
Modelo de datos panel para la evasión del sistema de protección social en las principales áreas metropolitanas del país: Estimación Frecuentista y Bayesiana

Panel data model for the evasion of the social protection system in the main metropolitan areas of the country: Frequentist and Bayesian Estimation

Linnethy Julieth Lambraño Pérez^a

Heivar Yesid Rodríguez Pinzón^b

Wilmer Dario Pineda Ríos^c

Resumen

En Colombia la evasión del sistema de protección social, principalmente en la salud y pensión, es un tema que se está abordando con mayor profundidad, ya que tiene cierta influencia en la economía del país, involucrando varios factores en diferentes contextos como el social, cultural, político entre otros, por lo anterior es importante buscar una manera de disminuir los porcentajes de este problemas, una de las alternativas es capturar aquella heterogeneidad no observable de agentes económicos en la dimensión temporal, debido a que los análisis de series de tiempo y análisis de corte transversal no logran controlar aquella heterogeneidad, obteniendo resultados sesgados, se decide optar por los modelos de datos panel. En el presente trabajo se propone comparar el comportamiento de los evasores de las principales áreas metropolitanas de Colombia a través de la estimación frecuentista y bayesiana, identificar el tipo de modelo a utilizar, escoger cuales áreas y variables explicativas tienen un efecto significativo en la variable respuesta.

Palabras clave: *Evasores, Datos Panel, Efectos positivos, Bayesiana.*

Abstract

In Colombia the evasion of the social protection system, mainly in health and pension, is an issue that is being addressed in greater depth, since it has some influence on the country's economy, involving several factors in different contexts such as social, cultural, political among others, for the foregoing it is important to look for a way to reduce the percentages of this problem, one of the alternatives is to capture that unobservable heterogeneity of economic agents in the temporal dimension, because the analyzes of time series and cross-sectional analysis fail to control that heterogeneity, obtaining biased results, it is decided to opt for panel data models. In the present work we propose to compare the behavior of the evaders of the main metropolitan areas of Colombia through frequentist and Bayesian estimation, identify the type of model to be used, choose which areas and explanatory variables have a significant effect on the response variable.

Keywords: *Evaders, Panel Data, Positive effect, Bayesian.*

^aEstudiante de Estadística Universidad Santo Tomás Bogotá

^bDocente de Estadística Universidad Santo Tomás Bogotá

^cDocente de Estadística Universidad Santo Tomás Bogotá

1. Introducción

El uso de modelos econométricos de datos panel tiene una gran importancia en los mercados, involucrando agentes económicos, teniendo en cuenta la dimensión temporal, con el objetivo de encontrar la heterogeneidad no observable y así lograr identificar un fenómeno que a simple vista no se está detectando. Este modelo tiene asociados dos efectos que son excluyentes entre ellos, los cuales son llamados efectos fijos o efectos aleatorios, cada uno de ellos relacionado con el comportamiento de los individuos.

Es este trabajo es de suma de importancia identificar cual estimación es más precisa, si la frecuentista o la bayesiana. En los últimos años, la estadística bayesiana ha sido utilizada en gran medida, cada vez estando más presente en investigaciones, aportando una gran precisión, haciendo que la hora de elegir una estimación esta sea la indicada, por lo anterior se quiere concluir cuál de estas estimaciones es tiene mayor exactitud para el caso de los modelos econométricos de datos panel y por medio de sus intervalos y predicciones decidir cual es la más precisa.

Para los modelos econométricos de datos panel, es necesario realizar la prueba de Hausman para identificar cual efecto sea justa a los datos, una vez encontrado el efecto, se procederá a realizar la estimación frecuentista caracterizando las 13 áreas metropolitanas y periodos trimestrales que aportan un efecto positivo en la variable de interés, obteniendo intervalos de confianza y predicciones, para lograr una comparación, se estimará de forma bayesiana el modelo encontrado, por medio del método MCMC (Markov Chain Monte Carlo), específicamente el Muesteador de Gibbs, calculando sus los intervalos de credibilidad y sus respectiva predicciones, así por último concluir cual estimación es la mejor.

2. Sistema de protección social

Al abordar el contexto económico que enmarca este trabajo, es importante mencionar que la evasión del sistema de protección social en Colombia, principalmente en salud y pensión cada vez va más en aumento, según la **Unidad de Gestión Pensional y Parafiscal (UGPP)**, del total de las persona que están en la obligación de contribuir a los aportes de protección social, en el 2015, solo el 82 %, es decir aproximadamente 12.4 millones de personas lo hicieron, mientras que el 18 % restante no efectuaron el aporte. Según la directora de UGPP Gloria Inés Cortes Arango, afirma lo siguiente “ *Los aportes a la seguridad social son mucho más que un impuesto, es un aseguramiento contra los riesgos de enfermedad, vejez, invalidez y muerte los más beneficiados con este aseguramiento son el cotizante y su familia, no es el Gobierno Nacional*”¹

El artículo 1º de la Ley 789 de 2002 se define el Sistema de la Protección Social como el “Conjunto de políticas públicas destinadas a disminuir la vulnerabilidad y a mejorar la calidad de vida de los colombianos, especialmente de los más desprotegidos para obtener como mínimo el derecho a la salud, la pensión y el trabajo.”

La financiación y administración de este Sistema de Protección Social en Colombia se lleva por medio del régimen contributivo de salud, con vinculaciones mediante el Fondo de Solidaridad y Garantías - FOSYGA, con lo cual, la evasión del pago en los aportes genera inestabilidad financiera del sistema, afectando la cobertura de las prestaciones de carácter económico, de salud y servicios complementarios por parte del Estado, la sociedad, las instituciones respectivas.

Este Sistema tiene dos grandes componentes: el Sistema de seguridad social integral y el Sistema de asistencia social; estos dos sistemas son complementarios, todos los colombianos deben estar incluidos en el Sistema de protección social: como cotizantes, aquellos individuos que tienen ingresos y aportan a la seguridad social y pertenecen al Régimen Contributivo, y como beneficiarios de la seguridad social o de

¹<https://www.incp.org.co/las-consecuencias-de-la-evasion-pensional-para-el-sistema-tributario-en-colombia/>

los programas de asistencia social, aquellos individuos que no tienen ingresos y pertenecen al Régimen Subsidiado.

El Sistema de Protección Social está conformado por seis subsistemas:

1. **Sistema General de Seguridad Social en Salud:** Es la forma como se brinda un seguro que cubre los gastos de salud a los habitantes del territorio nacional, colombianos y extranjeros. Usted se afilia a la Entidad Promotora de Salud (EPS) que prefiera, recibe su Carnet de Salud y entra a formar parte del Sistema de Seguridad Social en Salud. Los servicios los recibe a través de una Institución Prestadora de Servicios (IPS), que como verá más adelante, le traerá muchos beneficios.
2. **Sistema General de Seguridad Social en Pensiones:** Tiene por objeto garantizar a la población, el amparo contra las contingencias derivadas de la vejez, la invalidez y la muerte, mediante el reconocimiento de las pensiones y prestaciones que se determinan en la presente ley, así como propender por la ampliación progresiva de cobertura a los segmentos de población no cubiertos con un sistema de pensiones
3. **Sistema General de Riesgos Laborales:** Propósito principal del Sistema General de Riesgos Laborales (SGRL) es dar una respuesta estructurada a las necesidades de prevenir, proteger y atender a los trabajadores, frente a las enfermedades y los accidentes que puedan ocurrir con ocasión o como consecuencia del trabajo que desarrollan. El Sistema se basa en un esquema de aseguramiento para la población trabajadora, que cubre las contingencias generadas por accidentes de trabajo y enfermedades laborales (ATEL), mediante prestaciones asistenciales y económicas.
4. **Instituto Colombiano de Bienestar Familiar - ICBF:** Es la entidad del estado colombiano que trabaja por la prevención y protección integral de la primera infancia, la niñez, la adolescencia y el bienestar de las familias en Colombia. ICBF brinda atención a niños y niñas, adolescentes y familias, especialmente a aquellos en condiciones de amenaza, inobservancia o vulneración de sus derechos.
5. **Régimen del Subsidio Familiar - Cajas de Compensación Familiar:** Se ha encargado del reconocimiento de una prestación social para los trabajadores, administrado por las Cajas de Compensación Familiar, que cumple una importante labor en materia de equidad, inclusión social y redistribución del ingreso. El Sistema de Subsidio Familiar es el resultado de un pacto voluntario entre empresarios y sindicatos para mejorar la calidad de vida de la población trabajadora, constituyéndose como un escenario novedoso de carácter tripartito para la aplicación de políticas sociales.
6. **Servicio Nacional de Aprendizaje-SENA:** Es la institución pública colombiana encargada de la función que le corresponde al Estado de invertir en el desarrollo social y técnico de los colombianos ofreciendo y ejecutando la Formación Profesional gratuita, para la incorporación y el desarrollo de las personas en actividades productivas que contribuyan al desarrollo social, económico y tecnológico del país. El SENA es la principal institución del Sistema Nacional de Formación para el Trabajo. **Ruiz, A(2016)**

Para resumir un poco y tener una idea más clara, a continuación se presentará un mapa conceptual de lo que es el sistema de protección social.



Figura 1: Elaboración propia

3. Modelo econométrico de datos panel

Un modelo econométrico de datos panel es aquel que incluye una muestra de agentes económicos o de interés (individuos, empresas, bancos, áreas metropolitanas, países, entre otros) en un período de tiempo determinado, combinando dos tipos de datos (dimensión temporal y estructural). El principal objetivo de aplicar y estudiar los datos en panel, es capturar la heterogeneidad no observable, entre agentes económicos, así como también en el tiempo, debido a que esta heterogeneidad no se puede detectar con estudios de series temporales ni con estudios de corte transversal. La aplicación de esta metodología permite analizar dos aspectos importantes: *Efectos individuales específicos*² y *Efectos temporales*³ (Baronio, A. M.)

El modelo de regresión general de datos panel es:

$$Y_{it} = a_{it} + b_1X_{1it} + b_2X_{2it} + \dots + b_kX_{kit} + U_{it} \quad \text{con } i=1,\dots,n \text{ y } t=1,\dots,T \quad (1)$$

Donde:

- i se refiere al individuo o a la unidad de estudio.
- t a la dimensión tiempo, a es un vector de interceptos que puede contener entre 1 y $n+t$ parámetros.
- b es un vector de K parámetros.
- a Vector de interceptos que pueden contener entre 1 y $n+t$ parámetros.
- X_{it} es la i -ésima observación al momento t para las K variables explicativas X_1, X_2, \dots, X_K .
- U_{it} es el término del error.

- **Componentes del error:** El término del error U_{it} se puede descomponer de la siguiente forma:

$$U_{it} = u_i + v_t + w_{it} \quad (2)$$

²Son aquellos que afectan de manera desigual a cada uno de los agentes de estudio contenidos en la muestra (individuos, empresas, países) los cuales son invariables en el tiempo y que afectan de manera directa las decisiones que tomen dichas unidades.

³Son aquellos que afectan por igual a todas las unidades individuales del estudio.

Donde:

- u_i representa los efectos no observables que difieren entre las unidades de estudio pero no en el tiempo, que generalmente se asocia a la tecnología incorporada.
- v_i se le identifica con efectos no cuantificables que varían en el tiempo, pero no entre las unidades de estudio
- w_{it} se refiere al término de error puramente aleatorio

3.1. Modelo de efectos fijos

Este modelo considera que existe un término constante diferente para cada individuo y supone que los efectos individuales son independientes entre sí. Con este modelo se considera que las variables explicativas afectan por igual a las unidades de corte transversal y que éstas se diferencian por características propias de cada una de ellas, medidas por medio del intercepto. Es por ello que los n interceptos se asocian con variables dummy con coeficientes específicos para cada unidad, los cuales se deben estimar. Para la i -ésima unidad de corte transversal, la relación es la siguiente:

$$Y_{it} = ia_i + b_1X_{1i} + b_2X_{2i} + \dots + b_kX_{ki} + U_i \quad \text{con } i=1,\dots,n \quad (3)$$

Donde el subíndice i representa un vector columna de unos y U_i representa el componente del error.

3.2. Modelo de efectos aleatorios

Considera que los efectos individuales no son independientes entre sí, es decir que considera que los efectos individuales están distribuidos alrededor de un valor dado, sino que están distribuidos aleatoriamente alrededor de un valor dado. Comúnmente en el análisis de regresión se asume que el gran número de factores que afecta el valor de las variables son dependientes, pero que no han sido incluidas explícitamente como variables independientes del modelo, resumiéndose en la perturbación aleatoria. Así, con este modelo se considera que tanto el impacto de las variables explicativas como las características propias de cada unidad de corte transversal son diferentes. El modelo se expresa algebraicamente de la siguiente forma:

me d

$$Y_{it} = a + b_iX_{1it} + b_2X_{2it} + \dots + b_kX_{kit} + U_{it} \quad \text{con } i=1,\dots,n \text{ y } t=1,\dots,T \quad (4)$$

Donde:

$U_{it} = u_i + v_t + w_{it}$ se convierte en el nuevo término de la perturbación, donde U_{it} no es homocedástico, u_i , v_t , w_{it} corresponden al error asociado con las series de tiempo (v_t), a la perturbación de corte transversal (u_i) y el efecto aleatorio combinando ambas (w_{it}).

4. Test de Hausman

El test propuesto por Hausman en el año 1978, es un test Chi cuadrado que determina si las diferencias son sistemáticas y significativas entre dos estimaciones, comparando los β obtenidos por medio del estimador de efectos fijos y efectos aleatorios, identificando si las diferencias entre ellos son o no son significativas. En la hipótesis nula se comprueba la existencia de no correlación entre los α_i y las variables explicativas.

H_0 = No existe correlación entre las variables explicativas y los efectos individuales.

H_1 = Existe correlación entre las variables explicativas y los efectos individuales.

Criterio de rechazo

- Si el p-value < 0.05 rechazo H_0 , es decir que hay correlación entre los efectos individuales y las variables explicativas, lo que indica que el estimador fijo debe ser utilizado.
- Si el p-value > 0.05 , no se rechaza H_0 , es decir que no hay una correlación entre los efectos individuales y las variables explicativas, lo que indica que el estimador de efectos aleatorios es el que se debe utilizar.

5. Estimación bayesiana

Según lo expresado por Koop, G. (2003) la econometría bayesiana se basa en las reglas de probabilidad. Esto es una de las principales ventajas del enfoque bayesiano. Las reglas de probabilidad implican:

$$p(A, B) = p(A | B)p(B) \quad (5)$$

Donde $p(A, B)$ es la probabilidad conjunta de que A y B ocurran, $p(A | B)$ es la probabilidad de que ocurra A condicionada de B que haya ocurrido y $p(B)$ es la probabilidad marginal de B.

La expresión anterior, se puede revertir la posición de A y B y encontrar la probabilidad conjunta de A,B, que es la siguiente:

$$p(A, B) = p(B | A)p(A) \quad (6)$$

Al igualar la expresión (5) y (6) para $p(A, B)$ y reorganizarla da como resultado la Regla de Bayes que se muestra a continuación:

$$p(B | A) = \frac{p(A | B)p(B)}{p(A)} \quad (7)$$

En los modelos econométricos, el interés se centra en los coeficientes de la regresión, principalmente en la estimación de estos coeficientes, siendo los parámetros de estudio, Y un vector o matriz de los datos y θ un vector o matriz que contienen los parámetros.

La econometría bayesiana usa la regla Bayes que se expresa a continuación:

$$p(\theta | Y) = \frac{p(Y | \theta)p(\theta)}{p(Y)} \quad (8)$$

El interés se basa específicamente en $p(\theta | Y)$, donde se maneja a θ como una variable aleatorias, siendo esto refutado por los frecuentistas, ya que no la consideran como una variable aleatoria. La econometría bayesiana se basa en una visión subjetiva de la probabilidad, que argumenta la incertidumbre sobre cualquier cosa desconocida usando las reglas de la probabilidad.

Ignorando el $p(y)$ se puede escribir de la siguiente manera la ecuación presentada anteriormente:

$$p(\theta | Y) \propto p(Y | \theta)p(\theta) \tag{9}$$

Siendo $p(\theta | Y)$ se refiere a la densidad posterior, $p(Y | \theta)$ la función de verosimilitud y $p(\theta)$ la densidad prior

5.1. Muestreador de Gibbs

Según Ausín, C. (2012) el muestreador de Gibbs es un caso particular del algoritmo de Metropolis Hasting, que consiste en tener bloques las distribuciones propuestas que coincidan con las distribuciones posteriores condicionadas, de modo que la probabilidad de aceptación sea uno. Se supone que el conjunto de parámetros es $\beta = \beta_1, \dots, \beta_k$ y que para todo $i = 1, \dots, k$ es fácil de simular de la distribución posterior condicional $\pi(\beta_i | \beta_1, \dots, \beta_{i-1}, \beta_{i+1}, \dots, \beta_k)$ como se explica a continuación:

1. Se fija un valor inicial $\beta_1^0, \dots, \beta_k^0$
2. Se simula $\beta_1^{(t+1)} \sim \pi(\beta_1 | \beta_2^t, \dots, \beta_k^t)$
3. Se simula $\beta_2^{(t+1)} \sim \pi(\beta_2 | \beta_1^{(t+1)}, \beta_3^t, \dots, \beta_k^t)$
4. Se simula $\beta_3^{(t+1)} \sim \pi(\beta_3 | \beta_1^{(t+1)}, \beta_2^{(t+1)}, \beta_4^t, \dots, \beta_k^t)$
5. Así sucesivamente
6. Se simula $(\beta_k^{(t+1)} \sim \pi(\beta_k | \beta_1^{(t+1)}, \dots, \beta_{(k-1)}^{(t+1)}))$

5.2. Diagnósticos de convergencia

Posterior a la estimación bayesiana de los parámetros del modelo, se busca validar las estimaciones realizadas bajo la convergencia de las cadenas, siendo los diagnósticos de convergencia la herramienta que cumple con este objetivo. Por tanto, a continuación se presentan tres de los criterios más destacados dentro de la literatura:

- **Heidelberger and Welch:** Es un diagnóstico de control de longitud de ejecución basado en un criterio de precisión relativa para la estimación de la media. El ajuste predeterminado corresponde a una precisión relativa de dos dígitos significativos. También elimina hasta la mitad de la cadena para asegurar que los medios se estimen a partir de una cadena que ha convergido. (P PD 1981)
- **Raftery and Lewis:** Es un diagnóstico de control de longitud de ejecución basado en un criterio de exactitud de estimación del cuantil q. Está pensado para su uso en una prueba piloto corta de una cadena de Markov. También calcula el número de iteraciones de "quemado" que se descartarán al principio de la cadena. (Raftery Lewis 1995)
- **Geweke:** Diagnóstico de convergencia para cadenas de Markov basado en una prueba de igualdad de los medios de la primera y última parte de una cadena de Markov (por defecto el primer 10 % y el último 50 %). Si las muestras se extraen de la distribución estacionaria de la cadena, los dos medios son iguales y la estadística de Geweke tiene una distribución normal asintóticamente estándar. (Geweke 1992) **García, Repositorio Universidad Santo Tomás. (2017)**

6. Resultados y discusión

En las secciones anteriores se fundamentó la parte teórica que sustentan la estimación frecuentista y bayesiana de un modelo econométrico datos panel, por tanto en esta sección se evidenciarán los resultados

de la aplicación de los modelos iniciando con la contextualización de la información a utilizar, hasta los resultados de la estimación frecuentista y bayesiana de los parámetros del modelo con sus respectivos intervalos y predicciones.

6.1. Análisis Descriptivo

Los datos utilizados para la aplicación de este trabajo tienen como fuente la Base de Datos Única de Afiliados (BDUA). Para la aplicación del modelo se tomó la base de datos discriminada por las principales 13 áreas metropolitanas del país y por un total de 15 periodos trimestrales comenzando en el año 2012 y finalizando en el año 2015, las variables a tener en cuenta será el número de personas Evasoras, Contribuyentes, Subsidiadas, Ocupadas y Cesantes.

Dado lo anterior, se cuenta con la siguiente información:

- **Variable Respuesta (Evasores):** Cantidad de personas evasoras del sistema de protección social reportadas en el DANE a nivel de las principales áreas metropolitanas del país.
- Las 13 áreas metropolitanas⁴ de Colombia a tener en cuenta son: Barranquilla, Bogotá, Bucaramanga, Cali, Cartagena, Cúcuta, Ibagué, Manizales, Medellín, Montería, Pasto, Pereira y Villavicencio.

De esta manera, a continuación se presenta visualmente el comportamiento dada la descripción anterior:

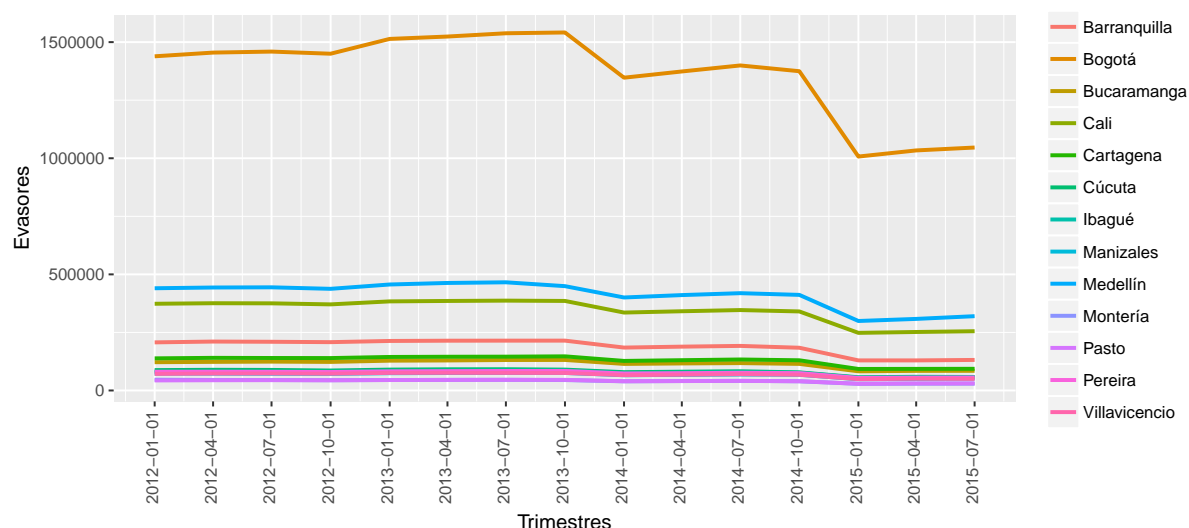


Figura 2: Evasores por áreas y trimestres

A través de *Figura 2*, se logra evidenciar un comportamiento destacado del área metropolitana de Bogotá frente a las demás áreas en cuanto a la cantidad de personas evasoras del sistema de protección social. Adicional a esto, se observa que las áreas metropolitanas de Medellín y Cali a pesar de tener un número inferior de personas evasoras pertenecen a las áreas metropolitanas con mayor número de evasores, estas tres áreas son consideradas como principales por tener un elevado número de habitantes. Las áreas metropolitanas de Villavicencio y Montería son las que presentan un bajo número de personas evasoras frente a las demás áreas metropolitanas observadas.

⁴Se define como el área de influencia que incluye municipios circundantes, que con la ciudad conforman un solo tejido urbano no discontinuo y han sido reconocidos legalmente.

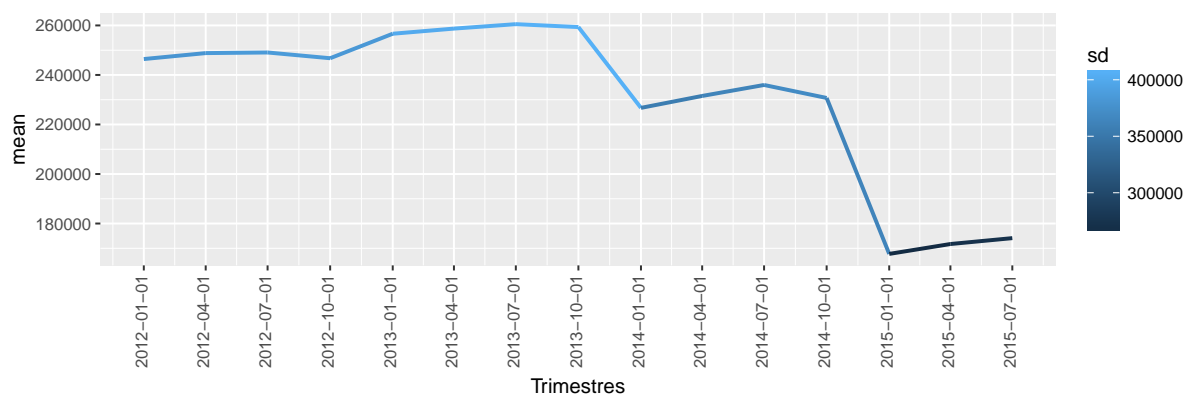


Figura 3: Heterogeneidad por trimestres

En la *Figura 3*, se observa que en el trimestre (Julio-Agosto-Septiembre) del año 2013 registra el promedio más alto de evasores frente a los demás periodos, se logra evidenciar que al pasar los periodos trimestrales el promedio de evasores va disminuyendo hasta el primer trimestre del 2015, en los próximos dos trimestres de este mismo años este promedio comienza aumentar.

Las covariables que se tendrán en cuenta para el planteamiento del modelo son:

■ **Subsidiados:**

Esta variable hace referencia al número de personas que están en el régimen subsidiado, la afiliación está dirigido a la población pobre y vulnerable del país en el Sistema General de Seguridad Social en Salud (SGSSS), el Estado colombiano ha definido al régimen subsidiado en salud como vía de acceso efectiva al ejercicio del derecho fundamental de la salud, los municipios, distritos y departamentos tienen funciones específicas frente a la identificación y afiliación de la población objeto, así como sobre la inversión, contratación y seguimiento de la ejecución de los recursos que financian el régimen (recursos de esfuerzo propio, de la nación (SGP) y del Fosyga).⁵

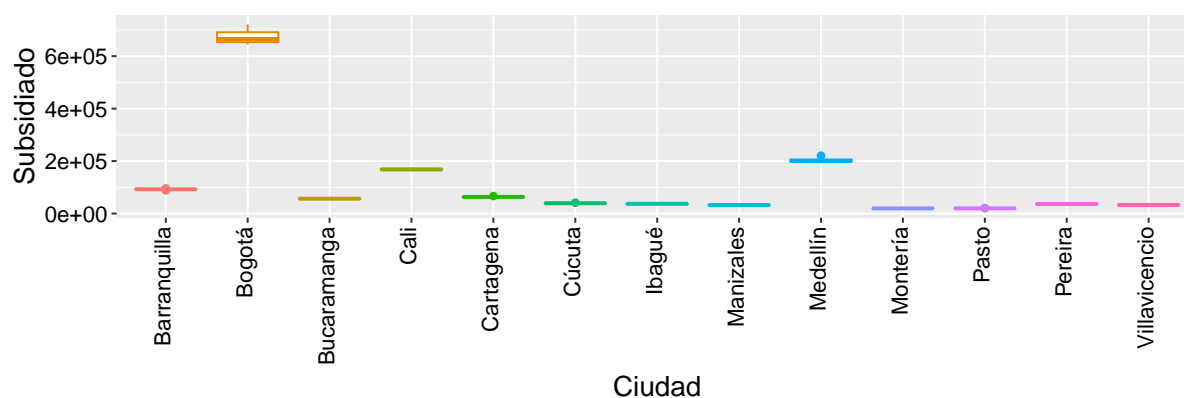


Figura 4: Subsidiados vs Áreas Metropolitanas

Con la contextualización sobre este régimen, se analiza que en la *Figura 4* el área en donde hay una mayor cantidad pertenecientes a este régimen es Bogotá, seguido del área de Medellín y Cali, las que menos reportan es Montería y Pasto.

⁵www.minsalud.gov.co/proteccionsocial/Regimensubsidiado/Paginas/regimen-subsidiado.aspx

■ **Contributivo:**

Esta variable hace referencia al número de personas que están en el régimen contributivo, definiéndose como: Las personas que tienen capacidad de pago, es decir aquellas vinculadas a través de contrato de trabajo, los servidores públicos, los pensionados, jubilados y los trabajadores independientes con capacidad de pago, entre los cuales están incluidos, madre comunitaria o sustituta, aprendices en etapa electiva, aprendices en etapa productiva.

Por regla general las personas afiliadas al sistema general de seguridad social en salud a través del régimen contributivo tienen la obligación de cotizar igualmente al régimen de pensiones y sobre el salario o ingresos por prestación de servicios que devenguen.

Estas personas deben hacer un aporte mensual (cotización) a una Entidad Promotora de Salud, para que ésta les garantice la atención en salud a través de las instituciones prestadoras de servicios de salud, conocidas como IPS.⁶

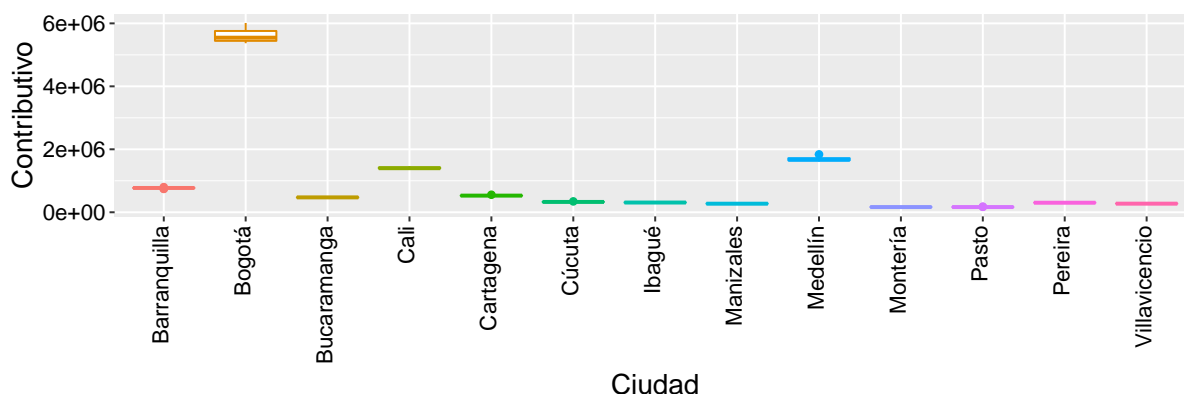


Figura 5: Contributivo vs Áreas Metropolitanas

En la *Figura 5* es importante resaltar que la variable contributivo está teniendo un comportamiento muy similar al de la variable subsidiado, en donde las tres áreas con mayor número de personas en regímenes contributivo es Bogotá, Medellín y Cali y permaneciendo Montería y Pasto como las que menos cantidad de personas tiene.

■ **Cesantes:**

Es la persona que habiendo trabajado antes por lo menos durante dos semanas consecutivas se encuentra desocupada buscando empleo.⁷

⁶ www.minsalud.gov.co/proteccionsocial/Regimensubsidiado/Paginas/aseguramiento-al-sistema-general-salud.aspx

⁷ https://www.dane.gov.co/files/faqs/faq_ech.pdf

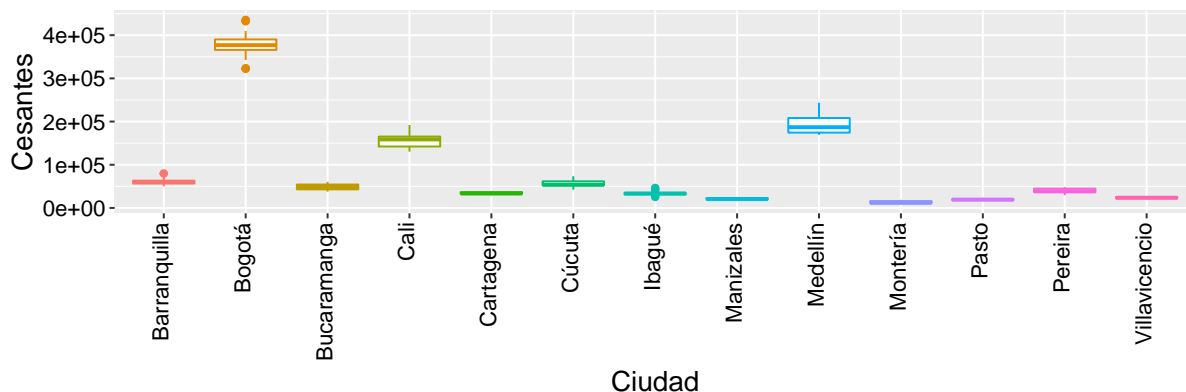


Figura 6: Cesantes vs Áreas Metropolitanas

En la *Figura 6* se observa que el área de Bogotá es la que tiene un mayor número de cesantes de seguridad social con respecto a las demás áreas, seguido de la ciudad de Medellín y Cali, estas siendo las principales áreas del país, aquellas que presentan menor cantidad de cesantes son Montería y Pasto.

■ **Ocupados:**

Son las personas que durante el período de referencia se encontraban en una de las siguientes situaciones:

1. Trabajó por lo menos una hora remunerada en dinero o en especie en la semana de referencia.
2. Los que no trabajaron la semana de referencia, pero tenían un trabajo.
3. Trabajadores familiares sin remuneración que trabajaron en la semana de referencia por lo menos 1 hora.⁸

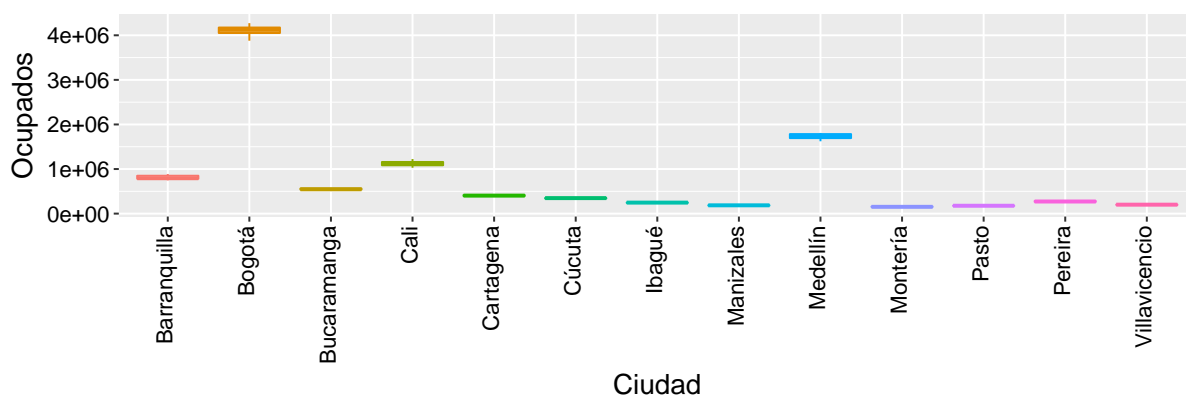


Figura 7: Ocupados vs Áreas Metropolitanas

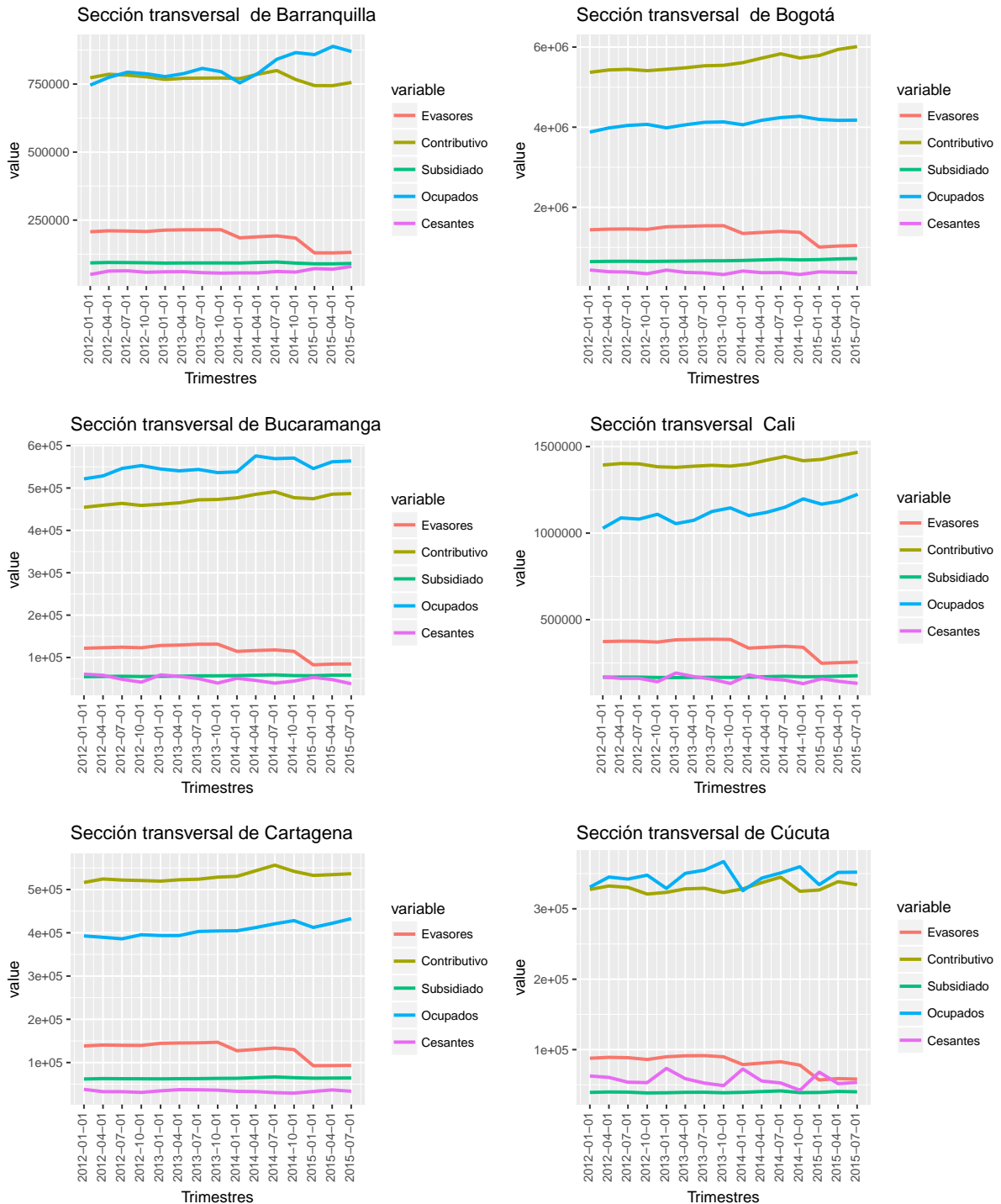
En la *Figura 7* se observa que el área de Bogotá es la que tiene un mayor número de personas ocupadas con respecto a las demás áreas, seguido de la ciudad de Medellín y Cali, estas siendo

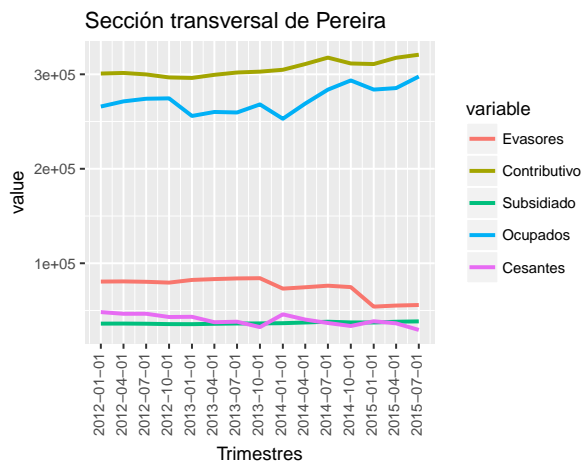
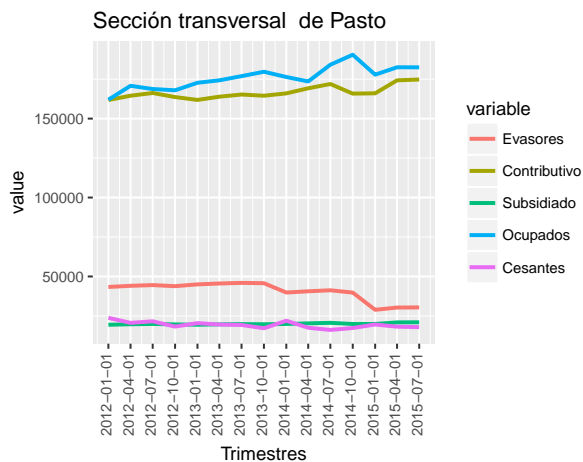
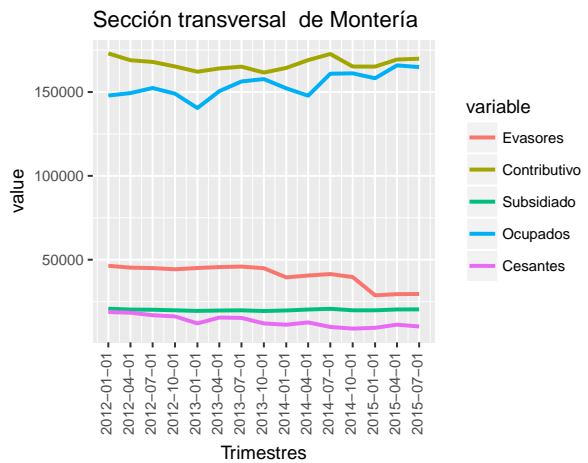
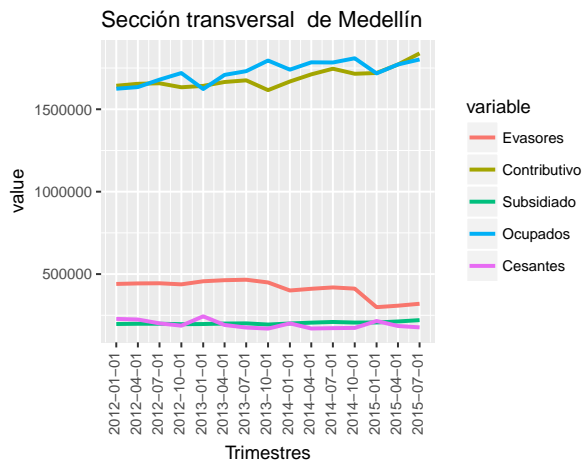
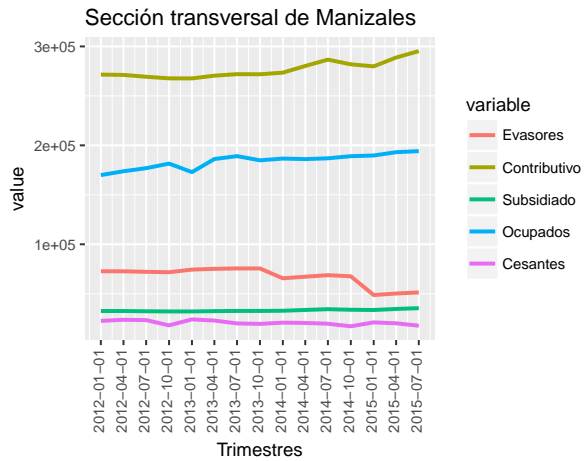
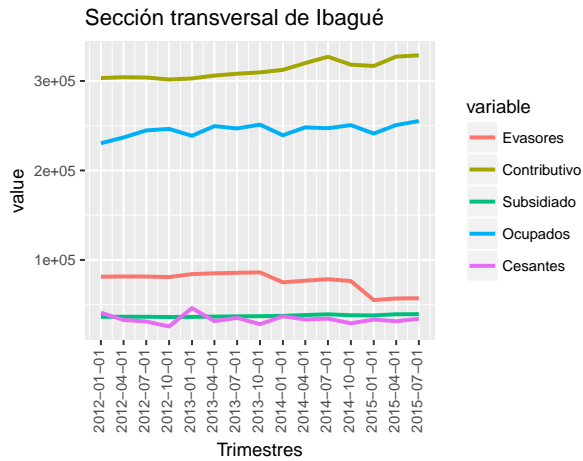
⁸https://www.dane.gov.co/files/faqs/faq_ch.pdf

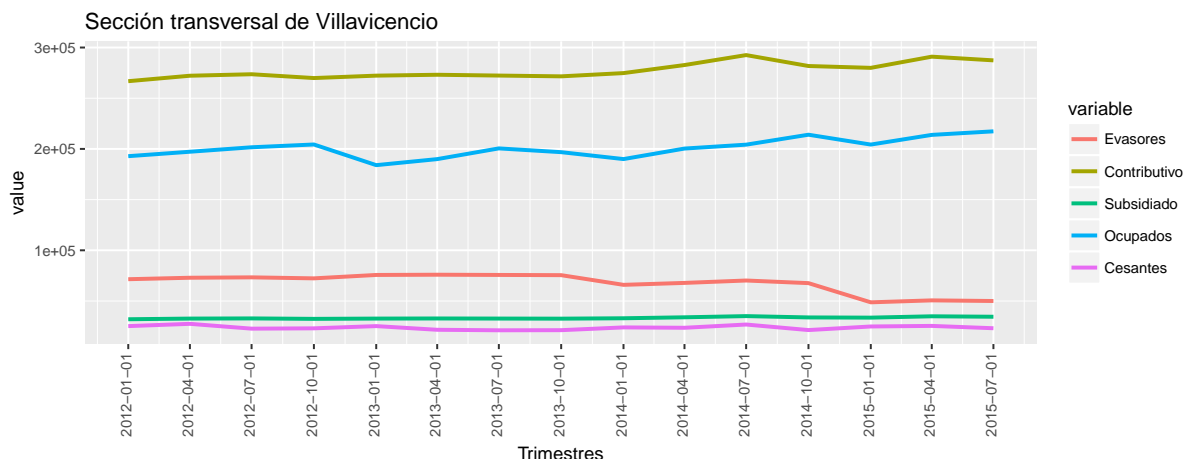
las principales áreas del país, las áreas que presentan menor cantidad de ocupados son Montería y Pasto.

6.1.1. Sección transversal individual

A continuación se presentará el comportamiento de las variables nombradas anteriormente para las 13 áreas metropolitanas del país.







En las gráficas presentadas con anterioridad, se evidencia que la cantidad de personas que pertenecen al régimen contributivo y que están ocupados son muy similar, esto tiene sentido ya que cada persona que tiene un trabajo, debe estar en el régimen contributivo, en estas 13 áreas el comportamiento de la evasión a lo largo de los periodos va disminuyendo, es importante resaltar que la cantidad de personas evasoras por áreas es menor al número de personas pertenecientes al régimen contributivo, lo ideal es que el número de personas evasoras fueran casi nulo, por último las personas que son de régimen subsidiado o cesantes son las que reporta una menor número tiene en cada ciudad.

6.2. Estimación Frecuentista

La estimación de este modelo se realizará por medio del software R, utilizando la función *plm()* se procederá a generar las estimaciones de los modelos con efectos fijos y aleatorios, después con la estimación de estos dos modelos se realizará el test de hausman con la función *phptest()* y así decidir cual de los efectos es el indicado para el modelo.

Contraste de hipótesis

$$H_0 : corr(\alpha_i, X) = 0$$

$$H_1 : corr(\alpha_i, X) \neq 0$$

Tabla 1: Test de Hausman
Hausman Test
dt = 4, chisq = 562.12, p < 2.2e-16

Con el resultado anterior, se evidencia que el valor p es menor a 0.05, lo cual se concluye que se debe utilizar un modelo de efectos fijos, es decir que se considera un término constante diferente para cada individuo e independiente entre sí.

6.2.1. Modelo de efectos fijos

$$Evasores_i = \alpha_i + b_1 Subsidiado_i + b_2 Contributivo_i + b_3 Ocupados_i + b_4 Cesantes_i + U_i \quad \text{con } i=1, \dots, n \quad (10)$$

a) Estimación del modelo de efectos fijos

Tabla 2: Modelo de Efectos Fijos

<i>Balanced Panel: n=13, T=15, N=195</i>				
Coefficientes	Estimate	Std. Error	t-value	Pr(> t)
Ocupados	-0.27897	0.10984	-2.5398	0.01195 *
Cesantes	-1.36273	0.26738	-5.0965	8.78e - 07 ***
Total Sum of Squares: 5.9304e+11				
Residual Sum of Squares: 2.0824e+11				
R-Squared: 0.64885				
Adj. R-Squared: 0.61729				
F Statistic 82.228*** (df = 4; 178)				
*p<0.1; **p<0.05; ***p<0.01				

Se evidencia que en el modelo las variable ocupados y cesantes son estadísticamente significativas, el 61 % de la variabilidad de los evasores son debido a las variables mencionadas anteriormente.

En este modelo se observa que cuando el número de personas ocupadas aumentan en mil, se espera que disminuya en aproximadamente 278 personas evasoras, para los cesantes cuando el número de personas en aumenta en mil, se espera que disminuya en aproximadamente en 1362 personas evasoras.

b) Prueba de efectos fijos en el tiempo

Es importante conocer si el modelo a ajustar debe tener un efecto temporal, a través de la función *plmtest()* se realizará la prueba de efectos individuales o temporales.

Contraste de hipótesis

$$H_0 : \text{Efectos fijos individuales}$$

$$H_1 : \text{Efectos fijos en el tiempo}$$

Tabla 3: Prueba sobre para efectos individuales o temporales

Lagrange Multiplier Test - time effects (Breusch-Pagan) for balanced panels
chisq = 70.93, df = 1, p-value < 2.2e-16

La *tabla 3* muestra el resultado de la comparación de los modelos indicados anteriormente, como el *valor - p* < 0.05, entonces no se acepta la hipótesis nula y se concluye que se debe utilizar efectos fijos en el tiempo.

c) Estimación del Modelo de efectos fijos temporales

Tabla 4: Modelo de efectos fijos temporales

<i>Balanced Panel: n=13, T=15, N=195</i>				
Coeficientes				
	Estimate	Std. Error	t-value	$Pr(> t)$
Cesantes	-1.05906	0.25230	-4.1977	4.41e-05 ***
Total Sum of Squares: 3.977e+11				
Residual Sum of Squares: 1.3756e+11				
R-Squared: 0.65411				
Adj. R-Squared: 0.59084				
F Statistic 77.5362 (df = 4; 164)				
*p<0.1; **p<0.05; ***p<0.01				

La *tabla 4* muestra el resultado de modelo de efectos fijos temporales, se evidencia que la variable que es significativa es cesantes, en donde nos indica que cuanto en número de personas cesantes aumentan en mil, se espera que los evasores disminuyan en aproximadamente en una persona.

d) Pruebas del modelo

■ Pruebas de heterocedasticidad

Contraste de hipótesis

H_0 : Los residuales tiene varianza constante
 H_1 : Los residuales no tiene varianza constante

Tabla 5: Test de heterocedasticidad

Studentized Breusch-Pagan test
BP = 100.68, df = 18, p-value = 1.662e-13

Al analizar el p-value, se concluye que existe presencia de heterogeneidad en los residuales del modelo

■ Prueba de Normalidad

Contraste de hipótesis

H_0 : Los residuales siguen normalidad
 H_1 : Los residuales no siguen normalidad

Tabla 6: Prueba de normalidad

Adjusted Jarque-Bera test for normality
AJB = 2163.8, p-value < 2.2e-16

Al analizar el p-value, se concluye que no hay normalidad en los residuales del modelo

■ Pruebas de la dependencia transversal

Contraste de hipótesis

H_0 : No hay dependencia en la sección transversal

H_1 : Hay dependencia en la sección transversal

Tabla 7: Prueba de dependencia transversal

Breusch-Pagan LM test for cross-sectional dependence in panels
chisq = 509.86, df = 78, p-value < 2.2e-16

Al analizar el p-value, se concluye los residuales si dependen de la sección transversal.

■ Pruebas de correlación serial

Contraste de hipótesis

H_0 : No hay correlación serial en los residuales

H_1 : No hay correlación serial en los residuales

Tabla 8: Prueba de correlacion serial

Breusch-Godfrey/Wooldridge test for serial correlation in panel models
chisq = 84.341, df = 15, p-value = 1.113e-11

Al analizar el p-value, se concluye existe correlación serial en los residuales

e) Efectos de las áreas metropolitanas sección transversal

A continuación, se muestra los efectos de sección transversal por áreas y por periodos trimestrales.

Tabla 9: Efectos por áreas metropolitanas

d	Áreas metropolitanas	Estimate	Std. Error	$Pr(> t)$
d_1	Barranquilla	-183559	77872	0.01959 *
d_2	Bogotá	4949089	405025	< 2.2e-16 ***
d_3	Bucaramanga	-499199	54669	2.432e-16 ***
d_4	Cali	537006	120009	1.425e-05 ***
d_5	Cartagena	-476148	39292	< 2.2e-16 ***
d_6	Cúcuta	-642681	39744	< 2.2e-16 ***
d_7	Ibagué	-692695	27041	< 2.2e-16 ***
d_8	Manizales	-747030	20565	< 2.2e-16 ***
d_9	Medellín	896991	178860	1.365e-06 ***
d_{10}	Montería	-857827	17274	< 2.2e-16 ***
d_{11}	Pasto	-851563	19873	< 2.2e-16 ***
d_{12}	Pereira	-690321	30389	< 2.2e-16 ***
d_{13}	Villavicencio	-742062	22180	< 2.2e-16 ***

En la *Tabla 9* se evidencia que las áreas que tiene un efecto positivo son Bogotá, Cali y Medellín, todas estas áreas son significativas.

f) Efectos de los periodos sección transversal

Tabla 10: Efectos por trimestres

f	Trimestres	Estimate	Std. Error	$Pr(> t)$
f_{12-1}	Ene - Mar 2012	3713.53	79098.98	0.9626
f_{12-2}	Abr - Jun 2012	8282.21	79818.90	0.9175
f_{12-3}	Jul - Sep 2012	6022.32	80005.95	0.9401
f_{12-4}	Oct - Dic 2012	-10756.73	79307.22	0.8923
f_{13-1}	Ene - Mar 2013	19342.72	81002.80	0.8116
f_{13-2}	Abr - Jun 2013	16375.07	80124.35	0.8383
f_{13-3}	Jul - Sep 2013	18659.71	80532.39	0.8171
f_{13-4}	Oct - Dic 2013	6847.18	79677.60	0.9316
f_{14-1}	Ene - Mar 2014	-2110.16	81092.43	0.9793
f_{14-2}	Abr - Jun 2014	8554.23	81115.47	0.9161
f_{14-3}	Jul - Sep 2014	27175.69	82091.42	0.7410
f_{14-4}	Oct - Dic 2014	524.64	82084.24	0.9949
f_{15-1}	Ene - Mar 2015	-47544.52	82950.69	0.5673
f_{15-2}	Abr - Jun 2015	-32121.48	82704.99	0.6982
f_{15-3}	Jul - Sep 2015	-22964.42	82713.30	0.7816

En la *Tabla 10* se observa que los trimestres Oct-Dic 2012, Ene-Mar 2014, Abr-Jun 2015 y Jul-Sep 2015, tienen un efecto negativo en la estimación de la evasión, se evidencia que ninguno de los periodos son significativos.

Con lo expuesto anteriormente, se puede decir que las áreas metropolitanas que tiene un efecto positivo en la evasión son Bogotá, Medellín y Cali, los trimestres que presentan un efecto negativo en la evasión son el tercer trimestre del 2012, el primer trimestre del 2014 y el segundo y tercer trimestre del 2015.

g) Ecuación para estimar la evasión

A continuación se presentará la ecuación encontrada para la estimación de la variable respuesta, es decir la evasión.

$$\begin{aligned}
\text{Evasión} = & -(1.05906 \times \text{Cesantes}) - (183559 \times d_1) + (4949089 \times d_2) - (499199 \times d_3) + \\
& (537006 \times d_4) - (476148 \times d_5) - (642681 \times d_6) - (692695 \times d_7) - (747030 \times d_8) + \\
& (896991 \times d_9) - (857827,7 \times d_{10}) - (851563 \times d_{11}) - (690321 \times d_{12}) - (742062 \times d_{13}) + \\
& (3713.53 \times f_{12-1}) + (8282.21 \times f_{12-2}) + (6022.32 \times f_{12-3}) - (10756.73 \times f_{12-4}) + \\
& (19342.72 \times f_{13-1}) + (16375.07 \times f_{13-2}) + (18659.71 \times f_{13-3}) + (6847.18 \times f_{13-4}) - \\
& (2110.16 \times f_{14-1}) + (8554.23 \times f_{14-2}) + (27175.69 \times f_{14-3}) - (524.64 \times f_{14-4}) - \\
& (47544.52 \times f_{15-1}) - (32121.48 \times f_{15-2}) - (22964.42 \times f_{15-3})
\end{aligned}$$

6.3. Estimación Bayesiana

Para continuar con el objetivo de este proyecto, se procederá a realizar la estimación bayesiana por medio del método de MCMC, específicamente el muestreador de Gibbs, para lograr realizar la simulación es importante encontrar la condicional posterior que a continuación se desarrollará para este tipo de datos en especial.

6.3.1. Definición y planteamiento

La función de verosimilitud para este modelo se basa en la siguiente ecuación de regresión:

$$Y_i = \alpha_{iT} + \tilde{X}\tilde{\beta} + \varepsilon_i \quad (1)$$

Donde α_i denota los interceptos de los i individuos y $\tilde{\beta}$ denota el vector de coeficientes de pendientes (que supone que es el mismo para todos los individuos) junto con las suposiciones del error, donde a continuación se evidencia la función de verosimilitud:

$$Y | \tilde{\beta}, h \sim N(\tilde{X}\tilde{\beta}, h^{-1}I_{TN}) \quad (2)$$

$$p(Y | \alpha, \tilde{\beta}, h) \propto \frac{1}{|h^{-1}I_{TN}|^{1/2}} \exp \left\{ -\frac{1}{2}(Y_i - \tilde{X}\tilde{\beta})^T (h^{-1}I_{TN})^{-1} (Y_i - \tilde{X}\tilde{\beta}) \right\}$$

$$p(Y | \alpha, \tilde{\beta}, h) \propto h^{\frac{NT}{2}} \exp \left\{ -\frac{h}{2}(Y_i - \tilde{X}\tilde{\beta})^T (Y_i - \tilde{X}\tilde{\beta}) \right\}$$

Donde

$$\tilde{X} = (1_T \otimes I_N)$$

$$\tilde{\beta} = [\alpha_1 \quad \alpha_2 \quad \dots \quad \alpha_n \quad \beta]^T$$

Para la estimación bayesiana, se hará uso de prior jerárquicas, en donde se utilizará una Normal-Gamma independiente, donde se supone que $\tilde{\beta}$ y \tilde{X} son prior diferentes entre sí

$$\tilde{\beta} \sim N(\underline{\tilde{\beta}}, \underline{V}) \quad (3)$$

$$h \sim G(\underline{s}^{-2}, \underline{v}) \quad (4)$$

Los modelos cada vez se centran en que los parámetros sean de alta dimensión, haciendo que la parte bayesiana con las priors jerárquicas solucione algunos problemas causados por espacios de parámetros de gran dimensión, en el modelo a utilizar contiene $N + k$ parámetros (N interceptos en α , $k - 1$ coeficientes de pendientes $\tilde{\beta}$ y la precisión del error h).

Esta prior jerarquica asume que para $i = 1, \dots, N$

$$\alpha_i \sim N(\mu_\alpha, V_\alpha) \quad (5)$$

Donde α_i y α_j siendo independiente para $i \neq j$, la estructura jerárquica de la ecuación anterior es μ_α y V_α como parámetros desconocidos, donde requiere su propia prior de la siguiente manera:

$$\mu_\alpha \sim N(\underline{\mu}_\alpha, \underline{\sigma}_\alpha^2)$$

$$V_\alpha^{-1} \sim G(V_\alpha^{-1}, \underline{v}_\alpha)$$

Las priors jerárquicas colocan más estructura en el sentido de que asume que todas los interceptos provienen de la misma distribución. Esta estructura extra (si es consistente con los patrones en los datos) permite una mayor precisión en la estimación.

Para la distribución posterior de $\tilde{\beta}$ y h condicionada en α se derivan de la misma manera que se realizó anteriormente

$$\tilde{\beta} | Y, h, \alpha, \mu_\alpha, V_\alpha \sim N(\bar{\beta}, \bar{V}_\beta) \quad (6)$$

Donde:

$$\bar{\beta} = \bar{V}_\beta (V_\beta^{-1} \underline{\beta} + h \sum_{i=1}^N \tilde{X}_i^t (Y_i - \alpha_{itT}))$$

$$\bar{V}_\beta = (V_\beta^{-1} + h \sum_{i=1}^N \tilde{X}_i^t \tilde{X}_i)$$

y

$$h | Y, \tilde{\beta}, \alpha, \mu_\alpha, V_\alpha \sim G(\overline{s^{-2}}, \bar{v}) \quad (7)$$

Donde:

$$\bar{v} = TN + v$$

$$\overline{s^{-2}} = \frac{\sum_{i=1}^N (Y_i - \alpha_{itT} - \tilde{X}_i) \tilde{\beta}^t (Y_i - \alpha_{itT} - \tilde{X}_i) \tilde{\beta}}{\bar{v}}$$

La posterior condicional para cada α_i es independiente de α_j para $i \neq j$, está dado por:

$$\alpha_i | Y, \tilde{\beta}, h, \mu_\alpha, V_\alpha \sim N(\bar{\alpha}_i, \bar{V}_i) \quad (8)$$

Donde:

$$\bar{\alpha}_i = \frac{V_\alpha (Y_i - \tilde{X}_i \tilde{\beta})_{tT}^t + h^{-1} \mu_\alpha}{(TV_\alpha + h^{-1})}$$

$$\bar{V}_i = \frac{V_\alpha h^{-1}}{(TV_\alpha + h^{-1})}$$

Finalmente la posterior condicional de los parámetros jerárquicos μ_α y V_α son:

$$\mu_\alpha | Y, \tilde{\beta}, h, \alpha, V_\alpha \sim N(\bar{\mu}_\alpha, \bar{\sigma}_\alpha^2) \quad (9)$$

Donde:

$$\bar{\mu}_\alpha = \frac{V_\alpha \underline{\mu}_\alpha + \underline{\sigma}_\alpha^2 + \sum_{i=1}^N \alpha_i}{V_\alpha + N \underline{\sigma}_\alpha^2}$$

$$\bar{\sigma}_\alpha^2 = \frac{V_\alpha \underline{\sigma}^2}{V_\alpha + N \underline{\sigma}_\alpha^2}$$

y

$$V_\alpha | Y, \tilde{\beta}, h, \alpha, \mu_\alpha \sim G(\bar{V}_\alpha^{-1}, \bar{v}_\alpha) \quad (10)$$

Donde:

$$\bar{v}_\alpha = \underline{v}_\alpha + N$$

$$\bar{V}_\alpha^{-1} = \frac{\sum_{i=1}^N (\alpha_i - \mu_\alpha)^2 + V_\alpha \underline{v}_\alpha}{\bar{v}_\alpha}$$

Ya teniendo las posteriors, se procederá a realizar la estimación bayesiana, por medio del muestreador de Gibbs.

6.3.2. Resultado estimación bayesiana

Como se mostró en la *subsección 6.3.1* se encontraron las posterior condiciones, ya conociendo la distribución que siguen, se procederá a realizar la simulación por el muestreador de Gibbs y se presentaran las pruebas correspondientes.

■ Diagnóstico de Heidelberg and Welch

- Cadenas y densidad de funciones posterior para Variables explicativas

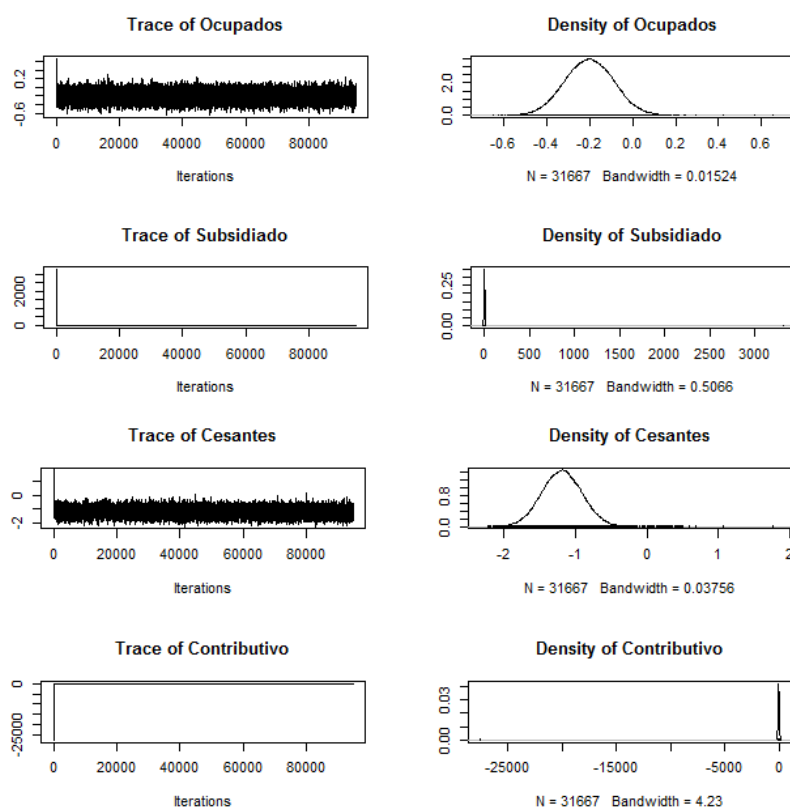


Figura 8: Traza de las cadenas de las variables del modelo ajustado

Tabla 11: Resultados Prueba de Heidelberg and Welch

<i>Variables</i>	<i>Stationary test</i>	<i>Start iteration</i>	<i>p-value</i>
Subsidiado	passed	3168	0.1063
Contributivo	passed	3168	0.1054
Ocupados	passed	1	0.1293
Cesantes	passed	3168	0.1269
<i>Variables</i>	<i>Halfwidth test</i>	<i>Mean</i>	<i>Halfwidth</i>
Subsidiado	passed	-0.650	0.04399
Contributivo	passed	-0.318	0.36653
Ocupados	passed	-0.203	0.00148
Cesantes	passed	-1.181	0.00322

En la *tabla 11*, se evidencia que el modelo pasa el supuesto de estacionariedad para las covariables, así como la prueba de longitud de la muestra

- Cadenas y densidad de funciones posterior para las áreas metropolitanas

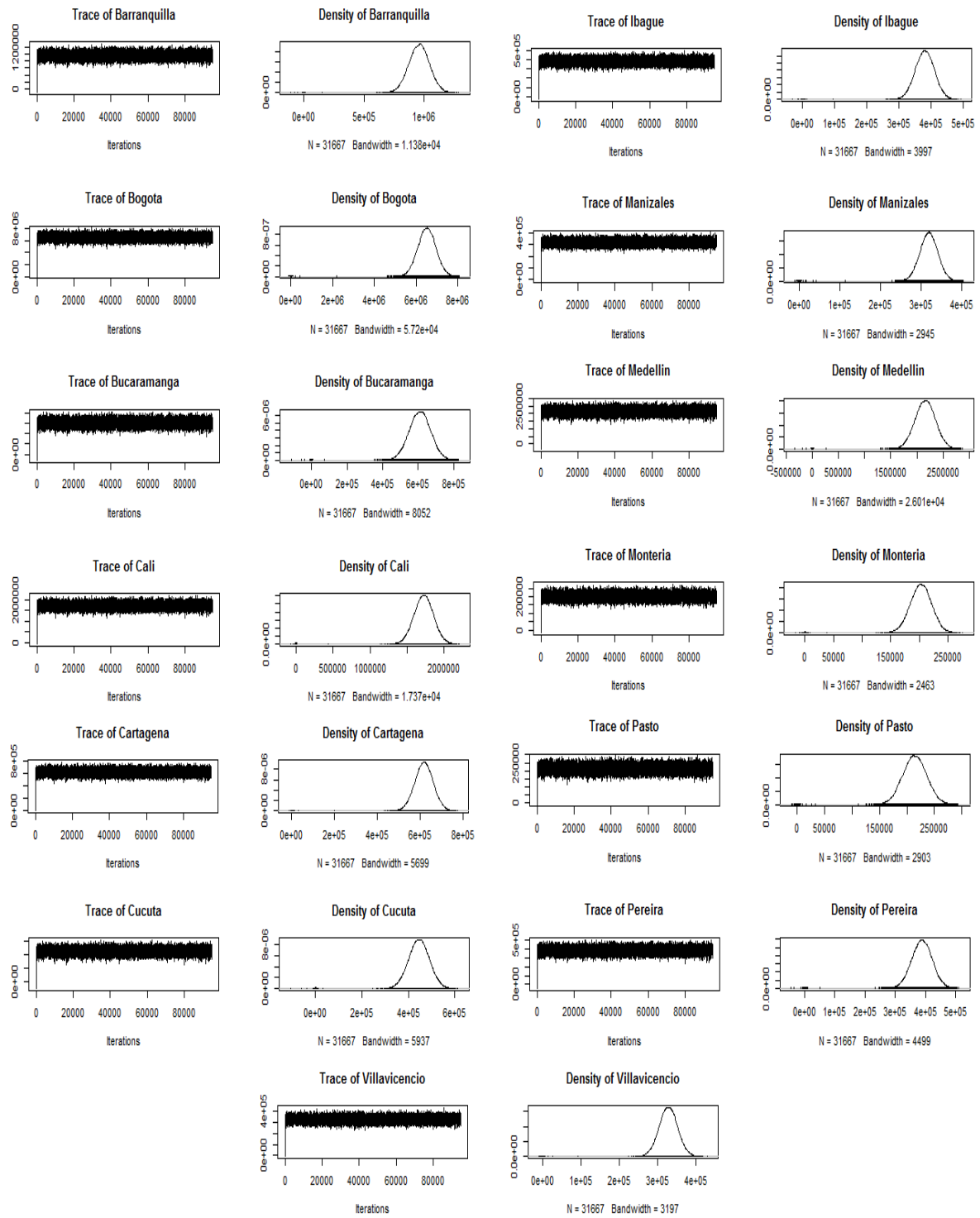


Tabla 12: Resultados

<i>Areas</i>	<i>Stationary test</i>	<i>Start iteration</i>	<i>p-value</i>
Barranquilla	passed	3168	0.1467
Bogota	passed	3168	0.1599
Bucaramanga	passed	3168	0.1255
Cali	passed	3168	0.1807
Cartagena	passed	3168	0.4588
Cucuta	passed	3168	0.1908
Ibague	passed	3168	0.0719
Manizales	passed	3168	0.2387
Medellin	passed	3168	0.1401
Monteria	passed	3168	0.3236
Pasto	passed	3168	0.7639
Pereira	passed	3168	0.1526
Villavicencio	passed	3168	0.2201

<i>Areas</i>	<i>Halfwidth test</i>	<i>Mean</i>	<i>Halfwidth</i>
Barranquilla	passed	957235.818	975.26976
Bogota	passed	6517802.413	4951.86485
Bucaramanga	passed	609907.812	698.06132
Cali	passed	1727109.334	1495.98053
Cartagena	passed	617254.583	492.65844
Cucuta	passed	445784.845	507.98899
Ibague	passed	381124.434	343.44737
Manizales	passed	319696.988	255.11326
Medellin	passed	2156636.278	2228.46730
Monteria	passed	202739.250	214.92469
Pasto	passed	213977.833	248.72678
Pereira	passed	388117.226	384.55548
Villavicencio	passed	327639.482	275.43759

Con lo presentado anteriormente en el modelo las variables pasan el supuesto de estacionariedad, así como la prueba de longitud de la muestra, donde se tiene que es suficientemente larga para estimar la media con buena precisión.

■ Resultado del modelo

A continuación se presentan las estimaciones bayesianas de los parámetros del modelo.

Tabla 13: Resultados del modelo

<i>Parameters</i>	<i>Mean</i>	<i>2.5 %</i>	<i>97.5 %</i>	<i>n.effective</i>
Subsidiados	-0.843	-61.924	61.643	63000
Contributivo	-0.531	-8.039	6.764	63000
Cesantes	-1.154	-1.719	-0.462	190
Ocupados	-0.200	-0.426	0.031	1200

Dic 5302.5 pD = 671.1

Dados los resultados del modelo se observa que la variable Cesantes se encuentra dentro del intervalo de credibilidad del 95 %, siendo la única variable significativa y por lo cual se concluye que al aumentar el número de personas cesantes en las áreas metropolitanas, se espera que disminuya en una persona los evasores.

6.4. Comparación

En esta sección se hará la respectivas comparaciones para lograr concluir cual estimación es mas precisa.

6.4.1. Estimaciones

Tabla 14: Estimaciones Frecuentistas vs Bayesianas

Áreas metropolitanas	Estimate	Std. Error	Pr(> t)
Barranquilla	858458	79514	< 2.2e-16 ***
Bogotá	5991107	406824	< 2.2e-16 ***
Bucaramanga	542818	56072	< 2.2e-16 ***
Cali	1579024	121462	< 2.2e-16 ***
Cartagena	565870	40325	< 2.2e-16 ***
Cúcuta	399336	40577	< 2.2e-16 ***
Ibagué	349322	27393	< 2.2e-16 ***
Manizales	294988	20449	< 2.2e-16 ***
Medellín	1939008	180589	< 2.2e-16 ***
Montería	184190	16933	< 2.2e-16 ***
Pasto	190454	19915	< 2.2e-16 ***
Pereira	351697	30950	< 2.2e-16 ***
Villavicencio	299955	22280	< 2.2e-16 ***

Áreas metropolitanas	Estimate	Std. Error	2.5 % - 97.5 %
Barranquilla	938406.497	158481	735384 - 1121681
Bogota	6388197.372	1008990	5391137 - 7351459
Bucaramanga	597977.124	104421	453198 - 726804
Cali	1692944.410	274547	1388405 - 1978358
Cartagena	605039.259	96435	506109 - 700591
Cucuta	437031.682	76328	331684 - 531006
Ibague	373644.899	61026	303402 - 438810
Manizales	313318.844	49842	262441 - 362384
Medellin	2114177.789	358602	1650151 - 2532224
Montería	198716.973	33803	154657 - 238348
Pasto	209792.675	36856	157682 - 256025
Pereira	380376.620	63785	301380 - 452654
Villavicencio	321137.548	51704	265351 - 374244

En la *Tabla 14* se evidencia que las estimaciones tanto frecuentista como bayesiana tiene valores muy similares, coincidiendo con las tres áreas metropolitanas más evasoras (Bogotá, Medellín y Cali).

6.4.2. Intervalos

A continuación se presentará, los intervalos de confianza para el caso de la estimación frecuentista y los intervalos de credibilidad para la estimación bayesiana.

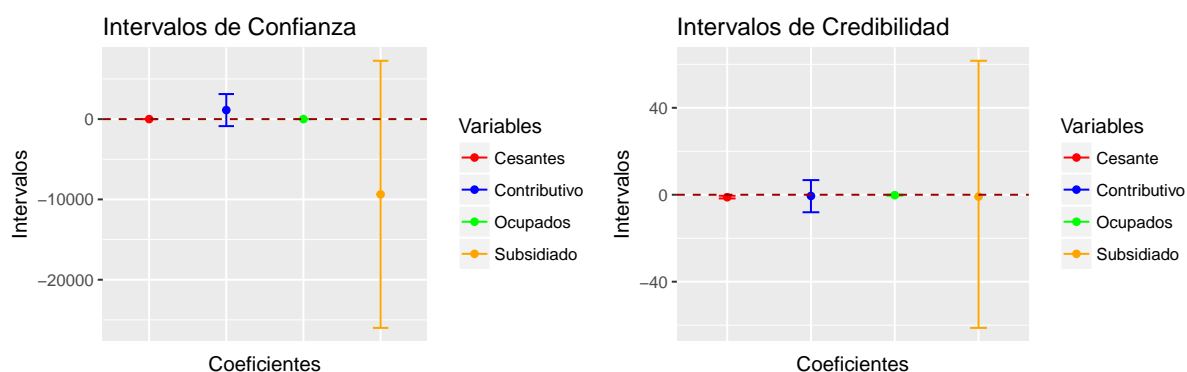


Figura 9: Intervalos estimación frecuentista vs bayesiana

Se evidencia que la anchura de los intervalos de la estimación frecuentista es mayor comparada con la bayesiana, con lo que nos da un indicio de que la estimación bayesiana tiene una mayor precisión.

Es de interés saber si realmente el intervalo de la variable cesante contiene el cero, lo ideal sería que no, ya que se ha evidenciado que esta variable es la única que es significativa en las dos estimaciones presentadas.

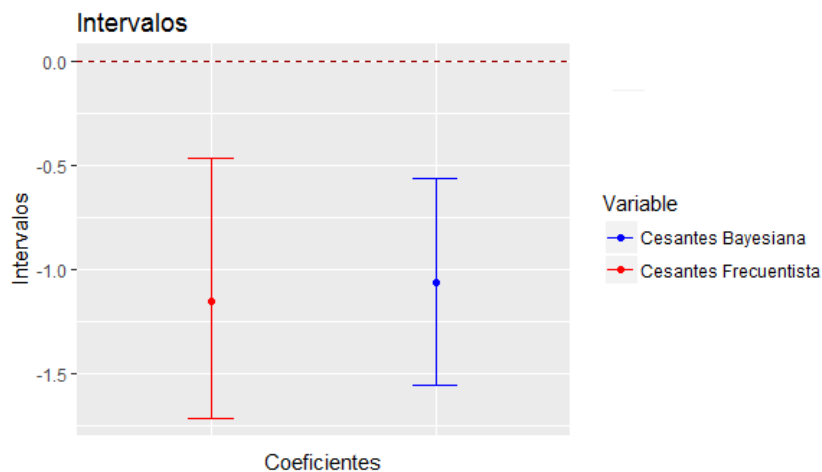
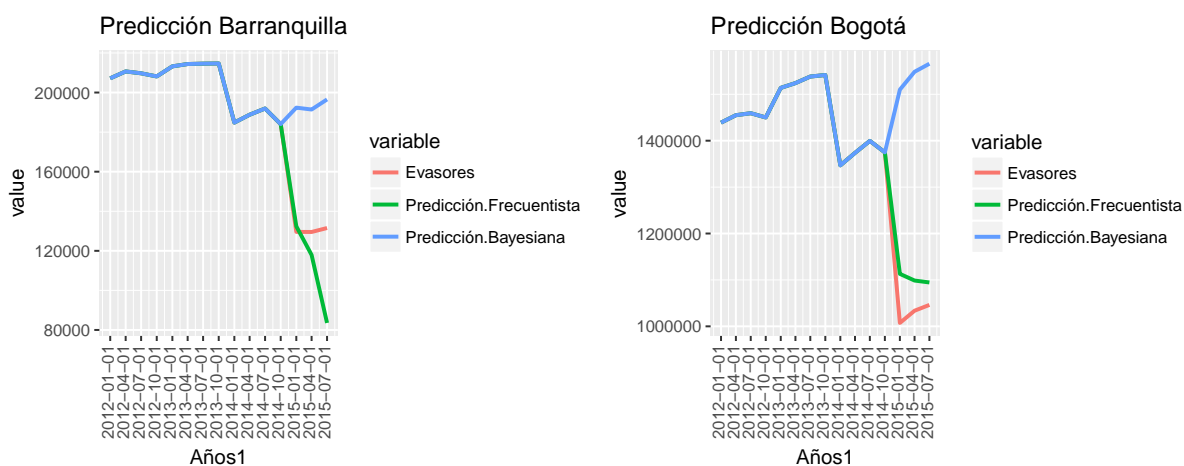


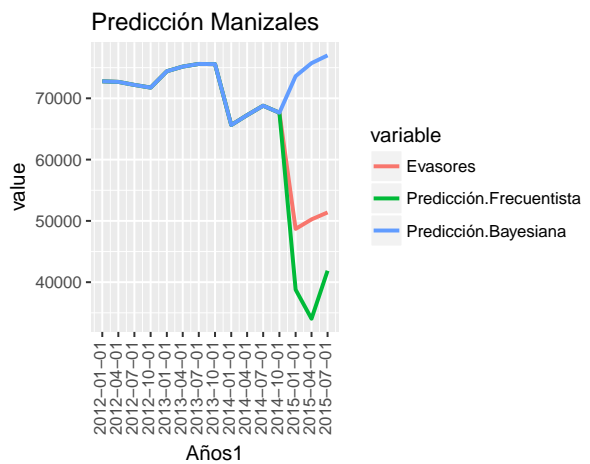
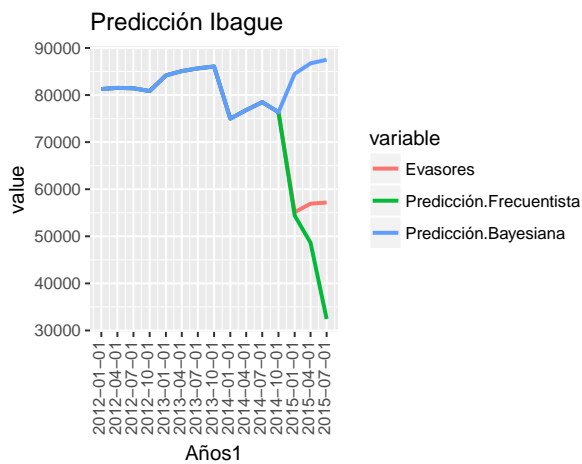
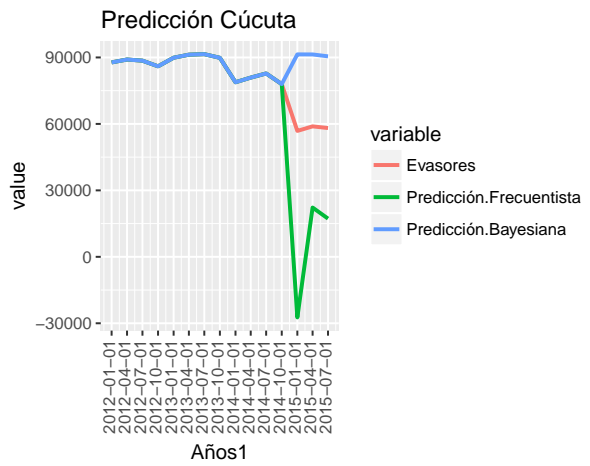
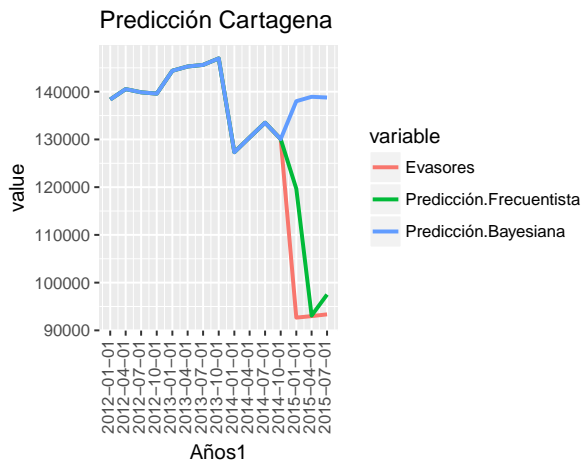
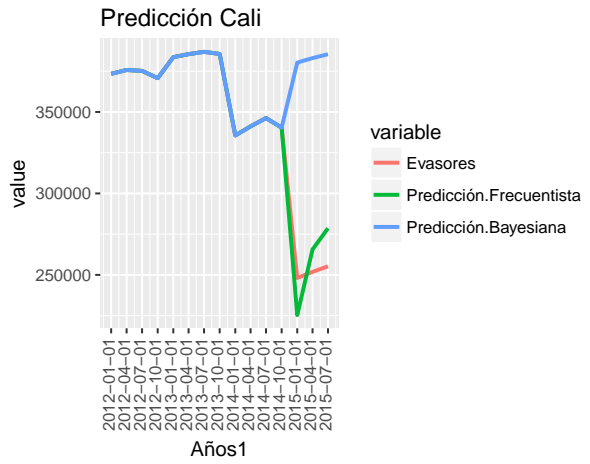
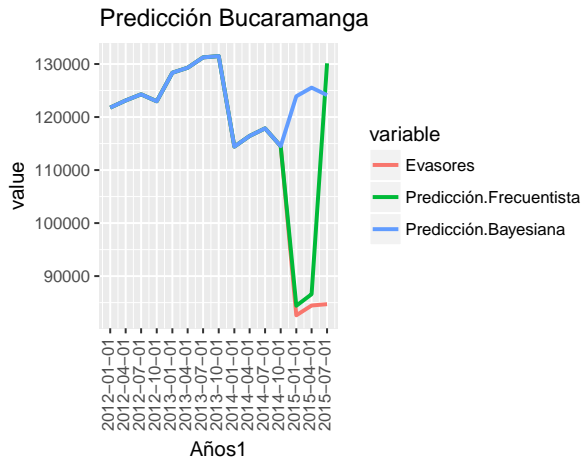
Figura 10: Intervalo de la variable Cesante

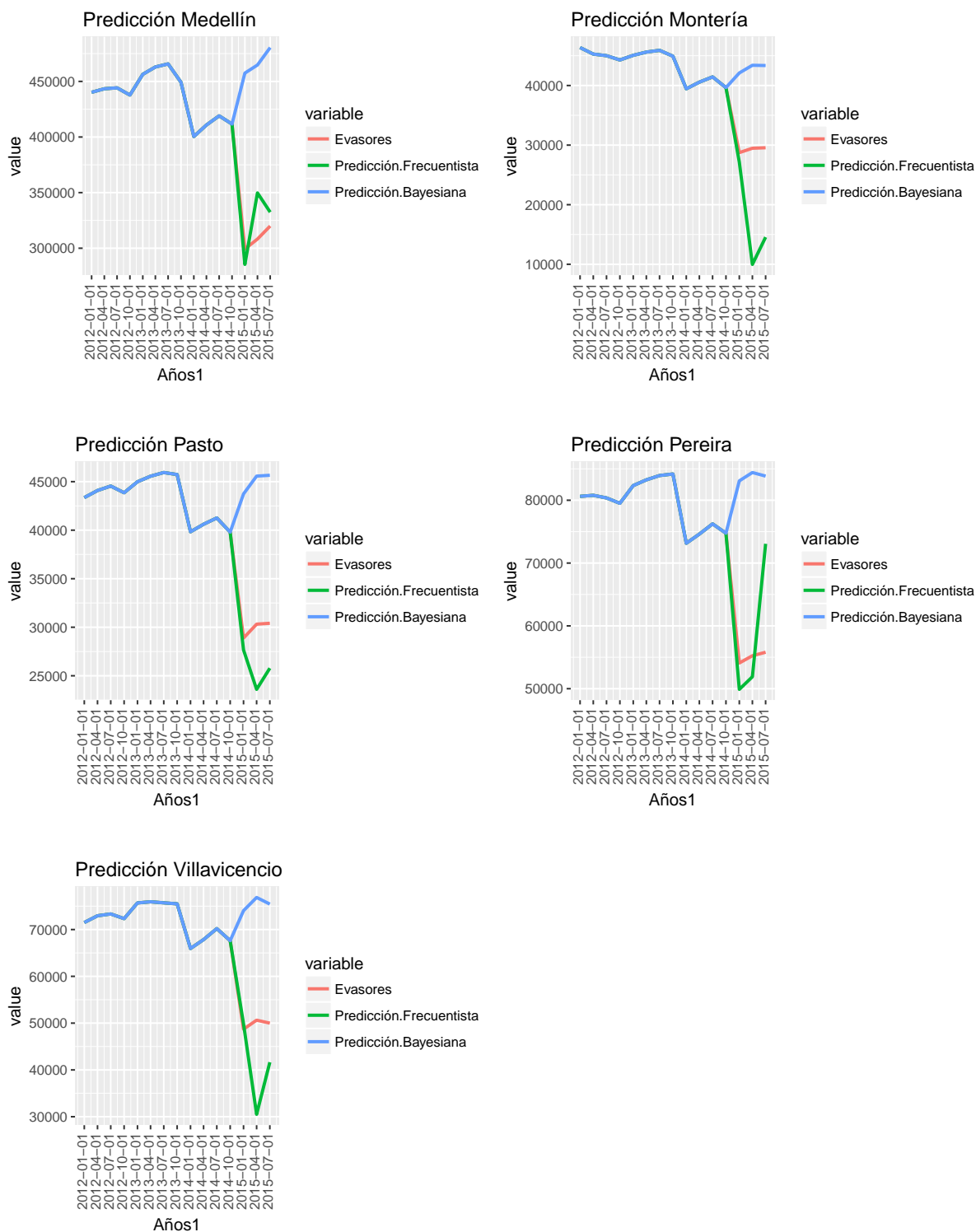
Al observar la gráfica del intervalo, se afirma que la variable explicativa cesante es significativa y cabe resaltar que el intervalo de credibilidad es más angosto que el de confianza.

6.4.3. Predicción

Para la saber cuál modelo tiene mejor estimación, con los datos que se tiene, se estimará la evasión en las áreas metropolitanas para el año 2015 y con este se verificará con cuál de las estimaciones es más precisa.







Con las predicciones presentadas anteriormente se evidencia que la estimación que más se acerca al verdadero valor de la evasión para los trimestres 2015-1 al 2015-3 es la estimación frecuentista, ya que por decirlo de alguna manera en la mayoría de las áreas metropolitanas adoptan un comportamiento similar en los trimestres mencionados, por el contrario la predicción bayesiana esta sobreestimando el valor de la evasión.

7. Conclusiones

Para finalizar se puede decir lo siguiente:

1. Analizando los efectos estimados tanto de la forma frecuentista como la bayesiana, se encontró que las áreas metropolitanas que tienen efectos positivos para la evasión son Bogotá, Medellín y Cali y las áreas metropolitanas que tienen un efecto negativo para la evasión son Montería, Pasto y Villavicencio; los trimestres en donde existe efecto positivo para la evasión es 2014-3, 2013-1 y 2013-4, los trimestres en donde existen efectos negativos para la evasión es 2015-1, 2015-2 y 2015-3.
2. Los intervalos presentados, evidenciamos que la estimación bayesiana fue más precisa, ya que los intervalos de credibilidad fueron más angostos, haciendo que la precisión fuese mayor.
3. Las predicciones realizadas a los trimestres 2015-1, 2015-2 y 2015-3 (valores ya conocidos), se evidencia que la estimación que más se acerca al verdadero valor de la evasión de estos trimestres, es la frecuentista, ya que como se observó en las gráficas, el comportamiento de esta predicción es muy similar a los periodos comparados.
4. En cuanto a las estimaciones presentadas, tanto frecuentista como bayesiana, en estas se llegó a la misma conclusión, en donde Cesantes fue la única variable significativa, concluyendo que si estas personas aumentan en mil, se espera que los evasores disminuyan aproximadamente en la misma cantidad.
5. Como conclusión general, se evidencia que cualquiera de las estimaciones brindan una buena precisión, si se quiere una estimación que genere intervalos más precisos, es conveniente elegir el camino bayesiano, ya que como se evidenció, este logra intervalos más angostos, pero si se quiere una predicción más acertada a los verdaderos valores, lo ideal sería escoger la estimación frecuentista, en este trabajo se logró encontrar que las dos estimaciones son buenas y que se llegaron a la misma conclusión en cuanto a las variables significativas, a los efectos tanto en las áreas metropolitanas como en los trimestres, finalmente dependiendo de interés de proyecto o trabajo, se debe escoger el tipo de estimación.

8. Agradecimientos

Quiero agradecer primeramente a mis directores de tesis Yesid Rodríguez y Wilmer Pineda por brindarme cada uno su apoyo, conocimiento y dedicación, le agradezco a mis papas y hermanita por apoyarme en este grato camino de estudiar estadística, en especial agradezco a mi novio Diego Armando Corredor Rivera, quien siempre me apoyó en todos los momentos del transcurso de la carrera, demostrándome su disciplina y perseverancia y finalmente gracias a Dios por permitirnos a los dos culminar uno de los muchos logros que vamos a cumplir juntos.

9. Referencias

1. Ausín, C. (2012), Métodos de cadenas de Markov Monte Carlo, Universidad Carlos III de Madrid.
2. Albert, J. (2009), Bayesian computation with R, Springer Science Business Media.
3. Baronio, A. M., Vianco, A. M. Cuadernos de Econometría.
4. Chib, S. (2008). Panel data modeling and inference: a Bayesian primer. *Advanced studies in theoretical and applied econometrics*, 46, 479.
5. Croissant, Y., Millo, G. (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software*, 27(2), 1-43.
6. García, Repositorio Universidad Santo Tomás. (2017). Modelo espacial con acercamiento bayesiano para el estudio de víctimas en Colombia en 2015.
7. Gujarati Damodar, N., Porter Dawn, C. (1997). *Econometría*. 3ª. Edición. México. Ed. Mc Graw Hill.
8. Hsiao, C. (2014), *Analysis of panel data*, number 54, Cambridge university press.
9. Huo, R. (2015). *Panel Data Models with Interactive Fixed Effects: A Bayesian Approach*. Columbia University.
10. Kennedy, P. (2003), *A guide to econometrics*, MIT press.
11. Kessler, L. M., Munkin, M. K. (2015). Bayesian estimation of panel data fractional response models with endogeneity: an application to standardized test rates. *Empirical Economics*, 49(1), 81-114.
12. Kleiber, C., Zeileis, A. (2008). *Applied econometrics with R*. Springer Science Business Media.
13. Koop, G. (2003), *Bayesian Econometrics*, Wiley.
14. Montero. R (2005): Test de Hausman. Documentos de Trabajo en Economía Aplicada. Universidad de Granada. España
15. Morawetz, U. (2006). Bayesian modelling of panel data with individual effects applied to simulated data. Univ. für Bodenkultur, Department für Wirtschafts- u. Sozialwiss., Inst. für Nachhaltige Wirtschaftsentwicklung.
16. Pesaran, M. H. (2015). *Time series and panel data econometrics*. Oxford University Press.
17. Ruiz, A. L. M. (2016), Estimar la evasión del sistema de protección social y caracterizar los efectos principales en Colombia.
18. Torrecillas, R. L. (2014). Guía CERO para datos de panel. Un enfoque. UAM-Accenture Chair on the Economics and Management of Innovation.
19. Trinh, K. D. Bayesian estimation and specification test for a time-varying stochastic frontier model.
20. Unidad de Gestión Pensional y Parafiscal(n.d.), <http://www.ugpp.gov.co/parafiscales/guias-del-sistema-de-la-proteccion-social.html>.
21. Wooldridge, J. M. (2006), *Introducción a la econometría: un enfoque moderno*, Editorial Paraninfo
22. Zheng, Y., Zhu, J., Li, D. (2008). Analyzing spatial panel data of cigarette demand: A Bayesian hierarchical modeling approach. *Journal of Data Science*, 6(4), 467-489.