

Proyecto Final Regresión Avanzada

Carlos Pérez, Juan Zinser

Diciembre 2016

Contents

| | |
|--|-----------|
| Introducción: Calidad del Aire en la Ciudad de México | 2 |
| Descripción de la Información | 2 |
| Análisis Exploratorio | 3 |
| Modelado e Implementación | 6 |
| Regresión Múltiple Dinámico Gamma | 7 |
| Resultados | 8 |
| Referencias | 13 |
| Apéndice | 14 |

Introducción: Calidad del Aire en la Ciudad de México

Dado que la calidad del aire es un tema primordial en la salud de los habitantes y cada vez más un tema central en la política ambiental en las megalópolis del planeta, es relevante analizar el efecto que tiene la acumulación vehicular sobre los componentes atmosféricos que ponen en riesgo la salud de los habitantes.

En el presente año existieron varios periodos que se declararon como contingencia ambiental en la Ciudad de México debido al exceso en la concentración de ciertos contaminantes en la atmósfera, de modo que es cada vez más necesario realizar monitoreos y estudios del impacto que tiene la cantidad de unidades vehiculares que circulan en las ciudades con respecto a los problemas ambientales que originan. En particular aunado al crecimiento urbano, el Valle de México no presenta una orografía favorable para la contaminación ambiental, ya que esta es propensa a que se produzca el efecto invernadero dentro de la cuenca del valle.

Por lo tanto, en el presente análisis se desea estimar la relación entre el parque vehicular, esto es el número registrado de autos, motocicletas, autobuses de pasajeros y camiones de carga en circulación, y la concentración de partículas suspendidas de menos de 2.5 micrómetros en los municipios de la Ciudad de México para el periodo 2004-2015.

Descripción de la Información

Las variables explicativas o regresores son parte del conjunto de datos *Vehículos de motor registrados en circulación* que forma parte de los registros administrativos en las fuentes estadísticas de INEGI. El nivel de desagregación más granular de los datos es anual y a nivel municipio.

La información es pública y puede ser descargada en el siguiente enlace [<http://www.inegi.org.mx/est/contenidos/proyectos/registros/economicas/vehiculos/default.aspx>], esta información surge a partir de los padres vehiculares de cada municipio y es consolidada por INEGI a través de los gobiernos estatales.

Este conjunto comprende categorías oficial, particular y público para cada tipo de vehículo, no obstante fueron agregadas para poseer datos con pocos valores nulos. Las variables resultantes **auto**, **cyc**, **moto**, **cam_pas** representan el total de autos, motocicletas, camiones de carga y camiones de pasajeros respectivamente.

Las variables representan el conteo total de vehículos, por lo que poseen una escala de medición *racional*. A continuación una breve descripción de cada una:

- **auto** : Comprende los vehículos a motor destinados principalmente al transporte de personas, que tengan hasta 7 asientos (incluyendo el del chofer).
- **cam_pas** : Comprende los autobuses urbanos y suburbanos, microbuses, camiones escolares, camionetas pick-up (utilizadas para el transporte de trabajadores), ómnibus y en general, los vehículos con más de 7 asientos, destinados al transporte público o privado de personas.
- **cyc** : Son los que están destinados exclusivamente al transporte de carga. También se incluyen los vehículos de tracción diseñados para el remolque como tráilers, auto-tanque, panel, cabinas, pipas, redilas, vanette, volteo y grúas. Es el transporte de pasajeros y/o carga que es utilizado por organismos gubernamentales para satisfacer sus propios requerimientos y/o atender necesidades de la población.
- **moto** : Vehículo automotor de dos o tres ruedas, cuyo peso no excede los cuatrocientos kilogramos, incluye triciclos, motonetas y motocicletas. Es el transporte de pasajeros y/o carga que es utilizado por organismos gubernamentales para satisfacer sus propios requerimientos y/o atender necesidades de la población.

La variable dependiente es la concentración de partículas suspendidas de menos de 2.5 micrómetros medida a través de la red manual de monitoreo atmosférico, la cual puede ser descargada del sitio de la calidad del aire de la Ciudad de México [http://148.243.232.112:8080/opendata/red_manual/red_manual_particulas_susp.csv]

Las partículas suspendidas **pm2.5** son medidas a partir de dos métodos principalmente: gravimetría, que conta en determinación de la masa de partículas presente en un flujo de aire, las partículas son separadas de

la corriente y depositadas sobre un filtro colocado en un elemento oscilante cuya variación en la frecuencia de oscilación es proporcional a la masa; por otro lado, radiación beta, que consiste en la atenuación en la intensidad de la radiación beta por las partículas depositadas sobre un filtro continuo, en la que se calcula la masa a partir de la emisión de electrones.

Los datos poseen una desagregación a nivel estación de monitoreo, es decir con coordenadas latitud y longitud, las cuales están georreferenciadas a los municipios correspondientes. La dimensionalidad temporal que tienen es a nivel diario y dichas las cifras corresponden al promedio diario de partículas suspendidas por lo que poseen una escala de medición *racional*. Estas cantidades fueron agregadas para obtener un cálculo aproximado del total de partículas suspendidas al año y poder cruzar la información con el total del parque vehicular.

Los datos se ven de la siguiente forma, se presenta un vistazo general de los mismos

Table 1: Vistazo del conjunto de datos

| year | pm2.5 | cam | cyc | auto | moto | id_cdmx |
|------------|-------|-------|-------|--------|-------|---------|
| 2004-01-01 | 4679 | 4711 | 16225 | 817516 | 23914 | 1 |
| 2004-01-01 | 3165 | 1308 | 40056 | 192442 | 933 | 0 |
| 2005-01-01 | 4922 | 9655 | 17504 | 726454 | 20209 | 1 |
| 2005-01-01 | 2941 | 4258 | 44719 | 230913 | 1743 | 0 |
| 2006-01-01 | 4167 | 9793 | 19388 | 847237 | 26790 | 1 |
| 2006-01-01 | 2641 | 19607 | 62837 | 331104 | 3703 | 0 |

Algunos hechos importantes sobre las fuentes que originan las partículas suspendidas Google :

- Ciertos tipo de camiones y los tractocamiones son los vehículos de mayor emisión de partículas finas debido al uso de diésel como combustible. Generan el 8% de las PM10 y el 11% de las PM2.5. espacio
- El tránsito vehicular sobre las vialidades genera emisiones de partículas, producidas por el paso de los vehículos sobre ellas y la resuspensión del polvo. Contribuyen con el 30% de las PM10 y el 13% de las PM2.5.
- Las actividades de la construcción (demolición, edificación y construcción de vialidades) aporta el 16% de las PM10 y el 25% de PM2.5 espacio
- La industria química y la fabricación de productos a base de minerales no metálicos (ej. cemento, concretos, cal) son los mayores emisores de partículas del sector industrial aportando el 7% de PM10 y 10% de PM2.5

No obstante el objetivo principal era realizar un análisis completo de la ciudad, la red de monitoreo presenta dos grandes problemas: no todas las estaciones activas recopilan información de todos los contaminantes de modo que hay una cobertura geográfica limitada; por otro lado, de estas estaciones que sí realizan mediciones sobre el tipo de contaminante en cuestión no todas están operando durante el periodo completo de modo que también existe una insuficiencia en términos de la cobertura temporal.

Dado lo anterior la muestra se redujo considerablemente a 5 municipios, 3 en CDMX y 2 en EDOMX, en los cuales se encuentra información detallada de forma completa.

La variable *id_entmun* representa la clave geoestadística de los municipios establecida por INEGI. Por otro lado se tiene la variable *year* que representa el año en que se efectuaron las mediciones.

Análisis Exploratorio

En seguida se muestra la Tabla 1 con un resumen estadístico, en el que se pueden apreciar algunos estadísticos que permiten conocer las distribuciones como son la media, la mediana y máximos y mínimos, el total de observaciones en el conjunto de datos es de 20 en total.

Table 2: Resumen Estadístico

| year | pm2.5 | cam | cyc | auto | moto |
|--------------|--------------|---------------|----------------|-----------------|---------------|
| Min. :2004 | Min. :2198 | Min. : 1308 | Min. : 16225 | Min. : 192442 | Min. : 933 |
| 1st Qu.:2006 | 1st Qu.:2621 | 1st Qu.: 3969 | 1st Qu.: 21070 | 1st Qu.: 449503 | 1st Qu.:12294 |
| Median :2008 | Median :3466 | Median : 9196 | Median : 33498 | Median : 741557 | Median :22948 |
| Mean :2008 | Mean :3449 | Mean : 8760 | Mean : 55610 | Mean : 735573 | Mean :22708 |
| 3rd Qu.:2011 | 3rd Qu.:4180 | 3rd Qu.:11186 | 3rd Qu.: 80701 | 3rd Qu.: 988387 | 3rd Qu.:31751 |
| Max. :2013 | Max. :4922 | Max. :19607 | Max. :157070 | Max. :1309829 | Max. :47896 |

Como se puede observar en la tabla del resumen estadístico se tienen fechas desde 2004 hasta 2013 lo que representan 10 años de observaciones para el Estado de México y 10 observaciones más para la Ciudad de México. Las variables **cam**, **cyc**, **auto** y **moto** tienen medias de distintas magnitudes. El promedio histórico para camiones de pasajeros, autobuses, microbuses, etc. se encuentra en 8,760, seguido del promedio histórico del transporte de carga con 33,498 unidades. Luego está la cantidad promedio de motocicletas a los largo de los últimos 10 años de 22,708 a lo que se le suma en parque vehicular la categoría más grande que es la de auto con una media de 735,573 autos.

La concentración de partículas varía relativamente poco comparado con las variables independientes ya que tiene mínimos y máximos anuales de 2198 y 4922 respectivamente. Es importante tener en cuenta que esta es la suma anual del promedio diario del total de partículas existentes en la composición del aire, por lo que resulta una buena variable dependiente que refleja el comportamiento de la contaminación atmosférica al ser sólo un múltiplo del total de partículas suspendidas de tamaño menor a 2.5 micrómetros.

La figura 2 presenta las relaciones existentes entre las variables consideradas en el conjunto de datos. Es importante apreciar que existe una relación positiva, como era de esperarse, entre el conteo total de vehiculos en sus distintas categorías y el total de partículas suspendidas.

En particular existe en este gráfico la distribuciones a lo largo del tiempo si se considera la columna de gráficos bajo la etiqueta *year*, en las que se puede observar un comportamiento creciente. En el gráfico se cuenta con una distinción entre estados de la Ciudad de México y el Estado de México por el color de los puntos. El negro corresponde a la Ciudad de México y el amarillo corresponde al Estado de México. Es necesario hacer un recordatorio ya que, si bien los datos son parte de la red de monitoreo manual y es muy posible que no se cuente con toda la información sobre contaminación atmosférica, es posible hacer un análisis cuantitativo con base en los municipios que pertenecen a esta para determinar cómo es la relación entre el padrón vehicular y la concentración de partículas suspendidas.

A continuación se presenta una gráfica exploratoria de la dependiente elegida contra las variables junto con su correlación lineal, en particular en los diagramas de dispersión se observan relaciones positivas de 0.71, 0.49, 0.12 entre la variable del conteo de autos, motos y camiones de pasajeros contra los niveles de concentración de partículas pm2.5 respectivamente. Por otro lado se tiene una correlación negativa de -.81 para la variable de camiones de carga cuando se analiza contra la dependiente.

En el título de las gráficas se exhibe el valor de la correlación lineal que existe entre ambas. En particular es importante mencionar que existe una acumulación natural en la zona de conteos vehiculares bajos y altas cantidades de concentración de partículas suspendidas de menos de 2.5 micrometros, por lo que se usa una transformacion logarítmica sobre las variables de modo que se dispersen un poco, disminuyan en esacala y además de aprovechar dicha transformacion de soporte para poder probar modelos normales sobre las variables.

Cada una de las gráficas anteriores muestra los diagramas de dispersión de las variables originales, la correlación lineal de menor magnitud es la de los camiones de pasajeros; no obstante es positiva, las observaciones de pm2.5 podrían estar sesgadas ya que se está utilizando solo la red manual de monitoreo atmosférico. Sin embargo presentan relaciones muy interesantes como el caso de los camiones de carga. Como se puede apreciar por los colores de la gráfica existe una distincion importante entre los dos estados, los amarillos(claros) representan el Estado de México y los negros representan la Ciudad de México. En particular se tiene que

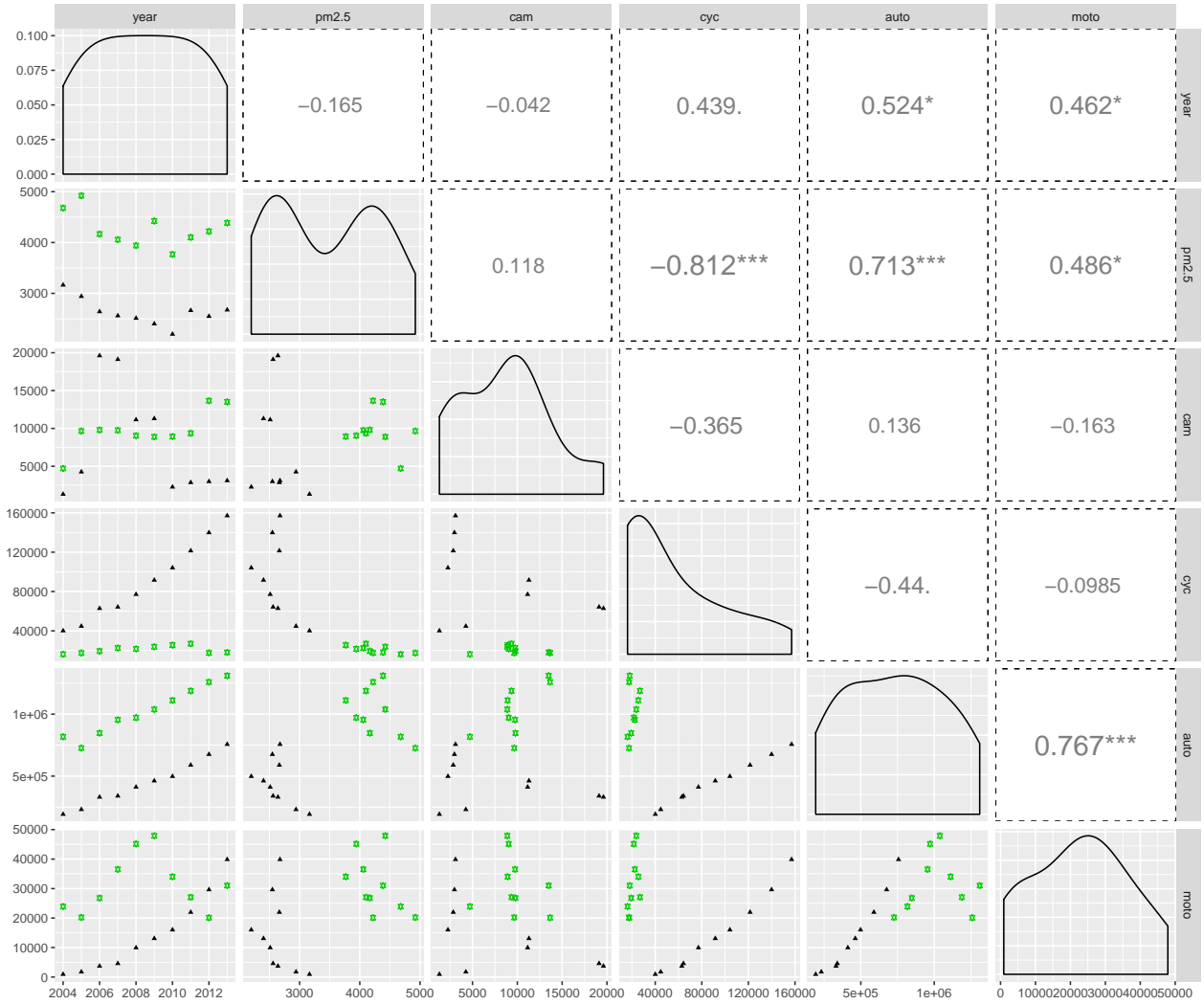


Figure 1: Matriz de Diagramas de Dispersión

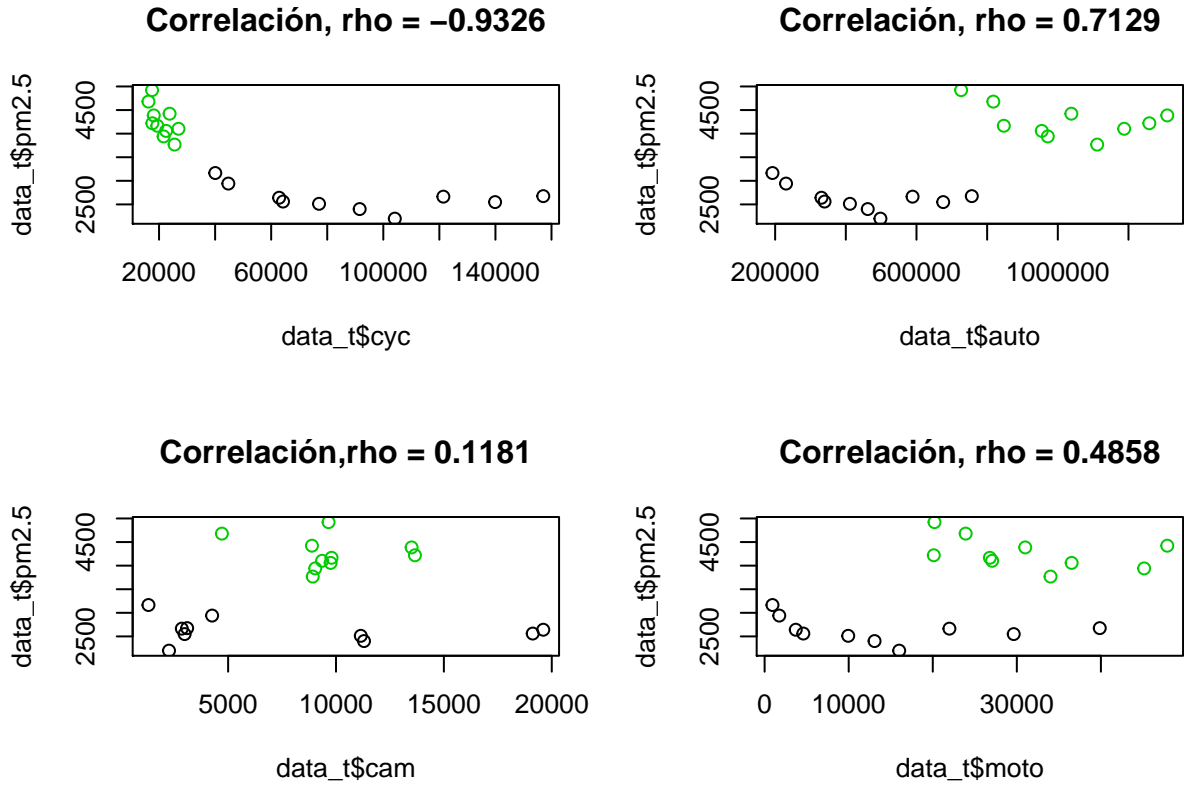


Figure 2: Correlaciones Específicas

hay claramente un comportamiento distinto en cada uno de los diagramas de dispersión como se muestra por los colores o tonos en las imágenes.

Modelado e Implementación

En este análisis se realizaron distintos experimentos que permitieron cada vez un mejor ajuste de los datos. Se consideraron cada vez mejores especificaciones que se ven reflejadas en las métricas de desempeño correspondientes. En un principio se ajustó un modelo simple de regresión normal usando como variable explicativa el total de autos. Fue un primer ejercicio que no resultó exitoso ya que no es posible que una recta ajuste bien este tipo de datos como se muestra gráficamente. Luego se intentó un modelo de regresión múltiple incluyendo los términos de interacción, lo cual conllevó a una mejora aunque no muy grande: de un DIC de 81 a un DIC de 74. Por último, una vez que ya se intentó el modelo con interacciones quedaba explotar la dimensión más importante, por lo que se realizó un ajuste a través de un modelo dinámico temporal, en el que se estableció inicialmente dinámica para todos los parámetros del modelo, de modo que se llegó a un modelo más general que logró capturar muy bien el comportamiento de las partículas suspendidas de menos de 2.5 micrómetros a partir del padrón vehicular de motos, autos, autobuses y camiones de carga registrados en los gobiernos municipales. Se generaron varios modelos con parametrizaciones normales que llevaron a buenos resultados DIC de 30 y 20, no obstante seguían muy lejos de ser buenos modelos para predecir la concentración de partículas. Al final se probó con otro modelo generalizado gamma, que llegó a vencer y sobrepasar muy bien a todos los anteriores.

Regresión Múltiple Dinámico Gamma

La forma funcional que se eligió fue la siguiente

$$\log(y_t) = \alpha + \sum_{i \in \{auto, cyc, cam, moto\}} \beta_{it} \log(x_{it}) + \alpha^{cdmx} x_{cdmx} + \sum_{i \in \{auto, cyc, cam, moto\}} \beta_{it}^{cdmx} \log(x_{it}) * x_{cdmx}$$

En términos de distribuciones se tiene que el mejor modelo resultó ser un modelo gamma, el cual tuvo que ser alimentado con las variables transformadas debido a problemas numéricos de escala. Si se utilizaban las variables originales se llegaba a crecer demasiado y BUGS marcaba error. Al ser variables de una gran magnitud, con la escala logarítmica no se pelagra de caer en valores negativos inadmisibles por la distribución gamma.

Se representa la cantidad de partículas de menos de 2.5 micrómetros en escala logarítmica como una gamma $\log(y_t) \sim \Gamma(\tau * \mu_t, \tau)$. La media de la distribución está ligada al predictor lineal a través de la función exponencial. La combinación lineal consta de la cantidad de vehículos (automóviles, transporte de carga, motocicletas y camiones urbanos) y su respectiva interacción geográfica, de la siguiente forma: $\eta_t = \alpha + \alpha^{cdmx} + \sum_{i \in I} \beta_{i,t} \log(x_{i,t}) + \sum_{i \in I} \beta_{i,t}^{cdmx} \log(x_{i,t}) * x_{cdmx}$. Donde $I = auto, cyc, moto, cam$. Las variables x_{cdmx} son variables indicadoras sobre si la observación corresponde al Estado de México o a la Ciudad de México. De igual forma las β_i^{cdmx} ayudan a separar el efecto de los vehículos en cada uno de los estados. También se utilizan α^{cdmx} para separar la ordenada al origen del modelo según el estado. No se incluye el efecto de los diferentes vehículos para distinguir diferentes ordenadas al origen. Las ordenadas al origen se interpretan de modo que cuando la observación pertenece a la Ciudad de México en un periodo t , la ordenada será $\alpha + \alpha^{cdmx}$, y cuando pertenezca al Estado de México será α .

Al ver en el modelo que variables están indizadas por el tiempo, se puede ver que las $\beta_{i,t}^{cdmx}$, $\beta_{i,t}$ tienen parámetros propios a su tiempo, mientras que α^{cdmx} y α son los únicos parámetros sin dinámica temporal. La especificación del modelo termina ligando las variables temporales entre periodos.

A continuación se presenta el resto de la especificación del modelo

Distribuciones de Observables

$$\begin{aligned} \log(y_t) &\sim \Gamma(\tau * \mu_t, \tau) \\ \mu_t &= \exp(\eta_t) \\ \eta_t &= \alpha + \sum_{i \in \{auto, cyc, cam, moto\}} \beta_{it} \log(x_{it}) + \alpha^{cdmx} x_{cdmx} \\ &\quad + \sum_{i \in \{auto, cyc, cam, moto\}} \beta_{it}^{cdmx} \log(x_{it}) * x_{cdmx} \end{aligned}$$

Distribuciones de Estado

$$\begin{aligned} \beta_{i,t}^{cdmx} &\sim N(\beta_{i,t-1}^{cdmx}, \tau_i^{cdmx}); \quad i \in \{auto, cyc, cam, moto\}; \quad 2 \leq t \leq 10 \\ \beta_{i,t} &\sim N(\beta_{i,t-1}, \tau_i); \quad i \in \{auto, cyc, cam, moto\}; \quad 2 \leq t \leq 10 \end{aligned}$$

Distribuciones Iniciales

$$\begin{aligned}
\tau &\sim \Gamma(0.001, 0.001); \quad i \in \{auto, cyc, cam, moto\} \\
\beta_{i,1} &\sim N(0, 0.001); \quad i \in \{auto, cyc, cam, moto\} \\
\beta_{i,1}^{cdmx} &\sim N(0, 0.001); \quad i \in \{auto, cyc, cam, moto\} \\
\alpha &\sim N(0, \tau_a) \\
\alpha^{cdmx} &\sim N(0, \tau_{cdmx})
\end{aligned}$$

Distribuciones Iniciales

$$\begin{aligned}
\tau &\sim \Gamma(0.001, 0.001); \quad i \in \{auto, cyc, cam, moto\} \\
\beta_{i,1} &\sim N(0, 0.001); \quad i \in \{auto, cyc, cam, moto\} \\
\beta_{i,1}^{cdmx} &\sim N(0, 0.001); \quad i \in \{auto, cyc, cam, moto\} \\
\alpha &\sim N(0, \tau_a) \\
\alpha^{cdmx} &\sim N(0, \tau_{cdmx})
\end{aligned}$$

Parámetros de Suavizamiento

$$\begin{aligned}
\lambda &= 15 \\
\tau_i &= \lambda * \tau; \quad i \in \{auto, cyc, cam, moto\} \\
\tau_a &= \lambda * \tau \\
\tau_{cdmx} &= \lambda * \tau \\
\tau_i^{cdmx} &= \lambda * \tau
\end{aligned}$$

Resultados

La convergencia de las Cadenas de Markov se presenta al final del análisis, en el apartado del apéndice. Para cada uno de los parámetros en el tiempo se presenta un diagnóstico básico a partir de gráficos. Constituido por un histograma alrededor de la estimación puntual, de un análisis de convergencia de los promedios así como un autocorrelograma para observar la dependencia de la cadena consigo misma. Asimismo se presenta la traza de la cadena que muestra el recorrido que realiza la cadena sobre el espacio parametral. En términos generales se ven bastante bien la mayoría de las cadenas, se analizaron durante varias iteraciones y se aumentaron los parámetros de la cadena para obtener mejores resultados cuando se observaba convergencia lenta en algunos casos. Los valores finales de la cadena de simulación para el modelo dinámico fueron número total de iteraciones: 120,000 con número de iteraciones para burnin: 60000. Se tuvo que incrementar tanto el número debido a que la convergencia de tantos parámetros no es tan rápida como para modelos más sencillos.

A continuación se muestran los gráficos temporales para cada uno de los parámetros que fueron estimados, asimismo presentan intervalos de confianza para cada uno de ellos a lo largo del tiempo.

En particular se probó la hipótesis de que fuera necesaria una dinámica para las ordenadas. No obstante, una vez ejecutados varios modelos tanto gamma como normales con la especificación de dinámica autorregresiva en las alfas, se llega a la conclusión que los parámetros no cambian en el tiempo y que por lo tanto no es

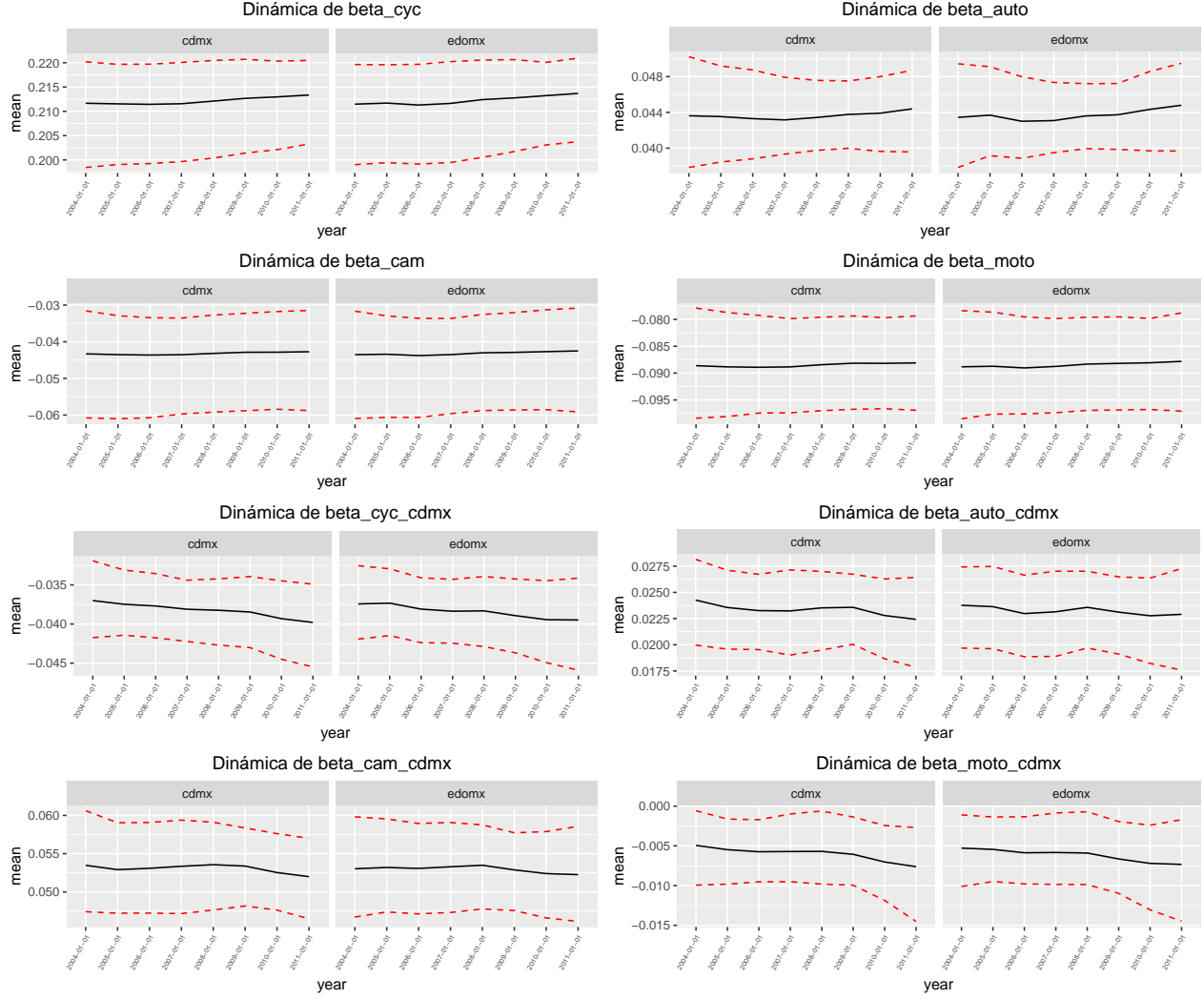
necesaria dicha dinámica. Más aun, cuando se restringen dichos parámetros a que la dinámica no exista como tal, el margen del DIC paso de -30 a -50 y un último esfuerzo aumentando las iteraciones de la cadena permitieron llegar a muy buenas predicciones para dos periodos en cada estado con un DIC de -65.

Table 3: Coeficientes de ordenadas

| | alpha_cdmx | alpha |
|-------|------------|------------|
| mean | 0.0000195 | 0.0000222 |
| sd | 0.0011916 | 0.0011602 |
| 2.5% | -0.0022949 | -0.0021006 |
| 25% | -0.0007849 | -0.0007182 |
| 50% | 0.0000019 | -0.0000453 |
| 75% | 0.0007773 | 0.0007750 |
| 97.5% | 0.0024709 | 0.0022466 |

En la tabla anterior se puede observar los valores para ambas pendientes, $\alpha_{cdmx} + \alpha$ para la Ciudad de Mexico y α para el Estado de México.

En términos cuantitativos no representan un valor muy distinto de cero, por lo que no se considera como algo significativo para el modelo, no obstante se considera incluido por completitud. A continuación se presentan los efectos que resultan de mayor interés.



En términos de la derivada, en el caso de la Ciudad de México esto es igual a

$$\frac{\partial E(\log(y_t)|\mathbf{x}, \theta)}{\partial x_i} = \exp(\eta_t) \frac{\beta_i + \beta_i^{cdmx}}{x_i} = E(\log(y_t)|x, \theta) * \frac{\beta_i + \beta_i^{cdmx}}{x_i} \quad i \in \{auto, cyc, cam, moto\}$$

De modo que con esta especificacion se tiene que existe un efector marginal relativo

$$\frac{\partial E(\log(y_t)|x, \theta)}{\partial x_i} / E(\log(y_t)|x, \theta) = \frac{\beta_{it} + \beta_{it}^{cdmx}}{x_{it}}.$$

Asi evaluando en la media histórica de los valores correspondientes se tiene que para el caso de la Ciudad de México, los cambios promedio relativo en escala logarítmica son los siguientes

$$\begin{aligned}
\frac{\beta_{auto,t} + \beta_{auto,t}^{cdmx}}{x_{auto,t}} &= 0.0048387 \\
\frac{\beta_{cyc,t} + \beta_{cyc,t}^{cdmx}}{x_{cyc,t}} &= 0.0174772 \\
\frac{\beta_{moto,t} + \beta_{moto,t}^{cdmx}}{x_{moto,t}} &= -0.0091395 \\
\frac{\beta_{cam,t} + \beta_{cam,t}^{cdmx}}{x_{cam,t}} &= 0.0010684
\end{aligned}$$

Es un caso análogo para el Estado de México, sin embargo ya no se toma en cuenta los β_i^{cdmx} .

$$\begin{aligned}
\frac{\beta_{auto,t}}{x_{auto,t}} &= 0.0033559 \\
\frac{\beta_{cyc,t}}{x_{cyc,t}} &= 0.0185979 \\
\frac{\beta_{moto,t}}{x_{moto,t}} &= -0.009259 \\
\frac{\beta_{cam,t}}{x_{cam,t}} &= -0.004819
\end{aligned}$$

De modo que al analizar los coeficientes, las combinaciones de betas representan el efecto relativo en escala logarítmica que tendría el aumento en la cantidad de vehículos sobre la cantidad de partículas pm2.5 en el aire. En este caso se debe exponenciar restar uno y multiplicar por cien para encontrar el cambio porcentual. Estos oscilan entre el .3 y .5 porciento para las variables de auto.

La mayor parte de las betas tuvieron un compartamiento con relativamente poca variación en el tiempo. No obstante, se tiene un comportamiento que es positivo pero decreciente, lo que iría acorde con la teoría de que mientras pasa el tiempo el efecto de cada vehículo sobre la contaminación debería ser menor. Esto se puede atribuir principalmente a mayores reglamentos o bien a los avances tecnológicos.

En términos prácticos la β_{cam}^{cdmx} es la única que contiene al cero dentro de los intervalos, de modo que es la única no contundentemente significativa. Las demás betas se pueden observar en las gráficas anteriores, en las que la línea negra denota el promedio obtenido a partir de las simulaciones de MCMC.

En lo subsecuente se analizará el desempeño de la predicción de las partículas a lo largo del tiempo. Para esto se ejecutó el modelo con sólo 8 observaciones para cada estado, y dos de ellas se utilizaron para medir la precisión que podía llegar a tener como instrumento predictivo.

Como se mencionó en la sección exploratoria $\log(y_t)$ tiene un comportamiento distinto en ambos estados (Edomex y Ciudad). Los niveles han sido mayores en la Ciudad de México durante todo el periodo y también en términos de las independientes. Para la predicción de los datos, los intervalos de confianza son más amplios para la Ciudad de México, lo que indica que hay más incertidumbre.

Las predicciones no son nada malas, sin embargo debido a la incertidumbre mayor que existe en este caso en la Ciudad de México, se tiene que no predice tan bien como en el caso del Estado de México, cuyo comportamiento resulta más estable en términos de las variables explicativas respecto al comportamiento que se tiene en la Ciudad de México que está compuesto de observaciones con mayor variabilidad.

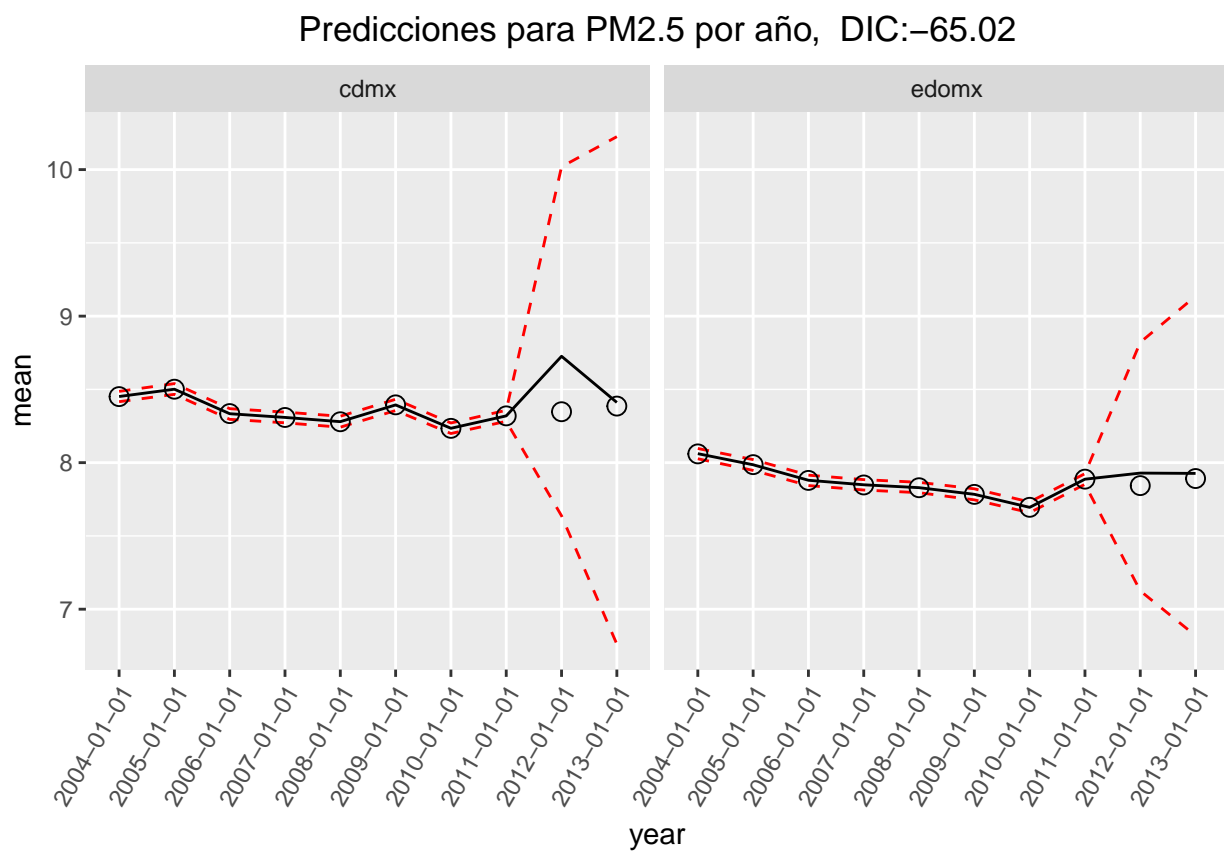


Figure 3: Predicción

En conclusión, se llegó a un modelo bueno a pesar de que fue largo el camino de prueba, error e imaginación, probando modelos simples y cada vez más complejos. Al grado de llegar a un modelo dinámico con el que se logró predecir de forma relativamente acertada. La especificación gamma junto con la transformación logarítmica representaron un salto hacia una mejor estimación tanto de los parámetros como las predicciones de la variable dependiente. Más aún se obtuvieron coeficientes que concuerdan con la teoría y que si bien la mayoría son positivos, estos decrecen a lo largo del tiempo lo que también explica el comportamiento de decaimiento en algunos casos como son los de la categoría de transporte de carga, sin perder de vista que tienen un efecto positivo en la concentración de las partículas suspendidas. Por otro lado se concluye que el efecto de los autos sobre la concentración de dichas partículas es el mayor y que por lo tanto representa un área de oportunidad para mejorar las condiciones ambientales en el Valle de México a través de estudios más especializados.

Referencias

- Congdon, P. (2001). Bayesian Statistical Modelling. Wiley: Chichester.
- Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. (2002). Bayesian Data Analysis, 2a. edición. Chapman & Hall: Boca Raton.
- Nieto-Barajas, L. E. & de Alba, E. (2014). Bayesian regression models. En Predictive Modeling Applications in Actuarial Science. E.W. Frees, R.A. Derrig & G. Meyers (eds.) Cambridge University Press, pp 334-366.
- Banerjee, S., Carlin, B. P. & Gelfand, A. (2014). Hierarchical Modeling and Analysis for Spatial Data, 2a. edición. Chapman & Hall: Boca Raton.
- R (<http://www.r-project.org/>)
- R Studio (<http://www.rstudio.com/>)
- OpenBUGS (<http://www.openbugs.net/>)
- JAGS (<http://sourceforge.net/projects/mcmc-jags/files/JAGS/>)
- <http://www.aire.df.gob.mx/>
- <http://www.inegi.org.mx/est/contenidos/proyectos/registros/default.aspx>
- http://www.ats.ucla.edu/stat/mult_pkg/faq/general/log_transformed_regression.htm

Apéndice

