. Si modelamos estos efectos en la serie podemos mejorar la precisión de la estimación de los parámetros y las previsiones. Al verse afectadas las series por estos fenómenos, el análisis de su dinámica estaría distorcionado si no los tuvieramos en cuenta dentro del modelo. Uno de los principales métodos para modelar estos fenómenos es el análisis de intervención denominado así por Box and Tiao (1975). Ésta técnica permite modelar las series junto con la consideración de aquellos fenómenos que, sin formar parte de la dinámica básica de la serie, afectan a los valores que toman a lo largo del tiempo.

6.1. Variables impulso

Supongamos una variable y_t que sigue un modelo Seasonal-t-QVAR

$$\mathbf{y}_t = \boldsymbol{\omega} + \boldsymbol{\mu}_t + \mathbf{s}_t + \mathbf{v}_t, \tag{6.1}$$

y que esta serie está afectada en un instante dado, $t = h$, por un suceso conocido. En consecuencia, no observamos en todos los instantes esta serie, ya que para $t = h$ observaremos una variable z_t que esta relacionada con $\mathbf{y}_t$ mediante

$$\mathbf{z}_h = \boldsymbol{\tau} + \mathbf{y}_h,$$

donde $\boldsymbol{\tau}$ es la magnitud del efecto sobre la serie.

Para representar el instante de ocurrencia de este suceso se define una variable I_t impulso

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

$$I_t^h = 0, \quad \text{si } t \neq h$$
$$I_t^h = 1, \quad \text{si } t = h$$

y supondremos que la serie observada, $\mathbf{z}_t$, sigue el modelo

$$\mathbf{z}_t = \boldsymbol{\tau} I_t^{(h)} + \mathbf{y}_t,$$

es decir, utilizando 6.1

$$\mathbf{z}_t = \boldsymbol{\tau} I_t^{(h)} + \boldsymbol{\omega} + \boldsymbol{\mu}_t + \mathbf{s}_t + \mathbf{v}_t. \tag{6.2}$$

Este modelo establece que la serie, z_t, sigue el modelo Seasonal-t-QVAR $\mathbf{z}_t = \boldsymbol{\omega} + \boldsymbol{\mu}_t + \mathbf{s}_t + \mathbf{v}_t$ en todos los instantes de tiempo en que la variable impulso, $I_t^{(h)}$, es cero. Sin embargo para $t = h$, observamos la suma del valor que proviene del modelo Seasonal-t-QVAR, y_h, que es desconocido, y del efecto efecto determinista de $\boldsymbol{\tau}$. La representación(6.2) es un ejemplo de función de trasferencia, es decir, de una función que describe el efecto sobre una variable respuesta, z_t, de una variable de entrada, $I_t^{(h)}$. La función de trasferencia en este caso es simplemente $\boldsymbol{\tau}$.

6.2. Variable escalón

Estas ideas pueden extenderse para modelar intervenciones que tienen un efecto permanente sobre la serie como por ejemplo un subida de precios. Estas intervenciones se modelan con variables escalón, que se definen mediante

$$\begin{aligned} E_t^{(h)} &= 0, \quad \text{si } t < h \\ E_t &= 1, \quad \text{si } t \geq h. \end{aligned} \tag{6.3}$$

El efecto de esta variable sobre un serie que sigue el modelo Seasonal-t-QVAR dado por la ecuación (6.1) puede representarse mediante el modelo de intervención

$$z_t = \boldsymbol{\tau} E_t^{(h)} + \boldsymbol{\omega} + \boldsymbol{\mu}_t + \mathbf{s}_t + \mathbf{v}_t,$$

que implica que todos los valores posteriores al instante h están afectados por una cantidad constante, $\boldsymbol{\tau}$. Este efecto equivale a un cambio de nivel en la serie a partir del instante h

$$\mathbf{z}_t = \mathbf{y}_t, \quad t < h$$
$$\mathbf{z}_{h+j} = \boldsymbol{\tau} + \mathbf{y}_{h+j}, \quad j \geq 0.$$

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

6.3. Relación entre impulsos y escalones

La diferencia básica entre una variable impulso y una escalón es que el efecto de la primera se reduce a un intervalo de tiempo mientras que el de la segunda se extiende hasta el final del periodo observado. Suele decirse que el efecto de un impulso en la serie es transitorio, mientras que el de un escalón es permanete.

En determinados casos, no se conoce a priori si un efecto es permanente o transitorio. Por ejemplo, un aumento de precios en h puede, en algunos productos, disminuir las ventas a largo plazo (efecto permanente), mientras que, en otros, provoca únicamente una leve disminución de ventas en los periodos posteriores a la subida, que desaparece en pocos periodos. Para diferenciar ambas respuestas, tendremos en cuenta que un impulso es la primera diferencia de un escalón

$$I_t^{(h)} = E_t^{(h)} - E_{t-1}^{(h)} = \nabla E_t^{(h)}. \tag{6.4}$$

Por lo tanto, podemos permitir que los datos nos indiquen si el efecto debe modelarse mediante un impulso o un escalón. Si estimamos el modelo

$$\mathbf{z}_t = \boldsymbol{\tau}_0 E_t^{(h)} - \boldsymbol{\tau}_1 E_{t-1}^{(h)} + \boldsymbol{\omega} + \boldsymbol{\mu}_t + \mathbf{s}_t + \mathbf{v}_t$$

y obtenemos que $\hat{\boldsymbol{\tau}}_\mathbf{0} \simeq \hat{\boldsymbol{\tau}}_\mathbf{1}$, resulta

$$\mathbf{z}_t = \hat{\boldsymbol{\tau}}_\mathbf{0} \nabla E_t^{(h)} + \boldsymbol{\omega} + \boldsymbol{\mu}_t + \mathbf{s}_t + \mathbf{v}_t, \tag{6.5}$$

que puede escribirse

$$\mathbf{z}_t = \hat{\boldsymbol{\tau}}_\mathbf{0} I_t^{(h)} + \boldsymbol{\omega} + \boldsymbol{\mu}_t + \mathbf{s}_t + \mathbf{v}_t,$$

indicando que la serie sufre únicamente un efecto transitorio. Por el contrario, si obtenemos $\hat{\boldsymbol{\tau}}_\mathbf{1} = 0$ o un valor de $\hat{\boldsymbol{\tau}}_\mathbf{0}$ con signo opuesto al de $\hat{\boldsymbol{\tau}}_\mathbf{1}$, esto implica que existe un efecto permanente.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Capítulo 7

Resultados

7.1. Aplicación del modelo Seasonal-t-QVAR

En este capítulo presentamos los principales hallazgos al aplicar el modelo Seasonal-t-QVAR(1) el cual se utiliza para estudiar los efectos de la interacción dinámica entre la tasa de inflación $y_{1,t}$ y la tasa de desempleo $y_{2,t}$ de Estados Unidos, que involucran componentes de estacionalidad anual. Se comparan los rendimientos estadísticos del modelo Seasonal-t-QVAR(1) y el modelo clásico Seasonal-Gaussian-VARMA(1,1), que es un caso especial limite del Seasonal-t-QVAR(1). Ademas, se vuelve a presentar el modelo Seasonal-t-QVAR utilizando un análisis de intervención para modelar el comportamiento de las series ante el SARS-CoV-2.

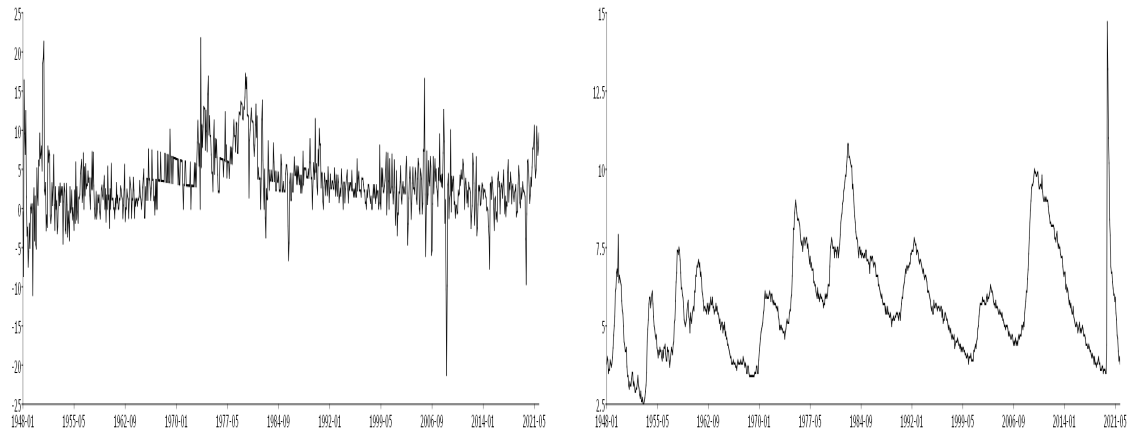
Los datos mensuales sobre la tasa de inflación $y_{1,t}$ y la tasa de desempleo de Estados Unidos $y_{2,t}$ se utilizan para el período de enero de 1948 a febrero de 2022 (fuente: Datos económicos de la Reserva Federal). Las estadísticas descriptivas de $y_{1,t}$ y $y_{2,t}$ se presentan en el Cuadro 7.1. La evolución de $y_{1,t}$ y $y_{2,t}$ se presenta en la Figura 7.1.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Cuadro 7.1: Estadística descriptiva (Tasa de inflación y tasa de desempleo en Estados Unidos)

Estadística descriptiva	Tasa de inflación $y_{1,t}$	Tasa de desempleo $y_{2,t}$
Fecha de inicio	Febrero 1948	Febrero 1948
Fecha de término	Febrero 2022	Febrero 2022
Tamaño de la muestra	889	889
Mínimo	-21.25	2.5
Máximo	21.72	14.7
Promedio	3.37	5.76
Mediana	2.94	5.60
SD	3.99	1.70
Asimetria	0.39	0.84
Curtosis	6.78	4.07

Figura 7.1: Evolución de la tasa de inflación y la tasa de desempleo de Estados Unidos para el período de febrero 1948 a febrero de 2022.



(a) Tasa de inflación, $y_{1,t}$

(b) Tasa de desempleo, $y_{2,t}$

7.1.1. Resultados de la estimación

Las estimaciones de parámetros, las condiciones de MV y los rendimientos estadísticos de los modelos Seasonal-Gaussian-VARMA(1,1), Seasonal-t-QVAR(1) y Seasonal-t-QVAR(1)+intervención se presentan en el Cuadro 7.2. Para los modelos, $\hat{C}_{\mu,1} < 1$, $\hat{C}_{\mu,2} < 1$ y $\hat{C}_{\mu,3} < 1$, por lo que se admiten las condiciones necesarias del estimador MV. Para el modelo Seasonal-Gaussian-VARMA(1,1), Seasonal-t-QVAR(1) y Seasonal-t QVAR(1)+intervención, los componentes de nivel local y los componentes de estacionalidad anual se presentan en la Figura 7.2 (c-j) y Figura 7.3 (a-d).

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Cuadro 7.2: Estimaciones de parámetros y diagnósticos de modelos (tasa de inflación y tasa de desempleo de Estados Unidos)

Seasenal-Gaussian-VARMA(1,1)	Parámetros	Seasenal-t-QVAR(1)	Parámetros	Seasenal-t-QVAR(1) + Intervención	Parámetros
ω_1	3.3238*** (0.9971)	ω_1	-1.1847* (0.6132)	ω_1	1.2328* (0.6884)
ω_2	4.1763*** (0.1618)	ω_2	4.0892*** (0.1516)	ω_2	3.9844*** (0.0745)
$\Phi_{1,1}$	0.9737*** (0.0519)	$\Phi_{1,1}$	0.9999*** (1.9342)	$\Phi_{1,1}$	0.9908*** (0.0183)
$\Phi_{1,2}$	0.0111 (0.0475)	$\Phi_{1,2}$	-0.0113 (0.0475)	$\Phi_{1,2}$	-0.0002 (0.0500)
$\Phi_{2,1}$	-0.0005 (0.0499)	$\Phi_{2,1}$	0.0093 (0.6999)	$\Phi_{2,1}$	0.0087 (0.6999)
$\Phi_{2,2}$	0.9819*** (0.0359)	$\Phi_{2,2}$	0.9873*** (0.0253)	$\Phi_{2,2}$	0.9922*** (0.0156)
$\kappa_{1,1}$	0.1794 (0.1786)	$\kappa_{1,1}$	0.1978*** (0.0043)	$\kappa_{1,1}$	0.1969*** (0.0061)
$\kappa_{1,2}$	-0.3400* (0.1857)	$\kappa_{1,2}$	-0.2730 (0.2467)	$\kappa_{1,2}$	-0.5257** (0.2489)
$\kappa_{2,1}$	-0.0022 (0.0189)	$\kappa_{2,1}$	-0.0090 (0.0484)	$\kappa_{2,1}$	-0.0057 (0.0493)
$\kappa_{2,2}$	1.0674*** (0.0364)	$\kappa_{2,2}$	1.6782*** (0.0469)	$\kappa_{2,2}$	1.6028*** (0.0422)
				τ_1	1.0832 (0.9249)
				τ_2	3.9498*** (0.1466)
$\kappa_{Ene,1}$	0.3924*** (0.1227)	$\kappa_{Ene,1}$	0.5255*** (0.1870)	$\kappa_{Ene,1}$	0.5024*** (0.1075)
$\kappa_{Feb,1}$	0.3425*** (0.1233)	$\kappa_{Feb,1}$	0.2991** (0.1346)	$\kappa_{Feb,1}$	0.2435** (0.0976)
$\kappa_{Mar,1}$	0.2844*** (0.1088)	$\kappa_{Mar,1}$	0.6116*** (0.1295)	$\kappa_{Mar,1}$	0.5588*** (0.1180)
$\kappa_{Abr,1}$	0.1566 (0.1003)	$\kappa_{Abr,1}$	0.5446*** (0.1426)	$\kappa_{Abr,1}$	0.4619*** (0.0928)
$\kappa_{May,1}$	0.1792** (0.0854)	$\kappa_{May,1}$	0.2634** (0.1337)	$\kappa_{May,1}$	0.2366*** (0.0688)
$\kappa_{Jun,1}$	0.5810*** (0.1266)	$\kappa_{Jun,1}$	0.7214*** (0.1404)	$\kappa_{Jun,1}$	0.6636*** (0.0667)
$\kappa_{Jul,1}$	0.2837** (0.1305)	$\kappa_{Jul,1}$	0.4564*** (0.1607)	$\kappa_{Jul,1}$	0.4150*** (0.1407)
$\kappa_{Ago,1}$	-0.1047 (0.1073)	$\kappa_{Ago,1}$	0.2147 (0.1498)	$\kappa_{Ago,1}$	0.1837 (0.1161)
$\kappa_{Sep,1}$	0.1077 (0.1087)	$\kappa_{Sep,1}$	0.3253* (0.1671)	$\kappa_{Sep,1}$	0.3142** (0.1416)
$\kappa_{Oct,1}$	0.0203 (0.0870)	$\kappa_{Oct,1}$	0.0942 (0.1579)	$\kappa_{Oct,1}$	0.0864 (0.1289)
$\kappa_{Nov,1}$	0.4808*** (0.1088)	$\kappa_{Nov,1}$	0.2126 (0.1573)	$\kappa_{Nov,1}$	0.2296** (0.0939)
$\kappa_{Dic,1}$	0.2942*** (0.0954)	$\kappa_{Dic,1}$	0.4768*** (0.1548)	$\kappa_{Dic,1}$	0.4473*** (0.1152)
$\kappa_{Ene,2}$	-0.0217 (0.0318)	$\kappa_{Ene,2}$	-0.2474** (0.1040)	$\kappa_{Ene,2}$	-0.0044 (0.0143)
$\kappa_{Feb,2}$	0.0700*** (0.0189)	$\kappa_{Feb,2}$	0.1482*** (0.0532)	$\kappa_{Feb,2}$	0.1398*** (0.0320)
$\kappa_{Mar,2}$	-0.1188*** (0.0335)	$\kappa_{Mar,2}$	-0.0416*** (0.0161)	$\kappa_{Mar,2}$	-0.0491*** (0.0181)
$\kappa_{Abr,2}$	0.3836*** (0.0926)	$\kappa_{Abr,2}$	-0.0162* (0.0086)	$\kappa_{Abr,2}$	-0.1520*** (0.0331)
$\kappa_{May,2}$	-0.0106 (0.0123)	$\kappa_{May,2}$	-0.0060*** (0.0017)	$\kappa_{May,2}$	0.0068 (0.0103)
$\kappa_{Jun,2}$	0.0172 (0.0243)	$\kappa_{Jun,2}$	0.0555** (0.0219)	$\kappa_{Jun,2}$	-0.0632** (0.0291)
$\kappa_{Jul,2}$	-0.0063 (0.0169)	$\kappa_{Jul,2}$	-0.1235*** (0.0315)	$\kappa_{Jul,2}$	-0.1218** (0.0484)
$\kappa_{Ago,2}$	-0.0110 (0.0128)	$\kappa_{Ago,2}$	0.0181 (0.0211)	$\kappa_{Ago,2}$	-0.0236*** (0.0076)
$\kappa_{Sep,2}$	-0.0185 (0.0129)	$\kappa_{Sep,2}$	-0.1327*** (0.0306)	$\kappa_{Sep,2}$	-0.0642*** (0.0145)
$\kappa_{Oct,2}$	-0.0370 (0.0242)	$\kappa_{Oct,2}$	-0.0132 (0.0203)	$\kappa_{Oct,2}$	-0.0908*** (0.0130)
$\kappa_{Nov,2}$	-0.0402** (0.0171)	$\kappa_{Nov,2}$	-0.0345*** (0.0121)	$\kappa_{Nov,2}$	0.0570*** (0.0144)
$\kappa_{Dic,2}$	-0.0280* (0.0170)	$\kappa_{Dic,2}$	-0.0241 (0.0186)	$\kappa_{Dic,2}$	0.0849 *** (0.0310)
$\Omega_{1,1}^{-1}$	3.0003*** (0.0713)	$\Omega_{1,1}^{-1}$	2.5692*** (0.0676)	$\Omega_{1,1}^{-1}$	2.5772***(0.0629)
$\Omega_{1,2}^{-1}$	-0.0675*** (0.0136)	$\Omega_{1,2}^{-1}$	-0.0232*** (0.0075)	$\Omega_{1,2}^{-1}$	-0.0168**(0.0074)
$\Omega_{2,2}^{-1}$	-0.3902*** (0.0093)	$\Omega_{2,2}^{-1}$	0.2279*** (0.0051)	$\Omega_{2,2}^{-1}$	0.2149***(0.0047)
		ν	8.6847*** (0.5915)	ν	10.8709*** (0.7784)
$C_{\mu,1}$	0.9810	$C_{\mu,1}$	0.9937	$C_{\mu,1}$	0.9915
		$C_{\mu,2}$	0.8154	$C_{\mu,2}$	0.8021
		$C_{\mu,3}$	0.6660	$C_{\mu,3}$	0.6439
LL	-2.9956	LL	-2.5332	LL	-2.4310
AIC	6.0744	AIC	5.1519	AIC	4.9520
BIC	6.2738	BIC	5.3567	BIC	5.1675
HQC	6.1507	HQC	5.2302	HQC	5.0344

Los rendimientos estadísticos de los tres modelos se comparan utilizando los siguientes criterios de rendimiento basados en la varosimilitud: LL, AIC, BIC y HQC. Estas métricas se presentan en el Cuadro 7.2. Todas las métricas basadas en la verosimilitud sugieren que el modelo Seasonal-t-QVAR(1) + intervención es superior al modelo Seasonal-t-QVAR(1) y el modelo Seasonal-Guassian-VARMA(1,1). De acuerdo con esos resultados, para la estimación de los efectos de interacción para la tasa de inflación y la tasa de desempleo , el modelo Seasonal-t-QVAR(1) + intervención es mejor en comparación a los otros dos modelos.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

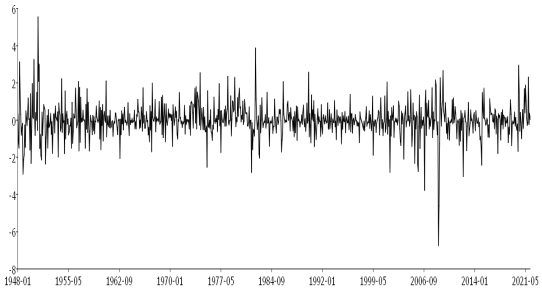
Figura 7.2: Evolución de μ_t y s_t para el modelo Seasonal-Gaussian-VARMA(1,1) (modelo clásico) y Seasonal-t-QVAR(1) (modelo DCS) para Estados Unidos en el período de febrero 1948 a febrero de 2022.



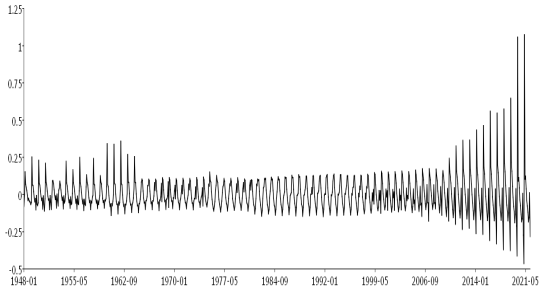
(a) Tasa de inflación, $\mu_{1,t}$ (Modelo clásico)



(b) Tasa de desempleo, $\mu_{2,t}$ (Modelo clásico)



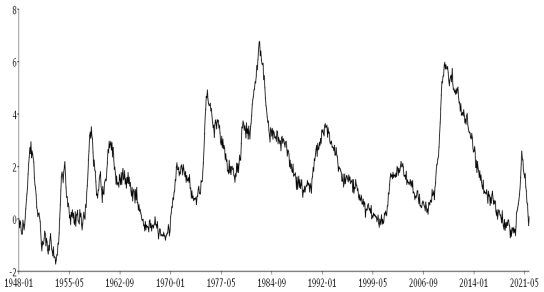
(c) Tasa de inflación, $s_{1,t}$ (Modelo clásico)



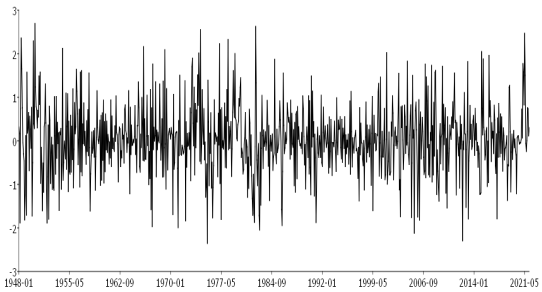
(d) Tasa de desempleo, $s_{2,t}$ (Modelo clásico)



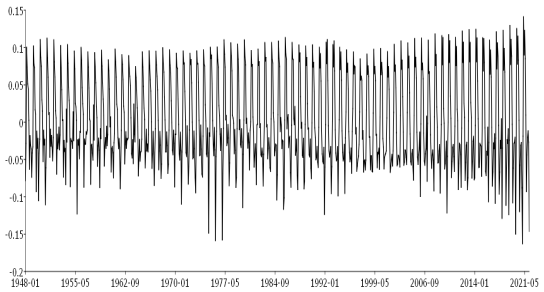
(e) Tasa de inflación, $\mu_{1,t}$ (Modelo DCS)



(f) Tasa de desempleo, $\mu_{2,t}$ (Modelo DCS)



(g) Tasa de inflación, $s_{1,t}$ (Modelo DCS)



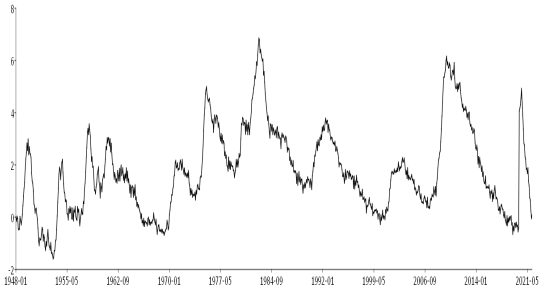
(h) Tasa de desempleo, $s_{2,t}$ (Modelo DCS)

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

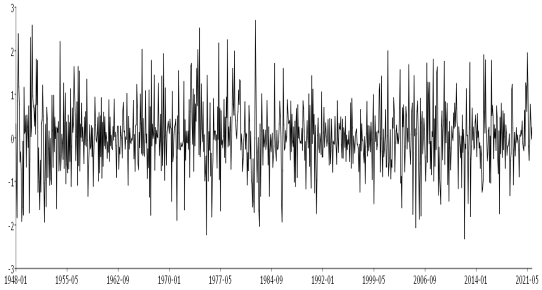
Figura 7.3: Evolución de μ_t y s_t para el modelo Seasonal-t-QVAR(1) + intervención (modelo DCS + intervención) para Estados Unidos en el período de febrero 1948 a febrero de 2022.



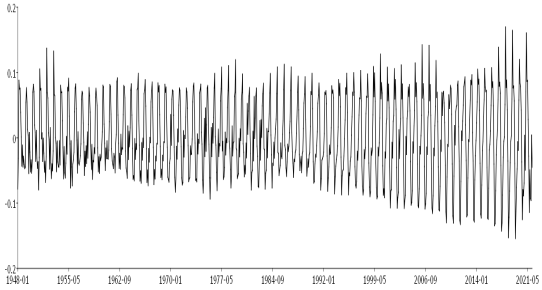
(a) Tasa de inflación, $\mu_{1,t}$ (Modelo DCS + intervención)



(b) Tasa de desempleo, $\mu_{2,t}$ (Modelo DCS + intervención)



(c) Tasa de inflación EE.UU, $s_{1,t}$ (Modelo DCS + intervención)



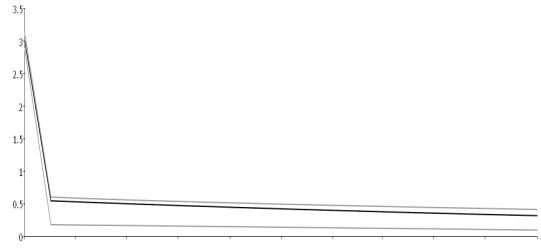
(d) Tasa de desempleo, $s_{2,t}$ (Modelo DCS + intervención)

Para los modelos Seasonal-Gaussian-VARMA(1,1), Seasonal-t-QVAR(1) y Seasonal-t-QVAR(1)+intervención, las IRF se presentan en las Figuras 7.4, 7.5 y 7.6, respectivamente. Al comparar las estimaciones de la IRF de esas cifras, la IRF para Seasonal-t-QVAR(1)+intervención brinda un apoyo más claro a la teoría presentada en Nordhaus (1975), con respecto a los efectos de interacción negativos entre la inflación y las tasas de desempleo de Estados Unidos. Los efectos 'Shock inflacionario → Tasa de desempleo' (Figura 7.4c, 7.5c y 7.6c) son significativamente negativos hasta dentro de cuatro meses para el modelo para Seasonal-t-QVAR(1)+intervención. Por otro lado, los mismos efectos son significativamente negativos durante veinte meses en el futuro para el modelo Seasonal-Gaussian-VARMA(1,1) y siete meses en el futuro para el modelo Seasonal-t-QVAR(1). Con respecto a los efectos 'Shock de desempleo → Tasa de inflación' (Figura 7.4b, 7.5b y 7.6b), la estimación de IRF es negativa para los modelos Seasonal-t-QVAR(1)+intervención y Seasonal-t-QVAR(1), aunque este ultimo no es significativo según los intervalos de confianza de Montecarlo. Para el modelo Seasonal-Guassian-VARMA(1,1) son positivos y no significativos. En definitiva,

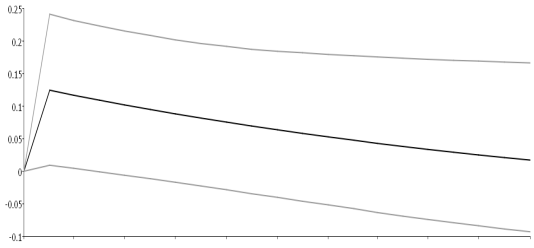
Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

de acuerdo con las Figuras 7.4b, 7.5b y 7.6b, las estimaciones IRF para el modelo Seasonal-t-QVAR(1)+intervención respaldan más consistentemente la teoría presentada en Nordhaus (1975), que las estimaciones IRF para el modelo Seasonal-t-QVAR(1) y Seasonal-Gaussian-VARMA (1,1).

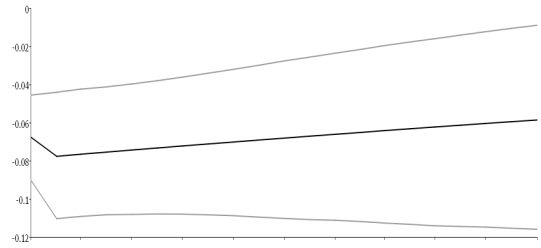
Figura 7.4: IRF hasta 20 meses para el modelo Seasonal-Gaussian-VARMA(1,1) (90 % intervalo de confianza Monte Carlo).



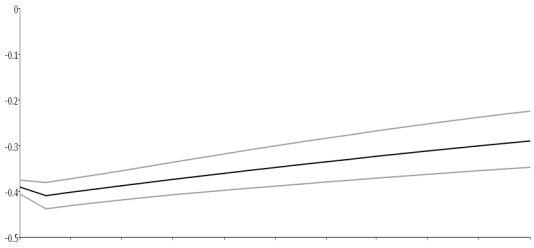
(a) Shock inflacionario → Tasa de inflación



(b) Shock de desempleo → Tasa de inflación



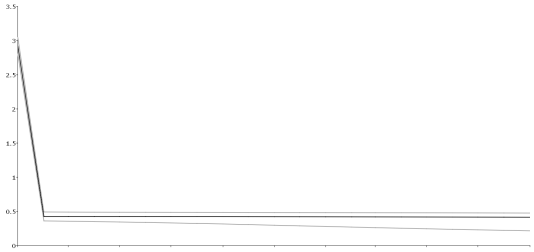
(c) Shock inflacionario → Tasa de desempelo



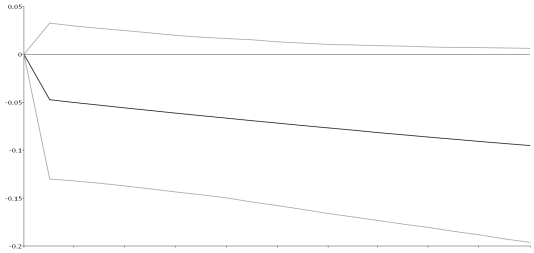
(d) Shock de desempleo → Tasa de desempleo

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

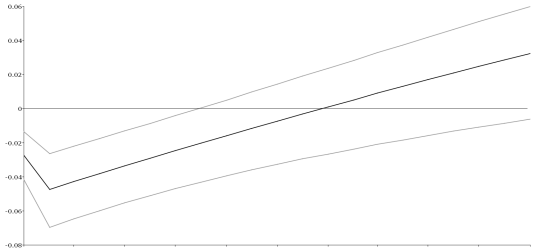
Figura 7.5: IRF hasta 20 meses para el modelo Seasonal-t-QVAR(1) (90 % intervalo de confianza Monte Carlo).



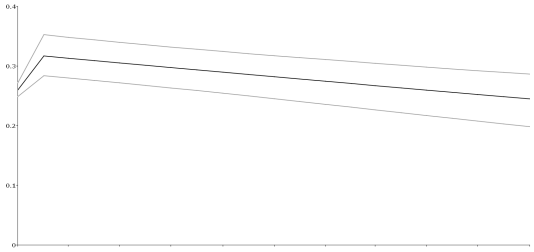
(a) Shock inflacionario → Tasa de inflación



(b) Shock de desempleo → Tasa de inflación



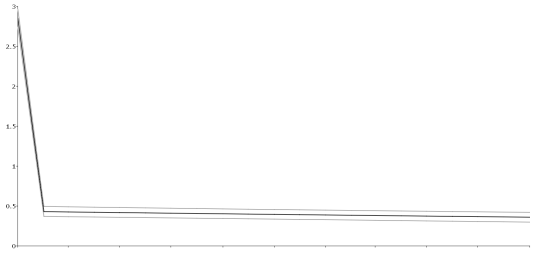
(c) Shock inflacionario → Tasa de desempelo



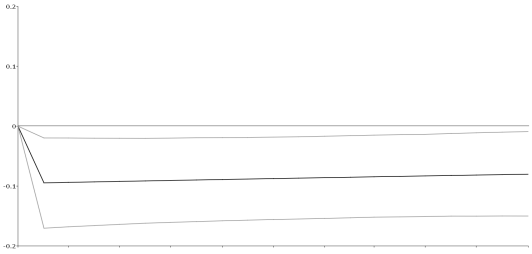
(d) Shock de desempleo → Tasa de desempleo

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

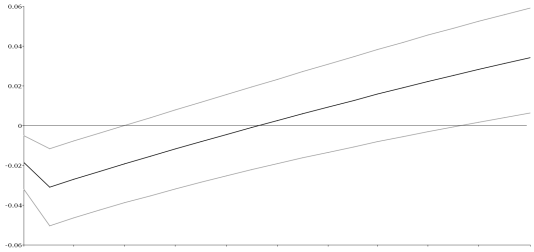
Figura 7.6: IRF hasta 20 meses para el modelo Seasonal-t-QVAR(1) + intervención (90 % intervalo de confianza Monte Carlo).



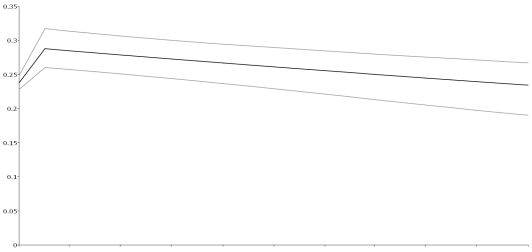
(a) Shock inflacionario → Tasa de inflación



(b) Shock de desempleo → Tasa de inflación



(c) Shock inflacionario → Tasa de desempelo



(d) Shock de desempleo → Tasa de desempleo

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

7.2. Conclusiones

En este capítulo se aplicaron los modelos DCS de series de tiempo revisados en capítulos anteriores. Los modelos DCS se pueden relacionar directamente con varios modelos clásicos de series de tiempo. Para resaltar las principales diferencias entre DCS y los modelos clásicos, se han comparado tres modelos de localización: el modelo Seasonal-Gaussian-VARMA(1,1) de señal más ruido gaussiano (es decir, un modelo clásico de series de tiempo), el modelo Seasonal-t-QVAR(1) y Seasonal-t-QVAR(1)+intervención (es decir, modelos DCS). Quizás, la principal diferencia entre DCS y los modelos clásicos es que los modelos DCS son robustos a valores atípicos (es decir, valores extremos en el componente irregular). Una propiedad interesante de los modelos DCS es que, en varios casos, esos modelos son generalizaciones de los modelos clásicos de series de tiempo. Por lo tanto, los modelos DCS pueden proporcionar un mejor ajuste a los datos de series temporales que los modelos clásicos. Este rendimiento estadístico superior se ha presentado en la aplicación de este capítulo que involucran datos de la tasa de inflación y tasa de desempleo de Estados Unidos. El rendimiento estadístico de los modelos DCS es superior al rendimiento estadístico de los modelos clásicos de series temporales y además observamos que el rendimento del modelo DCS con intervención es mejor que el modelo DCS ante la presencia de algun shock externo en las variables.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Bibliografía

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, 19(4):465–474.

Bernanke, B. S. (1986). Alternative explanations of the money-income correlation.

Blanchard, O. J. and Watson, M. W. (1986). Are business cycles all alike? In *The American business cycle: Continuity and change*, pages 123–180. University of Chicago Press.

Blasques, F., Gorgi, P., Koopman, S. J., and Wintenberger, O. (2018). Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics*, 12(1):1019–1052.

Blazsek, S., Escribano, Á., and Licht, A. (2017). Score-driven non-linear multivariate dynamic location models.

Blazsek, S., Escribano, Á., and Licht, A. (2018a). Seasonal quasi-vector autoregressive models with an application to crude oil production and economic activity in the united states and canada.

Blazsek, S., Escribano, A., and Licht, A. (2018b). Seasonality detection in small samples using score-driven nonlinear multivariate dynamic location models.

Blazsek, S. and Licht, A. (2020). Dynamic conditional score models: a review of their applications. *Applied Economics*, 52(11):1181–1199.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Box, G. E. and Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical association*, 70(349):70–79.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Brown, B. M. (1971). Martingale central limit theorems. *The Annals of Mathematical Statistics*, pages 59–66.

Casella, G. and Berger, R. L. (2021). *Statistical inference*. Cengage Learning.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, pages 573–578.

Cox, D. R., Gudmundsson, G., Lindgren, G., Bondesson, L., Harsaae, E., Laake, P., Juselius, K., and Lauritzen, S. L. (1981). Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics*, pages 93–115.

Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

Davidson, R., MacKinnon, J. G., et al. (2004). *Econometric theory and methods*, volume 5. Oxford University Press New York.

De Livera, A. M., Hyndman, R. J., and Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American statistical association*, 106(496):1513–1527.

Diebold, F. X., Tay, A., and Wallis, K. (1997). Evaluating density forecasts of inflation: the survey of professional forecasters.

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*, volume 38. OUP Oxford.

Engle, R. F., Hendry, D. F., and Richard, J.-F. (1983). Exogeneity. *Econometrica: Journal of the Econometric Society*, pages 277–304.

Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.

Hamilton, J. D. (2020). *Time series analysis*. Princeton university press.

Harvey, A. (2006). Forecasting with unobserved components time series models. *Handbook of economic forecasting*, 1:327–412.

Harvey, A. (2011). Modelling the phillips curve with unobserved components. *Applied Financial Economics*, 21(1-2):7–17.

Harvey, A. and Luati, A. (2012). Filtering with heavy tails.

Harvey, A. C. (1989). Forecasting, structural time series models and the kalman filter.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Harvey, A. C. (1993). Time series models. *Hemel Hempstead: Harvester-Wheatsheaf*.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*, volume 52. Cambridge University Press.

Jensen, S. T. and Rahbek, A. (2004). Asymptotic inference for nonstationary garch. *Econometric Theory*, 20(6):1203–1226.

Johnson, N. L. and Kotz, S. (1972). Distributions in statistics: Continuous multivariate distributions, john wiley & sons. *Inc. New York*.

Koenker, R. and Hallock, K. F. (2001). Quantile regression. *Journal of economic perspectives*, 15(4):143–156.

Kohn, R., Ansley, C. F., and Wong, C.-M. (1992). Nonparametric spline regression with autoregressive moving average errors. *Biometrika*, 79(2):335–346.

Koopman, S. J., Harvey, A. C., Doornik, J. A., and Shephard, N. (2009). Stamp 8.2: Structural time series analysis modeller and predictor. *London: Timberlake Consultants*.

Koopman, S. J., Lucas, A., and Scharth, M. (2016). Predicting time-varying parameters with parameter-driven and observation-driven models. *Review of Economics and Statistics*, 98(1):97–110.

Kotz, S. and Nadarajah, S. (2004). *Multivariate t-distributions and their applications*. Cambridge University Press.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1-3):159–178.

Lange, K. L., Little, R. J., and Taylor, J. M. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, 84(408):881–896.

Lin, T.-I. and Wang, Y.-J. (2009). A robust approach to joint modeling of mean and scale covariance for longitudinal data. *Journal of Statistical Planning and Inference*, 139(9):3013–3026.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Maronna, R. A., Martin, R. D., Yohai, V. J., and Salibián-Barrera, M. (2019). *Robust statistics: theory and methods (with R)*. John Wiley & Sons.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Masreliez, C. (1975). Approximate non-gaussian filtering with linear state and observation relations. *IEEE Transactions on Automatic Control*, 20(1):107–110.

McCullagh, P. and Nelder, J. (1989). Generalized linear models 2nd edition, vol. 37 760 of. *Monographs on Statistics and Applied Probability*, page 761.

Mitchell, J. and Wallis, K. F. (2011). Evaluating density forecasts: Forecast combinations, model mixtures, calibration and sharpness. *Journal of Applied Econometrics*, 26(6):1023–1040.

Muler, N., Pena, D., and Yohai, V. J. (2009). Robust estimation for arma models. *The Annals of Statistics*, 37(2):816–840.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245.

Nyblom, J. and Mäkeläinen, T. (1983). Comparisons of tests for the presence of random walk coefficients in a simple linear model. *Journal of the American Statistical Association*, 78(384):856–864.

Robert, C. P., Casella, G., and Casella, G. (2010). *Introducing monte carlo methods with r*, volume 18. Springer.

Schick, I. C. and Mitter, S. K. (1994). Robust recursive estimation in the presence of heavy-tailed observation noise. *The annals of statistics*, 22(2):1045–1080.

Shephard, N. G. and Harvey, A. C. (1990). On the probability of estimating a deterministic component in the local level model. *Journal of time series analysis*, 11(4):339–347.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, pages 1–48.

Sims, C. A. et al. (1986). Are forecasting models usable for policy analysis? *Quarterly Review*, 10(Win):2–16.

Sims, C. A., Goldfeld, S. M., and Sachs, J. D. (1982). Policy analysis with econometric models. *Brookings papers on economic activity*, 1982(1):107–164.

Slater, L. (1965). Handbook of mathematical functions, edited by m. abramowitz & ia stegun. *New York: Dover. Temme, NM (1983). Numer. Math*, 41:63–82.

Stock, J. H. and Watson, M. W. (2001). Vector autoregressions. *Journal of Economic perspectives*, 15(4):101–115.

Universidad del Bío-Bío. Sistema de Bibliotecas - Chile

Taylor, J. and Verbyla, A. (2004). Joint modelling of location and scale parameters of the t distribution. *Statistical Modelling*, 4(2):91–112.

Teräsvirta, T. (2006). Forecasting economic variables with nonlinear models. *Handbook of economic forecasting*, 1:413–457.

Tiao, G. C. and Tsay, R. S. (1989). Model specification in multivariate time series. *Journal of the Royal Statistical Society: Series B (Methodological)*, 51(2):157–195.

Uriel, E. and Peiro, A. (2000). *Introducción al análisis de series temporales*. Paraninfo.

Whittle, P. (1983). Prediction and regulation, 2nd edn., revised. *University of Minnesota*.

Zhu, D. and Zinde-Walsh, V. (2009). Properties and estimation of asymmetric exponential power distribution. *Journal of econometrics*, 148(1):86–99.