

Analítica

Propuesta de modelación basada en un enfoque
de redes probabilísticas: una aplicación a la
consistencia macroeconómica

Jaime Fernández



www.ecuadorencifras.gob.ec



Propuesta de modelación basada en un enfoque de redes probabilísticas: una aplicación a la consistencia macroeconómica¹

Jaime Fernández[†]

[†] Facultad Latinoamericana de Ciencias Sociales, Quito, Ecuador.

Resumen

Un marco de consistencia macroeconómica permite evaluar la factibilidad de ocurrencia simultánea de un conjunto de previsiones o supuestos macroeconómicos. Desde esta perspectiva se han desarrollado diversas aproximaciones, incluyendo: cuadros de consistencia contable, matrices de contabilidad social, modelos de equilibrio general, entre otros. En este documento se presenta un enfoque alternativo basado en modelos probabilísticos gráficos con variables latentes. Estas potentes herramientas estadísticas permiten recuperar, a partir de series históricas de datos, las relaciones de interdependencia existentes entre un grupo de variables aleatorias. Su principal ventaja es que no requieren fijar supuestos ex ante sobre el comportamiento de los agentes que componen un sistema; en contraparte, permiten testear hipótesis o teorías económicas a la luz de evidencias empíricas. En la investigación se prueban dos algoritmos y se construyen algunos escenarios para lograr el mejor “aprendizaje” de los modelos. Finalmente, se ensaya un ejercicio de inferencia sobre la tasa de crecimiento del PIB del Ecuador.

Palabras clave: consistencia macroeconómica, modelos probabilísticos gráficos, modelos de árbol, variables latentes

Abstract

A macroeconomic consistency framework assesses the possibility of the simultaneous occurrence of a set of macroeconomic forecasts or assumptions. For this, there are several approaches, including: accounting consistency charts, social accounting matrices, general equilibrium models, among others. This paper shows an alternative approach based on probabilistic graphical models and latent variables. These powerful

¹El autor agradece los valiosos comentarios al documento realizados por Wilson Pérez, Pablo Samaniego, Christian Rojas y los árbitros que han colaborado con la Revista Analitika en la revisión del artículo.

[†] jfernandez@innovar-ec.com

Statistical tools can retrieve, from historical data series, the interdependence existing within a group of random variables. The principal advantage is that they do not require making ex-ante assumptions about the behavior of the agents in the system. These tools are useful for testing hypotheses and economic theories in light of empirical evidence. Two algorithms are tested and several scenarios are built in order to achieve the best “learning” performance in the models. Finally, an inference exercise on Ecuador’s GDP growth rate is made

Palabras clave: macroeconomic consistency, probabilistic graphical models, tree models, latent variables.

Código JEL: E61, C11, C16, C19.

1 Introducción

En esta sección se presenta el estado actual de los desarrollos teóricos y empíricos relacionados con la consistencia macroeconómica, así como el contexto en el cual se ha originado la investigación. La primera parte muestra una breve reseña de la evolución de los modelos que se han utilizado para aproximarse a la consistencia macroeconómica, así como algunos aspectos que orientan el nuevo enfoque que se plantea en este documento. La segunda parte describe el alcance y los objetivos del estudio presentado en este documento.

La dinámica de una economía está en constante evolución, la modelación sobre esta área implica aceptar la complejidad de la predicción de sus estados futuros. La estrecha interrelación que existe entre los componentes de un sistema económico origina causalidades circulares, problemas de endogeneidad o procesos de retroalimentación. Estas interrelaciones pueden generar procesos de convergencia, transiciones o cambios estructurales en el sistema. Inclusive, éstas pueden originar comportamientos caóticos, es decir, exhibir trayectorias substancialmente divergentes explicadas por cambios relativamente pequeños en ciertos parámetros.

En este contexto de resultados tan amplios que se podrán generar a partir de diferentes “configuraciones” de un sistema económico, resulta fundamental contar con herramientas que permitan evaluar la factibilidad de ocurrencia de determinados estados futuros de dicho sistema, dado un conjunto de supuestos sobre el mismo. Un marco de consistencia macroeconómica permite, precisamente, evaluar la concordancia que tienen las diferentes previsiones de los grandes agregados macroeconómicos que, en el caso ecuatoriano, son realizadas por diferentes organismos estatales con altos niveles de asimetría de información entre sí. En este sentido, es importante determinar si las previsiones (o metas) que forman parte de la planificación macroeconómica de un país son factibles de ocurrir simultáneamente.

En Ecuador, de acuerdo a la institucionalidad existente en la actualidad, la responsabilidad de generar previsiones macroeconómicas para la proforma presupuestaria se encuentra

repartida entre el Ministerio de Coordinación de la Política Económica (sector financiero), Ministerio Coordinador de la Producción, Empleo y Competitividad (sector externo) Banco Central del Ecuador (sector real) y Ministerio de Finanzas (sector fiscal), principalmente. Estas instituciones, si bien es cierto, interactúan en cierta medida para la generación de previsiones conjuntas, no necesariamente estas son coherentes o verosímiles simultáneamente. Por otro lado, aun cuando las previsiones pueden haber sido generadas, de manera aislada, utilizando criterio de experto, los modelos tradicionales no necesariamente permiten incorporar este tipo de criterios de experto en el análisis conjunto de consistencia macroeconómica.

Un rápido vistazo a la historia de las previsiones macroeconómicas realizadas en Ecuador permite observar las claras limitaciones de las metodologías utilizadas. En el mejor de los casos, se han desarrollado unos pocos escenarios para los resultados macroeconómicos en función de unos cuantos parámetros, generalmente el precio del petróleo. El hecho de tener previsiones macroeconómicas con estimaciones puntuales hace que se pierda el foco de lo que podría suceder en los valores extremos de los parámetros, o en puntos intermedios entre los pocos escenarios que se generan. Este hecho representa un limitante para la toma de decisiones y la formulación de medidas de política, por ejemplo contracíclicas, ya que poco o nada se puede conocer respecto a un amplio espectro de escenarios que podrían resultar del movimiento simultáneo de varios parámetros. La elección del modelo adecuado para aproximarse al problema de la consistencia macroeconómica es particularmente importante. Como se ilustrará en la siguiente sección, históricamente no ha sido una tarea sencilla contar con herramientas matemáticas/estadísticas adecuadas que permitan abordar el problema de manera eficaz. La evolución de estas áreas del conocimiento, instrumentales a la economía, ha permitido también mejorar en cuanto a la bondad que los modelos escogidos ofrecen para explicar los marcos de consistencia macroeconómica.

1.1 El problema de la consistencia

Los primeros modelos de consistencia fueron los llamados modelos multisectoriales, orientados principalmente al sector real de la economía (Pérez y Acosta, 2005). Estos modelos suponían el cumplimiento del principio de la demanda efectiva y, la existencia de una oferta perfectamente elástica y una tecnología constante y plana. Debido a la relativa simplicidad de estos modelos, que intentaban reproducir un sistema económico en función de insumos y productos, permitían inferir una gran cantidad de información acerca de las interrelaciones de los agentes, aunque generalmente poco apegadas a su comportamiento real en el mercado.

El desarrollo de la informática hizo posible el diseño de modelos de simulaciones y resolución simultánea de ecuaciones basados en teorías económicas más elaboradas (incluyendo ecuaciones de comportamiento) que demandaban mayor poder de cálculo. Esto representaría el origen de los modelos de equilibrio general computable (MEGC), basados en el enfoque desarrollado tiempo atrás por Walras (1874). En estos modelos se consideran a los precios como argumentos dentro de las funciones de oferta y demanda, por lo cual, los precios y las

cantidades se definen simultáneamente en todos los mercados, ya que la función de producción depende también de los precios de los insumos que se definen en estos.

Tanto los modelos multisectoriales como los modelos de equilibrio general han sido ampliamente criticados, no solamente por las bases teóricas que los sustentan, sino por la rigidez de los supuestos sobre los cuales se asientan. Además, en el caso de los MEGC, la falta de disponibilidad de series históricas largas evita que se puedan realizar estimaciones de todos los parámetros que el modelo demanda, exigiendo así la utilización de parámetros o coeficientes técnicos “importados” o tomados de la literatura.

Además de las argumentaciones teóricas que se puedan plantear en torno a los modelos tradicionales de consistencia macroeconómica, en Ecuador, la disponibilidad y calidad de la información es un problema de fondo. Como se estableció en la sección introductoria, las previsiones macroeconómicas que vienen de las diversas instituciones del Estado, a cargo de realizarlas, utilizan diversas fuentes de información. Para algunas variables macroeconómicas existe poca o nula información y, para su estimación, se deben utilizar variables “proxy”. Otro grupo de variables de este tipo cuenta con varias fuentes de información que permiten su cálculo o estimación, pero estas no necesariamente son compatibles entre sí.

1.2 La propuesta metodológica

Los argumentos presentados en la sección anterior reflejan la necesidad de contar con una herramienta que permita aproximarse al problema de la consistencia macroeconómica de modo que, a la vez, minimice la cantidad y rigidez de los supuestos utilizados y reproduzca las principales relaciones macroeconómicas de un sistema con la mayor precisión posible. Además, es deseable que dicha herramienta permita la incorporación del criterio de experto en la evaluación de consistencia de los distintos escenarios que pudieran plantearse. Estas son, precisamente, las características que tiene la herramienta que se plantea utilizar en esta investigación para abordar la consistencia macroeconómica en Ecuador.

La base teórica sobre la que se asienta el enfoque utilizado corresponde a los modelos probabilísticos gráficos. Estas herramientas estadísticas permiten modelar sistemas complejos a partir de las relaciones observadas entre el conjunto de variables que conforman dichos sistemas. Además, una importante ventaja -y a la vez un fuerte justificativo de trabajar con estas herramientas en economía-, es que no se requiere de supuestos *ex ante* acerca del comportamiento de los agentes, sean estos en forma de funciones, ecuaciones o coeficientes técnicos.

En base a las consideraciones anteriores, el alcance del presente trabajo conlleva la aplicación de un nuevo enfoque (probabilístico Bayesiano y de redes) a un área de la economía: la evaluación de consistencia macroeconómica. Sin embargo, no se puede desconocer que ésta, al igual que otras metodologías existentes, tiene limitaciones, mismas que deben estar siempre presentes en la interpretación de resultados y, principalmente, la formulación de política pública. El marco de consistencia contable que se utiliza para la selección de variables y la

En el diseño de la política económica de un país es particularmente importante evaluar ex ante el efecto que una o varias medidas que se plantea adoptar tendrá en todo el aparato económico. Esto significa que se deben evaluar los posibles efectos que tendrían las diferentes opciones de política sobre la evolución de la economía en su conjunto. Esta tarea no necesariamente resulta sencilla, y por eso es que se debe recurrir a herramientas que ayuden a evaluar dichos efectos. Un modelo de consistencia macroeconómica, si bien no tiene como fin evaluar dichos impactos, permite rastrear las relaciones dentro del sistema económico que pudieran verse afectadas por la toma de medidas de política.

Por la naturaleza y complejidad que envuelve a un sistema económico definido sobre agentes sociales, no se pueden esperar previsiones ni evaluaciones de política suficientemente objetivas basadas solamente en el procesamiento estadístico de la información, o en modelos econométricos, debiendo integrarse el conocimiento y la experiencia recogida en los criterios de experto. La visión de consistencia macroeconómica adoptada en este trabajo también permite incorporar dicho conocimiento a priori.

Otra arista del enfoque conceptual utilizado parte de la crítica a la economía convencional, en cuanto adopta un enfoque sistémico que concibe a la economía como parte integrante del sistema socio-económico-político que evoluciona junto a un contexto institucional específico. En este sentido, se distingue de la corriente neoclásica, que mira a lo social como la suma de agentes homogéneos de conocimiento perfecto, que solo interactúan en el mercado, y más bien entiende al individuo como dotado de conocimiento local que responde a la incertidumbre con reglas de comportamiento no necesariamente consistentes con la optimización de alguna función objetiva o con la trayectoria definida por la solución de un conjunto de ecuaciones.

La única manera de conocer el estado futuro de una economía es observar efectivamente su evolución. Así que, si bien las herramientas de consistencia macroeconómica buscan determinar con un margen de precisión prudente la racionalidad de los escenarios económicos que se evalúen, no es de esperarse que éstas se consideren instrumentos formales de previsión. Esto es especialmente cierto para un sistema sujeto a procesos de sustanciales cambios, como ha estado ocurriendo en los últimos años con el sistema económico en Ecuador.

La idea central cuando se refiere a la consistencia macroeconómica es determinar la coherencia de determinados escenarios económicos planteados, ya sea como previsiones o como potenciales efectos de la política pública. Es decir, se trata de mirar el todo en su interacción con las partes, descartar trayectorias imposibles por inconsistentes, integrar el análisis que desde varias perspectivas se hace de la trayectoria socialmente deseable, y desde esa base definir y priorizar preguntas de investigación relevantes.

Con este antecedente, y citando a Dutt (2011), en concordancia con lo planteado por Easterly (1989), el punto de partida en toda evaluación de consistencia macroeconómica es la definición de “identidades contables que involucran flujos y niveles (incluyendo aquellas que enseñan cómo los niveles cambian debido a los flujos) y posteriormente añade ecuaciones que representan relaciones institucionales y de comportamiento relevantes que involucran variables de nivel y de flujos, de tal manera que podemos determinar los valores de las

variables del sistema dados sus parámetros”. Estas identidades deben incluir las variables de política que se quiere emplear.

Esto no significa asumir que estas identidades, la definición de las variables incluidas y sus mecanismos de medición, están objetivamente dados y además correctamente medidos (siendo esto último de particular importancia). Es verdad que estas definiciones están condicionadas por el contexto social, político e incluso histórico, en donde se desarrollan y que se reflejan, por ejemplo, en las cada vez más abundantes bifurcaciones de las corrientes del pensamiento económico. Por otro lado, la interacción con estos instrumentos de consistencia permite también encontrar nuevas definiciones y variables que, a menudo, no son medidas con el suficiente cuidado, pero que deberían serlo.

El último pilar teórico sobre el que se asienta el presente trabajo es el planteamiento de Bacha (1990) y su modelo de tres brechas con comportamientos no necesariamente racionales de los agentes (institucionales) que deben actuar ante escenarios de incertidumbre. Este modelo permite “rastrear” las restricciones fiscales y externas que frenan un potencial crecimiento más alto de las economías en desarrollo, en el mediano plazo. En el planteamiento de Bacha, se puede construir un marco de consistencia macroeconómica “cerrando” por el lado de la brecha privada (diferencia entre el ahorro y la inversión privada) a través del balance fiscal y la cuenta corriente.

2.2 Enfoque bayesiano

El teorema de Bayes (1763) se constituye en el punto de ruptura en la visión de la estadística tradicional. La estadística tradicional o frecuentista únicamente admite la estimación de probabilidades en función de experimentos que puedan ser repetibles y replicables, mientras que en la estadística bayesiana se concibe el concepto de probabilidades subjetivas a priori, mismas que pueden ser modificadas cuando se reciba información adicional o nueva evidencia empírica para tener probabilidades más “refinadas”. Esta revolución en la forma de hacer estadística ha permitido ampliar el horizonte de la modelación matemática, permitiendo combinar distintas fuentes de información para obtener mejores resultados en las estimaciones. Esta es la visión que, desde el lado de la estadística, guía esta investigación en las partes que corresponda.

Dadas estas ventajas de utilizar un enfoque bayesiano en la metodología propuesta en esta investigación y el nivel de desarrollo que la teoría estadística y las herramientas computacionales han alcanzado, es claro que un enfoque de consistencia macroeconómica planteado desde esta visión puede mejorar la explicación de diversos fenómenos económicos e incluso arrojar información de sobre nuevos campos dentro de esta área que hasta hoy no han sido explotados.

3 Modelos gráficos

3.1 Antecedentes

En esta sección se presenta una visión general de los principales modelos estadísticos gráficos existentes, sus aplicaciones y su relación entre sí. El objetivo de esta sección es justificar la utilización de modelos probabilísticos gráficos para modelar el marco de consistencia macroeconómica deseado y explicar sus principales aspectos teóricos. Además, se presentará el algoritmo BP (belief propagation), el cual será utilizado para abordar los problemas de inferencia que aparecerán en el modelamiento.

Los modelos probabilísticos gráficos son potentes herramientas estadísticas que permiten recuperar, a partir de series de datos, las relaciones de dependencia condicional existentes entre el grupo de variables que conforman un sistema, y éstas son representadas en un grafo. Una de las principales ventajas que ofrecen los modelos gráficos es la potencialidad de representar la distribución conjunta de todas las variables del sistema de manera compacta. Las redes bayesianas (grafos dirigidos) y los campos aleatorios de Markov (grafos no dirigidos) son los dos tipos de modelos gráficos más utilizados. Además, existen otros modelos gráficos un tanto menos difundidos pero igualmente generalizables como los grafos de factores y los grafos de Tanner. Si bien existen equivalencias matemáticas entre los distintos modelos que se presentarán, no necesariamente existen interpretaciones económicas “formales” para la dinámica de modelación de cada uno de ellos. Todos estos modelos adicionales se presentan con la única intención de demostrar la generalidad y variedad en cuanto a la aplicabilidad de los modelos probabilísticos gráficos.

Entre las principales propiedades que hacen de los modelos gráficos herramientas útiles, se pueden citar las siguientes:

- Permiten visualizar de manera simple la estructura de un modelo probabilístico y pueden ser utilizados como insumo para el diseño de nuevos modelos.
- Ciertos cálculos que podrían resultar complejos cuando se quiere realizar inferencia y aprendizaje en modelos un tanto sofisticados, pueden ser expresados de manera más sencilla como transformaciones gráficas de los modelos, que llevan consigo expresiones matemáticas de manera implícita.
- Con una simple inspección del modelo gráfico, se pueden obtener ideas bastante claras acerca de las propiedades del mismo, incluyendo, por ejemplo, las propiedades de independencia condicional de las variables consideradas.

Definición 1. *Un grafo es un conjunto de nodos (usualmente llamados vértices) y aristas (conocidas también como arcos). En los modelos gráficos probabilísticos, los nodos representan variables aleatorias y las aristas las relaciones probabilísticas existentes entre las mismas.*

La estructura de un grafo refleja de manera compacta la distribución conjunta de las variables aleatorias consideradas descompuestas como un producto de factores, cada uno de los cuales depende solo de un subconjunto de éstas variables, utilizando el teorema de factorización de probabilidades y la propiedades de independencia condicional (ver Anexos). Un concepto muy importante en esta perspectiva teórica es el de independencia condicional:

Definición 2. *Dos variables aleatorias X e Y se dicen condicionalmente independientes, dada una tercera variable aleatoria Z si:*

$$P(X, Y|Z) = P(X|Z)P(Y|Z)$$

3.2 Redes bayesianas

Una red bayesiana es un grafo dirigido (las aristas tienen dirección) acíclico (para ningún nodo existe una secuencia de aristas que comience y termine en él) en el que los nodos pueden ser variables aleatorias continuas, discretas, observables o no observables. Las aristas reflejan la relación de dependencia entre unas variables y otras.

Fácilmente se podría asegurar que las redes bayesianas son los modelos gráficos mas populares, pues son utilizadas para modelar sistemas expertos que involucran áreas como: diagnósticos médicos, aprendizaje en red, reconocimiento de voz, búsqueda heurística, etc.

Para entender de mejor manera el concepto, la estructura y la utilidad de una red bayesiana, se desarrollaría un ejemplo clásico (Lauritzen y Spiegelhalter, 1988) en el contexto del diagnóstico médico.

Ejemplo 1. *Supóngase que se desea construir una máquina que devuelva diagnósticos automáticamente para cada paciente, dada una serie de síntomas y resultados de exámenes médicos. Con esta información, se desea inferir la posibilidad de que una o varias enfermedades dadas sean las causantes de estos síntomas y resultados de exámenes médicos. Se supone conocida una estructura de relaciones probabilísticas entre los diferentes síntomas, exámenes médicos y enfermedades que, sin pérdida de generalidad, puede ser construida en base al criterio de experto (en este caso médico).*

Para el desarrollo del ejemplo, se definen las siguientes variables:

A *Haber viajado recientemente a Asia*

T *Tener tuberculosis*

F *Tener el hábito de fumar*

C *Tener cáncer de pulmón*

B *Tener bronquitis*

E *Tener alguna de las dos enfermedades: tuberculosis o cáncer de pulmón*

X Obtener un resultado positivo respecto mediante un examen de rayos X

D Tener disnea (falta de aire)

Además, se necesitan los lineamientos provenientes del criterio de experto para definir la estructura de relaciones probabilísticas entre las variables:

- *A* aumenta la probabilidad de *T*.
- *F* es un factor de riesgo para *C* y *B*.
- *E* puede ser detectado a través de *X* (no se puede distinguir de cual de las dos enfermedades se trata).
- *D* puede ser causada por *B* o *E*.

Dada da la información presentada arriba, el grafo asociado a este ejemplo se presenta en la Figura 1:

Para cada una de las variables consideradas en el ejemplo (listadas en letras mayúsculas) se definen, en letras, las variables asociadas que representan el estado de cada una de ellas. Para este ejemplo se asume que cada variable puede estar en un número finito de estados. Así, por ejemplo, *a* representa el estado de la variable *A*, es decir, *a* indica si el paciente ha viajado a o no recientemente a Asia. Esa misma lógica se aplica para el resto de variables. Como se dijo en la introducción, del grafo se pueden extraer las relaciones de probabilidad condicional asociadas a cada arista. Por ejemplo, $P(t|a)$ representa la probabilidad condicional de que un paciente tenga tuberculosis, dado que se conoce si ha viajado o no a Asia recientemente. En este caso se dice que *A* es un nodo padre de *T*. Los nodos de una grafo pueden tener varios nodos padre. El nodo *B*, por ejemplo tiene dos nodos padre: *E* y *F*. Entonces, la probabilidad condicional de tener disnea se escribiría como $P(d|e, b)$.

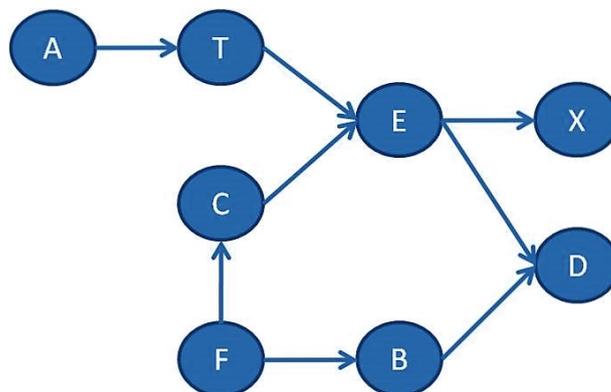


Figura 1: Grafo ejemplo diagnóstico médico

Utilizando el ejemplo planteado, se puede introducir la siguiente propiedad de las redes bayesianas: la probabilidad de que la variable asociada a un nodo se encuentre en un estado dado depende únicamente del estado de las variables que se encuentran en sus nodos padres. Otra propiedad que se puede ilustrar en este ejemplo es la de independencia condicional. Así, por ejemplo, C y B son condicionalmente independientes dada F , es decir, $P(b, c|f) = P(b|f)P(c|f)$. La intuición de esta afirmación sería que la probabilidad de tener cáncer de pulmón (C) no se ve afectada por el hecho de tener o no bronquitis (B), dado que se conoce si el paciente tiene o no el hábito de fumar (F), y viceversa.

En las redes bayesianas y, en general, en los modelos gráficos es deseable tener grafos dispersos, es decir, grafos en donde la mayor parte de los nodos no tengan una relación directa de dependencia probabilística, lo cual implica que no existan “demasiadas” aristas; los algoritmos, de solución exacta en unos casos y aproximada en otros, funcionan mejor y los tiempos de cálculo son eficientes cuando esto sucede.

Volviendo al ejemplo, se define la siguiente notación: $y = a, f, t, c, b, e, x, d$, es decir, y representa una combinación posible de síntomas, resultados de exámenes y enfermedades. Así, utilizando las propiedades de independencia condicional y el teorema de factorización de probabilidades, se tiene:

$$P(y) = P(a)P(f)P(t|a)P(c|f)P(b|f)P(e|t, c)P(d|b, e)P(x|e) \quad (1)$$

De manera general, en una red bayesiana formada por N variables: x_1, x_2, \dots, x_N , la función de probabilidad conjunta se puede escribir:

$$P(x_1, x_2, \dots, x_N) = \prod_{i=1}^N P(x_i|pa(x_i)) \quad (2)$$

donde $pa(x_i)$ se refiere a los estados de las variables que se encuentran en los nodos padre del nodo i :

Una vez estructuradas las redes, el objetivo principal será calcular las distribuciones marginales de algunas variables, a lo cual se conocerá como “inferencia” en este sentido. Estas probabilidades marginales para un nodo dado, en el caso de que todas las variables de la red tengan un número finito de estados, se calculan como la suma sobre todos los demás estados de las otras variables de la red.

En la red del ejemplo, una probabilidad marginal deseada sería la de tener cierta enfermedad, supóngase bronquitis. En este caso, la expresión para la distribución marginal sería:

$$P(b) = \sum_a \sum_f \sum_t \sum_c \sum_e \sum_x \sum_d P(a, f, t, c, b, e, x, d) \quad (3)$$

En el sentido tradicional del cálculo de las probabilidades marginales, esta sumatoria puede involucrar una cantidad demasiado grande de operaciones, sin embargo, al aprovechar

las propiedades antes mencionadas en una red dispersa (que es lo deseado), el número de operaciones y, por ende, el tiempo de cálculo se reducen notoriamente. En una red pequeña la inferencia exacta puede ser realizada en un tiempo razonable, pero a medida que aumenten los nodos en la red, el número de operaciones necesarias crecerá exponencialmente con el número de nodos, dejando de ser en cierto punto computacionalmente factible. Esta es precisamente la principal ventaja del algoritmo BP; se lo puede utilizar para calcular de manera aproximada las probabilidades marginales en una red en un tiempo que crece linealmente con el número de nodos de la red.

Aún aprovechando las ventajas de una red bayesiana, existen casos en los que la inferencia no se puede realizar de manera exacta. En este caso, las probabilidades marginales que se calculen de manera aproximada se conocerán como “beliefs”. La notación que se utilizará para el belief de un nodo i será $b(x_i)$.

En las redes bayesianas no siempre será necesario sumar sobre todas las variables, pues existe la posibilidad de que tengamos información a priori sobre el estado de algunas de ellas, estas variables se conocerán como “observables”. En el ejemplo anterior, se deseaba calcular la probabilidad de que el paciente tenga bronquitis (marginal), si además se conoce que el paciente no fuma, entonces dicha probabilidad marginal se reescribiría como:

$$P(b) = \sum_a \sum_t \sum_c \sum_e \sum_x \sum_d P(a, f, t, c, b, e, x, d) \quad (4)$$

A las demás variables se les conocería como “no observables”. Para distinguir gráficamente entre unas y otras se utilizará la siguiente codificación: los nodos de las variables observables serán círculos rellenos y los nodos de las no observables serán círculos vacíos.

3.3 Campos aleatorios de Markov

Un campo aleatorio de Markov (CAM) es, al igual que las redes bayesianas, un modelo estadístico gráfico. En este caso, las variables aleatorias representadas en los nodos poseen la propiedad Markoviana². En varios sentidos es similar a una red bayesiana: un campo aleatorio de Markov puede representar dependencias cíclicas (lo cual por definición de red bayesiana no es posible) pero no puede representar dependencias inducidas³ (una de las ventajas principales de una red bayesiana). El campo aleatorio de Markov clásico está representado por el modelo Ising (Dobrushin, 1968).

Las principales aplicaciones de estos campos aleatorios se encuentran enmarcadas en áreas como el procesamiento de imágenes y visión artificial: mejora de resolución, segmentación,

²La distribución de probabilidad condicional de los estados futuros de la variable, dado el estado presente y los estados pasados, depende únicamente del estado presente

³Si un nodo hijo es observado y tiene múltiples nodos padre, en una red bayesiana la información puede fluir entre dichos nodos padre, en cambio, en un campo aleatorio de Markov ese flujo de información no puede darse.

texturización y corrección de imágenes, reconocimiento de objetos, detección de movimiento, etc. En las citadas aplicaciones, la idea central es “enseñarle” al computador a realizar estas tareas en un tiempo razonable, dada únicamente una serie de arreglos bidimensionales con los valores de la intensidad de los píxeles. Un aspecto esencial para el desarrollo de la visión artificial consiste en encontrar modelos teóricos sólidos y computacionalmente factibles para llevar a cabo el aprendizaje y la inferencia. En este contexto, los CAM’s proveen dicho marco teórico para gran parte de los problemas involucrados en la visión artificial.

Para ilustrar la funcionalidad de los CAM’s, considérese un ejemplo en el que se desea inferir la distancia que hay desde el observador a los objetos que conforman cierta escena. Se dispone de una imagen en escala de grises de 1000 x 1000 píxeles y el problema consiste en inferir las distancias d_i correspondientes a los valores de intensidad de los píxeles $I_i, i = 1, 2, \dots, 1000000$; es decir, el índice recorre cada una de las posiciones de la matriz 1000 x 1000.

En el caso general, se asumirá que existen ciertas propiedades observables de la imagen notadas por y_i y que otras propiedades x_i deben ser inferidas. Los índices i no necesariamente representarán las posiciones de los píxeles individualmente, dependiendo de la aplicación podrán representar regiones vecinas de éstos. Se asume también la existencia de alguna dependencia estadística entre x_i y y_i para cada posición i . A dicha dependencia se le conocerá como la función de compatibilidad conjunta (también se le suele llamar la “evidencia” para x_i) y se la notará como $\phi(x_i, y_i)$. Por último, para que los problemas de visión artificial sean factibles de resolver, aunque sea de manera aproximada, es necesario hacer un supuesto acerca de la estructura para las x_i . Se codifica la estructura asumida de la escena suponiendo que los nodos i están organizados en una grilla bidimensional y que las variables x_i deberían, en la medida de lo posible, ser compatibles con las variables vecinas en la escena x_j . Dicha compatibilidad estará representada por la función notada $\psi_{ij}(x_i, x_j)$. Con estas consideraciones, se define la probabilidad conjunta total de una escena x (escondida) y una imagen y (observable) como:

$$P(\{x\}, \{y\}) = \frac{1}{z} \prod_{ij} \psi_{ij}(x_i, x_j) \prod_i \phi_i(x_i, y_i) \quad (5)$$

Donde Z es la constante de normalización y el producto sobre (ij) es sobre los vecinos más cercanos en el lattice cuadrado. La Figura 2 presenta una descripción gráfica del modelo planteado.

En la Figura 2, los círculos vacíos representan a los nodos ocultos (escena) x_i y los círculos rellenos representan a los nodos observables (imagen) y_i . Como se evidencia, en este modelo las aristas no son dirigidas, por lo cual, no existe la noción de causalidad o parentalidad entre nodos que si tiene lugar en las redes bayesianas. En lugar de dichas relaciones, en los CAM’s se dispone de la funciones de compatibilidad $\psi_{ij}(x_i, x_j)$ y no de las probabilidades condicionales $P(x_i|x_j)$.

Al igual que en el caso anterior, el objetivo es calcular los beliefs $b(x_i)$ para todas las posiciones i , es decir inferir las propiedades ocultas de la imagen. De manera similar, el

cálculo exacto las probabilidades marginales de los nodos crece exponencialmente con el número de ellos.

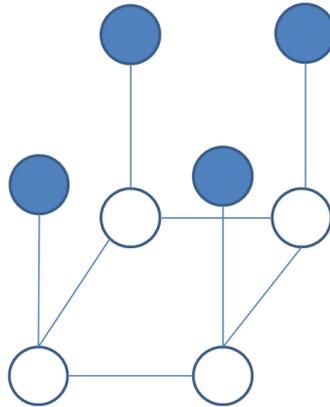


Figura 2: Campo aleatorio de Markov

3.4 El algoritmo BP (belief propagation)

Este algoritmo permite resolver de manera aproximada los problemas de inferencia que surgen a menudo en diferentes campos de la ciencia. Por este motivo, no es de sorprenderse que dicho algoritmo haya sido “descubierto” en diversas aplicaciones y con diversos nombres. Es así que, por ejemplo, los siguientes métodos pueden ser considerados casos particulares del algoritmo BP: algoritmo de Viterbi (encuentra la secuencia más probable de estados en una cadena de Markov oculta), algoritmo de avance-retroceso (calcula las distribuciones marginales a posteriori de las variables de estado ocultas en una cadena de Markov oculta), algoritmo de Pearl para redes bayesianas y el método de la matriz de transferencia (analiza la propagación de ondas acústicas o electromagnéticas). La explicación detallada del funcionamiento del algoritmo BP se presenta en la sección de Anexos.

3.5 Distancia de Kullback-Leibler

Todos los modelos gráficos (no solamente los presentados en esta sección) definen una función de probabilidad conjunta $P(\{x\})$. Si no se conoce con certeza dicha función, se podría disponer de alguna otra función de probabilidad conjunta aproximada, notada $b(\{x\})$. Con las dos funciones, se define entonces una distancia entre ellas, conocida como la distancia de Kullback-Leibler, de la siguiente manera:

$$D(b(\{x\})||P(\{x\})) = \sum_{\{x\}} b(\{x\}) \ln \frac{b(\{x\})}{P(\{x\})} \quad (6)$$

4 El modelo

En esta sección se desarrolla el modelo propuesto para brindar un nuevo enfoque a la consistencia macroeconómica en el Ecuador. La herramienta planteada contempla tres grandes etapas: en primer lugar, se estructura una serie de datos con la historia más larga posible de los principales agregados macroeconómicos y sus relaciones de consistencia desde un enfoque contable. En una segunda etapa, se realiza el aprendizaje del modelo gráfico utilizando datos económicos históricos del Ecuador. Finalmente, se realiza la inferencia del modelo gráfico propuesto (árbol con variables latentes), utilizando para esto una derivación del algoritmo “belief propagation”, mencionado previamente.

Como se recordará de la sección en la que se explicaron los modelos probabilísticos gráficos, las principales aplicaciones de este tipo de modelos están enmarcadas en la visión artificial y reconstrucción de imágenes y patrones. La idea de analizar la consistencia macroeconómica desde esta perspectiva radica precisamente en la riqueza que aportan las series históricas de datos económicos, es decir, utilizar la información de interrelación de las distintas variables que forman parte de este sistema y que, de algún modo, no es evidente a primera vista o no está plasmada en ecuaciones preestablecidas. La visión, entonces, del modelo propuesto es la de analizar un marco actual o futuro de consistencia macroeconómica a la luz de las “fotos” o “imágenes” económicas de la historia de un país.

4.1 Modelo de árbol con variables latentes

Existe un grupo de modelos probabilísticos llamados modelos de variables latentes. Estos modelos permiten relacionar un conjunto de variables observables con un conjunto de variables no observables. En estos modelos, las variables latentes explican las relaciones de dependencia estadística entre las variables observables, permitiendo así modelar sistemas de manera compacta y eficiente. Por otro lado, estos modelos permiten reducir la dimensionalidad de los datos y generar variables aleatorias condicionalmente independientes, lo cual simplifica considerablemente el análisis. Las aplicaciones de estos modelos gráficos son numerosas y abarcan muchas áreas del conocimiento como: psicología, sociología, economía, y también ciencias biológicas y áreas relacionadas a la inteligencia artificial (Mourad *et al.*, 2013).

Definición 3. *Un modelo de árbol con variables latentes (LTM, por sus siglas en inglés) es un modelo probabilístico gráfico con estructura de árbol (grafo en el que cualquier par de nodos está conectado exactamente por un único camino) en el que los nodos terminales representan a variables observables, mientras que los nodos internos pueden ser variables latentes o no observables.*

Estos modelos son poderosas herramientas de modelamiento estadístico ya que tienen una estructura simple que permite realizar inferencia de manera eficiente, y la flexibilidad de inclusión de variables latentes permite capturar relaciones complejas en el sistema que está siendo modelado.

Un LTM, entonces, está compuesto por un grafo con estructura de árbol $T(V, E)$ y un conjunto de parámetros θ . Si el LTM es un grafo dirigido, entonces se trata de una red bayesiana (Figura 3) y, si se trata de un grafo no dirigido, entonces se habla de campo aleatorio de Markov (Figura 4).

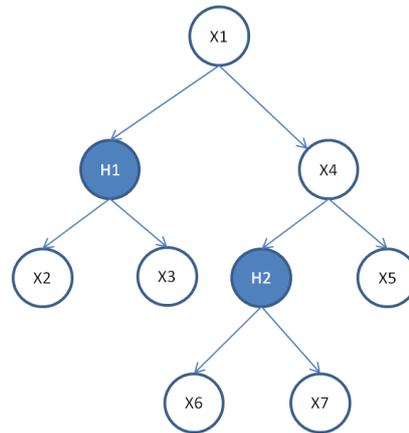


Figura 3: Red bayesiana con forma de árbol (grafo dirigido con estructura causal)

El conjunto de nodos $V = \{V_1, \dots, V_{n+m}\}$ representa a las $n + m$ variables observables y latentes. El conjunto $X = \{X_1, \dots, X_n\}$ contiene a las n variables observables y el conjunto $H = \{H_1, \dots, H_m\}$ a las m variables latentes. Finalmente, el conjunto de k aristas, notado $E = \{e_1, \dots, e_k\}$ captura las relaciones de dependencia directa entre estas variables.

Cuando el LTM tiene forma de red bayesiana (Figura 3), el conjunto de parámetros θ consiste en las distribuciones de probabilidad, una para cada variable. Como se recordará, además, de la sección en la que se explicaron los modelos gráficos, dada una variable V_i , cuyos nodos padres son $pa(V_i)$, se encuentra definida la distribución condicional $P(V_i|pa(V_i))$. Cuando se trata de una variable V_i cuyo conjunto $pa(V_i)$ es vacío, en lugar de la distribución condicional, se encuentra definida una distribución marginal $P(V_i)$. La definición de la distribución de probabilidad conjunta de todas las variables para el caso de un grafo dirigido se especificó ya en la sección 3.2. por lo que se desarrolla más bien en la explicación del grafo no dirigido.

4 El modelo

En esta sección se desarrolla el modelo propuesto para brindar un nuevo enfoque a la consistencia macroeconómica en el Ecuador. La herramienta planteada contempla tres grandes etapas: en primer lugar, se estructura una serie de datos con la historia más larga posible de los principales agregados macroeconómicos y sus relaciones de consistencia desde un enfoque contable. En una segunda etapa, se realiza el aprendizaje del modelo gráfico utilizando datos económicos históricos del Ecuador. Finalmente, se realiza la inferencia del modelo gráfico propuesto (árbol con variables latentes), utilizando para esto una derivación del algoritmo “belief propagation”, mencionado previamente.

Como se recordará de la sección en la que se explicaron los modelos probabilísticos gráficos, las principales aplicaciones de este tipo de modelos están enmarcadas en la visión artificial y reconstrucción de imágenes y patrones. La idea de analizar la consistencia macroeconómica desde esta perspectiva radica precisamente en la riqueza que aportan las series históricas de datos económicos, es decir, utilizar la información de interrelación de las distintas variables que forman parte de este sistema y que, de algún modo, no es evidente a primera vista o no está plasmada en ecuaciones preestablecidas. La visión, entonces, del modelo propuesto es la de analizar un marco actual o futuro de consistencia macroeconómica a la luz de las “fotos” o “imágenes” económicas de la historia de un país.

4.1 Modelo de árbol con variables latentes

Existe un grupo de modelos probabilísticos llamados modelos de variables latentes. Estos modelos permiten relacionar un conjunto de variables observables con un conjunto de variables no observables. En estos modelos, las variables latentes explican las relaciones de dependencia estadística entre las variables observables, permitiendo así modelar sistemas de manera compacta y eficiente. Por otro lado, estos modelos permiten reducir la dimensionalidad de los datos y generar variables aleatorias condicionalmente independientes, lo cual simplifica considerablemente el análisis. Las aplicaciones de estos modelos gráficos son numerosas y abarcan muchas áreas del conocimiento como: psicología, sociología, economía, y también ciencias biológicas y áreas relacionadas a la inteligencia artificial (Mourad *et al.*, 2013).

Definición 3. *Un modelo de árbol con variables latentes (LTM, por sus siglas en inglés) es un modelo probabilístico gráfico con estructura de árbol (grafo en el que cualquier par de nodos está conectado exactamente por un único camino) en el que los nodos terminales representan a variables observables, mientras que los nodos internos pueden ser variables latentes o no observables.*

Estos modelos son poderosas herramientas de modelamiento estadístico ya que tienen una estructura simple que permite realizar inferencia de manera eficiente, y la flexibilidad de inclusión de variables latentes permite capturar relaciones complejas en el sistema que está siendo modelado.

Un LTM, entonces, está compuesto por un grafo con estructura de árbol $T(V, E)$ y un conjunto de parámetros θ . Si el LTM es un grafo dirigido, entonces se trata de una red bayesiana (Figura 3) y, si se trata de un grafo no dirigido, entonces se habla de campo aleatorio de Markov (Figura 4).

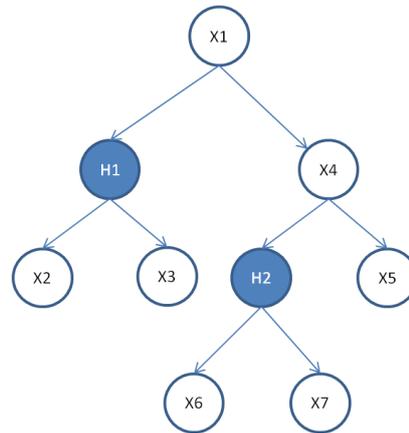


Figura 3: Red bayesiana con forma de árbol (grafo dirigido con estructura causal)

El conjunto de nodos $V = \{V_1, \dots, V_{n+m}\}$ representa a las $n + m$ variables observables y latentes. El conjunto $X = \{X_1, \dots, X_n\}$ contiene a las n variables observables y el conjunto $H = \{H_1, \dots, H_m\}$ a las m variables latentes. Finalmente, el conjunto de k aristas, notado $E = \{e_1, \dots, e_k\}$ captura las relaciones de dependencia directa entre estas variables.

Cuando el LTM tiene forma de red bayesiana (Figura 3), el conjunto de parámetros θ consiste en las distribuciones de probabilidad, una para cada variable. Como se recordará, además, de la sección en la que se explicaron los modelos gráficos, dada una variable V_i , cuyos nodos padres son $pa(V_i)$, se encuentra definida la distribución condicional $P(V_i|pa(V_i))$. Cuando se trata de una variable V_i cuyo conjunto $pa(V_i)$ es vacío, en lugar de la distribución condicional, se encuentra definida una distribución marginal $P(V_i)$. La definición de la distribución de probabilidad conjunta de todas las variables para el caso de un grafo dirigido se especificó ya en la sección 3.2. por lo que se desarrolla más bien en la explicación del grafo no dirigido.

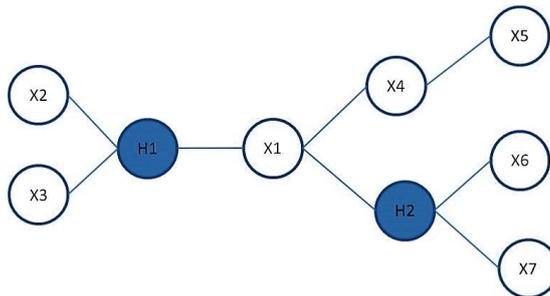


Figura 4: Campo aleatorio de Markov con forma de árbol (grafo no dirigido con estructura de interdependencias)

En el caso de que el LTM corresponda a un grafo no dirigido, el conjunto de parámetros θ corresponde a las distribuciones de probabilidad de las aristas y los nodos internos. Con esta consideración, la función de distribución conjunta se expresa como:

$$P(V) = \frac{\prod_{(V_a, V_b) \in E} P(V_a, V_b)}{\prod_{j=1}^M P(I_j)^{d(I_j)-1}} \tag{7}$$

Donde $d(I_j)$ es el grado del j -ésimo nodo interno I_j y (V_a, V_b) es la notación para referirse a las aristas del conjunto E .

En la sección de anexos se detallan algunos aspectos importantes de los LTM, como el criterio de selección del modelo y los procesos de aprendizaje e inferencia.

4.2 Propuesta de aproximación a la consistencia macroeconómica

Una vez establecidos todos los aspectos precedentes a la propuesta del modelo fruto de esta investigación, en esta sección se plantea un enfoque de consistencia macroeconómica para el Ecuador. Como se mencionaba al inicio de esta sección, el modelo se desarrolla en tres etapas, que se describen en las siguientes secciones.

4.2.1 Primera etapa: consistencia contable

La propuesta de aproximación a la consistencia macroeconómica parte de uno de los más recientes trabajos que se han desarrollado en Ecuador al respecto, el marco de consistencia propuesto por Pérez y Mancheno (2011) para la evaluación de la Agenda de Política Económica 2011-2013 del Ministerio Coordinador de la Política Económica. Esta investigación propone un análisis de la consistencia macroeconómica del país desde un enfoque contable, utilizando para ello un cierre de tres brechas (Bacha, 1990) por el lado de la brecha privada.

Este trabajo, sin ser necesariamente un instrumento con desarrollo econométrico, incorpora un grupo de identidades macro-contables que permiten acotar el conjunto de proyecciones posibles en los distintos escenarios de consistencia. El modelo de Pérez y Mancheno (2011) tiene desagregaciones para los principales sectores:

- Sector Público No Financiero
 - Ingresos / Gastos (sobre la línea)
 - Necesidades y fuentes de financiamiento (bajo la línea)
- Sector Monetario y Financiero
 - Créditos y depósitos
 - Reserva internacional de libre disponibilidad
 - Otras cuentas del BCE
- Sector Externo
 - Importaciones petroleras y no petroleras
 - Exportaciones petroleras y no petroleras
 - Balanza de servicios y renta
- Sector Real
 - Oferta total
 - * Producto interno bruto (PIB)
 - * Importaciones
 - Demanda total
 - * Exportaciones
 - * Consumo de los hogares
 - * Consumo del sector público
 - * Formación bruta de capital fijo
 - * Variación de existencias

Dentro del sector público no financiero (SPNF) se resume el movimiento de las principales cuentas agregadas de todas las instituciones y empresas públicas, exceptuando aquellas que realizan intermediación financiera de primer o segundo piso. Básicamente, dentro de los ingresos se recoge la información de los flujos correspondientes a tributos, petróleo, aranceles, seguridad social (contribuciones) y excedentes de empresas públicas. En el caso de los gastos,

los principales componentes son: intereses, sueldos y salarios, bienes y servicios, seguridad social y el gasto de capital. Las necesidades y fuentes de financiamiento del SPNF resumen las amortizaciones de los créditos internos y externos, así como el financiamiento identificado para cubrir la brecha de ingresos y gastos fiscales.

En el apartado del sector monetario y financiero, aparte de la obvia importancia de contar con la información de los créditos y depósitos, hay que destacar el aporte que la desagregación de algunas de estas cuentas tiene para el sistema que se busca modelar en esta investigación ya que, por ejemplo, el crédito productivo está fuertemente vinculado al crecimiento de economía, o el crédito de consumo al macroagregado “consumo de los hogares”. Dentro de las cuentas del Banco Central del Ecuador, a más de la reserva internacional de libre disponibilidad, se cuenta con información de los depósitos de los gobiernos locales, encajes y depósitos de sociedades públicas no financieras.

En una economía abierta, dolarizada y fundamentalmente tomadora de precios internacionales, como la ecuatoriana, la información del sector externo resulta sumamente importante para evaluar un marco de consistencia macroeconómica. Sin embargo, la excesiva desagregación de este sector puede resultar en una manera de “diluir” las relaciones con los otros grandes agregados macroeconómicos, evitando así un correcto aprendizaje del LTM.

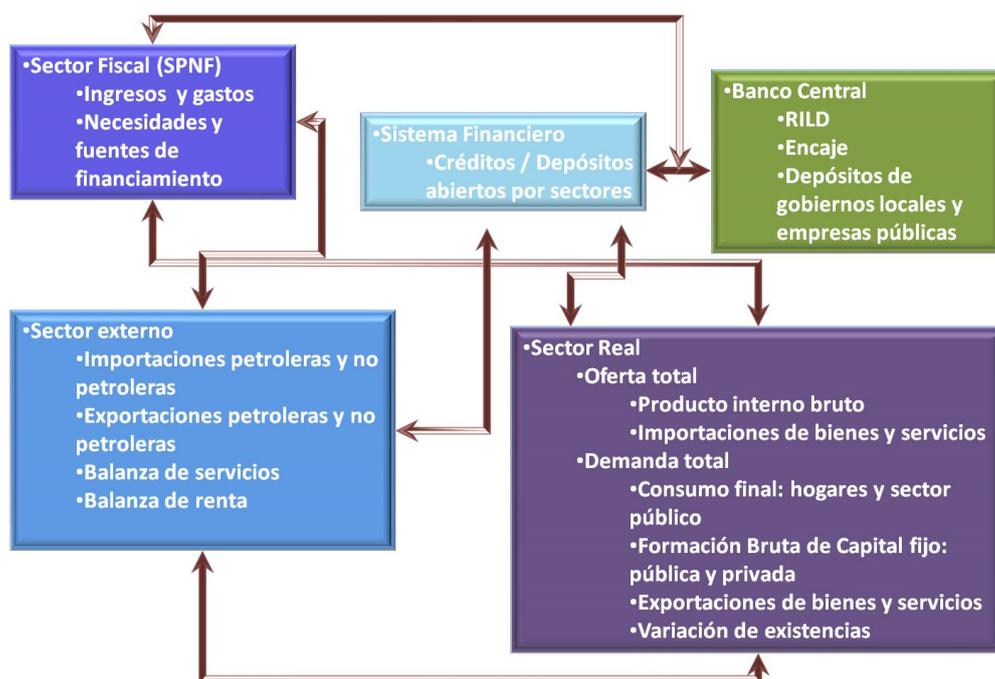


Figura 5: Relaciones intersectoriales

Fuente: Agenda de Política Económica 2011-2013 (MCPE)

Elaboración: propia

En este sentido, se han utilizado las importaciones y exportaciones diferenciadas entre

petroleras y no petroleras, así como los saldos de las balanzas de servicios y renta.

Finalmente, en el sector real se utilizan prácticamente las variables disponibles en la tabla oferta utilización, prescindiendo de algunas de ellas de forma sistemática durante el proceso de modelación para evitar problemas de sobreidentificación en los modelos explorados ya que, como se sabe, estas variables están vinculadas por una identidad contable.

La Figura 5 presenta de manera esquemática las principales relaciones intersectoriales que se plasman en este enfoque de consistencia.

La serie histórica de datos que se ha construido utilizando este marco de consistencia macroeconómica es la que sirve de insumo para la siguiente etapa: el aprendizaje de la estructura del modelo de árbol con variables latentes.

4.2.2 Segunda etapa: aprendizaje del LTM

Se han desarrollado diversos algoritmos que permiten hacer aprendizaje de la estructura de un modelo de árbol con variables latentes a partir de un conjunto de observaciones de las variables del sistema. Cada uno de estos algoritmos, encasillados en alguna de las tres categorías de modelos explicadas en la sección anterior: basados en búsqueda, clusterización de variables o basados en distancia.

De la revisión de la literatura especializada, se han seleccionado dos algoritmos a ser implementados y contrastar sus resultados en el aprendizaje de los LTM que se realizará en esta investigación. El principal criterio de selección ha sido la compatibilidad de la información económica disponible con los supuestos que cada uno de los algoritmos tiene a priori sobre el conjunto de datos a modelar. Con este antecedente, los algoritmos que se han seleccionado son el de agrupamiento recursivo (Choi *et al.*, 2011) y el de agrupamiento de Chow-Liu con unión de vecinos (Saitou y Nei, 1987). El primer algoritmo corresponde a la categoría de los modelos basados en distancia y el segundo al grupo de modelos de clusterización de variables. En ambos casos, la matriz de distancias es el input que se requiere para el aprendizaje del LTM. En la sección de Anexos se explica con detalle cada uno de estos algoritmos.

Como se sabe, en economía, la inclusión de variables no observables (latentes) se ha ido convirtiendo casi en un estándar en los modelos que se desarrollan para explicar o dar respuesta a los fenómenos estudiados en esta área del conocimiento. Existen varias explicaciones para que esta corriente haya tomado fuerza, sin embargo, a criterio del autor, una de las principales es el hecho de contar con información en distinto nivel de desagregación levantada, a su vez, con distintas periodicidades, lo cual dificulta la inclusión simultánea de todas las variables que se desea modelar en un mismo sistema.

Este es el caso, por ejemplo, del Producto Interno Bruto en Ecuador. La información de cuentas nacionales provista por el Banco Central es trimestral y, además, se publica con un trimestre de rezago. Esto, en la práctica, implica que existan períodos de hasta seis meses en los que no se tiene una idea clara de la evolución de la economía que permita tomar medidas correctivas tempranas, en caso de ser necesarias. Por otro lado, existen indicadores

mensuales que pueden dar alguna luz sobre lo que está ocurriendo en el sistema económico y que tienen relación con el PIB, como el consumo de energía eléctrica o el Índice de Actividad Económica Coyuntural (Ideac), por ejemplo, pero dicha relación no es fácilmente explicitada precisamente por la falta de herramientas que permitan hacerlo.

La idea, entonces, del aprendizaje del modelo que se plantea en este trabajo consiste en construir en primera instancia un LTM con la mayor cantidad de variables observables mensualmente posibles. Una vez seleccionado el mejor modelo en función de los criterios explicados anteriormente, el siguiente paso consiste en buscar una explicación económica a las variables latentes encontradas. Para esto, se plantea un barrido de todas las opciones posibles, descartando aquellas que por una simple aplicación de conceptos económicos no serían plausibles a la luz de las variables observables que están conectadas a las variables latentes en cuestión. Así, para continuar con el ejemplo anterior, si una variable latente H está conectada directamente en el LTM a las variables X_1 : empleo y X_2 : consumo total de energía eléctrica, una alternativa factible para explicar esa variable latente podría ser el PIB pero, por otro lado, sería muy poco probable que el efecto no observable común a X_1 y X_2 sea el consumo del sector público.

Más allá del aporte de esta investigación en cuanto a lo novedoso y las ya mencionadas potencialidades de utilizar modelos gráficos con variables latentes para tratar explicar un ámbito de consistencia macroeconómica, una contribución adicional consiste en buscar dar contenido a las variables latentes, precisamente por el hecho de estar modelando un sistema económico y se puede tener bastante claridad sobre el tema a la luz de la teoría económica. Resumiendo, en cuanto al aprendizaje del árbol, se construye el mejor LTM con un amplio conjunto de datos de la mayor periodicidad posible (mensual, en este caso) y luego se prueba la explicación de las variables latentes halladas “calzando” con los principales agregados macroeconómicos en el siguiente nivel de desagregación en el que dicha información esté disponible (trimestralmente en Ecuador). La comparación entre modelos se realiza utilizando la distancia de Kullback-Leibler (definida en una sección previa) entre las distribuciones resultantes de los mismos, bajo el supuesto de que la distribución original es la que se obtiene en el primer LTM, es decir, en el que las variables latentes no están asociadas a un indicador macroeconómico aún.

4.2.3 Tercera etapa: inferencia

Como ya se dijo antes, por inferencia se entiende responder a preguntas del tipo $P(X|E, M)$, donde X es la variable (o conjunto de variables) sobre la que se desea conocer algo, M es el modelo probabilístico que se desea utilizar para responder a la pregunta y E es la evidencia o información observable disponible para “refinar” dicha respuesta. Puntualmente, en el caso de los LTM, la inferencia sería de la forma $P(W|X, H; \theta^{ML}, T)$, donde W es el conjunto de variables sobre las que se desea inferir (responder) algo, X es el conjunto de las demás variables observables, H el conjunto de variables latentes, θ^{ML} el conjunto de parámetros

mensuales que pueden dar alguna luz sobre lo que está ocurriendo en el sistema económico y que tienen relación con el PIB, como el consumo de energía eléctrica o el Índice de Actividad Económica Coyuntural (Ideac), por ejemplo, pero dicha relación no es fácilmente explicitada precisamente por la falta de herramientas que permitan hacerlo.

La idea, entonces, del aprendizaje del modelo que se plantea en este trabajo consiste en construir en primera instancia un LTM con la mayor cantidad de variables observables mensualmente posibles. Una vez seleccionado el mejor modelo en función de los criterios explicados anteriormente, el siguiente paso consiste en buscar una explicación económica a las variables latentes encontradas. Para esto, se plantea un barrido de todas las opciones posibles, descartando aquellas que por una simple aplicación de conceptos económicos no serían plausibles a la luz de las variables observables que están conectadas a las variables latentes en cuestión. Así, para continuar con el ejemplo anterior, si una variable latente H está conectada directamente en el LTM a las variables X_1 : empleo y X_2 : consumo total de energía eléctrica, una alternativa factible para explicar esa variable latente podría ser el PIB pero, por otro lado, sería muy poco probable que el efecto no observable común a X_1 y X_2 sea el consumo del sector público.

Más allá del aporte de esta investigación en cuanto a lo novedoso y las ya mencionadas potencialidades de utilizar modelos gráficos con variables latentes para tratar explicar un ámbito de consistencia macroeconómica, una contribución adicional consiste en buscar dar contenido a las variables latentes, precisamente por el hecho de estar modelando un sistema económico y se puede tener bastante claridad sobre el tema a la luz de la teoría económica. Resumiendo, en cuanto al aprendizaje del árbol, se construye el mejor LTM con un amplio conjunto de datos de la mayor periodicidad posible (mensual, en este caso) y luego se prueba la explicación de las variables latentes halladas “calzando” con los principales agregados macroeconómicos en el siguiente nivel de desagregación en el que dicha información esté disponible (trimestralmente en Ecuador). La comparación entre modelos se realiza utilizando la distancia de Kullback-Leibler (definida en una sección previa) entre las distribuciones resultantes de los mismos, bajo el supuesto de que la distribución original es la que se obtiene en el primer LTM, es decir, en el que las variables latentes no están asociadas a un indicador macroeconómico aún.

4.2.3 Tercera etapa: inferencia

Como ya se dijo antes, por inferencia se entiende responder a preguntas del tipo $P(X|E, M)$, donde X es la variable (o conjunto de variables) sobre la que se desea conocer algo, M es el modelo probabilístico que se desea utilizar para responder a la pregunta y E es la evidencia o información observable disponible para “refinar” dicha respuesta. Puntualmente, en el caso de los LTM, la inferencia sería de la forma $P(W|X, H; \theta^{ML}, T)$, donde W es el conjunto de variables sobre las que se desea inferir (responder) algo, X es el conjunto de las demás variables observables, H el conjunto de variables latentes, θ^{ML} el conjunto de parámetros

estimados del modelo por máxima verosimilitud y T es el árbol que define la estructura de interdependencias de las variables en cuestión.

Sobre este punto no se hará mucho énfasis en la presente investigación pues es demasiado amplio el universo de temas que se podrían abordar, sin embargo, se hará un ejercicio de inferencia relacionado al Producto Interno Bruto (siempre en el ámbito de la consistencia macroeconómica) utilizando una implementación del algoritmo de pase de mensajes belief propagation (BP), explicado ya en una sección anterior.

El objetivo del trabajo en este punto sería, más bien, sentar el interés para futuras investigaciones que puedan mejorar lo realizado y explotar la potencialidad de esta herramienta para modelar sistemas económicos, no solamente en el ámbito de la consistencia sino pensando en previsiones o evaluaciones de medidas de política pública.

5 Resultados y conclusiones

En esta sección se presentan los principales resultados y conclusiones de la implementación de la metodología propuesta. En la primera parte se describe la información base utilizada para el modelamiento. Luego, se muestran los LTM que se han armado en función de los distintos algoritmos utilizados, con los respectivos criterios estadísticos que se aplicaron para hallarlos. En esta parte se presentan dos aplicaciones adicionales de la herramienta: en la primera se utilizan series mensuales rezagadas para armar los LTM, esto con el objeto de entender de mejor manera como se van construyendo las relaciones mensuales entre las distintas variables en una ventana de tiempo trimestral y abordar el siempre presente problema de la endogeneidad de las variables económicas; la segunda aplicación propone un ejercicio de construcción de un LTM con diversos agregados macroeconómicos anuales de una serie de datos larga. Finalmente, se presenta un ejercicio real de inferencia sobre el PIB, condicionado a un conjunto de supuestos macroeconómicos, para evaluar la consistencia de la previsión de crecimiento planteada en base a dichos supuestos con aquella que el modelo prevé.

La implementación de la metodología se ha realizado en la herramienta de software para modelamiento matemático Matlab, versión 7.6.0 (R2008a). Esto debido a la facilidad que brinda para la programación de modelos probabilísticos gráficos y la cantidad de herramientas preestablecidas con las que cuenta el programa, mismas que facilitan la tarea de implementación. Además, el Matlab en general permite escribir códigos mucho más cortos que otras herramientas comparables ya que no necesita hacer alocaiones de memoria o definiciones previas de tipos de datos, entre otras cosas.

Los algoritmos este trabajo utilizan algunas piezas de código y funciones desarrolladas en el proyecto “Bayes Net Toolbox for Matla” de Kevin Murphy⁴. También se han utilizado algunas piezas de código y funciones del proyecto “Kernel Belief Propagation: Tutorial Code”,

⁴Código libre bajo licencia GNU GPL v2, disponible en el sitio <https://code.google.com/p/bnt/> al 31 de julio de 2014

desarrollado por Le Song, Arthur Gretton, Danny Bickson, Yucheng Low y Carlos Guestrin⁵. Finalmente, las versiones base de los algoritmos que fueron modificados para el aprendizaje de los LTM de esta investigación se tomaron del proyecto “Learning Latent Tree Graphical Models” de Myung Choi, Vincent Tan, Animashree Anandkumar y Alan Willsky⁶.

5.1 Información utilizada

Para los distintos modelos que se prueban en el trabajo, se han utilizado cuatro conjuntos de datos con la siguientes especificaciones:

5.1.1 Conjunto de datos 1

Este conjunto de datos se ha utilizado para construir las principales relaciones de consistencia contable en los grandes agregados macroeconómicos. La estructura de estos datos corresponde a una simplificación del marco de consistencia presentado por Pérez y Mancheno (2011) para la evaluación de la Agenda de Política Económica 2011-2013. La serie corresponde a tasas de variación anuales y contiene datos entre el año 2000 y el 2014. Cabe señalar que las variables utilizadas no son todas aquellas que intervienen en la construcción de un marco contable de consistencia macroeconómica; del universo posible de indicadores macroeconómicos, este grupo de variables se ha seleccionado descartando aquellas que muestran bajos o nulos niveles de correlación con las demás del sistema. La Tabla 1 contiene la descripción de estas variables.

5.1.2 Conjunto de datos 2

Este conjunto de datos se utiliza para la construcción del LTM con la información de mayor periodicidad posible, en el que las variables latentes no tienen aún contenido económico. Los datos corresponden a información mensual entre enero de 2001 y diciembre de 2008 (96 observaciones). Las variables utilizadas en esta serie de datos se especifican en la Tabla 2 y corresponden a tasas de variación.

⁵Disponible en el sitio <http://www.gatsby.ucl.ac.uk/gretton/kernelBP/kernelBP.htm> al 31 de julio de 2014

⁶Disponible en el sitio <http://people.csail.mit.edu/myungjin/latentTree.html> al 31 de julio de 2014

Tabla 1: Variables utilizadas en la construcción del marco de consistencia contable

Sector	Variable
Real	Producto interno bruto Deflactor del PIB (índice) Producto interno bruto (USD de 2007) Exportaciones Importaciones Consumo hogares Consumo del sector público Inversión
Público no financiero	Ingresos petroleros Ingresos no petroleros Gasto corriente Gasto de capital y préstamo neto Resultado Global (sobre línea) Resultado primario Necesidades de Financiamiento Financiamiento identificado Brecha de financiamiento (bajo la línea)
Externo	Cuenta Corriente Balanza Comercial Total balanza Comercial no petrolera Exportaciones totales de bienes (sin petróleo) Importaciones totales de bienes (sin petróleo) Saldo de comercio no registrado Balanza de servicios y renta Servicios crédito Servicios débito Renta crédito Renta débito Transferencias corrientes Cuenta de Capitañ y Financiera
Monetario y financiero	Crédito al sector privado Crédito a otras sociedades no financieras Crédito a otros sectores residentes Reserva internacional de libre disponibilidad Depósitos gobiernos locales Depósitos sociedades públicas no financieras
Tributario	Impuesto a la Renta Global Impuesto al Valor Agregado Impuesto a los Consumos Especiales Impuesto a los Vehículos Motorizados Aranceles

5.1.3 Conjunto de datos 3

Este conjunto de datos contiene las series trimestrales de la tabla oferta-utilización entre el primer trimestre de 1990 y el cuarto trimestre de 2008. Combinando esta información con la del conjunto de datos 2 se busca dar contenido económico a las variables latentes que aparecen en el LTM asociado a dicho conjunto. La Tabla 3 contiene la descripción de las variables, cuyos valores se encuentran en dólares del año 2000.

Tabla 2: Variables mensuales utilizadas para el LTM

Variable	Código en los LTM
Ingresos petroleros	IngPet
Ingresos Tributarios por Impuesto a la Renta	Irent
Ingresos Tributarios por IVA	Iva
Gastos Corrientes en intereses	GCInt
Gastos Corrientes en sueldos y salarios	GCSyS
Gastos Corrientes en compras de bienes y servicios	GCBByS
Gastos de Capital	GC
Ideac	Ideac
Crédito total al sector privado	Cred
Ipc	Ipc
Ingreso real	IngReal
Desempleo	Desemp
Tasa de Ocupación Global	OcupG
Tasa de Ocupación Bruta	OcupB
Tasa de Subempleo	Subemp
Impuestos Directos	ImpDir
Impuestos Indirectos	ImpInd
Retenciones	Retenc
Iva Interno	IvaInt
Iva Import	IvaImpor
Ice Interno	IceInt
Ice Import	IceImport
Exportaciones (con petróleo)	ExpCP
Importaciones (con petróleo)	ImpSP
Términos de intercambio (con petróleo)	TdISP
Exportaciones (sin petróleo)	ExpSP
Importaciones (sin petróleo)	ImpSP
Términos de intercambio (sin petróleo)	TdICP
Tipo de cambio efectivo real	Tcer
Energía Facturada a Clientes Finales	Energ

5.1.4 Conjunto de datos 4

Esta serie de datos anuales entre 1971 y 2010 se utiliza para construir un LTM en el que se modela la economía ecuatoriana utilizando el mismo conjunto de variables con el que Aguiar *et al.* (2011) estimaron un modelo estructural de ecuaciones. En este conjunto existen variables que se han generado por los autores mencionados a partir de estimaciones propias, como es el caso del stock de capital por ejemplo. El objetivo de utilizar este conjunto de datos es crear un marco común de comparación entre modelos con fines similares, en una suerte de benchmark de metodologías. La Tabla 4 muestra la descripción de estas variables, todas corresponden a tasas de variación de sus logaritmos naturales.

5.2 Resultados

En esta sección se presentan los resultados de los distintos ejercicios que se han practicado sobre los conjuntos de datos descritos en la sección anterior. Como se explicó antes, el primer ejercicio consiste en intentar explicar las variables latentes que aparecen en el LTM

construido con información mensual mediante la extrapolación de la estructura de dicho árbol al conjunto de datos trimestral.

En todos los ensayos de construcción de un LTM que se realizaron, a más de probar con los dos algoritmos descritos anteriormente, se prueba la estructura de dos maneras: la primera mediante un algoritmo que escoge la variable con el mayor nivel de correlación con el resto de variables⁷, y la segunda fijando a priori la variable vinculada al nodo raíz (en base a criterio de experto, considerando las variables que están siendo modeladas).

Tabla 3: Variables de la tabla oferta utilización

Variable	Código en los LTM
Producto interno bruto	Pib
Importaciones de bienes y servicios	M
Consumo de los hogares	Ch
Consumo del sector público	Cg
Exportaciones de bienes y servicios	X
Inversión	Inv
Valor agregado no petrolero	VaNP
Valor agregado petrolero	VaP
Otros elementos del PIB	OEPib

Tabla 4: Variables anuales utilizadas para la construcción de un LTM

Variable	Abreviatura en los LTM
Consumo de los hogares	lnC
Pago total de salarios	lnLW
Salario	LnW
Excedente bruto de explotación	lnExB
Precios de la economía	lnP
Ingreso	lnY
Cantidad de trabajo (PEA ocupada)	lnL
Tasa de interés	lnr
Capital	lnK
Stock de capital	lnSK
Precios internacionales	lnPInt
Población económicamente activa	lnPea
Empleo	lnEmp
Importaciones	lnM
Gasto de Gobierno	lnG
Exportaciones	lnX
Inversión	lnInv
Variación de existencias	lnVK
Consumo intermedio	lnCInt

5.2.1 Caso 1

En este caso se utiliza el algoritmo de agrupamiento de Chow-Liu con unión de vecinos (ACLUV) para la construcción de un árbol trimestral a partir de un árbol con datos men-

⁷Para esto el criterio de selección es la suma de las correlaciones con todo el resto de variables

suales que contiene variables latentes. Este ejercicio se realiza dejando a elección del modelo la variable asociada al nodo raíz. La Figura 6 muestra el LTM resultante, en el cual el algoritmo detectó la existencia de cuatro variables latentes.

La Figura 7 presenta el LTM resultante al extrapolar la estructura del TM construido con datos mensuales al LTM de datos trimestrales (conjunto de datos 3) y, luego, optimizar el BIC sobre todas las combinaciones posibles de sustituciones de las variables observables trimestrales en lugar de las variables latentes encontradas.

En general, al modelar sistemas complejos es altamente improbable encontrar una correspondencia biunívoca entre todas las variables latentes y las nuevas variables observables incluidas, y más aún lo es al tratarse de un sistema socio-económico como el que se está abordando en esta investigación. No obstante de esto, si existen algunas pautas bastantes claras de conceptos económicos que podrían explicar estas relaciones no observables.

Así, por ejemplo, al observar las variables conectadas por la variable latente 3 en el LTM mensual (Figura 6), se puede presumir observando el LTM trimestral (Figura 7) que este efecto no observable está fuertemente incidido por el Consumo de los hogares y los Otros elementos del PIB.

Por otro lado, también hay relaciones entre variables observables manifestadas a través de variables latentes en el LTM mensual que luego se rompen al incluir las nuevas variables observables de periodicidad trimestral y se agrupan de distinta manera en el modelo gráfico. Esto se explicará por el nuevo orden de “fuerzas” y relaciones de dependencia que causa la inclusión de estas nuevas variables observables. Este es el caso, por ejemplo, de las variables conectadas por las variables latentes 1 y 2 en el LTM mensual (Figura 6): Gastos corrientes en bienes y servicios, Ingresos petroleros, Impuesto al valor agregado y Gasto de capital. Al observar el LTM trimestral (Figura 7), estas variables se han dispersado a través del árbol, incluso agrupándose con otras en un sentido más lógico desde el punto de vista de la teoría económica (véase por ejemplo que los Ingresos petroleros están directamente conectados al Consumo del gobierno y, por otro lado, los Gastos corrientes en sueldos y salarios tienen vínculo directo con el Impuesto a la renta).

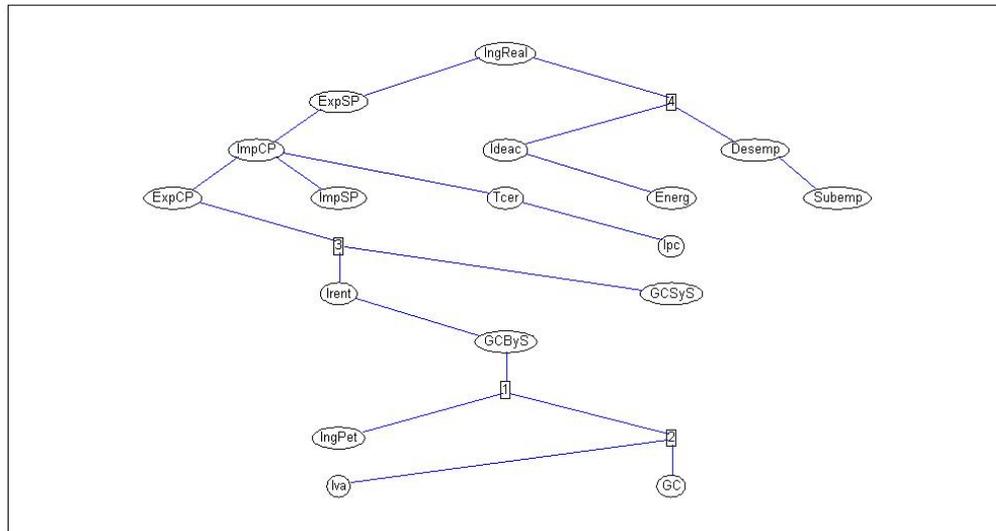


Figura 6: LTM mensual ACLUV, raíz libre

Nota: Los nodos etiquetados con números y encerrados en cuadrados corresponden a variables latentes.

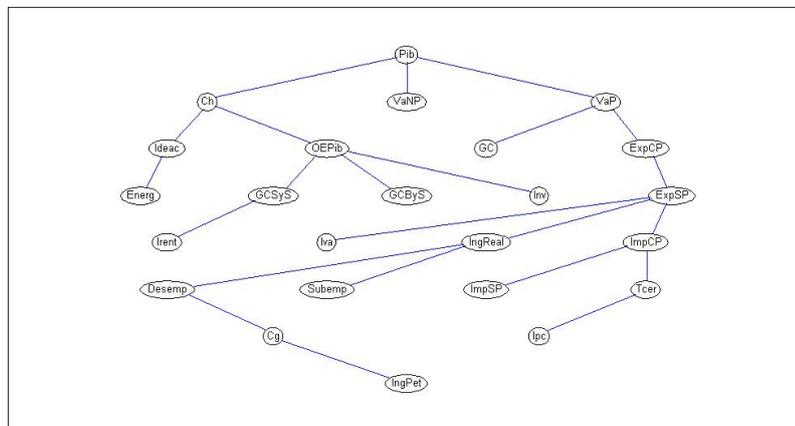


Figura 7: LTM trimestral ACLUV sin VL, raíz libre

Nota: Los nodos etiquetados con números y encerrados en cuadrados corresponden a variables latentes.

En el análisis previo, la construcción del LTM trimestral cerraba las puertas a la inclusión de nuevas variables latentes, pues el objetivo del ejercicio era buscar explicaciones con contenido económico a los efectos no observables detectados por el modelo en el LTM mensual. El LTM de la Figura 8, en contraparte, “flexibiliza” esta restricción y admite la inclusión de variables observables, siempre y cuando mejoren el score del LTM original. Esto tiene mucho sentido pues, a pesar de haber incluido las variables de las cuales si se dispone información trimestral para mejorar las estimaciones del modelo, es claro que en fenómenos económicos, como en muchas otras áreas de modelación, pueden seguir existiendo efectos no observables (y que tal vez siempre lo sean) que están determinando las relaciones dentro del sistema.

Tabla 5: Indicadores de ajuste de los LTM caso 1

Modelo	Log-likelihood	BIC
LTM mensual ACLUV, raíz libre	-12702.519535	-12771.597087b
LTM trimestral ACLUV sin VL, raíz libre	-14052.694872	-14132.134058
LTM trimestral ACLUV con VL, raíz libre	-15117.490654	-15207.291473

Si se observan tanto la verosimilitud de la estimación como el criterio de información bayesiana (BIC) en la Tabla 5 se ve claramente que existe una significativa mejora del modelo cuando se incluyen las variables trimestrales y, más aún, cuando se permite la inclusión de variables latentes en este último LTM.

Quizás el efecto más interesante que se puede detectar en el LTM de la Figura 8 se encuentra en el subárbol que se encuentra en la parte inferior derecha, en el que aparecen unidas las variables latentes 1 y 2 a variables observables como: Índice de actividad económica coyuntural, Consumo de energía eléctrica, Consumo de los hogares, Otros elementos del PIB, Inversión, Impuesto a la renta, entre otras. Por la naturaleza económica de las variables en cuestión, se podría presumir que estos efectos no observables están ligados a la producción potencial de la economía.

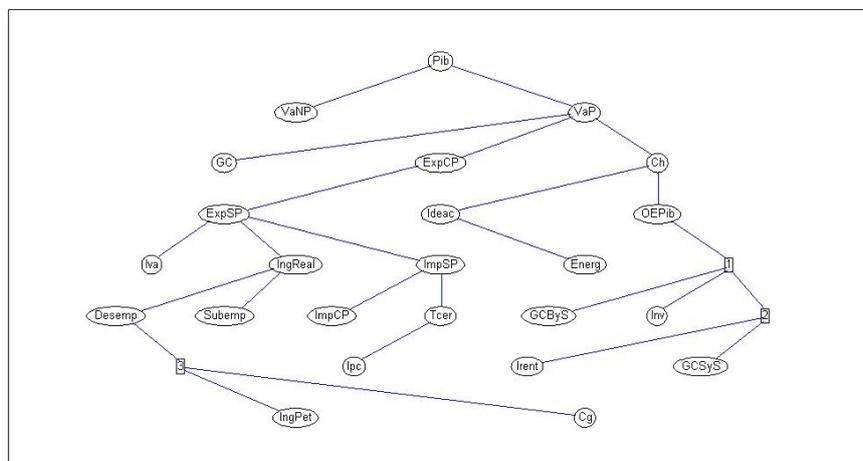


Figura 8: LTM trimestral ACLUV con VL, raíz libre

Nota: Los nodos etiquetados con números y encerrados en cuadrados corresponden a variables latentes

5.2.2 Caso 2

Este ejercicio busca dar una nueva visión al siempre presente problema de la endogeneidad⁸ de las variables económicas, mediante la inclusión de dos rezagos temporales en todas las variables mensuales y modelarlas a través de un LTM. En Ecuador, los principales agregados macroeconómicos se publican trimestralmente, impidiendo así tener una idea concreta de como son las relaciones de interdependencia de los indicadores de los que se dispone durante el trimestre. La idea, entonces, de este ensayo es intentar entender mediante un modelo probabilístico gráfico las relaciones de dependencia entre variables que aproximan al sistema económico con una periodicidad mensual.

En este caso se utiliza asimismo el algoritmo ACLUV pero, para alternar respecto del anterior, se elige a priori el nodo raíz. Por el objetivo de monitorear el desempeño de la economía con una periodicidad mayor a la de la publicación de las cuentas nacionales, la variable que se ha seleccionado para colocar en el nodo raíz es el crédito total al sector privado.

Se podría escribir mucho sobre las diversas relaciones que se manifiestan en el LTM de la Figura 9, sin embargo, pensando en el objetivo de haber introducido este ejercicio, obsérvense los tres subárboles marcados en las regiones encerradas con línea continua. En el primer subárbol (de arriba hacia abajo) se muestran efectos no observables (variable 12) que

⁸El tipo de endogeneidad al que hago referencia en este apartado es aquel que se deriva de la simultaneidad o el comportamiento circular en las variables en cuestión. La endogeneidad debida a variables omitidas no es abordada en este artículo.

marcan una relación entre el Crédito en $t - 2$, el Ingreso real en $t - 2$ y, el Desempleo y el Subempleo en $t - 1$. Por otro lado, el segundo subárbol muestra la estrecha relación que existe entre estas mismas variables observables, conectadas por dos variables latentes (5 y 7), pero con distinto rezago; la relación es entre el Crédito en t , el Ingreso Real en t y, el Desempleo y el Subempleo en $t - 2$. Finalmente, el tercer subárbol una vez más muestra las fuertes relaciones de interdependencia que existen entre las variables que conforman este subsistema, siempre enlazadas mediante efectos no observables (variables latentes 2 y 8); en este caso las variables vinculadas son el Crédito en $t - 1$, el Ingreso Real en $t - 1$ y, el Desempleo y el Subempleo en t .

Estas tres regiones marcadas en el LTM muestran claramente la endogeneidad de las variables en cuestión y, a la vez, ponen en evidencia el error que se cometería al modelarlas con un enfoque causa-efecto pues, como se dijo en la introducción, en los sistemas económicos las relaciones muchas veces son circulares y no de causa-efecto. Además de la circularidad existente en el subsistema formado por estas tres variables, es interesante destacar que incluso la periodicidad es distinta; el primer y tercer subárbol muestran que los efectos que el Crédito y el Ingreso Real (además estos dos muestran siempre patrones de comportamiento simultáneo) manifiestan en el Desempleo y el Subempleo tienen rezago de un mes, mientras que los efectos contrarios, es decir, de los niveles de Desempleo y Subempleo sobre el Crédito y el Ingreso Real, se dan con un rezago de dos meses.

Por otro lado, como se mencionó al inicio de este caso, la variable de crédito total al sector privado se seleccionó como una “proxy” del nivel de actividad económica. La bondad de utilizar dicha variable se ha visto confirmada en el hecho de que siempre está vinculada a las variables de empleo e ingreso real mediante efectos no observables (variables latentes). Precisamente, todo apunta a que este efecto no observable es el nivel de actividad económica, para el cual no existe una medición “real” con periodicidad mensual, por lo que no se puede incluir explícitamente en el modelo. Este hecho demuestra las ventajas de utilizar la aproximación de modelos probabilísticos gráficos para la modelación macroeconómica en ausencia de información con la periodicidad que se desearía. Finalmente, el Log-likelihood de este modelo es -36020.248054 y el BIC es -36234.388467.

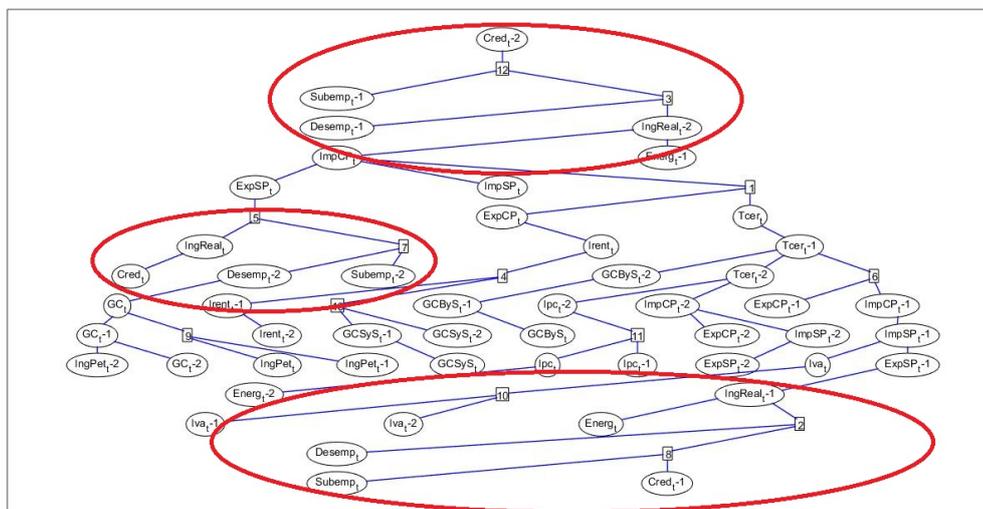


Figura 9: LTM mensual con rezagos ACLUV, raíz fija

Nota: Los nodos etiquetados con números y encerrados en cuadrados corresponden a variables latentes.

5.2.3 Caso 3

Este caso de estudio utiliza las series de datos anuales del conjunto de datos. La justificación de presentar este ensayo es que esta es una serie considerablemente larga (1971-2010) y se dispone información de los principales agregados macroeconómicos, con lo cual, el planteamiento de consistencia a la luz del tipo de modelo propuesto adquiere mayor robustez. Sobre este LTM se plantea el problema de inferencia en la siguiente y última sección.

Como se observa en la Figura 10, la variable con mayor nivel de relacionamiento escogida por el algoritmo es el PIB (Y). De las diversas regiones que se han formado en torno a las variables observables y latentes, hay algunas que cabe resaltar por los conceptos económicos que se relacionan. En este modelo el Log-likelihood es de -10877.377801 y el BIC es -10953.363109.

En primer lugar, obsérvese la variable latente 2, que relaciona el PIB con los Consumos intermedios y los Precios de la economía. La Brecha de producción (diferencia entre el PIB real y el PIB potencial) podría ser un muy buen candidato para explicar este efecto latente pues, además de determinar la producción efectiva y los consumos intermedios, esta brecha se explica fundamentalmente por presiones de demanda, lo que implica que es inflacionaria y, de ahí precisamente la relación directa con los precios de la economía.

Analizando el subárbol generado alrededor los nodos 1 y 3 (variables latentes) en el LTM de la Figura 10, se encuentra que están relacionadas directamente las variables observables:

Variación de capital, Stock de capital y Tasa de interés, por un lado, y por otro las Importaciones y el Empleo. Dadas estas relaciones, la explicación a estos efectos latentes podría incluir a variables no observables que se sabe tienen vinculación directa con las variables en mención como son: la tasa de interés natural (para el primer grupo), y el tipo de cambio real de equilibrio o la tasa natural de desempleo (para el segundo grupo).

Otro efecto no observable sumamente relevante para el análisis macroeconómico es el que se encuentra recogido en la variable latente 4. Al observar las variables observables que relaciona (PIB, gasto del gobierno y exportaciones totales) se puede pensar que dicho efecto no observable tiene estrecha relación con el comportamiento del mercado petrolero a nivel internacional. Los efectos de este mercado, que se transmiten a nuestro país principal pero no exclusivamente mediante el precio del petróleo, son claramente palpables en los niveles de exportaciones de la economía ecuatoriana, cuya generación de divisas por concepto de petróleo es la principal fuente que alimenta los recursos públicos destinados para inversión. Desde luego, esta inversión pública, posible gracias a los ingresos por exportaciones petroleras, se ve a su vez representada en el ingreso nacional.

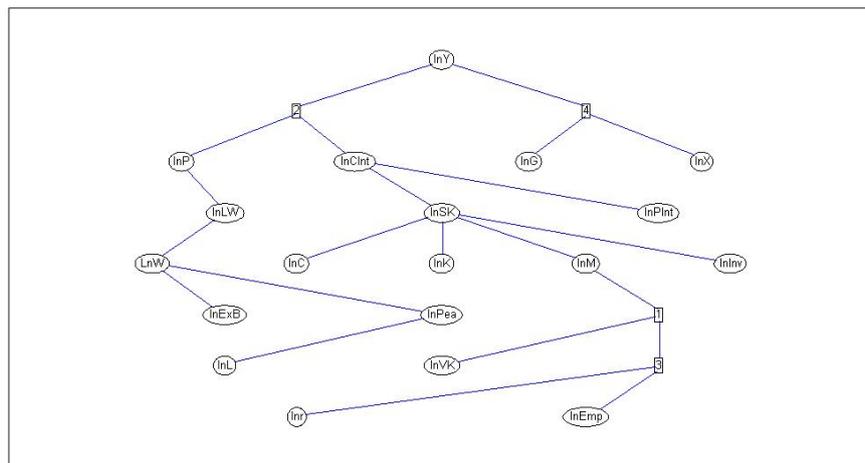


Figura 10: LTM anual ACLUV, raíz libre

Nota: Los nodos etiquetados con números y encerrados en cuadrados corresponden a variables latentes.

Son muchos los ejercicios interpretativos que se pueden iniciar a partir de los resultados de los modelos que se han planteado, sin embargo, la intencionalidad de este trabajo ha sido abrir un nuevo campo de investigación utilizando herramientas estadísticas modernas que permitan mejorar el entendimiento de sistemas complejos como lo es el socio-económico. Se pueden, además, refinar mucho más los conjuntos de datos utilizados aumentar la historia de las series disponibles para recoger de mejor manera la historia económica de un país

e intentar plasmarla (con la enorme simplificación de la realidad que esto implica) en un enfoque de consistencia macroeconómica como el planteado en esta investigación.

5.2.4 Inferencia sobre el PIB

Como se explicó antes, la última parte de la investigación consiste en hacer un ejercicio de inferencia sobre el producto interno bruto para ejemplificar la utilización del modelo propuesto como una herramienta de evaluación de la consistencia macroeconómica en el Ecuador. Para esta parte se utiliza el LTM con datos anuales (Figura 10), ya que éste utiliza un conjunto de datos macroeconómicos relativamente completo y una serie larga, lo cual permite que se recojan de mejor manera las relaciones económicas históricas del sistema.

El algoritmo de propagación de beliefs (BP) utiliza como insumo la estructura del LTM resultante del proceso de aprendizaje explicado en la sección anterior. Además, un supuesto importante de esta metodología es que las variables del sistema siguen una distribución gaussiana; sin embargo, este no es un supuesto inverosímil, pues se trabaja con la serie de variaciones interanuales. Una posible mejora a este trabajo es precisamente la generalización a cualquier tipo de distribución continua no gaussiana mediante la aplicación de estimadores de densidad kernel e incrustaciones en espacios de Hilbert (Song *et al.*, 2011)

La Figura 11 muestra las distribuciones marginales empíricas (a priori) de las variables observables del LTM aprendido con el algoritmo ACLUV para el conjunto de datos 4. A partir de esta estructura empieza el proceso de inferencia sobre cualquier nodo observable (que no sea un nodo terminal) mediante la rutina de paso de mensajes que involucra el algoritmo BP. Es decir, dada la distribución conjunta de todas las variables del sistema (observables y no observables) contenida en la estructura del LTM, el objetivo es encontrar la distribución a posteriori “refinada” de una variable observable, condicionada a un conjunto de valores observados de un grupo de variables.

En este parte se realiza un ejercicio de inferencia sobre el producto interno bruto para determinar cómo cambia su distribución marginal cuando se observan (o predicen, si se evalúa un escenario futuro de consistencia macroeconómica) otras variables del sistema. Puntualmente, el ejercicio consiste en suponer un escenario de contracción de la demanda agregada en el que el consumo de los hogares se contrae en un 1%, los Precios internacionales mantienen un crecimiento inercial en un 2.5% y el gasto del gobierno se contrae en un 2%.

La Figura 12 muestra como cambia la distribución a posteriori del PIB cuando se marginaliza respecto de las demás variables del sistema y se condiciona a la evidencia disponible (escenario planteado). Puntualmente, la Figura 12 responde a la pregunta de inferencia $P(Y|C = -0,01, PInt = 0,025, G = -0,02; X^*, H; \theta^{ML}, T)$, donde X^* corresponde al conjunto de las demás variables observables, H al conjunto de variables latentes, θ^{ML} son los parámetros asociados a las aristas (estimados en el proceso de aprendizaje) y T es el LTM construido a partir del conjunto de datos 4 con el nodo raíz libre.

La tasa promedio de crecimiento real del PIB anual en el período de análisis (1971-2010)

es de 4.36%. En el escenario planteado, al condicionar el sistema a la evidencia dada en las tres variables mencionadas anteriormente, la esperanza de crecimiento del PIB es de -0.26%. De esta manera se pueden utilizar las herramientas desarrolladas en esta investigación para evaluar escenarios de consistencia macroeconómica, por ejemplo en el caso ecuatoriano, aplicados al crecimiento de la economía dados los supuestos macroeconómicos con los que se construye la proforma presupuestaria.

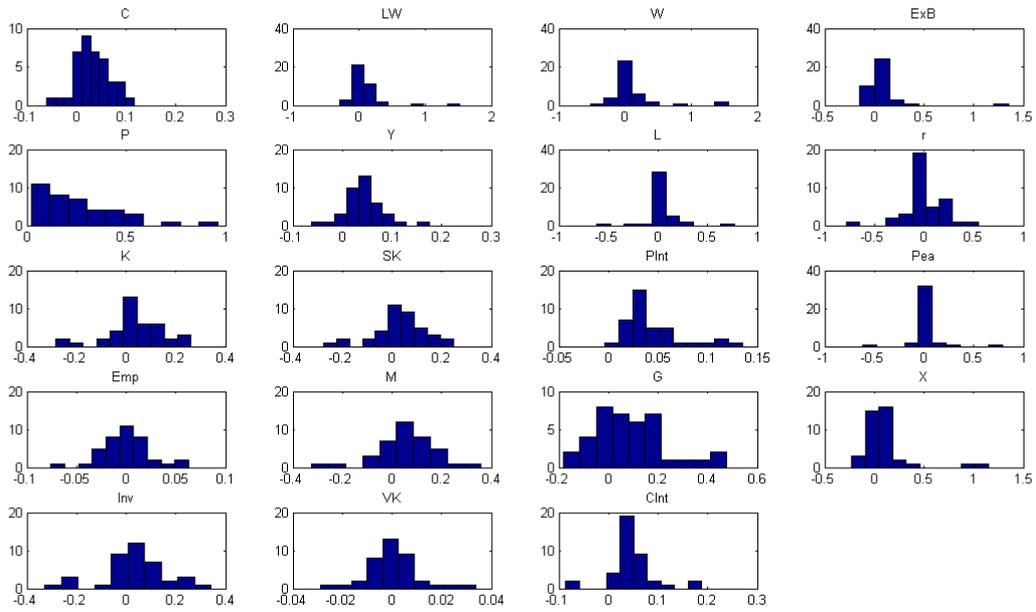


Figura 11: Distribuciones empíricas de las variables observables del LTM

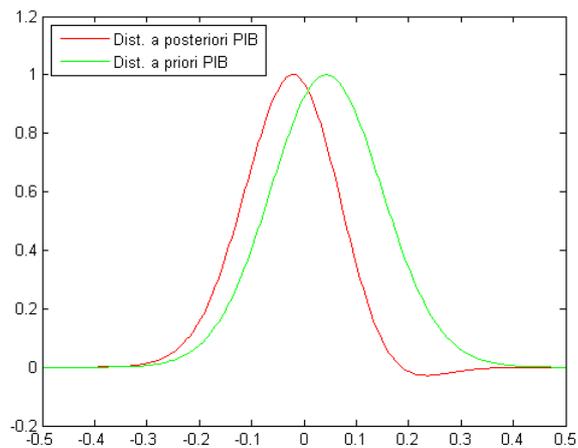


Figura 12: Distribución marginal a posteriori del PIB

5.3 Conclusiones

Son varias las conclusiones que se pueden obtener de la presente investigación; en esta sección se intentan resaltar los aspectos más importantes de este nuevo enfoque de modelación económica que, a criterio del autor, abre todo un nuevo campo de investigación y ha arrojado interesantes resultados en esta primera aproximación.

- Con esta nueva visión de modelamiento de sistemas económicos se rompe la visión clásica de modelamiento causa-efecto que llevan implícitos los modelos de regresiones y, principalmente, series temporales (ARIMA, VAR, etc.). En la práctica, en economía se encuentran fenómenos que no siempre obedecen a una lógica causal y, más bien, responden a comportamientos circulares y procesos de retroalimentación permanente con otros componentes del sistema. Los modelos de equilibrio general, al permitir la estimación simultánea de ecuaciones, no necesariamente reflejan esta lógica causa-efecto pero, en cambio, adolecen de los consabidos problemas de falta de información suficiente que permita estimar la gran cantidad de parámetros que involucran y, generalmente, también la baja capacidad de reproducir adecuadamente la compleja realidad de los sistemas económicos.
- Esta investigación ha puesto de manifiesto una nueva forma de abordar el siempre presente problema de la endogeneidad en los datos económicos, principalmente, pero también de sistemas vinculados a muchas otras áreas de conocimiento. La herramienta propuesta no solamente permite “comprobar” de algún modo esta endogeneidad a la luz de los propios datos, sino que también permite observar este fenómeno dentro de todo el sistema. Además, otra potencialidad de la herramienta en este sentido consiste

en la capacidad de determinar gran parte de los problemas de endogeneidad presentes en un conjunto de datos para que esto sea considerado en otro tipo de modelaciones.

- Los resultados obtenidos en segundo caso presentado muestran que la variable “crédito total al sector privado” es un buen indicador, de periodicidad mensual, del nivel de actividad económica. En un ejercicio previo, cuando el artículo estaba en realización, se probó con la variable Ideac (índice de actividad económica coyuntural) en lugar del crédito total y los resultados fueron menos concluyentes que los obtenidos en el ejercicio definitivo. Este resultado es relevante pues el Ideac, por su naturaleza, tiene como objetivo precisamente hacer las veces de un indicador temprano de la actividad económica del país. Por el contrario, lo que se observa en este ejercicio es que una variable como el crédito, que no requiere de un alto nivel de procesamiento, o responde a un modelo detrás, cumple de mejor manera esta función.
- De uno de los casos de análisis que se plantearon en esta investigación se puede concluir que, utilizando las variables adecuadas, los modelos probabilísticos gráficos como los LTM permiten entender el funcionamiento de un sistema económico en períodos de tiempo en los que usualmente la contabilidad nacional no lo permite. Tal es el caso de Ecuador, donde la información de las variables de la tabla oferta utilización se publican trimestralmente y, con un trimestre de rezago. En este caso, el ejercicio de construir un LTM con indicadores mensuales y luego intentar explicar los efectos no observables mediante la extrapolación de la estructura del árbol a las series trimestrales permitió esbozar algunas probables explicaciones a dichos efectos no observables, explicaciones que en ciertos casos pudieron ser entendidas a la luz de la teoría económica.
- En prácticamente todos los ensayos de aprendizaje de modelos de árbol con variables latentes, el algoritmo de agrupamiento de Chow-Liu con unión de vecinos tuvo un mejor desempeño que el otro algoritmo implementado (agrupamiento recursivo). Si bien en términos de eficiencia computacional este último tiene un mejor desempeño, la diferencia no es significativa y prevalece la capacidad del primer algoritmo de reproducir relaciones económicas coherentes dentro del sistema modelado. Además, generalmente las variables latentes que resultaban del aprendizaje con el algoritmo ACLUV tenían una explicación intuitiva dentro de la teoría económica, cosa que no necesariamente sucedía con el algoritmo AR. Es necesario recalcar que el análisis de la implicación o interpretación económica que dichas variables latentes podría tener es una tarea de grandes proporciones, que rebasa el alcance de este estudio. Como se evidencia en los LTM construidos, existe un sin número de relaciones que pasan por la existencia de dichos efectos no observables; siendo cada una de ellas en si misma objeto de una investigación más profunda. El objetivo del presente estudio fue abrir una puerta, hasta ahora poco explorada, de análisis de consistencia macroeconómica utilizando herramientas estadísticas de vanguardia.

- El problema de elección del nodo raíz es sumamente importante al modelar sistemas utilizando LTM. Este es un aspecto muy importante y determina en buena medida la bondad de ajuste de los modelos gráficos. Generalmente, es el criterio de experto el que prevalece para fijar el nodo raíz en un modelo de árbol con variables latentes, sin embargo, utilizando la subrutina propuesta para hallar la variable con mayor nivel de relacionamiento con el resto del sistema, los resultados no fueron muy distintos de la elección que presumiblemente se habría hecho únicamente en base al criterio de experto.
- Una conclusión que es necesaria resaltar respecto a este enfoque de modelamiento económico es la capacidad de extrapolar propiedades entre conjuntos de datos que tengan distinta periodicidad pero que tengan variables comunes. Como se sabe, la integración coherente y estructurada de información de diversas fuentes es un problema abierto pues, en la práctica, y sobre todo en economía, los conjuntos de datos suelen ofrecer información con distintas periodicidades, contradictoria o con problemas de consistencia entre los micro y los macro datos. Esta propuesta de aproximación a la consistencia macroeconómica ayuda en uno de estos tres problemas, permitiendo conjugar propiedades inherentes a conjuntos de datos de distintas periodicidades en un solo grafo que permita dar una mejor interpretación del sistema. Sin embargo, al momento de responder preguntas de inferencia se necesitarían de otras herramientas que ayuden en este propósito, además de la identificación y solución de los conflictos de información.
- La herramienta propuesta como herramienta de aproximación a la consistencia macroeconómica, si bien no enfoca el problema desde el lado contable o mediante la inclusión de ecuaciones de comportamiento, si recoge la memoria de un sistema económico y, como tal, permite hacer evaluaciones más “ágiles” y propias de una economía, sin tener que depender de resultados o parámetros importados de otras realidades. Claro está que la calidad de la información que se utilice para el ejercicio de aprendizaje es sumamente importante para que las relaciones económicas existentes se vean correctamente plasmadas en la distribución conjunta y las distribuciones marginales de las variables involucradas en el sistema. El enfoque estadístico planteado en esta investigación es pionero en el ámbito de la consistencia macroeconómica y, en general, bastante novedoso incluso si el ámbito de comparación se refiere al modelamiento económico en un sentido amplio.
- En Ecuador, generalmente las previsiones macroeconómicas se plantean en torno a unos cuantos escenarios discretos de algunos indicadores relevantes, como el precio del petróleo, inflación, inversión pública, entre otros. En la evaluación de consistencia utilizando este enfoque se dejan por fuera los efectos simultáneos no necesariamente aditivos que pueden generar los supuestos que al respecto se realizan. Por otro lado,

dar estimaciones puntuales de indicadores que responden a un complejo sistema socio-económico, como el PIB por ejemplo, es bastante aventurado, por decirlo de algún modo. Para la generación y evaluación de política pública es mucho más provechoso contar con escenarios probabilísticos en los que se puedan determinar regiones de mayor o menor probabilidad de ocurrencia y no solamente estimaciones puntuales que ni siquiera se conoce en que parte de la distribución están. La herramienta construida en este trabajo constituye un importante paso en busca de dar el paso entre las previsiones puntuales que se realizan en torno a unos cuantos escenarios de los grandes agregados macroeconómicos, pues al conocer la distribución conjunta de un sistema vinculado a un modelo gráfico, se puede obtener información más rica de regiones de mayor o menor probabilidad de ocurrencia dados unos ciertos supuestos o conjuntos de evidencias.

- Es claro que, por un lado, es beneficioso para el aprendizaje del modelo contar con series de datos lo suficientemente largas como para que se puedan recoger la mayor cantidad de relaciones entre las distintas variables del sistema, sin embargo, esto también puede ser una desventaja cuando una economía ha sido sujeta de profundos cambios estructurales, mismos que rompen las relaciones históricas observadas entre las variables del sistema. Una forma de abordar este problema puede ser la experimentación sobre distintas ventanas de tiempo para el entrenamiento del modelo, de manera que exista suficiente información para estimar los parámetros pero que, asimismo, una serie muy larga no permita que se recojan cambios estructurales en el aparato económico de un país.
- De la presente investigación se desprenden algunos trabajos futuros que bien podrían aportar a un mayor y mejor entendimiento del funcionamiento del sistema económico de Ecuador y, en general, de cualquier país. Un primer trabajo que sería importante conseguirlo es una descripción exhaustiva de todas las relaciones que se han encontrado en el aprendizaje de los distintos LTM construidos sobre los conjuntos de datos disponibles. Este espacio de trabajo ha resultado corto para la gran cantidad de información, hipótesis, conclusiones que se han desprendido del enfoque de consistencia propuesto. Por esto, resultaría sumamente importante explotar todos los análisis que quedan pendientes y, por supuesto, emprender nuevos análisis también con conjuntos de datos más amplios o refinados, o con nuevas preguntas de inferencia que se busquen responder.
- En cuanto a la metodología, algunas pautas que pueden guiar investigaciones futuras en este campo del conocimiento se describen a continuación:

Generalizar el algoritmo de inferencia para distribuciones continuas no gaussianas utilizando estimadores de densidad kernel con incrustaciones en espacios de Hilbert o con alguna otra metodología que permita conseguir este objetivo.

Trabajar en la estimación y construcción de macrosistemas (bosques), es decir, no intentando modelar todo el sistema económico con un modelo de árbol sino mediante varios modelos de árbol interconectados, que respondan cada uno de ellos a los distintos sectores de la economía.

Se puede trabajar en una mejor definición de distancias entre variables, que no se basen únicamente en los coeficientes de correlación, sino que utilicen algún (algunos) otro criterio que recoja de mejor manera las relaciones iniciales entre variables.

Referencias

- Aguiar, V., Castro, L., y y Barbaro, F. (2011). Modelo macroeconómico de tres brechas endógeno. *Technical report, Ministerio Coordinador de la Política Económica*.
- Bacha, E. (1990). A three-gap model of foreign transfers and the gdp growth rate in developing countries. *Journal of Development Economics*, 32(2):279–296.
- Choi, M., Tan, V., Anandkumar, A., y Willsky, A. (2011). Learning latent tree graphical models. *The Journal of Machine Learning Research*, 12:1771–1812.
- Chow, C. y Liu, C. (1968). Approximating discrete probability distributions with dependence trees. *Information Theory, IEEE Transactions on*, 14(3):462–467.
- Cooper, G. (1990). The computational complexity of probabilistic inference using bayesian belief networks. *Artificial intelligence*, 42(2):393–405.
- Dobrushin, R. (1968). The description of a random field by means of conditional probabilities and conditions of its regularity. *Theory of Probability & Its Applications*, 13(2):197–224.
- Dutt, A. (2011). Macroeconomic theory after the crisis. *Review of Radical Political Economics*.
- Easterly, W. (1989). *A consistency framework for macroeconomic analysis*, volumen 234.
- Garey, M. y Johnson, D. (1979). A guide to the theory of npcompleteness. *WH Freeman, Nueva York*, 3(1).
- Lauritzen, S. y Spiegelhalter, D. (1988). Local computations with probabilities on graphical structures and their application to expert systems (with dicusion). *Journal of the Royal Statistical Society*, 8:157–224.
- Mourad, R., Sinoquet, C., Zhang, N., Liu, T., y Leray, P. (2013). A survey on latent tree models and applications. *Journal of Artificial Intelligence Research*, 47:157–203.

- Pearl, J. (1988). Probabilistic reasoning in intelligent systems: networks of plausible inference. *Morgan Kaufmann*.
- Pérez, W. y Acosta, M. (2005). Modelo ecuatoriano de equilibrio general aplicado. *Cuestiones Económicas*, 22:5–43.
- Pérez, W. y Mancheno, D. (2011). Modelo de consistencia macroeconómica para la evaluación de la agenda de política económica 2011-2013. *Technical report, Ministerio Coordinador de la Política Económica*.
- Roch, S. (2006). A short proof that phylogenetic tree reconstruction by maximum likelihood is hard. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 3(1):92.
- Saitou, N. y Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, 4(4):406–425.
- Song, L., Xing, E., y Parikh, A. (2011). Kernel embeddings of latent tree graphical models. En J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, y K. Weinberger (Eds.). *Advances in Neural Information Processing Systems*, 24:2708–2716.
- Walras, L. (1874). Elements of pure economics or the theory of social wealth. En *Richard D. Irwin Inc*.
- Zhang, N. y Kocka, T. (2004). Effective dimensions of hierarchical latent class models. *J. Artif. Intell. Res. (JAIR)*, 21:1–17.

ANEXOS

Teorema 1. Factorización de probabilidades. Sean X_1, X_2, \dots, X_n variables aleatorias, entonces su distribución conjunta se puede escribir como:

$$P(X_1, X_2, \dots, X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \dots P(X_n|X_1, X_2, \dots, X_{n-1}) \quad (8)$$

A continuación se describen algunas propiedades básicas que se derivan de la definición de independencia condicional. Sean X, Y, Z y W variables aleatorias, y sea U una transformación de Y , entonces se cumple que:

- Si $X \perp Y | Z$, entonces $Y \perp X | Z$;
- Si $X \perp Y | Z$ y $U = g(Y)$, entonces $X \perp U | Z$;
- Si $X \perp Y | Z$ y $U = g(Y)$, entonces $X \perp Y | (Z, U)$;

- Si $X \perp Y|Z$ y $X \perp W|(Y, Z)$, entonces $X \perp (Y, W)|Z$;
- Si $X \perp Y|Z$ y $X \perp Z|Y$, entonces $X \perp (Y, Z)$; y
- $X \perp Y|Z \Leftrightarrow f(x, y, z)f(z) = f(x, z)f(y, z) \Leftrightarrow \exists a, b : f(x, y, z) = a(x, z)b(y, z)$.

Propiedades de la distancia de Kullback-Leibler

Esta distancia de Kullback-Leibler no cumple con las propiedades usuales de una distancia definida sobre un espacio métrico, pues, ni es simétrica, i.e. $D(b(\{x\})||P(\{x\})) \neq D(P(\{x\})||b(\{x\}))$, ni cumple con la desigualdad triangular: $D(b(\{x\})||P(\{x\})) + D(P(\{x\})||q(\{x\})) \not\leq D(b(\{x\})||q(\{x\}))$.

Por otro lado, las otras dos propiedades usuales de una distancia si son cumplidas por la distancia de Kullback-Leibler, estas son:

- $D(b(\{x\})||P(\{x\})) \geq 0$, para cualquiera b y P , y;
- $D(b(\{x\})||P(\{x\})) = 0$ si y solamente si las dos funciones de probabilidad b y P son iguales.

Algunos aspectos de los LTM

Scoring y elección del modelo

En los modelos gráficos un indicador de scoring es siempre deseable para poder comparar entre distintas alternativas para una elección final. Aunque varios tipos de score pueden ser utilizados para “medir” modelos de árbol con variables latentes, el criterio de información bayesiana (BIC, por sus siglas en inglés) es el que se utiliza de forma casi general en la literatura especializada y los trabajos experimentales desarrollados.

Considérese un conjunto de n variables observables $X = \{X_1, \dots, X_n\}$ y una colección de N observaciones independientes idénticamente distribuidas (i.i.d.) $D_x = \{x^1, \dots, x^N\}$, BIC del árbol relacionado a estas variables se calcularía como sigue:

$$BIC(T, D_x) = \log P(D_x | \theta^{ML}, T) - \frac{1}{2} \dim(T) \log N \quad (9)$$

Siendo θ^{ML} el conjunto de parámetros estimados por máxima verosimilitud y $\dim(T)$ la dimensión del modelo. El primer término del BIC evalúa la calidad de ajuste del modelo a los datos y el segundo término penaliza el indicador en función de la dimensión del modelo. En los LTM, al incluir variables latentes, la dimensión no puede ser calculada simplemente como el número de parámetros libres. En su lugar, se calcula una dimensión “efectiva” como el rango de la matriz jacobiana que mapea los parámetros del modelo a la distribución conjunta de las variables observables.

Hay que considerar que, según demuestra Roch (2006), la estimación de parámetros de un LTM por máxima verosimilitud es un problema NP-duro⁹. Por este motivo, como se explicará más adelante, existen diversos algoritmos heurísticos de aproximación al problema.

Otro aspecto importante a considerar en la comparación de LTM es la parsimonia del modelo, también conocida como minimalidad según lo plantea Pearl (1988). Antes de definir la parsimonia, se necesita definir la equivalencia marginal:

Definición 4. Sean los LTM: $M_1 = (T_1, \theta_1)$ y $M_2 = (T_2, \theta_2)$, contruidos sobre el mismo conjunto de n variables observables $X = \{X_1, \dots, X_n\}$, se dice que M_1 y M_2 son **marginalmente equivalentes** si sus distribuciones conjuntas son iguales:

$$P(X_1, \dots, X_n|T_1, \theta_1) = P(X_1, \dots, X_n|T_2, \theta_2) \quad (10)$$

Además, si dos modelos marginalmente equivalentes tienen la misma dimensión, entonces se dice que los modelos son equivalentes. Con estos conceptos, se puede ya definir la parsimonia:

Definición 5. Un modelo M se dice parsimonioso si no existe otro modelo M' que sea marginalmente equivalente y que tenga una dimensión menor.

Se puede demostrar que un modelo parsimonioso siempre tiene el mejor score posible. Además, una característica de los LTM parsimoniosos es que no contienen variables latentes redundantes.

Finalmente, existe una condición que asegura que un LTM no incluye clases latentes (cardinalidad de las variables latentes) redundantes. Sea H una variable latente en un LTM. Sea Z el conjunto de k variables que son vecinos de H , $Z = \{Z_1, \dots, Z_k\}$. Entonces, según establece Zhang y Kocka (2004), se dice que un LTM es regular si para cualquier variable latente H , se cumple que:

$$|H| \leq \frac{\prod_{i=1}^k |Z_i|}{\max_{i=1}^k |Z_i|} \quad (11)$$

Según demuestran Zhang y Kocka (2004), todos los modelos parsimoniosos son regulares. Eso permite que la búsqueda del mejor modelo quede restringida únicamente al espacio de modelos regulares.

Aprendizaje de la estructura

En general, y más aún cuando se modelan sistemas complejos, como el marco de consistencia macroeconómica de esta investigación, la estructura del LTM no es conocida a priori,

⁹En teoría de complejidad, los problemas NP-duros son aquellos cuya resolución es, al menos, tan “compleja” como la de un problema no determinísticamente polinomial (NP), es decir, que no pueden ser solucionados en un tiempo polinomial. Para mayor profundización, referirse al texto de Garey y Johnson (1979).

es decir: las relaciones de dependencia entre variables, cantidad de variables latentes, sus cardinalidades y los parámetros. Por este motivo, la gran mayoría de trabajos que se han desarrollado, sobre todo en la última década, cuando estos modelos han tomado considerable impulso (Mourad *et al.*, 2013) tienen como objetivo el aprendizaje de la estructura del LTM a partir de un conjunto de observaciones del sistema.

Según resumen Mourad *et al.* (2013), los métodos de aprendizaje de la estructura de un LTM se dividen en tres categorías: métodos basados en búsqueda (desarrollados fundamentalmente para el aprendizaje de redes bayesianas), métodos de clusterización de variables (relacionados a procesos jerárquicos) y los métodos basados en distancias (desarrollados mayoritariamente a la luz del campo de la filogenética).

Dentro de los métodos basados en búsqueda destacan el método *greedy* de Naive y el método de búsqueda avanzada. En el grupo de los métodos basados en clusterización de variables, se han desarrollado métodos específicos según los tipos de LTM; así, por ejemplo, para árboles binarios existe el método aglomerativo jerárquico, para árboles no binarios está el método bietapa aglomerativo-esperanza maximización, y para árboles planos existen el algoritmo Bridged Island. Dentro de los métodos basados en distancias, los principales desarrollos son: el algoritmo de unión de vecinos, el método de aprendizaje dedicado a LTM generales y los métodos espectrales. Más adelante se presenta el detalle de los métodos seleccionados para el aprendizaje de la estructura del LTM propuesto en este trabajo.

Inferencia

El otro gran objetivo cuando se trabaja con modelos probabilísticos gráficos es, por supuesto, hacer inferencia sobre las variables. La inferencia probabilística en una red bayesiana general es un problema NP-duro, según lo demuestra Cooper (1990). Este problema se enfrenta en la práctica con diferentes algoritmos heurísticos que se han ido mejorando a través del tiempo, siendo el desarrollo teórico pionero en este sentido el algoritmo de Chow-Liu (Chow y Liu, 1968). Aún cuando existen muchos métodos que se han desarrollado a partir del algoritmo original de Chow-Liu y que hasta la actualidad siguen en estudio, el algoritmo que más fuerza ha cobrado, tanto en nuevos estudios teóricos, como en aplicaciones prácticas es el de propagación de beliefs (BP), desarrollado inicialmente por Pearl (1988).

Por contraparte, los LTM ofrecen considerables ventajas al momento de hacer inferencia. Por un lado, debido a la estructura de árbol que poseen, el orden del número de cálculos es lineal con respecto al número de variables observables y, por otro, como se ha dicho reiteradamente, estos modelos tienen la potencialidad de representar complejas relaciones entre un conjunto de variables observables a través de la inclusión de variables latentes. Sin embargo, hay que considerar que la complejidad de la inferencia en los LTM no solamente depende en el número de variables observables, sino también en la cantidad y cardinalidad de las variables latentes.

Algoritmos para el aprendizaje de las estructuras de los LTM

Antes de especificar el funcionamiento de cada uno de estos algoritmos, a continuación se explicará el procedimiento para testear el relacionamiento entre nodos. Para esto, el siguiente

lema brinda algunos insumos:

Lema 1. Para las distancias d_{ij} relacionadas a cualquier par de nodos $i, j \in V$, donde V es un árbol de 3 o más nodos, se define la expresión $\Phi_{ijk} = d_{ik} - d_{jk}$, cumple las siguientes propiedades:

- i $\Phi_{ijk} = d_{ij}$ para todo $k \in V \setminus \{i, j\}$ si y solamente si i es nodo terminal y j es su antecesor.
- ii $\Phi_{ijk} = -d_{ij}$ para todo $k \in V \setminus \{i, j\}$ si y solamente si j es nodo terminal e i es su antecesor.
- iii $-d_{ij} < \Phi_{ijk} = \Phi_{ijk'} < d_{ij}$ para todo $k, k' \in V \setminus \{i, j\}$ si y solamente si i y j son nodos terminales y tienen un mismo nodo antecesor¹⁰.

En base a lo anterior, se puede determinar las relación entre los nodos i y j de la siguiente manera:

Fijos los nodos $i, j \in V$, considérense todos los demás nodos $k \in V \setminus \{i, j\}$. Entonces, hay tres posibles casos para el conjunto $\{\Phi_{ijk} : k \in V \setminus \{i, j\}\}$:

1. $\Phi_{ijk} = d_{ij}$ para todo $k \in V \setminus \{i, j\}$. En este caso, i es un nodo terminal y j su antecesor. De acuerdo al lema anterior, si el signo es negativo, la relación es exactamente la contraria.
2. Φ_{ijk} es constante para todo $k \in V \setminus \{i, j\}$ pero distinto de d_{ij} y de $-d_{ij}$. En este caso, i y j son nodos terminales y tienen un antecesor común (nodos hermanos).
3. Φ_{ijk} no es constante para todo $k \in V \setminus \{i, j\}$. Entonces, existen tres posibles subcasos mutuamente excluyentes:

Los nodos i y j no son nodos hermanos y tampoco tienen una relación antecesor-sucesor.

Los nodos i y j son nodos hermanos pero al menos uno de ellos no es un nodo terminal.

Los nodos i y j tienen una relación antecesor-sucesor pero el sucesor no es un nodo terminal.

Algoritmo de agrupamiento recursivo

A continuación se describe el algoritmo de agrupamiento recursivo:

1. Iniciar definiendo al conjunto $Y = V$, es decir, incluyendo a todas las variables observables.
2. Calcular $\Phi_{ijk} = d_{ik} - d_{jk}$, para todo $i, j, k \in Y$. Donde d_{ik} para este caso corresponde al coeficiente de correlación asociado a la arista que une los nodos i y k .

¹⁰La prueba del lema se puede encontrar en el Apéndice A.1 de Choi *et al.* (2011)

3. Utilizando el procedimiento de testeo de relaciones entre nodos explicado anteriormente, se define a $\{\Pi_l\}_{l=1}^L$ como la partición más gruesa de Y tal que para cada subconjunto Π_l (con $|\Pi_l| \geq 2$), cualquier par de nodos en Π_l son o bien nodos hermanos que son nodos terminales, o tienen una relación antecesor-sucesor. Notar que para alguno l , Π_l podría contener un solo nodo. Empezar a construir el nuevo conjunto activo agregando nodos en estas particiones que contienen un solo nodo, como sigue: $Y_N \leftarrow \bigcup_{l:|\Pi_l|=1} \Pi_l$
4. Para cada $l = 1, \dots, L$ con $|\Pi_l| \geq 2$, si Π_l contiene un nodo predecesor u , actualizar $Y^N \leftarrow Y^N \cup \{u\}$. De lo contrario, introducir un nuevo nodo oculto (variable latente) h , y luego conectar h (como un antecesor) a cada nodo contenido en Π_l , y actualizar $Y^N \leftarrow Y^N \cup \{h\}$.
5. Actualizar el conjunto activo: $Y^A \leftarrow Y$ y $Y \leftarrow Y^N$.
6. Para cada nodo oculto $h \in Y$, calcular las distancias d_{hl} , para todo $l \in Y$, utilizando las ecuaciones 12 y 13.
7. Si $|Y| \geq 3$, regresar al paso 2. Caso contrario, si $|Y| = 2$, conectar los dos nodos restantes en Y con una arista y parar. Si $|Y| = 1$, parar.

Sean $i, j \in C(h)$ dos nodos sucesores de h y sea $k \in Y^A \setminus \{i, j\}$ cualquier otro nodo en el conjunto activo previo. Del lema presentado anteriormente, se tiene que $d_{ih} - d_{jh} = d_{ik} - d_{jk} = \Phi_{ijk}$ y $d_{ih} + d_{jh} = d_{ij}$. Con estos antecedentes, se puede calcular la distancia entre un nodo previamente activo $i \in Y^A$ y su nuevo nodo predecesor oculto $h \in Y$ como sigue:

$$d_{ih} = \frac{1}{2}(d_{ij} + \Phi_{ijk}) \tag{12}$$

Para cualquier otro nodo activo $l \in Y$, se puede calcular d_{hl} utilizando un nodo sucesor $i \in C(h)$ de la siguiente manera:

$$d_{hl} = \begin{cases} d_{il} - d_{ih} & \text{si } l \in Y^A \\ d_{ik} - d_{jh} - d_{lk} & \text{si } l \in Y^A \end{cases} \tag{13}$$

Algoritmo de agrupamiento de Chow-Liu con unión de vecinos

El algoritmo de agrupamiento de Chow-Liu con unión de vecinos involucra dos etapas: en primer lugar, se construye el árbol de Chow-Liu sin variables latentes (Chow y Liu, 1968), notado $MST(V; D)$, donde D corresponde a la matriz de distancias. En segundo lugar, se aplica el método de unión de vecinos (Saitou y Nei, 1987) para reconstruir un subárbol latente sobre los vecinos cerrados¹¹ para cada nodo interno en $MST(V; D)$.

Con esta introducción, el algoritmo de agrupamiento de Chow-Liu con unión de vecinos sigue los siguientes pasos:

¹¹Ver más detalles en Choi *et al.* (2011) P.19

1. Construir el árbol de Chow-Liu $MST(V; D)$. Hacer $T \leftarrow MST(V; D)$
2. Identificar el conjunto de nodos internos del $MST(V; D)$.
3. Para cada nodo interno i , se define $nbd[i; T]$ como su vecindario cerrado (todos los nodos vecinos incluido i) en T . Sea $S \leftarrow AR(nbd[i; T], D)$ la salida al aplicar el algoritmo anterior (agrupamiento recursivo) con $nbd[i; T]$ como el conjunto de nodos observables de entrada.
4. Reemplazar el subárbol sobre el conjunto de nodos $nbd[i; T]$ en T con S . Notar con T al nuevo árbol.
5. Repetir los pasos 3 y 4 hasta que todos los nodos internos hayan sido operados.

Explicación del algoritmo BP

En lo que sigue, la explicación se hará en términos de los campos aleatorios de Markov, sin que esto implique una pérdida de generalidad pues, existe una equivalencia entre cualquier campo aleatorio de Markov, una red bayesiana o algún otro modelo gráfico. Además, como se especificó en el artículo, el enfoque de consistencia fue planteado utilizando modelos de árbol con variables latentes (latent tree models), que son casos particulares de campos aleatorios de Markov.

Al ser los nodos observables y_i fijos, se puede simplificar la escritura de $\phi_i(x_i, y_i)$ como $\phi_i(x_i)$, de manera que la atención se centre en la distribución de probabilidad conjunta de las variables desconocidas x_i :

$$P(\{x\}) = \frac{1}{z} \prod_{(ij)} \psi_{ij}(x_i, x_j) \prod_i \phi_i(x_i) \quad (14)$$

En el algoritmo BP, se utilizará la notación $m_{ij}(x_j)$ para referirse al “mensaje” que le manda un nodo escondido i a otro nodo escondido j “contándole” cuál debería ser su estado. En la Figura 13 se ilustra esta notación.

El mensaje $m_{ij}(x_j)$ será un vector de la misma dimensión de x_j donde el valor de cada componente es proporcional a cuan probable el nodo i “cree” que el nodo j esté en el estado correspondiente a esa componente. En el algoritmo BP estándar, el belief en un nodo i es proporcional al producto de la evidencia local en ese nodo, es decir $\phi_i(x_i)$, y todos los mensajes que llegan al nodo i :

$$b_i(x_i) = k\phi_i(x_i) \prod_{j \in N(i)} m_{ji}(x_i) \quad (15)$$

Donde $N(i)$ es el conjunto de todos los nodos comunicados con el nodo i (nodos vecinos) y k es una constante de normalización, pues los beliefs deben sumar 1. Los mensajes quedan determinados de manera recursiva por la siguiente regla de actualización:

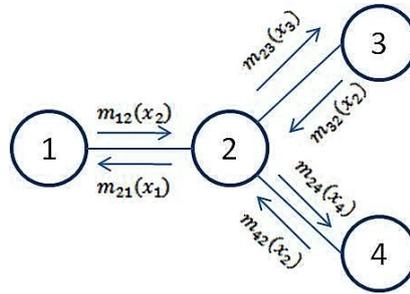


Figura 13: Ilustración del envío de mensajes entre nodos

$$\sum_{x_i} \phi_i(x_i) \psi_{ij}(x_i, x_j) \prod_{k \in N(i) \setminus j} m_{ki}(x_i) \rightarrow m_{ij}(x_j) \quad (16)$$

Donde la sumatoria sobre x_i quiere decir que se suma sobre todos los estados posibles del nodo i .