



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**DIVERSOS INDICADORES DE LAS FINANZAS Y DE
LA ECONOMÍA QUE INFLUYEN SOBRE EL INPC DE
MÉXICO**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

A C T U A R I O

P R E S E N T A:

HÉCTOR JAVIER CERVANTES RODRÍGUEZ



**DIRECTOR DE TESIS:
ACT. MARÍA AURORA VALDÉS MICHELL
2010**



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

DEDICATORIA

Dedico esta tesis a mis padres, Héctor Javier Cervantes Hernández y María Elena Rodríguez Rivera por darme la vida y principalmente por estar siempre a mi lado.

Gracias “Pa” por darme un ejemplo de vida y enseñarme a luchar por lo que se quiere y jamás rendirse.

Gracias “Ma” por siempre creer que puedo lograrlo y ayudarme a conseguir mis objetivos y estimularme a siempre terminar las cosas.

A mi hermano, César Cervantes Rodríguez, por darme su apoyo incondicional, por ser mi amigo y compartir grandes momentos con él.

A toda mi familia, mi tía Tere, mi tía “Chuy”, Gaby, Cris por estar conmigo siempre, en los mejores y en los peores momentos de mi vida, en las buenas y en las malas.

Un agradecimiento muy especial a Dany por darme tantas alegrías y ser esa “chispita” que ilumina el camino.

Un agradecimiento muy especial a mi abuela Alejandra Rivera Vargas, por haber compartido su tiempo con todos sus nietos y habernos inculcado grandes valores y un gran cariño a la familia.

Finalmente agradezco a todos aquellos que me han estado apoyando a lo largo de este proceso, algunos desde el inicio, otros un poco después pero finalmente todos están ahí, Paty; amigos de la secundaria, preparatoria y universidad; compañeros y amigos de trabajo.

AGRADECIMIENTOS

Quiero expresar mi agradecimiento a la UNAM por nutrir mis estudios de licenciatura.

A la Facultad de Ciencias, UNAM

Al Departamento de Matemáticas, UNAM

A los miembros del Jurado: Act. María Aurora Valdés Michell, M. en A. P. María del Pilar Alonso Reyes, M. en F. Alberto de la Rosa Elizalde, Act. Jorge Luis Silva Haro y al Dr. Arturo Lorenzo Valdés por realizar la revisión del presente escrito.

A la Act. María Aurora Valdés Michell por aceptar dirigir este proyecto y estar siempre dispuesta a compartir sus conocimientos y formarme como actuario.

Contenido

	Pág.
Introducción	1
1 Antecedentes	3
1.1 Inflación	3
1.2 Índice Nacional de Precios al Consumidor (INPC)	7
1.3 Regresión Lineal Múltiple	10
1.4 Planteamiento de la Hipótesis General	17
1.5 Metodología	17
1.6 Justificación de cada una de las variables	18
2 Algoritmos para la selección de variables y construcción del modelo	28
2.1 Criterios para facilitar la selección del modelo	29
2.1.1 Coeficiente de determinación múltiple	30
2.1.2 R^2 ajustada para “p” términos	30
2.1.3 Cuadrado Medio de Residuales	30
2.1.4 Estadística C_p de Mallows	31
2.2 Algoritmo de selección por segmentos	32
2.2.1 Regresión por segmentos (Forward Stepwise regression)	33
2.2.2 Selección hacia adelante (Forward Selection)	35
2.2.3 Eliminación hacia atrás (Backward Elimination)	35
3 Modelos seleccionados	37
3.1 Resultados obtenidos por el Método Backward	38
3.2 Resultados obtenidos por el Método Backward con ajuste de la variable “y”	42
Conclusiones	47
Anexo A. Regresiones obtenidas y sus comentarios	50
Anexo B. Observaciones atípicas	58
Anexo C. Análisis de multicolinealidad y correlación	60
Anexo D. Transformaciones estabilizadoras de varianza	67
Bibliografía	68

Introducción

La inquietud de realizar el presente trabajo surge de los problemas que se presentan a diario en la economía mexicana con relación al aumento continuo de precios tanto en los bienes de consumo básico como en los de lujo tomando únicamente el último sexenio del Partido Revolucionario Institucional (PRI) hasta el final del primer periodo del Partido Acción Nacional (PAN) al frente de la presidencia de la República Mexicana. Por tal motivo, las siguientes páginas son el resultado de un gran esfuerzo conjunto de investigación y dedicación por medio del cual se intentan analizar de manera sencilla pero concreta diversas propuestas de modelos econométricos con ciertas variables que, se considera, influyen en el comportamiento del INPC y por consecuencia en la inflación de México.

Debido a que la finalidad de esta investigación no es estudiar cada una de las variables que se consideran en esta tesis, sino tomarlos como base para un análisis preciso, se mencionarán de forma muy modesta ciertos conceptos de carácter *financiero, económico, monetario, etcétera*; tales como: tipo de cambio interbancario, inflación tanto de México como de Estados Unidos, tasa de interés interbancaria de equilibrio, reservas internacionales, índice de precios y cotizaciones, entre otras. Todos ellos tomados en cuenta debido a la sencillez de consulta, disponibilidad y obtención para todo público, pero sobretodo por la importancia que cada una significa dentro de la economía mexicana.

Para realizar el análisis se llevarán a cabo diversas regresiones lineales múltiples por diferentes algoritmos con la finalidad de poder encontrar el mejor, para esto se consideran 17 variables tomando como variable dependiente al Índice Nacional de Precios al Consumidor (INPC), que es el principal indicador de la inflación mexicana, y por consiguiente, las independientes serán las restantes. El periodo que se tomó en cuenta para este estudio, fue de forma mensual y abarca desde marzo de 1995 hasta febrero de 2007, habiendo un total de 144 observaciones para cada una de las variables.

La suposición que se hace desde un principio de que el INPC, y en consecuencia la Inflación, son predecibles radica en que se sabe a priori que este indicador cambia periódicamente, sin embargo lo que se desconoce es en qué magnitud lo hará.

Por todo esto es que el presente trabajo pretende responder a las siguientes cuestiones:

- Dentro de la gran diversidad de variables de carácter económico y financiero que existen, ¿cuáles son las que mayor influencia tienen sobre la Inflación de México?
- ¿Es posible construir un modelo capaz de pronosticar confiablemente el comportamiento del INPC?
- En caso de ser posible encontrar un modelo ¿Éste cumpliría con todas las condiciones necesarias para su liberación y finalmente su uso?

1. Antecedentes

1.1 Inflación

Cuando se habla de inflación se refiere al aumento continuo de los precios de bienes y servicios terminados que hay en el mercado. Para poder medirla, se tiene al Índice de Precios al Consumidor (IPC), sin embargo en México debido a la existencia del Índice de Precios y Cotizaciones (IPC) se le ha denominado Índice Nacional de Precios al Consumidor (INPC), por medio del cual se cuantifica el porcentaje de incremento de los precios de bienes y servicios de la canasta básica.

Así pues, para entender mejor la inflación se mencionan algunas de las tesis más importantes que pretenden explicarla:

- Los marxistas aseguran que la inflación es consecuencia de la desigual repartición de la riqueza y de la explotación de grupos favorecidos de capitalistas sobre los trabajadores.
- Los estructuralistas señalan que la inflación se debe a la agrupación del ingreso, el desempleo y la existencia de amplios sectores marginados, así como la gran diferencia que hay entre la producción de bienes y servicios y lo que realmente demanda la mayoría de la población.
- *“Los neoestructuralistas afirman que la inflación es solamente un fenómeno estructural de la economía y de la organización de la sociedad; dicen que la inflación es consecuencia de lo que han provocado los precios anteriores. Los neoestructuralistas no toman en cuenta el factor del circulante.”* [Echavarría, (1992)]
- En América Latina existe una importante escuela, la *Latinoamericana* que considera que la inflación en esta región proviene de ciertas características estructurales de sus economías y no del exceso de demanda.
- La teoría del aumento de costo y sobreprecio menciona que la inflación es el resultado del aumento de salario de la clase obrera y éste a su vez es mayor que la productividad de los trabajadores, cuando los proveedores de materias utilizadas para la producción de otros bienes y servicios aumentan sus precios.

- La escuela austriaca considera que lo que origina la inflación es el aumento del circulante por arriba de los artículos y servicios producidos, conduciendo a un aumento de precios de los mismos.

Por otro lado se tienen algunas tesis que pretenden decir cómo controlar este fenómeno.

- Los keynesianos dicen que para poder controlar la inflación es necesario que el Estado vigile la demanda efectiva. Explicada brevemente se puede decir que si falta demanda efectiva en el mercado, se deben reducir los impuestos, aumentar el gasto fiscal y elevar los salarios; por el contrario, si la demanda es excesiva, se deben elevar los impuestos, no permitir el alza de salarios y producir un superávit fiscal.
- Los neoestructuralistas expresan que para poder llevar a cabo el control de la inflación es necesario controlar los precios; dicen que este control es posible si al llevarlo en práctica los precios relativos están fijados de forma correcta. Sin embargo este método solamente aumenta la inflación, ya que si se congelan los precios durante este proceso, los salarios aumentan y lo que sucede es que la demanda de los consumidores aumenta provocando un incremento en la tasa inflacionaria.

De las tesis mencionadas se puede afirmar que la inflación sostenida es provocada por el aumento de precios y ese aumento es ocasionado por el exceso de demanda que el público ejerce sobre los bienes y servicios disponibles en el mercado ya que, al haber más demanda los precios suben, a menos que la obtención de los mismos aumente.

El alza de precios también se puede asociar con una reducción de los suministros, es decir se produce menos de lo que se necesita, y como la demanda es más alta, los precios incrementan para poder controlarla. Esta situación sirve como indicador para los productores de estos bienes que deben producir más de los mismos, ya que como es de esperarse, la inflación beneficia a los que más tienen los cuales pueden ser por ejemplo, los dueños de empresas a quienes les proporciona mayor poder de compra; y al mismo tiempo, este fenómeno perjudica a los más pobres, como pueden ser por ejemplo los trabajadores o asalariados, ya que aumentan más las utilidades que los salarios; y de forma inversa sucede

con los artículos que necesita la población más pobre, debido a la escasez de demanda dejan de producirse.

Debido a que los precios de mercado relacionan las utilidades marginales de todos los artículos que se ofrecen en el mercado, incluyendo la mano de obra, el movimiento de cualquier precio, afecta y provoca que se modifiquen todos los demás.

Una economía estable, que es cuando los indicadores monetarios permanecen estables, puede sufrir perturbaciones que la lleven a la inflación, ya que siguiendo en parte la tesis de Keynes, se tiene que si la demanda es escasa, el gobierno produce déficit fiscal, esto provoca que la demanda aumente y si ésta la llega a rebasar trae como consecuencia la inflación, a este tipo de inflación se le ha llamado “*demand-pull inflation*” o “*Demanda atrae inflación*”. También existe la contraparte, la llamada “*cost-push inflation*” o “*Precios pujan la inflación*”, que de forma semejante a la anterior si la demanda es muy alta y la mano de obra es escasa para elevar la producción, lo que sucede es que suben los precios de los bienes y por lo tanto se produce la inflación.

Durante el proceso de inflación las remuneraciones económicas que se obtienen de alguna fecha posterior después de la fecha en que se elaboró algún tipo de trabajo van perdiendo su valor debido a la desvalorización que sufre la moneda en este periodo. Por esta razón es que se puede decir que los más propensos a sufrir esta devaluación son los asalariados, ya que primero trabajan y tiempo después reciben su salario.

Es de importante mención que no siempre la inflación es mala ni perjudicial para las clases trabajadoras, pues como característica de las economías de mercado, cuando este fenómeno comienza a manifestarse, produce bienestar económico e incrementa el empleo, ya que como se habló anteriormente, aumenta la demanda de ciertos artículos y por consecuencia su producción, esto implica mayor necesidad de mano de obra y por lo tanto genera fuentes de trabajo; sin embargo, una vez que este hecho adquiere mayor fuerza y la economía comienza a indexarse, es decir, se ajustan los costos de acuerdo al índice general de precios, en el mejor de los casos solamente deja de producir empleos, pero si la inflación se convierte en *Hiperinflación*, que es cuando las tasas de inflación comienzan a crecer de forma muy rápida, sucede lo contrario que al principio, entre más aumento de la inflación se dé, mayor desempleo ocasiona. Pero si disminuye la inflación, la economía se contrae, se reduce el PIB y el desempleo se vuelve mucho mayor. Es por eso que una vez que un país

entra en el proceso de inflación es muy difícil que salga de la misma, lo más conveniente es controlarla, a menos que se adopte la solución que tomó Alemania, tal como lo menciona Hernán Echavarría “Una vez que se llega a la hiperinflación, la única solución es la que adoptó Alemania después de dos guerras mundiales: borrón y cuenta nueva. Se cambia la moneda, se suspenden las emisiones de medio circulante y sólo se dejan funcionando aquellas unidades de producción que sean económicas,..., para suplir la demanda efectiva del consumidor, a precios de equilibrio, es decir, a precios que limpien el mercado.” [Echavarría, 1992]

Como se sabe, mientras exista inflación la demanda de los consumidores aumenta hacia determinados productos, a esta demanda se le llama “*Demanda efectiva de los consumidores*”. Esta demanda se lleva a cabo con el medio circulante o moneda en circulación, la cual recibe este nombre debido a que toda venta implica una compra esto provoca que el dinero pase de compradores a vendedores de forma permanente, es decir “está en circulación”. Por esta razón es que cuando aumentan los salarios, arriendos y utilidades, hay más dinero en circulación en el mercado, esto conlleva a un aumento de precios de los bienes y por lo tanto los ingresos de las empresas aumentan en igual magnitud. De esto se puede reafirmar que quienes más se benefician con la inflación y con el aumento de salarios son los productores y comerciantes, que finalmente son los dueños de empresas y propiedades.

De este primer medio circulante se derivan algunos otros “circulantes”, se tiene por ejemplo el caso de circulante ampliado o depósitos de la gente e instituciones a entidades financieras, entre otros.

En todas las economías que existen, hay 2 tipos de demandas:

1. La que realizan los consumidores hacia los artículos y servicios ya terminados que requieren para satisfacer las necesidades básicas, los cuales son mejor conocidos como “*Canasta Básica*”.
2. La que es por parte de las empresas o personas inversionistas, los cuales son bienes que necesitan para poder producir más artículos.

Esta adquisición de bienes y servicios no es gratis, ambas partes, tanto consumidores como empresas, deben dar dinero a cambio, el cual se obtiene de diferentes formas según sea el caso. Los asalariados, como su nombre lo indica, reciben un salario o

suelo de los servicios que ofrecen a las unidades de producción. Por otro lado las personas que son dueñas de propiedades reciben arriendos y los dueños de capital industrial reciben utilidades. Todo esto es lo que constituye el Producto Interno Bruto (PIB), y esto a su vez es lo que produce anualmente cada país.

De todas estas personas las que tienen más influencia en la inflación son los asalariados ya que comprenden la mayoría de las personas que reciben ingresos, además de que como no pueden ahorrar cuantiosas cantidades de dinero son los que más flujo de efectivo aportan al mercado, por esta razón no es tan fácil aumentar los salarios, ya que cuando se hace en algún sector de la población, éstos puján los precios y los que no reciben dicho aumento tienen menos renta y deben dejar de comprar algunos bienes debido a la alza de precios que provocó la demanda de los que si fueron beneficiados.

La inflación es el incremento de precios producidos por el aumento de demanda, así que si lo que se quiere es controlarla, primero se debe rezagar y disminuir la demanda; cuando se ha logrado esto, la moneda deja de circular como lo hacía antes y de esta forma los consumidores cuentan con menos dinero para gastar y por lo tanto el proceso inflacionario comienza a volverse cada vez menor. Esto provoca que los productores no puedan vender sus productos a precios que ya estaban planeados, ocasionando una disminución en el crecimiento de la producción, y esto a su vez provoca que comiencen a disminuir los ingresos de la población. Sin embargo el no controlar la inflación puede conducir a la economía a la hiperinflación.

1.2 Índice Nacional de Precios al Consumidor (INPC)

Cuando se habla de inflación se refiere al aumento continuo de precios de los bienes y servicios terminados en alguna región específica, siendo más precisos, la inflación de México puede definirse como el aumento periódico de los precios de los bienes y servicios terminados dentro de la República Mexicana.

Este fenómeno es determinado y publicado por la principal institución financiera y económica del país, el Banco de México, el cual se obtiene por medio del Índice Nacional de Precios al Consumidor (INPC). Es necesario hacer una importante aclaración acerca de esta notación ya que puede confundirse con el Índice de Precios y Cotizaciones (IPC), puesto que en otros países al INPC se le define como "*Índice de Precios al Consumidor*

(IPC)” y puede llegar a haber problemas si es que se trabaja con ambos indicadores dentro del mismo trabajo de investigación.

El INPC es un indicador que mide la variación de los precios de los bienes que se encuentran dentro de una canasta representativa, la cual refleja el consumo promedio de las familias mexicanas, a través del tiempo. Entonces, se puede hacer la correcta relación entre “inflación” e “INPC” considerando que la *inflación* es el objeto a medir y el *INPC* es el instrumento apropiado para su medición.

La importancia de este indicador radica en la forma en que se recolectan los datos para determinar el subconjunto de bienes con los cuales se determina la canasta para finalmente construir el INPC. Dicha muestra es representativa de lo que la población en general está consumiendo, ya que cada estado de la República está representado dentro del índice por al menos una ciudad, además de que la ponderación de cada genérico representa la importancia relativa de su gasto en relación con el gasto total. Sin embargo, el INPC no solamente consiste en obtener datos de encuestas a lo largo y ancho del territorio nacional y transcribirlas en cifras; para obtenerlo es necesario evaluar todos los datos obtenidos y así obtener un indicador confiable, para lo cual el Banco de México ha adoptado la *Fórmula de Laspeyres* como medio para procesar los datos y determinar el índice requerido. Dentro de la fórmula de Laspeyres la canasta y ponderaciones permanecen fijas durante el tiempo que se mantenga vigente el índice.

Otro aspecto de importante mención acerca de este indicador es que dentro de él también se encuentran representados ciertos precios que los habitantes deben pagar, tal es el caso del Impuesto al Valor Agregado (IVA), Impuesto Sobre los Autos Nuevos (ISAN), precios de contado y cualquier otro impuesto que se tenga la obligación de pagar al momento de adquirir un bien o servicio.

Por estos motivos es que las autoridades monetarias, hablando más concretamente el Banco de México, analiza y estudia su comportamiento y de esta manera puede tomar las medidas necesarias para que la moneda nacional, el peso mexicano, no pierda poder adquisitivo, ya que dentro del INPC se encuentra representada una gran variedad de precios que los habitantes deben pagar en la vida cotidiana.

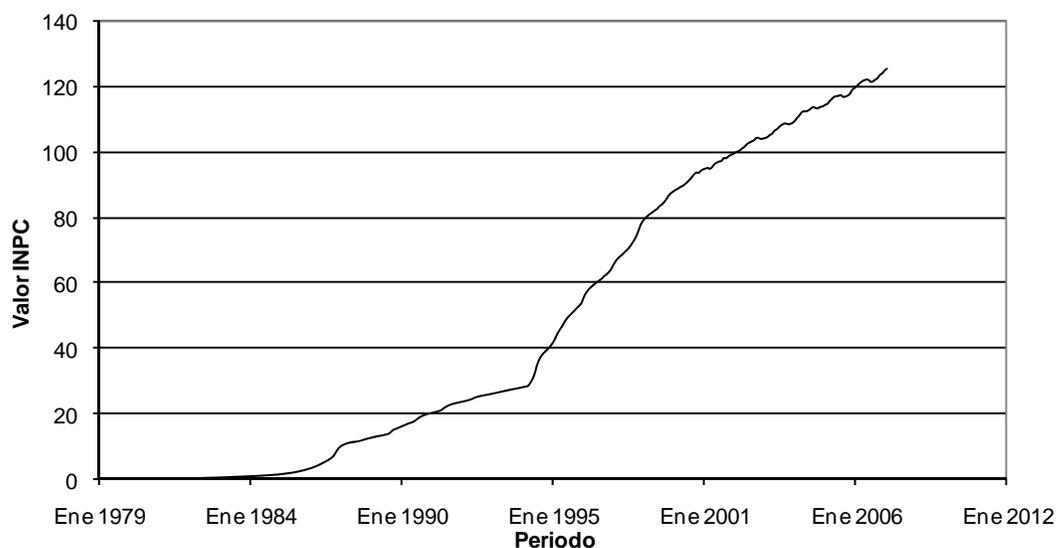
Por otro lado, existe otro indicador que merece mención en este trabajo ya que es mucho más recomendable usar si lo que se pretende es hacer un análisis concreto de la

inflación, este indicador alternativo es “*La Inflación Subyacente*”. Este índice, a diferencia de la inflación total, abarca únicamente un subconjunto de bienes y servicios de la canasta, excluyendo aquellos que estén sujetos a decisiones, sean estacionables o presenten alta volatilidad en sus precios.

Este último punto es con la finalidad de dar a conocer al lector que existen otros indicadores de mayor utilidad si lo que se desea es estudiar a la inflación como tal, sin embargo el objetivo principal de esta tesis no es éste, sino intentar construir un modelo con diversas variables independientes para poder predecir el INPC, es por esta razón que solamente se menciona la existencia de la inflación subyacente y no se tomará en consideración para la construcción del modelo ya que para obtenerlo, se toma un subconjunto de la canasta que se utiliza para obtener la inflación total, por lo cual se corre el riesgo de que exista correlación demasiado alta entre estos dos índices, por lo tanto para evitar este fenómeno se decidió tomar el INPC como variable dependiente.

Para un mejor entendimiento de lo que ha sucedido a través de los años con el INPC y por lo tanto con la inflación de México, se presenta a continuación una gráfica en la cual se observa la evolución y crecimiento de esta variable (Gráfica 1). Dichos datos se obtuvieron desde los más antiguos posibles hasta el más reciente según la página de Internet del Banco de México [Banxico.org.mx].

Grafica 1. Crecimiento del INPC



1.3 Regresión Lineal Múltiple

Para tener una idea más precisa del procedimiento que se va a realizar a lo largo de este trabajo, un concepto que se debe tener presente es el de “Regresión Lineal” puesto que es uno de los temas sobre los cuales se encuentra fundamentado el presente trabajo.

Como “Modelo de Regresión Lineal” se puede definir a aquel modelo matemático que ajusta a la mayoría de los datos u observaciones con “la mejor recta” cuyos parámetros poseen las mejores cualidades estadísticas para el caso donde únicamente interviene una única variable regresora, a este modelo se le llama “Modelo de Regresión Lineal Simple”. Sin embargo, éste no es el único caso posible, ya que en la mayoría de los casos intervienen más de una variable regresora, a estos modelos se les llama “Modelos de Regresión Lineal Múltiple”, donde en lugar de ajustarse una recta, se ajusta un hiperplano en el espacio de k dimensiones, donde k es el número de regresores o variables predictoras.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (1)$$

Los modelos de regresión lineal deben su nombre a que tienen como característica principal que la ecuación obtenida es una función lineal de los parámetros o betas (β) sin importar la superficie que genera el hiperplano descrito.

Los modelos de regresión lineal se usan como funciones de aproximación muy adecuada a la función verdadera desconocida.

Debido a que en la mayoría de los problemas reales se desconocen los parámetros de regresión (las β_i) así como la varianza del error, se estiman con base a observaciones recaudadas previamente.

Esta útil herramienta matemática tiene diversos usos, como pueden ser:

1. Descripción de los datos
2. Predicción y estimación de observaciones futuras
3. Estimación de parámetros
4. Control

De los cuales el de interés en este trabajo es el de predicción y estimación de observaciones futuras.

Para la estimación de los parámetros de regresión o betas se puede lograr por el método de mínimos cuadrados, en el cual se supone que el error del modelo (1), ε_i posee las siguientes características:

- $E(\varepsilon_i) = 0$
- $Var(\varepsilon_i) = \sigma^2$ constante
- Los errores no están correlacionados

Los datos observados se pueden ordenar de la siguiente manera:

Tabla 1. Datos para la regresión lineal múltiple

Observación	Respuesta	Regresores			
i	Y	X ₁	X ₂	...	X _k
1	y ₁	x ₁₁	x ₁₂	...	x _{1k}
2	y ₂	x ₂₁	x ₂₂	...	x _{2k}
.
.
.
N	y _n	x _{n1}	x _{n2}	...	x _{nk}

Por lo que el modelo muestral de regresiones se puede expresar como:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i \quad (2)$$

Sin embargo se pueden manejar en notación matricial para un mejor y más fácil manejo de los datos, por lo que el modelo anterior quedaría de la siguiente manera:

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon} \quad (3)$$

donde:

$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \quad \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Entonces para encontrar los estimadores $\hat{\underline{\beta}}$ se tiene:

$$S(\underline{\beta}) = \sum_{i=1}^k \varepsilon_i^2 = \underline{\varepsilon}' \underline{\varepsilon} = (\underline{y} - X \underline{\beta})' (\underline{y} - X \underline{\beta})$$

de lo cual se obtiene:

$$S(\underline{\beta}) = \underline{y}' \underline{y} - 2 \underline{\beta}' X' \underline{y} + \underline{\beta}' X' X \underline{\beta}$$

ya que $\underline{\beta}' X' \underline{y}$ es un escalar y su transpuesta $(\underline{\beta}' X' \underline{y})' = \underline{y}' X \underline{\beta}$ es el mismo escalar. Los estimadores por mínimos cuadrados deben cumplir la siguiente condición:

$$\left. \frac{\partial S}{\partial \underline{\beta}} \right|_{\hat{\underline{\beta}}} = -2 X' \underline{y} + 2 X' X \hat{\underline{\beta}} = 0$$

despejando se obtiene:

$$X' X \hat{\underline{\beta}} = X' \underline{y}$$

por lo tanto el estimador de β por mínimos cuadrados es:

$$\hat{\underline{\beta}} = (X' X)^{-1} X' \underline{y}$$

si y sólo si $(X' X)^{-1}$ existe, lo cual es cierto si los regresores son linealmente independientes.

De esto se deriva que el modelo de regresión correspondiente a los niveles de variables regresoras $\underline{x}' = [1, x_1, \dots, x_k]$ es:

$$\hat{y} = X \hat{\underline{\beta}} = \hat{\beta}_0 + \sum_{j=1}^k \hat{\beta}_j x_j$$

El vector de valores ajustados \hat{y}_i que corresponde a los valores observados y_i es

$$\hat{\underline{y}} = X \hat{\underline{\beta}} = X (X' X)^{-1} X' \underline{y}$$

La diferencia entre el valor observado y_i y el valor estimado \hat{y}_i es a lo que se le llama residual $e_i = y_i - \hat{y}_i$. Que en forma matricial se denota de la siguiente manera.

$$\underline{e} = \underline{y} - \hat{\underline{y}}$$

Una vez que se han encontrado los estimadores por el método de Mínimos Cuadrados, es recomendable mencionar las propiedades que poseen para determinar si es conveniente su uso final en la práctica en casos reales. Dichas propiedades son:

- “El estimador de mínimos cuadrados $\hat{\beta}$ es el mejor estimador lineal insesgado de β .”
- Si se supone que los errores ε_i tienen distribución Normal, entonces $\hat{\beta}$ es el estimador de máxima probabilidad de β . (Apéndice C4 para demostración)
- La varianza de $\hat{\beta}_j$ es $\sigma^2 C_{jj}$ donde $C = (X'X)^{-1}$ y la Covarianza entre $\hat{\beta}_i$ y $\hat{\beta}_j$ es $\sigma^2 C_{ij}$. Donde σ^2 se logra estimar a través de la suma de cuadrados de los residuales o SS_{Res} .” [Montgomery, (2002)]

La suma de cuadrados de los residuales (SS_{Res}) tiene $n - p$ grados de libertad, donde n es el número de observaciones recopiladas y p es el número de parámetros estimados en el modelo de regresión, donde $p = k + 1$. Esta suma de cuadrados de los residuales se puede encontrar de la siguiente manera:

$$SS_{Res} = y'y - \hat{\beta}'X'y$$

Si a esta suma de cuadrados, la dividimos entre sus grados de libertad se obtiene el “Cuadrado Medio de Residuales” o MS_{Res} por su nombre en inglés “*Mean Square Residuals*”

$$\frac{SS_{Res}}{n - p} = \hat{\sigma}^2 = MS_{Res}$$

por lo que $\hat{\sigma}^2$ es un estimador insesgado de σ^2 .

[Montgomery, (2002)] (Apéndice C3 para demostración)

Una vez que se ha obtenido una propuesta de modelo, el siguiente paso es determinar si al menos uno de los regresores x_1, x_2, \dots, x_k contribuye de manera significativa al modelo, en otras palabras, esto significa que al menos una $\beta_j \neq 0$.

Una herramienta que resume la información principal y que establece la prueba de hipótesis $H_0 : \beta_1 = 0, \beta_2 = 0, \dots, \beta_k = 0$ es el Análisis de Varianza o ANOVA por su nombre en inglés “*Analysis of Variance*” (Tabla 1).

Tabla 2. Construcción de ANOVA

A N O V A

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F ₀
Regresión	SS _R	k	MS _R	MS _R / MS _{Res}
Residuales	SS _{Res}	n - k - 1	MS _{Res}	
Total	SS _T	n - 1		

Donde:

SS_R = Suma de cuadrados de la regresión

$$SS_R = \underline{\hat{\beta}}' X' \underline{y} - \frac{(\sum_{i=1}^n y_i)^2}{n}$$

SS_{Res} = Suma de cuadrados de residuales

$$SS_{Res} = \underline{y}' \underline{y} - \underline{\hat{\beta}}' X' \underline{y}$$

SS_T = Suma total de cuadrados

$$SS_T = \underline{y}' \underline{y} - \frac{(\sum_{i=1}^n y_i)^2}{n}$$

o

$$SS_T = SS_{Res} + SS_R$$

MS_R = Cuadrado medio de la regresión

$$MS_R = \frac{SS_R}{k}$$

MS_{Res} = Cuadrado medio de residuales

$$MS_{Res} = \frac{SS_{Res}}{n - k - 1}$$

n = Tamaño de la muestra

k = Número de regresores estimados asociados a cada x_j para $j = 1, \dots, k$

Ya que se ha obtenido la información necesaria para determinar si al menos una $\beta_j \neq 0$ se realiza la prueba de hipótesis:

$$\begin{array}{l} H_0 : \beta_1 = \dots = \beta_k = 0 \\ \text{vs} \\ H_1 : \beta_j \neq 0 \text{ al menos para una } j \end{array}$$

El rechazo de H_0 implica que al menos un x_j para $j = 1, \dots, k$ contribuye al modelo de forma significativa.

Para que H_0 sea rechazada debe cumplirse que $F_0 > F_{\alpha, k, n-k-1}$, donde

α = Nivel de significancia que por lo regular es del 5%

k = Número de regresores estimados asociados a cada x_j para $j = 1, \dots, k$

n = Tamaño de la muestra

Una vez que se ha determinado si al menos uno de los coeficientes interviene de manera significativa en la regresión o no, lo ideal es saber cuál o cuáles son los que sirven de ellos, para esto es necesario realizar una prueba individual de los coeficientes de regresión, la cual puede ser la “prueba de una cola” o bien la “prueba de dos colas”, la diferencia entre las dos pruebas radica en cómo se haya planteado la hipótesis alternativa para cada β_j desde un principio.

La prueba de dos colas en x_j se realiza cuando la hipótesis alternativa (H_1) se ha definido como que el coeficiente β_j dentro del modelo de regresión lineal múltiple sea distinto a cero, es decir:

$$\begin{array}{l} H_0 : \beta_j = 0 \\ \text{vs} \\ H_1 : \beta_j \neq 0 \end{array}$$

La prueba “ t ” de una cola se realiza cuando se ha definido como hipótesis alternativa (H_1), para x_j , que su coeficiente β_j dentro del modelo de regresión lineal múltiple sea mayor o menor que cero, es decir:

$$\begin{array}{l} H_0 : \beta_j \leq 0 \\ \text{vs} \\ H_1 : \beta_j > 0 \end{array}$$

o bien

$$\begin{array}{l} H_0 : \beta_j \geq 0 \\ \text{vs} \\ H_1 : \beta_j < 0 \end{array}$$

El estadístico de prueba que toma para la realización de esta prueba de hipótesis es:

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

donde C_{jj} es el elemento de la diagonal de $(X'X)^{-1}$

Se rechaza la hipótesis $H_0 : \beta_j = 0$ si $|t_0| > t_{\alpha/2, n-k-1}$, donde de igual manera que en la prueba de significancia de la regresión, o prueba “F”,

α = Nivel de significancia que en la mayoría de los casos es del 5%

k = Número de regresores estimados asociado a cada x_j para $j = 1, \dots, k$

n = Tamaño de la muestra

Esta prueba individual de los coeficientes es una prueba de contribución de x_j asumiendo que los demás regresores se encuentran dentro del modelo, por lo que el no rechazo de H_0 implica que se puede eliminar x_j del modelo, ya que este coeficiente no posee valor para explicar la respuesta, además de que puede disminuir la utilidad del modelo puesto que se corre el riesgo de que esta variable aumente el cuadrado medio de residuales.

1.4 Planteamiento de la Hipótesis General

La hipótesis que se plantea para el desarrollo de esta tesis, es que dentro de la Bolsa Mexicana de Valores (BMV), el Banco de México (Banxico), el Instituto Nacional de Estadística y Geografía (INEGI) entre otras instituciones económicas y de información estadística del país existen diversas variables, índices, instrumentos, etcétera que afectan de forma directa el proceso inflacionario de la nación. Esta influencia que se supone existe y que además es persistente contribuye a que sea posible construir un modelo incluyendo las variables que sean más representativas de las que se sugieren y explican dentro de algunas líneas y que además teóricamente cuentan con dichos supuestos. Del modelo obtenido lo que se espera es poder predecir o pronosticar el comportamiento de la inflación de México, en otras palabras, la hipótesis supone que la inflación, que es cuantificada mediante el Índice Nacional de Precios al Consumidor (INPC), es predecible y por lo tanto se podrá conocer un valor aproximado del Índice Nacional de Precios al Consumidor en fechas futuras, con lo cual se pretende alcanzar el objetivo de conocer en qué proporción se modificarán los precios de los bienes y servicios terminados.

1.5 Metodología

Para poder encontrar el modelo óptimo se procederá a obtener diversas propuestas por 4 maneras diferentes que al final se compararán y de acuerdo a los resultados, las variables que se consideren significativas e importantes dentro de los cuatro modelos obtenidos serán las que finalmente se considerarán para obtener un modelo confiable y eficiente.

Las cuatro maneras a seguir para conseguir dichos modelos son las siguientes:

1. Modelo de regresión múltiple completo, es decir, que incluya todas las variables que se proponen desde un principio
2. Obtener el modelo a través del procedimiento “Stepwise”
3. Conseguir el modelo por medio del procedimiento “Backward”
4. Encontrar una nueva propuesta de modelo siguiendo el procedimiento “Forward”

Estos algoritmos se explicarán en las páginas subsecuentes de forma sencilla para un mejor entendimiento de los mismos.

Como se ha mencionado desde el inicio de la investigación, la variable dependiente es el Índice Nacional de Precios al Consumidor (INPC) al cierre mensual, que es el índice

que se encarga de medir la inflación de México, desde marzo de 1995 hasta febrero de 2007, por lo tanto se tomarán en cuenta 144 observaciones.

Las variables independientes que se utilizaron para el estudio se mencionan a continuación con su respectiva hipótesis alternativa, la cual para manejar un criterio más amplio y de esta forma se aprecien solamente las variables que son significativamente distintas de cero, se realizó la prueba “t” de dos colas, esto es, planteando $\beta_i \neq 0$ $\forall i = 1, \dots, 16$, sin que a priori se tuviera que proponer que debieran ser positivas o negativas (Tabla 2).

1.6 Justificación de cada una de las variables

INPC de los Estados Unidos de América

Como se ha mencionado anteriormente, la inflación es el aumento continuo de los precios de los bienes y servicios terminados y la forma de medirla es por medio del Índice Nacional de Precios al Consumidor, esta definición se aplica a todo tipo de economías.

Debido a que la economía mexicana depende en gran magnitud de la economía extranjera, sobre todo de la economía de los Estados Unidos de América, es que se puede hacer alguna suposición de que el INPC de la República Mexicana se ve influida por el INPC de los Estados Unidos de América, por tal motivo y debido a la influencia que tiene esta nación con México, se supone que si la inflación norteamericana cambia, esto afectará de alguna manera a la inflación mexicana.

Medio circulante

Esta es una variable que se ha tomado en cuenta para incluirla en el presente estudio ya que, al haber aumento de salarios se necesita mayor medio circulante para poder satisfacer la demanda de dinero, esto a su vez implica un aumento de precios en algunos productos y servicios terminados.

Importaciones totales de México

Las importaciones totales de México, mes a mes reflejan las necesidades de la economía nacional. Ya que una nación importa ciertos productos debido a que su producción dentro del territorio nacional puede no ser suficiente para satisfacer la demanda de los mismos, así

pues, se tiene que esta variable puede influir de manera significativa al proceso inflacionario de México, siendo que al importar distintos bienes y servicios se piensa complementar la demanda de los ya producidos en el país y con eso evitar que suban desmedidamente de precios.

Exportaciones totales de México

De forma análoga a la variable anterior, las Exportaciones totales de México también son una variable que es igual de importante, ya que cuando una nación exporta algún producto es porque teóricamente existe producción excesiva y entonces es posible sacar provecho de ésta vendiéndola a otras naciones con lo cual se espera obtener considerables remuneraciones económicas, las cuales se verán reflejadas en la economía nacional provocando una inflación menos agresiva hacia la economía individual.

Índice Nacional de Precios y Cotizaciones (IPC)

Dentro de las bolsas de valores del mundo se encuentran una gran diversidad de índices bursátiles que ayudan a los inversionistas a tomar decisiones en virtud de ahorro y/o inversión, ya que el objetivo de ciertos índices es resumir la información de los precios cotizados en algún mercado durante algún tiempo determinado en un sólo número de fácil comprensión. Así pues, se tiene que en la Bolsa Mexicana de Valores (BMV) también existen diversos índices bursátiles, siendo el Índice de Precios y Cotizaciones (IPC) su principal indicador.

Al IPC se le atribuye como el principal indicador de la BMV ya que para construcción se toma una muestra ponderada, actualmente de 35, emisoras líderes seleccionadas por sus niveles de bursatilidad, dinámica de valor de capitalización del mercado y algunas otras restricciones. Además de que la muestra es revisada anualmente para evitar cualquier anomalía y/o defecto con cualquier emisora que se encuentre dentro del cálculo del índice.

Debido a la importancia del IPC que tiene en la BMV, es que se le ha tomado en consideración para el desarrollo de esta investigación, ya que para la construcción de este indicador se toma en cuenta la participación de grandes empresas que repercuten de manera considerable y de manera periódica en la economía de los mexicanos, se tiene por ejemplo

que algunas de las empresas que se toman en cuenta para su construcción son: Bimbo, CEMEX, Elektra, Grupo Modelo, Telmex, Televisa, TV Azteca, por mencionar algunas. Se sabe que todas y cada una de ellas aportan constantemente bienes y servicios que los ciudadanos adquieren, quienes a su vez pagan por obtenerlos, es aquí donde nuevamente se ve el efecto de la inflación, ya que si dichos productos tienen buena aceptación entonces muy probablemente también incrementen su participación dentro de este índice y por lo tanto repercuta en el valor del IPC, y viceversa, si los productos no han sido aceptados de buena manera en general, posiblemente la participación de estas instituciones dentro del IPC se vería disminuida provocando algún cambio en el indicador IPC. Por tal motivo es que el comportamiento del IPC es muy probable que muestre de una manera aproximada como se está manifestando la Inflación en México.

Ingresos totales de PEMEX

La empresa Petróleos Mexicanos (PEMEX) es una de las más importantes para la economía mexicana y mundial, ya que produce una gran cantidad de barriles de petróleo diarios de los cuales un gran número son exportados y por ende recibe una gran cantidad de dinero por estas ventas. En consecuencia si los ingresos son superiores a los egresos o gastos, la empresa, teóricamente, no tiene la necesidad de incrementar el precio de sus productos. Si por el contrario, los ingresos son menores a los gastos, la empresa para evitar pérdidas caer en un déficit financiero, se ve en la necesidad de incrementar los precios de lo que produce.

Gastos totales de PEMEX

Al igual que sucedió con las variables “Importaciones Totales de México” y “Exportaciones Totales de México” que estaban relacionadas, los gastos totales de PEMEX están analógicamente relacionados con los *Ingresos Totales de PEMEX*, es decir, para que PEMEX pueda percibir ingresos antes debe hacer inversiones, dentro de las cuales se encuentra: compra de equipo necesario para extracción de petróleo, transporte para enviar los barriles extraídos a las debidas refinerías, salario de los empleados, etc. por lo que si aumentan todos estos gastos, la empresa estará obligada a subir de igual manera los precios de los productos que vende y en consecuencia todos los derivados de los mismos, por ejemplo: la gasolina, diesel, gas, etc. Pero si se toma en cuenta la variable anterior

“Ingresos Totales de PEMEX”, si los Gastos son menores que los Ingresos Totales, PEMEX podría crear reservas financieras que le ayuden a solventar demás gastos y por lo mismo no entrar en la necesidad de aumentar los precios, sino por el contrario podría darse el caso de disminuirlos en ciertos productos.

Tipo de cambio promedio mensual peso-dólar americano a la venta

El tipo de cambio se define como el precio de una divisa expresado en términos de la unidad de otra moneda. Este tipo de cambio puede ser a diferentes plazos, de los cuales los más comunes son: valor al mismo día, valor 24 horas y valor 48 horas, al mismo tiempo que existen tanto tipo de cambio a la compra y tipo de cambio a la venta. El primero se define como la cantidad de moneda local que las instituciones bancarias o financieras dan a cambio de la divisa para quien quiera venderlas, mientras que el segundo es el precio al que cada institución bancaria o financiera vende la divisa en moneda local a los que estén interesados en adquirirla.

El criterio que se tomó en cuenta para seleccionar el tipo de cambio a la venta sobre el tipo de cambio a la compra radica en la situación de que se cree que este valor del peso frente al dólar refleja mejor la situación de la inflación mexicana, ya que México al ser un país que, en gran parte, su economía está estrechamente relacionada a la de los Estados Unidos de América, una gran cantidad de transacciones se realizan en dólares, por lo que las empresas mexicanas o internacionales deben adquirir esta divisa para poder realizar sus negociaciones de manera más práctica y cómoda, en lugar de que las instituciones extranjeras hagan el cambio de dólares a pesos.

Tasa de interés interbancaria de equilibrio

“Con el objeto de establecer una tasa de interés interbancaria que refleje mejor las condiciones del mercado, el Banco de México decidió dar a conocer la Tasa de Interés Interbancaria de Equilibrio (TIIE).” [banxico.org.mx]

Por lo tanto, esta tasa de interés puede tener efecto sobre la inflación ya que las instituciones bancarias al ofrecer atractivos intereses sobre los diversos préstamos a distintos plazos o sobre autorizaciones de crédito, las personas pueden acudir con más confianza a diversas instituciones bancarias y/o de crédito para conseguir algún préstamo,

lo cual trae como consecuencia que las personas tengan al alcance más medio circulante y por lo tanto puedan adquirir más bienes y servicios; si por el contrario las tasas de crédito son extremadamente altas, lo que se espera es que las personas no recurran a este servicio y por ende el medio circulante sea menor al caso anterior y por consecuencia las ventas de diversos productos se vean estancadas o disminuidas, lo cual tiene como consecuencia que bajen de precio para aumentar la demanda; ya que como se comentó anteriormente, es una situación natural de la inflación que al haber escasa demanda de algún bien los precios bajan para que dicha demanda aumente y posteriormente cuando la demanda es suficientemente alta suben los precios.

Producción de vehículos de pasajeros

La producción de vehículos de pasajeros es un reflejo de lo que está ocurriendo con el poder adquisitivo de las personas, ya que si la gente ha cubierto sus necesidades principales, por ej.: los alimentos, vestido, calzado, por mencionar algunas, y aun así cuentan con una razonable cantidad de dinero, entonces en ciertas ocasiones la gente en lugar de ahorrarlo lo usa para adquirir artículos de más lujo como es el caso de vehículos. Por lo anterior las empresas que se encargan de producirlos no harán una sobreproducción ya que implicaría pérdidas financieras, por ende únicamente producen lo que esperan vender mes con mes. Por lo tanto, si aumenta la producción de vehículos es consecuencia de que se están vendiendo suficientes como para aumentarla y esto a su vez es consecuencia de que los aumentos de precios dentro de los artículos de primera necesidad no son lo suficientemente grandes como para que la población deje a un lado el consumo de artículos de lujo como los son los vehículos; por el contrario si la producción de vehículos disminuye probablemente se deba a que no se está vendiendo la cantidad estimada para dicho periodo, lo cual no necesariamente es consecuencia de que estén aumentando los precios del consumo básico sino puede ser que la población en lugar de gastar el medio circulante en artículos de lujo lo esté invirtiendo o ahorrando.

Reservas internacionales

Como reserva se entiende a una cantidad de dinero que debe poseer cada institución de carácter financiero o económico para hacer frente a sus obligaciones para con sus clientes. Con base a esto, se puede definir como “reservas internacionales” a la cantidad de recursos con los que cuenta cada país representados principalmente en moneda extranjera y otros activos como puede ser el oro; las cuales pueden ser utilizados para participar en el mercado cambiario, pagar gastos de importaciones y exportaciones así como para defender el tipo de cambio en caso de alguna posible devaluación de su moneda. Sin embargo, estas actividades no necesariamente dependen del todo de las Reservas Internacionales ya que los gastos que implica importar y exportar bienes, pueden pagarse también a través de las utilidades que cada empresa genera por la venta de sus productos.

National Association of Securities Dealers Automated Quotation (NASDAQ)

El índice NASDAQ puede entenderse como un mercado intermediario que crea oportunidades a un gran número de compañías para poder realizar negociaciones y/o transacciones.

Hablando más concretamente, es el mercado electrónico más grande del mundo cuyas negociaciones se realizan a través de complejos sistemas computacionales y canales de telecomunicaciones, gracias a esto, es que las negociaciones se pueden hacer en tiempo real y pueden haber datos de transacción a más de 1.3 millones de usuarios en 83 países alrededor del mundo. Al igual que el Dow Jones, el NASDAQ cuenta con subíndices para un mejor manejo del mismo, en este caso existen dos tipos:

- I. NASDAQ National Market: Cotizan empresas de gran capitalización.
- II. NASDAQ Smallcap Market: Cotizan empresas pequeñas con capacidad de crecimiento.

La mayoría de los títulos manejados en este indicador son: de carácter tecnológico, de telecomunicaciones, banca, minoristas y demás industrias en desarrollo, entre las que se encuentran desde muy pequeñas firmas hasta macro empresas como es el caso de Microsoft.

Las transacciones que se realizan a través del NASDAQ no tienen como límite una cantidad determinada de participantes, por lo cual una gran variedad de negocios pueden

participar sin ventajas o desventajas frente a otros ya que los participantes seleccionan con quien tratar.

La razón por la cual se ha decidido incluir este importante indicador financiero de los Estados Unidos de América en un modelo de regresión del Índice Nacional de Precios al Consumidor de México, es porque debido a la importancia que tiene en las finanzas de la nación norteamericana, un gran número de compañías de gran importancia a nivel internacional participan dentro del NASDAQ, como se ha mencionado el caso de Microsoft, así como también Yahoo, Starbucks Co., NVIDIA Co., Intel Co., Google Inc., entre otras, las cuales tienen una gran participación en México, por lo que se sospecha pueden tener una relación en el comportamiento del INPC, ya que al participar en el comercio mexicano, los habitantes adquieren sus productos y por lo tanto reciben a cambio medio circulante de nuestro país, lo cual se puede tomar como un hecho que justifica su participación en esta investigación.

Dow Jones industrial

El Dow Jones es un cuantificador del rendimiento del mercado en general cuya principal cualidad es que los promedios en cuestión se calculan sobre la base del precio y no de la capitalización, por tal motivo el peso de los componentes varía de acuerdo al precio de sus acciones únicamente sin tomar en cuenta el número de acciones en circulación.

Otra característica importante es que los títulos que maneja poseen una excelente reputación, a su vez que representan adecuadamente al promedio al que pertenecen, así como que son atractivos para varios inversionistas.

El Dow Jones, al igual que el NASDAQ, se subdivide en otros índices más específicos para un análisis más preciso de los promedios en cuestión, dichas subdivisiones son:

- I. Dow Jones de Promedio de Transporte (DJTA)
- II. Dow Jones de Promedio de Utilidades (DJUA)
- III. Dow Jones de Promedio Industrial (DJIA)

De este último, el Dow Jones de promedio industrial (DJIA), cabe recalcar que sus componentes pertenecen a todo tipo de industrias: servicios financieros, tecnología, minoristas, entre otras. Por esta razón éste es el más seguido y reconocido de los índices de

acciones y en consecuencia de los tres promedios que conforman al Dow Jones, ya que los dos restantes se limitan a títulos que pertenecen exclusivamente al tipo de cada uno.

La justificación para incluir esta variable es, al igual que la anterior, la participación en este índice de compañías de carácter internacional que tienen una gran interactividad en la vida diaria de los mexicanos como por ej.: American Express, AT&T, Coca-Cola, Hewlett-Packard, IBM, McDonald's, Microsoft, Pfizer, Walt-Mart, por mencionar algunas.

Sin embargo se desconoce cuál de las dos variables explica de mejor manera el INPC, por lo cual es que se decidió incluir ambas variables dentro de la tesis.

Bolsa de valores de EUA

Esta variable posee un comportamiento en gráficas muy semejante al de las variables anteriores, NASDAQ y Dow Jones industrial ya que los valores de este indicador son los de la bolsa de valores de Nueva York, una de las más importantes no sólo de la Unión Americana sino de todo el mundo, donde algunos de los principales indicadores que se encuentran presentes en esta institución son precisamente estos, no obstante se decidió incluir esta variable puesto que se desconoce si actuando conjuntamente con las demás variables puede explicar de manera certera a la variable dependiente en cuestión. Por estas razones, es que la incorporación de la Bolsa de Valores de EUA puede justificarse de la misma manera que se hizo con la variable "NASDAQ" y la variable "Dow Jones".

Tasa CETE real

“Los CETES son Certificados de la Tesorería de la Federación denominados en moneda nacional que emite el Gobierno Federal, los cuales pueden devengar intereses o no. Dichos certificados se emiten a plazos que por lo general son a 28, 91, 182 y 364 días, tiempo en el cual se convierten en liquidables también por el Gobierno Federal. La emisión de dichos certificados sirve para controlar el medio circulante así como para financiar el gasto público.” (Banxico.org.mx)

La tasa de rendimiento de los CETES se obtiene mediante la diferencia matemática entre el precio de compra y el valor nominal, este último se define como *“el importe de las acciones de una sociedad o de otros valores bursátiles.”* [Kohler, (1970)]

Se tiene entonces, que de acuerdo a la breve definición anteriormente enunciada, la tasa CETE, y los CETES en general pueden tener un efecto más directo con el INPC puesto que intervienen directamente con el efectivo que circula a lo largo del país, el cual es a su vez un elemento de radical importancia en la presencia del fenómeno inflacionario de la nación.

Personas de la industria manufacturera

Antes de justificar la adición de esta variable, se debe definir que es la industria manufacturera.

“Se entiende como industria manufacturera a las actividades orientadas a la transformación mecánica, física o química de bienes naturales o semiprocesados en artículos cualitativamente diferentes. Las actividades manufactureras pueden realizarse con diferentes niveles o escalas de producción, así como empleando diversas tecnologías, y diferentes niveles de complejidad o de integración económica al interior de la unidad de producción (sean plantas industriales, fábricas, talleres o incluso casa-habitación), lo cual no determina su carácter manufacturero.

De igual manera, debe tenerse presente que la producción manufacturera puede presentarse bajo diferentes modalidades: desde la producción por cuenta propia, en la cual el productor selecciona y adquiere las materias primas para su actividad y decide las características, volúmenes y destino de los productos que fabrica; de igual forma existe la producción sobre pedido; hasta la producción bajo la modalidad de maquila, en la cual se reciben las materias primas para su transformación bajo las especificaciones y requerimientos del cliente. Finalmente, debe recordarse que si bien toda la manufactura implica la transformación, no toda transformación es manufactura, pues existen actividades en las que igualmente se presenta una determinada transformación, como: en el beneficio de productos mineros, la construcción y en las actividades que acondicionan los bienes para poder ser utilizados o consumidos tales como: las actividades de los cerrajeros, la preparación de alimentos, entre otras.” (seplader.seige.qroo.gob.mx)

Por lo tanto, lo que se intenta alcanzar mediante la inclusión de las personas que laboran en la industria manufacturera en el modelo de regresión múltiple, es tener en consideración la mano de obra económicamente activa del periodo en cuestión, lo cual es

un elemento de suma importancia en la inflación de un país ya que si ésta incrementa puede traer de la mano un incremento en necesidad de mano de obra, debido al alza en la demanda de ciertos artículos, y por lo tanto generar empleos; así como también, estas personas son las que mantienen vigentes las economías de cada país.

Tabla 3. Forma de medición de las Variables

Variable	Forma de medición	Hipótesis Alternativa
Índice de Precios al Consumidor de los Estados Unidos de América	Tasa	$\beta_1 \neq 0$
Medio circulante mensual	Millones de pesos	$\beta_2 \neq 0$
Importaciones totales de México	Millones de dólares	$\beta_3 \neq 0$
Exportaciones totales de México	Millones de dólares	$\beta_4 \neq 0$
Índice de Precios y Cotizaciones (IPC)	Tasa	$\beta_5 \neq 0$
Ingresos totales de PEMEX	Millones de pesos	$\beta_6 \neq 0$
Gastos totales de PEMEX	Millones de pesos	$\beta_7 \neq 0$
Tipo de cambio promedio mensual Peso-Dólar americano	Peso por dólar	$\beta_8 \neq 0$
Tasa de interés interbancaria	Tasa	$\beta_9 \neq 0$
Producción de vehículos de pasajeros	Miles de unidades	$\beta_{10} \neq 0$
Reservas internacionales de México	Millones de dólares	$\beta_{11} \neq 0$
NASDAQ	Tasa	$\beta_{12} \neq 0$
Dow Jones	Tasa	$\beta_{13} \neq 0$
Bolsa de valores de los Estados Unidos de América	Tasa	$\beta_{14} \neq 0$
Tasa CETE real	Tasa	$\beta_{15} \neq 0$
Personas de la industria manufacturera	Unidades	$\beta_{16} \neq 0$

2. Algoritmos para la selección de variables y construcción del modelo

El problema de selección de variables y construcción del modelo no es nada fácil de resolver debido a la infinidad de variables independientes que se pueden tomar en cuenta para algún estudio en particular, lo cual se complica si se toma en consideración que existen 2^k ecuaciones en total, siendo k el número de regresores posibles y suponiendo que se incluye la ordenada al origen β_0 .

Un problema más, es que existe una contrariedad cuando se debe seleccionar un subconjunto de los regresores disponibles, esto es:

- “1) Incluir el mayor número de regresores posibles para que la influencia contenida sea influyente sobre la variable dependiente “y”, sin embargo, a su vez se desea
- 2) Incluir el menor número de regresores dentro de nuestras posibilidades para que la varianza del modelo sea lo menor posible” [Montgomery, (2002)]

Es por esto que el modelo que se encontrará es uno que sea el término medio de los 2 requerimientos antes mencionados, y esto a su vez se define como encontrar la “mejor ecuación de regresión”.

No obstante la “mejor ecuación de regresión” no siempre reditúa el éxito que se espera, incluso aunque lo hiciera, una gran parte del “peso” de la ecuación se puede concentrar en unas cuantas observaciones, las cuales se llaman “observaciones influyentes o atípicas”. Es por eso que aunado a esta ecuación o ecuaciones se debe realizar *el Análisis de Residuales*, por medio del cual salen a la luz formas funcionales de regresores, se encuentran nuevos regresores candidatos, o incluso se pueden identificar las observaciones atípicas.

A pesar de que se han desarrollado algunos algoritmos para seleccionar el modelo final de forma más rápida y eficiente, éstos no garantizan encontrar la mejor ecuación, sin embargo ayudan a optimizar el tiempo de búsqueda y a descartar ciertas variables que estén correlacionadas.

Si del subconjunto de variables se eliminan ciertas variables del modelo, se puede mejorar la precisión de los parámetros estimados en las variables que aún se encuentran en el mismo; incluso si algunas variables son omitidas no son despreciables. Al omitir variables se introduce sesgo potencial en los coeficientes estimados de las variables que se

han quedado dentro del modelo así como en las variables de respuesta; ahora, si las variables eliminadas no tienen un efecto significativo, entonces el MSE (Error Cuadrado Medio) de los estimadores sesgados será menor que la varianza de los estimadores insesgados.

Existe peligro si se retienen variables despreciables ya que aumentan las varianzas de los estimadores de los parámetros y de la respuesta predicha.

Otro tipo de problema que se puede encontrar en el desarrollo del análisis, son los defectos en los datos, ya que en la mayoría de los casos, se recolectan datos históricos dentro de los cuales pueden existir valores atípicos, inconsistencias en la muestra, incluso algunas de las observaciones pueden verse afectadas por factores externos o ajenos al control sobre los mismos, etcétera. Como menciona Montgomery en su libro, *“Estos defectos nos pueden traer impacto sobre el proceso de selección de variables y conducir a la mala especificación del modelo.”* [Montgomery, (2002)]

Otro problema que se puede presentar, es que el intervalo que se ha tomado en cuenta para realizar el modelo es muy limitado lo cual puede tener como consecuencia que el regresor parezca no muy importante si se ha utilizado el método de mínimos cuadrados para estimar los parámetros. Por todo esto es que se debe tener mucho cuidado al momento de seleccionar el modelo final, sin embargo existen algunos criterios que ayudan a que dicha selección sea más fácil.

2.1 Criterios para facilitar la selección del modelo

Como se mencionó anteriormente, existen diversos problemas para determinar el mejor o los mejores modelo(s) de regresión, pero gracias a los criterios que se mencionan a continuación la selección final es mucho más fácil y rápida. Hay que aclarar que el objetivo de esta tesis no es estudiar ni analizar los criterios que a continuación se mencionan así como tampoco del análisis de regresión, por lo cual se exhorta al lector consultar el libro *“Introducción al análisis de regresión lineal”* de Montgomery (2002) para profundizar y entender mejor los procedimientos que se mencionan en las siguientes líneas.

2.1.1 Coeficiente de determinación múltiple:

$$R_p^2 = 1 - \frac{SS_{Res}(p)}{SS_T}$$

El coeficiente de determinación múltiple aumenta a medida que el valor de p también lo hace, por esta razón es que se pueden agregar regresores al modelo hasta que se llegue al punto en que agregar otra variable ya no es útil debido a que se produce un mínimo aumento en R_p^2 .

2.1.2 R^2 ajustada para “ p ” términos:

$$R_{Aj,p}^2 = 1 - \left(\frac{n-1}{n-p} \right) (1 - R_p^2)$$

Debido a que este indicador no necesariamente aumenta al incluir más regresores al modelo, si se aumentan “ s ” regresores entonces $R_{Aj,p+s}^2 > R_{Aj,p}^2$ sí y sólo si $F_{parcial} > 1$. Entonces si se desea seleccionar un subconjunto óptimo por este criterio se debe seleccionar al que contenga el máximo $R_{Aj,p}^2$.

2.1.3 Cuadrado Medio de Residuales:

$$MS_{Res}(p) = \frac{SS_{Res}(p)}{n-p}$$

Si lo que se quiere es usar este criterio para seleccionar el mejor modelo se puede justificar la elección en lo siguiente:

”1.- $MS_{Res}(p)$ mínimo.

2.- Valores de p tales que $MS_{Res}(p)$ sea aproximadamente igual a MS_{Res} del modelo completo.

3.- Valor de p cercano al punto donde crece $MS_{Res}(p)$ mínimo.

De esto se obtiene, que el modelo de regresión para un subconjunto que minimiza $MS_{Res}(p)$, maximizará $R_{Aj,p}^2$. Entonces se puede concluir que $MS_{Res}(p)$ mínimo es equivalente a R^2 ajustada máxima.” [Montgomery, (2002)]

2.1.4 Estadística C_p de Mallows:

$$C_p = \frac{SS_{Res}(p)}{MSE(X_1, \dots, X_{p-1})} - (n - 2p)$$

Para usar este criterio de selección es de mucha utilidad observar la gráfica de C_p en función de p para cada ecuación de regresión. Bajo este método se prefieren aquellas ecuaciones con pequeños valores de C_p ; esto conlleva a que es preferible aceptar algo de sesgo en la ecuación siempre y cuando esta decisión ayude a disminuir el error promedio de predicción.

El hacer estos estudios y generar modelos puede tener diferentes aplicaciones y finalidades, según lo que se quiera investigar. Estos usos de la regresión pueden ser:

1. Descripción de los datos
2. Predicción y estimación
3. Estimación de parámetros
4. Control

Es por eso que no se puede aplicar el mismo criterio de selección para todos los modelos ni en todas las investigaciones; esto dependerá del objetivo de la investigación, por ejemplo, si lo que se desea es una buena descripción de los datos, entonces lo que se pretende encontrar son ecuaciones donde la suma de cuadrados de los residuales sea mínima. *“Como SS_{Res} se minimiza usando k regresores candidatos, es preferible eliminar algunas variables sólo si se obtiene un pequeño aumento en SS_{Res} ; lo deseable es describir el sistema con el mínimo posible de regresores, pero a la vez explicar la parte importante de la variabilidad de “y”*”. [Montgomery, (2002)]

Para encontrar un modelo con la finalidad de predecir observaciones, lo que se desea seleccionar son las variables regresoras que reduzcan al mínimo el error cuadrático medio de la predicción, lo cual puede llevar a eliminar del modelo los regresores con efectos pequeños. Incluso para la selección de este modelo es útil tomar en cuenta la estadística PRESS (Prediction Error Sum of Squares) para evaluar todos los modelos que son candidatos ya que discrimina modelos alternativos:

$$PRESS_p = \sum_{i=1}^n \left(\frac{e_i}{1 - h_{ii}} \right)^2$$

Donde se seleccionará(n) el/los modelo(s) con valor pequeño de $PRESS_p$.

Si lo que se quiere es estimar parámetros, se debe tomar en cuenta el sesgo resultante de omitir algunas variables, así como las varianzas de los coeficientes estimados para cada variable.

Por último para encontrar un modelo que sirva para control, Montgomery sugiere lo siguiente: “... es de suma importancia obtener estimadores exactos de los parámetros lo cual implica que sus errores estándar deben ser pequeños así como que los ajustes realizados por las x 's para controlar “y” serán proporcionales a las $\hat{\beta}$ los coeficientes de regresión deben representar con fidelidad los efectos de los regresores.” [Montgomery, (2002)]

Si después de realizar las pruebas necesarias para seleccionar el modelo adecuado, se obtiene más de uno, “el mejor modelo” será aquel que tenga la menor, o en el mejor de los casos, que no tenga colinealidad entre las variables.

2.2 Algoritmo de selección por segmentos.

Como se ha mencionado anteriormente, existen algunos procedimientos que simplifican la búsqueda del “modelo final”, los cuales evalúan sólo unos pocos subconjuntos de variables. Sin embargo, hay que reiterar el punto de que estos modelos que se obtienen no son necesariamente el mejor o los mejores.

A estos métodos de búsqueda se les suele llamar “*Procedimientos del tipo por segmentos*” los cuales se dividen principalmente en 3 tipos:

1. Regresión por segmentos (Forward Stepwise Regression)
2. Selección hacia delante (Forward Selection)
3. Eliminación hacia atrás (Backward Elimination)

2.2.1 Regresión por segmentos (Forward Stepwise Regression)

Una manera rápida y eficiente para encontrar una aproximación del modelo adecuado es a través del método “*Forward Stepwise Regression*”, el cual indica paso a paso cual de las variables que están esperando ser incluidas al modelo es la más apropiada para ser implantada al modelo y cual o cuales de las que ya se encuentran dentro no son tan significativas como para seguir las conservando. Para esto es necesario seguir una serie de pasos los cuales se mencionan a continuación y con los cuales se espera obtener el mejor modelo con las variables más representativas e influyentes para la variable independiente.

Este algoritmo es en términos del estadístico F^* para la prueba parcial F .

- i. Se obtiene un primer modelo de regresión lineal simple para cada una de las k variables potenciales en x .

Para cada uno de los modelos obtenidos se realiza una prueba F^* donde el estadístico de prueba para la variable k se obtiene como sigue:

$$F_k^* = \frac{MSR(x_k)}{MSE(x_k)}$$

Entonces, la variable con el mayor valor de F^* es la candidata a entrar al modelo de regresión lineal múltiple.

¿Cómo saber si dicha variable se puede incluir en el modelo?

Para esto es necesario fijar un valor predeterminado el cual normalmente es establecido mediante un valor cercano al valor en tablas de $F_{\alpha, m, p}$, donde m es el número de parámetros estimados, α es el nivel de significancia y p es un número muy cercano al tamaño de la muestra

El criterio para decidir si la variable entra o no es:

Si F_k^* es mayor que el valor predeterminado, entonces la variable k es agregada al modelo, de otra forma la variable no se incluye.

En este caso como es el primer paso si ninguna variable es propensa a entrar, se ha concluido el procedimiento ya que ninguna variable se consideró útil dentro del modelo de regresión.

- ii. Si la variable k ha permanecido dentro del modelo, entonces el siguiente paso es hacer una nueva regresión lineal, esta vez con dos variables pero siempre

incluyendo a la variable k . Ya obtenido el modelo bivariado, de nuevo se realiza una prueba F^* para determinar cuál de las variables que se encuentran fuera es la más indicada para incluirla en el modelo y si alguna de las que ya están dentro debe salir.

La prueba F^* que se realiza es la siguiente:

$$F_j^* = \frac{MSR(x_j | x_k)}{MSE(x_k, x_j)}$$

De forma semejante al paso anterior, la variable con el mayor valor de F^* es la candidata para entrar en el modelo. Nuevamente el criterio para determinar si entra o no es comparar F^* con un valor predeterminado.

Si F^* es mayor que este valor entonces la variable “ j ” es incluida al modelo. En caso de que no cumpla con esto, el algoritmo termina y no se agrega ninguna otra variable.

- iii. Supóngase que x_j es incluida en el modelo, entonces el siguiente paso es determinar si alguna de las variables ya incluidas es necesario que salga del modelo. En este caso la variable que recién entró no se le debe hacer dicha prueba sino a aquella(s) que ya se encontraba(n), según sea el caso de cuantas variables ya se hayan incluido en el modelo; en este caso sería hacer la prueba a x_k , entonces el estadístico de prueba es:

$$F_k^* = \frac{MSR(x_k | x_j)}{MSE(x_j, x_k)}$$

En forma contraria a lo que sucede cuando una variable es candidata a entrar, esta vez la variable con menor valor de F^* es la candidata para salir del modelo, sin embargo al igual que en los casos anteriores este valor debe ser menor a otro ya determinado con anterioridad.

Así pues si F_k^* es menor que dicho valor, la variable k debe ser removida del modelo, en caso contrario se mantiene dentro.

- iv. De nuevo se supone que ambas variables, tanto x_j como x_k , se han mantenido dentro del modelo, entonces el siguiente paso es determinar nuevamente cual de

las variables ausentes es la candidata a entrar y una vez que se ha encontrado alguna “variable prospecto” se procede a comprobar cuál de las ya presentes debe ser removida. Todo esto es siguiendo los procedimientos antes vistos y realizando las pruebas F^* correspondientes.

Este algoritmo se sigue hasta que ninguna variable pueda ser eliminada y ninguna pueda ser agregada. Es hasta este punto cuando se puede decir que se ha encontrado un modelo con variables altamente influyentes dentro del modelo y además se puede decir que ninguna variable que no se encuentra es necesaria dentro del mismo.

2.2.2 Selección hacia adelante (Forward Selection)

La *selección hacia adelante* parte con la hipótesis de que no hay regresores en el modelo además de la ordenada al origen (β_0). El procedimiento de este método es, en esencia, el mismo que se sigue en la *Regresión por segmentos “Forward Stepwise Regression”*, sólo que aquí se omite el paso de remover alguna variable que ya ha sido introducida al modelo. El estadístico de prueba que se usa para determinar si la variable candidata entra o no es:

$$F_{IN} = F_{\alpha,1,n-p} \text{ con } \alpha = 0.05$$

2.2.3 Eliminación hacia atrás (Backward Elimination)

La *eliminación hacia atrás* funciona en forma análoga al método de selección hacia adelante. Este método parte con la hipótesis de que todos los regresores candidatos se encuentran dentro del modelo y se procede a calcular el estadístico F^* para cada uno de los regresores. Por ejemplo para X_1

$$F_1^* = \frac{MSR(X_1 | X_2, \dots, X_{p-1})}{MSE(X_1, \dots, X_{p-1})}$$

Entonces el regresor con el mínimo valor de F_i^* es comparado con un valor $F_{OUT} = F_{\alpha,1,n-p}$ con $\alpha = 0.05$ previamente establecido, si $F_i^* < F_{OUT}$ entonces X_i es quitada del modelo. Se reajusta con $k-1$ regresores y se vuelve a repetir el procedimiento y se llega al modelo final cuando ya no se pueden quitar más regresores.

Montgomery recomienda “*aplicar todos los métodos para ver si se logra encontrar alguna concordancia.*” [Montgomery, (2002)]

Berk (1978) propone que la selección hacia delante es útil para tamaños de subconjuntos pequeños y, por el contrario, si se tiene un subconjunto grande es recomendable usar el método de eliminación hacia atrás, ya que en la mayoría de los casos se ve menos afectada por la correlación entre los regresores, que la selección hacia delante. [Mantel, (1970)]

Una última recomendación que hace Montgomery antes de decidir cuál de los modelos obtenidos por estos métodos es el que se quedará como “modelo final” es: “*Estos modelos los podemos evaluar mediante criterios C_p o MS_{Res} y finalmente determinar el modelo final.*” [Montgomery, (2002)]

3. Modelos seleccionados

Como se sabe, para poder seleccionar un modelo de entre las diversas propuestas que pueden surgir, éste debe cumplir ciertos aspectos los cuales a manera de resumen se mencionan a continuación:

- La prueba t sea significativa para cada uno de los coeficientes
- La prueba F resulte significativa para todo el modelo de regresión
- No exista autocorrelación, la cual puede ser determinada mediante el estadístico Durbin-Watson
- Tenga el menor valor del estadístico PRESS
- Posea un alto valor de R^2
- No exista multicolinealidad entre las variables
- Homocedasticidad
- Los residuales posean una distribución normal

Por lo tanto, después de obtener diversas alternativas de modelos de regresión por distintos algoritmos (Véase Anexo A), se observó que el modelo conseguido mediante el método Backward genera los resultados más estables de acuerdo a las pruebas establecidas para poder considerarlo como tal.

Para un mejor entendimiento se presentan tanto los resultados obtenidos como algunos comentarios.

3.1 Resultados obtenidos por el Método Backward

Tabla. 4 Resumen del Modelo

Modelo	R	R Cuadrada	R Cuadrada Ajustada	Error Estándar del Estimado	Durbin-Watson
5	.998 ^a	.997	.996	1.473598060	.938

e. Predictores: (Constante), Pers_Ind_Manuf, Gastos_PEMEX, Produc_Veh_Pas, Tasa_Int_Banc, TC_interb48h_Vta, IPC, Ingresos_PEMEX, Export_Mex, Circulante_mens, INPC_USA, Import_Mex, Dow_Jones

Tabla 5. ANOVA(f)

Modelo		Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F	Sig.
5	Regresión	87,643.563	12	7,303.630	3,363.417	9.783E-157
	Residual	284.465	131	2.171		
	Total	87,928.028	143			

e. Predictores: (Constante), Pers_Ind_Manuf, Gastos_PEMEX, Produc_Veh_Pas, Tasa_Int_Banc, TC_interb48h_Vta, IPC, Ingresos_PEMEX, Export_Mex, Circulante_mens, INPC_USA, Import_Mex, Dow_Jones

f. Variable Dependiente: INPC

Tabla 6. Coeficientes

Modelo		Coeficientes no estandarizados		Coeficientes Estandarizados	t	Sig.	Estadísticas de Colinealidad		Intervalos de Confianza para B al 95.0%	
		B	Error Estándar	Beta			Tolerancia	VIF	Cota Inferior	Cota Superior
5	(Constante)	-204.01	16.03		-12.7	1.19E-24			-2.3572E+02	-1.7230E+02
	INPC_USA	2.06	0.15	0.81	13.96	1.05E-27	0.01	135.1	1.7718E+00	2.3568E+00
	Circulante_mens	7.32E-05	1.11E-08	0.31	6.59	9.83E-10	0.01	87.69	5.1255E-05	9.5236E-05
	Import_Mex	8.64E-04	3.88E-04	0.16	2.23	2.75E-02	0.01	195.8	9.7256E-05	1.6314E-03
	Export_Mex	-1.34E-03	3.73E-04	-0.22	-3.59	4.69E-04	0.01	151.8	-2.0748E-03	-6.0018E-04
	IPC	-1.42E-03	7.03E-05	-0.33	-20.2	4.98E-42	0.09	10.56	-1.5576E-03	-1.2795E-03
	Ingresos_PEMEX	5.05E-05	9.62E-06	0.11	5.25	6.05E-07	0.05	18.66	3.1446E-05	6.9508E-05
	Gastos_PEMEX	-7.81E-05	1.42E-05	-0.12	-5.5	1.95E-07	0.05	20.29	-1.0618E-04	-4.9973E-05
	TC_interb48h_Vta	1.21	0.26	0.07	4.64	8.40E-06	0.11	8.93	6.9513E-01	1.7290E+00
	Tasa_Int_Banc	-0.1	0.02	-0.06	-5.91	2.83E-08	0.25	3.97	-1.3376E-01	-6.6649E-02
	Produc_Veh_Pas	3.46E-02	1.34E-02	0.03	2.58	1.09E-02	0.22	4.63	8.0978E-03	6.1154E-02
	Dow_Jones	1.94E-03	2.26E-04	0.16	8.58	2.35E-14	0.07	14.14	1.4899E-03	2.3827E-03
	Pers_Ind_Manuf	2.53E-05	4.80E-06	0.09	5.27	5.35E-07	0.09	10.98	1.5827E-05	3.4824E-05

a. Variable Dependiente: INPC

Valor crítico de la prueba t de dos colas 1.978239

Ecuación de Regresión Backward:

$$y_i = -240.011 + 2.064x_1 + 7.325E-05x_2 + 8.643E-04x_3 - 1.338E-03x_4 - 1.419E-03x_5 + 5.0477E-05x_6 - 7.80774E-05x_6 - 7.80774E-05x_7 + 1.212x_8 - 0.100x_9 + 0.035x_{10} + 1.936E-03x_{13} + 2.533E-05x_{16}$$

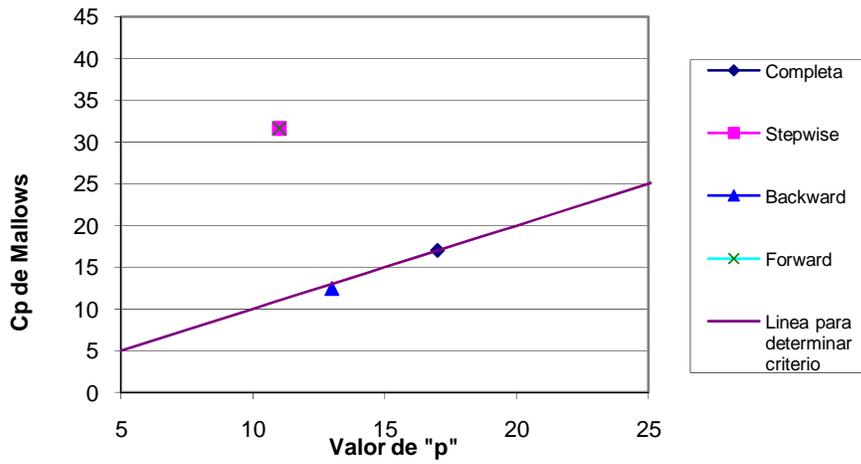
Tabla 7. Valores del Estadístico PRESS

VALORES ESTADÍSTICO PRESS			
Completa	Stepwise	Backward	Forward
1363.86508	401.382576	364.672461	401.382576

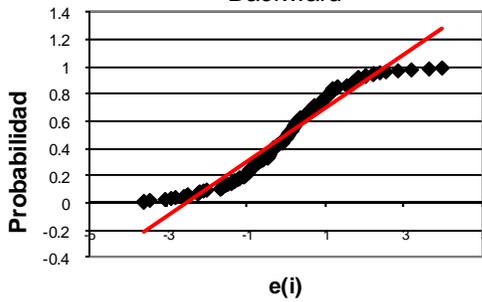
Tabla 8. Valor R² Predicción PRESS

VALOR R ² PREDICCIÓN PRESS			
Completa	Stepwise	Backward	Forward
0.98448885	0.9954351	0.9958526	0.9954351

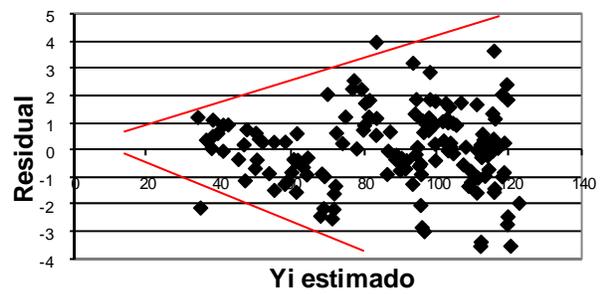
Gráfica 2. Cp de Mallows



Gráfica 3. Probabilidad normal Backward

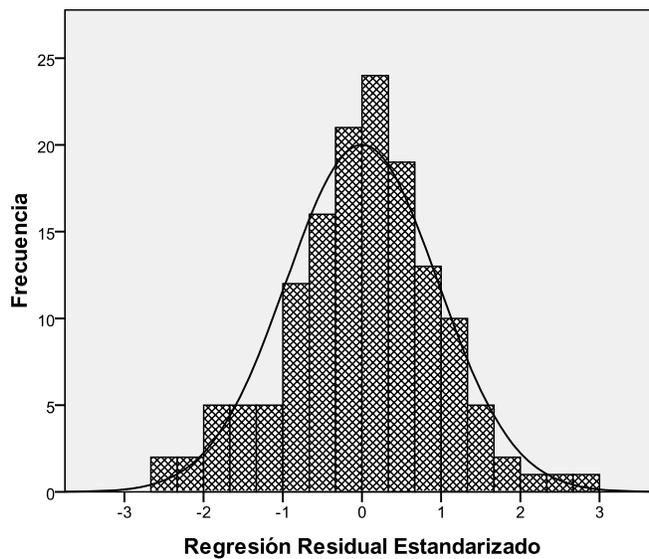


Gráfica 4. Residuales Backward



Gráfica 5. Histograma del primer modelo seleccionado

Variable Dependiente: INPC



Media = 6.14E-14
Desv. Est. = 0.957
N = 144

Esta propuesta se eligió de entre las cuatro primeras (Véase Anexo A para los demás modelos) debido a que es el modelo con menos inconsistencias en sus resultados y pruebas aplicadas, tal es el caso de que de las 12 variables significativas 6 no tuvieron índices que demostraran la presencia de dependencia lineal, así como la gráfica de la C_p de Mallows muestra que fue el modelo que cuenta con el mínimo valor del mismo y además fue el más cercano a la recta $C_p = p$, lo cual indica que es la ecuación con menor sesgo y esto es un requisito deseable en la construcción del modelo ideal.

“Cuando se usa el criterio C_p puede ser útil el visualizar la gráfica de C_p en función de p para cada ecuación de regresión,... Las ecuaciones de regresión con poco sesgo tendrán valores de C_p que caen cerca de la recta $C_p = p$, mientras que las que tengan sesgo apreciable quedaran arriba de esa recta... En general, se prefieren valores pequeños de C_p , ... Puede ser mejor aceptar cierto sesgo en la ecuación, para reducir el error promedio de predicción.” [Montgomery, (2002)]

A su vez, también presentó el MSRes mínimo y menor Error Total al ubicarse gráficamente por debajo de todos los otros modelos. Pero sobre todo, es el modelo que tiene el valor del estadístico PRESS más pequeño, siendo este valor el más importante de todos dentro de esta investigación, ya que el objetivo de encontrar el “mejor modelo de regresión” es con la finalidad de poder predecir observaciones futuras.

Sin embargo, como puede apreciarse en los resultados obtenidos, a pesar de que la prueba F resulto muy significativa para este modelo, existe una R^2 demasiado alta y además de que en esta propuesta de modelo resultaron significativas las 12 variables incluidas bajo el criterio de la prueba “t” de dos colas, no es conveniente su liberación y por lo tanto su utilización, ya que contiene un nivel significativo de multicolinealidad (Véase Anexo C).

Otra inconsistencia que se encontró es la presencia de autocorrelación negativa de acuerdo al estadístico Durbin-Watson, además en la gráfica de residuales se observa una varianza creciente así como la gráfica de probabilidad normal no se ajusta de manera correcta para poderse considerar como tal, por lo que para un análisis econométrico se debe descartar esta propuesta.

Pese a esto, existen diversos métodos para corregir ciertos problemas que se pueden presentar tanto en la construcción de un modelo de regresión lineal como para la construcción de un modelo econométrico, entre ellos se encuentra la posibilidad de realizar

una transformación sobre la variable dependiente con la finalidad de estabilizar la varianza. La transformación que se seleccionó para este proyecto es:

$$y^* = \frac{1}{y}$$

En el Anexo D se hace una mención más detallada acerca de las diferentes transformaciones que se pueden hacer para estos fines.

Con este ajuste de variable se obtuvieron nuevos modelos por los mismos algoritmos que en el caso anterior para poder hacer una comparación entre ambos resultados. De estos modelos nuevamente el modelo seleccionado es el que se obtuvo por el método Backward, a continuación se presentan sus resultados y en el Anexo A se muestran las demás propuestas.

3.2 Resultados obtenidos por el Método Backward con ajuste de la variable “y”

Tabla 9. Resumen del Modelo

Modelo	R	R Cuadrada	R Cuadrada Ajustada	Error Estándar del Estimado	Durbin-Watson
8	0.995	0.990	0.990	0.000510664	.903

h. Predictores: (Constante), Pers_Ind_Manuf, Tasa_Int_Banc, Nasdaq, TC_interb48h_Vta, IPC, Ingresos_PEMEX, INPC_USA, Import_Mex, Dow_Jones

Tabla 10. ANOVA(f)

Modelo		Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F	Sig.
8	Regresión	3.52E-03	9	3.91E-04	1,500.756	6.96E-130
	Residual	3.49E-05	134	2.61E-07		
	Total	3.56E-03	143			

h. Predictores: (Constante), Pers_Ind_Manuf, Tasa_Int_Banc, Nasdaq, TC_interb48h_Vta, IPC, Ingresos_PEMEX, INPC_USA, Import_Mex, Dow_Jones

f. Variable Dependiente: INPC

Tabla 11. Coeficientes

Modelo		Coeficientes no estandarizados		Coeficientes Estandarizados		t	Sig.	Estadísticas de Colinealidad		Intervalos de Confianza para B al 95.0%	
		B	Error Estandar	Beta				Tolerancia	VIF	Cota Inferior	Cota Superior
8	(Constante)	8.08E-02	4.11E-03			19.656	2.66E-41			7.2702E-02	8.8969E-02
	INPC_USA	-3.32E-04	3.38E-05	-0.645		-9.829	1.68E-17	0.017	58.65	-3.9865E-04	-2.6510E-04
	Import_Mex	1.57E-07	5.12E-08	0.14		3.071	2.58E-03	0.035	28.42	5.5960E-08	2.5847E-07
	IPC	1.40E-07	2.22E-08	0.16		6.303	3.91E-09	0.114	8.75	9.5909E-08	1.8362E-07
	Ingresos_PEMEX	-2.24E-09	1.08E-09	-0.025		-2.083	3.91E-02	0.515	1.942	-4.3680E-09	-1.1362E-10
	TC_interb48h_Vta	-1.00E-03	8.38E-05	-0.283		-11.96	7.08E-23	0.131	7.643	-1.1677E-03	-8.3622E-04
	Tasa_Int_Banc	1.05E-04	5.78E-06	0.305		18.191	5.15E-38	0.261	3.834	9.3688E-05	1.1655E-04
	Nasdaq	3.07E-07	1.09E-07	0.047		2.824	5.46E-03	0.269	3.715	9.2126E-08	5.2254E-07
	Dow_Jones	-4.09E-07	8.24E-08	-0.169		-4.966	2.05E-06	0.064	15.72	-5.7246E-07	-2.4633E-07
	Pers_Ind_Manuf	-1.73E-08	1.35E-09	-0.295		-12.8	5.41E-25	0.138	7.24	-1.9958E-08	-1.4614E-08

a. Variable Dependiente: INPC

Valor crítico de la prueba t de dos colas 1.977826

Ecuación de Regresión Backward:

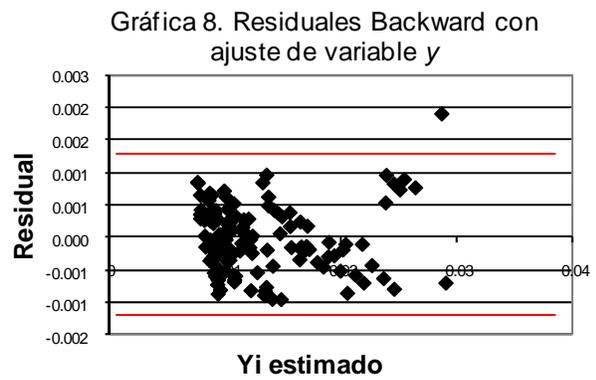
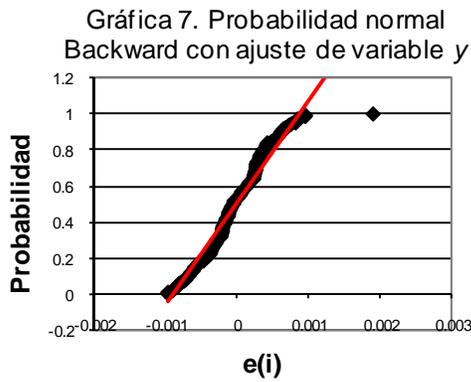
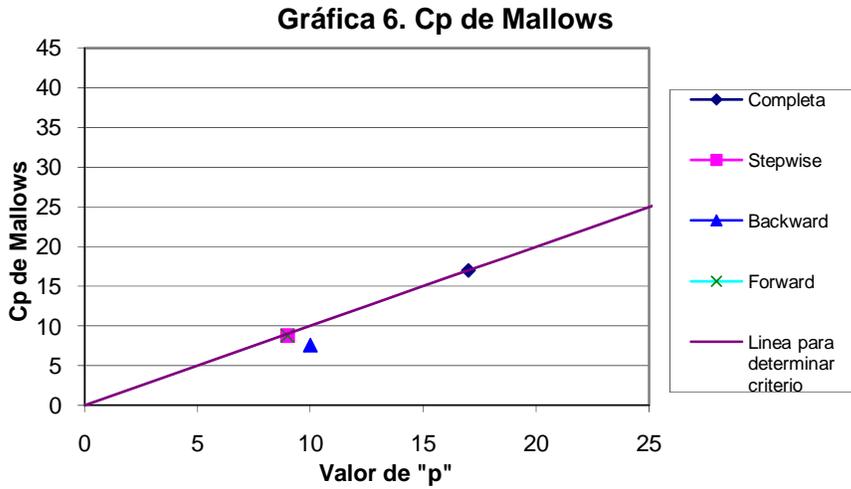
$$y_i = 8.08E-02 - 3.32E-04x_1 + 1.57E-07x_3 + 1.40E-07x_5 - 2.24E-09x_6 - 1.00E-03x_8 + 1.05E-04x_9 + 3.07E-07x_{12} - 4.09E-07x_{13} - 1.73E-08x_{16}$$

Tabla 12. Valores del estadístico PRESS con cambio de variable

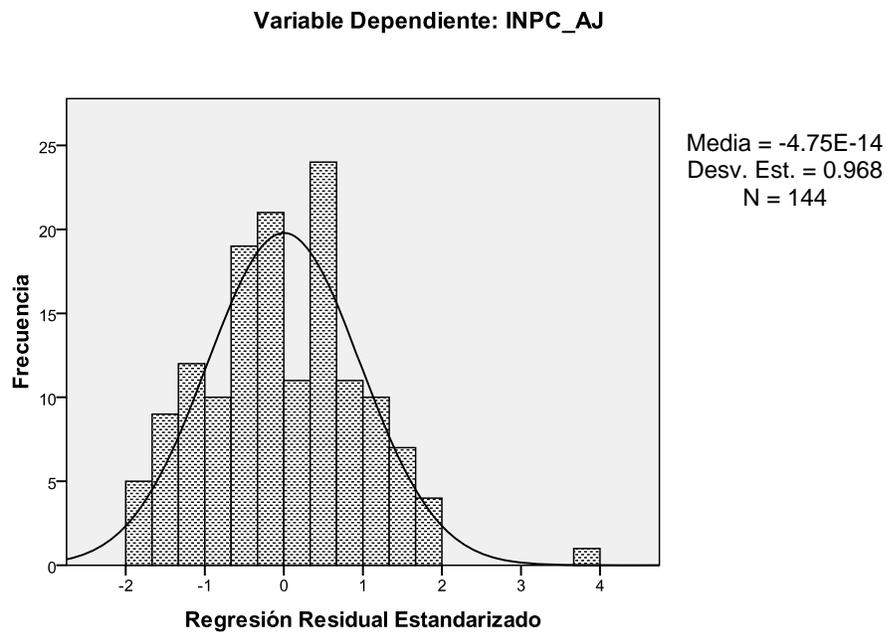
VALORES ESTADÍSTICO PRESS			
Completa	Stepwise	Backward	Forward
0.00010924	0.00004334	0.00004300	0.00004334

Tabla 13. Valor R² Predicción PRESS con cambio de variable

VALOR R ² PREDICCIÓN PRESS			
Completa	Stepwise	Backward	Forward
0.96929169	0.98781538	0.98791237	0.98781538



Gráfica 9. Histograma del segundo modelo seleccionado



La selección de este modelo se justifica con base a que los resultados obtenidos, así como las diversas pruebas estadísticas realizadas son las más satisfactorias, ya que nuevamente este modelo es el que presenta el mayor valor de R^2 , es el modelo que presenta el menor error promedio de predicción de acuerdo a la gráfica Cp de Mallows, pero lo más importante es que el estadístico PRESS resultó ser el menor de todos. Otro aspecto que surge a la vista después de haber realizado la transformación antes mencionada, es el de las “observaciones atípicas”, para este caso es solamente una observación la que puede considerarse como tal la cual se analiza en el Anexo B y se concluye que no es influyente, por lo que se decidió no omitirla de la base de datos y continuar con el modelo con todas las observaciones.

A pesar de haber estabilizado la varianza y de que un número considerable de las pruebas resultaron significativas no se puede considerar como un buen modelo de predicción esto debido a que la colinealidad no se ha podido solucionar, de igual manera sucede con el estadístico Durbin-Watson ya que sigue mostrando una autocorrelación negativa, mucho menos se puede considerar como un buen modelo econométrico ya que además de presentar estos problemas, surgen nuevos inconvenientes si lo que se quiere es considerarlo como tal, esto es, presenta síntomas de regresión espuria o falsa.

Como su nombre lo indica, en econometría la regresión espuria es toda aquella regresión que de acuerdo a los resultados obtenidos aparenta ser una regresión adecuada, sin embargo no puede ser utilizada como modelo de predicción puesto que las series utilizadas para la construcción del modelo se comportan de manera similar en el tiempo.

En econometría cuando se tienen dos series consideradas como modelos econométricos con cambios aleatorios

$$\begin{aligned}
 y_i &= y_{i-1} + u_i & u_i &\sim N(0, S^2_u) \\
 x_i &= x_{i-1} + e_i & e_i &\sim N(0, S^2_e) \\
 y_i &= \beta_0 + \beta_1 x_i + \varepsilon_i
 \end{aligned}$$

Algunos de los indicios que hacen suponer si una regresión es o no de este tipo son:

“1. La regresión de una caminata aleatoria en el tiempo por mínimos cuadrados producirá valores de R^2 de alrededor de 0.44 sin importar el tamaño de la muestra, si la media de la variable no tiene ninguna relación con el tiempo.

2. *En el caso de caminatas aleatorias con deriva, es decir, $\beta \neq 0$, R^2 será mayor y aumentará con el tamaño de la muestra, llegando a uno en el límite sin importar el valor de β .*
3. *El residuo de la regresión en el tiempo que se toma como la serie a la que se quita la tendencia, tiene el promedio sólo 14% de la varianza estocástica real de la serie original.*
4. *Los residuos de la regresión en el tiempo también están autocorrelacionados con alrededor de $(1-10/N)$ de correlación de orden 1, donde N es el tamaño de la muestra.*
5. *Las pruebas t convencionales para probar la significancia de algunos de los regresores no son válidas. Tienden a rechazar la hipótesis nula de no dependencia en el tiempo, con mucha frecuencia.*
6. *La regresión de una caminata aleatoria sobre otra, en la que se incluye el tiempo para la tendencia, está sujeta en gran medida a un fenómeno de regresión espuria. Es decir, la prueba t convencional tenderá a indicar una relación entre las variables cuando no existe ninguna de ellas.” [Maddala, (1996)]*

Algunas de las consecuencias que trae consigo esta regresión son:

- La probabilidad de obtener estimadores distintos de cero es muy alta debido a que el estadístico t es muy elevado
- El estadístico F es muy elevado, lo cual indica que la relación entre las variables es significativa
- Los valores de los estimadores pueden señalar relación significativa entre las variables

Si dos variables tienen cambio aleatorio indica que la varianza de ambas series aumenta con el tiempo. En otras palabras, la serie y_t se aleja de su media y en consecuencia genera valores de R^2 cercana a uno, lo cual crea una mala suposición de que el ajuste del modelo es muy bueno, sin embargo, esto es consecuencia de que las series se mueven juntas.

Una observación más acerca del problema de la regresión espuria es que el estadístico Durbin-Watson muestra un valor muy cercano a cero, esto debido a que como la serie es cambio aleatorio, los errores muestran un fuerte proceso de autocorrelación.

Para fines prácticos, existe una sencilla regla para determinar si la regresión en cuestión es falsa o no. Dicha regla es:

$$DW < R^2$$

Si esto ocurre la regresión es falsa y en consecuencia no se puede usar en la práctica ya que los resultados que emite no son confiables ya que las series se comportan de manera similar en el tiempo a su vez que los coeficientes no pueden utilizarse para un análisis o pronóstico económico. Ésta es la situación que se presenta en los modelos seleccionados anteriormente, ya que se tiene que en el primer caso $DW = 0.938 < 0.99 = R^2$; y para el segundo $DW = 0.903 < 0.99 = R^2$, con lo cual se concluye que ambos modelos son espurios y no se pueden utilizar en un análisis econométrico.

A pesar de que los modelos obtenidos presentan diversas anomalías, principalmente el de multicolinealidad, y no se pueden considerar como modelos econométricos, cabe mencionar que existen métodos alternativos para la corrección de la multicolinealidad. Dichos métodos son:

- Regresión Ridge
- Regresión Ridge General
- Regresión por Componentes Principales
- Estimación de Rango Fraccionario
- Análisis de Regresión de Raíz Latente

De igual manera se pueden aplicar más alternativas para poder obtener un modelo capaz de ser utilizado para predicción econométrica, entre otras existen:

- Trabajar con tasas de incremento
- Obtener el modelo mediante series de tiempo, es decir, verificar mediante el estadístico Dickey-Fuller si el modelo en cuestión es un Proceso de Tendencia Estacionaria o bien, un Proceso de Diferencia Estacionaria
- Trabajar con datos diferenciados en lugar de hacerlo con datos sobre niveles.

Conclusiones

La motivación para tomar una gran variedad de variables desde el inicio de la investigación que al parecer no tienen relación entre sí, como puede ser el caso, por ejemplo entre las variables llamadas “Personas de la Industria Manufacturera” y la llamada “Tipo de Cambio promedio mensual Peso-Dólar americano”; o el caso del “Índice Nacional de Precios y Cotizaciones” con la variable “Producción de Vehículos de Pasajeros” radica en dos puntos principalmente:

1. Se desconoce la capacidad potencialmente predictiva que cada una posee para poder explicar el comportamiento del INPC.
2. Se intentaron tomar variables de diferentes sectores de la economía para obtener un criterio más amplio sobre la investigación y poder determinar si ciertas variables que aparentemente no están en lo absoluto relacionadas en el comportamiento del INPC, al interactuar de forma conjunta con otras variables son buenos predictores de dicha variable.

Por tal motivo es que en este capítulo se retomarán las cuestiones que fueron propuestas en la Introducción, ya que en base a todos los resultados obtenidos e información recolectada se pretende dar una respuesta sólida, concreta y verídica a dichas interrogantes, las cuales son:

- Dentro de la gran variedad de variables de carácter económico que existen, ¿Cuáles son las que mayor influencia ejercen sobre la inflación de México?

Se sabe que dentro del inmenso espacio muestral de variables económicas que existen y que pueden ser utilizadas, no todas pueden representar de manera precisa otras variables, o puede que no estén tan relacionadas con otras, por tal motivo es que para este trabajo se consideraron las que se sospecha pueden ser algunas de las más útiles y funcionar de una mejor manera dentro de esta investigación. Así pues y de acuerdo a los resultados obtenidos por la paquetería SPSS y el nivel de significancia que se tomó del 5% para realizar las pruebas “t”, se podría afirmar que las variables más significativas en mayor número de ocasiones son:

1. Dow Jones
2. IPC
3. Tasa de interés Interbancaria de Equilibrio
4. Medio Circulante
5. Tipo de Cambio promedio mensual Peso-Dólar americano
6. Producción de Vehículos de Pasajeros
7. Reservas Internacionales
8. Exportaciones Totales de México

Sin embargo, esto sólo es una aproximación ya que pueden existir otros índices que sean mejores o igual de satisfactorios para construir modelos capaces de pronosticar el INPC.

- ¿Es posible construir un modelo con estas variables capaz de pronosticar confiablemente el comportamiento del INPC?

El Índice Nacional de Precios al Consumidor es un indicador económico que se obtiene mediante una fórmula matemática, la fórmula de Laspeyres, por lo que el resultado es exclusivamente matemático. Dicho valor tiene un incremento periódico a través del tiempo, sin descartar que puede tener altibajos, pero finalmente siempre incrementa (Grafica 1), sin embargo se desconoce en qué magnitud lo hará. Esto crea la suposición de que se puede pronosticar, y por lo tanto encontrar un modelo matemático capaz de hacerlo y con la ayuda de la Regresión Lineal es que dicha suposición es posible de realizar, sin embargo hay que tener en consideración cuales variables independientes son las que se deberán contemplar, ya que pueden existir indicadores numéricos, que no por el simple hecho de ser de este tipo son útiles para con la variable dependiente en cuestión ya que gracias a los diversos métodos que existen para obtener los modelos de regresión fue posible detectar su falta de significancia dentro de la regresión y por lo tanto descartarlas para el modelo final. Por lo tanto, después de tomar en cuenta un gran número de variables, se fueron descartando progresivamente las que se considera, son menos útiles en esta

investigación, llegando finalmente a las que ya se mencionaron a lo largo de esta tesis, y a su vez se fueron obteniendo diversos modelos con distintas combinaciones de variables de los cuales el que mejores resultados genera es:

$$y_i = 8.08E-02 - 3.32E-04x_1 + 1.57E-07x_3 + 1.4E-07x_5 - 2.24E-09x_6 - 1.0E-03x_8 + 1.05E-04x_9 + 3.07E-07x_{12} - 4.09E-07x_{13} - 1.73E-08x_{16}$$

- En caso de ser posible encontrar un modelo, ¿éste cumpliría con todas las condiciones necesarias para su liberación y finalmente su uso?

Retomando la pregunta anterior, se sabe que sí es posible encontrar, no sólo uno, sino una gran variedad de modelos, pero debido a las diversas pruebas que deben satisfacer para considerarse unos buenos modelos para predicción y análisis econométrico y poder considerarlos como tales, es que la mayoría de las propuestas no son útiles y no se pueden liberar ya que se pueden mostrar una serie de problemas, por ejemplo el caso de la regresión espuria, autocorrelación severa o multicolinealidad. Sin embargo, si al menos uno de los modelos presentados resulta mínimamente significativo, se puede llegar a la conclusión de que se puede perfeccionar con métodos matemáticos y estadísticos más complejos para finalmente desarrollar modelos más efectivos y de confiabilidad mayor que se puedan liberar y utilizar.

Ahora bien, la forma más adecuada de su uso sería mediante la planeación de escenarios tanto para el mejor como para el peor de los casos, esto es, plantear escenarios optimistas y pesimistas para cada variable independiente, los cuales deberían hacerse en tiempos futuros y deberían consistir en pronósticos sobre las situaciones que se espera tendrían cada una de las variables en estos tiempos, y estos a su vez son los que deberían ser utilizados en el modelo perfeccionado para finalmente obtener un pronóstico acertado de la variable dependiente en cuestión, en este caso, el INPC y por lo tanto de la Inflación de México.

Anexo A. Regresiones obtenidas y sus comentarios

A.1 Resultados obtenidos de la Regresión Completa:

Tabla A1. Resumen del Modelo

Modelo	R	R Cuadrada	R Cuadrada Ajustada	Error Estándar del Estimado	Durbin-Watson
1	0.998	0.997	0.996	1.476660758	0.962

a. Predictores: (Constante), Pers_Ind_Manuf, Bolsa_EUA, Tasa_CETE_real, Gastos_PEMEX, Produc_Veh_Pas, Tasa_Int_Banc, Nasdaq, TC_interb48h_Vta, IPC, Ingresos_PEMEX, Export_Mex, Circulante_mens, Rvas_Int, INPC_USA, Import_Mex, Dow_Jones

Tabla A2. ANOVA(b)

Modelo		Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F	Sig.
1	Regresión	87,651.101	16	5,478.194	2,512.326	1.698E-150
	Residual	276.927	127	2.181		
	Total	87,928.028	143			

a. Predictores: (Constante), Pers_Ind_Manuf, Bolsa_EUA, Tasa_CETE_real, Gastos_PEMEX, Produc_Veh_Pas, Tasa_Int_Banc, Nasdaq, TC_interb48h_Vta, IPC, Ingresos_PEMEX, Export_Mex, Circulante_mens, Rvas_Int, INPC_USA, Import_Mex, Dow_Jones

b. Variable Dependiente: INPC

Tabla A3. Coeficientes(a)

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados		Sig.	Estadísticas de Colinealidad		Intervalo de Confianza para B al 95.0%	
		B	Error Estándar	Beta	t		Tolerancia	VIF	Cota Inferior	Cota Superior
1	(Constante)	-211.351	16.799		-12.581	4.53E-24			-244.594	-178.108
	INPC_USA	2.16	0.162	0.844	13.303	7.89E-26	0.006	162.169	1.838	2.481
	Circulante_mens	7.20E-05	1.18E-05	0.302	6.089	1.25E-08	0.01	98.915	0	0
	Import_Mex	8.92E-04	3.92E-04	0.16	2.276	0.025	0.005	199.282	0	0.002
	Export_Mex	-1.37E-03	3.93E-04	-0.225	-3.488	6.69E-04	0.006	168.434	-0.002	0
	IPC	-1.39E-03	7.19E-05	-0.32	-19.376	9.66E-40	0.091	10.996	-0.002	-0.001
	Ingresos_PEMEX	4.80E-05	1.04E-05	0.107	4.595	1.03E-05	0.046	21.892	0	0
	Gastos_PEMEX	-7.60E-05	1.50E-05	-0.12	-5.058	1.45E-06	0.044	22.592	0	0
	TC_interb48h_Vta	1.453	0.319	0.083	4.553	1.22E-05	0.075	13.261	0.822	2.085
	Tasa_Int_Banc	-0.107	0.018	-0.062	-5.964	2.29E-08	0.228	4.39	-0.142	-0.071
	Produc_Veh_Pas	0.032	0.014	0.026	2.377	0.019	0.211	4.74	0.005	0.059
	Rvas_Int	-6.76E-05	5.64E-05	-0.05	-1.198	0.233	0.015	68.884	0	0
	Nasdaq	1.83E-04	3.48E-04	0.006	0.525	0.601	0.22	4.542	0	0.001
	Dow_Jones	3.87E-03	0.002	0.32	1.917	0.057	0.001	1,126.90	0	0.008
	Bolsa_EUA	-0.044	0.044	-0.166	-0.998	0.32	0.001	1,119.58	-0.13	0.043
	Tasa_CETE_real	-7.27E-04	2.25E-02	0	-0.032	0.974	0.776	1.289	-0.045	0.044
	Pers_Ind_Manuf	2.38E-05	4.98E-06	0.081	4.774	4.90E-06	0.085	11.751	0	0

a. Variable Dependiente: INPC

Valor crítico de la prueba t de dos colas 1.97882

A.2 Resultados obtenidos por el Método Stepwise:

Tabla A4. Resumen del Modelo

Modelo	R	R Cuadrada	R Cuadrada Ajustada	Error Estándar del Estimado	Durbin-Watson
10	0.998	0.996	0.996	1.586900572	0.957

j. Predictores: (Constante), INPC_USA, Pers_Ind_Manuf, IPC, Dow_Jones, Circulante_mens, Tasa_Int_Banc, TC_interb48h_Vta, Export_Mex, Produc_Veh_Pas, Rvas_Int

Tabla A5. ANOVA

Modelo		Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F	Sig.
10	Regresión	87,593.101	10	8,759.310	3,478.327	1.239E-155
	Residual	334.928	133	2.518		
	Total	87,928.028	143			

j. Predictores: (Constante), INPC_USA, Pers_Ind_Manuf, IPC, Dow_Jones, Circulante_mens, Tasa_Int_Banc, TC_interb48h_Vta, Export_Mex, Produc_Veh_Pas, Rvas_Int

k. Variable Dependiente: INPC

Tabla A6. Coeficientes

Modelo	Coeficientes no estandarizados		Coeficientes estandarizados		t	Sig.	Estadísticas de Colinealidad		Intervalo de Confianza para B al 95.0%	
	B	Error Estándar	Beta				Tolerancia	VIF	Cota Inferior	Cota Superior
10 (Constante)	-208.594	16.803			-12.414	5.59E-24			-241.829	-175.359
	INPC_USA	2.175	0.161	0.849	13.47	1.28E-26	0.007	138.87	1.855	2.494
	Pers_Ind_Manuf	1.75E-05	4.78E-06	0.06	3.651	3.74E-04	0.106	9.405	.000	.000
	IPC	-1.36E-03	7.53E-05	-0.313	-18.08	1.21E-37	0.096	10.456	-.002	-.001
	Dow_Jones	2.22E-03	2.32E-04	0.184	9.595	6.86E-17	0.078	12.848	.002	.003
	Circulante_mens	7.61E-05	1.11E-08	0.319	6.836	2.67E-10	0.013	75.857	.000	.000
	Tasa_Int_Banc	-0.119	0.018	-0.069	-6.43	2.11E-09	0.247	4.041	-.155	-.082
	TC_interb48h_Vta	1.888	0.313	0.107	6.03	1.52E-08	0.091	11.047	1.269	2.507
	Export_Mex	-0.001	1.61E-04	-0.113	-4.295	3.35E-05	0.041	24.362	-.001	.000
	Produc_Veh_Pas	0.046	0.014	0.036	3.291	1.28E-03	0.234	4.282	.018	.073
	Rvas_Int	-1.38E-04	5.44E-05	-0.101	-2.543	0.012	0.018	55.488	.000	.000

a. Variable Dependiente: INPC

Valor crítico de la prueba t de dos colas 1.977961

A.3 Resultados obtenidos por el Método Forward:

Tabla A7. Resumen del Modelo

Modelo	R	R Cuadrada	R Cuadrada Ajustada	Error Estándar del Estimado	Durbin-Watson
10	0.998	0.996	0.996	1.586900572	0.957

j. Predictores: (Constante), INPC_USA, Pers_Ind_Manuf, IPC, Dow_Jones, Circulante_mens, Tasa_Int_Banc, TC_interb48h_Vta, Export_Mex, Produc_Veh_Pas, Rvas_Int

Tabla A8. ANOVA(k)

Modelo		Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F	Sig.
10	Regresión	87,593.101	10	8,759.310	3,478.327	1.239E-155
	Residual	334.928	133	2.518		
	Total	87,928.028	143			

j. Predictores: (Constante), INPC_USA, Pers_Ind_Manuf, IPC, Dow_Jones, Circulante_mens, Tasa_Int_Banc, TC_interb48h_Vta, Export_Mex, Produc_Veh_Pas, Rvas_Int

k. Variable Dependiente: INPC

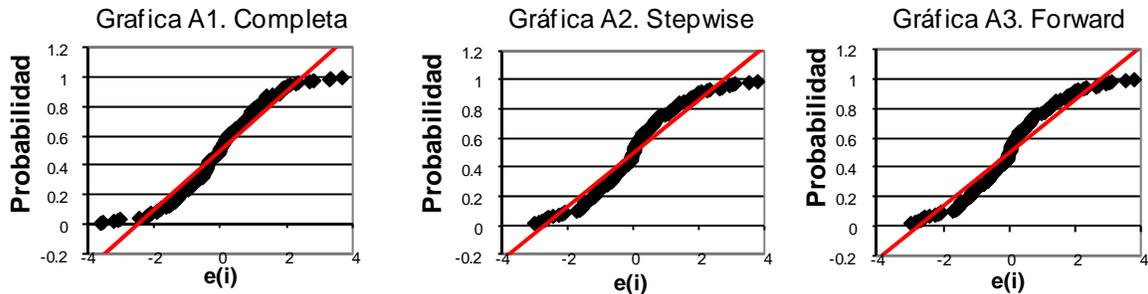
Tabla A9. Coeficientes

Modelo	Coeficientes no estandarizados		Coeficientes estandarizados		t	Sig.	Estadísticas de Colinealidad		Intervalo de Confianza para B al 95.0%	
	B	Error Estándar	Beta				Tolerancia	VIF	Cota Inferior	Cota Superior
10 (Constante)	-208.594	16.803			-12.414	5.59E-24			-241.829	-175.359
	INPC_USA	2.175	0.161	0.849	13.47	1.28E-26	0.007	138.87	1.855	2.494
	Pers_Ind_Manuf	1.75E-05	4.78E-06	0.06	3.651	3.74E-04	0.106	9.405	.000	.000
	IPC	-1.36E-03	7.53E-05	-0.313	-18.08	1.21E-37	0.096	10.456	-.002	-.001
	Dow_Jones	2.22E-03	2.32E-04	0.184	9.595	6.86E-17	0.078	12.848	.002	.003
	Circulante_mens	7.61E-05	1.11E-08	0.319	6.836	2.67E-10	0.013	75.857	.000	.000
	Tasa_Int_Banc	-0.119	0.018	-0.069	-6.43	2.11E-09	0.247	4.041	-.155	-.082
	TC_interb48h_Vta	1.888	0.313	0.107	6.03	1.52E-08	0.091	11.047	1.269	2.507
	Export_Mex	-0.001	1.61E-04	-0.113	-4.295	3.35E-05	0.041	24.362	-.001	.000
	Produc_Veh_Pas	0.046	0.014	0.036	3.291	1.28E-03	0.234	4.282	.018	.073
	Rvas_Int	-1.38E-04	5.44E-05	-0.101	-2.543	0.012	0.018	55.488	.000	.000

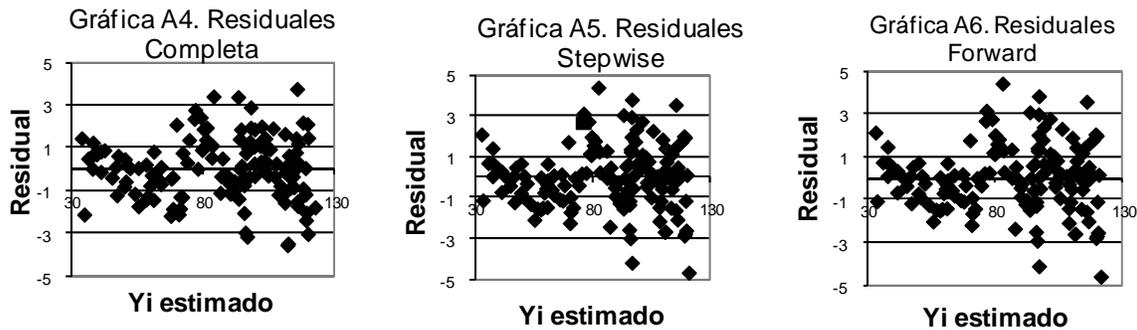
a. Variable Dependiente: INPC

Valor crítico de la prueba t de dos colas 1.97796

A.4 Gráficas de Probabilidad Normal de los Residuales de los primeros 4 modelos obtenidos



A.5 Gráficas de Residuales de los primeros cuatro modelos



Como se puede observar en cada una de las gráficas de residuales (eje Y) en función de los valores estimados (eje X), en todas y cada una de ellas (residuales por el método Completo (Gráfica A4), Stepwise (Gráfica A5), y Forward (Gráfica A6)) se llega a la conclusión de que la varianza de los mismos es creciente, es decir, a medida que incrementa el valor de los valores estimados, la varianza se vuelve más grande; lo cual es muy claro, ya que se observa que la dispersión de los residuales tiene forma de cono que abre hacia la derecha. Esto es muy insatisfactorio en cualquier modelo de regresión, ya que un requisito imprescindible es la premisa de homocedasticidad o varianza constante. Sin embargo, esto se puede solucionar mediante una transformación de la variable dependiente o bien utilizar el método de mínimos cuadrados ponderados.

A.6 Resultados obtenidos de la Regresión Completa con el ajuste en la variable INPC:

Tabla A10. Resumen del Modelo

Modelo	R	R Cuadrada	R Cuadrada Ajustada	Error Estándar del Estimado	Durbin-Watson
1	0.995	0.991	0.989	0.000515297	0.988

a. Predictores: (Constante), Pers_Ind_Manuf, Bolsa_EUA, Tasa_CETE_real, Gastos_PEMEX, Produc_Veh_Pas, Tasa_Int_Banc, Nasdaq, TC_interb48h_Vta, IPC, Ingresos_PEMEX, Export_Mex, Circulante_mens, Rvas_Int, INPC_USA, Import_Mex, Dow_Jones

Tabla A11. ANOVA(b)

Modelo		Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F	Sig.
1	Regresión	3.523E-03	16	2.202E-04	829.352	3.983E-120
	Residual	3.372E-05	127	2.655E-07		
	Total	3.557E-03	143			

a. Predictors: (Constante), Pers_Ind_Manuf, Bolsa_EUA, Tasa_CETE_real, Gastos_PEMEX, Produc_Veh_Pas, Tasa_Int_Banc, Nasdaq, TC_interb48h_Vta, IPC, Ingresos_PEMEX, Export_Mex, Circulante_mens, Rvas_Int, INPC_USA, Import_Mex, Dow_Jones

b. Variable Dependiente: INPC

Tabla A12. Coeficientes(a)

Modelo	Coeficientes no estandarizados		Coeficientes estandarizados		t	Sig.	Estadísticas de Colinealidad		Intervalo de Confianza para B al 95.0%	
	B	Error Estándar	Beta	t			Tolerancia	VIF	Cota Inferior	Cota Superior
1 (Constante)	8.51E-02	5.86E-03			14.52	0			7.352E-02	9.672E-02
INPC_USA	-3.79E-04	5.67E-05	-0.736	-6.694	0	0.006	162.169	-4.913E-04	-2.671E-04	
Circulante_mens	2.41E-09	4.13E-09	0.05	0.584	0.56	0.01	98.915	-5.758E-09	1.058E-08	
Import_Mex	7.50E-08	1.37E-07	0.067	0.548	0.585	0.005	199.282	-1.957E-07	3.456E-07	
Export_Mex	7.55E-08	1.37E-07	0.062	0.55	0.583	0.006	168.434	-1.961E-07	3.472E-07	
IPC	1.28E-07	2.51E-08	0.146	5.103	0	0.091	10.996	7.837E-08	1.776E-07	
Ingresos_PEMEX	-5.78E-09	3.64E-09	-0.064	-1.586	0.115	0.046	21.892	-1.299E-08	1.433E-09	
Gastos_PEMEX	5.88E-09	5.24E-09	0.046	1.121	0.264	0.044	22.592	-4.496E-09	1.625E-08	
TC_interb48h_Vta	-1.07E-03	1.11E-04	-0.301	-9.581	0	0.075	13.261	-1.287E-03	-8.467E-04	
Tasa_Int_Banc	1.06E-04	6.24E-06	0.306	16.907	0	0.228	4.39	9.315E-05	1.178E-04	
Produc_Veh_Pas	1.21E-06	4.74E-06	0.005	0.256	0.799	0.211	4.74	-8.172E-06	1.060E-05	
Rvas_Int	2.11E-08	1.97E-08	0.077	1.071	0.286	0.015	68.884	-1.788E-08	6.005E-08	
Nasdaq	3.34E-07	1.21E-07	0.051	2.751	0.007	0.22	4.542	9.369E-08	5.742E-07	
Dow_Jones	-1.14E-06	7.04E-07	-0.471	-1.623	0.107	0.001	1,126.89	-2.537E-06	2.506E-07	
Bolsa_EUA	1.60E-05	1.52E-05	0.304	1.052	0.295	0.001	1,119.57	-1.413E-05	4.617E-05	
Tasa_CETE_real	-2.22E-06	7.85E-06	-0.003	-0.283	0.778	0.776	1.289	-1.775E-05	1.331E-05	
Pers_Ind_Manuf	-1.71E-08	1.74E-09	-0.291	-9.838	0	0.085	11.751	-2.052E-08	-1.365E-08	

a. Variable Dependiente: INPC

Valor crítico de la prueba t de dos colas 1.97882

A.7 Resultados obtenidos por el método Stepwise con el ajuste en la variable INPC:

Tabla A13. Resumen del Modelo

Modelo	R	R Cuadrada	R Cuadrada Ajustada	Error Estándar del Estimado	Durbin-Watson
8	0.995	0.99	0.989	0.000514865	.852

h. Predictores: (Constante), TC_interb48h_Vta, Tasa_Int_Banc, Pers_Ind_Manuf, INPC_USA, IPC, Dow_Jones, Nasdaq, Export_Mex

Tabla A14. ANOVA

Modelo		Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F	Sig.
8	Regresión	3.52E-03	8	4.40E-04	1,660.51	8.15E-131
	Residual	3.58E-05	135	2.65E-07		
	Total	3.56E-03	143			

h. Predictores: (Constante), TC_interb48h_Vta, Tasa_Int_Banc, Pers_Ind_Manuf, INPC_USA, IPC, Dow_Jones, Nasdaq, Export_Mex

i. Variable Dependiente: INPC

Tabla A15. Coeficientes

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Estadísticas de Colinealidad		Intervalo de Confianza para B al 95.0%	
		B	Error Estándar	Beta			Tolerancia	VIF	Cota Inferior	Cota Superior
8	(Constante)	0.078	0.004		20.723	9.07E-44			7.035E-02	8.519E-02
	TC_interb48h_Vta	-1.02E-03	8.40E-05	-0.287	-12.103	2.72E-23	0.132	7.556	-1.183E-03	-8.505E-04
	Tasa_Int_Banc	1.07E-04	5.78E-06	0.309	18.45	1.01E-38	0.265	3.768	9.514E-05	1.180E-04
	Pers_Ind_Manuf	-1.64E-08	1.28E-09	-0.279	-12.808	4.47E-25	0.157	6.369	-1.889E-08	-1.383E-08
	INPC_USA	-3.12E-04	3.21E-05	-0.606	-9.722	2.93E-17	0.019	52.206	-3.758E-04	-2.487E-04
	IPC	1.37E-07	2.25E-08	0.156	6.085	1.13E-08	0.113	8.86	9.239E-08	1.814E-07
	Dow_Jones	-3.94E-07	8.37E-08	-0.162	-4.704	6.22E-06	0.063	15.94	-5.593E-07	-2.282E-07
	Nasdaq	3.04E-07	1.12E-07	0.046	2.726	7.25E-03	0.26	3.84	8.352E-08	5.247E-07
	Export_Mex	1.21E-07	4.78E-08	0.099	2.538	1.23E-02	0.049	20.453	2.678E-08	2.159E-07

a. Variable Dependiente: INPC

Valor crítico de la prueba t de dos colas 1.977692

A. 8 Resultados obtenidos por el método Forward con el ajuste en la variable INPC:

Tabla A16. Resumen del Modelo

Modelo	R	R Cuadrada	R Cuadrada Ajustada	Error Estándar del Estimado	Durbin-Watson
8	0.995	0.99	0.989	0.000514865	.852

h. Predictores: (Constante), TC_interb48h_Vta, Tasa_Int_Banc, Pers_Ind_Manuf, INPC_USA, IPC, Dow_Jones, Nasdaq, Export_Mex

Tabla A17. ANOVA

Modelo		Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F	Sig.
8	Regresión	3.52E-03	8	4.40E-04	1,660.51	8.15E-131
	Residual	3.58E-05	135	2.65E-07		
	Total	3.56E-03	143			

h. Predictores: (Constante), TC_interb48h_Vta, Tasa_Int_Banc, Pers_Ind_Manuf, INPC_USA, IPC, Dow_Jones, Nasdaq, Export_Mex

i. Variable Dependiente: INPC

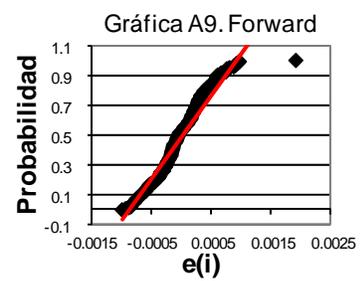
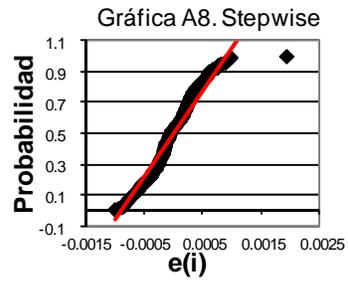
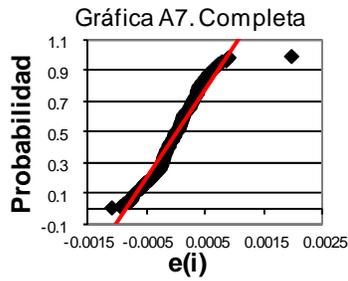
Tabla A18. Coeficientes

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Estadísticas de Colinealidad		Intervalo de Confianza para B al 95.0%	
		B	Error Estándar	Beta			Tolerancia	VIF	Cota Inferior	Cota Superior
8	(Constante)	0.078	0.004		20.723	9.07E-44			7.035E-02	8.519E-02
	TC_interb48h_Vta	-1.02E-03	8.40E-05	-0.287	-12.103	2.72E-23	0.132	7.556	-1.183E-03	-8.505E-04
	Tasa_Int_Banc	1.07E-04	5.78E-06	0.309	18.45	1.01E-38	0.265	3.768	9.514E-05	1.180E-04
	Pers_Ind_Manuf	-1.64E-08	1.28E-09	-0.279	-12.808	4.47E-25	0.157	6.369	-1.889E-08	-1.383E-08
	INPC_USA	-3.12E-04	3.21E-05	-0.606	-9.722	2.93E-17	0.019	52.206	-3.758E-04	-2.487E-04
	IPC	1.37E-07	2.25E-08	0.156	6.085	1.13E-08	0.113	8.86	9.239E-08	1.814E-07
	Dow_Jones	-3.94E-07	8.37E-08	-0.162	-4.704	6.22E-06	0.063	15.94	-5.593E-07	-2.282E-07
	Nasdaq	3.04E-07	1.12E-07	0.046	2.726	7.25E-03	0.26	3.84	8.352E-08	5.247E-07
	Export_Mex	1.21E-07	4.78E-08	0.099	2.538	1.23E-02	0.049	20.453	2.678E-08	2.159E-07

a. Variable Dependiente: INPC

Valor crítico de la prueba t de dos colas 1.97769

A.9 Gráficas de Probabilidad Normal de los Residuales de los modelos obtenidos con el ajuste de variable



A.10 Gráficas de residuales con el ajuste de la variable dependiente INPC



Anexo B. Observaciones atípicas

Una observación atípica se define como aquella observación que se encuentra muy alejada del resto del conjunto de datos. Para determinar si las observaciones atípicas son influyentes se debe repetir el análisis omitiendo estos datos y posteriormente observar los resultados obtenidos, si éstos varían de manera drástica las observaciones atípicas se denominan como influyentes, de lo contrario no son consideradas como tal y por lo tanto pueden mantenerse.

“El efecto de los valores atípicos sobre el modelo de regresión se puede comprobar con facilidad eliminándolos y volviendo a ajustar la ecuación de regresión. Se podría encontrar que los valores de los coeficientes de regresión, o de los estadísticos de resumen como “t”, F o R², y que el cuadrado medio de residuales pueden ser muy sensibles a los valores atípicos...” [Montgomery, (2002)]

Estas observaciones pueden identificarse de manera más sencilla en las gráficas de dispersión de los residuales así como en las de probabilidad normal (Gráficas A7, A8, A9, A10, A11 y A12).

En el caso de esta investigación se pudo identificar una observación atípica la cual corresponde al mes de Marzo de 1995 y en base a lo anterior se volvió a realizar el análisis para determinar si es influyente. Los resultados de este nuevo estudio se presentan a continuación:

Tabla B1. Estudio de Observación atípica

Regresión Completa			Regresión Método Backward		
	Con Observación 1	Sin observación 1		Con Observación 1	Sin observación 1
Valor de “F”	829.3515522	906.8373926	Valor de “F”	1500.756313	1474.65354
R ²	0.99052	0.991390737	R ²	0.990176534	0.991128159
MSRes	2.66E-07	2.21E-07	MSRes	2.61E-07	2.18E-07

Regresión Método Stepwise			Regresión Método Forward		
	Con Observación 1	Sin observación 1		Con Observación 1	Sin observación 1
Valor de “F”	1660.514065	1606.729141	Valor de “F”	1660.514065	1606.729141
R ²	0.989939722	0.990886392	R ²	0.989939722	0.990886392
MSRes	2.65E-07	2.22E-07	MSRes	2.65E-07	2.22E-07

Por lo tanto, después de observar los resultados Tabla B1 se puede concluir que la observación atípica de este conjunto de datos, no es influyente en el modelo ya que los nuevos cuadrados medios de residuales (MSRes), las nuevas “F” y las nuevas R^2 no difieren demasiado con los obtenidos con todas las observaciones iniciales, por tal motivo se puede seguir con el modelo incluyendo todos los datos sin la necesidad de omitir o replantear todo el modelo para ajustar algún cambio considerable.

Anexo C. Análisis de multicolinealidad y correlación

Muchas veces en los modelos de regresión se pueden obtener estadísticos que suponen que los modelos son idóneos, por ej., se puede obtener una R^2 muy alta, la prueba F puede resultar significativa para el modelo, etc. Sin embargo el cumplimiento de estos puntos no significa que sea un buen prospecto, ya que además de eso, no debe existir asociación lineal entre las variables independientes, es decir, que no exista multicolinealidad.

“La situación en la que las variables explicativas tienen una alta intercorrelación se conoce como multicolinealidad. Cuando las variables explicativas tienen una alta intercorrelación, puede ser difícil desenmarañar los efectos independientes de cada variable explicativa sobre la variable explicada” [Maddala, (1996)].

Para conocer dicha presencia, se debe observar el valor de los Factores de Inflación de Varianza (VIF), si éste resulta menor que 5 ó 10 en todas las variables entonces no hay multicolinealidad. A su vez si el Índice de Condición oscila entre 30 y 100 la colinearidad existe pero no es grave, si se encuentra entre 100 y 1000 es grave pero aún no es de preocupación sin embargo se deben tomar ciertas precauciones.

También se pueden observar las proporciones de varianza, las cuales para descartar por completo o no la multicolinealidad, deben ser menores que 0.5; este indicador es muy útil, ya que muestra las variables que están envueltas en este problema y de esta manera, poder tomar las medidas necesarias para determinar si se puede continuar o no con el modelo, ya que estas variables pueden ser altamente explicativas de la variable dependiente y por lo tanto no deberían ser omitidas del modelo.

A continuación se presentan los estudios de colinealidad de todos los modelos presentados, así como los coeficientes de correlación de las variables independientes.

Tabla C1. Diagnóstico de Colinealidad del modelo Completo

Modelo	Dimension	Eigenvalor	Índice de Condición	Proporciones de Varianza															
				(Constante)	INPC_USA	Circulante_mens	Import_Mex	Export_Mex	IPC	Ingresos_PEMEX	Gastos_PEMEX	TC_interb48h_Vta	Tasa_Int_Banc	Produc_Veh_Pas	Rvas_Int	Nasdaq	Dow_Jones	Bolsa_EUA	Tasa_CETE_real
1	1	14.68	1.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	1.162	3.555	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.10	0.00
	3	0.443	5.755	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.03	0.00	0.00	0.00	0.00	0.01	0.00
	4	0.390	6.137	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.74	0.00
	5	0.163	9.477	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.11	0.00	0.00	0.04	0.00	0.03	0.00
	6	0.084	13.221	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.05	0.00	0.00	0.15	0.00	0.00	0.00
	7	0.032	21.431	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.00	0.00	0.04	0.21	0.01	0.07	0.00	0.00	0.00
	8	0.017	29.040	0.00	0.00	0.01	0.00	0.00	0.38	0.03	0.02	0.00	0.27	0.03	0.01	0.03	0.00	0.00	0.00
	9	0.013	33.736	0.00	0.00	0.00	0.00	0.00	0.04	0.60	0.55	0.00	0.00	0.02	0.00	0.03	0.00	0.01	0.00
	10	0.007	46.483	0.00	0.00	0.00	0.00	0.00	0.03	0.05	0.04	0.00	0.24	0.02	0.00	0.39	0.00	0.00	0.00
	11	0.004	57.436	0.00	0.00	0.12	0.03	0.04	0.00	0.01	0.05	0.00	0.00	0.40	0.00	0.07	0.00	0.00	0.00
	12	0.003	72.506	0.00	0.00	0.27	0.00	0.00	0.18	0.01	0.00	0.10	0.03	0.09	0.30	0.03	0.00	0.00	0.00
	13	0.001	107.416	0.00	0.00	0.03	0.01	0.00	0.16	0.06	0.01	0.74	0.05	0.01	0.53	0.07	0.00	0.00	0.00
	14	0.000	205.502	0.00	0.01	0.01	0.41	0.30	0.01	0.16	0.27	0.12	0.01	0.08	0.11	0.00	0.00	0.00	0.34
	15	0.000	263.294	0.02	0.02	0.12	0.53	0.63	0.00	0.01	0.00	0.00	0.00	0.06	0.01	0.06	0.00	0.00	0.51
	16	0.000	766.829	0.96	0.96	0.39	0.01	0.02	0.03	0.01	0.00	0.02	0.10	0.08	0.03	0.05	0.01	0.00	0.14
	17	0.000	829.551	0.01	0.00	0.04	0.00	0.00	0.01	0.04	0.04	0.01	0.01	0.00	0.00	0.00	0.99	0.99	0.00

a. Variable Dependiente: INPC

Tabla C2. Diagnóstico de Colinealidad del modelo Stepwise

Modelo	Dimension	Eigenvalor	Índice de Condición	Proporciones de Varianza												
				(Constante)	INPC_USA	Pers_Ind_Manuf	IPC	Dow_Jones	Circulante_mens	Tasa_Int_Banc	TC_interb48h_Vta	Export_Mex	Produc_Veh_Pas	Rvas_Int		
10	1	9.974	1.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	0.812	3.504	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00
	3	0.137	8.531	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.21	0.00	0.00	0.00	0.00	0.00
	4	0.042	15.328	0.00	0.00	0.00	0.09	0.00	0.01	0.01	0.02	0.00	0.00	0.13	0.01	0.00
	5	0.016	24.674	0.00	0.00	0.00	0.33	0.03	0.02	0.02	0.17	0.00	0.02	0.20	0.01	0.00
	6	0.009	32.726	0.00	0.00	0.00	0.10	0.24	0.01	0.35	0.00	0.00	0.00	0.01	0.00	0.00
	7	0.004	51.600	0.00	0.00	0.00	0.01	0.00	0.10	0.00	0.00	0.00	0.74	0.33	0.00	0.00
	8	0.003	54.451	0.00	0.00	0.00	0.14	0.00	0.35	0.04	0.05	0.03	0.03	0.03	0.37	0.00
	9	0.001	81.875	0.00	0.00	0.00	0.18	0.10	0.01	0.05	0.90	0.00	0.01	0.01	0.37	0.00
	10	0.000	174.546	0.02	0.03	0.84	0.01	0.60	0.10	0.00	0.04	0.05	0.12	0.12	0.09	0.00
	11	0.000	588.226	0.98	0.97	0.15	0.04	0.02	0.40	0.09	0.00	0.15	0.16	0.16	0.14	0.00

a. Variable Dependiente: INPC

Tabla C3. Diagnóstico de Colinealidad del modelo Backward

Modelo	Dimension	Eigenvalor	índice de Condición	Proporciones de Varianza												
				(Constante)	INPC_USA	Circulante_mens	Import_Mex	Export_Mex	IPC	Ingresos_PEMEX	Gastos_PEMEX	TC_interb48h_Vta	Tasa_Int_Banc	Produc_Veh_Pas	Dow_Jones	Pers_Ind_Manuf
5	1	11.47	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	0.91	3.56	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00
	3	0.41	5.31	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.01	0.00	0.03	0.00	0.00	0.00
	4	0.14	9.16	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.20	0.00	0.00	0.00
	5	0.03	18.43	0.00	0.00	0.02	0.00	0.00	0.03	0.00	0.01	0.01	0.00	0.17	0.00	0.00
	6	0.02	26.76	0.00	0.00	0.04	0.00	0.00	0.45	0.02	0.00	0.00	0.22	0.04	0.02	0.00
	7	0.01	29.25	0.00	0.00	0.00	0.00	0.00	0.00	0.62	0.52	0.00	0.02	0.07	0.03	0.00
	8	0.01	35.85	0.00	0.00	0.00	0.00	0.00	0.06	0.19	0.17	0.00	0.26	0.02	0.20	0.00
	9	0	48.93	0.00	0.00	0.14	0.02	0.04	0.00	0.01	0.03	0.00	0.02	0.38	0.02	0.00
	10	0	76.07	0.00	0.00	0.08	0.00	0.00	0.30	0.02	0.02	0.90	0.00	0.06	0.05	0.00
	11	0	173.88	0.01	0.01	0.07	0.37	0.23	0.02	0.09	0.21	0.02	0.00	0.13	0.58	0.36
	12	0	220.64	0.01	0.02	0.17	0.58	0.72	0.00	0.00	0.01	0.00	0.00	0.05	0.10	0.40
	13	0	635.86	0.98	0.96	0.48	0.03	0.00	0.02	0.04	0.00	0.07	0.20	0.09	0.01	0.23

a. Variable Dependiente: INPC

Tabla C4. Diagnóstico de Colinealidad del modelo Forward

Modelo	Dimension	Eigenvalor	índice de Condición	Proporciones de Varianza											
				(Constante)	INPC_USA	Pers_Ind_Manuf	IPC	Dow_Jones	Circulante_mens	Tasa_Int_Banc	TC_interb48h_Vta	Export_Mex	Produc_Veh_Pas	Rvas_Int	
10	1	9.974	1.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	0.812	3.504	0.00	0.00	0.00	0.01	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00
	3	0.137	8.531	0.00	0.00	0.00	0.10	0.00	0.00	0.21	0.00	0.00	0.00	0.00	0.00
	4	0.042	15.328	0.00	0.00	0.00	0.09	0.00	0.01	0.02	0.00	0.00	0.13	0.01	0.01
	5	0.016	24.674	0.00	0.00	0.00	0.33	0.03	0.02	0.17	0.00	0.02	0.20	0.01	0.01
	6	0.009	32.726	0.00	0.00	0.00	0.10	0.24	0.01	0.35	0.00	0.00	0.01	0.00	0.00
	7	0.004	51.600	0.00	0.00	0.00	0.01	0.00	0.10	0.00	0.00	0.74	0.33	0.00	0.00
	8	0.003	54.451	0.00	0.00	0.00	0.14	0.00	0.35	0.04	0.05	0.03	0.03	0.37	0.37
	9	0.001	81.875	0.00	0.00	0.00	0.18	0.10	0.01	0.05	0.90	0.00	0.01	0.37	0.37
	10	0.000	174.546	0.02	0.03	0.84	0.01	0.60	0.10	0.00	0.04	0.05	0.12	0.09	0.09
	11	0.000	588.226	0.98	0.97	0.15	0.04	0.02	0.40	0.09	0.00	0.15	0.16	0.14	0.14

a. Variable Dependiente: INPC

Tabla C7. Diagnóstico de Colinealidad del modelo Backward con ajuste de variable

Modelo	Dimension	Eigenvalor	Índice de Condición	Proporciones de Varianza										
				(Constante)	INPC_USA	Import_Mex	IPC	Ingresos_PEMEX	TC_interb48h_Via	Tasa_Int_Banc	Nasdaq	Dow_Jones	Pers_Ind_Manuf	
8	1	8.776	1.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	0.732	3.461	0.00	0.00	0.00	0.01	0.05	0.00	0.07	0.00	0.00	0.00	0.00
	3	0.263	5.777	0.00	0.00	0.00	0.00	0.54	0.00	0.05	0.01	0.00	0.00	0.00
	4	0.139	7.943	0.00	0.00	0.00	0.17	0.09	0.00	0.10	0.03	0.00	0.00	0.00
	5	0.073	10.981	0.00	0.00	0.00	0.01	0.04	0.00	0.09	0.28	0.00	0.00	0.00
	6	0.008	32.913	0.00	0.00	0.26	0.36	0.09	0.01	0.34	0.02	0.00	0.01	0.01
	7	0.004	46.658	0.00	0.00	0.00	0.00	0.01	0.06	0.09	0.56	0.53	0.00	0.00
	8	0.004	47.496	0.00	0.00	0.19	0.02	0.01	0.40	0.07	0.00	0.04	0.02	0.02
	9	0.001	113.885	0.01	0.06	0.04	0.40	0.00	0.35	0.03	0.05	0.29	0.37	0.37
	10	0.000	397.768	0.98	0.94	0.51	0.02	0.15	0.17	0.16	0.05	0.13	0.60	0.60

Variable Dependiente: INPC

Tabla C8. Diagnóstico de Colinealidad del modelo Forward con ajuste de variable

Modelo	Dimension	Eigenvalor	Índice de Condición	Proporciones de Varianza									
				(Constante)	TC_interb48h_Via	Tasa_Int_Banc	Pers_Ind_Manuf	INPC_USA	IPC	Dow_Jones	Nasdaq	Export_Mex	
8	1	8.124	1.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	0.634	3.580	0.00	0.00	0.10	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	3	0.149	7.373	0.00	0.00	0.12	0.00	0.00	0.12	0.00	0.05	0.00	0.00
	4	0.076	10.331	0.00	0.00	0.09	0.00	0.00	0.03	0.00	0.25	0.00	0.00
	5	0.008	31.416	0.00	0.01	0.22	0.01	0.00	0.42	0.00	0.00	0.33	0.33
	6	0.005	42.140	0.00	0.13	0.29	0.01	0.00	0.00	0.30	0.20	0.07	0.07
	7	0.004	46.594	0.00	0.35	0.02	0.01	0.00	0.06	0.24	0.37	0.11	0.11
	8	0.001	112.243	0.02	0.34	0.02	0.45	0.07	0.33	0.26	0.06	0.07	0.07
	9	0.000	350.701	0.98	0.17	0.14	0.52	0.93	0.02	0.20	0.06	0.42	0.42

a. Variable Dependiente: INPC

Tabla C9. Coeficientes de Correlación

	Pers_Ind_Manuf	Bolsa_EUA	Tasa_CETE_real	Gastos_PEMEX	Produc_Veh_Pas	Tasa_Int_Banc	Nasdaq	TC_interb48h_Vta	IPC	Ingresos_PEMEX	Export_Mex	Circulante_mens	Rvas_Int	INPC_USA	Import_Mex	Dow_Jones	
Correlaciones	Pers_Ind_Manuf	1	-0.030	0.120	-0.297	-0.470	0.169	-0.058	-0.301	0.078	0.341	0.200	0.086	0.192	0.213	-0.192	-0.042
	Bolsa_EUA		1	0.047	0.202	0.020	-0.079	0.029	-0.103	-0.113	-0.189	-0.034	0.152	-0.066	-0.010	-0.006	-0.991
	Tasa_CETE_real			1	-0.092	-0.014	-0.034	0.061	-0.177	-0.047	0.128	-0.141	0.087	0.019	0.112	0.058	-0.063
	Gastos_PEMEX				1	0.198	-0.138	0.032	0.262	0.100	-0.927	0.236	0.020	-0.220	0.032	-0.344	-0.175
	Produc_Veh_Pas					1	-0.022	0.070	0.184	-0.036	-0.104	-0.131	0.108	0.141	-0.225	-0.083	0.006
	Tasa_Int_Banc						1	-0.005	-0.227	-0.273	0.212	-0.034	-0.232	0.291	0.286	-0.044	0.099
	Nasdaq							1	-0.179	0.025	0.028	-0.286	-0.248	0.263	0.267	0.120	-0.090
	TC_interb48h_Vta								1	0.436	-0.309	0.131	0.015	-0.530	-0.130	-0.095	0.107
	IPC									1	-0.089	-0.167	-0.365	-0.134	0.161	0.061	0.084
	Ingresos_PEMEX										1	-0.135	-0.125	0.314	0.078	0.142	0.168
	Export_Mex											1	0.248	-0.096	-0.198	-0.881	0.061
	Circulante_mens												1	-0.202	-0.680	-0.103	-0.155
	Rvas_Int													1	-0.209	0.033	0.048
	INPC_USA														1	-0.056	-0.005
	Import_Mex															1	-0.024
	Dow_Jones																

a. Variable Dependiente: INPC

El grado de correlación puede ser determinado de manera arbitraria por el investigador dependiendo del criterio que quiera imponer para una mejor precisión del modelo; para fines prácticos se puede tomar el punto intermedio, es decir 0.5, por lo tanto se pueden considerar como variables con alta correlación a aquellas que tienen un nivel, en valor absoluto, mayor que 0.5. En la Tabla C9 se puede observar que aparentemente no existe relación lineal significativa entre las variables independientes, por lo cual se podría pensar que las variables usadas son las indicadas para construir el modelo de regresión final, sin embargo esto no siempre es correcto, ya que puede no existir correlación por parejas de variables pero al actuar de forma conjunta en un modelo de regresión múltiple puede existir cierta colinealidad entre las variables que no la presentaron en un principio y viceversa.

De los estudios de colinearidad, se puede concluir que los modelos mostraron síntomas de que existe la misma sin importar el método que se utilice para obtenerlos,

aunque en ninguno de los casos se presentó de manera severa o grave, salvo el primer modelo que contenía todas las variables donde si hubo colinealidad severa.

A su vez, gracias al estudio de los Factores de Inflación de Varianza (VIF) es que se pudo revelar en el modelo completo que únicamente 4 variables mostraron tener un bajo grado de colinealidad con las otras variables, y solamente 3 variables por los demás métodos. En el estudio de los Eigenvalores y las proporciones de varianza solo 7 de las 16 variables independientes no mostraron problemas de multicolinealidad en el modelo completo; en el modelo obtenido por Stepwise y Forward, de las 10 variables que resultaron significativas para el modelo, solamente 4 estuvieron exentas de la colinealidad.

Anexo D. Transformaciones estabilizadoras de varianza

Muchas veces cuando se presenta el problema de que la varianza no es constante se puede recurrir a una transformación de variable y posteriormente replantear todo el modelo por todos los algoritmos para encontrar una nueva ecuación de regresión lineal múltiple.

Sin embargo, en la mayoría de las veces se desconoce cuál es la mejor transformación, por tal motivo se puede recurrir a seleccionarla de manera empírica.

*“A veces se puede recurrir a la experiencia o a consideraciones teóricas como guía para seleccionar una transformación adecuada. Sin embargo, en muchos casos no se tendrá un conocimiento a priori para creer que la varianza del error no sea constante. La primera indicación del problema se tiene al inspeccionar los diagramas de dispersión o con el análisis de residuales. En estos casos, la transformación adecuada se debe seleccionar **empíricamente**.”* [Montgomery, (2002)]

Algunas de las transformaciones a las que se puede recurrir se presentan en la siguiente tabla.

Tabla D1. Transformaciones útiles estabilizadoras de varianza [Montgomery, (2002)]

Relación entre σ^2 y $E(y)$	Transformación
$\sigma^2 \propto \text{constante}$	$y' = y$ (sin transformación)
$\sigma^2 \propto E(y)$	$y' = \sqrt{y}$ (raíz cuadrada, datos de Poisson)
$\sigma^2 \propto E(y)[1 - E(y)]$	$y' = \sin^{-1}(\sqrt{y})$ (arco seno; proporciones binomiales $0 \leq y_i \leq 1$)
$\sigma^2 \propto [E(y)]^2$	$y' = \ln(y)$ (logarítmica)
$\sigma^2 \propto [E(y)]^3$	$y' = y^{-1/2}$ (raíz cuadrada recíproca)
$\sigma^2 \propto [E(y)]^4$	$y' = y^{-1}$ (recíproca)

Bibliografía:

Banco de México (BANXICO). Diversos Índices. www.banxico.org.mx

Berk, K.N. (1978). "Comparing Subset Regression Procedures". Technometrics, 20. USA

Bolsa Mexicana de Valores (BMV). Diversos Índices. www.bmv.com.mx

Chou, Ya-Lun; Di Marco, Luis Eugenio. (1972). "Análisis estadístico". Nueva Editorial Interamericana. México.

Echavarría Hernán. (1992). "Inflación". Fondo Editorial Legis. 1992. Colombia

Instituto Nacional de Estadística y Geografía (INEGI). Diversos Indices. www.inegi.org.mx

Jacobson, Reed. (2000). "Programación con Microsoft Excel Versión 2000: Macros y Visual Basic para aplicaciones". McGraw-Hill. México

Kohler, Eric L. (1970). "Diccionario para contadores". Prentice-Hall, Inc. México

Maddala, G. S. (1996). "Introducción a la Econometría". Prentice-Hall Hispanoamericana, S.A. México.

Mansell Carstens Catherine. (1998) "Las Nuevas Finanzas en México". Editorial Milenio, S.A. de C.V. México.

Mantel, N. "Why Stepdown Procedures in Variable Selection", (1970) Technometrics, 12. USA

Montgomery, Douglas C.; Peck, Elizabeth A.; Vining, G. Geoffrey. (2002). "Introducción al análisis de regresión lineal". Compañía Editorial Continental (CECSA). México.

Neter, John; Wasserman, William; Kutner, Michael H. (1990). "Applied linear statistical models: Regression, analysis of variance, and experimental designs". Homewood, Illinois: R. D. Irwin. USA.

Secretaría de Planeación y Desarrollo Regional del Estado de Quintana Roo. Glosario.
seplader.seige.qroo.gob.mx

Bibliografía:

Banco de México (BANXICO). Diversos Índices. www.banxico.org.mx

Berk, K.N. (1978). "Comparing Subset Regression Procedures". *Technometrics*, 20. USA

Bolsa Mexicana de Valores (BMV). Diversos Índices. www.bmv.com.mx

Chou, Ya-Lun; Di Marco, Luis Eugenio. (1972). "Análisis estadístico". Nueva Editorial Interamericana. México.

Echavarría Hernán. (1992). "Inflación". Fondo Editorial Legis. 1992. Colombia

Instituto Nacional de Estadística y Geografía (INEGI). Diversos Indices. www.inegi.org.mx

Jacobson, Reed. (2000). "Programación con Microsoft Excel Versión 2000: Macros y Visual Basic para aplicaciones". McGraw-Hill. México

Kohler, Eric L. (1970). "Diccionario para contadores". Prentice-Hall, Inc. México

Maddala, G. S. (1996). "Introducción a la Econometría". Prentice-Hall Hispanoamericana, S.A. México.

Mansell Carstens Catherine. (1998) "Las Nuevas Finanzas en México". Editorial Milenio, S.A. de C.V. México.

Mantel, N. "Why Stepdown Procedures in Variable Selection", (1970) *Technometrics*, 12. USA

Montgomery, Douglas C.; Peck, Elizabeth A.; Vining, G. Geoffrey. (2002). "Introducción al análisis de regresión lineal". Compañía Editorial Continental (CECSA). México.

Neter, John; Wasserman, William; Kutner, Michael H. (1990). "Applied linear statistical models: Regression, analysis of variance, and experimental designs". Homewood, Illinois: R. D. Irwin. USA.

Secretaría de Planeación y Desarrollo Regional del Estado de Quintana Roo. Glosario.
seplader.seige.qroo.gob.mx