



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Facultad de Ingeniería

**“MODELADO DE LA DEMANDA DE TURBOSINA
EN MÉXICO”**

Tesis para obtener el título de Ingeniero Industrial

P R E S E N T A:

RODERIC RUIZ SANABRIA

DIRECTOR: DR. ESTHER SEGURA PÉREZ



CIUDAD UNIVERSITARIA, MÉXICO, D.F.
OCTUBRE 2015

Agradecimientos

A mis padres que con todo su amor y comprensión, que con cada desvelo, cada llamada de atención, cada abrazo y felicitación, han permitido que llegue a este punto de mi vida. Y que estoy seguro, me permitirán seguir adelante.

A mi hermano que siempre estuvo ahí, con algún comentario o consejo, permitiéndome crecer como persona.

A la familia Sanabria, aquella del 289 de la 229, por ser parte de mi vida y estar ahí en los momentos agradables y aquellos de aprendizaje.

A todas esas personitas que han formado parte de mi vida, en cada escuela, en cada lugar en donde he estado y que permanecen hasta el día de hoy.

A los doctores Esther y Wulfrano por su apoyo en la realización de este trabajo de tesis.

A mi universidad por brindarme las bases y conocimientos para ejercer una profesión y darme la oportunidad de utilizarlos para la transformación del país.

Roderic

Índice de Contenido

Índice de Contenido	1
Índice de Tablas	4
Índice de Ilustraciones	6
Introducción	7
Capítulo 1. Panorama en México del mercado de Turbosina	9
Turbosina y su entorno en México	10
Turbosina	10
Origen de la Turbosina	11
El mercado	14
Futuro del mercado	16
Capítulo 2. Marco Teórico	17
Pronóstico	18
Proceso de Pronóstico	18
Técnicas de Pronósticos	18
Series de tiempo	20
Métodos Causales	20
Métodos de regresión	20
Regresión Lineal	21
Análisis de Panel	21
Beneficios y limitaciones del Panel de Datos	22
Modelos de Panel de Datos	23
Estimadores de Panel de Datos	26
Métodos para la evaluación de pronósticos	28
Error estándar de estimación	28
Coeficiente de Determinación y Correlación	29
R ² Ajustada	30
Forecast Value Added Analysis (FVA)	31
Econometría	31
Capítulo 3. Identificación de variables	33

Selección de Variables.....	34
VARIABLES ECONOMÉTRICAS.....	34
PIB.....	35
INPC.....	36
Precio del petróleo.....	37
Precio de la Turbosina.....	38
VARIABLES DE INFRAESTRUCTURA.....	39
Nuevo Aeropuerto de la Ciudad de México.....	40
Matrícula de aeronaves.....	41
VARIABLES RELACIONADAS A LOS USUARIOS.....	41
Operaciones.....	42
Pasajeros.....	42
Carga.....	42
Capacidad Hotelera.....	42
Crecimiento poblacional.....	43
VARIABLES DEL MEDIO DE DESARROLLO.....	44
Guerras.....	44
Crisis económicas.....	45
CONSIDERACIONES IMPORTANTES.....	45
Multicolinealidad entre variables.....	45
Disponibilidad de información.....	46
Costos asociados.....	46
Principio de parsimonia.....	46
VARIABLES EXCLUIDAS U OMITIDAS.....	47
VARIABLES INCLUIDAS EN EL ANÁLISIS.....	47
Capítulo 4. Construcción y evaluación.....	49
¿Qué tipo de análisis utilizar?.....	50
Modelo Preliminar.....	50
¿Cuál es la mejor estimación?.....	51
Evaluación del modelo preliminar.....	53
Evaluación de las variables explicativas y construcción de un nuevo modelo.....	54

Construcción del modelo final.....	58
Evaluación del modelo final	59
Resultados del modelo final	63
Un Pronóstico a partir del modelo	66
Capítulo 5. <i>Discusión, Conclusiones y Recomendaciones</i>	69
Discusión	70
Conclusiones	71
Recomendaciones	72
Anexo A	74
Anexo B	75
Regresión lineal simple.....	75
Mínimos Cuadrados Ordinarios.....	76
Regresión Lineal Múltiple.....	79
Supuestos de la regresión	79
Anexo C	81
Pruebas de Hipótesis.....	81
Prueba de Breush-Pagan	82
Prueba de Hausman	83
Anexo D.....	84
Modelo Efectos Fijos 2007-2015	84
Series de Errores Estándar de Estimación	85
Pruebas de Hausman no Concluyentes.	86
Estimadores del modelo: $D_{turbosina} = \alpha_i + \beta_1 INPC + \beta_2 PIB + \beta_3 \text{Log}(P_{petroleo}) + \beta_4$ <i>Log(Pturbosina) + \beta_5 Operaciones + \beta_6 Pasajeros + \beta_7 Carga+ \beta_8 CapHotelera + uit. (Periodo</i> <i>2010-2015).</i>	87
Series Logarítmicas del Gasto de Turbosina	88
Mejor estimador para el modelo: $\text{Log}(\text{Gasto de turbosina}) = \alpha_i + \beta_1 INPC + \beta_2 Operaciones + \beta_3$ <i>Pasajeros + \beta_4 Carga + \beta_5 CapHotelera</i>	89
Combinaciones de modelos (variables Operaciones, Carga y CapHotelera)	92
Indicadores para un modelo Lin-Lin	94
Referencias	95

Índice de Tablas

Tabla 1. Histórico del origen y destino de la turbosina en territorio nacional (su diferencia es el déficit de inventarios). <i>*información en miles de barriles diarios.</i>	14
Fuente: Sistema de Información Energética con información de Petróleos Mexicanos.	
Tabla 2. Aeropuertos con mayor consumo de turbosina en México (representando el 80% de la venta). Elaboración propia con datos de ASA. *Los datos de 2015 están hasta febrero. Los datos completos se encuentran en el anexo A	15
Tabla 3. Resumen de los estimadores de panel de datos y su comportamiento con los modelos principales al estimar β (Cameron y Trivedi 2005, 699)	28
Tabla 4. Ponderaciones por grupo de objeto de gasto en por ciento. Las ponderaciones apuntadas se aplicarán a partir de la primera quincena de abril de 2013, y permanecerán constantes durante la vigencia del patrón de consumo de los hogares en la República Mexicana. (Diario Oficial de la Federación 2013)	37
Tabla 5. Estimación de Efectos fijos para el modelo $\text{Log (Gasto de turbosina)} = \alpha_i + \beta_1 \text{INPC} + \beta_2 \text{Operaciones} + \beta_3 \text{Pasajeros} + \beta_4 \text{Carga} + \beta_5 \text{CapHotelera}$	58
Tabla 6. Coeficientes de correlación entre las variables de operaciones, pasajeros y carga.	58
Tabla 7. Estimación Pooled OLS en Stata	60
Tabla 8. Estimación de efectos fijos en Stata	60
Tabla 9. Estimación de efectos aleatorios en Stata	61
Tabla 10. Prueba de Breush-Pagan calculada en Stata, que desestima al modelo Pooled OLS	61
Tabla 11. Prueba de Hausman con los coeficientes sin escalar	62
Tabla 12. Prueba de Hausman con los coeficientes escalados calculado en Stata	62
Tabla 13. Estimador de efectos fijos y cada uno de los efectos individuales para el modelo: $\text{Log (Gasto de turbosina)} = \alpha_i + \beta_1 \text{INPC} + \beta_2 \text{Pasajeros} + \beta_3 \text{Carga} + \beta_4 \text{CapHotelera}$ El aeropuerto cuyo intercepto fue tomado como referencia es el AICM.	63
Tabla 14. MAPE y error estándar de estimación (σ) para cada aeropuerto, las columnas con subíndice log se calcularon con los datos que entrega directamente el modelo. Las otras columnas se calcularon aplicando la transformación correspondiente. * Los datos de total, fueron calculados como promedio ponderado, de acuerdo a la participación que tiene cada aeropuerto en el consumo de turbosina.	65
Tabla 15. Pronósticos para el Log (Consumo de turbosina), Consumo de turbosina y la demanda de Turbosina para el mes de febrero de 2015, usando las variables explicativas pronosticadas mediante el método de Holt – Winters. Nótese que los errores de estimación son menores para el pronóstico del Log (Consumo de turbosina), esto es debido a que se trata de la variable que directamente pronostica el modelo.	67
Tabla 16. Pronóstico e intervalos de confianza con $\alpha = 0.05$ para el mes de febrero de 2015 de la variable de respuesta Log (Consumo de turbosina). Para este pronóstico se usaron los datos reales de las variables explicativas correspondientes al mes de febrero de 2015.	68
Tabla 17. Variables explicativas y su respectivo coeficiente. Elaboración propia con datos de cálculos en STATA	71
Tabla 18. Demanda anual de turbosina en los aeropuertos mexicanos 2007-2015*. Elaboración propia con datos de ASA. *Los datos de 2015 son hasta febrero.	74
Tabla 19. Modelo de efectos fijos para el periodo 2007-2015	84
Tabla 20. Ejemplos de la inconsistencia de la Prueba de Hausman para el modelo $D_{\text{turbosina}} = \alpha_i + \beta_1 \text{INPC} + \beta_2 \text{PIB} + \beta_3 \text{Log(Ppetroleo)} + \beta_4 \text{Log(Pturbosina)} + \beta_5 \text{Operaciones} + \beta_6 \text{Pasajeros} + \beta_7 \text{Carga} + \beta_8 \text{CapHotelera} + \text{uit. (Periodo 2010-2015)}$	86

<i>Tabla 21. Estimación Pooled OLS para el modelo $D_{turbosina} = \alpha_i + \beta_1 INPC + \beta_2 PIB + \beta_3 \text{Log}(P_{petroleo}) + \beta_4 \text{Log}(P_{turbosina}) + \beta_5 \text{Operaciones} + \beta_6 \text{Pasajeros} + \beta_7 \text{Carga} + \beta_8 \text{CapHotelera} + \text{uit. (Periodo 2010-2015)}$.</i>	87
<i>Tabla 22. Estimación de Efectos Fijos para el modelo $D_{turbosina} = \alpha_i + \beta_1 INPC + \beta_2 PIB + \beta_3 \text{Log}(P_{petroleo}) + \beta_4 \text{Log}(P_{turbosina}) + \beta_5 \text{Operaciones} + \beta_6 \text{Pasajeros} + \beta_7 \text{Carga} + \beta_8 \text{CapHotelera} + \text{uit. (Periodo 2010-2015)}$.</i>	87
<i>Tabla 23. Estimación de Efectos Aleatorios para el modelo $D_{turbosina} = \alpha_i + \beta_1 INPC + \beta_2 PIB + \beta_3 \text{Log}(P_{petroleo}) + \beta_4 \text{Log}(P_{turbosina}) + \beta_5 \text{Operaciones} + \beta_6 \text{Pasajeros} + \beta_7 \text{Carga} + \beta_8 \text{CapHotelera} + \text{uit. (Periodo 2010-2015)}$.</i>	87
<i>Tabla 24. Estimación Pooled OLS realizada en Stata</i>	89
<i>Tabla 25. Estimación de Efectos Fijos realizada en Stata</i>	89
<i>Tabla 26. Estimación de Efectos Aleatorios realizada en Stata</i>	90
<i>Tabla 27. Prueba de Breush-Pagan que compara la estimación de Efectos Aleatorios con la estimación Pooled OLS. Desestima al modelo Pooled OLS.</i>	90
<i>Tabla 28. Prueba de Hausman con los coeficientes sin escalar</i>	91
<i>Tabla 29. Prueba de Hausman con los coeficientes escalados.</i>	91
<i>Tabla 30. Estimación de efectos fijos para el modelo $\text{Log}(\text{Gasto de turbosina}) = \alpha_i + \beta_1 INPC + \beta_2 \text{Operaciones} + \beta_3 \text{Carga} + \beta_4 \text{CapHotelera}$</i>	92
<i>Tabla 31. Estimación de efectos fijos para el modelo $\text{Log}(\text{Gasto de turbosina}) = \alpha_i + \beta_1 INPC + \beta_2 \text{Operaciones} + \beta_3 \text{Pasajeros} + \beta_4 \text{CapHotelera}$</i>	92
<i>Tabla 32. Estimación de efectos fijos para el modelo $\text{Log}(\text{Gasto de turbosina}) = \alpha_i + \beta_1 INPC + \beta_2 \text{Pasajeros} + \beta_3 \text{Carga} + \beta_4 \text{CapHotelera}$</i>	93
<i>Tabla 33. Indicadores de desempeño para el modelo Lln-Lin de ecuación: $\text{Gasto de turbosina} = \alpha_i + \beta_1 INPC + \beta_2 \text{Pasajeros} + \beta_3 \text{Carga} + \beta_4 \text{CapHotelera}$</i>	94

Índice de Ilustraciones

<i>Ilustración 1. Participación en la venta de petrolíferos en México 2005-2014. Fuente: Elaboración propia con información de PEMEX-Refinación.</i>	10
<i>Ilustración 2. Consumo anual de turbosina en Millones de litros. Fuente: Elaboración propia con información de ASA.</i>	11
<i>Ilustración 3. Instalaciones petroleras en 2013. (PEMEX 2013)</i>	12
<i>Ilustración 4. Instalaciones dedicadas a la refinación y distribución de combustibles. (PEMEX 2013)</i>	12
<i>Ilustración 5. Producción de turbosina en miles de barriles diarios. Fuente: elaboración propia con datos del Sistema de Información Energética con información de Petróleos Mexicanos</i>	13
<i>Ilustración 6. Importaciones y exportaciones de turbosina en miles de barriles diarios. Fuente: elaboración propia con datos del Sistema de Información Energética con información de Petróleos Mexicanos</i>	13
<i>Ilustración 7. Camión de carga de turbosina de ASA. Fuente: Verónica Garzón. Aeropuerto Internacional de Chihuahua.</i>	14
<i>Ilustración 8. Red Aeroportuaria Mexicana. Elaboración propia con datos de PEMEX.</i>	15
<i>Ilustración 9. Técnicas de Pronósticos. (Escalante Membrillo 2005, 15)</i>	19
<i>Ilustración 10. Proyecto del Nuevo Aeropuerto internacional de la Ciudad de México (Redacción El Universal 2014)</i>	40
<i>Ilustración 11. Datos INEGI (Cuentame... población 2014)</i>	43
<i>Ilustración 12. Variables de precios de Turbosina y precio de Petróleo (Mezcla Mexicana) con su respectiva transformación logarítmica, en donde se puede apreciar la disminución de dispersión en los datos.</i>	51
<i>Ilustración 13. Algoritmo para realizar la selección del tipo de estimación que se tomará para realizar el pronóstico del gasto de turbosina. Elaboración propia.</i>	52
<i>Ilustración 14. Graficas de las variables de precios expresadas en unidades monetarias y en logaritmos a partir de enero de 2010 hasta enero de 2015. 1 barril = 159 litros.</i>	54
<i>Ilustración 15. Comportamiento del gasto de turbosina (en millones de pesos) en los 10 aeropuertos más importantes de México, para el periodo de Enero 2010 a Enero 2015.</i>	56
<i>Ilustración 16. Recta de ajuste de regresión lineal aplicada a una serie de datos (Gujarati y Porter 2009, 47)</i>	75
<i>Ilustración 17. Apreciación del tercer supuesto de regresión (Escalante Membrillo 2005, 34)</i>	76
<i>Ilustración 18. Valores observados y estimación lineal. (Gujarati y Porter 2009, 56)</i>	77
<i>Ilustración 19. Posibles errores cometidos en las pruebas de hipótesis. (Gujarati y Porter 2009)</i>	81
<i>Ilustración 20. Regiones de aceptación y rechazo de H_0 para el estadístico de prueba con distribución normal Z y un nivel de significancia de 5%. (Gujarati y Porter 2009)</i>	82
<i>Ilustración 21. Errores estándar de estimación para aeropuertos seleccionados, en el periodo 2007-2015. La mayoría de los aeropuertos muestran comportamientos atípicos durante los primeros 36 meses (periodo 2007-2009) que coincide con el periodo de crisis económica a nivel mundial.</i>	85
<i>Ilustración 22. Series logarítmicas del gasto de turbosina para los 10 principales aeropuertos de la República Mexicana, para el periodo 2010-2015</i>	88

Introducción

Objetivo: *Diseñar, analizar y evaluar un modelo que permita identificar las variables que impactan en la demanda y sirva para estimar el consumo de turbosina en cada aeropuerto mexicano. Lo anterior con base en estadística matemática (modelos econométricos) usando datos de panel.*

La propuesta de esta tesis se deriva del hecho que no existen modelos que pronostiquen el consumo de turbosina para cada aeropuerto en México. Metodologías usadas en informes prospectivos de petrolíferos¹ toman en cuenta la demanda global, dejando de lado el consumo local de cada terminal aérea.

Un punto interesante de esta tesis que sirve como hipótesis de la misma, es que además de impactar variables macroeconómicas en el consumo (como lo sugieren los modelos propuestos por la Secretaría de Energía²), también impactan variables microeconómicas.

El uso de pronósticos para la toma de decisiones es una de las actividades que se han vuelto de vital importancia en las empresas, ya que hacer un buen pronóstico, puede determinar el éxito o fracaso de una empresa. Por lo anterior, seleccionar el método de pronóstico es primordial para poseer información adecuada que proporcione criterios sólidos en la toma de decisiones.

Cuando se ocupan métodos causales, las implicaciones del análisis van más allá del pronóstico en sí mismo, pues al seleccionar adecuadamente las variables que impactan en el consumo de cualquier producto, y conocer su origen, es posible hacer estrategias de comercialización o estudios de mercado más precisos. De esta forma se tendría el conocimiento de hacia dónde se deben de dirigir los recursos de la organización y el rumbo que se seguirá.

Debido al crecimiento de la industria del transporte aéreo y en general de la aeronáutica en México y el mundo, el producto con el que se planteará el diseño de este modelo de pronósticos es turbosina. Siendo este el combustible más usual en dicha industria, se vuelve de gran interés conocer su comportamiento en los próximos años.

La construcción de un modelo que estime un comportamiento futuro es una tarea desafiante, misma que se encuentra en el desarrollo del presente trabajo de tesis. De esta forma esta tesis se divide en cinco capítulos. El primero presenta un panorama general de la situación de la turbosina en México. Mientras que el segundo capítulo aborda los principales tópicos que permitan una justificación matemática y econométrica al análisis que se realiza. El tercer capítulo presenta la forma en la que se seleccionaron las variables involucradas en el modelo y la formación de un modelo econométrico preliminar.

¹ Prospectiva de Petrolíferos 2004-2013 (SENER)
Prospectiva de Petrolíferos 2010-2025 (SENER)
Prospectiva de Petróleo Crudo y Petrolíferos 2013-2027 (SENER)

² Nota 1.

Por lo anterior el cuarto capítulo muestra el proceso de validación y obtención de un modelo final para su estudio y los resultados obtenidos. Finalmente el quinto apartado servirá para la discusión de lo obtenido en el capítulo cuarto y las conclusiones de esta tesis.

Capítulo 1.

*Panorama en México del
mercado de Turbosina*



Taca Intl Airlines y Embraer ERJ-190-100IGW (TI-BCF) con atardecer al fondo.
Foto: Jorge Rafael Solano Rodríguez. Fuente: www.aviationcorner.net

En este capítulo se explica qué es la turbosina, sus usos, además de hacer un breve recuento de su situación frente a otros combustibles en México. Se habla además, de las empresas productoras y distribuidoras de este combustible en el país, así como la normatividad vigente.

Turbosina y su entorno en México

Turbosina

El Jet A-1 (Catálogo de productos ECOPetrol VSM-01 2010), también conocido como turbo combustible, turbosina o JP-1A, es un destilado medio proveniente de la destilación atmosférica del petróleo, tiene características especiales de calidad, y es tratado químicamente para eliminar compuestos azufrados tales como sulfuros y ácidos nafténicos³, que pueden tener un comportamiento corrosivo.

La turbosina (PEMEX Refinación s.f.), es el combustible más utilizado en los aviones con motores de turbina o a reacción. Por lo que el sector aeronáutico es prácticamente el único sector donde se comercializa turbosina.

Para el mercado de petrolíferos en México, la turbosina tiene una participación del 4.09%, lo que representó un promedio diario de 60 mil barriles diarios en el periodo de 2005 a 2014.

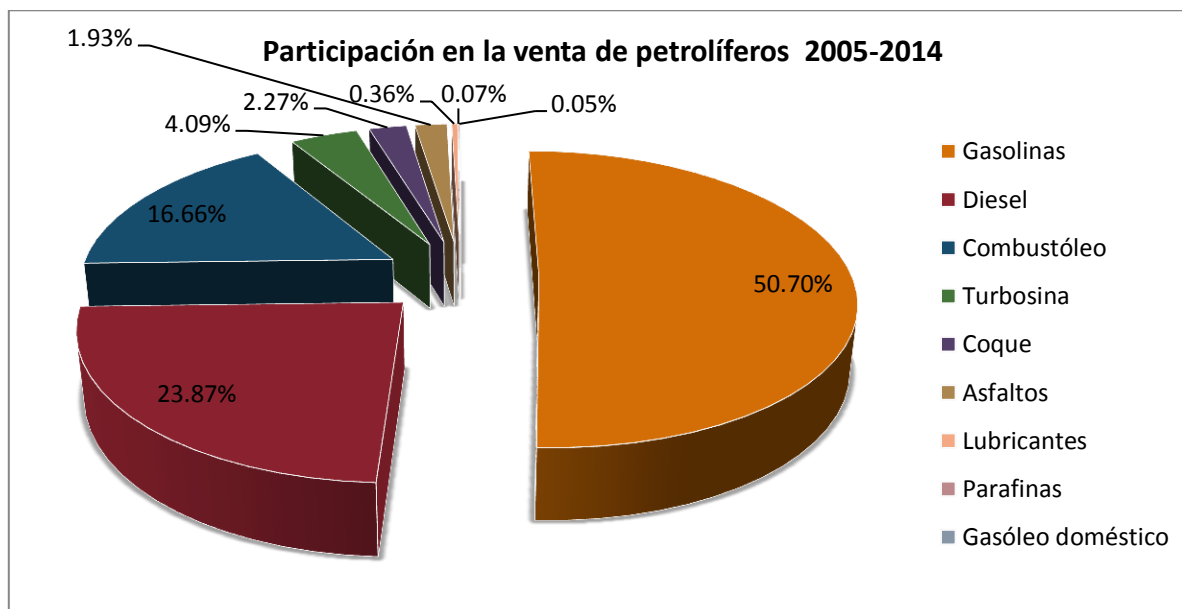


Ilustración 1. Participación en la venta de petrolíferos en México 2005-2014. Fuente: Elaboración propia con información de PEMEX-Refinación.

Sin embargo a pesar de eso, no es posible decir que el consumo de combustibles, en particular de turbosina se comporte constante a lo largo del tiempo. Su consumo ha

³ Los ácidos nafténicos son de origen orgánico y de tipo carboxílico. Estos ácidos están presentes en los crudos y han sido usados ampliamente como materia prima de varios productos; sin embargo, al momento de procesar crudos con altas concentraciones de estos, se convierten en un problema debido a la alta corrosividad que presentan a temperaturas elevadas (Pabón , y otros 2009).

variado a lo largo de los años, al sufrir una estrepitosa caída entre 2008 y 2010 al pasar de 231 millones de litros vendidos a poco menos de 157 millones de litros. Por otro lado también se observa una tendencia de crecimiento en los últimos años al pasar de 173 millones de litros vendidos en 2012 a poco más de 204 millones de litros vendidos en 2014.

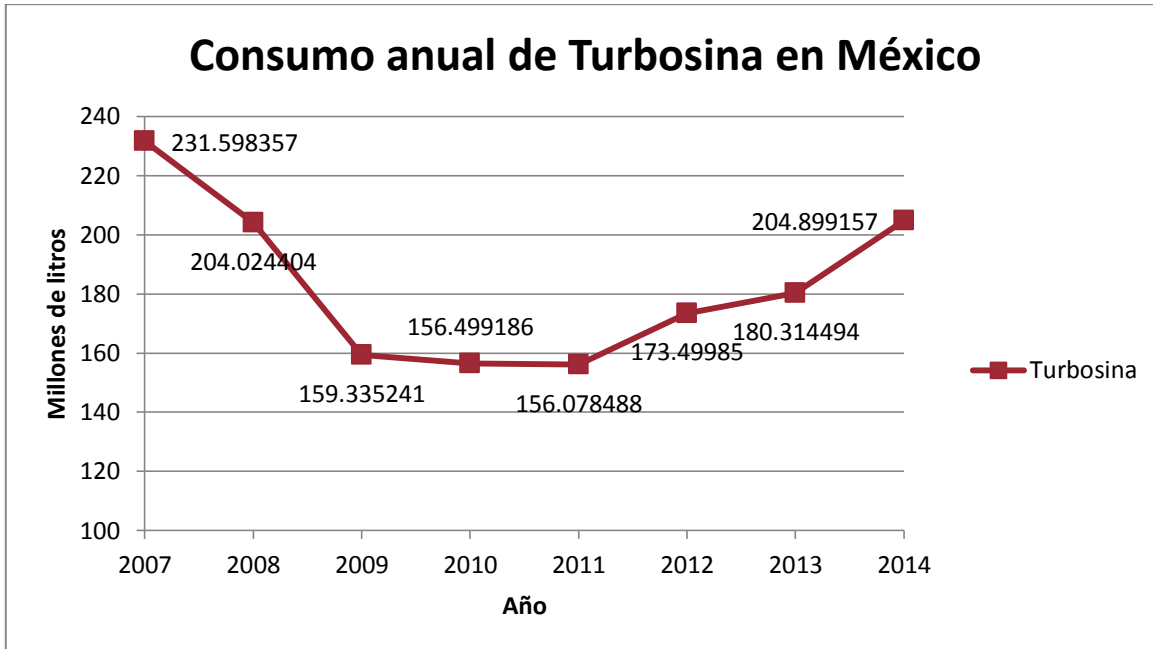


Ilustración 2. Consumo anual de turbosina en Millones de litros. Fuente: Elaboración propia con información de ASA.

Origen de la Turbosina

La turbosina que es utilizada en los aeropuertos nacionales es producida en su mayoría por Petróleos Mexicanos (PEMEX), mientras que el resto es importada. Esta paraestatal es la que se encarga de proveer el combustible en toda la república mexicana. De este modo, PEMEX se encarga tanto de la producción como la importación y exportación de combustibles en territorio nacional.

El abastecimiento de combustible es un problema de carácter mundial que lleva a los países a considerar diferentes alternativas para satisfacerla. En el caso de México, se cuenta con dos formas de abastecer a la demanda nacional; la primera de ellas es a partir de la producción realizada en las refinerías de PEMEX y por otro lado en caso de no contar con la capacidad necesaria para satisfacer la demanda, se recurre a importaciones.

México se caracteriza por ser un país petrolero, sin embargo eso no lo exenta de recurrir a la importación de combustibles para satisfacer la demanda nacional. La principal vía para la obtención de turbosina es la producción nacional.

La red con que cuenta PEMEX se encuentra distribuida en todo el país, ubicando a la zona de extracción de crudo en el Golfo de México; los centros de refinación se distribuyen en la zona del Golfo y centro del país. Algunos de los transportes de crudo y combustible se deben de realizar vía marítima, principalmente para abastecer a las penínsulas de Baja California y Yucatán.



Ilustración 3. Instalaciones petroleras en 2013. (PEMEX 2013)

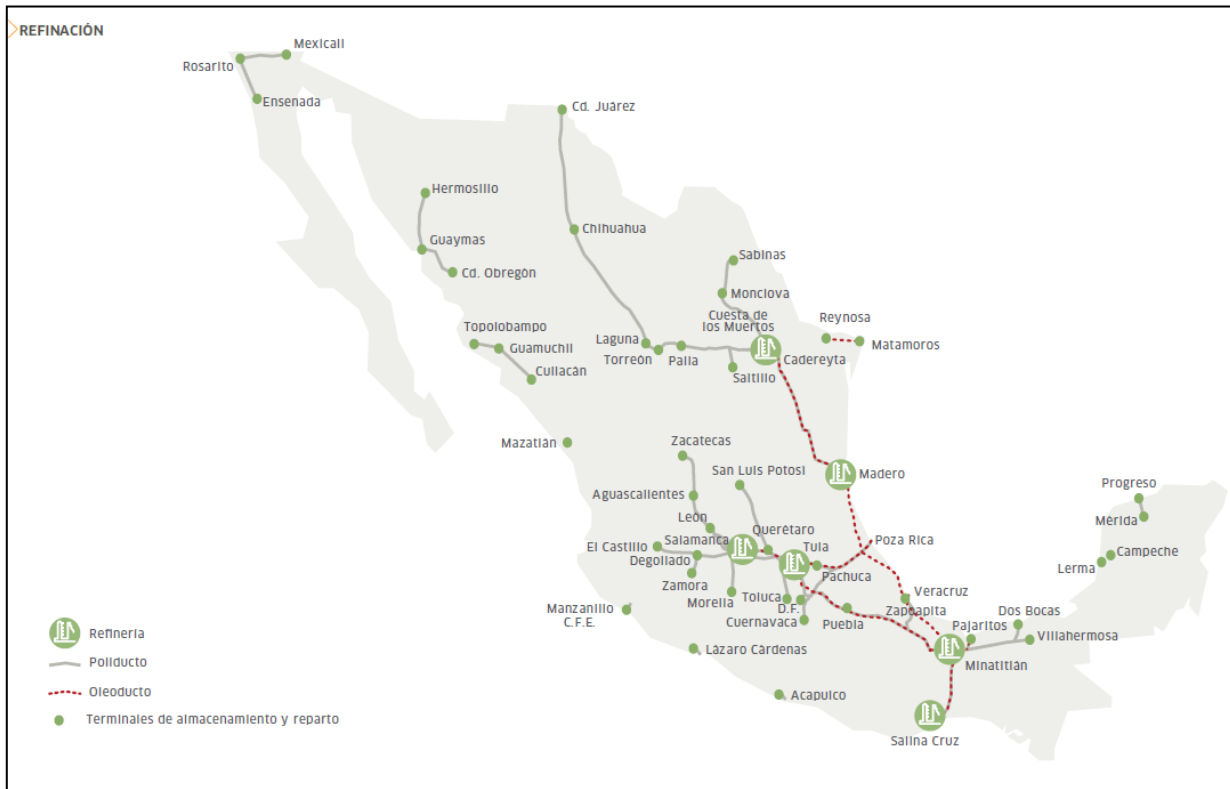


Ilustración 4. Instalaciones dedicadas a la refinación y distribución de combustibles. (PEMEX 2013)

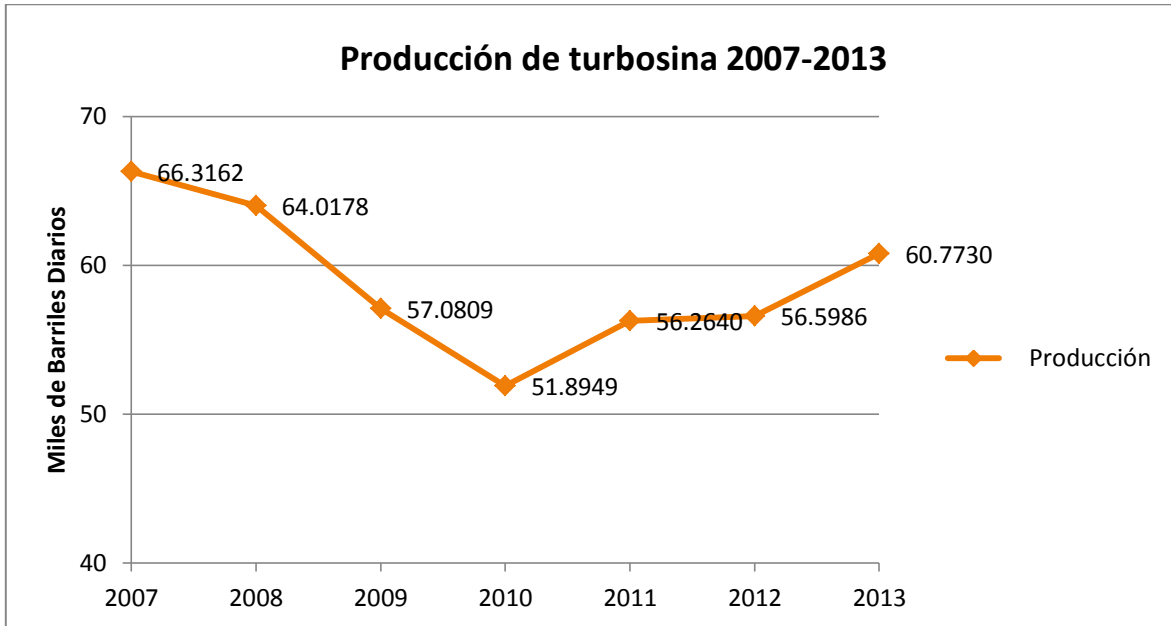


Ilustración 5. Producción de turbosina en miles de barriles diarios. Fuente: elaboración propia con datos del Sistema de Información Energética con información de Petróleos Mexicanos

Un análisis visual de las importaciones y exportaciones de turbosina permite observar que a pesar de que las importaciones se comportan de manera oscilante, tienen una tendencia negativa, y al mismo tiempo las exportaciones presentan una caída marcada al ser comparada con éstas.

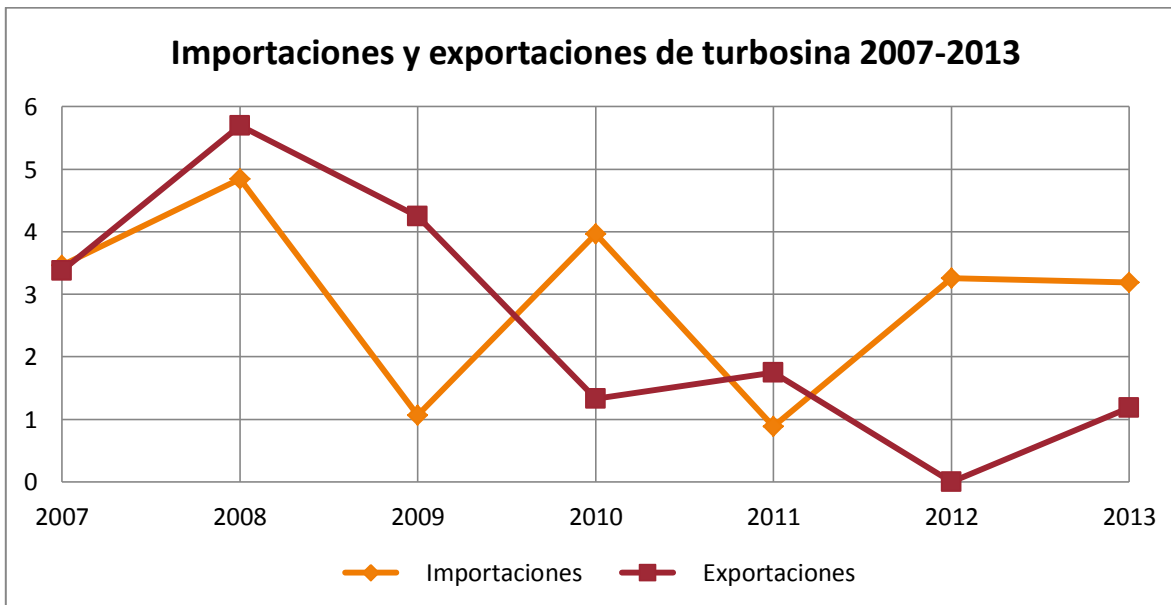


Ilustración 6. Importaciones y exportaciones de turbosina en miles de barriles diarios. Fuente: elaboración propia con datos del Sistema de Información Energética con información de Petróleos Mexicanos

Tabla 1. Histórico del origen y destino de la turbosina en territorio nacional (su diferencia es el déficit de inventarios).

*información en miles de barriles diarios.

Fuente: Sistema de Información Energética con información de Petróleos Mexicanos.

Turbosina	2007	2008	2009	2010	2011	2012	2013
Origen	66.3162	64.0178	57.0809	51.8949	56.2640	56.5986	60.7730
Producción	66.3162	64.0178	57.0809	51.8949	56.2640	56.5986	60.7730
Cadereyta	9.8308	6.7714	5.2322	4.2962	5.3393	4.7485	3.8322
Madero	4.8957	6.8270	5.9882	5.4892	5.9938	6.3781	7.4616
Tula	24.1907	22.6919	22.2107	22.1783	23.0749	23.8650	22.4299
Salamanca	10.4979	10.4919	8.7045	8.1050	7.5340	8.2782	10.6067
Minatitlán	0.3181	0.1821	0.0604	0.0000	0.0242	2.1220	0.0667
Salina Cruz	16.5830	17.0536	14.8849	13.2583	16.0776	11.7382	16.6952
Importaciones	3.4656	4.8455	1.0641	3.9630	0.8871	3.2597	3.1927
Destino	71.2703	70.6599	59.2519	57.1514	57.8720	59.2975	63.4079
Demanda interna	67.8903	64.9673	55.0051	55.8214	56.1267	59.2975	62.2256
Sector transporte	67.8903	64.9673	55.0051	55.8214	56.1267	59.2975	62.2256
Exportaciones	3.3800	5.6926	4.2468	1.3300	1.7453	0.0000	1.1823
Déficit de inventarios	-4.9541	-6.6421	-2.1710	-5.2565	-1.6079	-2.6990	-2.6349

Un análisis adecuado de la demanda de turbosina sin lugar a dudas ayudaría a reducir el déficit en inventarios que presenta PEMEX con respecto a la cantidad de combustible que dispone y la que es demandada, tal como se observa en la tabla anterior.

El mercado

La turbosina es uno de los costos principales de las aerolíneas y podrían representar hasta una tercera parte de sus gastos de operación. Algunas aerolíneas pueden cubrirse, pero por la misma naturaleza de la variación de la turbosina, debido a la volatilidad del precio del petróleo, muchas la dejan como cobertura normal para lo que cueste en el mercado pero no para tener un precio fijo (Cantera 2014).

La red aeroportuaria mexicana civil está compuesta por 61 aeropuertos tanto nacionales como internacionales, de los que 18 son operados en su totalidad por Aeropuertos y Servicios Auxiliares (ASA). Los demás aeropuertos tienen gobiernos independientes como el Aeropuerto Internacional de la Ciudad de México (AICM), o son operados por grupos aeroportuarios tales como el Grupo Aeroportuario del Pacífico y Aeropuertos del Sureste (ASUR). Sin embargo algunos servicios complementarios, en específico el suministro de combustible, son operados en su totalidad por ASA.



Ilustración 7, Camión de carga de turbosina de ASA. Fuente: Verónica Garzón. Aeropuerto Internacional de Chihuahua.



Ilustración 8. Red Aeroportuaria Mexicana. Elaboración propia con datos de PEMEX.

Aeropuertos y Servicios Auxiliares (ASA) es un Organismo descentralizado del Gobierno Federal con personalidad jurídica y patrimonios propios, que opera, administra y construye aeropuertos; presta servicios de suministro de combustibles ofrece asistencia técnica y consultoría, así como instrucción e investigación en materia aeronáutica y aeroportuaria participa en el desarrollo tecnológico y colabora con la Secretaría de Comunicaciones y Transportes en materia de regulación, verificación y supervisión de aeropuertos (Aeropuertos y Servicios Auxiliares s.f.).

Desde su creación en los 60's, ASA ha sido la compañía que controla el suministro de combustible en toda la red de aeropuertos mexicanos. De esta forma los precios de turbosina eran enteramente control de esta compañía. A pesar de ese control del precio, la oportunidad de las aerolíneas de comprar combustible en aeropuertos fuera del país, ha permitido que el precio de la turbosina fluctúe junto con los precios del mercado internacional, obligando a ASA a tener precios competitivos.

Tabla 2. Aeropuertos con mayor consumo de turbosina en México (representando el 80% de la venta). Elaboración propia con datos de ASA. *Los datos de 2015 están hasta febrero. Los datos completos se encuentran en el anexo A

Estación	2010	2011	2012	2013	2014	2015*
MEX México	1,252,424,694	1,306,942,967	1,382,562,604	1,423,724,441	1,498,339,140	236,305,922
CUN Cancún	527,932,460	554,808,670	625,592,589	690,810,457	730,645,668	137,796,307
GDL Guadalajara	250,885,454	240,635,712	234,052,558	246,720,769	264,702,118	43,834,480
MTY Monterrey	147,694,744	139,990,793	156,757,860	156,858,470	178,279,841	30,834,563
TIJ Tijuana	147,538,981	134,368,903	143,829,755	162,289,881	167,383,514	24,710,523
TLC Toluca	113,515,404	95,498,492	73,440,147	79,590,320	75,789,521	12,109,122
SJD San José Del Cabo	96,350,862	97,894,813	107,652,787	111,241,863	103,549,037	18,558,627
PVR Puerto Vallarta	94,845,643	87,362,388	88,569,302	86,587,801	105,947,322	27,879,488
MID Mérida	47,651,747	52,822,877	46,148,878	42,657,702	46,219,525	7,952,891
HMO Hermosillo	38,841,623	39,725,215	44,226,878	43,421,333	41,930,664	6,320,023

Futuro del mercado

El control del suministro de combustible que tiene ASA en la red aeroportuaria de México, se espera que termine en los próximos años con la llegada de compañías extranjeras y la creación de empresas nacionales que participen en la producción y distribución de combustibles.

La normatividad jurídica (después de las reformas establecidas en el Decreto publicado en el Diario Oficial de la Federación el 26 de enero de 2015) que rige la operación de aeropuertos y prestadoras de servicios establece que:

Artículo 17. La Secretaría otorgará permisos a personas físicas, o personas morales constituidas conforme a las leyes mexicanas, para la administración, operación, explotación y, en su caso, construcción de aeródromos civiles distintos a los aeropuertos.

Para aeródromos de servicio general, el permiso se otorgará exclusivamente a sociedades mercantiles mexicanas, e incluirán las actividades de administración, operación, explotación y, en su caso, construcción.

Los permisos se otorgarán previo cumplimiento de los requisitos exigidos en esta Ley y sus reglamentos; por los plazos que señale el permiso respectivo, pero en ningún caso podrán exceder de treinta años y podrán ser prorrogados por tiempo determinado, siempre que se hubiese cumplido con lo previsto en el título y se acepten las nuevas condiciones que establezca la Secretaría. (Ley de Aeropuertos 2015)

De este modo las operaciones de ASA y las concesiones con las que cuenta podrán cambiar en lo subsecuente como lo determina el artículo Tercero Transitorio de dicha ley:

Tercero. El organismo público descentralizado Aeropuertos y Servicios Auxiliares podrá continuar administrando aeropuertos en los términos de su Decreto de creación de fecha 10 de junio de 1965, publicado en el Diario Oficial de la Federación el 12 del mismo mes y año, y de sus decretos modificatorios, hasta en tanto la Secretaría, de conformidad con lo previsto en esta Ley, otorgue concesiones respecto de los aeropuertos administrados por dicho organismo.

Lo anterior, en el entendido de que Aeropuertos y Servicios Auxiliares deberá ajustarse a lo dispuesto en esta Ley, en lo relativo a la construcción, administración, operación y explotación de aeropuertos y en la prestación de los servicios (Ley de Aeropuertos 2015).

Con lo anterior se da la oportunidad de apertura a nuevos competidores que en podrán permitir un mercado más competitivo ofreciendo mejores tarifas y calidad a sus clientes. El generar competencia es uno de los mecanismos naturales para incentivar a la economía nacional y dar a los usuarios menores precios en los transportes o viajes realizados.

Capítulo 2.

Marco Teórico



Spanair Airbus A320-232 (EC-JNC). Aeropuerto Adolfo Suárez Madrid-Barajas.
Fuente: www.aviationcorner.net

En este capítulo se abordan los antecedentes matemáticos, econométricos y estadísticos que permitirán entender el desarrollo del presente trabajo de tesis. Estos antecedentes abarcan ciertas definiciones de conceptos básicos y necesarios para comprender este texto, e inclusive el desarrollo de algunas fórmulas y métodos considerados como pertinentes para el lector.

Pronóstico

La palabra pronóstico (Escalante Membrillo 2005, 11) se deriva del griego *prognôstikon* (latín *pronosticum*), que significa conjetura acerca de lo que puede suceder.

Se entiende por pronóstico a toda acción de predecir alguna tendencia, suceso o acontecimiento futuro. De esta manera el objetivo de un pronóstico es el de proveer información cercana a la realidad a partir de la información disponible, que posteriormente servirá para tomar decisiones sobre las acciones a tomar en el presente y en el futuro.

En el mundo actual, la mayoría de las decisiones de negocio se toman con la ayuda de alguna estimación o pronóstico. Lo anterior sirve para determinar la factibilidad y rentabilidad de un nuevo producto, o el comportamiento de las ventas y del mercado para un determinado bien o servicio. Así se pueden tomar determinaciones estratégicas en los negocios.

El valor de un pronóstico radica en la importancia de las decisiones que se pueden tomar a partir de él. De este modo, un error podría costar la propia vida de la empresa o el proyecto.

Proceso de Pronóstico.

De acuerdo con Lahiri & Sheng (Stekler 2014) el proceso de pronóstico es un proceso Bayesiano⁴ de decisión. Haciendo un marco de trabajo en donde los encargados del pronóstico empiezan con datos históricos (antecedentes), para entonces modificar sus predicciones después de algún número de meses en que se obtuvo mayor información.

Tomando en cuenta lo anterior, se han formulado diversas metodologías para poder realizar el pronóstico, en donde se toman en cuenta diferentes variables y elementos que permitan dar una estimación.

Técnicas de Pronósticos

El uso de una técnica o metodología para la realización de un pronóstico, permite que la incertidumbre alrededor de la estimación disminuya. Sin embargo, cada metodología cuenta con una efectividad diferente, de acuerdo al tipo de situación en la que se aplique.

Principalmente, la metodología empleada dependerá de cuatro cosas:

- Calidad y disponibilidad de datos históricos

⁴ La metodología bayesiana está basada en la interpretación subjetiva de la probabilidad y tiene como punto central el Teorema de Bayes.

- Exactitud con la que se quiere pronosticar
- Periodo de tiempo que se quiere pronosticar.
- Costo de la metodología empleada.

En general, cuando el tiempo del pronóstico aumenta, la exactitud del pronóstico se verá afectada de manera negativa. Por lo que escoger los alcances del pronóstico se convierte en pieza fundamental a seleccionar.

De acuerdo a lo anterior se pueden identificar dos principales tipos de técnicas de pronósticos: **cualitativas y cuantitativas**. Estos a su vez se subdividen en más subgrupos quedando organizados como se muestra a continuación:

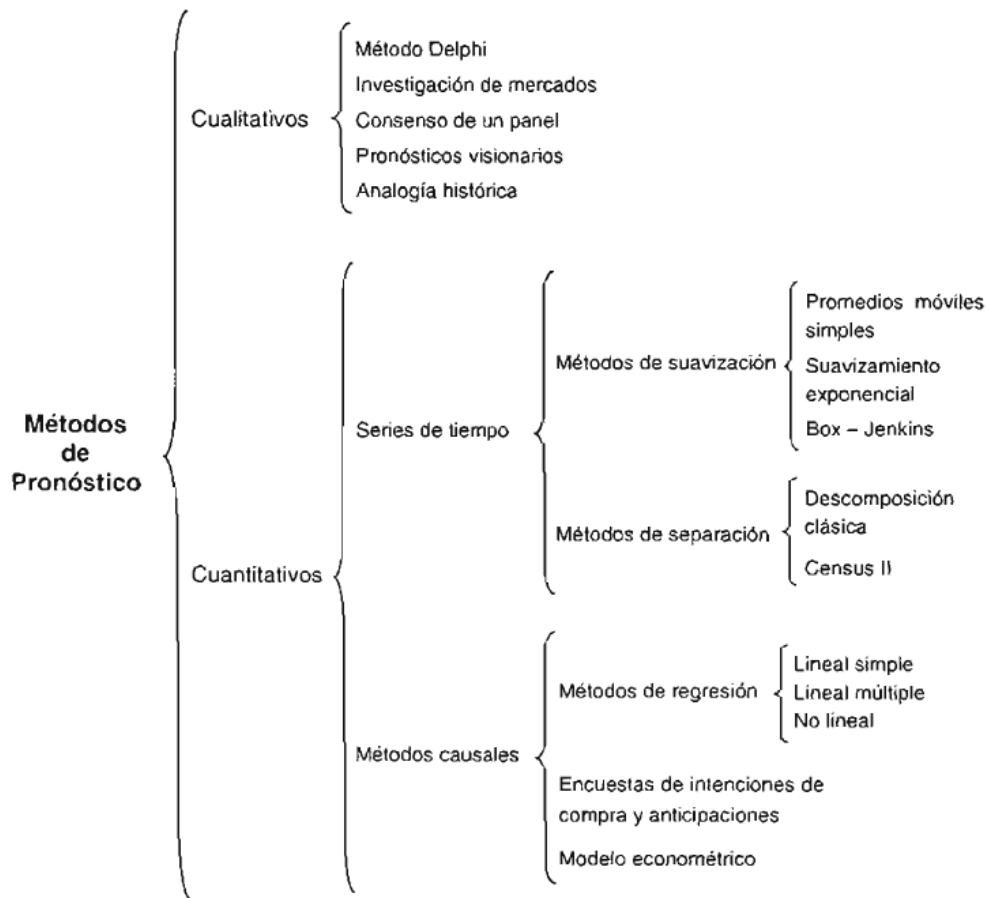


Ilustración 9. Técnicas de Pronósticos. (Escalante Membrillo 2005, 15)

Las técnicas cualitativas (Escalante Membrillo 2005, 20) de pronósticos, se utilizan generalmente cuando la información con la que se cuenta es escasa o cuando es de procedencia poco confiable. Lo anterior, se puede ver reflejado en aquella situación, cuando el producto del que se va a realizar el pronóstico es nuevo, o no existe una metodología establecida para la toma de datos. Estas técnicas, dependen de la experiencia, conocimiento, juicio y apreciación del que está realizando el pronóstico. En ocasiones al usar todas estas cualidades, se transforman los indicadores cualitativos en cuantitativos, sin embargo, es imposible dar un juicio a cerca de la efectividad de estos métodos.

Las técnicas cuantitativas de pronósticos (Escalante Membrillo 2005, 18), dan en la mayoría de los casos información más concreta y fiable que los métodos cualitativos. Se busca disminuir la incertidumbre que estos últimos provocan. Los métodos cuantitativos son utilizados cuando la información que se posee es veraz y confiable, además se cuenta con los suficientes datos históricos medidos de manera uniforme.

Así mismo se parte de la hipótesis de que los factores externos en los que se dieron los datos históricos prevalecerán. En la realidad existe un cambio que ocurre generalmente de forma gradual y que se debe de tener en consideración al momento de realizar el pronóstico.

Dentro de los métodos cuantitativos de pronósticos encontramos a las series de tiempo, los métodos causales y la combinación de ambos (Datos de panel o Análisis de panel).

Series de tiempo

En un modelo de series de tiempo (Hernández G. y Téllez B. 2006) dos factores son importantes: la serie de datos que se va a pronosticar y el periodo de tiempo a utilizarse. Un modelo de series de tiempo supone siempre que algún patrón o combinación de patrones es recurrente a través del tiempo. De esta manera, al identificar y extrapolar dicho patrón, se pueden desarrollar pronósticos para periodos subsecuentes.

En estos modelos se identifican cuatro efectos principales, tendencia, estacionalidad, efecto cíclico y aleatoriedad.

Métodos Causales.

También conocidos como modelos explicativos, buscan “explicar” las variables que provocan cambios en el comportamiento de la variable de estudio. Este tipo de modelos no solo nos proporcionan un pronóstico, si no que al contar con las variables principales del proceso, se pueden detectar y plantear estrategias que permitan focalizar los recursos de la empresa en acciones precisas que den más beneficios. Con lo mencionado, inclusive es factible considerar estas variables para un estudio de mercado o de viabilidad para la expansión o entrada de un producto.

Métodos de regresión

Dichos métodos son empleados ampliamente debido a su relativa simpleza, además de su efectividad. Y se han posicionado como una de las herramientas más eficientes de su tipo.

La idea general es la de ajustar un conjunto de datos a una ecuación que permita describir el comportamiento de dichos datos de la manera más realista posible.

Es de interés señalar que el ajuste de un modelo de regresión no se limita a analizar la relación entre dos variables (regresión lineal simple), sino que se buscarán relaciones del tipo:

$$Y = f(x_0, x_1, \dots, x_n) + e$$

relacionando de manera aproximada los valores de Y con los que toma una función f determinada por las variables x_0, x_1, \dots, x_n . e es el error o la diferencia entre el valor real y el estimado.

En general se destacan cuatro tipos de modelo:

- Regresión lineal simple.
- Regresión lineal múltiple.
- Regresión no lineal.
- Modelos Logit y Probit

Regresión Lineal

La ecuación de regresión lineal⁵ es la fórmula algebraica de un hiperplano en R^n con la cual se determina un valor estimado para la variable dependiente. Dicha ecuación es de la forma:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n + \varepsilon_t$$

donde $b_0, b_1, b_2, \dots, b_n$ son parámetros desconocidos, El error ε_t para el tiempo t, es una variable aleatoria que explica por qué la estimación del modelo no se ajusta exactamente a los datos.

Las constantes $b_0, b_1, b_2, \dots, b_n$ son determinadas a partir de los datos mediante el método de mínimos cuadrados, dicho método busca minimizar la suma del cuadrado de las diferencias entre los valores observados y los pronosticados. El desarrollo de dicho tema se encuentra en el Anexo B.

Análisis de Panel

En general los datos que se pueden encontrar son de dos tipos, series de tiempo y de corte transversal. En los datos de series de tiempo son observables una o diversas variables a lo largo del tiempo, mientras que los datos de corte transversal son observaciones de diferentes variables en un solo periodo de tiempo. Sin embargo cuando hay datos que cumplen con ambas condiciones, estos datos reciben el nombre de datos de panel (Gujarati y Porter 2009, 591); combinación de datos de series de tiempo y de corte transversal; datos de micropanel; datos longitudinales, entre otros. En otras palabras estos datos son observaciones de corte transversal a lo largo del tiempo.

Un modelo lineal general para datos de panel permite al intercepto (término independiente) y a los coeficientes de las variables explicativas variar para cada individuo⁶ y para cada periodo de tiempo. Expresado generalmente como:

⁵ Se dice que una función $Y = f(X)$ es lineal en X si X aparece elevado a una potencia o índice 1 solamente, y dicha variable no está multiplicada ni dividida por alguna otra variable. Si Y sólo depende de X. Otra forma de plantear que Y se relaciona linealmente con X es que la tasa de cambio de Y respecto de X (es decir, la pendiente, o la derivada, de Y respecto de X, dY/dX) es independiente del valor de X (Gujarati y Porter 2009).

⁶ Se conoce como individuo a la entidad de la que se extrae información para su análisis, para el caso de esta tesis, los individuos son los aeropuertos mexicanos.

$$y_{it} = \alpha_{it} + x_{1it}\beta_{1it} + x_{2it}\beta_{2it} + \dots + x_{kit}\beta_{kit} + u_{it}; \quad i = 1, \dots, N; \quad t = 1, \dots, T;$$

O matricialmente:

$$y_{it} = \alpha_{it} + x'_{it}\beta_{it} + u_{it}; \quad i = 1, \dots, N; \quad t = 1, \dots, T;$$

Donde y_{it} es la variable dependiente, α_{it} es el intercepto o término lineal, x_{kit} es un vector de $k \times 1$ de variables independientes, β_{it} es un vector de $k \times 1$ de coeficientes de las variables independientes, u_{it} es el término de perturbación (error), i referencia al sujeto en un corte transversal y t referencia al periodo de tiempo.

Sin embargo este modelo es demasiado genérico (Cameron y Trivedi 2005, 698-699) y no es estimable debido a que presenta más parámetros para estimar que observaciones. Por lo que adicionalmente es necesario colocar restricciones a medida que varían α_{it} y β_{it} con i y t , además del comportamiento del término de error u_{it} .

Beneficios y limitaciones del Panel de Datos

Como beneficio podemos encontrar un conjunto de datos mucho mayor con más variabilidad y menos colinealidad entre las variables que es típico en datos de corte transversal y series de tiempo. Con datos adicionales y más informativos, se puede obtener estimaciones más fiables y evaluar modelos de comportamiento más sofisticados con supuestos menos restrictivos. Otra ventaja de los datos de panel es su habilidad para controlar la heterogeneidad individual. No controlar estos efectos específicos individuales no observados conduce a un sesgo en las estimaciones resultantes. Los datos de panel además, son capaces de identificar y estimar mejor los efectos que son simplemente no detectables en datos de corte transversal o series de tiempo puras. En particular, los datos de panel son capaces de estudiar mejor cuestiones complejas del comportamiento dinámico (Baltagi 1998, 2).

Una característica adicional de los datos de panel (Frees 2004) es su habilidad para lidiar con algunos tipos de variables omitidas. Sin embargo al mismo tiempo puede existir “attrition bias”⁷ lo que puede permitir que la estimación de resultados sufra por el sesgo en la selección de observaciones.

En general las limitaciones del análisis con datos de panel giran alrededor de dos grandes vertientes, el modelo en sí mismo y los datos (Baltagi 1998, 3). Se dan por una mala especificación del modelo; y por otro lado por la dificultad para obtener y registrar todas las observaciones de cada individuo a lo largo del tiempo. Es posible que se registre algún tipo de sesgo al seleccionar la muestra, lo que puede provocar sesgo en la inferencia esbozada a través de la muestra. Los datos de panel también son susceptibles a los errores de medición realizados en cualquier momento. Es muy probable que la recolección de datos esté asociada a altos costos de muestreo y seguimiento de los individuos. Al incrementar los periodos de tiempo estudiados, la disponibilidad de información puede disminuir debido a que algunos individuos ya no se encuentran disponibles para realizar el muestreo.

⁷ Se conoce como “Attrition bias” al sesgo que se puede originar debido a la exclusión u omisión de variables en la estimación.

Modelos de Panel de Datos

Modelo Agrupado (Pooled Model).

El modelo agrupado es el modelo más restrictivo ya que especifica coeficientes constantes para todos los individuos, la suposición habitual del análisis de corte transversal teniendo como modelo:

$$y_{it} = \alpha + x'_{it}\beta + u_{it} \quad (A)$$

Si el modelo es correctamente especificado y los regresores no están correlacionados con el error, entonces éste puede ser estimado consistentemente usando MCO agrupados (*Pooled OLS*). El término de error está probablemente correlacionado con el tiempo para un individuo dado, sin embargo, en ese caso los errores estándar usualmente reportados no deberían utilizarse pues podrían ser fuertemente sesgados.

Dicotomía de individuo y tiempo (Individual and time dummies).

Una simple variante del modelo anterior permite al término independiente variar a través de los individuos y del tiempo mientras los coeficientes de las variables dependientes no (Cameron y Trivedi 2005, 699-700). Por lo que se tiene:

$$y_{it} = \alpha_i + \gamma_t + x'_{it}\beta + u_{it} \quad (B)$$

Cuando solo se considera α_i recibe el nombre de modelo de una vía (*One-Way*), mientras que al considerar α_i, γ_t recibe el nombre de modelo de dos vías (*Two-Way*). Escrito de otra forma:

$$y_{it} = \sum_{j=1}^N \alpha_j d_{j,it} + \sum_{s=1}^T \gamma_s d_{s,it} + x'_{it}\beta$$

Donde las N variables dicótomas⁸ de individuo $d_{j,it}$ son igual a uno si $i = j$ e igual a cero en otro caso, las $(T-1)$ variables dicótomas de tiempo $d_{s,it}$ son igual a uno si $t = s$ y a cero en otro caso, y se asume que x_{it} no incluye un término independiente. (En caso de ser incluido, entonces una de las N variables dicótomas de individuo tiene que ser desechada).

Este modelo tiene $N + (T - 1) + \dim[x]$ ⁹ parámetros que pueden ser estimados consistentemente si tanto $N \rightarrow \infty$ y $T \rightarrow \infty$. En general nos enfocamos en paneles cortos

⁸ Una variable dicótoma es un recurso para clasificar datos en categorías mutuamente excluyentes, como masculino o femenino. No es absolutamente esencial que las variables adquieran los valores 0 y 1. El par $(0,1)$ puede transformarse en cualquier otro par mediante una función lineal tal que $Z = a + bD$ ($b \neq 0$), donde a y b son constantes y donde $D = 1$ o 0 . Cuando $D = 1$ se tiene $Z = a + b$, y cuando $D = 0$, se tiene $Z = a$. Así, el par $(0,1)$ se convierte en $(a, a + b)$. Por ejemplo, si $a = 1$ y $b = 2$, las variables dicótomas serán $(1, 3)$. Esta expresión muestra que las variables cualitativas o dicótomas no tienen una escala natural de medición. Esto se debe a que se describen como variables en escala nominal. (Gujarati y Porter 2009, 278)

⁹ $\dim[x]$ es la dimensión de la matriz x'_{it}

donde $N \rightarrow \infty$ pero T no. Entonces γ_t puede ser estimado consistentemente, así que las $(T - 1)$ variables dicótomas de tiempo son simplemente incorporadas a los regresores x_{it} . El reto entonces recae en estimar los parámetros β controlando los N términos independientes (interceptos) de individuo α_i .

Modelo de efectos fijos y efectos aleatorios (Fixed Effects and Random Effects Model).

El modelo de efectos individuales y específicos permite a cada unidad de corte transversal tener un término diferente de intercepto manteniendo a los coeficientes de las variables regresoras iguales, asumiendo:

$$y_{it} = \alpha_i + x'_{it}\beta + \varepsilon_{it} \quad (C)$$

Donde ε_{it} es el error aleatorio. Esta es la forma más sencilla de expresar al modelo (B) con cualquier variable dicótoma de tiempo incluida en los regresores x_{it} . Los interceptos α_i son variables aleatorias que capturan la heterogeneidad inobservada entre individuos.

Una variante del modelo (C) trata a α_i como variables aleatorias no observadas que están potencialmente correlacionadas con las observaciones x_{it} . Esta variante es llamada **Modelo de Efectos Fijos**. A pesar de tener interceptos diferentes para cada entidad, estos no varían con el tiempo, es decir, "fijos".

La otra variante del modelo (C) asume que los efectos aleatorios individuales α_i son variables aleatorias que están distribuidas independientemente de las regresoras. Este modelo es llamado **Modelo de Efectos Aleatorios**. Para completar la especificación del modelo, se asume que α_i están idénticamente e independientemente distribuidas con media cero y varianza σ_α^2 . Adicionalmente se asume que α_i son independientes de las variables de error aleatorias ε_{it} (Frees 2004, 67-68).

$$y_{it} = x'_{it}\beta + \alpha_i + \varepsilon_{it}$$

Pudiéndose reescribir como:

$$y_{it} = x'_{it}\beta + v_{it} \quad (C.1)$$

De este modo α_i se parametriza como un componente más del término de error, y pasa a ser una perturbación aleatoria adicional.

El modelo de efectos aleatorios identifica el parámetro poblacional que describe el nivel de heterogeneidad individual, infiere a través del modelo de efectos aleatorios perteneciente a la subyacente población de individuos. En contraste debido a que el modelo de efectos fijos no puede estimar los parámetros que describen el nivel de heterogeneidad individual, la inferencia a partir del modelo de efectos fijos es condicional a los efectos fijos en la propia muestra. Por lo que el modelo de efectos aleatorios es más eficiente y permite una gama más amplia de inferencia estadística (Baum 2006).

En otras palabras, a partir del modelo de efectos fijos, solo es posible hacer inferencia de los individuos pertenecientes a la muestra. Mientras que con el modelo de efectos aleatorios, es posible hacer estimaciones de toda la población de individuos.

Modelo de panel dinámico (Dynamic Model). (Frees 2004, 218-220)

Debido a que los datos varían con el tiempo así como en cortes transversales se tiene la oportunidad de modelar los patrones dinámicos o temporales en los datos. En lo consecuente un primer punto que se debe de cubrir para tener un buen análisis, es un rango de datos amplio. Posteriormente, escoger un modelo estadístico y los tipos de aproximaciones usadas para establecer las propiedades de los parámetros estimados.

En particular los modelos dinámicos de panel se caracterizan por incluir una variable explicatoria indexada al tiempo.

Un primer modelo sería introducir una variable dicótoma para indicar la presencia o ausencia de efectos en un periodo; es un caso similar del modelo (B) de dos vías (*Two-way*).

$$y_{it} = \alpha_i + \gamma_t + x'_{it}\beta + u_{it} \quad (B.1)$$

Aquí los parámetros γ_t son cantidades específicas del tiempo que no dependen de los individuos, mientras que en el modelo (B) se consideraba el caso en el que γ_t son variables fijas.

Un de las técnicas más populares identifica al cambio o diferencia de la variable de interés en periodos sucesivos obteniendo un modelo que introduce un parámetro autoregresivo.

$$\Delta y_{it} = \alpha_i + x'_{it}\beta + \varepsilon_{it}$$

$$y_{it} = \alpha_i + y_{i,t-1} + x'_{it}\beta + \varepsilon_{it}$$

Generalizando:

$$y_{it} = \alpha_i + \gamma y_{i,t-1} + x'_{it}\beta + \varepsilon_{it} \quad (D)$$

Donde γ es un parámetro a ser estimado. Si $\gamma = 0$, entonces se está en presencia de un modelo de una vía (*One-Way*). En este caso γ es una medida de la relación de $y_{i,t-1}$ sobre y_{it} .

Otra forma de formular un modelo autoregresivo es

$$y_{it} = \alpha_i + x'_{it}\beta + \varepsilon_{it}$$

Donde $\varepsilon_{it} = \delta\varepsilon_{i,t-1} + \mu_i$ (E), para este modelo la autoregresión¹⁰ se encuentra en el término de error y no en la respuesta como en el modelo (D).

Estimadores de Panel de Datos

Estimador de MCO agrupados (*Pooled OLS*)

El modelo *Pooled OLS* es obtenido al apilar los datos a través de i y t dentro de una gran regresión con NT observaciones y estimado por Mínimos Cuadrados Ordinarios (MCO):

$$y_{it} = \alpha + x'_{it}\beta + u_{it}; \quad i = 1, \dots, N; \quad t = 1, \dots, T;$$

Si $cov[u_{it}, x_{it}] = 0$ entonces con que se cumpla $N \rightarrow \infty$ o $T \rightarrow \infty$ es suficiente para consistencia en la estimación (Cameron y Trivedi 2005).

El estimador *Pooled OLS* es claramente consistente si el modelo (A) es apropiado y los regresores no están correlacionados con el término de error. Sin embargo la matriz de varianzas basada en los errores indexados al individuo, no es apropiada debido a que los errores dados por los individuos i están casi con toda seguridad correlacionados positivamente a través de t .

Este estimador es inconsistente si el modelo verdadero es el modelo de efectos fijos. Por lo que la regresión mediante *Pooled OLS* de y_{it} sobre x_{it} y un intercepto conduce a una inconsistencia del estimador de β si los efectos individuales α_i están correlacionados con los regresores x_{it} , puesto que tal relación implica que el término combinado de error está correlacionado con los regresores.

Estimador “entre individuos” (*Between Estimator*).

A diferencia del estimador *Pooled OLS* que usa la variación de corte transversal y de tiempo para estimar β . El estimador “*Between*” (Cameron y Trivedi 2005, 703) generalmente utilizado para paneles cortos, solo toma en cuenta la variación de corte transversal. Para este estimador se parte del modelo (C) con la diferencia que se promedian los valores de cada variable quedando el modelo como:

$$\bar{y}_i = \bar{x}_i\beta + \alpha_i + \bar{\varepsilon}_i; \quad i = 1, \dots, N; \quad (F)$$

Donde

$$\bar{y}_i = T^{-1} \sum_t y_{it}; \quad \bar{\varepsilon}_i = T^{-1} \sum_t \varepsilon_{it}; \quad \bar{x}_i = T^{-1} \sum_t x_{it};$$

¹⁰ Los modelos autoregresivos pueden describirse como aquéllos en los que una variable o conjunto de variables se explican, al menos en parte, en función de los valores pasados de esa misma variable o conjunto de variables (Alvarez De Toledo, y otros 2006).

El estimador “*Between*” es el *Pooled OLS* de la regresión de \bar{y}_i a través de un intercepto y \bar{x}_i . Usa la variación entre diferentes individuos y es un análogo de la regresión para corte transversal para el caso especial de $T = 1$.

Este modelo es consistente si los regresores \bar{x}_i son independientes del término de error compuesto en (F). Es aplicable para los modelos de coeficientes constantes (A) y para el de efectos aleatorios. El estimador “*Between*” es inconsistente para el modelo de efectos fijos.

Estimador “dentro de los individuos” (*Within Estimator*)

El modelo busca eliminar al efecto del intercepto, pero a su vez explora la relación entre las desviaciones de las observaciones de la variable de respuesta con su respectivo valor medio y las desviaciones de las variables independientes con su respectiva media (Gujarati y Porter 2009, 599). De ésta manera se forma el modelo “*Within*”:

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)' \beta + (\varepsilon_{it} - \bar{\varepsilon}_i); \quad i = 1, \dots, N; \quad t = 1, \dots, T; \quad (G)$$

El estimador “*Within*” es el estimador *Pooled OLS* para (G). Se observa que este estimador es consistente para el modelo de efectos fijos mientras que los estimadores “*Pooled*” (A) y “*Between*” (F) no. El estimador “*Within*” es consistente para modelos *Pooled OLS*, Efectos Fijos y Efectos Aleatorios.

Entre sus limitaciones se encuentra que a pesar de que los coeficientes sean consistentes para el modelo *Pooled OLS*, son ineficientes¹¹ en comparación con los resultados de la estimación *Pooled OLS*. Una mayor limitación del modelo es que los coeficientes de los regresores que son invariantes con el tiempo (Cameron y Trivedi 2005, 704) no son reconocidos en el modelo “*Within*”.

Primeras Diferencias (*First-Differences Estimator*)

El estimador de primeras diferencias así como el estimador “*Within*”, explora características especiales del panel (Cameron y Trivedi 2005). En general mide la asociación que existe entre los cambios específicos del individuo en un periodo con respecto al anterior, tanto para variables independientes como la variable dependiente. Quedando el modelo como:

$$y_{it} - y_{i,t-1} = (x_{it} - x_{i,t-1})' \beta + (\varepsilon_{it} - \varepsilon_{i,t-1}); \quad i = 1, \dots, N; \quad t = 2, \dots, T; \quad (H)$$

El estimador de primeras diferencias es el estimador *Pooled OLS* para el modelo (H). Es consistente para el modelo de efectos fijos y no admite coeficientes de regresores invariantes con el tiempo. Este estimador es menos eficiente que el estimador “*Within*” para $T > 2$ si ε_{it} es indexada al individuo.

¹¹ La razón de esto es que, cuando las variables se expresan como desviaciones de sus valores medios, la variación en estos valores corregidos por la media es mucho más pequeña que la variación de los valores originales de las variables. En ese caso, la variación del término de perturbación ε_{it} puede ser relativamente grande, lo que produce errores estándar más altos de los coeficientes estimados. (Gujarati y Porter 2009, 600).

Estimador de Efectos Aleatorios (Random Effects Estimator)

Para este estimador se parte del modelo (C.1), para el cual el estimador *Pooled OLS* es consistente sin embargo un estimador de Mínimos Cuadrados Generalizados¹² (MCG) agrupados es más eficiente. Este estimador consiste en una regresión de desviaciones parciales de las y_{it} sobre las mismas desviaciones parciales de las x_{it} (Cobacho Tornel y Bosch Mossi 2004, 5). Concretamente:

$$y_{it} - \hat{\lambda}\bar{y}_i = (x_{it} - \hat{\lambda}\bar{x}_i)\beta + v_{it}; \quad i = 1, \dots, N; \quad t = 1, \dots, T;$$

Donde

$$v_{it} = (1 - \hat{\lambda})\alpha_i + (\varepsilon_{it} - \hat{\lambda}\bar{\varepsilon}_i)$$

Y $\hat{\lambda}$ es consistente para:

$$\lambda = 1 - \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + T\sigma_\alpha^2}}$$

Basado en estimaciones consistentes para σ_ε^2 y σ_α^2 . Hay que destacar que para $\hat{\lambda} = 0$ corresponde a un estimador *Pooled OLS*, $\hat{\lambda} = 1$ corresponde al estimador “*Within*”, y $\hat{\lambda} \rightarrow 1$ para $T \rightarrow \infty$.

Tabla 3. Resumen de los estimadores de panel de datos y su comportamiento con los modelos principales al estimar β (Cameron y Trivedi 2005, 699)

Estimador de β	Modelo Asumido		
	Pooled	Efectos Fijos	Efectos Aleatorios
Pooled OLS	Consistente	Inconsistente	Consistente
Between	Consistente	Inconsistente	Consistente
Within (o Efectos Fijos)	Consistente	Consistente	Consistente
Primeras Diferencias	Consistente	Consistente	Consistente
Efectos Aleatorios	Consistente	Inconsistente	Consistente

Métodos para la evaluación de pronósticos

Error estándar de estimación

El error estándar de estimación es una medida que califica la variabilidad o dispersión de los datos pronosticados, con respecto a los datos reales.

¹²Se aplica cuando las varianzas de las observaciones no son constantes (heterocedasticidad), o cuando existe algún grado de correlación entre las observaciones. Para más información revisar: Cameron, A. Colin, y Pravin K. Trivedi. *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press, 2005 cap. 4 p. 81-85. Frees, Edward W. . *Longitudinal and Panel Data: Analysis and Applications for the Social Sciences*. Cambridge: Cambridge University Press, 2004 p. 388, 421, 447.

Para el cálculo del error estándar de la estimación, es necesario calcular la diferencia entre los valores reales y los valores ajustados del pronóstico, esta diferencia entregará al llamado vector residual. Si $e_i = (e_1, \dots, e_n)$, los e_i (para $i = 1, \dots, n$) son los residuales, entonces la Suma del Cuadrado de los Residuales (SCR) es:

$$SCR = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2$$

Con la SCR es posible calcular un primer estimador del error estándar σ^2 , dividiéndolo entre el tamaño de la muestra n (Bingham y Fry 2010, 66-67). Sin embargo este estimador es sesgado, para obtener un estimador insesgado¹³, se tiene que reemplazar n en el denominador por $n-1$ quedando de la siguiente manera:

$$\sigma^2 \cong \frac{1}{n-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

El problema de realizar esta estimación de σ^2 , es que cada vez que se estima un parámetro a partir de los datos, se pierde un grado de libertad¹⁴. Este es un problema unidimensional por lo que el grado de libertad decrece de n a $n-1$ (caso de la regresión lineal simple).

Para el caso general se tiene que estimar p parámetros $(\beta_1, \dots, \beta_p)$ en consecuencia se pierden p grados de libertad y se obtiene un estimador insesgado al tener que dividir, no solo por n como anteriormente, sino por $(n-p)$, quedando el estimador como:

$$\sigma^2 \cong \hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \frac{SCR}{n-p}$$

$$\sigma = \sqrt{\frac{SCR}{n-p}}$$

Coefficiente de Determinación y Correlación

Estos coeficientes pueden ser calculados para regresiones simples o múltiples pero con diferentes acepciones para cada una.

Para el caso de dos variables, el **coeficiente de determinación** (R^2), expresa la proporción de la variación total en la variable dependiente y explicada por un incremento en la variable explicativa x . Esta idea es extendida al caso general con más de dos

¹³ Se dice que un estimador es insesgado cuando la diferencia entre el valor esperado de este estimador y el valor numérico del parámetro que se estima es cero.

¹⁴ Los grados de libertad son una medida de la cantidad de información disponible en el cálculo. Para este caso los grados se calculan como $n-p$, donde n es el tamaño de la muestra y p el número de parámetros que se calcularán. (Bingham y Fry 2010, 67).

variables; así el **coeficiente de determinación múltiple** denotado por R^2 , indica la proporción de la variación de y explicada por las variables explicativas x_1, x_2, \dots, x_n conjuntamente.

$$R^2 = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$$R^2 = 1 - \frac{\text{Suma del Cuadrado de los Residuos (SCR)}}{\text{Suma de Cuadrados Total (SCT)}}$$

Donde $\sum_{i=1}^n (y_i - \bar{y})^2$, es la variación total de los valores reales de y respecto de su media muestral.

El valor del coeficiente de determinación lineal o múltiple se encuentra entre 0 y 1. Si es 1, la línea de regresión ajustada explica 100% de la variación en y ; si es 0, el modelo no explica nada de la variación en y . Lo más frecuente es que R^2 se encuentre entre estos valores.

El **coeficiente de correlación** (R) está definido como la raíz cuadrada del coeficiente de determinación. Para el caso de dos variables el coeficiente de correlación mide el grado de asociación lineal entre las variables. El caso general es el **coeficiente de correlación múltiple**, denotado por R , el cual es una medida del grado de asociación entre y , y todas las variables explicativas en conjunto.

El valor de R puede ser positivo o negativo, por otro lado R siempre se considera positivo. En la práctica, sin embargo, R tiene poca importancia; R^2 es el que tiene mayor significado. (Gujarati y Porter 2009, 197 y 198).

R² Ajustada

Una propiedad importante de R^2 es que es una función no decreciente del número de variables explicativas o de regresoras presentes en el modelo; a medida que aumenta el número de regresoras, R^2 aumenta casi invariablemente y nunca disminuye. Planteado de otra forma, una variable adicional X no reduce R^2 . Debido a esta propiedad, al aumentar las variables X en el modelo R^2 aumentará a pesar de que alguna variable X aporte o no información al modelo (Gujarati y Porter 2009, 201).

De esa forma se ideó una forma de considerar el número de variables que están involucradas en la regresión al calcular R^2 y así se ideó R^2 ajustada (\bar{R}^2) definida como:

$$\bar{R}^2 = 1 - \frac{(SCR)(n-1)}{(SCT)(n-k)}$$

donde k es el número de parámetros en el modelo incluyendo el término independiente. El término ajustado significa ajustado por los grados de libertad asociados a las sumas de cuadrados (SCR y SCT) (Gujarati y Porter 2009, 202).

Forecast Value Added Analysis (FVA).

El Análisis del valor agregado al pronóstico o FVA (Gilliland 2013) por sus siglas en inglés es una métrica para la evaluación del rendimiento de cada paso y cada participante en el proceso de pronóstico. Se define como el cambio en una métrica de rendimiento del pronóstico (cualquier métrica puede ser usada, como el MAPE¹⁵, bias, etc.) que puede ser atribuida a cada paso y participante particular en el proceso de pronóstico.

En otras palabras el FVA expresa el resultado de hacer algo contra no haber hecho nada. Pudiendo ser tanto negativo como positivo. Comunicando si los esfuerzos realizados están agregando valor al hacer el pronóstico mejor, o en cambio, si se están haciendo las cosas peor.

Para el cálculo del FVA solo es necesario utilizar el llamado “naïve forecast”¹⁶ como base de comparación. Utilizando este pronóstico se calcula una métrica, y se medirá el mejoramiento de dicha métrica, a cada paso del pronóstico (una nueva variable integrada al pronóstico) con el objetivo de medir el comportamiento en el rendimiento del pronóstico.

La importancia del FVA radica en que ayuda a identificar los desperdicios o gastos innecesarios en el proceso de pronóstico. Al identificar y eliminar las actividades que no agregan valor (aquellas actividades que no hacen al pronóstico mejor) se puede mejorar el rendimiento del pronóstico. En términos económicos, asegura que cada recurso que es invertido en el proceso de pronóstico es de ayuda.

Econometría

En términos generales econometría significa “medición económica” (Gujarati y Porter 2009, 1); sin embargo esta disciplina va mucho más allá de una simple medición.

“La econometría se define como la ciencia social en la cual las herramientas de la teoría económica, las matemáticas y la inferencia estadística se aplican al análisis de los fenómenos económicos” (Goldberger 1964, 1).

“El arte del econométrico consiste en encontrar un conjunto de supuestos lo bastante específicos y realistas para que le permitan aprovechar de la mejor manera los datos con que cuenta” (Malinvaud 1966, 514).

Tradicionalmente, la teoría económica se basa en la realización de supuestos, o formulación de hipótesis de naturaleza un tanto cualitativa. Sin embargo, es por esta razón que la teoría por sí sola no postula alguna medida cuantitativa de la relación que pudiese existir entre dos o más variables.

¹⁵ Media Absoluta del Porcentaje de Error (MAPE), mide el tamaño del error global de la estimación en términos porcentuales. Es el promedio del valor absoluto del error porcentual de cada observación:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|x_i - x_i^*|}{x_i} \cdot 100$$

¹⁶ De acuerdo al glosario del Institute of Business Forecasting (IBF), a naïve forecasting es un pronóstico fácil de calcular, que requiere el mínimo de esfuerzo y manipulación para prepararlo. Algunos ejemplos son los métodos de último dato, promedio móvil o suavizado exponencial.

Desde el punto de vista económico la conducta humana es sumamente compleja, por lo que es prácticamente imposible explicar esa conducta por medio de uno o varios modelos, ya que estos revelan únicamente las relaciones existentes entre un reducido número de fenómenos económicos; pero a pesar de esto, un modelo inteligentemente planeado, puede representar en forma clara esas relaciones, que expresadas de otra forma serían muy confusas (García Pérez 1960).

Es por esto que la principal preocupación de la econometría es la verificación empírica de la teoría económica. El proceso que lleva la econometría (Gujarati y Porter 2009, 3) tradicionalmente es:

1. Planteamiento de la teoría o de la hipótesis.
2. Especificación del modelo matemático de la teoría.
3. Especificación del modelo econométrico o estadístico de la teoría.
4. Obtención de datos.
5. Estimación de los parámetros del modelo econométrico.
6. Pruebas de hipótesis.
7. Pronóstico o predicción.
8. Utilización del modelo para fines de control o de políticas

Una característica importante es que el modelo matemático es simple y presenta una solución determinista, mientras que el modelo econométrico, toma en cuenta la aleatoriedad e involucra un término de error a la ecuación matemática,

En general para los modelos econométricos se utilizan modelos de regresión de alguna índole. Por lo que el primer paso es la construcción del modelo econométrico, lo que conlleva a una selección minuciosa de las variables involucradas.

Capítulo 3.

Identificación de variables



KMV Tupolev Tu-154B-2 (RA-85457) Moscú - Domodedovo (Rusia)
Foto: Alexander Nikolaev - RuSpotters Team. Fuente: www.aviationcorner.net

Hasta el momento se ha hablado de la turbosina y su entorno en el país, además de las bases teóricas necesarias para la construcción y análisis de un modelo de pronósticos. En este capítulo servirá como introducción de la construcción del modelo de demanda de turbosina planteado como objetivo de esta tesis.

Cabe señalar que identificar las variables que impactan en el comportamiento de la demanda es uno de los objetivos establecidos, esto exige que el tipo de análisis que se deba realizar sea causal. Este capítulo, toca precisamente dicho tema, pues habla de la naturaleza de las variables, que pueden ser económicas, relacionados a los usuarios o a la infraestructura. Se parte de la idea de que todas las variables expuestas son viables a ser incluírlas en el estudio. Por lo anterior, son abordadas algunas dificultades de obtención y medición de variables, que permitirán establecer criterios que ayuden a reducir las variables que se introducirán al modelo. Sobre el final se hace un recuento de las variables que serán descartadas junto con una breve explicación.

Selección de Variables

La selección de variables es quizá el punto más importante al realizar un pronóstico mediante métodos causales, pues como el nombre del método lo indica, se trata de identificar las variables que impactan en la variable que se quiere pronosticar.

A diferencia de fenómenos físicos como el comportamiento de los gases o el movimiento de los cuerpos, que se rigen por leyes físicas (que indican las variables que afectan su comportamiento), en el caso de la demanda de turbosina no existe alguna regla o ley que explique las variables que la afectan.

La selección de variables se hará partiendo de la hipótesis que la demanda de turbosina se comporta como un sistema. La importancia de la analogía de este fenómeno con un sistema parte de la idea de identificar aquellos factores, causas, eventos o variables cuya ocurrencia individual o conjunta, influyen en la ocurrencia y comportamiento de la demanda de manera significativa; así como el tipo y lógica de la relación que mantienen entre sí (Ordaz Aguilar 2005, 15-16).

A partir de este enfoque sistémico es posible disgregar las áreas que forman al sistema. Existen muchos enfoques que se han utilizado para el pronóstico de ésta demanda y de otros combustibles, hablo del enfoque de ingeniería, estadístico y econométrico, (Wang 1999). Para esta tesis, las áreas que se estudiarán con el fin de encontrar a las variables que impactan sobre la demanda de combustibles son: econométricas, de infraestructura, de los usuarios y del medio de desarrollo; en las variables de usuarios queda inmerso el enfoque de ingeniería al considerar variables como el número de servicios realizados. Además se consideran variables como la estacionalidad del mercado y las condiciones sociopolíticas.

VARIABLES ECONOMÉTRICAS

Como se comentó en el Capítulo 1, la econometría utiliza hipótesis de naturaleza empírica o cualitativa para establecer una relación entre variables. Sin embargo estas hipótesis no

brindan una medida cuantificable que permita establecer el grado de relación y causalidad entre las variables. La teoría económica por sí sola no determina relaciones numéricas para las relaciones entre variables, pero sí indica la forma de la relación.

Si tomamos a la Teoría Keynesiana del consumo¹⁷ en consideración, será necesario utilizar indicadores económicos que nos permitan tener una medida acerca del ingreso que se percibe. Así mismo existen otros indicadores relativos al estado de crecimiento de la economía y su comportamiento, tales como el Producto Interno Bruto o el Índice de Precios al Consumidor. Este último permite establecer una medida al poder adquisitivo que se tiene en una región.

Tradicionalmente el precio de los productos juega un papel importante en la demanda de los mismos. La teoría económica clásica determina una relación entre la oferta, la demanda y el precio del mercado (Stonier y Hague 1970). Se dice que el precio de un bien determina el comportamiento del mercado, pues al bajar el precio, la demanda aumentará y la oferta decaerá; mientras que efectos contrarios se observarán si el precio incrementara.

De acuerdo con estas consideraciones se tomarán en cuenta para su análisis algunos indicadores económicos como el INPC y datos del mercado de hidrocarburos como lo son los precios del petróleo y turbosina.

PIB

El Producto Interno Bruto (PIB) es la medida estándar del valor final de los bienes y servicios producidos en una nación durante un periodo, menos el valor de las importaciones. Mientras que el PIB es el indicador más importante de las actividades económicas, éste no es una buena medida de las condiciones de vida y bienestar de las sociedades y solo una medida limitada de los estándares materiales de vida de las personas (OECD 2015).

El PIB tradicionalmente era considerado por excelencia como un indicador del nivel de desarrollo de una nación, sin embargo solo es una medida del nivel de riqueza de una nación; pero no de la distribución de la misma o del nivel de vida de las sociedades que conforman dicha nación.

Existen típicamente dos clasificaciones para el PIB. La primera de ellas se conoce como PIB nominal, y representa el valor financiero que se obtiene al sumar los servicios y bienes producidos por un país a valores corrientes del año en que fueron fabricados o generados. La segunda de ellas es el PIB real, que es una valoración monetaria absoluta a valores constantes, esto quiere decir que se utilizan los precios sin tener en cuenta la inflación ocurrida en el periodo respectivo.

¹⁷ La teoría keynesiana del consumo expone principalmente que el consumo depende fundamentalmente del ingreso y que si bien existen otros determinantes, éstos no poseen relevancia suficiente. (Gujarati y Porter 2009, 3).

“los hombres están dispuestos, por regla general y en promedio, a aumentar su consumo a medida que su ingreso crece, aunque no tanto como el crecimiento de su ingreso” (Keynes 1992, 90).

El PIB real presenta claras ventajas sobre el PIB nominal, pues cuando los precios suben por acción de la inflación, el valor del PIB también se ve elevado sin que en realidad haya un incremento efectivo en la producción nacional. Estas distorsiones causadas por la inflación son eliminadas en el PIB real.

Al ser el indicador por excelencia de riqueza en un país, sería lógico pensar que un incremento en él, representaría un incremento en el consumo de turbosina. Mayor riqueza en el país es indicador de que las actividades económicas como el transporte de pasajeros y/o mercancía por vía aérea se llevan a cabo de manera adecuada, propiciando el consumo de turbosina.

INPC

El Índice Nacional de Precios al Consumidor (INPC) es un indicador económico que se emplea recurrentemente, cuya finalidad es la de medir a través del tiempo la variación de los precios de una canasta fija de bienes y servicios representativa del consumo de los hogares. El INPC es el instrumento estadístico por medio del cual se mide el fenómeno económico que se conoce como inflación¹⁸ (Banco de México 2015). Dicho indicador se publica quincenalmente.

La precisión con la que se calcula el INPC, depende principalmente de dos factores que son su representatividad y su comparabilidad a través del tiempo. La representatividad se logra en la medida que la canasta de bienes y servicios que se utiliza para dar seguimiento a los precios, refleje los patrones de consumo de los hogares. La comparabilidad temporal requiere que la medición en la evolución de los precios se realice respecto a un punto o periodo base en el tiempo, es decir, cuál ha sido el cambio en el nivel de precios respecto de los que se tenían en una fecha, mes o año determinado. Lo anterior se logra en la medida en que todos los elementos utilizados en la medición del índice, excepto los precios, permanezcan estables a lo largo del tiempo, de manera que las variaciones que registre el INPC sean sólo debidas a cambios en precios.

Así, el INPC se elabora dando seguimiento a los precios de una canasta de bienes y servicios representativa del consumo de los hogares en un momento dado del tiempo. En la determinación de la canasta de consumo se utiliza como principal fuente de información la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) que recaba el INEGI. A partir de dicha información se construye para el INPC una canasta de bienes y servicios que agrupa el total del gasto en consumo de los hogares urbanos en conceptos genéricos representativos, a cada uno de los cuales se le asigna una ponderación que corresponde al peso que tiene dentro del gasto total.

En adición a lo anterior, es necesario seleccionar una muestra representativa de productos o servicios específicos correspondientes a cada concepto genérico a cuyos precios se da seguimiento. La muestra de productos específicos se utiliza para construir un índice de precios a nivel de genérico. Los índices de precios de los genéricos se

¹⁸ Se entiende por inflación, el crecimiento continuo y generalizado de los precios de los bienes y servicios que se expenden en una economía. Actualmente la inflación en México, es calculada por el INEGI y controlada por el Banco de México.

agregan conforme a la fórmula de índice de precios de Laspeyres¹⁹ utilizando la estructura de ponderación derivada de la ENIGH (Banco de México 2011).

Tabla 4. Ponderaciones por grupo de objeto de gasto en por ciento. Las ponderaciones apuntadas se aplicarán a partir de la primera quincena de abril de 2013, y permanecerán constantes durante la vigencia del patrón de consumo de los hogares en la República Mexicana. (Diario Oficial de la Federación 2013)

Índice Nacional de Precios al Consumidor	100.00000
Alimentos, bebidas y tabaco	23.14657
Ropa, calzado y accesorios	5.65823
Vivienda	28.48845
Muebles, aparatos y enseres domésticos	4.40139
Salud y cuidado personal	7.28882
Transporte	13.57985
Educación y esparcimiento	9.69586
Otros servicios	7.74084

Los grupos de consumo anteriormente señalados están compuestos por 283 conceptos de consumo genérico.

La importancia de la inclusión de este indicador en un modelo de pronóstico de la demanda para la turbosina, radica en el impacto que representa tener un mejor nivel de bienestar para el consumo de dicho combustible. Un mayor bienestar social y económico puede representar un incremento en la demanda de viajes por vía aérea, lo que se traduce en un mayor consumo de combustible. Caso contrario, al tener un nivel de bienestar menor, los viajes de tipo aéreo se vuelven innecesarios o incosteables, provocando una baja en el consumo de combustible.

Precio del petróleo

El petróleo es el principal insumo para la elaboración de la turbosina, por lo que la relación que existe entre ambos es muy estrecha. El petróleo es un combustible muy importante para la economía de las diferentes naciones; en el caso de México cerca de una tercera parte del PIB es por la venta de crudo.

El precio del petróleo se encuentra en sus niveles más bajos desde 2009, lo que representa grandes pérdidas para las naciones productoras de crudo. Este panorama en la caída de los precios que se viene presentando desde junio de 2014, de acuerdo a la

¹⁹ La fórmula para el cálculo de del índice de Laspeyres es:

$$I_{t/0} = \sum_{i=1}^n \frac{p_{i0}q_{i0}}{\sum_{i=1}^n p_{i0}q_{i0}} (p_{it}/q_{i0})$$

Donde $I_{t/0}$ es el índice que mide la variación promedio ponderado de los precios del periodo (t) en relación con los del periodo 0. (p_{it}/q_{i0}) es el índice de precios relativos del i-ésimo producto en el periodo (t), mide la variación del precio del periodo (t) en relación con el periodo base (0). $\frac{p_{i0}q_{i0}}{\sum_{i=1}^n p_{i0}q_{i0}}$ es la ponderación del precio relativo del i-ésimo producto en el índice y se obtiene dividiendo el valor de las (q) unidades del i-ésimo producto en el periodo base (0) entre el valor total de los (n) productos del índice en el mismo periodo base (0).

Organización de Países Exportadores de Petróleo (OPEP) decaerá en una desaceleración en la producción y en consecuencia un freno a la economía global. (Reuters 2015). Lo anterior a pesar de que la OPEP aumentó su pronóstico en la demanda de crudo para el primer semestre de 2015.

Petróleos Mexicanos produce 2.41 millones de barriles al día. El petróleo se clasifica por su nivel de pureza, entre más limpio, se puede aprovechar mejor la extracción de líquidos como las gasolinas, de los cuales se obtienen una rentabilidad en el mercado.

El American Petroleum Institute estableció la unidad API como una escala que sirve para medir el nivel de pureza e impureza de este hidrocarburo. Así, la pureza del petróleo se mide en grados API.

De acuerdo con lo establecido por el instituto, entre menos grados API contenga un crudo es más pesado que el agua y se califica como extra pesado y entre más grados API contenga es más ligero.

La pureza del petróleo determina el precio, en el mercado es más valioso un barril de crudo más liviano. (Martínez Medina 2014)

En México existen 4 tipos de petróleo que componen la llamada Mezcla Mexicana, estos son (PMI Comercio Internacional 2015):

- a) **Maya:** Es un crudo pesado (21-22° API) y amargo (3.4-3.8% de azufre en peso) por lo que brinda menores rendimientos de gasolina y diesel en esquemas de refinación simples en comparación con crudos más ligeros.
- b) **Istmo:** Es un crudo medio (32-33° API) y amargo (1.8% de azufre en peso) con buenos rendimientos de gasolina y destilados intermedios (diesel y jet fuel/keroseno). El mayor valor económico de este crudo se obtiene en refinerías con unidades FCC (Fluid Catalitic Cracker).
- c) **Olmecca:** Es el más ligero de los crudos mexicanos, con una gravedad de 38-39° API y un contenido de azufre de 0.73% a 0.95% en peso. Buen productor de lubricantes y petroquímicos.
- d) **Altamira:** Es un crudo pesado, con una gravedad de 15.5° a 16.5° API y un contenido de azufre en el rango de 5.5% a 6.0% en peso. Brinda menores rendimientos de gasolina y diesel. Sus características físico-químicas lo hacen adecuado para la producción de asfalto.

Una caída en los precios del crudo, se vería reflejado en una caída en el precio de producción de la turbosina, lo que se traduciría en un combustible más económico para las aerolíneas.

Precio de la Turbosina

El precio de la turbosina a diferencia del precio de otros combustibles que se comercializan en México va de la mano con las tendencias internacionales, debido a que las aerolíneas tienen la oportunidad de comprar turbosina en el extranjero. Adicionalmente

si bien el precio del crudo impacta directamente en el costo de producción del combustible, no es el único factor que determina su precio.

En México la comercialización de turbosina así como su transporte y entrega, no la hace directamente PEMEX, de este proceso se encarga Aeropuertos y Servicios Auxiliares (ASA). Lo que lo convierte en un monopolio para este servicio, pues no existe otro intermediario entre las aerolíneas y PEMEX.

El precio con el que Pemex vende la turbosina a ASA se calcula cada martes siguiendo la siguiente fórmula (Estudio-501 Aeropuertos y Servicios Auxiliares 2007):

$$\text{Precio de Lista} = \left[\left(\left(\text{Precio de Referencia} + \text{Costo de Manejo} \right) \times \text{Tipo de Cambio} \right) + \text{Costo de Transporte} \right] - \text{Descuento} + \text{IVA}$$

en donde

- a. Precio de lista de turbosina centro embarcador. Vigencia a partir del día martes de cada semana y cada día primero.
- b. Precio de referencia
- c. Costo de manejo: determinado por Pemex
- d. Tipo de Cambio: promedio de las 15 cotizaciones anteriores al lunes previo a la publicación del precio. Fuente: Diario Oficial de la Federación.
- e. Descuento y Costo de transporte: de acuerdo al lugar de suministro.

Por lo anterior, no es de esperarse que el descuento que se le otorga a ASA no se transfiera a las aerolíneas, y que al costo final sea necesario agregar la utilidad que ASA recibirá por las transacciones realizadas.

El precio de la turbosina, juega un papel muy importante en su consumo, pues en muchas ocasiones, el combustible representa más de un 30% de los gastos de operación de una aerolínea (Mccartney 2012). Lo que a su vez puede impactar significativamente en los precios de los viajes y como consecuencia, en la demanda de vuelos y específicamente del combustible.

Variables de Infraestructura

Cuando se habla de infraestructura, se habla de todo el conjunto de medios, elementos, servicios e instalaciones que son necesarios para que se lleve a cabo alguna actividad en particular. A partir de ésta idea es fácil identificar que la infraestructura aeroportuaria es aquel conjunto de elementos que permiten llevar a cabo el transporte aéreo.

La infraestructura aeroportuaria es posible clasificarla en terrestre y aérea. Siendo la primera todas las construcciones como lo son aeródromos y aeropuertos, así como los servicios necesarios para que estos operen (como el abastecimiento de combustible, migración y aduanas, etc.). Mientras que la infraestructura aérea se refiere a las aeronaves utilizadas en el transporte,

Es en este sentido en que conocer el comportamiento de la infraestructura ayudará a pronosticar la demanda de combustible. Se parte del planteamiento que si se incrementa

el número de aeronaves en operación, también se incrementará la demanda de combustible, y viceversa. Así mismo la ampliación y construcción de nuevos aeropuertos, representará una variable que impactará en el comportamiento del mercado de transporte aéreo. Pues con ésta última acción se abre el panorama al crecimiento de la flota aérea.

Nuevo Aeropuerto de la Ciudad de México.

Es un proyecto que plantea poner a la vanguardia a México y posicionarlo como un punto de entrada al continente y al comercio internacional. El Nuevo Aeropuerto Internacional de la Ciudad de México será el tercer aeropuerto más importante del mundo (Proyecto del Nuevo Aeropuerto de la Ciudad de Mexico 2014) y se tiene planeado que tenga una capacidad anual en su primera etapa de 50 millones de pasajeros y de 120 millones en su máxima expansión.



Ilustración 10. Proyecto del Nuevo Aeropuerto internacional de la Ciudad de México (Redacción El Universal 2014)

La construcción de ese nuevo aeropuerto fue planteada debido a la próxima saturación del aeropuerto actual y a su incapacidad para incrementar su aforo de operación. Actualmente el Aeropuerto Internacional de la Ciudad de México está alcanzado su máximo técnico de operaciones (394 mil operaciones al año), lo que se traduce en 32 millones de pasajeros al año, presentando un crecimiento aproximado desde 2010 de más del 3% en sus operaciones (Proyecto del Nuevo Aeropuerto de la Ciudad de Mexico 2014).

El nuevo aeropuerto será un parte aguas para el crecimiento económico del país, al tratarse de la mayor puerta para el intercambio comercial con el mundo. Por ser una nueva oportunidad de crecimiento, es de esperarse que se siga una tendencia positiva en el crecimiento de las operaciones aéreas para México.

Si bien ha existido mucha controversia alrededor de esta obra de infraestructura, su construcción afectará las condiciones económicas sobre las que se rige actualmente la Ciudad de México y el país entero. De esta forma se busca darle competitividad al país y a las relaciones comerciales con otras naciones, además de servir como una fuerte base que permita el desarrollo económico.

Un incremento en las operaciones, claramente impactará en la demanda de combustible necesario para el funcionamiento de las aeronaves. De esta manera, es como un proyecto de construcción de infraestructura puede impactar directamente en la demanda de turbosina en el país.

Matrícula de aeronaves.

Así como las instalaciones aeroportuarias son vitales para la operación del transporte aéreo, las aeronaves disponibles para realizar los servicios juegan un papel igual de importante.

Conocer la tendencia que lleva la adquisición de aviones por parte de las aerolíneas, es importante, porque esto da un panorama de la tendencia que sigue el mercado, lo que significativamente impactará en el consumo de turbosina.

La adquisición de aviones por parte de aerolíneas extranjeras será un buen indicador de las tendencias internacionales, mientras que las adquisiciones de aerolíneas nacionales nos revelaran el comportamiento del mercado nacional.

No basta con tomar en cuenta solo las aerolíneas que están en operación, ya que a lo largo de los años al igual que se han integrado nuevas aerolíneas como fue el caso de Interjet o Volaris, también han desaparecido aerolíneas tan importantes como Mexicana de Aviación, que llegó a ser la cuarta aerolínea más importante del mundo y que terminó operaciones en agosto de 2010 (Aerolíneas Mexicanas 2014).

Variables Relacionadas a los Usuarios

Tomando en cuenta que para efectos de ésta tesis solo se tomarán en cuenta Aeropuertos pertenecientes al Sistema Aeroportuario Mexicano (únicamente aeropuertos de carácter civil), no hay que pensar de más al momento de preguntarse la razón de ser de las aerolíneas o de los aviones, eso es simple: se trata de transportar. Como cualquier actividad económica, las empresas encargadas de esto buscan una remuneración en forma de ingresos para las mismas. Para lograrlo, es necesario conocer las necesidades de sus clientes,

Primordialmente la industria aérea da dos tipos de servicio; el transporte de pasajeros y el de carga (cantidad de bienes, llamados mercancías, que se transportan con el objetivo de obtener algún beneficio), Si bien el transporte de pasajeros estrictamente involucra un transporte de carga (equipaje, suministros, etc.), este último es secundario con respecto al pasajero en sí. Se considera transporte de carga a aquel que no incluye pasajeros.

Operaciones

El número de operaciones registradas en un periodo, es considerado como uno de los indicadores de actividad más importantes para las aerolíneas. Ya que evidencia la cantidad de vuelos que se realizaron durante un periodo. Este indicador conjunta a las operaciones de transporte regular de pasajeros y de transporte regular de fletamento. Para efectos de la demanda de turbosina, es un indicador que muestra la cantidad de veces que fue utilizada en el periodo, en otras palabras, es la frecuencia con la que se utiliza turbosina en el transporte aéreo.

Pasajeros

El transporte de pasajeros es quizá el mercado que domina al transporte aéreo y con el que la mayoría de las personas se siente más familiarizado. La evolución tecnológica ha permitido transportar más pasajeros en un solo viaje. A mediados de los años setenta uno de los íconos de la compañía Airbus el A300 era puesto al servicio de las aerolíneas, con capacidad de 200 a 300 pasajeros mientras que para 1995, el Boeing 777 era capaz de transportar más de 300 pasajeros y para 2007 fue introducido el Airbus 380 con una capacidad teórica de más de 800 pasajeros, lo que lo convierte en el avión comercial más grande del mundo. (Airbus 2015). Aunque por otro lado, aviones más grandes representan mayor consumo de combustible en el viaje.

La cantidad de personas que son transportadas, es una cantidad que va muy ligada a las operaciones realizadas y a la capacidad de los aviones utilizados. Ésta última varía considerablemente de acuerdo a la aerolínea y al equipo utilizado para el viaje.

Carga

El transporte de carga o fletamento, es un servicio con el que no todas las personas se sienten familiarizados, pero que se ha posicionado con el tiempo. Debido a su alto costo con respecto a los transportes terrestre y marítimo, el transporte aéreo había sido relegado. A pesar de que solo el 1% de toda la carga es transportada por vía aérea, el valor de ésta mercancía puede llegar a representar hasta una tercera parte de la carga total (Ruiz Olmedo 2012).

El desarrollo que ha tenido el transporte de carga aérea en los últimos años ha sido positivo y alentador. A nivel global el transporte de carga aérea presentó en 2014 un crecimiento del 4.5% con respecto 2013, lo que es un aceleración del 1.3% comparando 2013 con 2012 (IATA 2015). Este crecimiento no se dio de igual forma en todo el mundo, siendo la región de Asia y el Pacífico la de mayor crecimiento, mientras que Latinoamérica se dio un crecimiento inferior al 1%.

A diferencia de lo que pasa a nivel global, en México, la carga aérea transportada aumentó 11.9% comparada con el mes de enero en 2014, según reportó la Dirección General de Aeronáutica Civil en su informe mensual (Guerrero 2015).

Capacidad Hotelera

Uno de los fenómenos que impacta en la cantidad de viajes que se realizan durante cierto tiempo, es la temporada vacacional. Es en estas fechas cuando una mayor cantidad de personas aprovechan para viajar. Los viajes se incrementan considerablemente a lugares

con atractivos turísticos como lo son playas, sitios históricos, entre otros. La razón por la que las personas se desplazan a puntos específicos en la República Mexicana o fuera de ella, es por que ofrecen algún atractivo o sitios de interés.

Además de las vacaciones, visitar a alguien y la búsqueda de nuevas experiencias al visitar nuevos lugares, los viajes pueden ser por negocios. Los viajes ejecutivos son una constante en las aerolíneas a lo largo de todo el año y representan una valiosa fuente de ingresos.

Una de las principales características derivada de la gran afluencia de personas que se ven durante todo el año, es la capacidad hotelera. Ya sea por vacaciones a lugares como Cancún, Los Cabos o Acapulco, o de negocios a lugares como Monterrey o Guadalajara, poseer una fuerte infraestructura hotelera da la facilidad para que el usuario tenga la comodidad de desplazarse y permanecer ahí.

El crecimiento hotelero en muchos casos ha propiciado al turismo, permitiendo un desarrollo en la zona y en sus vías de comunicación.

Crecimiento poblacional

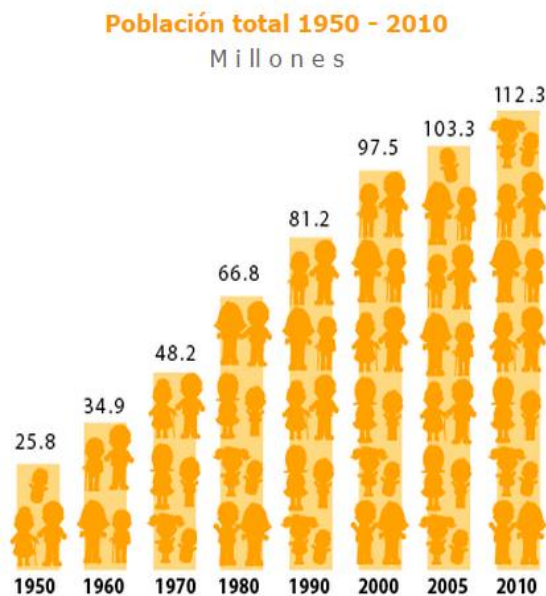


Ilustración 11. Datos INEGI (Cuentame... población 2014)

El crecimiento poblacional como factor de desarrollo económico, es uno de los más importantes, pues se trata de la materia prima que cada economía necesita. Típicamente, cada cambio económico que sucede va ligado en gran medida a los cambios que sufre la población en sí misma,

En México, así como en otros países, el aumento de la población fomenta el desarrollo y la inversión en infraestructura, en educación, salud, y un sin fin de proyectos. De alguna manera el crecimiento poblacional es un detonante económico, que provoca la inyección de capital e inversión.

Desde ese punto de vista, el crecimiento poblacional, es un indicador de las posibles necesidades que se deberán cumplir en el futuro. Esto detonará en propuestas de proyectos para satisfacer las necesidades de una población cambiante y creciente.

Específicamente para el caso de nuestro interés, un aumento de la población significa un aumento en el mercado al que las aerolíneas pueden llegar. Un mercado más grande aumenta las posibilidades de compra de un pasaje aéreo o la adquisición de un servicio para el transporte de carga, incrementando así el consumo de combustible al pasar de los años.

Variables del Medio de Desarrollo

La construcción de un modelo que explique el consumo es un proyecto ambicioso. Tradicionalmente se utilizaba una teoría que permitiera explicar el fenómeno y con base en él se calculaban los parámetros necesarios. Sin embargo en la actualidad es importante identificar que ninguna teoría económica se ajusta exactamente al comportamiento de la economía, lo que implica un reto aún más grande al momento de hacer un pronóstico.

Los econométricos algunas veces están dispuestos a tomar un enfoque ateórico, esto es, a desarrollar modelos sin una teoría implícita. Aun ante la ausencia de una teoría satisfactoria, si un número de variables económicas ha seguido alguna relación aritmética particular entre sí en el pasado, pueden seguir haciéndolo en el futuro. A condición de que la información utilizada sea lo suficientemente confiable, por lo general es preferible utilizar técnicas econométricas para descubrir y confirmar (o rechazar) tales relaciones, en vez de afirmar que esas relaciones existen, o de fijarse en estadísticas para ver qué tan estrecha es la relación entre dos variables. (Price 1998, 4-5)

Sin embargo la aseveración de que la relación que cumplen las variables se conserve, solo es válida cuando el medio en que se desarrolla presenta condiciones similares a lo largo del tiempo. En otras palabras las condiciones externas a las que es sometido el modelo se mantienen prácticamente constantes.

Para el caso de estudio de ésta tesis es importante identificar que se conceptualiza una economía estable, en donde las reformas políticas y económicas no impactarán de manera radical en el comportamiento del mercado de turbosina.

Guerras

La guerra es definida como el conflicto armado entre un grupo de individuos, este conflicto puede tener diferentes magnitudes, impactando de diferente manera. Ejemplos de este tipo de manifestaciones se tiene a lo largo de toda la historia del hombre y sus repercusiones han sido diversas. En la actualidad los conflictos bélicos han pasado a ser una fuente de ingresos económicos para los países productores y comercializadores de armamento. Sin embargo algo que no ha cambiado a lo largo de la historia, es el impacto negativo que tiene en la región geográfica donde se desarrolla el conflicto.

En el caso de México el último gran conflicto armado que involucró a toda la nación fue la Revolución Mexicana. *La revolución hizo sentir su impacto inmediato de desorden y caos en los escenarios de la guerra, ahí donde tenían lugar hechos de armas o donde algún ejército imponía temporalmente su dominio. Pero, por su misma naturaleza, este impacto no se extendió simultáneamente a todo el país, y sus efectos no fueron particularmente duraderos. Esto último puede explicarse porque la guerra civil no impuso una destrucción severa del aparato productivo: los revolucionarios descubrieron tempranamente que era más redituable someter las actividades productivas a la lógica y las necesidades de la guerra (cobrando impuestos forzosos, administrando las empresas privadas o confiscando parte de su producción) que acabar con ellas. Hubo, por supuesto, excepciones, como las protagonizadas por los villistas en las horas de ira de su dirigente. El efecto directo más notable de la revolución en el corto plazo fue la fragmentación del mercado nacional, resultado de dos fenómenos directamente asociados con el estado de guerra civil: la*

quiebra de sistema monetario nacional y la disrupción de la red ferroviaria. Ambos fenómenos provocaron situaciones de escasez, inflación rampante e incluso hambre en puntos localizados y momentos específicos del conflicto' Pero, en términos generales, la guerra no destruyó una gran cantidad de activos físicos, lo cual explica la pronta reactivación de la economía al final del conflicto (Kuntz Ficker 2002, 24).

Una guerra podría representar una desestabilidad económica y social tal, que las operaciones comerciales y de transporte de pasajeros se verían sesadas por completo, desplomando el consumo de combustible. En un escenario hipotético, el ejército sería el único consumidor de dicho producto.

Crisis económicas

Las crisis económicas son momentos del ciclo económico que se presentan de manera periódica en la economía, sin embargo, algunas de ellas pueden causar severos daños en la economía de las naciones. Como ejemplos claros de esto, se encuentra la gran crisis de 1929 (la Gran Depresión), que azotó a gran parte de las economías occidentales y que tuvo su mayor impacto en la Europa de la Posguerra. En la época actual, encontramos a la última gran crisis económica que impactó al mundo en 2009 y que tuvo su comienzo con la recesión de la mayor economía del mundo: EEUU.

México, al ser un país petrolero, vive momentos de tensión con la caída en los precios del petróleo en los últimos meses, pasando de \$92.25USD el barril de crudo en noviembre de 2014 a poco menos de \$40.00USD en enero de 2015 y con una leve recuperación en marzo promediando \$46.23USD el barril de la mezcla mexicana (Servicio Geológico Mexicano 2015).

Las crisis son un efecto de la desaceleración económica que vive el país; esto provoca una baja en la demanda de productos y servicios como lo son el transporte aéreo, ya sea de pasajeros o de carga. Si bien una caída en los precios de los hidrocarburos representaría una baja en las tarifas de las aerolíneas, el efecto global de la crisis económica afectaría la cantidad de operaciones aéreas demandadas

Consideraciones importantes

Multicolinealidad entre variables

La multicolinealidad no viola los supuestos básicos de la regresión. Se presentarán estimaciones consistentes e insesgadas y sus errores estándar se estimarán en la forma correcta. El único efecto de la multicolinealidad tiene que ver con la dificultad de obtener los coeficientes estimados con errores estándar pequeños. Sin embargo, se presenta el mismo problema al contar con un número reducido de observaciones o al tener variables independientes con varianzas pequeñas. Al respecto no hay una respuesta estadística. (Gujarati y Porter 2009, 326).

Cuando alguna de las variables del modelo se encuentra correlacionada con otra (presenta multicolinealidad), es difícil separar el efecto parcial de cada una de las variables sobre la variable dependiente (Esteban, y otros s.f.). La información muestral que incorpora una de estas variables es prácticamente la misma que el resto de las variables que presentan correlación con ellas.

De ésta forma las consecuencias de tener variables correlacionadas son:

- Varianzas y covarianzas grandes
- Intervalos de confianza más amplios
- Algunos coeficientes tienden a ser estadísticamente no significativos
- La medida global de bondad de ajuste R^2 puede ser muy alta.

Disponibilidad de información

La disponibilidad de información juega un papel muy importante al momento de realizar el modelado de un fenómeno. Esto se debe a que no siempre se dispone de la información en el momento o forma en que se necesita.

Un ejemplo de ello es el PIB, una variable que es evaluada trimestralmente, a diferencia de otras variables como el precio del petróleo, el precio de la turbosina y en particular del consumo de turbosina (la variable de interés) que es mensual. El caso de la variable de la capacidad hotelera (número de habitaciones promedio disponibles) es diferente, debido a que no se cuenta con la información de todos los años para todas las regiones ya que no se encontraron en la base de datos consultada (<http://www.datatur.sectur.gob.mx>), posiblemente porque no existió un registro en su momento.

Costos asociados

Como factor que se debe de considerar al momento de escoger a una variable para ingresarla dentro del modelo, es el costo asociado que tiene la obtención de dicha variable.

En algunas ocasiones, el costo de la obtención de una variable, es mayor al beneficio que representaría el de añadir dicha variable al modelo. Por lo que en esos casos se prescinde de la información y se prosigue con la formulación del modelo y la integración de nuevas variables.

Otro posible costo asociado va ligado al tiempo de espera en la obtención de información y las repercusiones que se tienen al no tomar una decisión a tiempo. Tomar una decisión con la mejor información disponible a tiempo se hace de vital importancia al seleccionar variables. Es necesario evaluar si vale la pena aguardar el tiempo necesario para la obtención de toda la información deseada.

Principio de parsimonia

La forma moderna del admirable precepto conocido en estadística como Principio de Parsimonia establece el uso de modelos simples de preferencia, de ésta manera se busca claridad y protección en contra del peligro de sobre-interpretar el modelo. Además sugiere la idea de penalizar modelos debido a su complejidad (Bingham y Fry 2010, 102-103).

De acuerdo con el principio parsimonia o de la navaja de Occam²⁰, conviene mantener el modelo de regresión lo más sencillo posible. Si se explica “sustancialmente” el

²⁰ “That descriptions be kept as simple as possible until proved inadequate”, The World of Mathematics, vol. 2, J.R. Newman, Simon & Schuster, Nueva York, 1956, p. 1247, or “Entities should not be multiplied beyond necessity”, Donald F. Morrison, Applied Linear Statistical Methods, Prentice Hall, Englewood Cliffs, Nueva Jersey, 1983, p. 58.

comportamiento de la variable dependiente con dos o tres variables explicativas, y si la teoría no resulta lo suficientemente fuerte para indicar que otras variables pueden incluirse, ¿Por qué se hace necesario introducir más variables? El Término de error ε_t , puede representar a todas las demás variables. Cabe destacar que, no se deben excluir variables pertinentes e importantes sólo por el hecho de mantener el modelo de regresión simple.

Variables Excluidas u Omitidas

Para la selección de variables se evaluaron todas las presentadas en este capítulo, cabe señalar que se buscó que las variables fueran de corte mensual, debido a que la información disponible para el consumo de turbosina en los aeropuertos tiene ésta característica. Además de utilizar información con un corte mayor (semestral, anual etc.) implicaría perder información y tener que retroceder muchos años para obtener una cantidad adecuada de observaciones y llevar a cabo el análisis, dificultando así su obtención (en caso de ser anualizadas las observaciones, se tendría que obtener información de por lo menos ¡30 años atrás!). De ésta manera las variables que no intervendrán en el modelo son:

- **Construcción de nuevos aeropuertos:** Debido a que este estudio de tesis tomará en cuenta que cada aeropuerto es un individuo diferente, por lo que no es viable incluir proyectos futuros como nuevos individuos al no conocer fehacientemente el posible impacto que tendrán. Además de no contar con la información de su impacto a través del tiempo.
- **Matrícula de aeronaves:** No se tomará en cuenta debido a que las capacidades de carga de cada avión varían de acuerdo al modelo de aeronave (incluso siendo el mismo modelo, varía su capacidad de acuerdo a la aerolínea) y la información disponible solo es la cantidad de aviones en operación de forma anual. Por otro lado, solo se cuenta con la información por aerolínea, imposibilitando hacer una estimación de aeronaves por aeropuerto.
- **Crecimiento poblacional:** Debido a que solo se tiene información de ésta variable en periodos de 5 años, se vuelve prácticamente imposible tener una aproximación mensual del valor de la población.
- **Guerras:** El país no se encuentra en periodo de guerra, siendo la Revolución Mexicana de 1910 reconocida como el último movimiento armado en el país. No se espera entrar en una guerra en los próximos años.
- **Crisis económicas:** el efecto de las crisis económicas se puede ver reflejado en la mayoría de las variables macroeconómicas como el INPC o PIB. a pesar de no tener una métrica establecida para cuantificar las crisis, el incluir una variable dicótoma que indicara la presencia o no de las mismas, podría ocasionar que se sobre especificara el modelo al incluir otras variables macroeconómicas.

Variables incluidas en el análisis

Las variables elegidas para el análisis que culminará con la formación de un modelo que explique la demanda de turbosina son las siguientes:

VARIABLES ECONOMÉTRICAS

- I. INPC (mensual).
- II. PIB (trimestral).
- III. Precio por barril de crudo de la mezcla mexicana en USD (promedio mensual).
- IV. Precio de venta por litro de turbosina en MXN (promedio mensual).

VARIABLES RELACIONADAS A LOS USUARIOS

- V. Número de operaciones en aeropuertos por aeropuerto (mensual).
- VI. Número de pasajeros transportados por aeropuerto (mensual).
- VII. Cantidad de carga transportada por aeropuerto en kg (mensual).
- VIII. Capacidad hotelera (número de habitaciones promedio disponibles) por destino turístico (promedio mensual).

Estas variables fueron elegidas por lo expuesto de cada una de ellas en este capítulo, adicionalmente cumplen con la facilidad de obtención y la periodicidad en la medición deseada para este análisis.

Es posible afirmar que las variables de Infraestructura y del Medio de Desarrollo que estuvieron propuestas en este trabajo, fueron más difíciles de cuantificar de una manera que pudieran ser integradas al modelo.

Capítulo 4.

construcción y evaluación



LAN Colombia. Cabina de Control
Fuente: aerolatinnews.com

En el presente capítulo se aborda la construcción del modelo explicativo del consumo de turbosina. Dicho modelo tendrá que ser de tipo causal para cumplir con el objetivo de identificar las variables que impactan en el consumo de combustible. Por lo tanto, se parte de la propuesta de un modelo preliminar utilizando todas las variables elegidas en el capítulo anterior; se prosigue explicando el proceso de depuración de variables, hasta llegar a un modelo final que explique el consumo.

¿Qué tipo de análisis utilizar?

Al ser **necesaria la construcción de un modelo** que nos **permita identificar las variables que afectan el consumo** de turbosina, y admita sacar conclusiones de cada aeropuerto, los datos para el análisis deben cumplir con ciertas características. Esto quiere decir que se deben identificar las posibles variables que construyan el modelo (lo que se realizó en el capítulo anterior de manera “teórica”), tener los **datos de estas variables para cada uno de los aeropuertos y además su evolución histórica** (esto último debido a que para inferir el comportamiento futuro, es necesario conocer el comportamiento pasado y actual de las variables).

De esta manera, **la naturaleza de los datos** necesarios para llevar a cabo este análisis **debe de corresponder** al de un **panel de datos** (descrito en el segundo capítulo).

Modelo Preliminar

A partir de esto, se decidió utilizar a los aeropuertos que tuvieran la información completa para todas sus variables (y así formar un panel de datos balanceado²¹). Cabe mencionar que a pesar de no contar con la información completa de la variable “Capacidad Hotelera” para el Aeropuerto Internacional de la Ciudad de México, se tomó la decisión de incluirlo y completar su información²² (ya que representa poco más de un tercio del consumo de turbosina a nivel nacional).

Se tomaron en cuenta a 43 de los más de 60 aeropuertos que pertenecen al Sistema Mexicano Aeroportuario, mismos que concentran más del 97% del consumo de turbosina en el país.

En la formación de este primer modelo causal, se decidió utilizar a la demanda de turbosina como variable dependiente y las variables PIB, INPC, Capacidad Hotelera, Operaciones, Pasajeros, Carga y precios de Turbosina y Petróleo como variables explicativas de la demanda. Lo anterior forma un modelo Lin-Lin (Lineal-Lineal); en donde el coeficiente de cada variable explicativa representa la variación proporcional que recibirá la demanda de turbosina por cada aumento o disminución unitaria en la variable.

²¹ Un panel de datos balanceado es aquel que tiene todos los datos de corte transversal y series de tiempo correspondientes. En oposición, los datos no balanceados, presentan omisión en alguno o más datos (los análisis para datos de panel incompletos son más complicados y el objetivo de ésta tesis no busca explorar esos métodos).

²² El AICM no contaba con la serie de datos del periodo 2007-2012 para la capacidad hotelera; por lo que se decidió completarla con la información disponible (periodo 2013- enero 2015) mediante una regresión lineal simple entre la variable regresora (capacidad hotelera) y el tiempo.

Sin embargo, previo a la conformación de este primer modelo, se estudió el comportamiento de las variables de precios, y se encontró que ambas series de datos presentaban un comportamiento inusual. Debido a que los datos de las variables son los mismos para todos los aeropuertos, se tomó la decisión de aplicarles una transformación logarítmica. Esto además de “aplanar los datos” disminuye su dispersión, además permitiría conocer de manera directa la elasticidad de la demanda de turbosina con respecto de los precios de turbosina y del petróleo (Mezcla Mexicana).

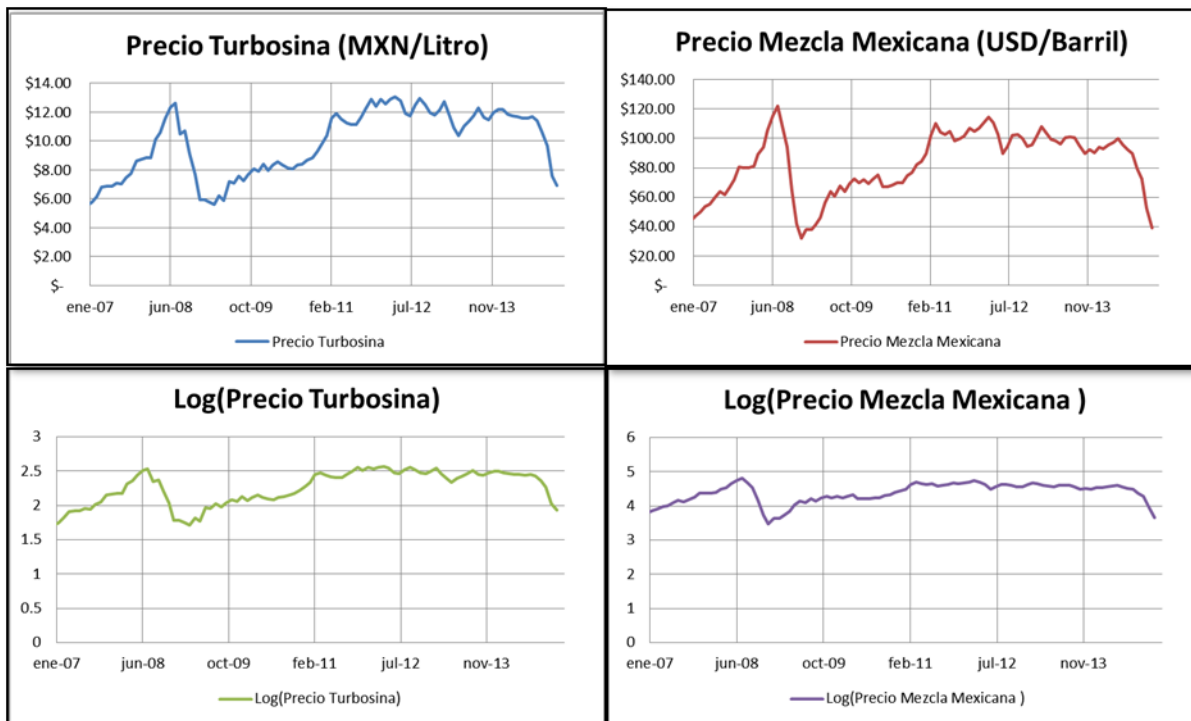


Ilustración 12. Variables de precios de Turbosina y precio de Petróleo (Mezcla Mexicana) con su respectiva transformación logarítmica, en donde se puede apreciar la disminución de dispersión en los datos.

No obstante que se le aplicó una transformación logarítmica a las variables de precios, se optó por mantener las demás variables sin transformación. Las variables que se mantuvieron igual para todos los aeropuertos (PIB e INPC) no presentaron el mismo comportamiento inusual que las variables de precios. Por otro lado, las demás variables, al tener diferentes comportamientos para cada aeropuerto, no fue posible identificar algún comportamiento atípico que justificara realizar un transformación logarítmica o de alguna otra índole.

De ésta manera el modelo de estudio preliminar quedó conformado como:

$$D_{turbosina} = \alpha_i + \beta_1 INPC + \beta_2 PIB + \beta_3 \text{Log}(P_{petroleo}) + \beta_4 \text{Log}(P_{turbosina}) + \beta_5 \text{Operaciones} + \beta_6 \text{Pasajeros} + \beta_7 \text{Carga} + \beta_8 \text{CapHotelera} + u_{it}$$

¿Cuál es la mejor estimación?

Una vez que se obtuvo al modelo que sería nuestro objeto de estudio, fue necesario identificar qué estimación (*Pooled OLS*, Efectos fijos o aleatorios) era la más adecuada y se acercaba a describir mejor el comportamiento de los aeropuertos.

Dicha selección se realizó llevando a cabo diferentes pruebas formuladas para discernir al modelo de estimación más adecuado. El proceso genérico se describe a continuación:

- Se calculan las tres estimaciones para su posterior evaluación.
- Al estimar el modelo de efectos fijos en Stata²³ también calcula una prueba restrictiva F, en la cual la hipótesis nula es $\alpha_1 = \alpha_2 = \dots = \alpha_i = 0$ (lo que es indicio de que no hay efectos fijos asociados a los individuos); aquí se contrastan los modelos *Pooled OLS* y de efectos fijos.
- Por otro lado se corre la prueba de Breusch-Pagan para efectos aleatorios (Montero G. 2011); consiste en realizar una regresión para la ecuación $y_{it} = x'_{it}\beta + u_i + e_{it}$; la hipótesis nula es $var(u_i) = 0$ con una χ^2 de contraste. Son comparados los estimadores de efectos aleatorios y *Pooled OLS*.
- Si se desestiman a los estimadores de efectos fijos y efectos aleatorios entre ambas pruebas, se toma al estimador *Pooled OLS* como verdadero.
- Finalmente en caso de que se desestime al modelo *Pooled OLS*, se realiza la prueba de Hausman²⁴ para seleccionar, ya sea el modelo de efectos fijos o el modelo de efectos aleatorios.

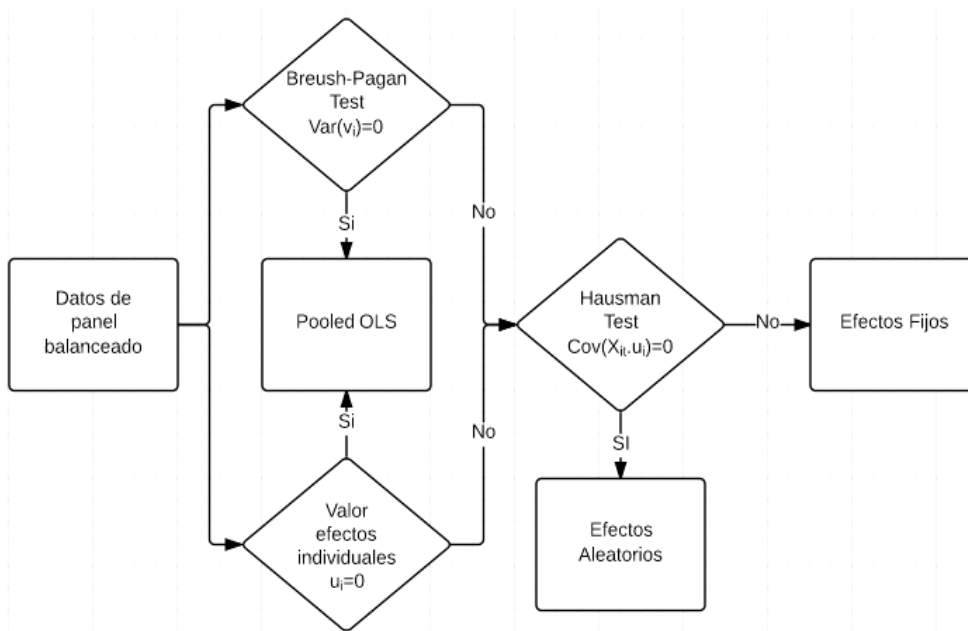


Ilustración 13. Algoritmo para realizar la selección del tipo de estimación que se tomará para realizar el pronóstico del gasto de turbosina. Elaboración propia.

²³ Stata es un software especializado para el análisis de datos y estadística aplicada.

²⁴ Esta prueba de especificación compara un estimador $\hat{\theta}_1$ que es conocido por ser consistente, con un estimador $\hat{\theta}_2$ que es eficiente bajo el supuesto que se está probando. La hipótesis nula es que el estimador $\hat{\theta}_2$ es en efecto un estimador eficiente (y consistente) del parámetro verdadero. Si este es el caso no debería haber diferencia sistemática entre ambos estimadores y se escoge al estimador eficiente como verdadero ($\hat{\theta}_2$). Si existiera una diferencia sistemática entre las estimaciones, el estimador escogido es $\hat{\theta}_1$. (Hausman specification test s.f.).

Para mayor referencia consultar anexo C.

Evaluación del modelo preliminar

Como nota antes de continuar:

Para el desarrollo de este modelo se había escogido como ventana de tiempo al periodo de enero de 2007 a enero de 2015, sin embargo se llegaron a resultados de heteroscedasticidad²⁵ presumiblemente ocasionados por los eventos atípicos de la crisis de 2008 (ver anexo D, “Modelo Efectos Fijos 2007-2015” y “Series de Errores Estándar de Estimación”). Lo anterior sirvió como antecedente para recortar la ventana de tiempo de este estudio al periodo 2010-2015.

Una vez que se realizó el procedimiento descrito en la sección anterior, donde se corrieron las diferentes estimaciones en Stata, se encontraron efectos que a primera vista parecieron inusuales. En las tres estimaciones, los valores β para las variables del PIB, precio del petróleo y carga salieron negativos.

Al mismo tiempo, a pesar de que la prueba de Breush-Pagan y la prueba F fueron concluyentes excluyendo al modelo *Pooled OLS*, el resultado de la prueba de Hausman rechazó la hipótesis nula con un valor p casi cero. Esto último sugería que el modelo de efectos fijos es preferible. Sin embargo arrojó una nota al principio del cálculo, cuya indicación era que la escala de al menos tres de los coeficientes es incompatible para realizar el test (Anexo D, “Pruebas de Hausman no concluyentes”).

Para aumentar la escala del coeficiente, basta con dividir todos los valores por una constante, adicionalmente para disminuir la escala de un coeficiente basta con multiplicar por una constante (Handout: Panel Data 2008).

Sin embargo a pesar de escalar las variables, la prueba de Hausman, no fue concluyente (Anexo D, “Pruebas de Hausman no concluyentes”).

Por otro lado, cada estimación arrojó diferentes variables estadísticamente no significativas. En el caso de la estimación *Pooled*, tanto el INPC como las variables de precios no fueron estadísticamente significantes (Tabla 21, columnas 1 y 5, Anexo D). Mientras que en el caso de la estimación de efectos fijos, todas las variables fueron estadísticamente significantes (Tabla 22, columnas 1 y 5, Anexo D). Finalmente, en la estimación de efectos aleatorios, tanto la carga y el INPC no fueron significantes para $\alpha = 1\%$, mientras que el INPC se volvía estadísticamente significativo para un nivel de significancia del 4.2% (Tabla 23, columnas 1 y 5, Anexo D).

Esto reveló la necesidad de analizar de manera más profunda a las variables involucradas, e identificar si efectivamente en la “práctica” eran las adecuadas para la conformación del modelo. Dicho análisis permitiría discernir qué variables era adecuado conservar y cuales sería necesario desechar del modelo.

²⁵ Heteroscedasticidad es el fenómeno cuando la varianza de los errores es diferente a lo largo de la regresión (no se cumple el supuesto de homoscedasticidad). Matemáticamente se expresa como:

$$\text{Var}(\varepsilon_{it}) \neq \sigma^2$$

Evaluación de las variables explicativas y construcción de un nuevo modelo.

De esta forma las primeras variables en ser analizadas fueron las de precios de turbosina y petróleo debido a la sospecha de colinealidad entre ellas.

Los logaritmos de ambas variables comparten un coeficiente de correlación R igual a 0.9193; mientras que para las variables sin alterar, el coeficiente de correlación R es de 0.9224. En cualquiera que sea el caso, ambos coeficientes son prácticamente iguales y muy cercanos a uno. Lo que confirma la presencia de colinealidad entre estas variables.

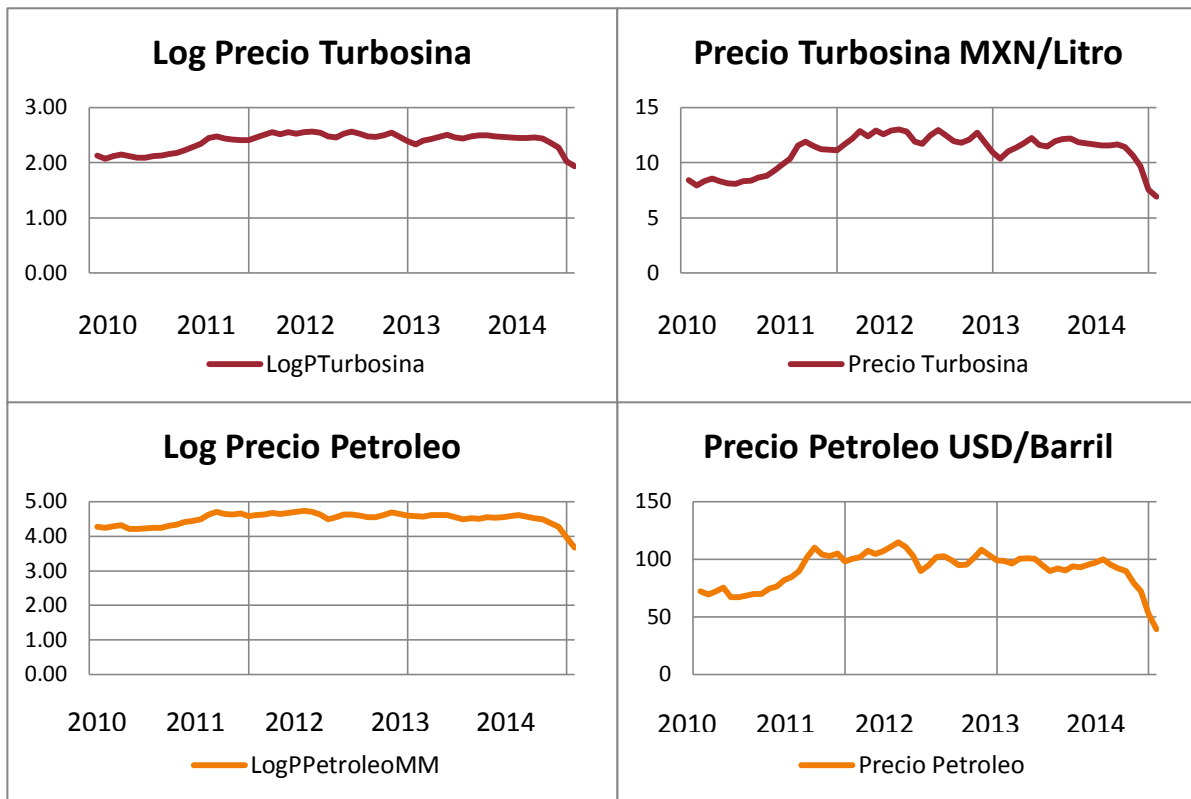


Ilustración 14. Graficas de las variables de precios expresadas en unidades monetarias y en logaritmos a partir de enero de 2010 hasta enero de 2015. 1 barril = 159 litros.

Se concluyó mantener la variable del precio de la turbosina por ser la variable que está más directamente vinculada con el producto que se quiere pronosticar, a diferencia del precio de la mezcla mexicana, que está vinculada indirectamente.

Por otro lado, al revisar la literatura, fue apreciable que las variables que son invariantes entre los individuos (INPC, PIB, Precio de la Turbosina), no contribuyen sustancialmente a encontrar la heterogeneidad inobservable entre los individuos estudiados (una ventaja del análisis de panel). Lo que significa que no aportan mucho valor al modelo.

“Es además posible incorporar parámetros invariantes entre los individuos, para contabilizar la variación temporal. Para la mayoría de datos de panel, esto no tiene

el mismo impacto de variabilidad que los parámetros que varían entre los individuos.

Debido a que la contabilización de heterogeneidad entre los individuos reduce la media del cuadrado de los errores y los errores estándar asociados con la estimación de parámetros, se tiene que la estimación de parámetros es más eficiente, que en el caso de tener variables que sean invariantes entre los individuos” (Frees 2004, 20-21).

En consecuencia se presentó la problemática de buscar alternativas que permitieran conservar estas variables, pero que ofrecieran mayor información e impacto al modelo.

Así se decidió sustituir al INPC nacional por el INPC local, repitiendo el INPC de la región más cercana para las localidades en las que no se halló un INPC registrado. En el caso del PIB se decidió utilizar el PIB estatal. No obstante, este último solo se encontró de manera anual y únicamente para el periodo 2010 - 2013, considerándose así, que no era conveniente introducirla al modelo (por el “ruido” que se introduciría al completar la serie con alguna estimación).

Paralelamente se evaluó la posibilidad de descartar al precio de la turbosina del modelo, esto debido a que es invariante entre los aeropuertos pudiendo no agregar más valor al modelo.

No obstante, ya que el precio de la turbosina es considerada una variable fundamental para la estimación (Wang 1999), fue necesario buscar una alternativa para mantenerla dentro del modelo. Así se decidió integrar el precio de la turbosina a la variable de demanda de turbosina; obteniendo así como variable dependiente al gasto de turbosina.

Esta integración se realizó de la siguiente manera:

$$Gasto\ de\ Turbosina_t = \frac{(Demanda\ de\ turbosina_t)(Precio\ de\ la\ turbosina_t)}{(1 + \sum_{k=1}^t Inflación_t)}$$

Donde t es el periodo de tiempo para el que se está haciendo el cálculo.

Una de las consideraciones que se puede observar en la fórmula anterior, es que el precio de la turbosina se encuentra deflactado (sin inflación), esto con el fin de eliminar el “ruido” que se podría introducir con la inflación. La inflación para el primer periodo, es la que se presentó en enero de 2010 (inflación = 1.09%).

De ésta forma el modelo que se obtuvo fue:

$$\begin{aligned} Gasto\ de\ turbosina \\ = \alpha_i + \beta_1 INPC + \beta_2 Operaciones + \beta_3 Pasajeros + \beta_4 Carga \\ + \beta_5 CapHotelera \end{aligned}$$

Antes de evaluar este modelo, se graficó el comportamiento del gasto para algunos aeropuertos, con el objetivo de determinar si era necesario realizar algún ajuste adicional al mismo (Ilustración 15).

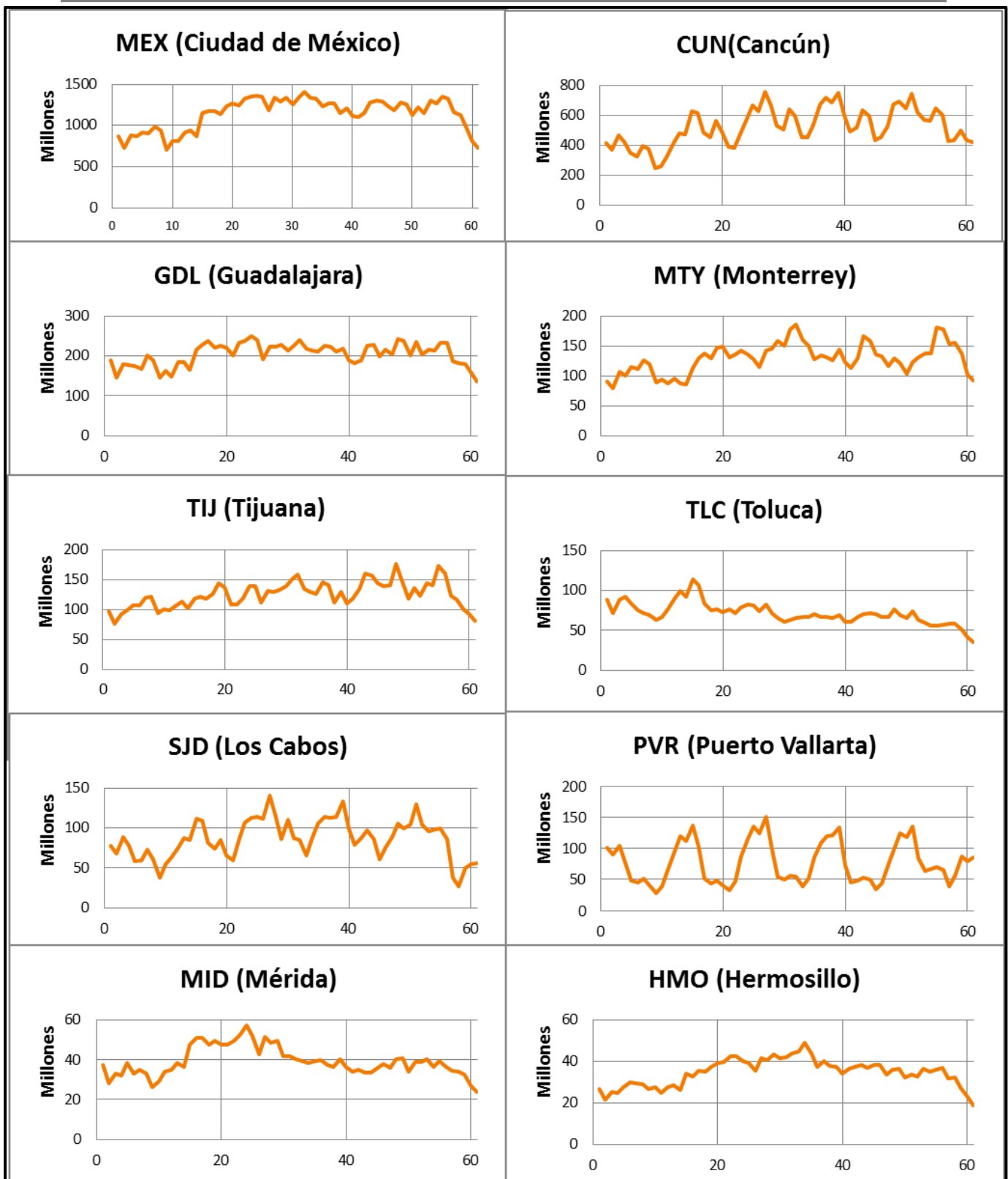


Ilustración 15. Comportamiento del gasto de turbosina (en millones de pesos) en los 10 aeropuertos más importantes de México, para el periodo de Enero 2010 a Enero 2015.

Lo que se observó fue que el gasto no se comportaba de la misma forma para los diferentes aeropuertos. Lugares como Cancún, Puerto Vallarta o los Cabos presentaron una clara tendencia cíclica, que se puede explicar fácilmente debido a que son parte de los principales destinos turísticos en el país, siendo altamente demandados en temporada vacacional. No obstante, aeropuertos como el de Toluca, Mérida o Hermosillo no tienen una forma clara, aunque se puede hablar de una tendencia decreciente para los últimos años. Otros aeropuertos como los de Tijuana, Ciudad de México y Guadalajara presentan tendencias crecientes en el gasto de turbosina, pero sin presentar un patrón claro.

Todo lo anterior, fue prueba de la heterogeneidad de comportamientos existente entre los aeropuertos, y que de alguna manera pudiera contribuir a que los intervalos en las estimaciones sean sobreestimados. Esto solo sería un síntoma de la presencia de heteroscedasticidad dentro del modelo.

La sobreestimación de los intervalos de confianza es uno de los principales problemas de la heteroscedasticidad en un modelo de estimación. Afortunadamente, esto no afecta la consistencia de la estimación. Por lo que el objeto de estudio no se pierde.

Con el fin de minimizar el riesgo de presentar heteroscedasticidad en la estimación, y de alguna manera homologar el comportamiento del gasto de turbosina, se optó por “aplanar” las series aplicando una transformación logarítmica al gasto de turbosina (Anexo D, “Series Logarítmicas del Gasto de Turbosina”). De ésta forma quedó el siguiente modelo:

$$\begin{aligned} \text{LogGasto de turbosina} \\ = \alpha_i + \beta_1 INPC + \beta_2 \text{Operaciones} + \beta_3 \text{Pasajeros} + \beta_4 \text{Carga} \\ + \beta_5 \text{CapHotelera} \end{aligned}$$

Los modelos de la forma $\log Y_t = \alpha + \beta'x + u$ son llamados modelos semilogarítmicos porque solo una variable (en este caso la variable dependiente) aparece en forma logarítmica. En la literatura, los modelos en los que únicamente la variable de respuesta es logarítmica son denominados modelos Log-Lin (Logarítmico-Lineal).

En este modelo, los coeficientes β miden la constante proporcional o el cambio relativo en Y_t para un cambio absoluto en el valor de las variables regresoras x (Gujarati y Porter 2009).

$$\beta = \frac{\text{Cambio relativo en la variable dependiente}}{\text{Cambio absoluto en la variable independiente (regresora)}}$$

En otras palabras los coeficientes β al ser multiplicados por 100 nos darán en términos porcentuales el cambio que tendría la variable dependiente por cada aumento o disminución unitaria en las variables regresoras. Esto es conocido como semielasticidad de Y con respecto de X .

Construcción del modelo final

Se utilizó el procedimiento descrito al principio de este capítulo (sección “¿Cuál es la mejor estimación?”), para identificar al mejor estimador. Los resultados entregaron que el mejor estimador era el modelo de efectos fijos (Anexo D, “Mejor estimador para el modelo: $\text{Log}(\text{Gasto de turbosina}) = \alpha_i + \beta_1 \text{INPC} + \beta_2 \text{Operaciones} + \beta_3 \text{Pasajeros} + \beta_4 \text{Carga} + \beta_5 \text{CapHotelera}$ ”).

Tabla 5. Estimación de Efectos fijos para el modelo
 $\text{Log}(\text{Gasto de turbosina}) = \alpha_i + \beta_1 \text{INPC} + \beta_2 \text{Operaciones} + \beta_3 \text{Pasajeros} + \beta_4 \text{Carga} + \beta_5 \text{CapHotelera}$

Fixed-effects (within) regression		Number of obs	=	2623
Group variable: id		Number of groups	=	43
R-sq: within	= 0.1052	Obs per group: min	=	61
between	= 0.0353	avg	=	61.0
overall	= 0.0381	max	=	61
corr(u_i, Xb)	= 0.0265	F(5,2575)	=	60.57
		Prob > F	=	0.0000

loggstotu~a	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
inpc	.0160781	.001252	12.84	0.000	.013623 .0185331
operaciones	.0001233	.0000197	6.27	0.000	.0000847 .0001618
pasajeros	-2.38e-07	1.66e-07	-1.43	0.152	-5.63e-07 8.72e-08
carga	6.61e-09	1.48e-08	0.45	0.654	-2.23e-08 3.56e-08
caphotelera	-.0000616	.0000252	-2.45	0.014	-.0001109 -.0000123
_cons	14.84184	.1563871	94.90	0.000	14.53519 15.1485

sigma_u	1.6276762
sigma_e	.32740787
rho	.96111195 (fraction of variance due to u_i)

F test that all u_i=0:	F(42, 2575) = 580.87	Prob > F = 0.0000
------------------------	----------------------	-------------------

Sin embargo, las variables de Pasajeros y Carga presentaron ciertas peculiaridades, mismas que sugerían un estudio a mayor profundidad de las mismas. En dicha estimación, tanto los coeficientes de pasajeros y carga son no significativos (Tabla 5, columnas 1 y 5). A pesar de ello, ¿esto es suficiente para excluirlas del modelo?; fue así que se decidió sacar la correlación global de las variables.

Tabla 6. Coeficientes de correlación entre las variables de operaciones, pasajeros y carga.

	operac~s	pasaje~s	carga
operaciones	1.0000		
pasajeros	0.9632	1.0000	
carga	0.9408	0.8902	1.0000

Como se puede observar, la correlación de la variable Operaciones con Pasajeros y Carga es muy alta, por otro lado la correlación entre las variables de Carga y Pasajeros a pesar de ser elevada, no es tan grande en comparación con las otras. De esta manera se confirmó el supuesto de correlación entre variables.

Lo anterior abre la interrogación acerca de, ¿qué variables se deben incluir en el modelo? Ya que al hacer pruebas con diferentes combinaciones de las tres variables para ser

incluidas en el modelo (Anexo D, “Combinaciones de modelos (variables Operaciones, Carga y CapHotelera)”), se obtiene que al ser incluidas dos variables, una de ellas es estadísticamente significativa, mientras que la otra no. Por otro lado a pesar de que los coeficientes sean casi cero (confirmado por la no significancia estadística), éstos al estar multiplicados por grandes cantidades pasan a tener un impacto sustancial en el pronóstico.

La variable Operaciones fue descartada por:

- Tratarse de una variable que engloba las operaciones de pasajeros y de transporte de carga, cuando este análisis podría hacerse de manera directa al incorporar las variables de pasajeros y carga,
- La correlación que existe entre esta variable y las otras dos es la más alta.

La variable Pasajeros, fue conservada por:

- Presentar la correlación más baja en conjunto con la variable carga.
- Representar al sector más importante de la industria aeronáutica en México (el transporte de pasajeros).
- Permitir estudiar el impacto individual del sector de transporte de pasajeros en la demanda de turbosina.

La variable Carga, fue conservada por:

- Su correlación con la variable Operaciones es menor que la correlación entre la variable Operaciones con la variable Pasajeros
- Es posible caer en el error tipo II; que es aceptar la hipótesis nula cuando esta es falsa.
- Representar una parte del mercado de la industria aeronáutica nacional que ha tenido un crecimiento importante en los últimos años²⁶

Incluir a las variables de pasajeros y carga permite determinar por separado cual es el impacto que tienen ambos sectores en el consumo de turbosina. Mientras que al incluir la variable operaciones, se estaría haciendo un análisis global del comportamiento del mercado y se podría perder información.

Por lo que el modelo final quedó como:

LogGasto de turbosina

$$= \alpha_i + \beta_1 INPC + \beta_2 Pasajeros + \beta_3 Carga + \beta_4 CapHotelera$$

Evaluación del modelo final

Utilizando el modelo formulado en la sección anterior, se llevó a cabo la identificación del mejor estimador para el modelo, cuyo primer paso fue validar o desestimar al estimador *Pooled OLS* como se describe en las siguientes líneas:

²⁶ “De acuerdo con cifras registradas por la Dirección General de Aviación Civil (DGAC) de la Secretaría de Comunicaciones y Transportes (SCT), durante el primer semestre de 2015 la carga total transportada en México fue de 321 mil 916 toneladas, 7.81% más que las 298 mil 606 toneladas de 2014}”. (García Morales 2015)

Tabla 7. Estimación Pooled OLS en Stata

Source	SS	df	MS			
Model	3405.07052	4	851.267631	Number of obs =	2623	
Residual	3933.18058	2618	1.5023608	F(4, 2618) =	566.62	
Total	7338.2511	2622	2.79872277	Prob > F =	0.0000	
				R-squared =	0.4640	
				Adj R-squared =	0.4632	
				Root MSE =	1.2257	

loggastotu~a	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
inpc	.0032354	.0040258	0.80	0.422	-.0046586	.0111295
pasajeros	2.69e-06	1.81e-07	14.91	0.000	2.34e-06	3.05e-06
carga	-9.26e-08	1.01e-08	-9.14	0.000	-1.12e-07	-7.27e-08
caphotelera	.0000536	7.24e-06	7.40	0.000	.0000394	.0000678
_cons	15.49526	.4243094	36.52	0.000	14.66324	16.32727

En la prueba F que realiza la estimación de efectos fijos cuya hipótesis nula es $\alpha_i = 0$, el valor de p es cero, por lo que H_0 se rechaza, lo que indica que algunos interceptos son significativos y no se puede usar un modelo *Pooled OLS* (que asume un intercepto común a todos los individuos). Dicha prueba desestima al modelo *Pooled OLS*.

Tabla 8. Estimación de efectos fijos en Stata

Fixed-effects (within) regression	Number of obs =	2623
Group variable: id	Number of groups =	43
R-sq: within = 0.0916	Obs per group: min =	61
between = 0.3102	avg =	61.0
overall = 0.2402	max =	61
corr(u_i, Xb) = -0.6191	F(4, 2576) =	64.92
	Prob > F =	0.0000

loggastotu~a	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
inpc	.0159191	.001261	12.62	0.000	.0134464	.0183919
pasajeros	6.21e-07	9.42e-08	6.59	0.000	4.36e-07	8.06e-07
carga	7.25e-09	1.49e-08	0.49	0.626	-2.19e-08	3.64e-08
caphotelera	-.000064	.0000253	-2.53	0.012	-.0001137	-.0000143
_cons	15.09507	.1522054	99.18	0.000	14.79661	15.39353

sigma_u	1.8306069	
sigma_e	.3298351	
rho	.96855667	(fraction of variance due to u_i)

F test that all u_i=0: F(42, 2576) = 799.46 Prob > F = 0.0000

Tabla 9. Estimación de efectos aleatorios en Stata

```

Random-effects GLS regression           Number of obs   =   2623
Group variable: id                     Number of groups =    43

R-sq:  within = 0.0873                 Obs per group:  min =    61
        between = 0.4403                avg =   61.0
        overall = 0.4199                max =    61

Random effects u_i ~ Gaussian          Wald chi2(4)    =   268.44
corr(u_i, X) = 0 (assumed)            Prob > chi2     =    0.0000
theta = .96582493
    
```

loggastotu~a	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
inpc	.0139121	.0011954	11.64	0.000	.0115692	.0162549
pasajeros	5.78e-07	9.38e-08	6.16	0.000	3.94e-07	7.62e-07
carga	1.91e-08	1.43e-08	1.34	0.181	-8.88e-09	4.71e-08
caphoteler	.0000211	.000018	1.17	0.243	-.0000143	.0000564
_cons	14.76713	.2328347	63.42	0.000	14.31079	15.22348
sigma_u	1.2350053					
sigma_e	.3298351					
rho	.93342152	(fraction of variance due to u_i)				

La prueba de Breush-Pagan que contrasta al modelo de efectos aleatorios con el modelo *Pooled OLS*, tiene los mismos resultados que la prueba anterior, ya que desestima al modelo *Pooled OLS*. El valor de p es cero por lo que se rechaza la hipótesis nula.

Tabla 10. Prueba de Breush-Pagan calculada en Stata, que desestima al modelo *Pooled OLS*

Breusch and Pagan Lagrangian multiplier test for random effects

```

loggastoturbosina[id,t] = Xb + u[id] + e[id,t]

Estimated results:
          |          Var          sd = sqrt(Var)
-----|-----
loggast~a | 2.798723  1.672938
e         | .1087912  .3298351
u         | 1.525238  1.235005

Test:   Var(u) = 0
        chi2(1) = 65116.22
        Prob > chi2 = 0.0000
    
```

Con ambos resultados se procedió a descartar al modelo *Pooled OLS* como mejor estimador y se procedió a revisar la prueba de Hausman para seleccionar entre efectos fijos o aleatorios.

La hipótesis nula dice que no hay diferencia entre ambos estimadores y en caso de ser aceptada se escoge al estimador de efectos aleatorios (en este caso a pesar de que ambos estimadores son consistentes, el estimador de efectos aleatorios es más eficiente).

La hipótesis alternativa indica que si hay diferencia entre ambas estimaciones y que la estimación de efectos aleatorios es inconsistente. Por lo anterior, en caso de ser aceptada esta hipótesis, se escogería al estimador de efectos fijos como verdadero

Tabla 11. Prueba de Hausman con los coeficientes sin escalar

Note: the rank of the differenced variance matrix (2) does not equal the number of coefficients being tested (4); be sure this is what you expect, or there may be problems computing the test. Examine the output of your estimators for anything unexpected and possibly consider scaling your variables so that the coefficients are on a similar scale.

	Coefficients		(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
	(b) fixed	(B) random		
inpc	.0159191	.0139121	.0020071	.0004165
pasajeros	6.21e-07	5.78e-07	4.28e-08	1.21e-08
carga	7.25e-09	1.91e-08	-1.19e-08	4.33e-09
caphotelera	-.000064	.0000211	-.000085	.0000179

b = consistent under Ho and Ha; obtained from xtreg
 B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

$$\text{chi2}(2) = (b-B)'[(V_b-V_B)^{-1}](b-B)$$

$$= 24.67$$

Prob>chi2 = 0.0000

La Prueba de Hausman rechazó la hipótesis nula (de que no hay diferencia entre los estimadores), por lo que el modelo más apropiado es el de efectos fijos. Sin embargo, al principio hubo una nota que indicó que no todos los coeficientes del modelo fueron utilizados para realizar la prueba, debido a que la escala de los coeficientes varía mucho. Como se puede observar el coeficiente de INPC es muy grande comparado con los demás; caso contrario al de la carga que es un coeficiente muy pequeño.

Por esta razón se decidió reducir al coeficiente del INPC, de esta forma la variable INPC fue multiplicada por la constante 100.

Tabla 12. Prueba de Hausman con los coeficientes escalados calculado en Stata

	Coefficients		(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
	(b) fixed	(B) random		
inpcm100	.0001592	.0001391	.0000201	4.16e-06
pasajeros	6.21e-07	5.78e-07	4.28e-08	1.21e-08
carga	7.25e-09	1.91e-08	-1.19e-08	4.33e-09
caphotelera	-.000064	.0000211	-.000085	.0000179

b = consistent under Ho and Ha; obtained from xtreg
 B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

$$\text{chi2}(4) = (b-B)'[(V_b-V_B)^{-1}](b-B)$$

$$= 26.40$$

Prob>chi2 = 0.0000

Al volver a correr la prueba de Hausman, la hipótesis nula fue rechazada nuevamente; por lo que el mejor estimador es el de efectos fijos.

Como se puede observar en la estimación de efectos fijos, Stata calcula una constante que es el promedio de los interceptos individuales, mismo que es estadísticamente significativo. Esto, reforzado con el hecho que la hipótesis nula es $\alpha_i = 0$ de la prueba global F de significancia para los efectos fijos es rechazada, permite confirmar el estatus de “mejor estimador” al modelo de efectos fijos.

Resultados del modelo final

Así el modelo final queda completo con los coeficientes de la estimación de efectos fijos, de la siguiente manera:

LogGasto de turbosina

$$= \alpha_i + 0.0159INPC + 6.21 \times 10^{-7} Pasajeros + 7.25 \times 10^{-9} Carga - 0.000064 CapHotelera$$

Para analizar la heterogeneidad de los aeropuertos (efectos fijos individuales, α_i) en Stata es necesario volver a correr el modelo agregando una variable dicótoma para cada individuo.

Tabla 13. Estimador de efectos fijos y cada uno de los efectos individuales para el modelo:
Log (Gasto de turbosina) = $\alpha_i + \beta_1 INPC + \beta_2 Pasajeros + \beta_3 Carga + \beta_4 CapHotelera$
El aeropuerto cuyo intercepto fue tomado como referencia es el AICM.

Source	SS	df	MS	Number of obs = 2623		
Model	7058.00499	46	153.434891	F(46, 2576)	= 1410.36	
Residual	280.246117	2576	.108791194	Prob > F	= 0.0000	
				R-squared	= 0.9618	
				Adj R-squared	= 0.9611	
				Root MSE	= .32984	
Total	7338.2511	2622	2.79872277			

loggastotu~a	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
inpc	.0159191	.001261	12.62	0.000	.0134464	.0183919
pasajeros	6.21e-07	9.42e-08	6.59	0.000	4.36e-07	8.06e-07
carga	7.25e-09	1.49e-08	0.49	0.626	-2.19e-08	3.64e-08
caphotelera	-.000064	.0000253	-2.53	0.012	-.0001137	-.0000143
aca	-4.498864	.8355562	-5.38	0.000	-6.137294	-2.860434
tgz	-5.342158	1.159605	-4.61	0.000	-7.61601	-3.068307
agu	-5.679717	1.145896	-4.96	0.000	-7.926687	-3.432746
bjx	-4.707201	1.117862	-4.21	0.000	-6.8992	-2.515202
cjs	-5.193436	1.142914	-4.54	0.000	-7.43456	-2.952313
clq	-7.486371	1.215318	-6.16	0.000	-9.86947	-5.103271
cpe	-6.700239	1.205739	-5.56	0.000	-9.064554	-4.335924
ctm	-6.846347	1.19635	-5.72	0.000	-9.192252	-4.500442
cul	-4.819705	1.182845	-4.07	0.000	-7.139127	-2.500282
cun	-.9262372	.613605	-1.51	0.131	-2.129446	-.2769719
cuu	-5.104976	1.145797	-4.46	0.000	-7.351754	-2.858199
cvj	-7.738338	1.161179	-6.66	0.000	-10.01528	-5.461399
czm	-5.573825	1.134345	-4.91	0.000	-7.798144	-3.349505
dgo	-6.415805	1.200961	-5.34	0.000	-8.770751	-4.060859
gdl	-2.343296	.7900943	-2.97	0.003	-3.892581	-.7940118
hmo	-4.605598	1.17042	-3.93	0.000	-6.900658	-2.310538
hux	-5.753638	1.164063	-4.94	0.000	-8.036232	-3.471044
lap	-5.51218	1.197329	-4.60	0.000	-7.860006	-3.164355
lmm	-6.376738	1.206147	-5.29	0.000	-8.741855	-4.011622
lto	-8.064674	1.217797	-6.62	0.000	-10.45263	-5.676713
mex	20.4197	1.184911	17.23	0.000	18.09622	22.74317
mid	-4.346403	1.101159	-3.95	0.000	-6.505651	-2.187156
mlm	-5.39062	1.14772	-4.70	0.000	-7.641168	-3.140072
mty	-3.008919	.947741	-3.17	0.002	-4.86733	-1.150508
mxl	-5.454347	1.183477	-4.61	0.000	-7.775009	-3.133685
mzt	-4.836305	1.029282	-4.70	0.000	-6.854609	-2.818001
oax	-5.719276	1.114677	-5.13	0.000	-7.905029	-3.533523
pbx	-5.607998	1.096426	-5.11	0.000	-7.757964	-3.458031
pqm	-10.61953	1.184321	-8.97	0.000	-12.94184	-8.29721
pvr	-3.445209	.9834874	-3.50	0.000	-5.373715	-1.516703
pxm	-7.88133	1.189533	-6.63	0.000	-10.21387	-5.548791
qro	-5.326733	1.115759	-4.77	0.000	-7.514608	-3.138858
sjd	-3.831809	1.175886	-3.26	0.001	-6.137587	-1.526032
slp	-5.637169	1.131839	-4.98	0.000	-7.856576	-3.417762
tam	-6.000532	1.154431	-5.20	0.000	-8.264238	-3.736826
tap	-6.553343	1.195203	-5.48	0.000	-8.896998	-4.209687
tij	-3.377995	1.124167	-3.00	0.003	-5.582357	-1.173634
tlc	-3.914764	1.149616	-3.41	0.001	-6.169029	-1.660499
ver	-4.88767	1.039681	-4.70	0.000	-6.926366	-2.848973
vsj	-5.079837	1.157803	-4.39	0.000	-7.350156	-2.809518
zcl	-6.101511	1.168656	-5.22	0.000	-8.393111	-3.80991
zih	-5.794432	1.121648	-5.17	0.000	-7.993856	-3.595008
zlo	-6.454009	1.156419	-5.58	0.000	-8.721613	-4.186404

Este procedimiento utiliza como referencia al intercepto de algún aeropuerto, mientras que los demás coeficientes α representan el grado en que los valores de los interceptos de las demás aerolíneas difieren del valor del intercepto del aeropuerto tomado como referencia. El Aeropuerto Internacional de la Ciudad de México (AICM) fue tomado como referencia.

Los coeficientes β a los que se llegan representan tasas de crecimiento mensuales, que al multiplicarlas por 100 son tasas de crecimiento porcentual en Y para cada aumento unitario en X. Por lo que para efectos de este análisis es posible decir que, la variable que más impacto tiene en el gasto de turbosina en términos porcentuales es el INPC, con una tasa del 1.59% mensual (Tabla 13, columnas 1 y 2), lo que aproximadamente sería un 19.08% anual. La variable de pasajeros tiene un impacto sobre el gasto del 0.0007452% anual. Mientras que la carga afecta aproximadamente en un 0.0000087% de forma anual. Sorprendentemente, la variable Capacidad Hotelera (número de habitaciones promedio disponibles) tiene un efecto negativo en el Gasto de turbosina, esto significa que por cada nueva habitación disponible en un destino turístico, el gasto de turbosina se reducirá aproximadamente en un 0.0768% anual (entre más cuartos disponibles exista, significa que menos personas viajaron a estos destinos).

A pesar de que estas cantidades parezcan muy pequeñas en términos porcentuales, debido a las grandes cantidades de combustible que se están pronosticando, estas tasas de crecimiento tienen un efecto que es necesario considerar.

Los coeficientes de las variables dicótomas calculados para este modelo representan el logaritmo de la diferencia entre el intercepto del AICM y el aeropuerto en cuestión. Como se puede inferir, el coeficiente para el AICM es el propio valor del intercepto (fue usado como valor de referencia). En este sentido, para tener el valor del intercepto de cada aeropuerto, será necesario sumar el valor diferencial y el intercepto para el AICM.

Los coeficientes de las variables dicótomas son en su mayoría estadísticamente significativos. En este sentido, el valor del intercepto del aeropuerto de Cancún fue el único no significativo con una probabilidad de 0.131 (Tabla 13, columnas 1 y 5), siendo esta la más alta. Los demás interceptos son estadísticamente significativos para un nivel de significancia del 1%. Con lo anterior se decidió conservar todos los coeficientes para fines del pronóstico y análisis.

Con los coeficientes obtenidos, se calcularon el error estándar σ y el MAPE, para proceder con la evaluación del modelo (Tabla 14). Ambos se calculan con los valores logarítmicos que entrega el modelo y con su respectiva transformación (para tener los resultados expresados en unidades del gasto de turbosina).

Lo que se observa es que al usar los valores a los que se les ha aplicado el logaritmo, los términos de error son más pequeños que los obtenidos al usar los valores corregidos. También se aprecia que en los aeropuertos con los porcentajes de error más grandes, como en el caso de Palenque (MAPE=119.2%) y Cuernavaca (MAPE=59.28%), la participación que tiene en el consumo de turbosina es muy pequeña (0.00% y 0.07% respectivamente). Mientras que en aeropuertos con mayor participación (AICM, Cancún y Guadalajara) el porcentaje de error en la estimación es menor (21.89%, 13.03% y 13.19% respectivamente).

Tabla 14. MAPE y error estándar de estimación (σ) para cada aeropuerto, las columnas con subíndice log se calcularon con los datos que entrega directamente el modelo. Las otras columnas se calcularon aplicando la transformación correspondiente. * Los datos de total, fueron calculados como promedio ponderado, de acuerdo a la participación que tiene cada aeropuerto en el consumo de turbosina.

Aeropuerto	Sigla	Participación por aeropuerto	MAPE _{log}	MAPE	σ_{log}	σ
Acapulco	ACA	0.62%	1.14%	19.25%	0.248334715	\$ 3,540,303.98
Aguascalientes	AGU	0.39%	1.23%	21.11%	0.251679631	\$ 2,437,119.79
Bahías de Huatulco	HUX	0.37%	1.73%	29.19%	0.355380826	\$ 3,887,489.08
Bajío (León, Gto.)	BJX	1.04%	0.71%	12.60%	0.161688239	\$ 4,154,035.11
Campeche	CPE	0.15%	1.38%	21.85%	0.260882997	\$ 1,219,655.57
Cancún	CUN	17.27%	0.63%	13.03%	0.170144838	\$ 89,879,653.02
Cd. Juárez	CJS	0.63%	0.93%	15.91%	0.208855203	\$ 3,495,349.68
Chetumal	CTM	0.14%	1.48%	23.87%	0.312279978	\$ 1,143,115.49
Chihuahua	CUU	0.72%	0.87%	14.98%	0.187811847	\$ 3,681,026.21
Colima	CLQ	0.09%	2.38%	38.03%	0.422688342	\$ 864,093.17
Cozumel	CZM	0.49%	2.33%	42.03%	0.493697674	\$ 6,133,871.67
Cuernavaca	CVJ	0.07%	3.56%	59.28%	0.603245851	\$ 809,645.03
Culiacán	CUL	1.05%	0.67%	12.14%	0.171372576	\$ 4,672,449.13
Durango	DGO	0.23%	1.06%	17.03%	0.212501915	\$ 1,285,515.28
Guadalajara	GDL	7.53%	0.67%	13.19%	0.169001152	\$ 33,449,826.65
Hermosillo	HMO	1.24%	0.98%	17.70%	0.237900355	\$ 7,862,680.48
La Paz	LAP	0.54%	0.86%	14.52%	0.184369144	\$ 2,744,814.96
Loreto	LTO	0.05%	2.09%	32.09%	0.393571633	\$ 477,585.95
Los Mochis	LMM	0.25%	0.78%	12.28%	0.16048714	\$ 987,760.96
Manzanillo	ZLO	0.23%	2.63%	41.92%	0.498784852	\$ 2,973,469.88
Mazatlán	MZT	0.69%	1.08%	18.65%	0.234489034	\$ 4,052,327.58
Mérida	MID	1.42%	0.91%	16.18%	0.216830502	\$ 8,592,728.48
Mexicali	MXL	0.54%	1.01%	17.11%	0.229330183	\$ 3,263,625.54
México	MEX	38.21%	0.96%	21.89%	0.274738493	\$ 346,111,959.60
Monterrey	MTY	4.84%	0.74%	14.06%	0.175922731	\$ 22,274,045.32
Morelia	MLM	0.59%	0.94%	16.04%	0.202960018	\$ 2,896,756.63
Oaxaca	OAX	0.34%	0.78%	12.78%	0.15904101	\$ 1,487,558.09
Palenque	PQM	0.00%	7.61%	119.20%	1.159413298	\$ 255,167.33
Puebla	PBC	0.43%	1.03%	16.67%	0.236660834	\$ 2,467,942.14
Puerto Escondido	PXM	0.05%	2.65%	44.90%	0.516826969	\$ 702,485.68
Puerto Vallarta	PVR	2.83%	1.93%	35.87%	0.413777168	\$ 32,301,095.13
Querétaro	QRO	0.44%	1.78%	33.05%	0.373547281	\$ 4,719,213.74
San José Del Cabo	SJD	2.89%	1.22%	25.20%	0.324910256	\$ 24,480,680.80
San Luis Potosí	SLP	0.38%	1.02%	16.89%	0.204254158	\$ 2,298,300.19
Tampico	TAM	0.28%	0.87%	14.22%	0.177322097	\$ 1,334,031.57
Tapachula	TAP	0.21%	0.82%	13.06%	0.168645373	\$ 893,970.29
Tijuana	TIJ	4.58%	0.62%	11.96%	0.159187944	\$ 19,187,015.08
Toluca	TLC	2.93%	0.95%	18.05%	0.243080162	\$ 17,060,362.67
Tuxtla Gutiérrez	TGZ	0.59%	0.94%	16.27%	0.202166217	\$ 2,981,188.28
Veracruz	VER	0.64%	0.76%	13.07%	0.167321665	\$ 2,872,193.08
Villahermosa	VSA	0.76%	0.76%	13.19%	0.173154215	\$ 3,435,218.08
Zacatecas	ZCL	0.29%	1.02%	16.51%	0.21299338	\$ 1,622,075.42
Zihuatanejo	ZIH	0.43%	2.92%	48.38%	0.550146417	\$ 5,963,634.48
Total general		97.45%	0.89%	18.21%	0.23008397	\$ 155,031,045.47

Un Pronóstico a partir del modelo

Con dicho modelo, se procedió a realizar el pronóstico para el mes de febrero del consumo de turbosina en cada aeropuerto, para ello se pronosticaron cada una de las variables explicativas (INPC, Pasajeros, Carga y Capacidad Hotelera) mediante el método de Holt-Winters Aditivo²⁷. Este método fue seleccionado para la estimación de las variables causales debido a que se considera como un método genérico, que permite hacer una estimación independientemente del comportamiento que presente la serie de datos.

El pronóstico, fue comparado posteriormente con los datos reales del mes de febrero. Como se puede observar en la Tabla 15, el error del pronóstico es muy elevado, esto se puede explicar debido a los errores que se “arrastran” de las estimaciones de las variables explicativas. Es decir, el modelo formulado para esta tesis, tiene un error inherente a la estimación; este error es aumentado debido a los errores que presentan las estimaciones de las variables explicativas.

Adicionalmente, se realizó una estimación con los valores reales del mes de febrero, con el objetivo de validar al modelo, y también se calcularon intervalos de confianza para cada estimación (Tabla 16).

Cabe aclarar que no se llevó a cabo un pronóstico para periodos de tiempo mayores debido a que este tipo de análisis son dinámicos, Esto quiere decir que conforme pasa el tiempo y se obtenga nueva información, se debe de integrar la nueva información para recalculando los parámetros. Esta constante alimentación del panel con nueva información, permite hacer mejores análisis, al adecuar el modelo a las condiciones que se van presentando.

²⁷ Se trata de un método de pronósticos a través de series de tiempo, este método descompone la serie en tres indicadores principales. El nivel de tendencia, la pendiente de tendencia y el factor de estacionalidad. La peculiaridad de este método es que es un modelo genérico, que permite realizar una estimación de cualquier serie de tiempo sin importar su comportamiento.

Para ahondar más en el tema y en el algoritmo del método se sugieren los siguientes libros:

- Shumway, Robert H. and David S. Stoffer. *Time Series Analysis and Its Applications*. London: Springer, 2011. Chapter 1: Characteristics of Time Series - Stationary Time Series.
- Brockwell Peter J. and Rochard A. Davis. *Introduction to Time Series*. New York: Springer- Verlag, 2002. Chapter 9 Forecasting Techniques - The Holt-Winters Seasonal Algorithm.
- Tsay, Ruey. *Analysis of financial time series*. Chicago: Wiley-Interscience, 2005. Chapter 2: Linear Time Series Analysis and Its Applications - Seasonal Models. .

Tabla 15. Pronósticos para el Log (Consumo de turbosina), Consumo de turbosina y la demanda de Turbosina para el mes de febrero de 2015, usando las variables explicativas pronosticadas mediante el método de Holt – Winters. Nótese que los errores de estimación son menores para el pronóstico del Log (Consumo de turbosina), esto es debido a que se trata de la variable que directamente pronostica el modelo.

Aeropuerto	Demanda Real [I]	Pronóstico Demanda	% error	Consumo real	Pronóstico Consumo	%error	Log (Consumo Real)	Pronóstico Log (Consumo)	% error
ACA	1705727	2916729.25	71%	9552446.41	16334325.4	71%	16.0723078	16.6087793	3%
AGU	1333910	2168982.49	63%	7470189.42	12146779.1	63%	15.8264309	16.3125746	3%
BJX	3366388	5649905.24	68%	18852513.3	31640712.2	68%	16.7521568	17.2699552	3%
CJS	1814848	3345286.67	84%	10163548	18734341.2	84%	16.1343182	16.7458688	4%
CLQ	395935	450619.388	14%	2217323.1	2523567.7	14%	14.6118112	14.7411842	1%
CPE	528770	977132.19	85%	2961228.32	5472155.21	85%	14.9011147	15.5151831	4%
CTM	372581	796509.614	114%	2086535.56	4460629.05	114%	14.5510156	15.3108004	5%
CUL	2683281	5975906.5	123%	15026963.8	33466390.9	123%	16.5253567	17.3260522	5%
CUN	64847451	114816649	77%	363159990	642998493	77%	19.710354	20.2816529	3%
CUU	2246313	4086575.6	82%	12579846.9	22885722.3	82%	16.3476066	16.9460238	4%
CVJ	138162	300953.898	118%	773737.592	1685408.03	118%	13.5589881	14.3375182	6%
CZM	1807550	2404062.12	33%	10122677.6	13463276.7	33%	16.1302888	16.4154763	2%
DGO	664879	1275060.27	92%	3723468.66	7140617.98	92%	15.1301662	15.7813099	4%
GDL	20127159	40706664.2	102%	112716518	227966274	102%	18.5403865	19.2447083	4%
HMO	3075994	7100993.05	131%	17226243	39767123.1	131%	16.6619445	17.4985511	5%
HUX	1838589	2164784.51	18%	10296502.8	12123269.5	18%	16.1473149	16.3106373	1%
LAP	1528181	2904346.76	90%	8558149.76	16264980.8	90%	15.9623946	16.6045249	4%
LMM	689664	1293113.18	87%	3862270.11	7241718.26	87%	15.1667657	15.7953691	4%
LTO	156798	243251.004	55%	878103.291	1362259.13	55%	13.6855195	14.124655	3%
MEX	109616086	250567901	129%	613874194	1403235363	129%	20.2353006	21.0620464	4%
MID	3811533	7918855.14	108%	21345423.2	44347330.7	108%	16.8763479	17.6075631	4%
MLM	1314231	3006464.07	129%	7359982.7	16836860.1	129%	15.8115681	16.6390811	5%
MTY	14758928	25888509.9	75%	82653243.4	144981350	75%	18.2301646	18.7921157	3%
MXL	1317230	2977417.98	126%	7376777.76	16674195.6	126%	15.8138475	16.6293729	5%
MZT	2196544	3793951.89	73%	12301129.6	21246965.2	73%	16.3252017	16.8717246	3%
OAX	1207238	1957867.9	62%	6760798.36	10964490.9	62%	15.7266515	16.2101725	3%
PBC	853890	1994407.31	134%	4781971.83	11169119.7	134%	15.3803635	16.2286634	6%
PQM	42747	16883.0406	-61%	239392.603	94548.7413	-61%	12.3858602	11.4568708	-8%
PVR	13024589	15130210.7	16%	72940563.5	84732508.3	16%	18.1051555	18.2550099	1%
PXM	256889	273704.541	7%	1438634.91	1532805.64	7%	14.1792052	14.2426104	-0%
QRO	2181646	2027795.57	-7%	12217697.5	11356101.2	-7%	16.3183961	16.2452657	-0%
SJD	8773572	18639210.3	112%	49133933.1	104383678	112%	17.7100605	18.4635839	4%
SLP	1257098	2373503.28	89%	7040025.33	13292140.5	89%	15.7671223	16.4026835	4%
TAM	864881	1660625.21	92%	4843523.85	9299866.45	92%	15.3931531	16.0455106	4%
TAP	573764	1117166.16	95%	3213204.61	6256376.22	95%	14.9827793	15.6491117	4%
TGZ	2054237	3515342.4	71%	11504179.1	19686690.7	71%	16.2582209	16.7954534	3%
TIJ	10735308	24006255.6	124%	60120086.3	134440312	124%	17.9118546	18.7166309	4%
TLC	5890986	14097274.9	139%	32990817.5	78947840.7	139%	17.3117398	18.1842979	5%
VER	1897995	3883102.94	105%	10629189.5	21746230.6	105%	16.1791145	16.894951	4%
VSA	2554166	4398943.41	72%	14303891.4	24635050.8	72%	16.4760422	17.0196808	3%
ZCL	679764	1570796.98	131%	3806827.93	8796808.64	131%	15.1523068	15.9898996	6%
ZIH	2468610	1882641.18	-24%	13824759	10543204.7	-24%	16.4419717	16.1709921	-2%
ZLO	1134239	1081542.35	-5%	6351987.9	6056875.06	-5%	15.6642784	15.6167046	-0%
Errores Globales			101%			101%			4%

Aeropuerto	Pronóstico	Min	Max	Valor Real
ACA	16.5917967	16.1050696	17.0785238	16.219019
AGU	16.3137009	15.8204179	16.8069839	15.9731421
BJX	17.271086	16.9541829	17.5879891	16.898868
CJS	16.7649007	16.355552	17.1742493	16.2810293
CLQ	14.7353353	13.9068814	15.5637892	14.7585224
CPE	15.5125984	15.0012771	16.0239197	15.0478259
CTM	15.2976753	14.6856178	15.9097328	14.6977268
CUL	17.3470179	17.0111338	17.682902	16.6720679
CUN	20.2654871	19.9320093	20.5989648	19.8570652
CUU	16.9494281	16.5813237	17.3175326	16.4943178
CVJ	14.3844419	13.2021018	15.5667821	13.7056992
CZM	16.4277153	15.4600856	17.3953449	16.277
DOO	15.7748258	15.3583297	16.1913219	15.2768774
GDL	19.2242378	18.8930017	19.555474	18.6870977
HMO	17.5141882	17.047912	17.9804643	16.8086557
HUX	16.306565	15.6100314	17.0030986	16.294026
LAP	16.6178957	16.2565388	16.9792526	16.1091058
LMM	15.8134098	15.4988608	16.1279588	15.3134769
LTO	14.0986229	13.3272367	14.8700091	13.8322307
MEX	21.0352737	20.4967962	21.5737513	20.3820117
MID	17.5966484	17.1716685	18.0216284	17.0230591
MLM	16.634985	16.2371907	17.0327793	15.9582793
MTY	18.8068906	18.4620883	19.1516928	18.3768758
MXL	16.6314592	16.1819803	17.0809381	15.9605587
MZT	16.8858192	16.4262291	17.3454092	16.4719128
OAX	16.1978434	15.8861288	16.5095581	15.8733627
PBC	16.2170359	15.7531892	16.6808826	15.5270747
PQM	11.5338119	9.26140361	13.8062202	12.5325714
PVR	18.2684431	17.4574547	19.0794314	18.2518666
PXM	14.2413803	13.2284181	15.2543426	14.3259164
QRO	16.2649507	15.5328114	16.9970899	16.4651073
SJD	18.4247262	17.7879138	19.0615386	17.8567716
SLP	16.4030316	16.0027008	16.8033624	15.9138335
TAM	16.0329402	15.6853953	16.3804851	15.5398643
TAP	15.641899	15.3113602	15.9724379	15.1294905
TGZ	16.7826578	16.3864193	17.1788963	16.4049321
TIJ	18.7236732	18.4116705	19.0356758	18.0585657
TLC	18.1836499	17.7072216	18.6600783	17.458451
VER	16.8906306	16.5626862	17.2185751	16.3258257
VSA	17.0224527	16.6830767	17.3618287	16.6227534
ZCL	15.9797127	15.5622533	16.397172	15.299018
ZIH	16.1695868	15.0913196	17.2478539	16.5886829
ZLO	15.6110391	14.6334388	16.5886395	15.8109896

Tabla 16. Pronóstico e intervalos de confianza con $\alpha = 0.05$ para el mes de febrero de 2015 de la variable de respuesta Log (Consumo de turbosina). Para este pronóstico se usaron los datos reales de las variables explicativas correspondientes al mes de febrero de 2015.

Capítulo 5.

Discusión, Conclusiones y Recomendaciones



Aeropuerto Internacional de Cancún-México 2013. Carga de mercancía.
Fuente: www.skyscrapercity.com

Discusión

En un mundo con una dinámica como la actual, el éxito de las grandes empresas radica en estar un paso delante de sus competidores y de la propia demanda del mercado; es aquí donde la labor de pronosticar cobra vida. En este trabajo, se parte de conocer la demanda de turbosina y algunas variables que impactan en su consumo. La importancia de dicho pronóstico radica primeramente en la planeación de producción del combustible y su posible importación. Por otro lado permitirá realizar una mejor estrategia de distribución por región y plantear alternativas para la reducción de costos, mejorando la competitividad. Esto último de vital importancia debido a la próxima abertura del mercado nacional. En este trabajo, se proponen algunas variables macroeconómicas y otras específicas de la industria aeronáutica, que evaluadas con análisis de panel explican y predicen el consumo de la turbosina en 43 aeropuertos. Dichos aeropuertos representan más del 97% del consumo nacional de turbosina.

No se cuenta con información de que existan estudios similares previos que permitan calcular una demanda de combustible para cada aeropuerto y que mucho menos tomen en cuenta la interacción entre ellos. Hasta el momento solo se han considerado estudios en donde la apreciación del fenómeno de consumo de combustible se aborde de manera global (a nivel república).

En esta tesis se aborda un enfoque que permite sacar estimaciones para cada aeropuerto perteneciente al Sistema Aeroportuario Mexicano. Este modelo pronostica el consumo de turbosina, a través del gasto de la misma. Al tener involucrado el costo del combustible en la obtención del gasto, no se obtiene de manera directa la demanda de turbosina en unidades volumétricas de combustible. Sin embargo se llega a un modelo formado por cuatro variables explicativas y que queda conformado de la siguiente manera:

LogGasto de turbosina

$$= \alpha_i + 0.0159INPC + 6.21 \times 10^{-7} Pasajeros + 7.25 \times 10^{-9} Carga - 0.000064 CapHotelera$$

Donde α_i es el intercepto individual correspondiente a cada aeropuerto i .

De las anteriores, únicamente el INPC puede ser considerado como variable macroeconómica. Mientras que las demás, son variables específicas de la industria aérea en México, siendo en particular la capacidad hotelera (número de habitaciones promedio disponibles) parte de la industria turística (que va muy ligada al sector de transporte de pasajeros en el país).

Como prueba de verisimilitud de este supuesto se encuentra la prueba de significancia realizada a cada una de las variables explicativas (Tablas 8 y 13, columna 5). Para esta prueba la carga no es estadísticamente significativa con una probabilidad de 0.626, sin embargo se decidió conservar este coeficiente. Cabe recordar que esta prueba solo nos dice que tan probable es que el valor real del coeficiente sea cero, y es de destacar que el valor del coeficiente calculado es muy cercano a cero (7.25×10^{-9}). Por otro lado, la capacidad hotelera es significativa para niveles de significancia mayores al 2.5%.

Los resultados arrojaron que el modelo más adecuado para la estimación, así como el que mejor describe el comportamiento del consumo de turbosina en los diferentes

aeropuertos es el de efectos fijos. Este modelo entre otras cosas permite visualizar la heterogeneidad de cada individuo a través del término independiente o interceptos de la ecuación. Sin embargo, solo permite hacer estimaciones de los individuos que formaron parte de la muestra. Es decir que el modelo de esta tesis solo permite hacer estimaciones e inferencias de los 43 aeropuertos tomados en cuenta para el estudio; y no de la totalidad de los aeropuertos mexicanos.

Debido a que la variable de Gasto de Turbosina se encuentra afectada por una transformación logarítmica, es necesario realizarle una transformación exponencial con el objetivo de obtener datos en unidades volumétricas (litros de combustible).

Al obtener los estadísticos MAPE y σ para un modelo de la forma Lin-Lin, (Anexo D, tabla 33) el MAPE global se reduce en 4 puntos porcentuales mientras que la desviación estándar aumenta más del 300% (Comparando con los resultados de la tabla 14, columnas 5 y 7), confirmándose así que la decisión de usar un modelo Log-Lin para reducir el error estándar es adecuada.

Conclusiones

Se obtiene un modelo que explica el consumo de turbosina en 43 aeropuertos mexicanos a través del gasto que se genera en cada uno. Este modelo, es el resultado de un análisis causal, que depende positivamente del INPC local, los pasajeros y la cantidad de carga transportada; y negativamente de la variable capacidad hotelera (número de habitaciones promedio disponibles) en las ciudades donde se encuentran los aeropuertos.

Tabla 17. Variables explicativas y su respectivo coeficiente. Elaboración propia con datos de cálculos en STATA

Variable	INPC	Pasajeros	Carga	Capacidad Hotelera
Coeficiente	0.0159	6.21×10^{-7}	7.25×10^{-9}	- 0.000064

El modelo no estima directamente la demanda de turbosina; el consumo se estima a través del gasto, que involucra a la demanda y al precio deflactado del combustible.

Al utilizar una estimación de efectos fijos, únicamente es viable realizar análisis y sacar conclusiones referentes a los 43 aeropuertos que componen la muestra de estudio.

Debido a que se utilizó un modelo Log-Lin, los coeficientes de las variables explicativas son constantes de proporcionalidad. En otras palabras, representan el cambio porcentual que sufre la variable dependiente por un cambio en una unidad en las variables explicativas.

Esto quiere decir que un incremento en una unidad del INPC representará un incremento ubicado entre el 1.34% y el 1.84%, siendo 1.59% el promedio en el gasto de turbosina, lo que evidencia que el fenómeno de inflación por si solo provoca un aumento en el gasto de combustible. Considerando que el gasto de turbosina se calcula con precios deflactados, la inflación no está ligada al precio si no, a la demanda de combustible. Esto es información valiosa para la regulación monetaria que realiza el Banco de México, ya que un aumento o disminución en la inflación, impactaría sensiblemente la demanda de turbosina.

El coeficiente de la variable pasajeros establece que cada pasajero representa un aumento aproximado del 6.21×10^{-5} % promedio en el consumo de turbosina. Dicho aumento se podrá ubicar entre el 4.36×10^{-5} % y el 8.06×10^{-5} %. Este indicador, ofrece a las aerolíneas y distribuidores de combustible, la oportunidad de conocer el impacto global en el consumo de combustible de un solo pasajero. Dicho indicador permite evaluar de mejor manera las estrategias de fidelización del cliente y servicios que se pueden ofrecer, sirviendo como análisis para un sector que ha presentado un crecimiento atractivo en el último año (García Morales 2015).

Con la variable carga ocurre algo similar, cada kilogramo de carga representa un aumento aproximado del 7.91×10^{-7} % en el consumo de turbosina. Estimado entre el -2.19×10^{-8} % y el 3.64×10^{-8} %. Este indicador proporciona una visión integral del impacto de la carga en cada operación. El intervalo de confianza sugiere que es factible que un aumento en la carga pueda representar una disminución del consumo, esto se explica con la economía de escala.

Por otro lado, cabe destacar el impacto negativo que tiene la variable de capacidad hotelera (número de habitaciones promedio disponibles) en el modelo, este impacto es del -0.0064% por cada habitación disponible, con un intervalo de confianza del -0.0113% al -0.0014%. Al analizar la estimación puntual del coeficiente, se entiende que al haber mayor número de cuartos disponibles (sin ocupar) el consumo disminuye. En otras palabras, está explicando lo que se deja de consumir de combustible por cada habitación que no es ocupada en un destino.

Se puede concluir, que una inversión en el sector hotelero repercutiría con un incremento en el consumo de turbosina en México.

Las variables de pasajeros y carga permiten visualizar el impacto que tienen estos mercados en la industria del transporte aéreo. En México, el mayor mercado de la industria aérea lo tiene el traslado de pasajeros.

Más del 80% del gasto de turbosina generado en México se realiza en 10 aeropuertos, siendo el AICM el que mayor consumo tiene (poco más del 38% del consumo nacional). Cualquier cambio que tenga el consumo del AICM afecta irremediablemente la demanda de los demás aeropuertos, lo cual puede ser benéfico o nocivo al sistema aeroportuario nacional, debido a que la concentración de actividades en un solo punto puede representar ventajas en el control de las mismas o desventajas cuando se está en presencia de una saturación en la capacidad del aeropuerto (situación por la que atraviesa el actual AICM).

Recomendaciones

Es recomendable hacer más test como la prueba de Goldfield-Quandt, el Rango de Correlación de Spearman, o la prueba de White para descartar la presencia de heteroscedasticidad y de ser así utilizar algún otro método de pronóstico o corrección de datos como el de Mínimos Cuadrados Ponderados (pronóstico) o el filtro de Hodrik Prescott (corrección de datos).

Se recomienda realizar una evaluación del modelo periódicamente para identificar posibles cambios que pudieran mejorar la estimación alcanzada en esta tesis, y la posible inclusión de otras variables.

Se propone revisar el sistema de distribución de combustible actual, y revisar si cumple con los requerimientos del mercado eficientemente. Probar si la inclusión de otras variables como el precio de los boletos de pasajeros y/o transporte de carga por destino podría explicar mejor al modelo.

Este estudio sirve como precursor para estudios de factibilidad de una reestructuración en la distribución del hidrocarburo. Al ser capaz de dar una estimación a largo plazo, se pueden estimar los requerimientos de combustible futuros en cada región y evaluar otras alternativas de distribución de combustible.

Para efectos del pronóstico, se sugiere realizar un contraste con series de tiempo y el presente modelo causal, con el fin de medir alcances y limitaciones de ambos métodos de pronóstico para la industria del combustible para aviones.

Anexo B

Regresión lineal simple

Este es el caso particular de regresión lineal, en donde solo interviene la variable de interés y una variable explicativa, es el método más sencillo de regresión. Se hará que, una serie de datos ubicados en el plano se ajusten a una línea recta.

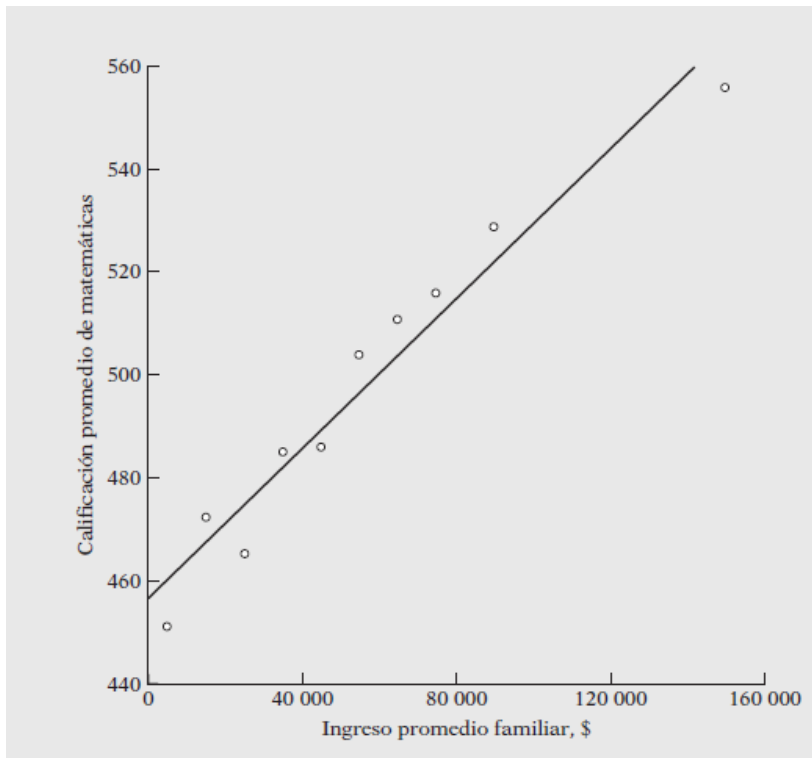


Ilustración 16. Recta de ajuste de regresión lineal aplicada a una serie de datos (Gujarati y Porter 2009, 47)

De esta forma la ecuación de regresión se escribe de la forma $y = b_0 + b_1x + \varepsilon$ (1), en donde

- y = variable dependiente
- x_1 = variable independiente o explicativa
- b_n = coeficientes de regresión
- ε = término de error.

Las consideraciones básicas del modelo de regresión son:

1. El error aleatorio ε tiene una media cero y varianza σ^2
2. Los errores son variables aleatorias no correlacionadas (un error no depende del otro)
3. La variable de respuesta y es una variable aleatoria cuya media para cada valor de x es

$$E(y|x) = b_0 + b_1x$$

con varianza

$$\text{Var}(y|x) = \text{Var}(b_0 + b_1x + \varepsilon) = \sigma^2 \quad (2)$$

Por lo anterior se espera cada observación de la variable dependiente tenga una distribución normal con los parámetros antes mencionados.

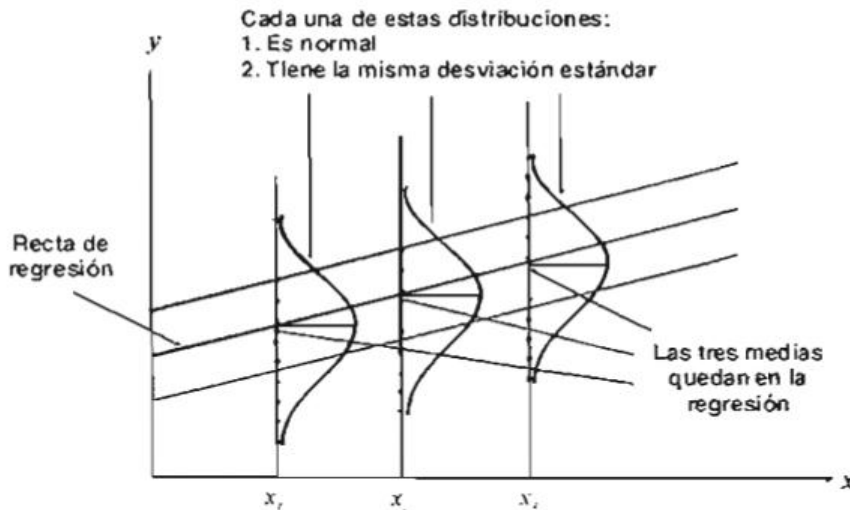


Ilustración 17. Apreciación del tercer supuesto de regresión (Escalante Membrillo 2005, 34)

Mínimos Cuadrados Ordinarios

Dados n pares de observaciones de Y y X , nos interesa determinar la función de regresión de manera que quede lo más cerca posible de la Y observada. Con este fin, se adopta el siguiente criterio: seleccionar la función de regresión de modo que la suma de los residuos $\sum \hat{u}_i = \sum (Y_i - \hat{Y}_i)$ sea la menor posible. Este criterio, aunque es intuitivamente atractivo, no es muy bueno. (Gujarati y Porter 2009, 56).

El criterio anterior no toma en cuenta la dispersión que existe entre los datos, de esta forma, todos los residuos reciben el mismo peso específico sin considerar el grado de dispersión a la función de regresión. Pues es muy probable que la suma algebraica de los residuos de un valor muy pequeño o cercano a cero al compensarse los errores positivos con lo negativos. Por lo que es necesario identificar un criterio que nos permita involucrar a la dispersión y no solo la suma de residuales.

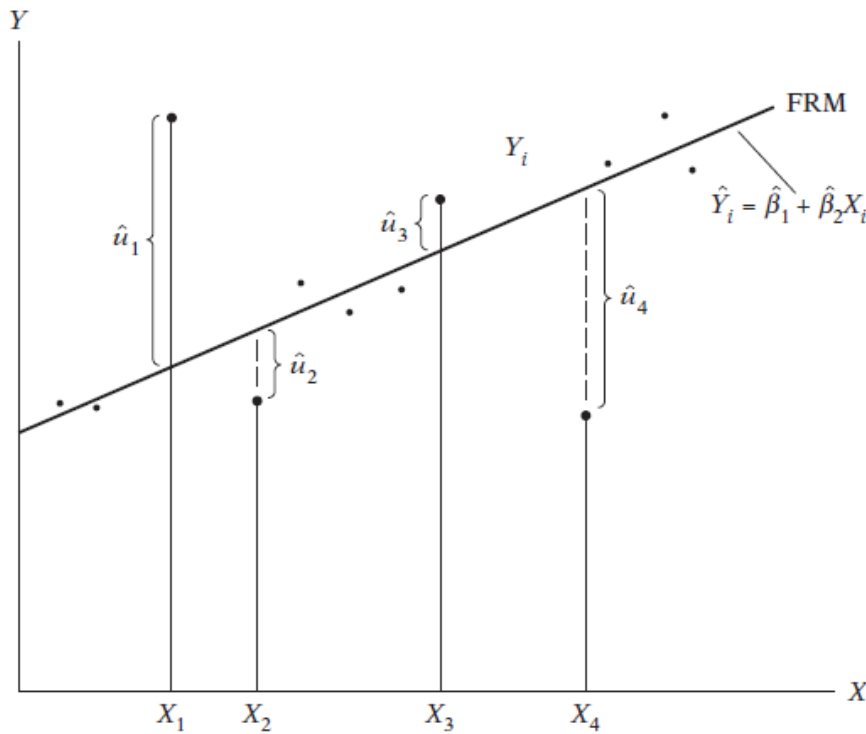


Ilustración 18. Valores observados y estimación lineal. (Gujarati y Porter 2009, 56)

El problema anterior se elimina al introducir el criterio de minimización de la suma de los cuadrados de los residuos (SCR), para encontrar los valores β_0 y β_1 que permitan resolver el problema. (Novales 2010, 22).

$$\underset{\widehat{\beta}_0, \widehat{\beta}_1}{Min} SCR = \sum_{i=1}^n \widehat{u}_i^2 \quad (3)$$

Como \widehat{u}_i puede escribirse como $\widehat{u}_i = y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_i)$, la ecuación anterior puede escribirse como

$$\underset{\widehat{\beta}_0, \widehat{\beta}_1}{Min} SCR = \sum_{i=1}^n (y_i - \widehat{\beta}_0 + \widehat{\beta}_1 x_i)^2 \quad (4)$$

La solución a este problema de optimización se denota por $\widehat{\beta}_0, \widehat{\beta}_1$, y se denomina estimador de Mínimos Cuadrados Ordinarios (que abreviaremos como MCO) de los coeficientes del modelo de regresión lineal simple. El estimador MCO (Novales 2010, 23) escoge, de entre todas las posibles, la recta que minimiza la suma de los cuadrados de las distancias entre cada punto de la nube generada por las observaciones muestrales y el asignado por la recta.

Derivando SCR con respecto a ambas variables (β_0 y β_1) e igualando dichas derivadas a cero, tenemos:

$$\frac{\partial SCR}{\partial \beta_0} = -2 \sum_{i=1}^n (y_i - \widehat{\beta}_0 + \widehat{\beta}_1 x_i) = 0 \quad (5)$$

$$\frac{\partial SCR}{\partial \beta_1} = -2 \sum_{i=1}^n (y_i - \widehat{\beta}_0 + \widehat{\beta}_1 x_i) x_i = 0 \quad (6)$$

con matriz de derivadas segundas:

$$\frac{\partial^2 SCR}{\partial \beta_0 \partial \beta_1} = \begin{pmatrix} 2n & 2 \sum_{i=1}^n x_i \\ 2 \sum_{i=1}^n x_i & 2 \sum_{i=1}^n x_i^2 \end{pmatrix}$$

que tiene por determinante:

$$DET = 4 \left(n \sum_{i=1}^n (x_i^2) - \left(\sum_{i=1}^n (x_i) \right)^2 \right) = n^2 \left(\frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2 \right) = n^2 \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = n^2 S_x^2$$

Siendo S_x^2 la varianza muestral de X : $S_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$. A partir de esos resultados, es posible plantear que al resolver el sistema planteado en las ecuaciones (5) y (6) serán los coeficientes β_0 y β_1 , que serán los coeficientes de nuestro modelo de regresión lineal.

Al resolver el sistema se obtiene

$$n \sum_{i=1}^n (y_i) = n \widehat{\beta}_0 + \widehat{\beta}_1 \sum_{i=1}^n (x_i) \quad (7)$$

$$n \sum_{i=1}^n (y_i x_i) = \widehat{\beta}_0 \sum_{i=1}^n (x_i) + \widehat{\beta}_1 \sum_{i=1}^n (x_i^2) \quad (8)$$

que constituyen un par de ecuaciones simultáneas en las incógnitas, $\widehat{\beta}_0, \widehat{\beta}_1$. Este sistema se conoce como sistema de ecuaciones normales. Al despejar las ecuaciones, los valores de $\widehat{\beta}_0, \widehat{\beta}_1$, servirán como estimadores de β_0 y β_1 que son los coeficientes de (1).

Si se despeja $\widehat{\beta}_0$ en (7), se obtiene:

$$\widehat{\beta}_0 = \frac{\sum_{i=1}^n (y_i) - \widehat{\beta}_1 \sum_{i=1}^n (x_i)}{n} = \bar{y} - \widehat{\beta}_1 \bar{x} \quad (9)$$

Que se puede utilizar para obtener el estimador de MCO de β_0 , una vez que se tenga el estimador de β_1 por lo que sustituyendo (9) en (8):

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n (x_i y_i) - \frac{1}{n} (\sum_{i=1}^n x_i) (\sum_{i=1}^n y_i)}{\sum_{i=1}^n (x_i^2) - \frac{1}{n} (\sum_{i=1}^n x_i)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x}) (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{n S_{xy}}{n S_x^2} = \rho_{xy} \frac{S_y}{S_x} \quad (10)$$

Donde S_{xy} , S_x , S_y , S_x^2 , S_y^2 , denotan respectivamente, la covarianza, varianzas y desviaciones típicas muestrales de X e Y. Las expresiones (9) y (910) son útiles, pues

proporcionan directamente las estimaciones MCO como función de estadísticos muestrales, sin necesidad de resolver el sistema de ecuaciones normales. Primero se calcula $\widehat{\beta}_1$ y, luego, se obtiene $\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}$.

Regresión Lineal Múltiple

El modelo de regresión lineal múltiple es la generalización del modelo de regresión lineal simple. La gran diferencia que existe entre ambas regresiones, es la presencia de múltiples variables explicativas en el modelo de regresión múltiple. Esta generalización surge al encontrar la necesidad de incluir otras variables que permitan mejorar el cálculo realizado cuando solo existe una variable explicativa.

De forma general la ecuación de regresión lineal múltiple es $y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n + \varepsilon_t$, que fue introducida al principio de esta sección.

Para determinar los valores de los coeficientes $b_0, b_1, b_2, \dots, b_n$ así como el del error estándar asociado, se utiliza el método de Mínimos Cuadrados Ordinarios (MCO) (ver apéndice A). Éstos cálculos suelen ser muy complejos y extensos, al verse involucrada el álgebra matricial. Por lo que para facilitar el cálculo, se dispone software especializado (Minitab, Excel, Matlab, etc.) para realizarlos.

Ciertamente, casi nunca se conocen los valores exactos de los parámetros de regresión, sino que solo se da una aproximación a ellos. A partir de los datos muestrales, se encuentran las estimaciones de los coeficientes y se determina el hiperplano en el espacio $(k+1)$ -dimensional de las variables regresoras $\{x_k\}$ y de la variable de interés $\{y\}$, que mejor ajuste al conjunto de datos, llamado hiperplano de regresión muestral. Así la ecuación de regresión múltiple ajustada quedará:

$$\hat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2 + \dots + \widehat{\beta}_n x_n .$$

Donde \hat{y} es el valor estimado de la variable dependiente, x_1, x_2, \dots, x_n las variables explicativas del modelo y $\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \dots, \widehat{\beta}_n$ son las estimaciones muestrales de los parámetros de regresión (Escalante Membrillo 2005, 64).

Supuestos de la regresión

Los supuestos que se utilizan en la regresión múltiple, son en principio los mismos utilizados para la regresión simple (del I al VII), adicionalmente se consideran otros supuestos más (VIII y IX).

- I. Modelo de regresión lineal, o lineal en los parámetros²⁸
- II. Valores fijos de X o valores de X independientes del término de error. En este caso, esto significa que se requiere covarianza cero entre ε (término de error) y cada variable X.
- III. Valor medio del término de error ε_t igual a cero. $E(\varepsilon_t | x_{1t}, x_{2t}, \dots, x_{nt}) = 0$, para el periodo t.

²⁸ Se dice que una función es lineal en el parámetro, digamos β_1 , si β_1 aparece elevado a una potencia de 1 solamente y no está multiplicado ni dividido por ningún otro parámetro (por ejemplo, $\beta_1\beta_2$, β_2/β_1 , etcétera). (Gujarati y Porter 2009).

- IV. Homoscedasticidad o varianza constante de ε_t . $Var(\varepsilon_t) = \sigma_\varepsilon^2$ para todo $t = 1, 2, \dots, k$.
- V. No autocorrelación, o correlación serial entre los términos de error correspondientes a dos observaciones muestrales cualesquiera. $cov(\varepsilon_i \varepsilon_j) = 0$ con $i \neq j$
- VI. El número de observaciones t debe ser mayor que el de parámetros a estimar. $k > n$.
- VII. Debe haber variación en los valores de las variables X.
- VIII. No debe haber colinealidad exacta entre las variables X.
- IX. No hay sesgo de especificación. El modelo está especificado correctamente.

Anexo C

Pruebas de Hipótesis

En todo proceso de investigación se formulan una o varias hipótesis que se buscan aceptar o rechazar para así dar validez a la investigación. En estadística una de las herramientas más usuales para hacer ésta validación de datos recibe el nombre de prueba de hipótesis.

El objetivo de una prueba de hipótesis es verificar si existe suficiente evidencia estadística para rechazar una presunta hipótesis en favor de una hipótesis alternativa. Tradicionalmente estas dos hipótesis reciben el nombre de hipótesis nula (H_0) e hipótesis alternativa (H_A). H_A siempre debe de coincidir con la hipótesis de la investigación, razón por la que se busca rechazar H_0 . Cuando se rechaza H_0 significa que el factor estudiado ha influido significativamente en los resultado de la investigación y se acepta la hipótesis alternativa.

En prueba de hipótesis existe la posibilidad de cometer algún tipo de error al momento de aceptar o rechazar la hipótesis nula. A estos errores se les conoce como Error Tipo I y Error Tipo II.

El error tipo I es considerado como el más serio de los dos errores potenciales (Gujarati y Porter 2009, 833). La probabilidad de dicho error es denotada por α y es también conocido como nivel de significancia de la prueba. Para estos efectos α representa el máximo riesgo aceptable de rechazar H_0 cuando en realidad es verdadera.

El error tipo II es considerado el menos “serio” de los dos y es denotado con la probabilidad β , siendo este error cuando se acepta la hipótesis nula siendo que en realidad es falsa. Es conocido como potencia de la prueba y poniéndolo de otra manera, la potencia de una prueba es su capacidad para rechazar una hipótesis nula falsa.

Idealmente la lógica diría que es necesario reducir ambos tipos de errores, sin embargo no es posible reducir ambos errores de manera simultánea, por lo que en la práctica se busca mantener al mínimo el error Tipo I (el más grave), aceptando valores como 0.01 y 0.05, y tratar de reducir al máximo la probabilidad de incurrir en un error tipo II.

Decisión	Estado de naturaleza	
	H_0 es verdadera	H_0 es falsa
Rechazar	Error tipo I	No hay error
No rechazar	No hay error	Error tipo II

Ilustración 19. Posibles errores cometidos en las pruebas de hipótesis. (Gujarati y Porter 2009)

Los pasos a seguir en la prueba de hipótesis son:

- I. La formulación de la hipótesis nula y alternativa. Que pueden ser de las siguientes formas:

$$\begin{aligned}
 H_0: P = p & \quad H_0: P < p & \quad H_0: P > p \\
 H_A: P \neq p & \quad H_A: P \geq p & \quad H_A: P \leq p
 \end{aligned}$$

- II. La selección del estadístico de prueba (p.ej. Z)
- III. Determinar la distribución de probabilidad del estadístico de prueba (p. ej. $t = \frac{z_1\sqrt{k}}{\sqrt{z_2}}$)
- IV. Escoger el nivel de significancia de la prueba α (comúnmente se utiliza un nivel $\alpha = 0.05$).
- V. Identificar la región de aceptación o rechazo de la distribución escogida y calcular el estadístico de prueba.
- VI. Aceptar o rechazar H_0 de acuerdo a la región en la que se encuentre el estadístico de prueba.

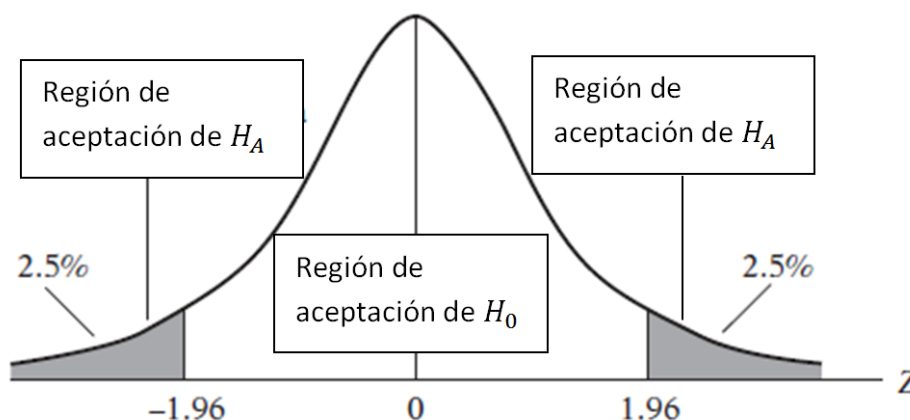


Ilustración 20. Regiones de aceptación y rechazo de H_0 para el estadístico de prueba con distribución normal Z y un nivel de significancia de 5%. (Gujarati y Porter 2009)

Prueba de Breush-Pagan

La prueba de Breusch-Pagan para efectos aleatorios (Montero G. 2011); consiste en realizar la regresión $y_{it} = x'_{it}\beta + u_i + e_{it}$, y en donde la hipótesis nula es $var(u_i) = 0$ con una χ^2 de contraste.

Si se rechaza la hipótesis nula, implica que efectivamente existe un componente inobservable de la varianza asociada a cada individuo y que el resultado de la regresión estará sesgado. Es decir el test de regresión anidada versus regresión agrupada (pooled) consiste en estimar si cada individuo tiene un origen en ordenadas distinto mediante la estimación de si (u_i) tiene una distribución distinta de cero. Nótese que tanto en el caso de efectos fijos (donde u_i tiene un valor constante para cada individuo pero una distribución para toda la muestra) como en el caso de efectos variables (donde u_i tiene una distribución para cada individuo) u_i siempre tiene que tener una cierta distribución (un valor y una desviación). Ojo lo importante no es que tenga un valor, ya que el valor fijo se estima en la constante del modelo, sino que lo relevante es que tenga una varianza, una distribución, significativamente distinta de cero (Montero G. 2011). De esta forma se contrasta la estimación Pooled OLS con la estimación de efectos aleatorios.

Prueba de Hausman

Al momento de especificar un modelo, juega un papel muy importante la existencia de correlación entre los regresores y los términos de error, por lo que no resulta conveniente establecer que $E[v_{it}|x_{it}] = 0$ (para el modelo $y_{it} = x'_{it}\beta + v_{it}$ de efectos aleatorios), puesto que $v_{it} = \alpha_i + \varepsilon_{it}$ contiene al efecto fijo inobservado, que puede estar correlacionado con los regresores x_{it} y de esta forma conducir a estimadores inconsistentes. Cuando $E[\alpha_i|x_{it}] \neq 0$, solo el estimador de efectos fijos es consistente, mientras que, bajo $H_0: E[\alpha_i|x_{it}] = 0$, tanto el estimador de efectos fijos y efectos aleatorios son consistentes, siendo el estimador de efectos aleatorios el estimador lineal insesgado óptimo (Cobacho Tornel y Bosch Mossi 2005).

Un contraste de Hausman se utiliza para analizar la posible correlación entre los α_i y los regresores y poder así decidir entre una estimación por efectos fijos o efectos aleatorios. Bajo $H_0: E[\alpha_i|x_{it}] = 0$, el estadístico de Hausman, converge en distribución a una χ^2_{NT} :

$$Q_{EF,EA} = (\widehat{\beta}_{EF} - \widehat{\beta}_{EA})'(\widehat{\sigma}_{EF}^2 - \widehat{\sigma}_{EA}^2)^{-1}(\widehat{\beta}_{EF} - \widehat{\beta}_{EA})$$

Donde $Q_{EF,EA}$ es el estadístico de prueba de Hausman, $\widehat{\beta}_{EF}$ es el vector de coeficientes del estimador de efectos fijos (consistente bajo H_0), $\widehat{\beta}_{EA}$ es el vector de coeficientes del estimador de efectos aleatorios (consistente y eficiente bajo H_0), $\widehat{\sigma}_{EF}^2$ la matriz de covarianzas del estimador de EF y $\widehat{\sigma}_{EA}^2$ la matriz de covarianzas del estimador de EA (Hausman specification test s.f.).

De esta forma si se rechaza la hipótesis nula, el mejor estimador es el de efectos fijos, si se acepta el mejor estimador es el de efectos aleatorios.

Anexo D

dturbosina	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
inpc	10429.72	2788.649	3.74	0.000	4962.436 15897
pib	-.3720437	.1101254	-3.38	0.001	-.58795 -.1561374
logppetroleo	-214637.9	145005.7	-1.48	0.139	-498928.8 69652.97
logpturbosina	272354.2	206037.6	1.32	0.186	-131592.6 676301
operaciones	1386.036	35.43286	39.12	0.000	1316.568 1455.504
pasajeros	26.47225	.3462616	76.45	0.000	25.79339 27.15111
carga	-.3893738	.0402641	-9.67	0.000	-.4683136 -.3104339
caphotelera	-419.86	39.6809	-10.58	0.000	-497.6564 -342.0637
aca	4045230	638948.2	6.33	0.000	2792541 5297920
tgz	-789528.3	175777.7	-4.49	0.000	-1134149 -444907.4
agu	-235927.9	180676.6	-1.31	0.192	-590153.2 118297.4
bjx	-739442.9	211047.8	-3.50	0.000	-1153213 -325673.3
cjs	-516732.9	182050.2	-2.84	0.005	-873651.3 -159814.5
clq	-766355.4	179097	-4.28	0.000	-1117484 -415227
cpe	-421834.5	175115.7	-2.41	0.016	-765157.4 -78511.52
ctm	-341919.8	172209	-1.99	0.047	-679544 -4295.575
cul	-4145292	198064.6	-20.93	0.000	-4533607 -3756977
cun	1.75e+07	1107644	15.77	0.000	1.53e+07 1.96e+07
cuu	-2485881	194407.2	-12.79	0.000	-2867026 -2104736
cvj	-773153.2	177609.1	-4.35	0.000	-1121365 -424941.8
czm	-29045.2	192957.5	-0.15	0.880	-407347.9 349257.5
dgo	-1781361	175884.9	-10.13	0.000	-2126192 -1436530
gdl	-11520.46	689587.1	-0.02	0.987	-1363490 1340449
hmo	-3575957	198497.1	-18.02	0.000	-3965120 -3186793
hux	-12800.15	171614.7	-0.07	0.941	-349259.2 323658.9
lap	-1697123	175571.3	-9.67	0.000	-2041339 -1352907
lmm	-1630921	177579.7	-9.18	0.000	-1979074 -1282767
lto	-844984.8	182721	-4.62	0.000	-1203218 -486751.3
mex	3.49e+07	2060927	16.93	0.000	3.08e+07 3.89e+07
mid	126402.2	216121.6	0.58	0.559	-297314.8 550119.1
mlm	-40665.51	181536.4	-0.22	0.823	-396576.5 315245.5
mty	-4638688	452012.8	-10.26	0.000	-5524881 -3752495
mxl	-413107.3	170889.6	-2.42	0.016	-748144.7 -78069.79
mzt	1085257	315943.2	3.43	0.001	465835 1704679
oax	-256237.6	207361.2	-1.24	0.217	-662779.4 150304.1
pbm	172378.2	224701.1	0.77	0.443	-268159.2 612915.7
pqm	273632.7	170528.4	1.60	0.109	-60696.55 607962
pvr	1074472	398959.2	2.69	0.007	292292.4 1856651
pxm	-421207.6	170732	-2.47	0.014	-755936 -86479.17
qro	138125.3	198126.9	0.70	0.486	-250312.3 526562.8
sjd	-1918734	185901.3	-10.32	0.000	-2283203 -1554266
slp	-41060.55	197863.8	-0.21	0.836	-428982.2 346861.1
tam	-2054588	183560.6	-11.19	0.000	-2414467 -1694708
tap	-511023.3	171821.7	-2.97	0.003	-847888.3 -174158.4
tij	1017952	216598.7	4.70	0.000	593299.7 1442604
tlc	-4394849	279797.1	-15.71	0.000	-4943405 -3846293
ver	-728417.2	321093.7	-2.27	0.023	-1357937 -98897.22
vsa	-1459028	179853.9	-8.11	0.000	-1811640 -1106415
zcl	-93892.54	172070.3	-0.55	0.585	-431244.9 243459.8
zih	130962.2	201697.2	0.65	0.516	-264475.1 526399.4
zlo	327844.5	176532	1.86	0.063	-18255.18 673944.3
_cons	1405239	456908.2	3.08	0.002	509447.7 2301030

Modelo Efectos Fijos 2007-2015

Tabla 19. Modelo de efectos fijos para el periodo 2007-2015

Series de Errores Estándar de Estimación

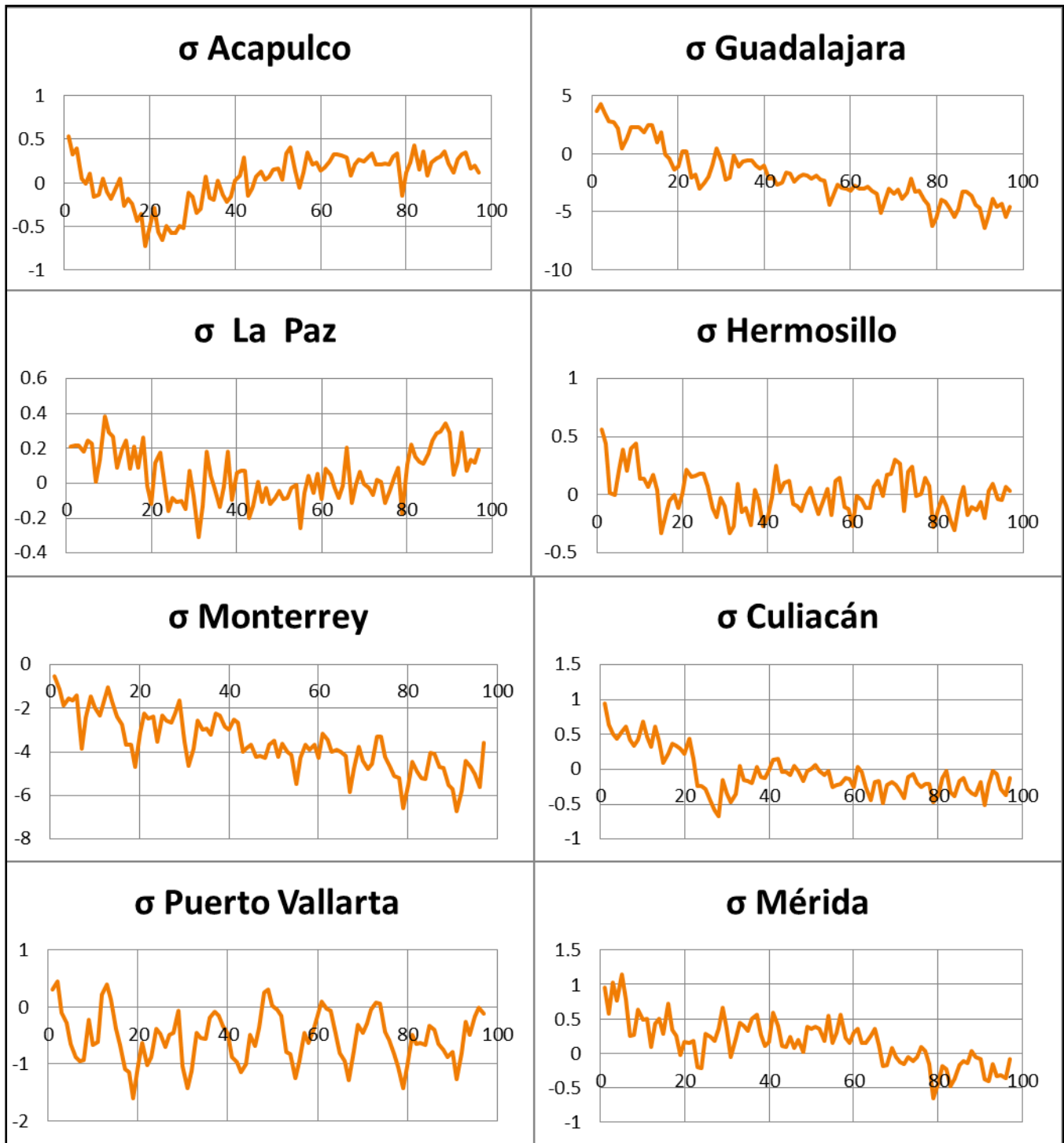


Ilustración 21. Errores estándar de estimación para aeropuertos seleccionados, en el periodo 2007-2015. La mayoría de los aeropuertos muestran comportamientos atípicos durante los primeros 36 meses (periodo 2007-2009) que coincide con el periodo de crisis económica a nivel mundial.

Pruebas de Hausman no Concluyentes.

Tabla 20. Ejemplos de la inconsistencia de la Prueba de Hausman para el modelo $D_{turbosina} = \alpha_i + \beta_1 INPC + \beta_2 PIB + \beta_3 \text{Log}(P_{petroleo}) + \beta_4 \text{Log}(P_{turbosina}) + \beta_5 \text{Operaciones} + \beta_6 \text{Pasajeros} + \beta_7 \text{Carga} + \beta_8 \text{CapHotelera} + \text{uit. (Periodo 2010-2015)}$

Note: the rank of the differenced variance matrix (4) does not equal the number of coefficients being tested (8); be sure this is what you expect, or there may be problems computing the test. Examine the output of your estimators for anything unexpected and possibly consider scaling your variables so that the coefficients are on a similar scale.

	Coefficients		(b-B) Difference	sqrt(diag(V _b -V _B)) S.E.
	(b) fixed	(B) random		
inpc	22251.57	14407.03	7844.539	641.1378
pib	-.9660299	-1.249464	.283434	.0202377
logppetro1~m	-964102.2	-1006257	42155.05	3104.1
logpturbos~a	1315909	1296222	19686.77	5927.922
operaciones	956.462	1050.5	-94.03836	14.03129
pasajeros	24.65227	23.6326	1.019669	.1179694
carga	-.1235155	-.0372387	-.1607542	.0157155
caphotelera	-616.0704	135.2823	-751.3527	57.13101

b = consistent under Ho and Ha; obtained from xtreg
B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

$$\begin{aligned} \chi^2(4) &= (b-B)'[(V_b-V_B)^{-1}](b-B) \\ &= 224.05 \\ \text{Prob}>\chi^2 &= 0.0000 \\ (V_b-V_B &\text{ is not positive definite}) \end{aligned}$$

Note: the rank of the differenced variance matrix (4) does not equal the number of coefficients being tested (8); be sure this is what you expect, or there may be problems computing the test. Examine the output of your estimators for anything unexpected and possibly consider scaling your variables so that the coefficients are on a similar scale.

	Coefficients		(b-B) Difference	sqrt(diag(V _b -V _B)) S.E.
	(b) fixed	(B) random		
inpcm100	222.5157	144.0703	78.44539	6.411378
pib	-.96603	-1.249464	.283434	.0202377
logppetro1~m	-964102.2	-1006257	42155.04	3104.1
logpturbos~a	1315909	1296222	19686.79	5927.923
operacio~m10	95.6462	105.05	-9.403836	1.403129
pasajeros	24.65227	23.6326	1.019669	.1179694
carga	-.1235155	-.0372387	-.1607542	.0157155
caphotel~m10	-61.60705	13.52823	-75.13527	5.713101

b = consistent under Ho and Ha; obtained from xtreg
B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

$$\begin{aligned} \chi^2(4) &= (b-B)'[(V_b-V_B)^{-1}](b-B) \\ &= 224.05 \\ \text{Prob}>\chi^2 &= 0.0000 \\ (V_b-V_B &\text{ is not positive definite}) \end{aligned}$$

Note: the rank of the differenced variance matrix (4) does not equal the number of coefficients being tested (8); be sure this is what you expect, or there may be problems computing the test. Examine the output of your estimators for anything unexpected and possibly consider scaling your variables so that the coefficients are on a similar scale.

	Coefficients		(b-B) Difference	sqrt(diag(V _b -V _B)) S.E.
	(b) fixed	(B) random		
inpcm100	219.2638	140.792	78.47177	6.404385
pibd100	-96.65721	-125.0031	28.34589	2.023527
logppe~m1.000	-984.4749	-1027.337	42.86205	3.20777
logptu~m1.000	1346.025	1326.788	19.23724	6.178602
operacio~m10	95.6168	105.0189	-9.402097	1.402818
pasajeros	24.65602	23.6364	1.019613	.1179524
cargad100	-12.317	3.755137	-16.07214	1.571422
caphotel~m10	-61.63235	13.50702	-75.13937	5.712697

b = consistent under Ho and Ha; obtained from xtreg
B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

$$\begin{aligned} \chi^2(4) &= (b-B)'[(V_b-V_B)^{-1}](b-B) \\ &= 224.07 \\ \text{Prob}>\chi^2 &= 0.0000 \\ (V_b-V_B &\text{ is not positive definite}) \end{aligned}$$

Estimadores del modelo: $D_{turbosina} = \alpha_i + \beta_1 INPC + \beta_2 PIB + \beta_3 \text{Log}(P_{petroleo}) + \beta_4 \text{Log}(P_{turbosina}) + \beta_5 \text{Operaciones} + \beta_6 \text{Pasajeros} + \beta_7 \text{Carga} + \beta_8 \text{CapHotelera} + u_{it}$. (Periodo 2010-2015).

Tabla 21. Estimación Pooled OLS para el modelo $D_{turbosina} = \alpha_i + \beta_1 INPC + \beta_2 PIB + \beta_3 \text{Log}(P_{petroleo}) + \beta_4 \text{Log}(P_{turbosina}) + \beta_5 \text{Operaciones} + \beta_6 \text{Pasajeros} + \beta_7 \text{Carga} + \beta_8 \text{CapHotelera} + u_{it}$. (Periodo 2010-2015).

Source	SS	df	MS				
Model	9.1969e+17	8	1.1496e+17	Number of obs =	2623		
Residual	1.5264e+16	2614	5.8393e+12	F(8, 2614) =	19687.51		
Total	9.3495e+17	2622	3.5658e+14	Prob > F =	0.0000		
				R-squared =	0.9837		
				Adj R-squared =	0.9836		
				Root MSE =	2.4e+06		

dturbosina	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
inpc	1469.535	15906.7	0.09	0.926	-29721.46	32660.53
plib	-1.839934	.5160679	-3.57	0.000	-2.851877	-.8279905
logppetrol~m	-1659805	873086.5	-1.90	0.057	-3371816	52205.41
logpturbos~a	2119075	1114457	1.90	0.057	-66232.28	4304382
operaciones	-264.9537	47.67919	-5.56	0.000	-358.4465	-171.4609
pasajeros	41.32776	.4847989	85.25	0.000	40.37713	42.27839
carga	.5044145	.0272922	18.48	0.000	.450898	.5579309
caphotelera	53.47976	14.3398	3.73	0.000	25.36125	81.59826
_cons	9709368	2648044	3.67	0.000	4516892	1.49e+07

Tabla 22. Estimación de Efectos Fijos para el modelo $D_{turbosina} = \alpha_i + \beta_1 INPC + \beta_2 PIB + \beta_3 \text{Log}(P_{petroleo}) + \beta_4 \text{Log}(P_{turbosina}) + \beta_5 \text{Operaciones} + \beta_6 \text{Pasajeros} + \beta_7 \text{Carga} + \beta_8 \text{CapHotelera} + u_{it}$. (Periodo 2010-2015)

Fixed-effects (within) regression	Number of obs =	2623	
Group variable: id	Number of groups =	43	
R-sq: within = 0.8672	Obs per group: min =	61	
between = 0.9421	avg =	61.0	
overall = 0.9303	max =	61	
corr(u_i, Xb) = 0.8600	F(8, 2572) =	2100.25	
	Prob > F =	0.0000	

dturbosina	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
inpc	22251.57	6011.625	3.70	0.000	10463.45	34039.69
plib	-.9660299	.1939686	-4.98	0.000	-1.34638	-.5856793
logppetrol~m	-964102.2	324780.1	-2.97	0.003	-1600959	-327245.3
logpturbos~a	1315909	414636.9	3.17	0.002	502852.5	2128965
operaciones	956.462	54.11811	17.67	0.000	850.3425	1062.581
pasajeros	24.65227	.4553193	54.14	0.000	23.75944	25.5451
carga	-.1235155	.0407714	-3.03	0.002	-.2034635	-.0435674
caphotelera	-616.0704	69.63727	-8.85	0.000	-752.6213	-479.5196
_cons	6700101	1022527	6.55	0.000	4695041	8705161
sigma_u	9727912.2					
sigma_e	897897.39					
rho	.99155247	(fraction of variance due to u_i)				

F test that all u_i=0: F(42, 2572) = 389.54 Prob > F = 0.0000

Tabla 23. Estimación de Efectos Aleatorios para el modelo $D_{turbosina} = \alpha_i + \beta_1 INPC + \beta_2 PIB + \beta_3 \text{Log}(P_{petroleo}) + \beta_4 \text{Log}(P_{turbosina}) + \beta_5 \text{Operaciones} + \beta_6 \text{Pasajeros} + \beta_7 \text{Carga} + \beta_8 \text{CapHotelera} + u_{it}$. (Periodo 2010-2015)

Random-effects GLS regression	Number of obs =	2623	
Group variable: id	Number of groups =	43	
R-sq: within = 0.8604	Obs per group: min =	61	
between = 0.9793	avg =	61.0	
overall = 0.9769	max =	61	
Random effects u_i ~ Gaussian	wald chi2(8) =	18223.66	
corr(u_i, X) = 0 (assumed)	Prob > chi2 =	0.0000	
theta = .94687839			

dturbosina	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
inpc	14407.03	6223.846	2.31	0.021	2208.516	26605.54
plib	-1.249464	.2008656	-6.22	0.000	-1.643153	-.8557744
logppetrol~m	-1006257	338158.6	-2.98	0.003	-1669036	-343478.5
logpturbos~a	1296222	431692.5	3.00	0.003	450120	2142323
operaciones	1050.5	54.42303	19.30	0.000	943.8332	1157.168
pasajeros	23.6326	.4579077	51.61	0.000	22.73512	24.53008
carga	-.0372387	.0391723	0.95	0.342	-.0395376	-.1140151
caphotelera	135.2823	41.45912	3.26	0.001	54.02388	216.5407
_cons	3991988	1097232	3.64	0.000	1841453	6142523
sigma_u	2161110.2					
sigma_e	897897.39					
rho	.85278874	(fraction of variance due to u_i)				

Series Logarítmicas del Gasto de Turbosina

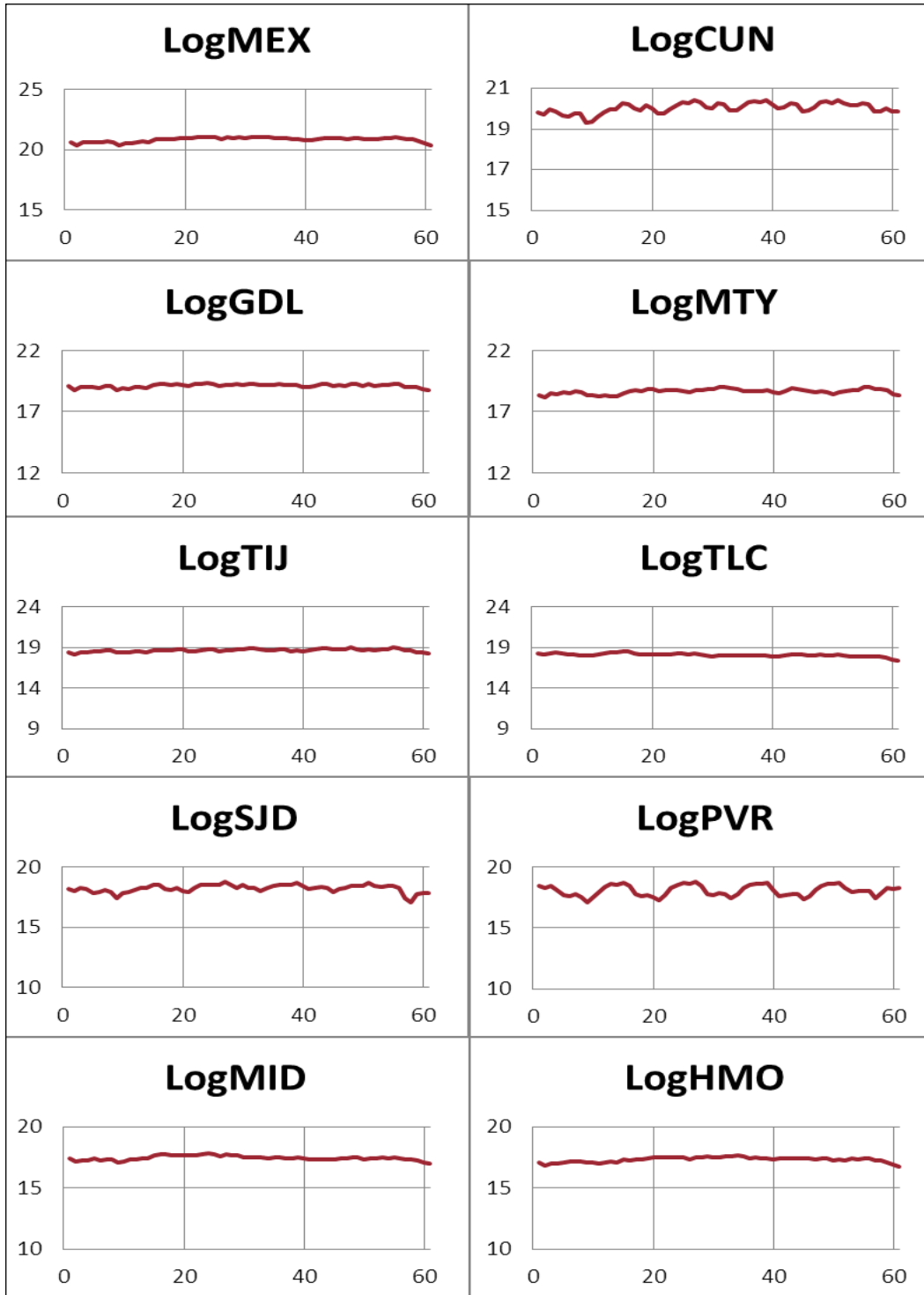


Ilustración 22. Series logarítmicas del gasto de turbosina para los 10 principales aeropuertos de la República Mexicana, para el periodo 2010-2015

Mejor estimador para el modelo: $\text{Log}(\text{Gasto de turbosina}) = \alpha_i + \beta_1 \text{INPC} + \beta_2 \text{Operaciones} + \beta_3 \text{Pasajeros} + \beta_4 \text{Carga} + \beta_5 \text{CapHotelera}$

Se calculan los tres estimadores, *Pooled OLS*, Efectos Fijos y Efectos Aleatorios, para posteriormente realizar las pruebas correspondientes para identificar al mejor estimador.

Tabla 24. Estimación Pooled OLS realizada en Stata

Source	SS	df	MS			
Model	4446.99352	5	889.398703	Number of obs =	2623	
Residual	2891.25759	2617	1.10479847	F(5, 2617) =	805.03	
				Prob > F =	0.0000	
				R-squared =	0.6060	
				Adj R-squared =	0.6052	
Total	7338.2511	2622	2.79872277	Root MSE =	1.0511	

loggastotu~a	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
inpc	.0070098	.0034545	2.03	0.043	.000236	.0137835
operaciones	.0006367	.0000207	30.71	0.000	.000596	.0006773
pasajeros	-1.69e-06	2.11e-07	-8.02	0.000	-2.10e-06	-1.28e-06
carga	-3.40e-07	1.19e-08	-28.72	0.000	-3.64e-07	-3.17e-07
caphotelera	.0000377	6.23e-06	6.04	0.000	.0000254	.0000499
_cons	14.39712	.3656152	39.38	0.000	13.6802	15.11404

Tabla 25. Estimación de Efectos Fijos realizada en Stata

Fixed-effects (within) regression	Number of obs =	2623
Group variable: id	Number of groups =	43
R-sq: within = 0.1052	Obs per group: min =	61
between = 0.0353	avg =	61.0
overall = 0.0381	max =	61
corr(u_i, Xb) = 0.0265	F(5, 2575) =	60.57
	Prob > F =	0.0000

loggastotu~a	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
inpc	.0160781	.001252	12.84	0.000	.013623	.0185331
operaciones	.0001233	.0000197	6.27	0.000	.0000847	.0001618
pasajeros	-2.38e-07	1.66e-07	-1.43	0.152	-5.63e-07	8.72e-08
carga	6.61e-09	1.48e-08	0.45	0.654	-2.23e-08	3.56e-08
caphotelera	-.0000616	.0000252	-2.45	0.014	-.0001109	-.0000123
_cons	14.84184	.1563871	94.90	0.000	14.53519	15.1485
sigma_u	1.6276762					
sigma_e	.32740787					
rho	.96111195	(fraction of variance due to u_i)				

F test that all u_i=0: F(42, 2575) = 580.87 Prob > F = 0.0000

Tabla 26. Estimación de Efectos Aleatorios realizada en Stata

```

Random-effects GLS regression           Number of obs   =   2623
Group variable: id                    Number of groups =    43

R-sq:  within = 0.1015                Obs per group:  min =    61
      between = 0.4807                  avg   =   61.0
      overall  = 0.4634                  max   =    61

Random effects u_i ~ Gaussian          Wald chi2(5)    =   333.91
corr(u_i, X) = 0 (assumed)            Prob > chi2     =    0.0000
theta = .95920524
    
```

loggastotu~a	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
inpc	.0142168	.0011806	12.04	0.000	.0119029	.0165308
operaciones	.0001385	.0000193	7.19	0.000	.0001008	.0001763
pasajeros	-3.82e-07	1.62e-07	-2.35	0.019	-7.00e-07	-6.39e-08
carga	1.29e-08	1.41e-08	0.91	0.360	-1.47e-08	4.04e-08
caphotelera	.0000184	.0000166	1.11	0.268	-.0000142	.000051
_cons	14.50981	.2076901	69.86	0.000	14.10274	14.91687
sigma_u	1.0267343					
sigma_e	.32740787					
rho	.90769948	(fraction of variance due to u_i)				

Tanto las estimaciones de efectos fijos como aleatorios, tiene estimaciones de los coeficientes muy parecidas entre sí, en comparación con las observadas en la estimación *Pooled*.

En la prueba F que realiza la estimación de efectos fijos cuya hipótesis nula es $\alpha_i = 0$, el valor de p es cero, por lo que H_0 se rechaza, lo que indica que algunos interceptos son significativos y no se puede usar un modelo *Pooled OLS* (que asume un intercepto común a todos los individuos). Dicha prueba desestima al modelo *Pooled OLS*.

Adicionalmente se corrió la prueba de Breush-Pagan que contrasta al modelo de efectos aleatorios con el modelo *Pooled OLS*, tiene los mismos resultados que la prueba anterior, ya que desestima al modelo *Pooled OLS*. El valor de p es cero por lo que se rechaza la hipótesis nula.

Tabla 27. Prueba de Breush-Pagan que compara la estimación de Efectos Aleatorios con la estimación *Pooled OLS*. Desestima al modelo *Pooled OLS*.

Breusch and Pagan Lagrangian multiplier test for random effects

$$\text{loggastoturbo}[\text{id},t] = Xb + u[\text{id}] + e[\text{id},t]$$

Estimated results:

	Var	sd = sqrt(Var)
loggast~a	2.798723	1.672938
e	.1071959	.3274079
u	1.054183	1.026734

Test: Var(u) = 0

chi2(1) = 55307.33
 Prob > chi2 = 0.0000

Con ambos resultados se procedió a descartar al modelo *Pooled OLS* como mejor estimador y se procedió a revisar la prueba de Hausman para seleccionar entre efectos fijos o aleatorios.

La hipótesis nula dice que no hay diferencia entre ambos estimadores y en caso de ser aceptada se escoge al estimador de efectos aleatorios (en este caso a pesar de que ambos estimadores son consistentes, el estimador de efectos aleatorios es más eficiente). La hipótesis alternativa indica que si hay diferencia entre ambas estimaciones y que la estimación de efectos aleatorios es inconsistente. Por lo anterior, en caso de ser aceptada esta hipótesis, se escogería al estimador de efectos fijos como verdadero

Tabla 28. Prueba de Hausman con los coeficientes sin escalar

Note: the rank of the differenced variance matrix (3) does not equal the number of coefficients being tested (5); be sure this is what you expect, or there may be problems computing the test. Examine the output of your estimators for anything unexpected and possibly consider scaling your variables so that the coefficients are on a similar scale.

	— Coefficients —		(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
	(b) fixed	(B) random		
inpc	.0160781	.0142168	.0018612	.0004364
operaciones	.0001233	.0001385	-.0000153	4.39e-06
pasajeros	-2.38e-07	-3.82e-07	1.44e-07	3.81e-08
carga	6.61e-09	1.29e-08	-6.25e-09	4.74e-09
caphoteleria	-.0000616	.0000184	-.00008	.000019

b = consistent under Ho and Ha; obtained from xtreg
 B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

$$\begin{aligned} \chi^2(3) &= (b-B)'[(V_b-V_B)^{-1}](b-B) \\ &= 19.89 \\ \text{Prob}>\chi^2 &= 0.0002 \end{aligned}$$

La Prueba de Hausman rechazó la hipótesis nula (de que no hay diferencia entre los estimadores), por lo que el modelo más apropiado es el de efectos fijos. Sin embargo, al principio hubo una nota que indicó que no todos los coeficientes del modelo fueron utilizados para realizar la prueba, debido a que la escala de los coeficientes varía mucho. Como se puede observar el coeficiente de INPC es muy grande comparado con los demás; caso contrario al de la carga que es un coeficiente muy pequeño.

Por esta razón se decidió reducir al coeficiente del INPC, así la variable INPC fue multiplicada por la constante 100.

Tabla 29. Prueba de Hausman con los coeficientes escalados.

	— Coefficients —		(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
	(b) fixed	(B) random		
inpcm100	.0001608	.0001422	.0000186	4.36e-06
operaciones	.0001233	.0001385	-.0000153	4.39e-06
pasajeros	-2.38e-07	-3.82e-07	1.44e-07	3.81e-08
carga	6.61e-09	1.29e-08	-6.25e-09	4.74e-09
caphoteleria	-.0000616	.0000184	-.00008	.000019

b = consistent under Ho and Ha; obtained from xtreg
 B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

$$\begin{aligned} \chi^2(5) &= (b-B)'[(V_b-V_B)^{-1}](b-B) \\ &= 37.02 \\ \text{Prob}>\chi^2 &= 0.0000 \end{aligned}$$

Al volver a correr la prueba de Hausman, la hipótesis nula fue rechazada nuevamente; por lo que el mejor estimador es el de efectos fijos.

Como se puede observar en la estimación de efectos fijos, Stata calcula una constante que es el promedio de los interceptos individuales, mismo que es estadísticamente significativo. Esto, reforzado con el hecho que la hipótesis nula de la prueba global F de significancia para los efectos fijos es rechazada, permite confirmar el estatus de “mejor estimador” al modelo de efectos fijos.

Combinaciones de modelos (variables Operaciones, Carga y CapHotelera)

Tabla 30. Estimación de efectos fijos para el modelo Log (Gasto de turbosina) = $\alpha_i + \beta_1$ INPC + β_2 Operaciones + β_3 Carga + β_4 CapHotelera

```
Fixed-effects (within) regression
Group variable: id
Number of obs      =    2623
Number of groups   =     43
R-sq:  within = 0.1045
      between = 0.0004
      overall  = 0.0001
Obs per group: min =     61
              avg  =    61.0
              max  =     61
corr(u_i, Xb) = -0.1547
F(4, 2576)        =    75.17
Prob > F          =    0.0000
```

loggastotu~a	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
inpc	.0159729	.0012501	12.78	0.000	.0135216	.0184243
operaciones	.0001	.0000111	9.02	0.000	.0000782	.0001217
carga	5.19e-09	1.47e-08	0.35	0.725	-2.37e-08	3.41e-08
caphotelera	-.0000678	.0000248	-2.74	0.006	-.0001164	-.0000192
_cons	14.92315	.145786	102.36	0.000	14.63728	15.20902
sigma_u	1.6803963					
sigma_e	.32747512					
rho	.96341147	(fraction of variance due to u_i)				

F test that all u_i=0: F(42, 2576) = 596.38 Prob > F = 0.0000

Tabla 31. Estimación de efectos fijos para el modelo Log (Gasto de turbosina) = $\alpha_i + \beta_1$ INPC + β_2 Operaciones + β_3 Pasajeros + β_4 CapHotelera

```
Fixed-effects (within) regression
Group variable: id
Number of obs      =    2623
Number of groups   =     43
R-sq:  within = 0.1052
      between = 0.0174
      overall  = 0.0211
Obs per group: min =     61
              avg  =    61.0
              max  =     61
corr(u_i, Xb) = -0.0179
F(4, 2576)        =    75.69
Prob > F          =    0.0000
```

loggastotu~a	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
inpc	.0160724	.0012517	12.84	0.000	.0136179	.0185269
operaciones	.0001233	.0000197	6.28	0.000	.0000848	.0001619
pasajeros	-2.33e-07	1.65e-07	-1.41	0.159	-5.57e-07	9.14e-08
caphotelera	-.0000614	.0000251	-2.44	0.015	-.0001108	-.0000121
_cons	14.85017	.1552532	95.65	0.000	14.54574	15.15461
sigma_u	1.6422629					
sigma_e	.32735707					
rho	.96178481	(fraction of variance due to u_i)				

F test that all u_i=0: F(42, 2576) = 783.46 Prob > F = 0.0000

Tabla 32. Estimación de efectos fijos para el modelo $\text{Log}(\text{Gasto de turbosina}) = \alpha_i + \beta_1 \text{INPC} + \beta_2 \text{Pasajeros} + \beta_3 \text{Carga} + \beta_4 \text{CapHotelera}$

Fixed-effects (within) regression		Number of obs	=	2623	
Group variable: id		Number of groups	=	43	
R-sq: within	= 0.0916	Obs per group: min	=	61	
between	= 0.3102	avg	=	61.0	
overall	= 0.2402	max	=	61	
corr(u_i, Xb)	= -0.6191	F(4, 2576)	=	64.92	
		Prob > F	=	0.0000	
loggastotu~a	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
inpc	.0159191	.001261	12.62	0.000	.0134464 .0183919
pasajeros	6.21e-07	9.42e-08	6.59	0.000	4.36e-07 8.06e-07
carga	7.25e-09	1.49e-08	0.49	0.626	-2.19e-08 3.64e-08
caphotelera	-.000064	.0000253	-2.53	0.012	-.0001137 -.0000143
_cons	15.09507	.1522054	99.18	0.000	14.79661 15.39353
sigma_u	1.8306069				
sigma_e	.3298351				
rho	.96855667	(fraction of variance due to u_i)			
F test that all u_i=0:		F(42, 2576) =	799.46	Prob > F =	0.0000

En todas las tablas anteriores, al excluir algunas de las variables Operaciones, Carga y CapHotelera, de las dos restantes una presentó ser estadísticamente insignificantes.

Indicadores para un modelo Lin-Lin

Aeropuerto	σ	MAPE
Acapulco	\$ 4,504,615.06	21.92%
Aguascalientes	\$ 1,574,962.58	12.35%
Bahías de Huatulco	\$ 3,551,165.22	26.99%
Bajío (León,Gto)	\$ 5,017,397.36	15.50%
Campeche	\$ 797,792.74	13.75%
Cancún	\$ 90,404,422.79	14.31%
Cd. Juárez	\$ 3,468,451.21	17.23%
Chetumal	\$ 1,330,652.67	22.32%
Chihuahua	\$ 3,852,166.07	16.75%
Colima	\$ 529,736.23	20.63%
Cozumel	\$ 3,096,969.43	20.01%
Cuernavaca	\$ 2,633,543.45	139.04%
Culiacán	\$ 3,729,541.80	10.69%
Durango	\$ 1,210,715.79	16.21%
Guadalajara	\$ 31,603,829.17	12.53%
Hermosillo	\$ 5,798,015.20	14.49%
La Paz	\$ 3,028,435.57	16.32%
Loreto	\$ 1,454,082.47	102.93%
Los Mochis	\$ 1,244,818.25	15.19%
Manzanillo	\$ 1,466,016.72	19.96%
Mazatlán	\$ 3,537,242.08	15.40%
Mérida	\$ 6,984,206.28	14.19%
Mexicali	\$ 2,291,541.73	12.40%
México	\$ 173,540,471.20	12.57%
Monterrey	\$ 20,782,504.07	12.51%
Morelia	\$ 2,616,630.99	13.69%
Oaxaca	\$ 1,658,620.79	14.12%
Palenque	\$ 946,475.82	1840.42%
Puebla	\$ 3,763,913.47	32.29%
Puerto Escondido	\$ 1,098,059.38	71.29%
Puerto Vallarta	\$ 19,117,733.26	20.99%
Querétaro	\$ 6,147,379.10	46.96%
San José Del Cabo	\$ 17,637,384.52	17.59%
San Luis Potosí	\$ 2,371,102.08	17.85%
Tampico	\$ 2,008,946.27	17.80%
Tapachula	\$ 1,100,756.05	15.81%
Tijuana	\$ 18,054,047.48	11.79%
Toluca	\$ 12,580,310.64	13.15%
Tuxtla Gutiérrez	\$ 3,299,264.01	16.15%
Veracruz	\$ 4,003,767.24	16.83%
Villahermosa	\$ 3,178,001.64	11.37%
Zacatecas	\$ 1,126,741.98	11.39%
Zihuatanejo	\$ 2,992,467.98	22.86%
Total general	\$ 481,134,897.83	13.88%

Tabla 33. Indicadores de desempeño para el modelo Lin-Lin de ecuación: Gasto de turbosina = $\alpha_i + \beta_1$ INPC + β_2 Pasajeros + β_3 Carga + β_4 CapHotelera

Referencias

- Baum, Christopher F. *An Introduction to Modern Econometrics Using Stata*. Stata Press, 2006.
- Aerolíneas Mexicanas*. 2014. <http://aerolineasmexicanas.mx/> (accessed Febrero 25, 2015).
- Aeropuertos y Servicios Auxiliares*. n.d. <http://www.asa.gob.mx/> (accessed Febrero 12, 2015).
- Airbus*. 2015. <http://www.airbus.com/company/history/the-narrative/> (accessed Marzo 5, 2015).
- Alvarez De Toledo, Pablo , Adolfo Crespo , Fernando Núñez, and Carlos Usabiaga. "Introducción de elementos autorregresivos." *Revista de Dinámica de Sistemas en modelos de dinámica de sistemas* 2, no. 1 (Marzo 2006): 37-66.
- Baltagi , Badi H. "Panel Data Methods." In *Handbook of Applied Economic Statistics*. Nueva York: CRC Press, 1998.
- Banco de México*. 2015. <http://www.banxico.org.mx> (accessed Febrero 9, 2015).
- Banco de México. *Documento*. México: Banco de México, 2011.
- Bingham, N.H., and John M. Fry. *Regression: Linear models in statics*. London: Springer, 2010.
- Cameron, A. Colin , and Pravin K. Trivedi. *Microeconometrics: Methods and Aplications*. Cambridge: Cambridge University Press, 2005.
- Cantera, Sara. "Aerolíneas padecen alza de combustible." *CNN Expansión*, Julio 2014.
- Catálogo de productos ECOPetrol VSM-01*. Técnico, combustibles, Brasilia: ECOPetrol, 2010.
- Cobacho Tornel, María Belén, and Mariano Bosch Mossi. "Contrastes de hipótesis en datos de panel." *XII Jornadas ASEPUMA*. Murcia, 2005.
- . "Métodos lineales de estimación con datos de panel: una aplicación al estudio de los efectos de la inversión en México." *XII Jornadas de ASEPUMA*. Murcia, 2004.
- Cuentame... economía*. 2015. <http://cuentame.inegi.org.mx/economia/> (accessed Febrero 09, 2015).
- Cuentame... población*. 2014. <http://cuentame.inegi.org.mx/poblacion/> (accessed Abril 20, 2015).
- Diario Oficial de la Federación*. Abril 25, 2013. <http://dof.gob.mx/> (accessed Febrero 10, 2015).
- Escalante Membrillo, María I. "Tesis: Pronósticos Mediante Análisis de Regresión". Ciudad de México: Facultad de Ingeniería, División de Estudios de Posgrado, 2005.
- Esteban, M. Victoria, M. Paz Moral, Susan Orbe, Marta Regúlez, Ainhoa Zarraga, and Marian Zubia. *Análisis de Regresión con Gretl*. España: Euskal Herriko Unibertsitatea, n.d.
- "Estudio-501 Aeropuertos y Servicios Auxiliares." Respuesta de la Unidad de Enlace ASA a la solicitud de información 0908500002807 , México, 2007.

- Frees, Edward W. . *Longitudinal and Panel Data: Analysis and Applications for the Social Sciences*. Cambridge: Cambridge University Press, 2004.
- García Morales, Irais . *T21 mx*. Julio 27, 2015. <http://t21.com.mx/aereo/2015/07/27/despega-trafico-carga-mexico-1s-ano> (accessed Agosto 9, 2015).
- García Pérez, Andres. *Apuntes de Econometría*. México: Escuela Nacional de Economía, UNAM, 1960.
- Gilliland , Michael . *Forecast Value Added Analysis: Step by Step*. Estados Unidos : SAS Institute Inc., 2013.
- Goldberger, Arthur S. *Econometric Theory*. John Wiley & Sons, 1964.
- Guerrero, Salvador . "Lenta expansión del transporte de carga aérea." *La Prensa*, Marzo 4, 2015.
- Gujarati, Damodar N., and Dawn C. Porter . *Econometría*. México: McGraw-Hill, 2009.
- Handout: Panel Data*. PAD 705 , Rockefeller College, University at Albany , 2008.
- Hausman specification test*. Manual , Stata.com, n.d.
- Hernández G., Silvina, and Susana C. Téllez B. *Apuntes de planeación y control de la demanda*. México: Facultad de Ingeniería, DIMEI., 2006.
- "IATA." *Air Cargo Ends 2014 on a Positive Note*. Febrero 4, 2015. <http://www.iata.org/pressroom/pr/Pages/2015-02-04-01.aspx> (accessed Marzo 11, 2015).
- Keynes, John Maynard . *Teoría general de la ocupación, el interés y el dinero*. Buenos Aires, Argentina: Fondo de Cultura Económica, 1992.
- Kuntz Ficker, Sandra . "¿Qué nos dejó la Revolución Mexicana?" México: Universidad de México, Noviembre 2002. 23-30.
- "Ley de Aeropuertos ." Diario Oficial de la Federación, 2015.
- Malinvaud, E. *Statistical Methods of Econometrics*. Rand McNally, 1966.
- Martínez Medina , Mayra. "Características de un crudo pesado y uno ligero." *Milenio Diario*, 01 06, 2014.
- Mccartney, Scott. "¿Cómo gastan las aerolíneas lo que usted paga por un pasaje?" *The Wall Street Journal*, Junio 15, 2012.
- Montero G., Roberto. *Efectos fijos o aleatorios: test de especificación*. Documentos de Trabajo en Economía Aplicada, España: Universidad de Granada, 2011.
- Mundial, Grupo del Banco. *El Banco Mundial BIRF-AIF*. 2015. <http://datos.bancomundial.org/indicador/NY.GDP.MKTP.CD> (accessed Febrero 8, 2015).
- Novalés, Alfonso. *Análisis de Regresión*. España: Departamento de Economía Cunitativa. Universidad Complutense, 2010.
- OECD*. 2015. <http://data.oecd.org/gdp/gross-domestic-product-gdp.htm> (accessed Febrero 8, 2015).

- Ordaz Aguilar, Erubiel . *Optimización del riesgo factorial* . Cuadernillo de Divulgación , México : Facultad de Ingeniería UNAM, 2005.
- Pabón , Carlos, Dario Peña, Carlos Mejía, and Haydee Quiroga. "Determinación de la corrosividad por ácido nafténico en un acero ATSM-A106 grado B, bajo condiciones de flijo dinámico." *Revista Lationamericana de Metalurgia y Materiales* (Universidad Simón Bolívar) "Volumen Suplemento" S1 (2009).
- PEMEX. *Anuario Estadístico*. PEMEX Dirección Corporativa de Finanzas, 2013.
- PEMEX Refinación . n.d. <http://www.ref.pemex.com/index.cfm?action=content§ionID=18&catID=1082> (accessed mayo 2014).
- PMI Comercio Internacional . Enero 6, 2015. <http://www.pmi.com.mx/> (accessed Febrero 10, 2015).
- Price, Lionel. *El análisis económico en un banco central. Modelos versus criterio personal*. Ensayo, México: Centro de Estudios Monetarios Latinoamericanos, 1998.
- Proyecto del Nuevo Aeropuerto de la Ciudad de Mexico. 2014. <http://www.aeropuerto.gob.mx/> (accessed Febrero 17, 2015).
- Redacción El Universal. "Listo el proyecto del nuevo aeropuerto de la ciudad de México." *El Universal*, Septiembre 3, 2014.
- Reuters. "OPEP eleva pronóstico de demanda de crudo para 2015." *El Economista*, Febrero 9, 2015.
- Ruiz Olmedo, Sergio A. *Énfasis Logística*. Junio 18, 2012. <http://www.logisticamx.enfasis.com/articulos/64301-limitaciones-la-carga-aerea-mexico> (accessed Marzo 11, 2015).
- Servicio Geológico Mexicano. 2015. <http://portalweb.sgm.gob.mx/economia/es/energeticos/695-seguimiento-precio-del-petroleo-mezcla-mexicana-mme-datos.html#mensual> (accessed Marzo 24, 2015).
- Stekler, H. O. "Forecasting- Yesterday, Today and Tomorrow." *International Journal of Forecasting*, 2014.
- Stonier, Alfred W., and Douglas C. Hague. *Manual de Teoría Económica*. Madrid-España: Aguilar S. A. de Ediciones, 1970.
- Wang, Haijiang Henry . *China's Oil Industry and Market*. Elsevier Ltd., 1999.