

Martínez-Ovando, Juan Carlos; Walker, Stephen G.

**Working Paper**

## Modelación de series de tiempo, estacionariedad y métodos Bayesianos no paramétricos

Working Papers, No. 2011-08

**Provided in Cooperation with:**

Bank of Mexico, Mexico City

*Suggested Citation:* Martínez-Ovando, Juan Carlos; Walker, Stephen G. (2011) : Modelación de series de tiempo, estacionariedad y métodos Bayesianos no paramétricos, Working Papers, No. 2011-08, Banco de México, Ciudad de México

This Version is available at:

<https://hdl.handle.net/10419/83714>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

**Banco de México**  
**Documentos de Investigación**

**Banco de México**  
**Working Papers**

**N° 2011-08**

**Modelación de Series de Tiempo, Estacionariedad y  
Métodos Bayesianos No Paramétricos**

**Juan Carlos Martínez-Ovando**  
Banco de México

**Stephen G. Walker**  
University of Kent

Septiembre 2011

La serie de Documentos de Investigación del Banco de México divulga resultados preliminares de trabajos de investigación económica realizados en el Banco de México con la finalidad de propiciar el intercambio y debate de ideas. El contenido de los Documentos de Investigación, así como las conclusiones que de ellos se derivan, son responsabilidad exclusiva de los autores y no reflejan necesariamente las del Banco de México.

The Working Papers series of Banco de México disseminates preliminary results of economic research conducted at Banco de México in order to promote the exchange and debate of ideas. The views and conclusions presented in the Working Papers are exclusively of the authors and do not necessarily reflect those of Banco de México.

# Modelación de Series de Tiempo, Estacionariedad y Métodos Bayesianos No Paramétricos\*

Juan Carlos Martínez-Ovando<sup>†</sup>  
Banco de México

Stephen G. Walker<sup>‡</sup>  
University of Kent

**Resumen:** En este artículo introducimos dos modelos generales no paramétricos estacionarios de primer orden para series de tiempo, para los cuales las distribuciones marginales (invariantes) y de transiciones son expresadas como mezclas de dimensión infinita. Esta característica los hacen los primeros modelos Bayesianos estacionarios completamente no paramétricos desarrollados a la fecha. Contribuimos a la discusión sobre el uso de modelos estacionarios en la práctica, como una motivación, y apoyamos la visión que, modelos estacionarios flexibles (no paramétricos) pueden ser una fuente confiable para inferencias y predicciones. Se observa que nuestros modelos se acoplan apropiadamente al esquema Bayesiano de inferencia debido a un adecuado teorema de representación. Un modelo estacionario de mezcla de escalas es desarrollado como caso particular junto con una estrategia computacional para realizar inferencia posterior y predicciones. La utilidad de este modelo es ilustrado con el análisis de los log-retornos del tipo de cambio euro-dólar E.U.A.

**Palabras Clave:** Estacionariedad, Procesos de Markov, Modelos de mezclas dinámicos, Medidas aleatorias de probabilidad, Medidas aleatorias de probabilidad condicional, Procesos latentes.

**Abstract:** In this paper we introduce two general non-parametric first-order stationary time-series models for which marginal (invariant) and transition distributions are expressed as infinite-dimensional mixtures. That feature makes them the first Bayesian stationary fully non-parametric models developed so far. We draw on the discussion of using stationary models in practice, as a motivation, and advocate the view that flexible (non-parametric) stationary models might be a source for reliable inferences and predictions. It will be noticed that our models adequately fit in the Bayesian inference framework due to a suitable representation theorem. A stationary scale-mixture model is developed as a particular case along with a computational strategy for posterior inference and predictions. The usefulness of that model is illustrated with the analysis of Euro/USD exchange rate log-returns.

**Keywords:** Stationarity, Markov processes, Dynamic mixture models, Random probability measures, Conditional random probability measures, Latent processes.

**JEL Classification:** C11, C14, C15, C22, C51.

---

\*Los autores agradecen a dos dictaminadores anónimos por sus valiosos comentarios y sugerencias. El primer autor agradece, entre otros, el apoyo financiero de CONACYT (México), University of Kent (UK), Regione Piemonte, Università degli Studi di Torino y Collegio Carlo Alberto (Italia). Se agradece la asistencia de Nadja Thürbeck en la traducción al español del artículo.

<sup>†</sup>Dirección General de Investigación Económica. Email: juan.martinez@banxico.org.mx. (Autor correspondiente).

<sup>‡</sup>School of Mathematics, Statistics and Actuarial Science. Email: S.G.Walker@kent.ac.uk.

# 1. Introducción

Gran parte de la teoría estadística conocida se basa en cierto tipo de simetría entre las observaciones (véase, por ejemplo [Zabell, 2005](#)). En el contexto de análisis de series de tiempo, la noción de simetría se ha expresado en términos de estacionariedad de la cual se derivan propiedades de muestras grandes y otros resultados teóricos. Sin embargo, una desventaja exhibida por la mayoría de los modelos estacionarios, que causa dudas a los usuarios acerca de la utilidad de los modelos, ha sido expuesta por la naturaleza estructural restrictiva de sus mecanismos de transición o distribuciones marginales. La tesis de los autores es que el uso de modelos estacionarios flexibles puede producir inferencias y predicciones razonables. En el resto de esta sección, se revisan algunos problemas generales encontrados en modelos estacionarios cuando se trata de datos reales y se comenta acerca de algunos enfoques que han resultado ser útiles para superar la falta de flexibilidad de modelos estacionarios existentes. También, se argumenta que los modelos estacionarios, cuando están conectados con los modelos bayesianos no paramétricos, pueden ser tan flexibles como muchos de sus contrapartes no estacionarios, o aún más, mientras que preservan todas las propiedades deseadas que un modelo estadístico puede tener. A continuación, se enfoca en la conexión entre la estacionariedad y los métodos bayesianos no paramétricos.

## 1.1. Estacionariedad y métodos bayesianos no paramétricos

El mundo en que vivimos está expuesto a cambios permanentes y a veces imprevistos, igual como los datos que se obtienen de él. Debido a cambios no anticipados observados en los datos, particularmente, en el contexto de series de tiempo, usuarios han desarrollado incertidumbre respecto al uso de los modelos estacionarios. En la literatura econométrica, una teoría influyente para realizar pronósticos, que logra resolver algunas fallas de modelos econométricos paramétricos existentes, se puede caracterizar básicamente por utilizar mezclas de los modelos cuyas formas estructurales no difieren en más que un valor específico de los parámetros involucrados. Véase [Clements y Hendry \(2008\)](#) para obtener un resumen reciente. Notemos que estos modelos se especifican a menudo como no estacionarios; sin embargo, combinan diferentes mecanismos que describen la dinámica de transición de los

datos y (posiblemente) anticipen valores futuros y cambios repentinos. También, véase [Krolzig \(1997\)](#) y [Frühwirth-Schnatter \(2006\)](#), donde se impone una estructura dinámica de Markov para describir las transiciones de datos de un componente del modelo a otro. Nótese que modelos de regímenes cambiantes Markov se propusieron por [Hamilton \(1989\)](#) para modelar patrones no estacionarios de datos de series de tiempo económicos. La relevancia y el buen desempeño en términos de predicción/ pronóstico de diferentes tipos de modelos de mezclas en econometría empírica han sido demostrados en [Clements y Hendry \(1999\)](#), incluso si estos modelos sufren de una falta de soporte teórico que pueden proporcionar los modelos estacionarios. De manera similar, modelos no estacionarios han sido utilizados en Econometría Financiera y Estadística Financiera (véase, por ejemplo [Giacominia et al., 2008](#)), con el objetivo de capturar de manera flexible el mecanismo dinámico del proceso generador de datos, dado que existe tal mecanismo. También, véase [Haas et al. \(2004\)](#) y [Barry y Winkler \(1976\)](#) donde se discuten algunas implicaciones de usar modelos no estacionarios en el análisis de portafolio.

Relacionado al problema de modelar cambios estructurales en Estadística, también se prestó atención a desarrollar procedimientos estadísticos con el fin de identificar (y probar) si los cambios observados en los datos representan cambios estructurales reales en el modelo. Véase, por ejemplo [Chu et al. \(1996\)](#) y [Zeileis et al. \(2005\)](#), para una discusión. Pero, como ya es bien sabido, procedimientos estadísticos destinados a identificar cambios estructurales sufren de la falla de identificar sólo cambios estructurales en retrospectiva y no se puede obtener mucha información acerca de cambios futuros. Por lo tanto, sería óptimo concebir un modelo estadístico que integre distintas posibilidades estructurales en una sola especificación, ya que capturaré una gran variedad de escenarios de datos. En cierto modo, lo anterior ha sido el objetivo de modelos de mezclas generales y de mezclas de Markov, existentes en la literatura.

Al preservar la idea de integrar formas estructurales distintas en un solo modelo, es razonable pensar que la dinámica de transición de mezclas de dimensiones infinitas sería una alternativa sensible y adecuada para modelar la dinámica de transición de datos de series de tiempo. Este tipo de modelos potencialmente capturaré características persistentes de los datos, así como eventos transitorios, pero este modelo también permitiré considerar

escenarios alternativos para inferencias futuras y, más importante, predicciones. Un modelo de dimensión infinita idealmente aprendería y asignaría mayores probabilidades a eventos recurrentes y menores probabilidades a eventos raramente observados. Otra característica deseable que pueden tener los mecanismos de transición de dimensión infinita, y que modelos no estacionarios probablemente no logren, es la de reproducir el comportamiento marginal de los datos para cualquier tiempo, lo que también ya sido una falla de la mayoría de modelos de mezclas de dimensión finita desarrollados anteriormente. Entonces, métodos bayesianos no paramétricos pueden llevar un papel importante. Refiérase a [Lijoi y Prünster \(2010\)](#) para una revisión reciente. Algunos modelos de mezcla de dimensión infinita para series de tiempo han sido desarrollados recientemente, pero hasta ahora ninguno de ellos puede reproducir el comportamiento marginal de los datos de manera flexible (véase algunos de estos documentos incluidos en [Barber et al., 2008, 2011](#)). Lo que contribuye este documento a la literatura es la introducción de modelos completamente estacionarios no paramétricos que permiten distribuciones flexibles marginales y de transición, que se representan como mezclas de dimensión infinita, así haciendo estos modelos las únicas alternativas que logran este objetivo dentro de un enfoque bayesiano no paramétrico. La construcción del modelo que se adopta en el presente documento sigue el enfoque de proceso latente de [Pitt et al. \(2002\)](#), pero se explota una estructura estocástica más flexible alrededor de la cual se forma la dependencia del modelo estacionario.

## 1.2. Acerca de estacionariedad y teoremas de representación

Adicionalmente, dentro del enfoque estadístico bayesiano, la noción de estacionariedad condicional y ergodicidad condicional de un modelo resulta relevante ya que da lugar a una versión dinámica análoga al teorema de representación de Finetti. ([de Finetti, 1937](#)). Tal resultado, de hecho, se establece bajo la noción más general de simetría estocástica bajo medidas invariantes debido a [Maitra \(1977\)](#) (también véase, [Aldous, 1985](#)). La representación de [Maitra](#) explica que para una secuencia contable de variables aleatorias estacionarias, por ejemplo  $\{Y_t\}_{t=1}^{\infty}$ , su medida de probabilidad asociada,  $\mathbb{P}$ , permite la siguiente representación para cualquier sub-segmento finito de la secuencia con longitud  $n$  y cualquier conjunto

medible  $(A_t)_{t=1}^n$ ,

$$\mathbb{P}(Y_1 \in A_1, \dots, Y_n \in A_n) = \int_{\Theta} \prod_{t=1}^n F_{\theta}(y_t \in A_t | y_{t-1}, \dots, y_1) \Pi(d\theta), \quad (1.1)$$

donde  $\theta$  es un elemento aleatorio que toma valores en el espacio  $\Theta$  (que puede ser de dimensión finita o infinita),  $\Pi$  es una medida de probabilidad con soporte en  $\Theta$ , y  $F_{\theta}(y_t \in A_t | y_{t-1}, \dots, y_1)$  representa una colección de distribuciones con estacionariedad condicional y ergodicidad condicional indexadas por  $\theta$ , con  $F_{\theta}(y_1 \in A_1)$  siendo la distribución marginal del modelo en el tiempo  $t = 1$ . Cabe destacar que la representación (1.1) abarca el teorema de la representación de [de Finetti](#) para variables aleatorias intercambiables, en cuyo caso la estacionariedad condicional es reemplazada con independencia condicional dada  $\theta$ . El teorema de representación de [de Finetti](#) ha motivado el enfoque subjetivo de intercambiabilidad.

Un caso particular de (1.1) se obtiene cuando  $F_{\theta}$  son de Markov, en cuyo caso el modelo completo se caracteriza por  $F_{\theta}(\cdot)$  marginal y transiciones de primer orden,  $F_{\theta}(\cdot | \cdot)$ . De misma manera como en el caso de intercambiabilidad, la medida de probabilidad  $\Pi$  representaría las creencias priores acerca del modelo. Otros modelos dependientes también pueden acomodar la representación de [Maitra](#). Por lo tanto, desde un punto de vista subjetivo, el teorema de representación de [Maitra](#) puede entenderse como motivación para el desarrollo de inferencia subjetiva bayesiana para el análisis de series de tiempo bajo estacionariedad. También se nota que para modelos no estacionarios no hay un resultado de representación como el de (1.1).

Al reconsiderar (1.1), cabe mencionar que cuando se escogen  $F_{\theta}$  y  $\Pi(d\theta)$  automáticamente se pondría una distribución prior en un espacio dado de modelos estacionarios. Por ejemplo, cuando  $F_{\theta}$  son paramétricos y  $\theta$  es de dimensión finita, el modelo se reducirá a un modelo paramétrico tradicional. En este caso, se asignaría cero probabilidad a muchos conjuntos grandes de modelos alternativos (es decir, modelos paramétricos cuya forma funcional se ubica fuera de la clase de modelos donde los  $F_{\theta}$  pertenecen), por lo tanto, la probabilidad posterior para aquellos conjuntos dados los datos también será cero. Entonces, parece razonable buscar especificaciones robustas de  $F_{\theta}$  y  $\Pi(d\theta)$  que abarcan una clase grande de modelos, para evitar sorpresas después de observar los datos. Así, las especificaciones no paramétricas bayesianas, donde  $\Theta$  es típicamente de dimensión infinita, parece ser una alternativa atractiva para tales

especificaciones.

En el presente documento, se introducen dos procedimientos constructivos que buscan definir modelos completamente no paramétricos estacionarios de Markov al permitir la representación de Maitra. Se anticipa que uno de ellos, el modelo de referencia, no se puede implementar en práctica con la maquinaria computacional disponible. Sin embargo, el segundo modelo extendido supera dichas fallas. Cabe destacar que los modelos extendidos tienen representaciones alternativas de estado-espacio. Estos modelos son alternativas atractivas en comparación con modelos tradicionales de estado-espacio, ya que la parte observable de los modelos se puede recuperar analíticamente dado que componentes de observación y de estado son distribuciones regulares condicionales. Un modelo estacionario de mezcla de escalas se introduce como caso particular del anterior y se desarrolla un algoritmo computacional para inferencia posterior. También, cabe destacar que los modelos completamente estacionarios no paramétricos introducidos en el presente documento pueden ser adaptados a una variedad de circunstancias, simplemente al especificar algunos de sus componentes de acuerdo con necesidades distintas (por ejemplo, tener soportes distintos).

### **1.3. Contenido del documento**

El documento está estructurado de la siguiente manera. La Sección 2 presenta un resumen general de modelos estacionarios de primer orden. Se introduce un modelo de referencia, que generaliza el estimador no paramétrico de tipo kernel de distribuciones marginales y de transición de un proceso de Markov; también se comenta acerca de una limitación práctica. También, se resumen ideas generales relacionadas a un enfoque de construcción de modelos de procesos latentes. En la Sección 3, se introduce un enfoque más amplio para especificar modelos estacionarios de primer orden, permitiendo densidades no paramétricas marginales y de transición que superan la limitación del modelo de referencia. En la Sección 4, se desarrolla un modelo particular estacionario de mezcla de escalas, basado en mezclas dobles no paramétricos de procesos de beta-Stacy y stick-breaking. En la Sección 5, se especifican priores para el modelo estacionario de mezclas de escalas y se desarrollan estrategias analíticas y computacionales para la inferencia posterior. La aplicabilidad del modelo estacionario de mezcla de escalas se ilustra en la Sección 6 con el análisis de un segmento corto de los log-retornos



del tipo de cambio euro-dólar E.U.A. El documento concluye con algunas consideraciones y comentarios finales en la Sección 7. El documento se complementa con apéndices donde se presenta un resumen general de la construcción de modelo de procesos latentes basada en mezclas y se proporcionan detalles acerca de la implementación de un modelo dinámica de mezclas de escala.

## 2. Modelos estacionarios de primer orden

### 2.1. Enfoques previos

Algunos enfoques para construir modelos estacionarios de primer orden para series de tiempo prestan atención a especificar una forma analítica explícita para la distribución marginal, la cual es compatible con el mecanismo de transición del modelo. Entre estos, se encuentran los enfoques basados en convoluciones (Jacobs y Lewis, 1978) y los modelos basados en procesos latentes (Pitt et al., 2002).

El enfoque basado en convolución de Jacobs y Lewis (1978) utiliza el operador “thinning” para definir modelos estacionarios para series de tiempo con valor discreto con distribuciones marginales dadas. Gaver y Lewis (1980) exploran la idea y definen un modelo estacionario con marginales gamma. Joe (1996) aprovechan la propiedad de proximidad bajo convolución de la familia de dispersión exponencial de las funciones de distribución y definen modelos estacionarios con marginales en la familia exponencial cerrada bajo convolución. Jørgensen y Song (1998), Jørgensen y Song (2006) y Zhu y Joe (2006), entre otros, estudiaron extensiones relacionadas. Generalizaciones más recientes que consideran el operador “thinning” binominal también se estudiaron por Bouzar y Jayakumar (2008) y Bouzar (2010).

El enfoque de procesos latentes se desarrolló por Pitt et al. (2002). La construcción del modelo se enfocó en modelos que permiten marginales en la familia exponencial conjugada y la familia exponencial de dispersión, mientras que preservan la linealidad de las expectativas condicionales de la dinámica de transición. Extensiones adicionales admitiendo marginales y dinámica de transición más estructuradas fueron desarrolladas después por Pitt y Walker (2005, 2006), donde también se desarrollaron versiones estacionarias de modelos ARCH y

GARCH con marginales dadas (véase, por ejemplo [Engle, 2001](#)). El alcance del enfoque de procesos latentes se amplió más por [Mena y Walker \(2007\)](#) al considerar medidas latentes de probabilidad, las cuales podrían resultar en mecanismos de transición (flexibles) más estructurados, mientras que conservan la distribución marginal dada. Esta idea se aplicó en [Contreras-Cristán et al. \(2009\)](#) donde se definen modelos alternativos estacionarios de valor para series de tiempo.

Sin embargo, a pesar de la flexibilidad que puedan tener ambos enfoques para construir modelos, hasta el momento las distribuciones (invariantes) marginales todavía pertenecen a una familia paramétrica dada de funciones de distribución. En el presente documento, se introducen dos modelos estacionarios de primer orden más flexibles para los cuales ambas la distribución marginal y la distribución de transición son no paramétricas. Empecemos con la especificación del modelo de referencia.

## 2.2. Un modelo de referencia

Modelos estacionarios de primer orden se especifican por distribuciones conjuntas marginales y de segundo orden. En general, se puede considerar un componente estocástico bivariado observable,  $(Y_{t-1}, Y_t)$ , cuyas entradas toman valores en el espacio común  $\mathcal{Y}$ . Adicionalmente, se supone que cualquier par  $(Y_{t-1}, Y_t)$  sigue una distribución conjunta simétrica de dimensión finita que se denota por  $K(y_{t-1}, y_t|\theta)$ , que depende de un parámetro  $\theta$ , tal que se cumpla la siguiente relación

$$p(y_t|\theta) = \int p(y_t|y_{t-1}; \theta)p(y_{t-1}|\theta)dy_{t-1}, \quad (2.1)$$

donde  $p(y_t|y_{t-1}; \theta) = K(y_{t-1}, y_t|\theta)/p(y_{t-1}|\theta)$  con  $p(\cdot|\theta)$  es la distribución marginal común inducida por  $K$ . El parámetro  $\theta$  toma valores en el espacio  $\Theta$ .

Una extensión natural bayesiana no paramétrica de la densidad conjunta para  $(Y_{t-1}, Y_t)$  se daría en términos de un modelo mixto de dimensión infinita, en el sentido de [Lo \(1984\)](#) que tomaría la forma de

$$p(y_{t-1}, y_t) = \sum_{j=1}^{\infty} W_j K(y_{t-1}, y_t|\theta_j). \quad (2.2)$$

Entonces, la distribución bivariada se caracterizaría por los pesos  $\{W_j\}_{j=1}^{\infty}$  y los parámetros  $\{\theta_j\}_{j=1}^{\infty}$ , cuyos elementos comparten un espacio de parámetros de dimensión finita  $\Theta$ . En

consecuencia, la distribución marginal común inducida por (2.2) se representaría por

$$\begin{aligned} p(y_t) &= \sum_{j=1}^{\infty} W_j \int K(dy_{t-1}, y_t | \theta_j) \\ &= \sum_{j=1}^{\infty} W_j K(y_t | \theta_j). \end{aligned} \tag{2.3}$$

Entonces, una versión no paramétrica de la densidad de transición de primer orden tomaría la forma

$$\begin{aligned} p(y_t | y_{t-1}) &= \frac{p(y_{t-1}, y_t)}{p(y_{t-1})} \\ &= \frac{\sum_{j=1}^{\infty} W_j K(y_{t-1}, y_t | \theta_j)}{\sum_{l=1}^{\infty} W_l K(y_{t-1} | \theta_l)}, \end{aligned} \tag{2.4}$$

para cualquier  $t$ .

Antes de proceder, se destaca que el mecanismo de transición del modelo de referencia (2.4) parece –y en términos de flexibilidad extiende– el estimador tipo kernel para la dinámica de transición de un proceso Markov con estado-espacio general, lo que propuso originalmente Roussas (1969) que sirvió como modelo de referencia para la estimación no paramétrica de procesos de Markov en el enfoque frecuentista. A lo largo de años, estimadores tipo kernel para procesos de Markov recibieron considerable atención. Véase, por ejemplo, Athreya y Atuncar (1998), Hili (2001), Campos y Dorea (2005) y Lacour (2008).

Pero nótese que a pesar de la simplicidad de la expresión (2.4) y el hecho de ser una generalización flexible de la distribución de transición en comparación con los estimadores tradicionales tipo kernel, su tratabilidad práctica está limitada debido al problema de incluir una suma infinita en su denominador. Por lo tanto, la implementación computacional del modelo de referencia no es posible, en este momento. Sin embargo, en la siguiente sección, se deriva una generalización flexible que supera la dificultad del denominador involucrado en representaciones similares a (2.4). Los modelos de estacionariedad que se derivan, admiten representaciones de mezcla de dimensión infinita para la marginal y la de transición, simultáneamente. Lo anterior se logra al entender la distribución de transición  $p(y_t | y_{t-1})$  como mezcla, en el sentido de Pitt et al. (2002), más bien que un cociente, como se trató anteriormente. Se refiere a tal enfoque de construcción del modelo como muestreo de Gibbs.

### 2.3. Construcción de modelo con muestreo de Gibbs

El enfoque de construcción de modelo con muestreo de Gibbs se desarrolló originalmente por [Pitt et al. \(2002\)](#), y [Mena y Walker \(2005\)](#) se refirió a este enfoque en un marco más amplio. La construcción depende de la incorporación de medidas de probabilidad latentes como instrumento para inducir una estructura de dependencia y un comportamiento marginal deseado del modelo. A continuación se establece una noción. Sea  $Y$  una variable aleatoria observable que toma valores en el espacio de la muestra  $\mathcal{Y}$ , y sea  $f$  la densidad (o la función de probabilidad) para  $Y$ , la cual se asume que es aleatoria. La definición del modelo se completa con la incorporación de una medida de probabilidad  $\Pi$  para  $f$ , que se define sobre el espacio funcional de dimensión infinita  $\mathcal{F}$  que incluye a  $f$ . Correspondientemente, la construcción del modelo se deriva por la siguiente estructura:

$$y|f \sim f, \quad y \quad f \sim \Pi. \quad (2.5)$$

Como se asume a menudo, la medida de probabilidad  $\Pi$  se define tal que la relación  $\mathbb{E}_\Pi(f) = g_0$  se mantenga, para cualquier función de densidad,  $g_0$ . Por lo tanto, se observa simplemente que la densidad marginal para  $Y$  se representaría por  $g_0$ , es decir,

$$p(y_t) = \int f(y_t)\Pi(df) = \mathbb{E}_\Pi[f(y_t)] = g_0(y_t). \quad (2.6)$$

La idea clave de la construcción del modelo con muestreo de Gibbs consiste en expresar la densidad de transición desde el tiempo  $(t - 1)$  a  $t$  como una mezcla de la forma

$$p(y_t|y_{t-1}) = \int f(y_t)\Pi(df|y_{t-1}) = \mathbb{E}_\Pi(f(y_t)|y_{t-1}), \quad (2.7)$$

para la cual existen expresiones analíticas explícitas para casi todas las medidas de probabilidad  $\Pi$  conocidas en la literatura (asumiendo que  $Y_{t-1}$  y  $Y_t$  son condicionalmente independientes dada  $f$ ; véase, por ejemplo [Lijoi y Prünster, 2010](#)). También, bajo especificaciones adecuadas de  $\Pi$ , será posible encontrar una representación flexible para la dinámica de transición (2.7), como lo estudió [Mena y Walker \(2005\)](#).

Una desventaja potencial de la construcción del modelo con el muestreo de Gibbs es que la forma de densidad marginal, (2.6), es típicamente paramétrica. Por lo tanto, la posibilidad de definir una alternativa para el modelo de referencia descrito antes, que sea completamente no

paramétrica, parece ser inviable. Sin embargo, un enfoque general que evita esta desventaja, se desarrolla en la Sección 3. La idea consiste en extender la construcción del modelo base con muestreo de Gibbs en dos direcciones. Por un lado, se deberían explorar representaciones adecuadas de doble mezcla para  $\Pi$ , que permite expresar la densidad marginal como una mezcla de dimensión infinita. Por otro lado, se debería definir una representación adecuada conjunta observable y latente de la dinámica de transición del proceso que permitirá evitar la suma infinita incluida en el denominador de la distribución de transición del modelo de referencia. Por lo tanto, el mecanismo de transición del modelo se escribiría como una mezcla de dimensión infinita.

Antes de continuar, se nota que en el fondo del enfoque de la construcción del modelo con muestreo de Gibbs se ubica la especificación de la medida de probabilidad conjunta inducida por  $f$  y  $\Pi$ , por ejemplo,  $\mathbb{P}$ , que se presenta por

$$\mathbb{P}(dy, df) = f(y)\Pi(df)dy. \quad (2.8)$$

El soporte de  $\mathbb{P}$  es dado por el producto de espacios  $\mathcal{Y} \times \mathcal{F}$ . La noción de  $\mathbb{P}$  será fundamental para la construcción del modelo y se describirá en la Sección 3, cuando se derivarán mecanismos marginales y de transición deseados a partir de unas versiones estructuradas alternativas de (2.8).

Antes de presentar los detalles de la construcción del modelo extendido con muestreo de Gibbs, se establecerá una conexión notable con el modelo de referencia. Puede ser, como se mostrará en adelante, que el modelo de referencia se deriva como caso particular de modelos estacionarios con muestreo de Gibbs.

### 2.3.1. Derivación del modelo de referencia vía muestreo de Gibbs

El modelo de referencia se presenta como caso particular de la construcción de modelo con muestreo de Gibbs cuando  $K(y_{t-1}, y_t | \theta_j) = K(y_{t-1} | \theta_j)K(y_t | \theta_j)$  y se define la siguiente medida de probabilidad conjunta sobre el producto de espacio  $\mathcal{Y} \times \mathcal{J}$

$$\mathbb{P}(dy_t, J = d) = W_d K(dy_t | \theta_d), \quad (2.9)$$

donde  $\mathcal{Y}$  es el espacio de la muestra de  $Y$  y  $\mathcal{J} = \{1, 2, \dots\}$ . Dentro de la construcción, la variable latente  $J$  lleva el papel de una imputación latente o de una variable indexada latente

asociada con un modelo de mezcla. Por lo tanto, de (2.9) sigue que la medida condicional de probabilidad de  $J = d$ , dada  $y_{t-1}$ , sería

$$\mathbb{P}(J = d|y_{t-1}) \propto W_d K(y_{t-1}|\theta_d), \quad (2.10)$$

con la constante de proporcionalidad dada por  $\sum_{l=1}^{\infty} W_l K(y_{t-1}|\theta_l)$ . Entonces, al imputar (2.10) en la versión análoga de (2.7) da lugar al mecanismo alternativo de transición bayesiano no paramétrico tipo kernel

$$\begin{aligned} p(y_t|y_{t-1}) &= \sum_{d=1}^{\infty} \mathbb{P}(y_t|d) \mathbb{P}(d|y_{t-1}) \\ &= \frac{\sum_{d=1}^{\infty} W_d K(y_t|\theta_d) K(y_{t-1}|\theta_d)}{\sum_{l=1}^{\infty} W_l K(y_{t-1}|\theta_l)}. \end{aligned} \quad (2.11)$$

Por ende, al seguir dicha derivación, se observa fácilmente que la distribución estacionaria marginal asociada con (2.11) coincidiría con

$$p(y_t) = \sum_{l=1}^{\infty} W_l K(y_t|\theta_l), \quad (2.12)$$

para cualquier  $t = 1, 2, \dots$ , como es deseado. Procedemos con la construcción del modelo extendido.

### 3. Modelos dinámicos estacionarios extendidos de mezclas

La construcción del modelo extendido con muestreo de Gibbs supera la limitación práctica que tiene el modelo de referencia, mientras que representar mecanismos marginales y de transición del modelo vía mezclas de dimensión infinita. La construcción se basa en la representación extendida de la medida de probabilidad conjunta  $\mathbb{P}$ , como en (2.8), pero esta vez se incorpora una medida de probabilidad latente adicional que llevará un papel fundamental en la descripción del mecanismo de transición de las variables observables y latentes. Una derivación general basada en mezclas se describe en el Apéndice A.

### 3.1. Distribución invariante

De manera similar como en el modelo de referencia introducido anteriormente, se buscan modelos estacionarios que admiten una densidad marginal (invariante) como mezcla de dimensión infinita. La forma explícita de tal distribución se representará a continuación, después de establecer algunas notaciones. Sea  $Y$  una variable aleatoria observable que toma valores en el espacio medible (espacio poláco)  $(\mathcal{Y}, \mathcal{B}_Y)$ . Los modelos estacionarios de interés admiten una densidad marginal (invariante) (o función de probabilidad, si es necesario) con la forma

$$p(y) = \sum_{j=1}^{\infty} W_j \int_{\mathcal{L}} K(y|\lambda) G_0(d\lambda; \theta_j), \quad (3.1)$$

donde  $\{W_j\}_{j=1}^{\infty}$  es una secuencia de pesos de probabilidad, tal que  $0 \leq W_j \leq 1$ , para cualquier  $j$ , y  $\sum_{j=1}^{\infty} W_j = 1$ ;  $\lambda$  es una variable de mezclas que toma valores en el espacio medible  $(\mathcal{L}, \mathcal{B}_{\mathcal{L}})$ ;  $K(y|\lambda)$  es un kernel paramétrico definido sobre  $\mathcal{Y} \times \mathcal{L}$ ;  $G_0(\cdot|\theta_j)$  es una función de distribución dada definida sobre el espacio medible  $(\mathcal{L}, \mathcal{B}_{\mathcal{L}})$ , que está indexada por el parámetro  $\theta_j$ , para cada  $j$ ; y  $\{\theta_j\}_{j=1}^{\infty}$  es una secuencia de parámetros que toma valores en un espacio común dado  $\Theta$  (de dimensión finita). También, se asume que las distribuciones de probabilidad  $G_0(\cdot|\theta_j)$  tienen exactamente la misma forma funcional a través de  $j$ 's, difieren solamente en términos de los parámetros  $\theta_j$ 's. Se asumirá que  $G_0$  es una medida de probabilidad difusa sobre  $(\mathcal{L}, \mathcal{B}_{\mathcal{L}})$  con la función de densidad  $g_0$ .

Como se aplica típicamente para modelos de mezclas, el alcance de (3.1) se puede extender a través de introducir una variable latente de imputación,  $J$ , que indica de cuál componente de la mezcla viene la variable aleatoria observable  $Y$ . En este caso, la versión aumentada a (3.1), después de incluir  $J = d$  como variable latente, se presenta de la siguiente forma

$$p(y, d) = W_d \int_{\mathcal{L}} K(y|\lambda) G_0(d\lambda; \theta_d), \quad (3.2)$$

para cualquier  $d \in \mathcal{J} = \{1, 2, \dots\}$ .

### 3.2. Definición estocástica

A partir de la definición aumentada derivada en (3.2), también se puede suponer que existe una función de distribución aleatoria,  $F$ , que toma valores en  $\mathcal{F}$ , que se define en el espacio me-

dible  $(\mathcal{L}, \mathcal{B}_{\mathcal{L}})$ . Además se supone que para cada  $j$  existe una medida de probabilidad indexada por  $\theta_j$ , por ejemplo  $\Pi(\cdot; \theta_j)$ , definida en el espacio  $(\mathcal{F}, \mathcal{B}_{\mathcal{F}})$  tal que  $G_0(\cdot|\theta_j) = \mathbb{E}_{\Pi}\{F(\cdot)|\theta_j\}$ . Como antes, se supone que  $G_0$  tiene la misma forma funcional a través de  $j$ , difiriendo solamente en términos de  $\theta_j$ . La medida de probabilidad  $\Pi(\cdot; \theta_j)$  tendrá la misma forma estructural, pero esta vez solamente se distingue en términos de su distribución de probabilidad base correspondiente  $G_0(\cdot|\theta_j)$ 's. La razón de incluir  $F$  explícitamente en la definición estocástica se explicará más adelante en el documento. Por el momento, anticipemos que llevará un papel fundamental en la especificación de la dinámica de transición del modelo, ya que el mecanismo de transición se definirá por  $F$  en sí.

Por lo tanto, mediante la expansión de alcance de (3.2), después de incluir  $F$  y  $\lambda$  como componente adicional, sería posible definir una estructura estocástica extendida,  $\mathbb{P}$ , que se define por el producto de espacio aumentado  $\mathcal{Y} \times \mathcal{L} \times \mathcal{F} \times \mathcal{J}$ , dado por

$$\mathbb{P}(dy, d\lambda, dF, J = d) = K(dy|\lambda)F(d\lambda)\Pi(dF|G_0(; \theta_d))W_d. \quad (3.3)$$

Esta estructura estocástica puede tener distintas interpretaciones o derivaciones. En opinión de los autores, entre las posibles alternativas, la próxima representación jerárquica análoga de (2.5) sería más representativa para el presente contexto:

$$\begin{aligned} Y|\lambda &\sim K(y|\lambda), & \lambda|F &\sim F(\lambda), \\ F &\sim \tilde{\Pi}(\cdot) = \sum_{j=1}^{\infty} W_j \Pi(\cdot|G_0(; \theta_j)), \end{aligned}$$

donde se nota explícitamente que la medida de probabilidad elegida  $\tilde{\Pi}$  se representa como mezcla doble no paramétrica. Véase [Teh et al. \(2006\)](#) y [Rodríguez et al. \(2008\)](#) para definiciones relacionadas estructuradas bayesianas no paramétricas.

La medida de probabilidad aumentada servirá como la definición estocástica para la construcción del modelo. La siguiente sección desarrolla la dinámica de transición del modelo inducida por (3.3). El marco inicial para especificar la dinámica de transición es la definición aumentada en la cual la dinámica de transición será dada en términos del triplete de variables observables y latentes,  $(y, \lambda, d)$ , y la estructura de dependencia se define por  $F$  a través de su medida de probabilidad asociada  $\Pi$ .



### 3.3. Dinámica de transición

Para definir la dinámica de transición, se considera el estado de tiempo  $(t - 1)$  para el triplete de variables observables y latentes  $(y_{t-1}, \lambda_{t-1}, d_{t-1})$  que, dada  $F$ , tendrá la medida de probabilidad conjunta

$$\mathbb{P}(dy_{t-1}, d\lambda_{t-1}, J_{t-1} = d_{t-1} | F) = W_{d_{t-1}} K(dy_{t-1} | \lambda_{t-1}) F(d\lambda_{t-1}). \quad (3.4)$$

Correspondientemente, el mecanismo de transición entre  $(t - 1)$  y  $t$  sería caracterizado por la regla de transición de probabilidad

$$\begin{aligned} \mathbb{P}(dy_t, d\lambda_t, d_t | y_{t-1}, \lambda_{t-1}, d_{t-1}) &= \int_{\mathcal{F}} \mathbb{P}(dy_t, d\lambda_t, d_t | F) \mathbb{P}(dF | y_{t-1}, \lambda_{t-1}, d_{t-1}) \\ &= W_{d_t} K(dy_t | \lambda_t) G(d\lambda_t | \lambda_{t-1}, d_{t-1}), \end{aligned} \quad (3.5)$$

con

$$\begin{aligned} \mathbb{P}(dF | y_{t-1}, \lambda_{t-1}, d_{t-1}) &= \Pi(dF | \lambda_{t-1}, d_{t-1}), \\ G(d\lambda_t | \lambda_{t-1}, d_{t-1}) &= \mathbb{E}_{\Pi}\{F(d\lambda_t) | \lambda_{t-1}, d_{t-1}\}, \end{aligned} \quad (3.6)$$

donde  $\Pi(dF | \lambda_{t-1}, d_{t-1})$  representa la medida condicional de probabilidad de  $F$  dada  $(\lambda_{t-1}, d_{t-1})$ , mientras que  $G(d\lambda_t | \lambda_{t-1}, d_{t-1})$  representa la regla predictiva de un paso bajo  $\Pi$  para  $\lambda_t$  condicional a  $(\lambda_{t-1}, d_{t-1})$ .

#### 3.3.1. Expresiones explícitas de las reglas condicionales y predictivas de un paso para $F$

Expresiones explícitas para la medida condicional de probabilidad para  $F$  dadas  $(\lambda_{t-1}, d_{t-1})$  se pueden obtener de acuerdo con los valores del índice latente  $d_{t-1}$  y los valores de la variable latente indexada  $J_t$ , es decir

$$\mathbb{P}(dF | \lambda_{t-1}, d_{t-1}) \propto \begin{cases} \Pi(dF | G_0(; \theta_{d_{t-1}}), \lambda_{t-1}) & , \quad j = d_{t-1}, \\ \Pi(dF | G_0(; \theta_j)) & , \quad j \neq d_{t-1}, \end{cases} \quad (3.7)$$

donde la constante de proporcionalidad se define en términos de la secuencia  $\{W_j\}_{j=1}^{\infty}$  y los demás componentes involucrados en el modelo. Nótese que la dependencia de  $F$  con respecto

a  $\lambda_{t-1}$  se volverá explícita solamente en el caso  $d_t = d_{t-1}$ . De otra manera,  $\Pi(dF|\lambda_{t-1}, d_{t-1})$  se reducirá a la medida de probabilidad marginal  $\Pi$  dado sólo el valor del índice  $d_{t-1}$ .

De acuerdo con lo anterior, la regla de transición de un paso para  $\lambda_t$ 's se dará por

$$G(d\lambda_t|\lambda_{t-1}, d_{t-1}) \propto \begin{cases} G_1(d\lambda_t|\lambda_{t-1}, d_{t-1}; \theta_{d_{t-1}}) & , \quad j = d_{t-1}, \\ G_0(d\lambda_t; \theta_j) & , \quad j \neq d_{t-1}, \end{cases} \quad (3.8)$$

donde  $G_1(d\lambda_t|\lambda_{t-1}, d_{t-1}; \theta_{d_{t-1}}) = \mathbb{E}_\Pi\{F(d\lambda_t)|\lambda_{t-1}, d_{t-1}; \theta_{d_{t-1}}\}$  corresponde a la regla de transición de un paso para  $\lambda_t$  bajo  $\Pi$ , condicional al par  $(\lambda_{t-1}, d_{t-1})$ . Cabe mencionar que la regla de transición se puede calcular explícitamente para la mayoría de las medidas de probabilidad  $\Pi$  conocidas en la literatura, incluso en el caso  $d_t = d_{t-1}$  (véase, por ejemplo [Lijoi y Prünster, 2010](#)).

### 3.3.2. Forma reducida del mecanismo de transición

En resumen, se puede decir que la dinámica de la secuencia  $\{(y_t, \lambda_t, d_t)\}_{t=1}^\infty$  se caracteriza por la regla de transición:

$$\mathbb{P}(dy_t, d\lambda_t, d_t|y_{t-1}, \lambda_{t-1}, d_{t-1}) \propto \begin{cases} W_{d_{t-1}} K(dy_t|\lambda_t) G_1(d\lambda_t|\lambda_{t-1}, d_{t-1}; \theta_{d_{t-1}}) & , \quad d_t = d_{t-1}, \\ W_{d_t} K(dy_t|\lambda_t) G_0(d\lambda_t; \theta_{d_t}) & , \quad d_t \neq d_{t-1}, \end{cases} \quad (3.9)$$

con todos componentes incluidos dados como antes.

Una característica distinta de la dinámica de transición de este modelo es que teóricamente sea posible recuperar la dinámica de transición de las observables,  $\{y_t\}$ , requiriendo solamente la computación de

$$\begin{aligned} \mathbb{P}(dy_t|y_{t-1}) &= \int \mathbb{P}(dy_t|F) \mathbb{P}(dF|y_{t-1}) \\ &= \int \int K(dy_t|\lambda_t) F(d\lambda_t) \mathbb{P}(dF|y_{t-1}), \end{aligned} \quad (3.10)$$

donde

$$\mathbb{P}(dF|y_{t-1}) \propto \begin{cases} W_{d_{t-1}} \int \Pi(dF|G_0(\cdot; \theta_{d_{t-1}}), \lambda_{t-1}) G_0(d\lambda_{t-1}|y_{t-1}; \theta_{d_{t-1}}) & , \quad j = d_{t-1}, \\ W_j \Pi(dF|G_0(\cdot; \theta_j)) & , \quad j \neq d_{t-1}, \end{cases}$$

con  $G_0(\lambda_{t-1}|y_{t-1}; \theta_{d_{t-1}}) \propto K(y_{t-1}|\lambda_{t-1})G_0(\lambda_{t-1}|\theta_{d_{t-1}})$ .

Es comprobable que el par (3.1) y (3.10) satisfacen la ecuación balance de un proceso Markov. Por lo tanto, ambas la distribución marginal invariante y la de transición de primer orden definen una secuencia estacionaria de primer orden  $\{Y_t\}_{t=1}^{\infty}$ .

Aunque es posible encontrar una expresión analítica para la regla de transición (3.10), para fines de inferencia, será suficiente abrir la regla de transición para variables observables y latentes simultáneamente. De esta manera, se pueden implementar algunos procedimientos estadísticos desarrollados para hacer inferencias sobre modelos estado-espacio.

## 4. Un modelo estacionario de mezcla de escalas

En esta sección se desarrolla un modelo dinámico de mezcla de escalas como caso particular de la construcción extendida introducida en la Sección 3. Los elementos del modelo estarían dados por el kernel paramétrico, que se asume siendo Gaussiano, es decir,  $K(y|\lambda) = N(y|0, \lambda)$ , donde  $\lambda$  es un parámetro de escala. Por lo tanto, el estado-espacio para  $\lambda$  es  $\mathbb{R}_+$ . Como la medida base de la probabilidad de  $\Pi$  se considera la distribución Weibull con parámetro de forma,  $\theta > 0$ , i.e.  $G_0(\lambda|\theta) = \text{We}(\lambda|\theta)$ , con  $\theta > 0$ , y función de densidad  $g_0(\lambda|\theta) \propto \theta\lambda^{\theta-1}e^{-\lambda^\theta}$ . La elección de  $G_0$  como una distribución Weibull resulta de la elección de  $\Pi$ , la medida de probabilidad para  $F$ , la cual se toma siendo un proceso beta-Stacy (véase, Walker y Muliere, 1997). El proceso beta-Stacy se eligió debido a la estructura flexible de su distribución condicional esperada de un paso, e incluirá expresiones explícitas de la función hazard,  $h_0$ , asociada con  $G_0$ . Escoger una familia paramétrica para  $G_0$  con funciones hazard explícitas es relevante para la especificación del modelo. A continuación, se resumen detalles con respecto al proceso beta-Stacy.

### 4.1. El proceso beta-Stacy

El proceso beta-Stacy (BS) se introdujo por Walker y Muliere (1997) como un proceso particular neutral-hacia-la-derecha (Doksum, 1974). Una característica atractiva del proceso BS, relevante para la construcción del modelo, se debe a la flexibilidad que se mantiene después de calcular distribuciones esperadas condicionales de un paso, es decir,  $G_1$  de acuerdo

con la noción usada. Resumimos la definición y las características relevantes del proceso beta-Stacy.

Una distribución aleatoria de probabilidad  $F$ , con soporte en la línea real positiva, se denomina proceso beta-stacy si se puede escribir como transformación adecuada de un proceso creciente de Lévy que tiene la siguiente forma

$$F(\lambda) \stackrel{d}{=} 1 - \exp\{-Z(\lambda)\}, \quad (4.1)$$

donde  $\{Z(\lambda) : \lambda \geq 0\}$  es un proceso con incrementos independientes no decrecientes, tal que: i)  $Z(\lambda)$  es no decreciente casi seguramente (c. s.), ii)  $Z(\lambda)$  es continua por la derecha c. s., iii)  $Z(0) = 0$  c. s., y iv)  $\lim_{\lambda \rightarrow \infty} Z(\lambda) = +\infty$  c. s.. El proceso beta-Stacy resulta cuando  $Z(\lambda)$  se toma como proceso log-beta caracterizado por la medida de Lévy no homogénea

$$\nu(du, ds) = \frac{1}{1 - e^{-s}} e^{-s\beta(u)} ds \alpha(du), \quad (4.2)$$

para  $u > 0$  y  $s \geq 0$ , con  $\alpha$  siendo una medida continua finita en  $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$  y  $\beta$  una función no negativa en la línea real positiva, tal que  $\lim_{\lambda \rightarrow +\infty} \int_0^\lambda \frac{\alpha(du)}{\beta(u)} = +\infty$ .

La parametrización particular considerada para el proceso BS que acomoda  $G_0$  (la distribución Weibull) como medida no condicional de probabilidad esperada, es

$$\alpha(du) = \beta(u)\theta u^{\theta-1} du, \quad (4.3)$$

con  $\theta > 0$  y  $\beta$  siendo un parámetro funcional positivo con soporte en la línea real positiva.

El proceso beta-Stacy es adecuado para la construcción del modelo debido a varias características. Su conjugación estructural garantiza que si  $F \sim \text{BS}(\alpha, \beta)$  y  $\lambda|F \sim F$ , entonces la medida condicional de probabilidad para  $F$  dada  $\lambda'$  (un valor exacto de  $\lambda$ ) también sería un proceso beta-Stacy caracterizado por un proceso log-beta actualizado,  $Z^*$ , con un cambio aleatorio de tamaño  $S_1$  en  $\lambda'$ , tal que

$$1 - \exp\{-S_{\lambda'}\}|\lambda' \sim \text{Be}(1, \beta(\lambda')), \quad (4.4)$$

y un componente continuo con medida Lévy actualizada  $\nu^*$  dada por

$$\nu^*(du, ds) = \frac{1}{1 - e^{-s}} e^{-s[\beta(u) + \mathbf{1}(\lambda' \geq u)]} ds \alpha(du). \quad (4.5)$$

Véase Corolario 2 en [Walker y Muliere \(1997\)](#) para obtener más detalles.

Adicionalmente, debido al hecho de que el proceso esté caracterizado por una medida de Lévy no homogénea y por la parametrización adoptada, el proceso beta-Stacy es más atractivo ya que admite distribuciones esperadas condicionales de un paso más flexibles las cuales, condicional a una observación exacta  $\lambda'$ , se pueden escribir como

$$\begin{aligned} G_1(\lambda|\lambda') &= \mathbb{E}_{\text{BS}}[F(\lambda)|\lambda'] = 1 - \mathbb{E}[e^{-Z^*(\lambda)}] \\ &= 1 - \exp \left\{ - \int_0^\lambda [\mu(u) \mathbf{1}_{(0,\lambda']}(u) + \mathbf{1}_{(\lambda',\infty)}(u)] h_0(u) du \right\} \\ &\quad \times [\mathbf{1}_{(0,\lambda')}(λ) + \mu(\lambda') \mathbf{1}_{[\lambda',\infty)}(\lambda)], \end{aligned} \tag{4.6}$$

con la última expresión derivada después de considerar la reparametrización  $\mu(u) = \frac{\beta(u)}{\beta(u)+1}$ . A primera vista, la forma funcional para  $G_1$  parece demasiado compleja. Pero, de hecho, es tratable, y representa una distribución tipo mixto para  $\lambda$  que se caracteriza por : i) una forma funcional moderada (es decir, reescalada uniformemente) por  $\mu$  en la región  $(0, \lambda')$ , ii) una masa de probabilidad en  $\lambda'$ , con probabilidad dependiendo de  $\mu$ , y iii) un ajuste de comportamiento de cola derecha con  $G_0$  en la región  $(\lambda', \infty)$ . Con el fin de hacer (4.6) computacional y analíticamente manejable, es conveniente considerar una parametrización adecuada  $\mu$ , la cual se describe a continuación.

## 4.2. Una parametrización adecuada para $\mu$

Sin perder flexibilidad, además podemos asumir que el parámetro funcional  $\mu$  pertenece a una clase de funciones en piezas y continuas por la derecha con valor  $(0, 1)$

$$\mu(u) = \sum_{k=0}^{\infty} \mu_k \mathbf{1}_{E_k}(u), \tag{4.7}$$

para  $u \geq 0$ , donde  $\{\mu_k\}_{k=0}^{\infty}$  es una secuencia con valor  $(0, 1)$  y  $\{E_k\}_{k=0}^{\infty}$  es una partición monotónica de la línea positiva real. Por partición monotónica se refiere a una partición inducida por una secuencia monotónica creciente  $0 = \xi_0 < \xi_1 < \dots < \xi_k < \dots$ , con  $E_k = (\xi_k, \xi_{k+1}]$ , para  $k \geq 0$ . Previamente, se consideraron parametrizaciones similares de los parámetros funcionales en [Arjas y Gasbarra \(1994\)](#) y [Nieto-Barajas y Walker \(2002\)](#), en contextos distintos.

Adicionalmente, se introduce otra variable indicador latente,  $z$ , con la cual se puede

escribir la función de densidad/probabilidad  $g_1$ , derivada de  $G_1$ , como la siguiente mezcla

$$g_1(\lambda|\lambda') = \sum_{z=-1}^1 g_{1,z}(\lambda|\lambda')p(z|\lambda'), \quad (4.8)$$

donde

$$z|\lambda' \sim \begin{cases} -1 & \text{con } p(z = -1|\lambda') = P\{(0, \lambda')|\lambda'\} \\ 0 & \text{con } p(z = 0|\lambda') = P\{\lambda'\} \\ 1 & \text{con } p(z = 1|\lambda') = P\{(\lambda', \infty)|\lambda'\}. \end{cases} \quad (4.9)$$

El Apéndice B presenta expresiones analíticas explícitas para los componentes relevantes involucrados en (4.8).

### 4.3. La función de verosimilitud y esquema de aumento de datos

Dada una colección de observaciones  $\mathbf{y} = \{y_t\}_{t=1}^T$ , la función de verosimilitud para el modelo dinámico estacionario de mezcla de escalas incluiría componentes en función de si los dos índices latentes consecutivos  $d_t$  y  $d_{t-1}$  son iguales o diferentes. Componentes específicos de la verosimilitud asociados con los casos  $d_t = d_{t-1}$  se manejan difícilmente de manera analítica. Por lo tanto, se requiere un esquema de aumento de datos. En particular, se incorporan las variables indicadoras incluidas en la especificación  $g_1$  como variables latentes adicionales. Así, la verosimilitud aumentada después de incorporar  $z_t$ 's se escribiría como

$$lik(Q, \mu, \boldsymbol{\lambda}, \mathbf{z}, \mathbf{d}; \mathbf{y}) \propto \left\{ \prod_{\{t:d_t=d_{t-1}\}} W_{d_t} \text{N}(y_t|0, \lambda_t) g_{1,z_t}(\lambda|\lambda_{t-1}; \theta_{d_t}, \mu) p(z_t|\lambda_{t-1}; \theta_{d_t}, \mu) \right\} \cdot \left\{ \prod_{\{t:d_t \neq d_{t-1}\}} W_{d_t} \text{N}(y_t|0, \lambda_t) \text{We}(\lambda_t|\theta_{d_t}) \right\}, \quad (4.10)$$

donde  $Q = \sum_{j=1}^{\infty} W_j \delta_{\theta_j}$  que se caracteriza por una secuencia conjunta  $\{(W_j, \theta_j)\}_{j=1}^{\infty}$ , y  $\mu$  es una función  $(0, 1)$  en piezas caracterizada por la secuencia conjunta  $\{(\xi_k, \mu_k)\}_{k=1}^{\infty}$ . Adicionalmente,  $\boldsymbol{\lambda} = (\lambda_t)_{t=1}^{\infty}$  corresponde a un proceso de escala latente, y  $\mathbf{d}$  corresponde a una secuencia latente de variables indexadas. Como ya se mencionó, las expresiones explícitas para  $p(z_t|\lambda_{t-1}; \theta_{d_t}, \mu)$  y  $g_{1,z_t}(\lambda|\lambda_{t-1}; \theta_{d_t}, \mu)$  se encuentran en el Apéndice B. Lo que se debe destacar acerca de la verosimilitud aumentada (4.10), con respecto al componente latente

$\mathbf{z} = (z_t)$ , es que la dimensión de  $\mathbf{z}$  en la implementación computacional del modelo cambia con las iteraciones de acuerdo con los casos en los cuales dos variables latentes indexadas coinciden. Procedemos con una descripción del análisis de la prior-a-posterior.

## 5. Análisis posterior

### 5.1. Truncamiento estocástico vía variables latentes

Un paso necesario para implementar el modelo estacionario de mezcla de escalas consiste en tratar con la suma infinita todavía incluida en la verosimilitud extendida (4.10) para cada  $d_t$ . De manera similar como en el enfoque utilizado por Walker (2007) para modelos de mezclas de dimensión infinita, se introduce una variable latente de truncamiento para restringir el rango de cada  $d_t$  a un conjunto finito. Pero, de manera diferente a Walker (2007), no utilizamos el muestreo de la rebanada (slice sampler). En lugar de eso, truncamos directamente el rango de cada  $d_t$  vía una variable latente,  $U_t$ .

Al abreviar la noción, el componente de verosimilitud (4.10), incluyendo cada  $d_t$ , se puede escribir como

$$f(\lambda_t, d_t) = W_{d_t} f(\lambda_t | \theta_{d_t}), \quad (5.1)$$

para  $d_t = 1, 2, \dots$ . La forma específica de  $f(\lambda_t | \theta_{d_t})$  en nuestro contexto dependerá de todas las combinaciones posibles de los conjuntos de índices  $(d_{t-1}, d_{t+1})$ , en conjunción con  $d_t$ . [Nótese que para un modelo de mezclas tradicionales,  $f(\lambda_t | \theta_{d_t})$  se expresaría solamente por el componente kernel de la mezcla.] El truncamiento estocástico se lleva a cabo en (4.10), en cada tiempo  $t$ , a través de (5.1) mediante la incorporación de la variable latente de truncamiento,  $U_t$ , tal que su distribución condicional, dado  $d_t$ , sería una variable aleatoria geométrica desplazada, es decir

$$U_t | d_t \sim \phi(1 - \phi)^{u-d_t} \mathbf{1}(u \geq d_t), \quad (5.2)$$

con  $0 < \phi < 1$ . Se puede observar que la distribución marginal para  $U_t$  será geométrica con parámetros  $\phi$ . Consecuentemente, la versión aumentada de (5.1), después de introducir  $u_t$ , será

$$f(\lambda_t, d_t, u_t) = W_{d_t} f(\lambda_t | \theta_{d_t}) \phi(1 - \phi)^{u_t - d_t} \mathbf{1}(u_t \geq d_t). \quad (5.3)$$

También es verificable que el modelo original (5.1) se puede recuperar de (5.3) después de integrar  $U_t$ .

Por lo tanto, el truncamiento estocástico se implementa en el muestreo de Gibbs de la posterior donde se requiere el muestreo de las distribuciones condicionales para cada  $d_t$  dado  $u_t$ , y vice-versa. Detalles acerca de la verosimilitud aumentada, después de incorporar la secuencia de otras variables latentes de truncamiento  $\mathbf{u} = \{u_t\}_{t=1}^T$ , y la implementación del muestreo Gibbs de la posterior se encuentran en los apéndices.

## 5.2. Especificación de la inicial

En la presente sección se describen dos prototipos de distribuciones iniciales bayesianas no paramétricas para los dos parámetros incluidos del modelo,  $\mu$  y  $Q$ . Además, se supone que  $\mu$  y  $Q$  son independientes.

### 5.2.1. Prior para $\mu$

Recuérdese que  $\mu$ , o la secuencia  $\{(\xi_k, \mu_k)\}_{k=1}^\infty$ , se pueden entender como una trayectoria realizada de un proceso con saltos, el cual se asumió impulsado por un proceso marcado Poisson homogéneo, cuya ley de probabilidad permite la descomposición.  $\mathbb{P}\{(\xi_k, \mu_k)\} = \Pi\{(\xi_k)\} \cdot \Pi\{(\mu_k)|(\xi_k)\}$ , con  $\Pi\{(\xi_k)\}$  siendo la ley de probabilidad de la parte del proceso de Poisson homogéneo con intensidad  $\nu_\xi > 0$ ; y  $\Pi\{(\mu_k)|(\xi_k)\}$  siendo la ley de probabilidad de la parte marcada que se especifica en términos del siguiente sistema de probabilidades condicionales.

$$p(\mu_0|\xi_0) = \text{Be}(\mu_0|a_\mu, b_\mu), \quad (5.4)$$

$$p(\mu_k|\xi_k, \dots, \xi_0, \mu_{k-1}, \dots, \mu_0) = \text{Be}(\mu_k|a_\mu + k, c_\mu), \quad (5.5)$$

para  $k = 1, 2, \dots$ . De acuerdo con lo anterior, la media de la secuencia de marcas aleatorias incrementará. En esta especificación de la inicial, el parámetro  $c_\mu$  regula qué tan rápido se puede esperar que la función  $\mu$  se acercará a 1. Mientras menor sea  $c_\mu$ , más lenta será la función  $\mu$  en aumentar en la media. Una especificación similar de la inicial en el espacio de funciones aleatorias constantes en piezas se utilizó por [Arjas y Gasbarra \(1994\)](#) para estimar funciones hazard aleatorias en el análisis de supervivencia.



### 5.2.2. Prior para $Q$

Recuérdese que  $Q$  se caracteriza por la secuencia  $\{(W_j, \theta_j)\}_{j=1}^{\infty}$ . Por lo tanto, de acuerdo con la representación discreta de  $Q$ , es adecuado suponer que  $\{(W_j, \theta_j)\}_{j=1}^{\infty}$  es motivada por un proceso Poisson-Dirichlet con dos parámetros (Pitman y Yor, 1997) que se puede escribir en la forma stick-breaking (véase, por ejemplo Ishwaran y James, 2001) como

$$\begin{aligned} W_1 = V_1, \quad \text{y} \quad W_j = V_j \prod_{l=1}^{j-1} (1 - V_l), \quad \text{for } j \geq 2, \\ \theta_j \stackrel{\text{i.i.d.}}{\sim} Q_0(\theta) \quad \text{y} \quad V_j \stackrel{\text{ind}}{\sim} \text{Be}(v_j | 1 - \alpha_{\text{PDP}}, j\alpha_{\text{PDP}}), \end{aligned} \quad (5.6)$$

para  $j \in \mathcal{J}$ , donde  $Q_0$  (la medida base de probabilidad de prior) es elegida para ser una distribución gamma con densidad  $q_0(\theta) \propto \theta^{a_{Q_0}-1} \exp\{-\theta b_{Q_0}\}$ , para algunos parámetros fijos positivos  $a_{Q_0}$  y  $b_{Q_0}$ . El parámetro  $\alpha_{\text{PDP}}$  también se fijaría para cierto valor en el intervalo  $[0, 1)$ .

### 5.3. Muestreo de Gibbs posterior

Implementar un esquema de muestreo posterior para el modelo todavía requiere un tratamiento adicional para manejar el soporte contable de las variables latentes indexadas,  $(d_t)$ , y la suma infinita involucrada en el denominador de la distribución condicional completa. Podemos resolver este problema al introducir otras variables latentes de truncamiento. Detalles acerca del esquema de truncamiento y la derivación de las distribuciones condicionales completas se encuentran en el Apéndice C.

El muestreo de la posterior se puede llevar a cabo vía el muestreo de Gibbs, con algunos pasos de muestreo trans-dimensionales. El muestreo se aplica para las siguientes distribuciones condicionales completas:

- $\mathbf{u} | \dots$  que depende solamente de  $\mathbf{d}$
- $\mathbf{d} | \dots$  que está determinada por las condicionales completas particulares para

$$d_t | W_{d_t}, \theta_{d_t}, \lambda_t, \lambda_{t-1}, \mu, u_t, z_t$$

- $\mu | \dots$  que incluirá un esquema adecuado de muestreo trans-dimensional

- $Q | \dots$  que sigue el esquema de muestreo de Gibbs en bloque
- $z | \dots$  que solamente aparece en los casos donde  $d_t = d_{t-1}$
- $\lambda | \dots$  que será sensitiva a combinaciones distintas inducidas por  $(d'_{t-1}, d'_t, d'_{t+1})$ .

Detalles sobre la implementación del muestreo de la posterior se encuentran en el Apéndice [C](#).

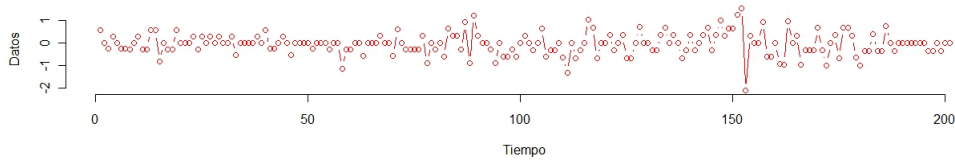
## 6. Resultados empíricos

El modelo de mezcla de escalas se ilustra en la práctica mediante un análisis de un segmento corto de log-retornos semanales de tipo de cambio euro-dólar E.U.A. que consiste en 200 observaciones<sup>1</sup>. Los datos de serie de tiempo se muestran en la Figura [1](#), en el panel (a). Ninguna especificidad sobre el periodo de tiempo bajo estudio es relevante en este punto, ya que nuestro objetivo en esta sección es mostrar el desempeño del modelo y el esquema del muestreo de la posterior. Se puede ver que los datos de serie de tiempo exhiben un periodo inicial de estabilidad seguido por fluctuaciones severas hacia el final del periodo de observación que se puede asociar con periodos de alta volatilidad. Todos los resultados producidos y reportados a continuación se obtuvieron vía el muestreo de Gibbs posterior con 15,000 iteraciones y considerando un periodo calentamiento de 10,000 iteraciones.

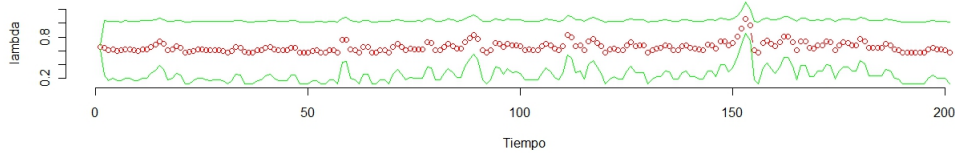
El modelo se parametriza por la distribución aleatoria  $Q$  que se supone ser impulsada por el proceso de Poisson-Dirichlet de dos parámetros, y una función de valor  $(0, 1)$  en piezas  $\mu$ , que es impulsada por un proceso homogéneo marcado de Poisson creciente en la media. Entonces, consideremos más detalles acerca de la intuición respecto a la especificación de la inicial del proceso Poisson-Dirichlet de dos parámetros, se puede ver que pertenece a la clase de modelos de especies de muestreo (species sampling models)<sup>2</sup>. Como tal modelo, tiene asociado un esquema de agrupamiento de la posterior que se determina por un parámetro  $0 \leq \alpha_{\text{PDP}} < 1$ , lo que se puede utilizar para afinar el esquema de agrupamiento y dar más

<sup>1</sup>La fuente de los datos es el Banco Central Europeo, URL: <http://www.ecb.int>.

<sup>2</sup>Modelos de especies de muestreo son funciones de distribución aleatoria discreta cuyas locaciones y pesos aleatorios son independientes. Véase [Pitman \(1996\)](#) y [Lijoi y Prünster \(2010\)](#).



(a) Datos



(b) Procesos latentes  $\{\lambda_t\}$

Figura 1: Serie de datos de log-retornos semanales de tipo de cambio euro-dólar E.U.A. y estimación de la posterior para el proceso latente (media de la posterior – línea de puntos, intervalo de confianza al 95 % – línea solida).

probabilidad para observar un nuevo agrupamiento cuando llega una nueva observación. En nuestro caso, se fija un parámetro en 0,5. El parámetro adicional corresponde a la distribución base gamma,  $Q_0$ , con parámetros de forma y escala dados por  $a_{Q_0}$  y  $b_{Q_0}$ , respectivamente. Recuerdese que las derivaciones de la distribución base determinan las distribuciones de las mezclas incluidas en la densidad marginal de mezclas (3.1). Se fija  $a_{Q_0} = 2$  y  $b_{Q_0} = 3,3$  para reflejar la información inicial con respecto a la dispersión marginal de los datos. Relacionado con el esquema de muestreo de la posterior definido alrededor  $Q$  es el parámetro de muestreo  $\phi$ , que básicamente sirve para fijar el número esperado de componentes en el modelo de mezcla que se actualizará en cada iteración. Este parámetro se pone en  $\phi = 0,18$ , que dará el algoritmo de la posibilidad de explorar hasta 5 componentes adicionales en cada iteración más allá del estado actual. El esquema del agrupamiento se puede visualizar a través del número y de la frecuencia de los agrupamientos resumidos en cada iteración a través del vector de pesos de probabilidad ( $W_j$ ). En la Figura 2 se muestran algunos de las trayectorias muestrales MCMC para ( $W_j$ ). Se puede observar que se identifican alrededor de diez agrupamientos en los datos en interacciones distintas, con la probabilidad de membresía determinada por la

magnitud de cada  $W_j$ . También, se puede observar que se pueden identificar más agrupamientos, pero con baja probabilidad. Estos agrupamientos tienen la posibilidad de capturar datos futuros y por ende, tienen un peso mayor o el peso se reduce cuando los datos nuevos parecen más a los datos ya observados.

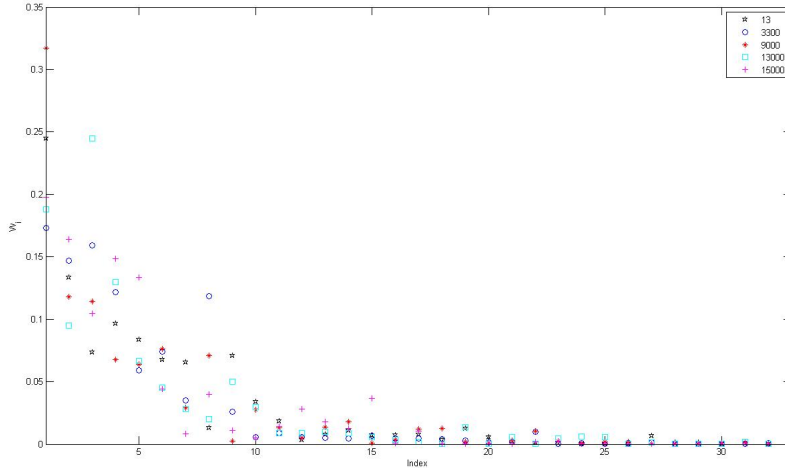


Figura 2: Algunas trayectorias muestrales de  $(W_j)$  para el modelo estacionario de mezcla de escalas presentados por iteraciones distintas del esquema de muestreo de la posterior.

Con respecto al procesos homogéneo marcado de Poisson, su regla de probabilidad se caracteriza por los parámetros  $\nu_\xi$ , determinando la intensidad del proceso de Poisson subyacente, y el triplete  $(a_\mu, b_\mu, c_\mu)$  que determina la realización de la parte marcada del proceso. Para obtener un número esperado de particiones razonablemente grande en el dominio de  $\mu$ , se fija  $\nu_\xi = 23$ . Por un lado, la magnitud de la parte marcada del proceso que mide el nivel de la función aleatoria  $\mu$ , se puede fijar en  $a_\mu = 3,6$ ,  $b_\mu = 1$  y  $c_\mu = 4,5$ . Recuérdese que  $a_\mu$  y  $b_\mu$  determinan la distribución de la posición marcada inicial del proceso. Por otro lado,  $a_\mu$  y  $c_\mu$  determinan la ley de la evolución de la parte marcada. En particular,  $c_\mu$  regula qué tan rápida la función  $\mu$  toma el valor 1 en la media<sup>3</sup>. Por lo tanto, con la especificación anterior, aseguramos que las trayectorias muestrales de  $\mu$  no llegarán tan rápido a 1. El procedimiento

<sup>3</sup>Impusimos la restricción para  $\mu$  que vaya hasta 1 en la media para garantizar que los caminos realizados de la medida aleatoria condicional de la probabilidad  $G_1$  satisficiera condiciones de regularidad de funciones de distribución.

de muestreo relacionado al paso trans-dimensional incluido en  $\mu$  depende de  $q_{rj}$  que básicamente regula la probabilidad de incrementar o disminuir el número de componentes incluidos en  $\mu$  en una unidad con respecto al estado reciente de la cadena de este parámetro. Para dar flexibilidad y equilibrio a estos movimientos, se fija  $q_{rj} = 0,5$ . Algunas trayectorias muestrales derivadas del esquema de muestreo de la posterior se muestran en la Figura 3. Se puede observar las trayectorias muestrales de la función  $\mu$  se adaptan en cada iteración. El gráfico resume la distribución de la distribución posterior para  $\mu$ , y por lo tanto, para  $G_1$ , lo que tiene baja dispersión hacia la región con valores altos para la variable latente  $\lambda$ .

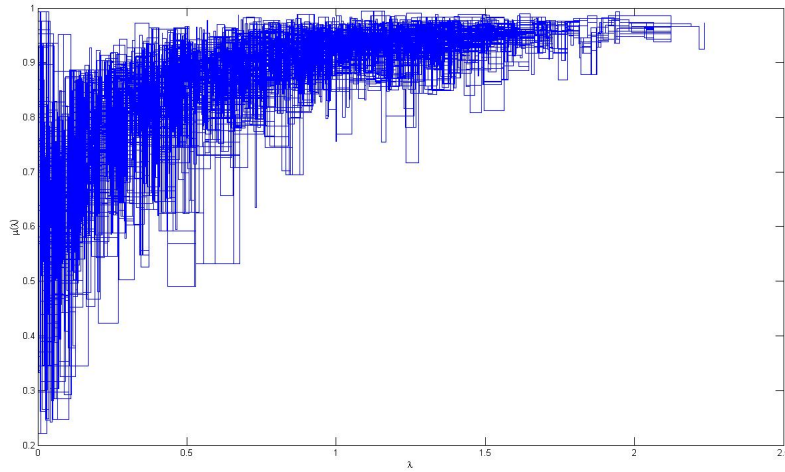


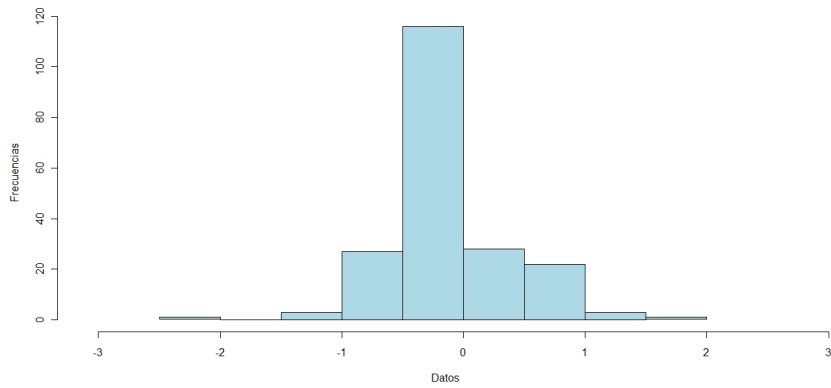
Figura 3: Trayectorias muestrales de la función  $\mu$  para el modelo estacionario de mezcla de escalas mostrado en distintas iteraciones del esquema del muestreo de la posterior.

Con respecto al proceso latente ( $\lambda_t$ ), se puede observar en la Figura 1, panel (b) que el modelo captura las características de volatilidad de los datos observados. Es decir, el proceso latente refleja periodos de alta volatilidad y también periodos con estabilidad relativa en los datos. La misma Figura resume la secuencia de las distribuciones de transición de la posterior  $G_1(\lambda_t|\lambda_{t-1})$ 's a través de la secuencia de las medias de la posterior, es decir,  $\mathbb{E}(\lambda_t|\lambda_{t-1}, \text{data})$  (líneas con puntos en el gráfico) y sus intervalos de confianza al 95 % asociados con la posterior (líneas solidas superiores e inferiores). Recordemos que  $G_1$  está especificada no paramétrica a través de  $Q$  y  $\mu$ , por lo tanto, no se imponen formas estructurales específicas para  $G_1$ , es decir, la distribución se adaptará ante nueva información. Por lo tanto, no sorprende que

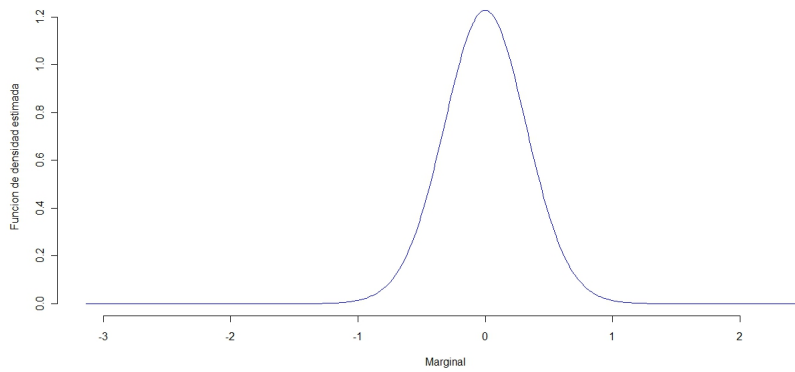
se observen intervalos de confianza asimétricos alrededor de las medias de la posterior del proceso latente.

En la Figura 4, panel (a), se muestra el histograma de las 200 observaciones de los log-retornos semanales de euro-dólar E.U.A. Como se puede apreciar, los datos muestran colas relativamente pesadas y parecen ser asimétricos. El histograma se muestra considerando las observaciones como si fueran independientes, para obtener cierta información acerca de su comportamiento marginal, ya que es la única evidencia empírica disponible para eso. Sin embargo, esta información no es tan adecuada, dado que los datos son dependientes. Por otro lado, el panel (b) de la misma figura muestra que la mezcla de escalas estacionaria captura parte del comportamiento de la cola pesada de los datos, sin embargo, no se refleja bien la asimetría. Lo anterior se esperaba porque, por su construcción, la distribución marginal del modelo sería unimodal y simétrica, como es el caso para el componente de mezcla kernel. Sería posible proporcionar el modelo con más estructura para capturar el sesgo en los datos, pero eso es tema de otro estudio. La distribución predictiva marginal en panel (b) de la misma figura se calculó vía métodos Monte Carlo.

El comportamiento predictivo del modelo estacionario de mezcla de escalas se resume en una secuencia de distribuciones predictivas de un paso para el periodo de  $t = 201$  a  $t = 206$ . Las distribuciones predictivas se computan como las estimaciones de la posterior para  $p(y_{t+1}|y_t)$  dados los datos, es decir, se definen como  $\mathbb{E}\{p(y_{t+1}|y_t)|\mathbf{y}_t\}$  para  $t = 200, \dots, 205$  con  $\mathbf{y}_t = (y_i : i = 1, \dots, t)$ . Las estimaciones Monte Carlo para las densidades predictivas se muestran en la Figura 5 juntas con los datos actuales de cada  $t$ . Como las marginales, las densidades predictivas del modelo de mezcla de escalas son unimodales y simétricas, ya que la dinámica de transición impulsa la brecha o dispersión del modelo condicionalmente de la observación previa. El modelo se puede adaptar para capturar las asimetrías o multimodalidades en la dinámica de transición de los datos, si es necesario. En este momento, solamente modelamos la dispersión condicional en los datos. Como se puede observar, las densidades predictivas adaptan su comportamiento de dispersión y de cola de acuerdo con los datos observados, y logran reflejar la incertidumbre para valores futuros de los datos un periodo en adelante. También, se puede decir que el periodo predicho mostró baja volatilidad. Por lo tanto, se muestra que el modelo estacionario de mezcla de escalas se desempeña



(a) log-retornos semanales de euro-dólar E.U.A.



(b) Estimación de la densidad marginal

Figura 4: Histograma de los log-retornos semanales de euro-dólar E.U.A. y estimación posterior de la densidad marginal.

bastante bien en términos de predicción. Tal aserción se puede validar fácilmente al calcular resúmenes predictivos como los propuestos por [Gelfand et al. \(1992\)](#). Sin embargo, las densidades predictivas que se muestran en la Figura 5 proporcionan evidencia sólida para el periodo bajo estudio<sup>4</sup>.

En esta sección hemos mostrado que se puede implementar el modelo estacionario de mezcla de escalas en la práctica y que produce resultados adecuados con datos actuales. Sin embargo, se requieren algunas consideraciones acerca de su implementación. Modelos bayesianos no paramétricos, en general, se implementan utilizando algoritmos computacionales intensos y sofisticados que requieren tiempo para producir resultados. Con el advenimiento de nueva tecnología, se puede resolver dicho problema. Por otro lado, aunque el modelo estacionario de mezcla de escala es un modelo completamente no paramétrico bien definido, tiene el objetivo de capturar la dispersión marginal y condicional en los datos, como se comentó anteriormente. También, es relativamente simple adaptar el marco del modelo para distintas necesidades, lo que es el enfoque de nuestro trabajo actual y futuro.

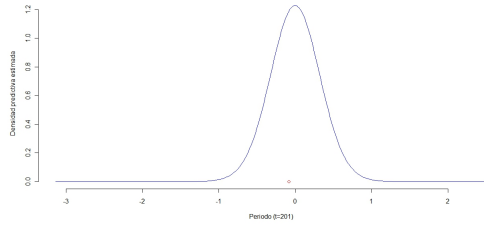
## 7. Discusión

En el presente documento, desarrollamos dos modelos generales no paramétricos estrictamente estacionarios para series de tiempo. Se presentan argumentos que motivan el uso de modelos estacionarios no paramétricos en la práctica. Sin embargo, la contribución principal que aporta este documento a la literatura es la introducción de modelos estacionarios de primer orden que permiten representaciones de mezclas de dimensión infinita para sus distribuciones marginales (invariantes) y de transición. Según nuestro conocimiento, hasta

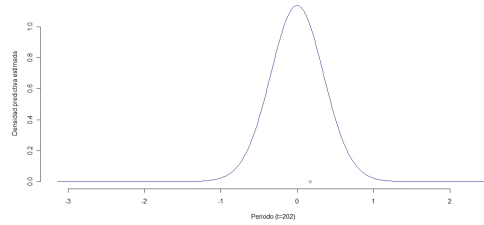
---

<sup>4</sup>La evaluación del modelo, dentro del enfoque bayesiano, se puede realizar al comparar los valores actuales de los datos futuros,  $y_{t+1}$ 's, con sus distribuciones predictivas correspondientes,  $P(y_{t+1}|y_t, \text{data})$ 's. Recuérdese que  $P(y_{t+1}|y_t, \text{data})$  incluye todas expectativas actualizadas sobre  $y_{t+1}$ . Alternativamente, se puede realizar un análisis de residuos (como se usa tradicionalmente en la práctica) al comparar  $y_{t+1}$ 's con sus resultados esperados,  $\mathbb{E}(y_{t+1}|y_t)$ 's (predictivo o ajustado), por ejemplo. Sin embargo, el análisis de residuos puede ser restrictivo ya que se refiere solamente a una característica de  $P(y_{t+1}|y_t)$ . El resumen estadístico para la última comparación fue propuesto por [Gelman et al. \(1996\)](#), entre otros; sin embargo, resúmenes alternativos para modelos bayesianos no paramétricos, en un modelo general, no están disponibles hasta el momento.

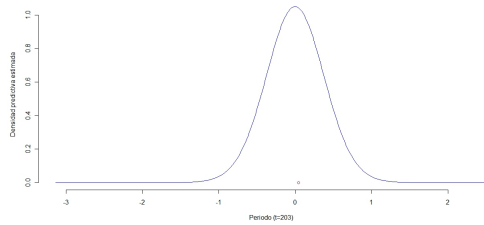




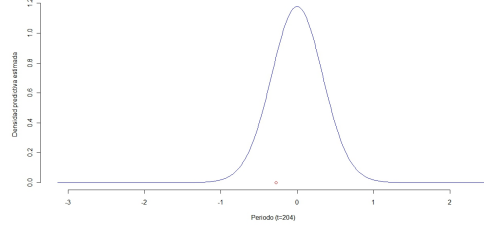
(a)  $t = 201$



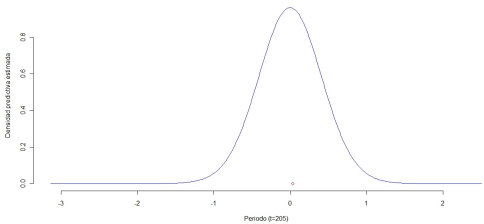
(b)  $t = 202$



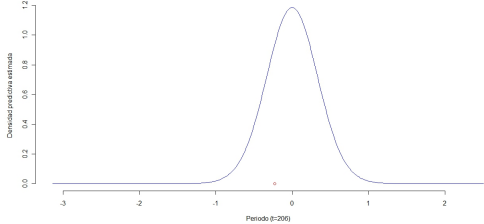
(c)  $t = 203$



(d)  $t = 204$



(e)  $t = 205$



(f)  $t = 206$

Figura 5: Secuencia de las densidades predictivas de un paso para los log-retornos semanales de euro-dólar E.U.A. de  $t = 201$  a  $t = 206$ . El valor actual de  $y_{t-1}$  para (a) es 0, para (b) es  $-0,0791$ , para (c) es  $0,16983$ , para (d) es  $0,04288$ , para (e) es  $-0,27747$  y para (f) es  $0,036403$ . Los valores actuales de  $y_t$  se marcan con “o” en los gráficos.

el momento, éstos son los primeros modelos estacionarios en la literatura que logran dicho objetivo.

Ambos modelos, el de referencia y del muestreador de Gibbs extendido, pueden adaptarse a un amplio rango de escenarios, ya que distintas especificaciones de sus componentes pueden cumplir con distintas necesidades. Por ejemplo, se pueden obtener diferentes soportes en variables observables y latentes al cambiar la especificación de la distribución de mezcla kernel, en ambos modelos, y de la distribución de mezcla base, en el caso del modelo extendido, respectivamente. Con respecto a la construcción del modelo extendido, también se pueden derivar diferentes mecanismos de transición al considerar medidas alternativas de probabilidad determinando la distribución aleatoria alrededor de la cual se induce la estructura de dependencia del modelo. La estructura inducida por el procesos beta-Stacy procesos es bastante flexible, pero será conveniente explorar otros tipos de medidas alternativas de probabilidad determinadas por un proceso de Lévy no homogéneo. Con respecto a los modelos de referencia, todavía se investigan las estrategias computacionales respecto a la inferencia posterior.

El modelo estacionario de mezcla de escalas introducido en la Sección 4 corresponde a un caso particular de la construcción del modelo extendido. Desarrollamos un algoritmo eficiente para la inferencia posterior. Predicciones se obtienen vía métodos tradicionales Monte Carlo (Robert y Casella, 1999). La distribución marginal que incluye este modelo es unimodal y su comportamiento de colas se debe a las colas de la distribución base de mezcla, en este caso, la distribución Weibull. La dinámica de transición de este modelo puede capturar colas pesadas debido a la especificación flexible de  $G_1$ . Especificaciones más generales que pretenden capturar el sesgo en los datos son posibles, por ejemplo, al considerar parametrizaciones más estructuradas.

En general, los modelos derivados de la construcción extendida (el modelo estacionario de mezcla de escalas siendo un caso particular) tienen una representación estado-espacio alternativa que se aparenta a la de los modelos de Markov ocultos (hidden Markov models) (véase, por ejemplo Cappé et al., 2007), en particular, el modelo infinito de Markov oculto (van Gael y Ghahramani, 2010). Una diferencia clave de los modelos desarrollados en este documento con respecto a los modelos de Markov ocultos tradicionales es que la regla de

transición incluida para variables observables y latentes son distribuciones regulares condicionales. Por lo tanto, a diferencia de los modelos de Markov ocultos tradicionales, nuestros modelos permiten recuperar la dinámica marginal para el proceso observable.

En el presente, es bien sabido que se requiere establecer condiciones bajo las cuales los modelos bayesianos no paramétricos, en general, obtienen consistencia posterior (véase, por ejemplo [Diaconis y Freedman, 1986](#)). Se puede verificar que los modelos estacionarios completamente no paramétricos introducidos en este documento son consistentes con el tipo de Doob (véase, [Lijoi et al., 2006](#)). Sin embargo, todavía se tienen que investigar condiciones para obtener consistencia fuerte, no solamente para nuestros modelos, pero para la mayoría de los modelos dependientes bayesianos no paramétricos que se desarrollaron anteriormente. Estudiar la consistencia posterior es un área abierta y activa en este ámbito. El reto principal que enfrentan nuestros modelos en este respecto basa en la representación no paramétrica de su distribución marginal, ya que los enfoques tradicionales para estudiar las propiedades de muestras grandes (asintóticas) no aplican en este contexto. Sin embargo, este problema se investiga actualmente.

Se pueden considerar varias generalizaciones de los modelos desarrollados en este documento. Por ejemplo, sería recomendable explorar generalizaciones multivariadas, la incorporación de covariables o construir modelos con mayor orden de dependencias. También, sería interesante estudiar la conexión de estos modelos con el caso en tiempo continuo.

## Referencias

- Aldous, D. (1985). Exchangeability and related topics. In Hennequin, P. L., editor, *École d'Été de Probabilités de Saint-Flour XII*, Lecture Notes in Mathematics No. 117. Springer.
- Arjas, E. y Gasbarra, D. (1994). Nonparametric Bayesian inference from right censored survival data using the Gibbs sampler. *Statistica Sinica*, 4:505–524.
- Athreya, K. B. y Atuncar, G. S. (1998). Kernel estimation for real-valued Markov chains. *Sankhya: The Indian Journal of Statistics*, 60(A):1–17.
- Barber, D., Cemgil, A. T., y Chiappa, S. (2008). *Proceedings of the Workshop Inference and*

- Estimation in Probabilistic Time-Series Models*. Isaac Newton Institute for Mathematical Sciences, Cambridge.
- Barber, D., Cemgil, A. T., y Chiappa, S. (2011). *Bayesian Time Series Models*. Cambridge University Press, Cambridge.
- Barry, C. B. y Winkler, R. L. (1976). Nonstationarity and portfolio choice. *Journal of Financial and Quantitative Analysis*, 11:217–235.
- Bouzar, N. (2010). Autoregressive sequences via Lévy processes. *REVSTAT - Statistical Journal*, 8(2):81–103.
- Bouzar, N. y Jayakumar, K. (2008). Time series with discrete semistable marginals. *Statistical Papers*, 49(4):619–635.
- Campos, V. S. M. y Dorea, C. C. Y. (2005). Kernel estimation for stationary density of Markov chains with general state space. *Annals of the Institute of Statistical Mathematics*, 57(3):443–453.
- Cappé, O., Moulines, E., y Rydén, T. (2007). *Inference in Hidden Markov Models*. Springer, New York.
- Chu, C.-S. J., Stinchcombe, M., y White, H. (1996). Monitoring structural change. *Econometrica*, 64(5):1045–1065.
- Clements, M. P. y Hendry, D. F. (1999). *Forecasting Non-Stationary Economic Time Series*. The MIT Press, Cambridge, Mass.
- Clements, M. P. y Hendry, D. F. (2008). Economic forecasting in a changing world. *Capitalism and Society*, 3(2):Article 1.
- Contreras-Cristán, A., Mena, R. H., y Walker, S. G. (2009). On the construction of stationary AR(1) models via random distributions. *Statistics*, 43(3):227–240.
- de Finetti, B. (1937). La previsión: Ses logiques, ses sources subjectives. *Annals of the Institute Henri Poincaré*, 7:1–68.

- Diaconis, P. W. y Freedman, D. (1986). On the consistency of bayes estimates (with discussion). *The Annals of Statistics*, 14(1):1–67.
- Doksum, K. (1974). Tailfree and neutral-to-the-right random probabilities and their posterior distributions. *The Annals of Probability*, 2(2):183–201.
- Engle, R. F. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives*, 15(4):157–168.
- Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer, New York.
- Gaver, D. P. y Lewis, P. A. W. (1980). First-order autoregressive gamma sequences and point processes. *Advances in Applied Probability*, 12(3):727–745.
- Gelfand, A. E., Dey, D. K., y Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. In Bernardo, J. M., Berger, J. O., Dawid, A. P., y Smith, A. F. M., editors, *Bayesian Statistics 5*, pages 147–167. Oxford University Press, Oxford.
- Gelman, A., Meng, X. L., y Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica*, 6:733–807.
- Giacominia, R., Gottschling, A., Haefke, C., y White, H. (2008). Mixtures of  $t$ -distributions for finance and forecasting. *Journal of Econometrics*, 144:175–192.
- Godsill, S. T. (2001). On the relationship between MCMC model uncertainty methods. *Journal of Computational and Graphical Statistics*, 10(2):230–248.
- Haas, M., Mittnik, S., y Paoletta, M. S. (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics*, 2(4):493–530.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384.

- Hili, O. (2001). Estimation non paramétrique de la densité d'un processus stationnaire mélangé. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 332(1):841–844.
- Ishwaran, H. y James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453):161–173.
- Jacobs, P. A. y Lewis, P. A. (1978). Discrete time series generated by mixtures. I: Correlational and runs properties. *Journal of the Royal Statistical Society, Series B*, 40(1):94–105.
- Joe, H. (1996). Time series with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability*, 33(3):664–677.
- Jørgensen, B. y Song, P. X.-K. (1998). Stationary time series models with exponential dispersion model margins. *Journal of Applied Probability*, 35(1):78–92.
- Jørgensen, B. y Song, P. X.-K. (2006). Binomial thinning models for integer time series. *Statistical Modelling*, 6:81–96.
- Krolzig, M.-H. (1997). *Markov Switching Vector Autoregressions: Modelling, Statistical Inference and Application to Business Cycle Analysis*. Springer, Berlin.
- Lacour, C. (2008). Nonparametric estimation of the stationary density and the transition density of a Markov chain. *Stochastic Processes and their Applications*, 118(2):232–260.
- Lijoi, A. y Prünster, I. (2010). Models beyond the Dirichlet process. In Hjort, N. L., Holmes, C. C., Müller, P., y Walker, S. G., editors, *Bayesian Nonparametrics*, pages 80–130. Cambridge University Press, Cambridge.
- Lijoi, A., Prünster, I., y Walker, S. G. (2006). Bayesian consistency for stationary models. *Econometric Theory*, 23:479–759.
- Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics*, 12(1):351–357.
- Maitra, A. (1977). Integral representations of invariant measures. *Transactions of the American Mathematical Society*, 229:209–225.

- Martínez-Ovando, J. C. (2011). *Contributions to the Bayesian Nonparametric Modelling of Time-Series Data*. PhD thesis, School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, United Kingdom.
- Mena, R. H. y Walker, S. G. (2005). Stationary autoregressive models via a Bayesian non-parametric approach. *Journal of Time Series Analysis*, 26(6):789–805.
- Mena, R. H. y Walker, S. G. (2007). On the stationary version of the generalized hyperbolic ARCH model. *Annals of the Institute of Statistical Mathematics*, 59:325–348.
- Neal, R. M. (2003). Slice sampling (with discussion). *The Annals of Statistics*, 31(3):705–767.
- Nieto-Barajas, L. E. y Walker, S. G. (2002). Markov beta and gamma processes for modelling hazard rates. *Scandinavian Journal of Statistics*, 29:413–424.
- Pitman, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In Ferguson, T. S., Shapley, L. S., and MacQueen, J. B., editors, *Statistics, Probability and Game Theory*, pages 245–267. Institute of Mathematics and Statistics, Hayward, CA.
- Pitman, J. y Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Annals of Probability*, 25(2):855–900.
- Pitt, M. K., Chatfield, C., y Walker, S. G. (2002). Constructing first order stationary autoregressive models via latent processes. *Scandinavian Journal of Statistics*, 29:657–663.
- Pitt, M. K. y Walker, S. G. (2005). Constructing stationary time series models using auxiliary variables with applications. *Journal of the American Statistical Association*, 100(470):554–564.
- Pitt, M. K. y Walker, S. G. (2006). Extended construction of stationary autoregressive process. *Statistics & Probability Letters*, 76(12):1219–1224.
- Robert, C. P. y Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer-Verlag, London.
- Rodríguez, A., Dunson, D. B., y Gelfand, A. E. (2008). The nested Dirichlet process. *Journal of the American Statistical Association*, 103(483):1131–1154.

- Roussas, G. C. (1969). Nonparametric estimation of the transition distribution function of a Markov process. *The Annals of Mathematical Statistics*, 40(4):1386–1400.
- Teh, Y. W., Jordan, M. I., Beal, M. J., y Blei, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581.
- van Gael, J. y Ghahramani, Z. (2010). Nonparametric hidden Markov models. In Barber, D., Cemgil, A. T., y Chiappa, S., editors, *Bayesian Time-Series Models*. Cambridge University Press, Cambridge.
- Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics*, 36:45–54.
- Walker, S. G. y Muliere, P. (1997). Beta-Stacy processes and a generalization of the Pólya-urn scheme. *The Annals of Statistics*, 25(4):1762–1780.
- Zabell, S. L. (2005). *Symmetry and Its Discontents: Essays on the History of Inductive Probability*. Cambridge Studies in Probability, Induction, and Decision Theory. Cambridge University Press, Cambridge.
- Zeileis, A., Leisch, F., Kleiber, C., y Hornik, K. (2005). An efficient method for estimating seemingly unrelated nonlinear regressions and tests for aggregation bias. *Journal of Applied Econometrics*, 20:99–121.
- Zhu, R. y Joe, H. (2006). Modelling count data time series with Markov processes based on binomial thinning. *Journal of Time Series Analysis*, 27(5):725–738.

## A. Construcción del modelo con el muestreador de Gibbs basado en mezclas

El modelo derivado en la Sección 3 se puede tratar bajo un enfoque más general basado en mezclas. Es posible partir de una medida conjunta de probabilidad derivada de la siguiente



función aleatoria de densidad (o distribución)

$$f_H(y) = \int K(y|\theta)H(d\theta), \quad (\text{A.1})$$

donde  $\theta$  es una variable aleatoria de mezcla,  $K(y|\theta)$  es un kernel paramétrico fijo, y  $H$  es una función aleatoria de distribución. Entonces, la medida de probabilidad para  $f_H$  se caracterizaría por la medida de probabilidad para  $H$ . Por lo tanto, la definición estocástica análoga a (2.5) se daría por

$$y|f_H \sim f_H, \quad \text{y} \quad H \sim \Pi, \quad (\text{A.2})$$

donde  $\Pi$  es la medida de probabilidad para  $H$ , con el supuesto adicional de  $\mathbb{E}_\Pi(H) = G_0$ , para una función difusa de distribución  $G_0$ . Después de aplicar el enfoque de construcción del modelo con el muestreador de Gibbs, las distribuciones marginales y de transición de primer orden del modelo estacionario serían

$$p(y_t) = \int f_H(y_t)\Pi(dH) = \int K(y_t|\theta)G_0(d\theta), \quad (\text{A.3})$$

$$p(y_t|y_{t-1}) = \int f_H(y_t)\Pi(dH|y_{t-1}) = \int K(y_t|\theta)G_1(d\theta|\theta')G_0(d\theta'|y_{t-1}), \quad (\text{A.4})$$

donde  $\theta$  y  $\theta'$  son dos variables auxiliares de mezcla,  $G_1(d\theta|\theta') = \mathbb{E}_\Pi(H(d\theta)|\theta')$  y  $G_0(\theta'|y_1) \propto K(y_1|\theta')G_0(\theta')$ . Como consecuencia, el par (A.3) y (A.4) caracterizan un proceso estrictamente estacionario de Markov.

Una característica clave de modelos estacionarios derivados con mezclas es que realmente es posible expresar su dinámica de transición a través de variables observables y latentes en conjunto al extender el enfoque de (A.2) a una medida aumentada de probabilidad inducida por

$$y|\theta \sim K(y|\theta), \quad \theta|H \sim H, \quad H \sim \Pi. \quad (\text{A.5})$$

Por lo tanto, distribuciones conjuntas marginal (invariante) y de transición para el par  $(Y, \theta)$  se puede escribir como

$$p(y_t, \theta_t) = \int K(y_t|\theta_t)H(\theta_t)\Pi(dH) = K(y_t|\theta_t)G_0(\theta_t), \quad (\text{A.6})$$

$$p(y_t, \theta_t|y_{t-1}, \theta_{t-1}) = \int K(y_t|\theta_t)H(\theta_t)\Pi(dH|\theta_{t-1}) = K(y_t|\theta_t)G_1(\theta_t|\theta_{t-1}). \quad (\text{A.7})$$

La última relación sigue de la estructura de independiencia condicional entre  $Y$  y  $H$ , dada  $\theta$ . También se puede comprobar que (A.6)-(A.7) caracterizan la ley de probabilidad de la

secuencia conjunta de Markov  $\{(Y_t, \theta_t)\}_{t=1}^{\infty}$ . Esta construcción se parece a la especificación de modelos estado-espacio. Sin embargo, en nuestro contexto, la parte marginal del modelo conjunto corresponde a las observables, es decir,  $\{Y_t\}_{t=1}^{\infty}$ , se puede recuperar desde (A.7) y coincide con (A.4), dado que las distribuciones condicionales involucradas en nuestro modelo son regulares.

Finalmente, se puede observar que la construcción del modelo extendido desarrollada en la Sección 3 se modificó de una versión más estructurada a (A.5). Procedemos con una descripción más detallada de la formula involucrada en el modelo estacionario de mezcla de escalas presentado en la Sección 4.

## B. Modelo estacionario de mezcla de escalas (formulas relevantes)

En la presente sección, se presenta formulas relevantes con respecto al modelo estacionario de mezcla de escalas desarrollado en la Sección 4. Se proporcionan expresiones analíticas explícitas para la dinámica de transición del modelo y se desarrollan detalles adicionales con respecto al esquema de truncamiento estocástico.

### B.1. Dinámica de transición

La dinámica de transición del modelo estacionario de mezcla de escalas se caracteriza por la distribución dada en (4.8) que corresponde a una transición de tipo mixto cuya función de densidad/probabilidad permite la siguiente descomposición

$$g_1(\lambda|\lambda') = \sum_{z=-1}^1 g_{1,z}(\lambda|\lambda')p(z|\lambda'), \quad (\text{B.1})$$

donde  $\lambda$  y  $\lambda'$  son dos variables latentes y  $z$  es otra variables latente que indica la región donde  $\lambda$  toman valores después de considerar la partición inducida en su soporte al condicionarla a

$\lambda'$ . Los otros componentes involucrados en la mezcla son

$$\begin{aligned} p(z = -1|\lambda') &= 1 - S_\mu(\lambda'), & g_{1,z=-1}(\lambda|\lambda') &= \frac{g_\mu(\lambda|\lambda')}{1 - S_\mu(\lambda')} \mathbf{1}(\lambda < \lambda'), \\ p(z = 0|\lambda') &= [1 - \mu(\lambda')] S_\mu(\lambda'), & g_{1,z=0}(\lambda|\lambda') &= \delta_{\lambda'}(\lambda), \\ p(z = 1|\lambda') &= \mu(\lambda') S_\mu(\lambda'), & g_{1,z=1}(\lambda|\lambda') &= \frac{g_0(\lambda)}{S_0(\lambda')} \mathbf{1}(\lambda > \lambda'); \end{aligned} \quad (\text{B.2})$$

con,

$$S_\mu(\lambda) = \exp \left\{ - \sum_{\{k:\xi_k \leq \lambda\}} \mu_k h_{0,k}(\lambda) \right\} \quad (\text{B.3})$$

$$h_{0,k}(\lambda) = \int_{E_k \cap (0,\lambda]} h_0(s) ds, \quad (\text{B.4})$$

$$S_0(\lambda) = \exp \left\{ - \int_0^\lambda h_0(u) du \right\}, \quad (\text{B.5})$$

y

$$g_\mu(\lambda|\lambda') = \mu(\lambda) h_0(\lambda) S_\mu(\lambda) \mathbf{1}(\lambda < \lambda') \quad (\text{B.6})$$

$$= \mu_{k(\lambda)} h_0(\lambda) \cdot \exp \left\{ - \sum_{\{k:\xi_k \leq \lambda'\}} \mu_k h_{0,k}(\lambda) \right\}, \quad (\text{B.7})$$

$$g_0(\lambda) = h_0(\lambda) S_0(\lambda). \quad (\text{B.8})$$

El índice  $k(\lambda)$  incluido en (B.7) representa el índice de la partición asociada con  $\mu$  donde  $\lambda \in E_k$  y  $\mu_{k(\lambda)}$  representa su nivel asociado, con  $h_{0,k}(\lambda)$  dada. La función  $h_0$  refiere a la función hazard de la distribución base  $G_0$  –es decir, la distribución Weibull en este caso–, y  $g_0$  a su densidad asociada.

## B.2. Truncamiento estocástico vía variables latentes

El esquema del truncamiento estocástico parece ser una forma adecuada para tratar las sumas infinitas involucradas en la verosimilitud (4.10). El truncamiento realizado a través de (5.2), donde la variable latente  $U_t$  tiene soporte en el conjunto  $\{0, 1, \dots\}$ . De acuerdo con

eso, la verosimilitud aumentada se vuelve

$$\begin{aligned}
lik \propto & \left\{ \prod_{t=1}^T N(y_t | 0, \lambda_t) \right\} \cdot \left\{ \prod_{\{t=1\} \cup \{t: d_t \neq d_{t-1}\}} W_{d_t} (1 - \phi)^{u_t - d_t} \mathbf{1}(u_t \geq d_t) \theta_{d_t} \lambda_t^{\theta_{d_t} - 1} e^{-\lambda_t^{\theta_{d_t}}} \right\} \\
& \cdot \left\{ \prod_{\{t: d_t = d_{t-1}, z_t = -1\}} W_{d_t} (1 - \phi)^{u_t - d_t} \mathbf{1}(u_t \geq d_t) \mu_{k(\lambda_t)} \theta_{d_t} \lambda_t^{\theta_{d_t} - 1} \right\} \\
& \cdot \left\{ \prod_{\{t: d_t = d_{t-1}, z_t = 0\}} W_{d_t} (1 - \phi)^{u_t - d_t} \mathbf{1}(u_t \geq d_t) (1 - \mu_{k(\lambda_{t-1})}) \right\} \\
& \cdot \left\{ \prod_{\{t: d_t = d_{t-1}, z_t = 1\}} W_{d_t} (1 - \phi)^{u_t - d_t} \mathbf{1}(u_t \geq d_t) \mu_{k(\lambda_{t-1})} \theta_{d_t} \lambda_t^{\theta_{d_t} - 1} e^{-(\lambda_t^{\theta_{d_t}} - \lambda_{t-1}^{\theta_{d_t}})} \right\} \\
& \cdot \exp \left\{ - \sum_{\{k: \xi_k < \lambda_t\}} \sum_{\{t: d_t = d_{t-1}, z_t = -1\}} \mu_k h_{0,k}(\lambda_t) \right\} \\
& \cdot \exp \left\{ - \sum_{\{k: \xi_k < \lambda_{t-1}\}} \sum_{\{t: d_t = d_{t-1}, z_t \in \{0,1\}\}} \mu_k h_{0,k}(\lambda_{t-1}) \right\}, \tag{B.9}
\end{aligned}$$

que se define para  $(Q, \mu, \boldsymbol{\lambda}, \mathbf{z}, \mathbf{d}, \mathbf{u})$ , donde  $\mathbf{u} = \{u_t\}_{t=1}^T$  representa la secuencia de variables latentes de truncamiento, y  $k(\lambda_t)$  y  $h_{0,k}(\lambda_t)$  están dadas. Más detalles acerca de este y otros esquemas estocásticos de truncamiento, incluyendo modelos de mezclas infinitas, se pueden encontrar en [Martínez-Ovando \(2011\)](#). En el siguiente apéndice, explicaremos el algoritmo para el muestreo posterior desarrollado con base en (B.9) para el modelo estacionario de mezcla de escalas.

## C. Detalles adicionales del muestreo de la posterior

Se proporcionan detalles con respecto a las distribuciones condicionales completas y el esquema de muestreo de posterior diseñando en la Sección 5.3, relacionado al modelo estacionario de mezcla de escalas.

### C.1. Distribuciones condicionales completas para $\mathbf{u}$

La distribución condicional completa para  $\mathbf{u}$  sólo depende de  $\mathbf{d}$ . Además, las  $u_t$ 's son condicionalmente independientes dados los  $d_t$ 's. Por lo tanto, solamente necesitamos especificar

la distribución genérica condicional completa para cada  $u_t$  que, condicional a  $d_t$ , sigue la distribución geométrica desplazada

$$\pi(u_t|d_t) = \phi(1 - \phi)^{u_t - d_t}, \quad (\text{C.1})$$

con soporte en el conjunto  $\{d_t, d_t + 1, \dots\}$ , como es dado en (5.2). Se puede realizar el muestreo con esta distribución condicional completa al utilizar procedimientos estándar.

## C.2. Distribución condicional completa para $\mathbf{d}$

La distribución condicional completa para  $\mathbf{d}$  se define por las distribuciones condicionales completas individuales para cada  $d_t$  que se deriven con (5.3). Pero, a diferencia de modelos bayesianos de mezcla no paramétricos estándar, la estructura condicional independiente entre las  $d_t$ 's no prevalece en nuestro modelo. En lugar de eso, los índices latentes exhiben una estructura de dependencia markoviana. Por lo tanto, la distribución condicional completa para cada  $d_t$  dependerá de  $(W_{d_t}, \theta_{d_t}, \lambda_t, \lambda_{t-1}, \mu, u_t)$  y el par de índices  $(d_{t-1}, d_{t+1})$ . Por lo tanto,  $d_t$  se distribuiría por

$$\pi(d_t | \dots) \propto g(d_t)(1 - \phi)^{-d_t} \mathbf{1}(d_t \leq u_t), \quad (\text{C.2})$$

para cada  $t$ , donde  $g(d_t)$  es una función medible que representa el componente general de la verosimilitud en la cual está incluida  $d_t$ . [Nótese que en modelos generales de mezcla la función  $g(d_t)$  se daría por  $W_{d_t} f(\lambda_t | \theta_{d_t})$ , donde  $f(\lambda_t | \theta_{d_t})$  sería el componente kernel de la mezcla. Sin embargo, lo anterior no aplica en nuestro caso.]

Antes de continuar, nótese que al imponer condiciones para  $u_t$ , el rango de posibles valores que  $d_t$  puede tomar se trunca al conjunto  $\{1, \dots, u_t\}$ . Por lo tanto, la distribución condicional completa para  $d_t$  incluirá calcular sumas finitas de la siguiente forma

$$\sum_{d_t=1}^{u_t} g(d_t)(1 - \phi)^{-d_t}.$$

El factor  $g(d_t)$  corresponde al producto de kernel de mezclas y el peso aleatorio asociado con el  $d_t$ -ésimo componente. Como se mencionó antes, la contribución de  $d_t$ 's a la verosimilitud

llevaría a las funciones  $g(d_t)$  definidas como

$$g(d_t = j) = \begin{cases} W_j g_1(\lambda_{t+1}|\lambda_t; \theta_j, \mu) g_1(\lambda_t|\lambda_{t-1}; \theta_j, \mu) & , \quad j = d_{t+1} = d_t = d_{t-1}, \\ W_j g_1(\lambda_{t+1}|\lambda_t; \theta_j, \mu), & , \quad j = d_{t+1} = d_t, \quad d_t \neq d_{t-1}, \\ W_j g_1(\lambda_t|\lambda_{t-1}; \theta_j, \mu), & , \quad d_{t+1} \neq d_t, \quad j = d_t = d_{t-1}, \\ W_j \text{We}(\lambda_t|\theta_j), & , \quad j = d_t, \quad d_{t+1} \neq d_t \neq d_{t-1}, \end{cases} \quad (\text{C.3})$$

para  $j = 1, 2, \dots, u_t$ . Por lo tanto, definir un nuevo muestreo,  $d_t$ , condicional a  $u_t$ , incluirá tomar muestras de la distribución discreta con soporte en  $\{1, \dots, u_t\}$  y peso de probabilidad proporcionales a los factores  $g(d)(1 - \phi)^{-d}$ , para diferentes valores de  $d$ .

Dos excepciones obvias de la especificación de  $g(d_t)$  dada arriba están en orden para el tiempo  $t = 1$  y  $T$ . Para el tiempo inicial,  $t = 1$ , esta función tomaría la forma

$$g(d_1 = j) = W_j \text{We}(\lambda_1|\theta_j),$$

para  $j = 1, 2, \dots, u_1$ , como ocurre con modelos tradicionales de mezcla. Respecto al tiempo  $t = T$ , resulta la función  $g$

$$g(d_T = j) = \begin{cases} W_j g_1(\lambda_T|\lambda_{T-1}; \theta_j, \mu), & , \quad j = d_T = d_{T-1}, \\ W_j \text{We}(\lambda_T|\theta_j), & , \quad j = d_T, \quad d_T \neq d_{T-1}, \end{cases}$$

para  $j = 1, 2, \dots, u_T$ . Nótese que la distribución anterior se especificaría en términos de combinaciones diferentes que también podría generar el par  $(z_t, z_{t+1})$ . Para el caso donde  $d_t = d_{t-1}$  o  $d_t = d_{t+1}$ , tal distribución se daría en términos de  $z_t$  o  $z_{t+1}$ , respectivamente.

### C.3. Distribución condicional completa para $\mu$

La distribución condicional completa para  $\mu$  no tiene una forma analítica cerrada. Para actualizar  $\mu$  se requiere la implementación de muestreos trans-dimensionales. El muestreo que adoptamos, actualiza la secuencia  $\{(\xi_k, \mu_k)\}$  de acuerdo con las variables latentes actuales  $\{\lambda_t\}$  que corresponden a los casos  $\{t : d_t = d_{t-1}\}$ . Para cada iteración, la forma de la distribución condicional completa para  $\mu$  se dará por

$$\pi(\{(\xi_k, \mu_k)\}_{k=0}^{\infty} | (\lambda_t, z_t, \theta_{d_t}, d_t : t = 1 \dots, T)) \propto \text{lik}(\{(\xi_k, \mu_k)\}_{k=0}^m; (\lambda_t, z_t, \theta_{d_t}, d_t : t = 1 \dots, T)) \cdot \pi(\{(\xi_k, \mu_k)\}_{k=0}^m) \pi(m|\lambda^*) \quad (\text{C.4})$$

donde  $m = \#\{\xi_k \leq \lambda^*\}$ , con  $\lambda^* = \max_t \{\lambda_t\}$ . Con respecto a la especificación de la inicial para  $\mu$  dada en la Sección 5.2.1, después de condicionar por los datos, inducirá la representación alternativa

$$\begin{aligned} \pi(m|\lambda^*) &\propto \nu_\xi^m e^{-\nu_\xi \lambda^*}, \\ \pi(\{(\xi_k, \mu_k)\}_{k=0}^m | m) &\propto \text{Be}(\mu_0 | a_\mu, b_\mu) \cdot \left\{ \prod_{k=1}^m \text{Be}(\mu_k | a_\mu + k, c_\mu) \right\}, \end{aligned} \quad (\text{C.5})$$

donde  $m$  es aleatoria. Por lo tanto, la parte de la verosimilitud (C.4) incluyendo  $\mu$  será

$$\begin{aligned} \text{lik}(\{(\xi_k, \mu_k)\}_{k=0}^m; (\lambda_t, z_t, \theta_{d_t}, d_t : t = 1 \dots, T)) &\propto \\ &\propto \left\{ \prod_{\{t: d_t = d_{t-1}, z_t \in \{-1, 1\}\}} \mu_{k(\lambda_t)} \right\} \cdot \left\{ \prod_{\{t: d_t = d_{t-1}, z_{-t} = 0\}} (1 - \mu_{k(\lambda_{t-1})}) \right\} \\ &\cdot \exp \left\{ - \sum_{\{k: \xi_k < \lambda_t\}} \sum_{\{t: d_t = d_{t-1}, z_t = -1\}} \mu_k h_{0,k}(\lambda_t) \right\} \\ &\cdot \exp \left\{ - \sum_{\{k: \xi_k < \lambda_{t-1}\}} \sum_{\{t: d_t = d_{t-1}, z_t \in \{0, 1\}\}} \mu_k h_{0,k}(\lambda_{t-1}) \right\}, \end{aligned} \quad (\text{C.6})$$

donde  $k(\lambda_t)$  y  $h_{0,k}(\lambda_t)$  están dadas como antes, para  $k = 1, 2, \dots$

El muestreo trans-dimensional a implementar sigue las ideas desarrolladas por [Godsill \(2001\)](#) que consiste en definir un muestreo Gibbs-Metropolis para un espacio de parámetros de dimensión infinita. Este esquema requiere extender el alcance a (C.6) al incluir otra variable latente,  $v$ , que indica el componente de  $\{(\xi_k, \mu_k)\}_{k=0}^\infty$  con cual se cambiaría el muestreo. Condicional al número actual de componentes  $m$  involucrados en  $\mu$  y  $\lambda$ . La distribución para  $v$  se dará por

$$v|m = \begin{cases} m, & \text{con prob. } 1 - q_{rj}, \\ m + 1, & \text{con prob. } q_{rj}, \end{cases} \quad (\text{C.7})$$

donde  $0 < q_{rj} < 1$  es un parámetro de muestreo pre-definido.

Por lo tanto, el esquema de actualización para  $\mu$  consistirá en mover el estado actual de la cadena  $\{(\xi_k, \mu_k)\}_{k=0}^m$  a  $\{\xi'_k, \mu'_k\}'_{k=0}$ , con  $m'$  definida sobre  $\{m - 1, m, m + 1\}$ . Este paso está restringido por el intervalo  $(0, \lambda^*]$  que varía entre las iteraciones. El procedimiento de actualización incluye los siguientes pasos:

- i) Condicional a  $m$ , estados propuestos  $\{(\xi'_k, \mu'_k)\}_{k=1}^l$  se derivan de  $l = m - 1, m, m + 1$ . Este paso se describe a continuación.
- ii) Condicional a  $m$ , el parámetro  $v'$  se actualiza con (C.7).
- iii) Condicional a  $v'$ , el índice  $m'$  se actualiza al tomar una muestra de su distribución condicional completa que se define con el valor actualizado de  $v'$ , es decir: a)  $v' = m$ , y b)  $v' = m + 1$ . Estas distribuciones se muestran a continuación.

a) En el caso donde  $v' = m$ , la distribución condicional completa para  $m'$  se vuelve

$$m'|v' = m = \begin{cases} m - 1, & \text{con prob. } \propto q_{rj} p(\{(\xi'_k, \mu'_k)\}_{k=0}^{m-1} | \dots) \\ & \cdot p(\{(\xi'_k, \mu'_k)\}_{k=0}^m | \{(\xi'_k, \mu'_k)\}_{k=0}^{m-1}, m), \\ m, & \text{con prob. } \propto (1 - q_{rj}) p(\{(\xi'_k, \mu'_k)\}_{k=0}^m | \dots) \\ & \cdot p(\{(\xi'_k, \mu'_k)\}_{k=0}^{m-1} | \{(\xi'_k, \mu'_k)\}_{k=0}^m, m). \end{cases} \quad (\text{C.8})$$

b) El caso  $v' = m + 1$  da lugar a la siguiente distribución para  $m'$ ,

$$m'|v' = m + 1 = \begin{cases} m, & \text{con prob. } \propto q_{rj} p(\{(\xi'_k, \mu'_k)\}_{k=0}^m | \dots) \\ & \cdot p(\{(\xi'_k, \mu'_k)\}_{k=0}^{m+1} | \{(\xi'_k, \mu'_k)\}_{k=0}^m, m), \\ m + 1, & \text{con prob. } \propto (1 - q_{rj}) p(\{(\xi'_k, \mu'_k)\}_{k=0}^{m+1} | \dots) \\ & \cdot p(\{(\xi'_k, \mu'_k)\}_{k=0}^m | \{(\xi'_k, \mu'_k)\}_{k=0}^{m+1}, m). \end{cases} \quad (\text{C.9})$$

- iv) Dado el valor actual de  $m'$ , se obtienen los estados propuestos  $\{(\xi'_k, \mu'_k)\}_{k=0}^{m'}$  y se mantienen como el estado actualizado de  $m'$ .

En la expresión anterior,  $p(\{(\xi'_k, \mu'_k)\}_{k=0}^l | \dots) = p(\{(\xi'_k, \mu'_k)\}_{k=0}^l | (\lambda'_t, z'_t, \theta'_{d_t}, d'_t : t = 1 \dots, T))$ , para  $l = m - 1, m, m + 1$ .

Con respecto al paso (i) descrito antes, los tres movimientos propuestos se definen para obtener nuevos estados  $\{(\xi'_k, \mu'_k)\}_{k=0}^l$  para  $l = m - 1, m, m + 1$ : i) Actualizar un punto de cambio actual; ii) incorporar un nuevo punto de cambio; y iii) borrar un punto de cambio. Más adelante se presentan detalles al respecto.



### C.3.1. Actualizar un punto de cambio actual.

Proponer  $\{(\xi'_k, \mu'_k)\}_{k=0}^m$ , dada  $m$ , consiste en escoger un índice  $k'$  aleatoriamente de un conjunto  $\{0, 1, \dots, m\}$ . Entonces, un nuevo nivel  $\mu'_{k'}$  se deriva de la distribución

$$p(\mu'_{k'} | \mu_{k'-1}, \mu_{k'+1}, \xi_{k'}) = U(\mu'_{k'} | \min\{\mu_{k'-1}, \mu_{k'+1}\}, \max\{\mu_{k'-1}, \mu_{k'+1}\}), \quad (\text{C.10})$$

y el índice actual  $\xi'_k = \xi_{k'}$  se mantiene. Los otros componentes se actualizan como  $\xi'_k = \xi_k$  y  $\mu'_k = \mu_k$ , para  $k \neq k'$ . Entonces, la secuencia actualizada  $\{(\xi'_k, \mu'_k)\}_{k=0}^m$  se acepta con probabilidad

$$\min \left\{ 1, \frac{p(\boldsymbol{\lambda} | \{(\xi'_k, \mu'_k)\}_{k=0}^m; \boldsymbol{z}, \theta) \pi(\{(\xi'_k, \mu'_k)\}_{k=0}^m | m)}{p(\boldsymbol{\lambda} | \{(\xi_k, \mu_k)\}_{k=0}^m; \boldsymbol{z}, \theta) \pi(\{(\xi_k, \mu_k)\}_{k=0}^m | m)} \right\}, \quad (\text{C.11})$$

donde  $p(\boldsymbol{\lambda} | \{(\xi_k, \mu_k)\}_{k=0}^m, \boldsymbol{z}, \theta)$  y  $\pi(\{(\xi_k, \mu_k)\}_{k=0}^m | m)$  están dadas como en (C.6) y (C.5), respectivamente.

### C.3.2. Incorporar un nuevo punto de cambio.

Definir  $\{(\xi'_k, \mu'_k)\}_{k=0}^{m+1}$ , dada  $m$ , consiste en tomar la muestra para un nuevo punto de cambio  $\xi'$  de la distribución  $U(0, \lambda^*)$  y, dada  $\xi'$ , un nuevo índice  $k(\xi') = \{k : \xi' \in E_k, k = 1, \dots, m\}$  se define. Entonces, se propone un nuevo nivel  $\mu'$  asociado con  $\xi'$  al tomar una muestra de

$$p(\mu' | \xi', \mu_{k(\xi')}, \mu_{k(\xi')+1}) = U(\mu' | \min\{\mu_{k(\xi')-1}, \mu_{k(\xi')+1}\}, \max\{\mu_{k(\xi')-1}, \mu_{k(\xi')+1}\}). \quad (\text{C.12})$$

Se asume que  $k(\xi') = l$ . Entonces, la secuencia propuesta  $\{(\xi'_k, \mu'_k)\}_{k=0}^{m+1}$  será  $\xi'_k = \xi_k$  y  $\mu'_k = \mu_k$ , for  $k = 0, \dots, l$ ;  $\xi'_{l+1} = \xi'$  y  $\mu'_{l+1} = \mu$ ; y  $\xi'_k = \xi_{k-1}$  y  $\mu'_k = \mu_{k-1}$  para  $k = l+2, \dots, m+1$ .

### C.3.3. Eliminar un punto de cambio

Se propone  $\{(\xi'_k, \mu'_k)\}_{k=0}^{m-1}$ , dada  $m$ , lo que se realiza al definir un valor de índice  $k'$  uniformemente de un conjunto  $\{1, \dots, m\}$  y definir  $\xi'_k = \xi_k$ , para  $k = 0, \dots, k' - 1$ , y  $\xi'_k = \xi_{k+1}$ , para  $k = k', \dots$ . Para actualizar los estados de los niveles aleatorios se requieren los siguientes pasos: i) para  $k = 0, \dots, k' - 1$ , se define  $\mu'_k = \mu_k$ , y ii) para  $k = k', \dots, \mu'_k = \mu_{k+1}$  es definido. Recuérdese que  $\lambda^*$  variará con cada iteración del muestreador de Gibbs, entonces,  $\mu$  se debería actualizar en el intervalo  $(0, \lambda^*]$  solamente, manteniendo los demás componentes de  $\mu$ , con  $\lambda^*$  sin cambio hasta la próxima iteración.

Si en el muestreo, el conjunto  $\{t : d_t = d_{t-1}\}$  es vacío (no es un caso común), entonces no tenemos información acerca  $\mu$  en esta iteración. Entonces, en ausencia de evidencia estadística sobre  $\mu$ , la muestra mantendrá los mismo estados como en la iteración previa de la cadena. Sin embargo, no esperamos observar muchos escenarios de este tipo, ya que la verosimilitud debe ser impulsada por los datos, impulsada por el dinámica de proceso, por lo tanto, en muy pocas ocasiones o ninguna, el conjunto  $\{t : d_t = d_{t-1}\}$  será vacío.

## C.4. Distribución condicional completa para $Q$

Después de imponer condiciones para  $\mathbf{u}$ , la función  $Q$  se caracteriza por la colección finita de variables latentes  $\{W_j\}_{j=1}^{u^*}$  y  $\{\theta_j\}_{j=1}^{u^*}$ , con  $u^* = \max\{u_t : t = 1, \dots, T\}$ , para las cuales se obtienen las distribuciones condicionales completas debido a la representación stick-breaking de  $Q$ . Se obtienen expresiones analíticas a través del muestreo de Gibbs por bloques (Ishwaran y James, 2001).

### C.4.1. Distribución condicional completa para $W_j$

Las  $W_j$ 's actualizadas dependen de un conjunto de variables indicador latentes actuales  $\{d_t\}$  y, implícitamente, del conjunto  $\{d_1^*, \dots, d_m^*\}$  de valores distintos de  $\{d_t\}$ . Entonces, al seguir el esquema multinomial de actualización, la distribución condicional completa para  $\{W_j\}_{j=1}^{u^*}$  se dará en términos de la distribución condicional completa asociada de variables stick-breaking  $\{V_j\}_{j=1}^{u^*}$ , con el siguiente esquema de actualización

$$W'_1 = V'_1 \quad \text{y} \quad W'_j = V'_j \prod_{l < j} (1 - V'_l),$$

donde

$$V'_l \stackrel{\text{ind}}{\sim} \text{Be}(1 - \alpha_{\text{PDP}} + M_l, l\alpha_{\text{PDP}} + \sum_{i=l+1}^{u^*} M_i),$$

para  $l = 1, \dots, u^*$ , y  $M_l = \#\{d_t : d_t = l\}$ .

### C.4.2. Distribución condicional completa para $\theta_j$

De manera similar, cuando se imponen condiciones para  $\{u_t\}$ , solamente un conjunto finito de parámetros  $\{\theta_j\}_{j=1}^{u^*}$  se tiene que actualizar. La distribución condicional completa

para cada variable latente se da por

$$p(\theta_j | \dots) \propto \text{Ga}(\theta_j | a_{Q_0}, b_{Q_0}) \cdot \text{We}(\lambda'_1 | \theta_j) \cdot \left\{ \prod_{t: d'_t \neq d'_{t-1}, d'_t = j} \text{We}(\lambda'_t | \theta_j) \right\} \cdot \left\{ \prod_{t: d'_t = d'_{t-1}, d'_t = j} g_1(\lambda'_t | \lambda'_{t-1}; \mu', \theta_j) \right\} \quad (\text{C.13})$$

para  $j = 1, \dots, u^*$ , con  $(d'_t, \lambda'_t, \mu')$  siendo los estados actuales de la cadena de Markov. Nótese que los primeros componentes de la expresión anterior coinciden con el kernel de posterior del modelo paramétrica Weibull-gamma para el parámetro de forma  $\theta_t$ . Tal distribución no tiene una forma analítica cerrada, así que se requiere otro método de muestreo para esa parte, por ejemplo, muestreador de la rebanada. Con respecto a  $\theta_j$ 's para las cuales no hay  $d'_t$ 's que coinciden en valor con el índice  $j$ ,  $\theta'_j$  actualizada se tiene que obtener de su distribución marginal  $Q_0$ , es decir, de  $\text{Ga}(a_{Q_0}, b_{Q_0})$  en nuestro caso.

## C.5. Distribución condicional completa para $\mathbf{z}$

Con cada iteración del muestreo de posterior, los estados actuales para  $\{d_t\}$  definirán un vector latente de variable de dimensión,  $\mathbf{z} = (z_t : d_t = d_{t-1}, t = 2, \dots, T)$ , de acuerdo con los casos donde  $d_t = d_{t-1}$ . Para todas estas variables latentes, la distribución condicional completa sería

$$\pi(z_t | \mu, \mathbf{z}_{-t}, \boldsymbol{\lambda}, \dots) \propto g_{1, z_t}(\lambda_t | \lambda_{t-1}; \theta_{d_t}, \mu) p(z_t | \lambda_{t-1}; \theta_{d_t}, \mu), \quad (\text{C.14})$$

donde  $g_{1, z_t}(\lambda_t | \dots)$  y  $p(z_t | \dots)$  están dadas.

Por lo tanto, actualizar  $z_t$  en cada tiempo  $t \in \{t : d_t = d_{t-1}\}$  se realiza de acuerdo con dos casos inducidos por el triplete  $(d_{t-1}, d_t, d_{t+1})$ : 1)  $d_{t-1} = d_t$  y  $d_t = d_{t+1}$ , y 2)  $d_{t-1} = d_t$  y  $d_t \neq d_{t+1}$ . Dichos casos darán lugar al siguiente esquema de muestreo:

- 1) Para el caso  $d_{t-1} = d_t$  y  $d_t = d_{t+1}$ , obtener  $z_t$  depende de los estados actuales de  $\theta_{d_t}$ . En términos de las  $\lambda$ 's involucradas, se deben considerar los tres casos siguientes: i)  $\lambda'_{t-1} < \lambda_{t+1}$ , ii)  $\lambda'_{t-1} = \lambda_{t+1}$ , y iii)  $\lambda'_{t-1} > \lambda_{t+1}$ . Tales casos, combinados con los estados actuales de  $z_{t+1}$  y  $\lambda_t$  llevan al siguiente esquema de muestreo:

i) El caso  $\lambda'_{t-1} < \lambda_{t+1}$  combinado con el valor actual de  $z_{t+1}$  llega al siguiente esquema de muestreo:

- a) Si  $z_{t+1} = 1$ , entonces, se toman la muestra aleatoriamente de  $z'_t$  del conjunto  $\{-1, 0, 1\}$  de acuerdo con la distribución

$$\pi(z_t|\mu, \mathbf{z}_{-t}, \boldsymbol{\lambda}, \dots),$$

que se describe a continuación.

- b) Si  $z_{t+1} = 0$ , entonces, el único valor admisible para  $z'_t$  es que sea igual a 1.
- c) De manera similar, si  $z_{t+1} = -1$ , entonces, la única opción admisible para  $z'_t$  es siendo igual a 1.

ii) El caso  $\lambda'_{t-1} < \lambda_{t+1}$  combinado con el valor actual de  $z_{t+1}$  da lugar al siguiente esquema de muestreo:

- a) Si  $z_{t+1} = 1$ , entonces  $z'_t = -1$ .
- b) Si  $z_{t+1} = 0$ , entonces  $z'_t = 0$ .
- c) Si  $z_{t+1} = -1$ , entonces  $z'_t = 1$ .

iii) Finalmente, el caso  $\lambda'_{t-1} > \lambda_{t+1}$  combinado con el valor actual de  $z_{t+1}$  da lugar al siguiente esquema de muestreo:

- a) Si  $z_{t+1} = 1$ , entonces  $z'_t = -1$ .
- b) Si  $z_{t+1} = 0$ , entonces  $z'_t = -1$ .
- c) De manera similar, si  $z_{t+1} = -1$ , entonces se requiere tomar un muestra de  $z'_t$  en el conjunto  $\{-1, 0, 1\}$  de acuerdo con la distribución

$$\pi(z_t|\mu, z, \boldsymbol{\lambda}),$$

que se da en adelante.

Con respecto a los casos donde se requiere obtener una muestra de  $z'_t$  del conjunto  $\{-1, 0, 1\}$ , la distribución condicional completa  $\pi(z_t|\mu, z, \boldsymbol{\lambda})$  se definirá típicamente por

$$\pi(z_t|\mu, z, \boldsymbol{\lambda}) \propto g_{1,z_t}(\lambda_t|\lambda_{t-1}; \theta, \mu)p(z_t|\lambda_{t-1}; \theta, \mu). \quad (\text{C.15})$$

La forma funcional de esta distribución condicional completa será sensitiva según cada caso. Los casos a explorar son: i)  $\lambda_t < \lambda_{t-1}$ , ii)  $\lambda_t = \lambda_{t-1}$ , y iii)  $\lambda_t > \lambda_{t-1}$ .

i) El caso  $\lambda_t < \lambda_{t-1}$  da lugar a la siguiente distribución condicional completa para  $z'_t$ ,

$$z'_t | \lambda_t, \lambda_{t-1}, \theta, \mu = \begin{cases} -1 & \text{con prob. } \propto g_{1,-1}(\lambda_t | \lambda_{t-1}; \theta, \mu) p(z'_t = -1 | \lambda_{t-1}, \theta, \mu) \\ 0 & \text{con prob. } \propto p(z'_t = 0 | \lambda_{t-1}, \theta, \mu) \\ 1 & \text{con prob. } \propto p(z'_t = 1 | \lambda_{t-1}, \theta, \mu). \end{cases} \quad (\text{C.16})$$

ii) El caso  $\lambda_t = \lambda_{t-1}$  da lugar a la distribución condicional completa dada por,

$$z'_t | \lambda_t, \lambda_{t-1}, \theta, \mu, y_t = \begin{cases} -1 & \text{con prob. } \propto p(z'_t = -1 | \lambda_{t-1}, \theta, \mu) \\ 0 & \text{con prob. } \propto N(y_t | 0, \lambda_{t-1}) p(z'_t = 0 | \lambda_{t-1}, \theta, \mu) \\ 1 & \text{con prob. } \propto p(z'_t = 1 | \lambda_{t-1}, \theta, \mu). \end{cases} \quad (\text{C.17})$$

iii) La distribución condicional completa que corresponde al tercer caso, es decir,  $\lambda_t > \lambda_{t-1}$ , resulta en ,

$$z'_t | \lambda_t, \lambda_{t-1}, \theta, \mu = \begin{cases} -1 & \text{con prob. } \propto p(z'_t = -1 | \lambda_{t-1}, \theta, \mu) \\ 0 & \text{con prob. } \propto p(z'_t = 0 | \lambda_{t-1}, \theta, \mu) \\ 1 & \text{con prob. } \propto g_{1,1}(\lambda_t | \lambda_{t-1}; \theta, \mu) p(z'_t = 1 | \lambda_{t-1}, \theta, \mu). \end{cases} \quad (\text{C.18})$$

Tales expresiones son validas para  $t = 2, \dots, T$ . Nótese que en todas las expresiones, las funciones  $p(z|\lambda, \theta, \mu)$  y  $g_{1,z}(\lambda_t | \lambda_{t-1}; \mu, \theta)$  se especifican de acuerdo con las definiciones utilizadas anteriormente.

2) Actualizar  $z_t$  en el caso de tener  $d'_{t-1} = d'_t$  y  $d'_t \neq d'_{t+1}$  se puede realizar al calcular las probabilidades dadas anteriormente que dependen del parámetro actual  $\theta_{d_t}$ .

## C.6. Distribución condicional completa para $\lambda$

La distribución condicional completa para  $\lambda$  se caracteriza por las distribuciones condicionales completas individuales para  $\lambda_t$ , cuya especificación y esquema de actualización será sensitiva a cada caso de acuerdo con las distintas combinaciones inducidas por las variables de índice latentes actuales ( $d_{t-1}, d_t, d_{t+1}$ ): I)  $d_{t-1} \neq d_t$  y II)  $d_{t-1} = d'_t$ , para  $t = 2, \dots, T$ .

Las distribuciones condicionales completas y el esquema de muestreo de actualización, inducidas por dichos casos se pueden dividir en:

I) Para los casos  $d_{t-1} \neq d_t$ , la distribución condicional completa para  $\lambda_t$  se dará por

$$p(\lambda_t|y_t; \theta_{d_t}) \propto N(y_t|0, \lambda_t) \text{We}(\lambda_t|\theta_{d_t}).$$

Actualizar  $\lambda_t$  se puede realizar utilizando muestreador de la rebanada.

II) El caso  $d_{t-1} = d_t$  será sensitiva a la relación adicional entre  $d_t$  y  $d_{t+1}$ . Por lo tanto, actualizar  $\lambda_t$  en este caso se puede realizar de acuerdo con:

IIa) El caso de tener  $d_t \neq d_{t+1}$ , para el cual  $\lambda_t$  actualizada puede obtenerse de la distribución condicional

$$p(\lambda_t|y_t, z_t, \lambda_{t-1}; \theta_{d_t}) \propto N(y_t|0, \lambda_t) g_{1,z_t}(\lambda_t|\lambda_{t-1}; \mu, \theta_{d_t}),$$

donde  $g_{1,z_t}(\lambda_t|\lambda_{t-1}; \mu, \theta_{d_t})$  ya está dada.

IIb) El caso de tener  $d_t = d_{t+1}$  inducirá la siguiente distribución condicional completa para  $\lambda_t$ ,

$$p(\lambda_t|y_t, z_t, z_{t+1}, \lambda_{t-1}, \lambda_{t+1}; \theta_{d_t}) \propto N(y_t|0, \lambda_t) g_{1,z_t}(\lambda_t|\lambda_{t-1}; \mu, \theta_{d_t}) \\ \cdot g_{1,z_{t+1}}(\lambda_{t+1}|\lambda_t; \mu, \theta_{d_t}) p(z_{t+1}|\lambda_t; \mu, \theta_{d_t}),$$

con  $g_{1,z_t}(\lambda_t|\lambda_{t-1}; \mu, \theta_{d_t})$  y  $p(z_{t+1}|\lambda_t; \mu, \theta_{d_t})$  también dadas.

En los casos anteriores, actualizar  $\lambda_t$  requerirá pasos de muestreador de la rebanada.

### C.6.1. Detalles adicionales acerca de la distribución condicional completa para $\lambda_t$

Debido a la manera en que se define cada  $z_t$  en el modelo, y que en cada  $t$  la variable latente previa  $\lambda_{t-1}$  divide el soporte de  $\lambda_t$ , y de manera similar  $\lambda_t$  divide el soporte de  $\lambda_{t+1}$ , las distribuciones condicionales completas derivadas anteriormente son realmente sensitivas según cada caso. Los casos a explorar dependen de  $(\lambda'_{t-1}, z'_t, z_{t+1}, \lambda_{t+1}, y_t)$ , específicamente de las combinaciones distintas de acuerdo a la relación entre el par  $\lambda'_{t-1}$  y  $\lambda_{t+1}$ , es decir: i)  $\lambda'_{t-1} < \lambda_{t+1}$ , ii)  $\lambda'_{t-1} = \lambda_{t+1}$ , y iii)  $\lambda'_{t-1} > \lambda_{t+1}$ . Asociado con cada uno de los casos anteriores,

también se deberían explorar las combinaciones derivadas entre los valores que pueden tomar  $z'_t$  y  $z_{t+1}$ .

También, las distribuciones condicionales completas dependerán del periodo de tiempo en el cual se encuentra el muestreo. Consideremos la distribución condicional completa para los periodos  $t = 2, \dots, T - 1$ .

- i) El caso  $\lambda'_{t-1} < \lambda_{t+1}$  requerirá explorar cinco sub-casos distintos en términos de  $z$ 's: a)  $z'_t = -1$  y  $z_{t+1} = 1$ , b)  $z'_t = 0$  y  $z_{t+1} = 1$ , c)  $z'_t = 1$  y  $z_{t+1} = 1$ , d)  $z'_t = 1$  y  $z_{t+1} = 0$ , y e)  $z'_t = 1$  y  $z_{t+1} = -1$ , para los cuales es posible derivar las distribuciones condicionales completas:

- a) El sub-caso  $z'_t = -1$  y  $z_{t+1} = 1$  dará lugar a la distribución condicional completa para  $\lambda'_t$  dada por

$$p(\lambda'_t | \dots) \propto N(y_t | 0, \lambda'_t) \frac{g_\mu(\lambda'_t | \lambda'_{t-1}) \mu_k(\lambda'_t) S_\mu(\lambda'_t)}{S_0(\lambda'_t)} \mathbf{1}(\lambda'_t \in (0, \lambda'_{t-1})).$$

- b) La combinación  $z'_t = 0$  y  $z_{t+1} = 1$  dará lugar a la distribución condicional completa degenerada, dada por,

$$p(\lambda'_t | \dots) \propto \delta_{\lambda_{t-1}}(\lambda'_t).$$

- c) La combinación  $z'_t = 1$  y  $z_{t+1} = 1$  induce la siguiente distribución condicional completa,

$$p(\lambda'_t | \dots) \propto N(y_t | 0, \lambda'_t) \frac{g_0(\lambda'_t) \mu_k(\lambda'_t) S_\mu(\lambda'_t)}{S_0(\lambda'_t)} \mathbf{1}(\lambda'_t \in (\lambda'_{t-1}, \lambda_{t+1})).$$

- d) La combinación  $z'_t = 1$  y  $z_{t+1} = 0$  dará lugar a la distribución condicional completa degenerada, dada por,

$$p(\lambda'_t | \dots) \propto \delta_{\lambda_{t+1}}(\lambda'_t).$$

- e) El último sub-caso corresponde a la combinación  $z'_t = 1$  y  $z_{t+1} = -1$ , la cual inducirá la siguiente distribución condicional completa

$$p(\lambda'_t | \dots) \propto N(y_t | 0, \lambda'_t) g_0(\lambda'_t) \mathbf{1}(\lambda'_t \in (\lambda_{t+1}, \infty)).$$

- ii) El segundo caso para las  $\lambda$ 's, es decir, cuando  $\lambda'_{t-1} = \lambda_{t+1}$ , requiere la exploración de tres sub-casos de  $z$ 's, es decir: i)  $z'_t = -1$  y  $z_{t+1} = 1$ , b)  $z'_t = 0$  y  $z_{t+1} = 0$ , y iii)  $z'_t = 1$  y  $z_{t+1} = -1$ . Explorémoslos en detalle al asumir que  $\hat{\lambda} = \lambda'_{t-1} = \lambda_{t+1}$ .

a) El sub-caso  $z'_t = -1$  y  $z_{t+1} = 1$  induce la siguiente distribución condicional para  $\lambda'_t$ ,

$$p(\lambda'_t | \dots) \propto N(y_t | 0, \lambda'_t) \frac{g_\mu(\lambda'_t | \lambda'_{t-1}) \mu_k(\lambda'_t) S_\mu(\lambda'_t)}{S_0(\lambda'_t)} \mathbf{1}(\lambda'_t \in (0, \hat{\lambda})).$$

b) La combinación  $z'_t = 0$  y  $z_{t+1} = 0$  inducirá la siguiente distribución degenerada

$$p(\lambda'_t | \dots) \propto \delta_{\hat{\lambda}}(\lambda'_t).$$

c) El último sub-caso corresponde a (ii) es asociado con la combinación  $z'_t = 1$  y  $z_{t+1} = -1$ , que inducirá la siguiente distribución condicional

$$p(\lambda'_t | \dots) \propto N(y_t | 0, \lambda'_t) g_0(\lambda'_t) \mathbf{1}(\lambda'_t \in (\hat{\lambda}, \infty)).$$

iii) Como en el caso (i), la combinación  $\lambda'_{t-1} > \lambda_{t+1}$  tendrá asociada cinco sub-casos distintos en términos de  $z$ 's. Es decir: a)  $z'_t = -1$  y  $z_{t+1} = 1$ , b)  $z'_t = -1$  y  $z_{t+1} = 0$ , c)  $z'_t = -1$  y  $z_{t+1} = -1$ , d)  $z'_t = 0$  y  $z_{t+1} = -1$ , y e)  $z'_t = 1$  y  $z_{t+1} = -1$ , que inducen las siguientes distribuciones condicionales completas:

a) El sub-caso  $z'_t = -1$  y  $z_{t+1} = 1$  dará lugar a una distribución condicionales completas para  $\lambda'_t$  dada por

$$p(\lambda'_t | \dots) \propto N(y_t | 0, \lambda'_t) \frac{g_\mu(\lambda'_t | \lambda'_{t-1}) \mu_k(\lambda'_t) S_\mu(\lambda'_t)}{S_0(\lambda'_t)} \mathbf{1}(\lambda'_t \in (0, \lambda_{t+1})).$$

b) La combinación  $z'_t = -1$  y  $z_{t+1} = 0$  dará lugar a una distribución degenerada, dada por

$$p(\lambda'_t | \dots) \propto \delta_{\lambda_{t+1}}(\lambda'_t).$$

c) La combinación  $z'_t = -1$  y  $z_{t+1} = -1$  inducirá la siguiente distribución

$$p(\lambda'_t | \dots) \propto N(y_t | 0, \lambda'_t) g_\mu(\lambda'_t) \mathbf{1}(\lambda'_t \in (\lambda_{t+1}, \lambda'_{t-1})).$$

d) La combinación  $z'_t = 0$  y  $z_{t+1} = -1$  dará lugar a otra distribución degenerada

$$p(\lambda'_t | \dots) \propto \delta_{\lambda'_{t-1}}(\lambda'_t).$$

e) El sub-caso final a explorar en (i) corresponde a una combinación  $z'_t = 1$  y  $z_{t+1} = -1$ , que inducirá la siguiente distribución condicional

$$p(\lambda'_t | \dots) \propto N(y_t | 0, \lambda'_t) g_0(\lambda'_t) \mathbf{1}(\lambda'_t \in (\lambda'_{t-1}, \infty)).$$



Entonces, para el periodo  $t = T$  la distribución condicional completa de  $\lambda'_T$  se define en términos de tres casos inducidos por  $z'_T$ . Es decir: a)  $z'_T = -1$ , b)  $z'_T = 0$ , y c)  $z'_T = 1$ . Las distribuciones condicionales completas para estos casos se definen abajo.

a) Para el caso  $z'_T = -1$ , la distribución condicional completa para  $\lambda'_T$  será,

$$p(\lambda'_T | \dots) \propto N(y_T | 0, \lambda'_T) g_\mu(\lambda'_T | \lambda'_{T-1}) \mathbf{1}(\lambda'_T \in (\lambda'_{T-1}, \infty)).$$

b) El caso  $z'_T = 0$  da lugar a la distribución degenerada para  $\lambda'_T$ ,

$$p(\lambda'_T | \dots) \propto \delta_{\lambda'_{T-1}}(\lambda'_T).$$

c) Finalmente, el caso  $z'_T = 1$  da lugar a la siguiente distribución condicional completa para  $\lambda'_T$ ,

$$p(\lambda'_T | \dots) \propto N(y_T | 0, \lambda'_T) g_0(\lambda'_T) \mathbf{1}(\lambda'_T \in (\lambda'_{T-1}, \infty)).$$

En el periodo  $t = 1$ , la distribución condicional completa para  $\lambda'_1$  será

$$p(\lambda'_1 | \dots) \propto N(y_1 | 0, \lambda'_1) g_0(\lambda'_1) \mathbf{1}(\lambda'_1 \in (0, \infty)).$$

Nótese que ninguna de las distribuciones condicionales completas no degeneradas tiene una forma analítica cerrada de acuerdo con su soporte. Por lo tanto, generar muestras de ellas debería requerir esfuerzos adicionales. Entre alternativas distintas, aquí se adopta el método de muestreo de la rebanada (“slice sampler”) para tomar muestras de estas distribuciones (véase, [Neal, 2003](#), para una introducción general a los métodos de muestreo de la rebanada).