Sarriko-On

Técnicas de predicción económica

ISBN: 978-84-692-3815-8

María Pilar González Casimiro

05-09

Facultad de Ciencias Económicas y Empresariales
Ekonomia eta Enpresal-Zientzien Fakultatea

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

Técnicas de predicción económica

Pilar González Casimiro
Departamento de Economía Aplicada III (Econometría y Estadística)
Facultad de Ciencias Económicas y Empresariales
Universidad del País Vasco (UPV-EHU)

Contenido

1. La predicción económica **3**

1.1. Predicción económica y toma de decisiones 3

1.2. Técnicas de Predicción 9

1.3. Evaluación de las predicciones 15

1.4. Predicción con series temporales 20

2. Modelos de Componentes No Observados **25**

3. Análisis de una serie con tendencia **33**

3.1. Modelos globales: ajuste de funciones matemáticas 33

3.2. Métodos de alisado 42

3.2.1. Medias móviles 43

3.2.2. Métodos de Alisado Exponencial 45

3.3. Diferenciación 54

4. Análisis de una serie con estacionalidad **57**

4.1. Modelos globales. Variables ficticias estacionales. 57

4.2. Métodos de alisado 62

4.2.1. Método de Relación a la Media Móvil 62

4.2.2. Alisado exponencial de Holt-Winters con estacionalidad 65

4.3. Diferenciación 67

5. Modelos Estructurales de Series Temporales **69**

5.1. Especificación del modelo estructural 69
5.1.1. Principales Modelos Estructurales 71
5.2. Estimación máximo verosímil 75
5.2.1. Modelos en el Espacio de los Estados 75
5.2.2. El Filtro de Kalman 78
5.2.3. Estimación por Máxima Verosimilitud 87
5.3. Predicción 89
5.4. Extracción de señales 94

6. Predicción con modelos ARIMA **99**
6.1. Predicción con modelos estacionarios 100
6.1.1. Predicción con modelos MA(q). 102
6.1.2. Predicción con modelos AR(p) 104
6.1.3. Predicción con modelos ARMA(p,q). 106
6.1.4. Predicciones con modelos estacionarios estimados 108
6.2. Predicción con modelos no estacionarios. 111

7. Ejercicios **115**
7.1. Cuestiones 115
7.2. Problemas 118

A. Modelo de Regresión Lineal **129**

B. Medias Móviles **133**

C. Operador de retardos **137**

D. Mínimos Cuadrados Descontados **139**

Presentación

Tanto en economía de la empresa como en el campo macroeconómico, se plantea el problema de la toma de decisiones, es decir, la elección de una opción entre diversas alternativas. A la hora de tomar una decisión es muy importante contar con una visión de lo que va a suceder en el futuro: tomar una decisión exige considerar todas aquellas alteraciones que pueden producirse durante el horizonte temporal relevante para el tema en cuestión. En buena lógica no se debería tomar una decisión sin considerar la evolución futura de todos aquellos acontecimientos que la condicionan. En los últimos años se ha puesto un gran énfasis en mejorar el proceso de toma de decisiones y aquí es donde entra la idea de la predicción.

Predecir es una tarea compleja. Al predecir se trata de calcular algún hecho futuro, en general, como resultado de un análisis racional o de un estudio de los datos existentes. El objetivo de este documento es presentar las técnicas de predicción más utilizadas en el área de la economía y la empresa. De hecho, está pensado para su utilización como material docente en cursos de predicción en licenciaturas como Economía y Administración y Dirección de Empresas. Los métodos desarrollados son métodos cuantitativos de análisis de series temporales, en particular, de análisis de series temporales univariantes.

Los métodos de predicción de series temporales univariantes se basan en dos nociones: los componentes no observados (tendencias, estacionalidad, ciclos,...) y los modelos ARIMA que son modelos paramétricos que tratan de obtener predicciones de la serie temporal en términos de la interrelación temporal de sus elementos. Si bien este documento va a tratar ambos tipos de modelos de series temporales, el hincapié se va a poner en los modelos de componentes no observados. Así, trás introducir en el capítulo 1 la necesidad de la predicción en el área de la economía, así como describir las principales técnicas de predicción y sus limitaciones, el capítulo 2 pasa a definir y explicitar la noción de modelos de componentes no observados. Estos modelos se desarrollan con más detalle en el capítulo 3 que analiza las series con tendencia, el capítulo 4 que trata de las series con estacionalidad y el capítulo 5 que estudia los Modelos Estructurales de series temporales que son modelos basados en la idea de los componentes no observados pero especificados de forma estocástica. Por último, en el capítulo 6 se presenta de forma sucinta la teoría de la predicción con modelos ARIMA y el capítulo 7 ofrece al lector una colección de ejercicios como apoyo para trabajar los distintos temas planteados.

El enfoque del documento es tanto teórico como práctico. Cada capítulo incluye ejemplos empíricos de aplicación de las distintas técnicas de predicción que se proponen, planteando algunos ejercicios sencillos para que el lector pueda trabajar de forma autónoma.

Para poder predecir con series temporales es preciso disponer del software adecuado. En la actualidad, existen un buen número de paquetes estadístico-econométricos que permiten trabajar con series temporales. En los ejemplos que se presentan en este documento se han utilizado varios paquetes para el análisis de series temporales:

- EViews. Quantitative Microsoft Software. http://www.eviews.com/

  Este software se ha utilizado en los capítulos 2, 3 y 4.

- STAMP (Structural Time series Analyser, Modeller and Predictor). Las primeras versiones de este programa fueron escritas por Andrew Harvey y Simon Peters. http://www.stamp-software.com/

  Este software se ha utilizado en el capítulo 5.

- GRETL (Gnu Regression, Econometrics and Time-series Library)

  Este paquete es software libre y se puede descargar en español desde la página http://gretl.sourceforge.net/gretl_espanol.html

  Este software se ha utilizado en el capítulo 6.

Capítulo 1

La predicción económica

1.1. Predicción económica y toma de decisiones

En un sentido amplio se puede decir que el objeto de la ciencia económica es el estudio de la forma en que los agentes económicos toman sus decisiones y el análisis de las consecuencias que se derivan de la adopción de dichas decisiones. Tanto en economía de la empresa como en el campo macroeconómico, se plantea el problema de la toma de decisiones, es decir, la elección de una opción entre diversas alternativas. Cada opción dará lugar a un resultado distinto que puede ser medido en términos de utilidad, coste, beneficio, o cualquier otra magnitud, dependiendo del problema que se esté considerando. Ahora bien, el resultado concreto que se obtenga dependerá de situaciones que se puedan producir fuera del ámbito de influencia del decisor. A la hora de tomar una decisión es muy importante contar con una visión de lo que va a suceder en el futuro: tomar una decisión exige considerar todas aquellas alteraciones que pueden producirse durante el horizonte temporal relevante para el tema en cuestión. En buena lógica no se debería tomar una decisión sin considerar la evolución futura de todos aquellos acontecimientos que la condicionan. En los últimos años se ha puesto un gran énfasis en mejorar el proceso de toma de decisiones y aquí es donde entra la idea de la predicción.

Cuando se toman decisiones, el decisor se encuentra, en general, en ambiente de incertidumbre respecto a los sucesos que se pueden producir en el futuro. El problema con el que se enfrenta el decisor es elegir entre decisiones alternativas, teniendo en cuenta la utilidad de sus decisiones ante cada uno de los sucesos posibles. Estos sucesos son hechos, situados generalmente en el futuro, o que el decisor desconoce. En cualquier caso, el decisor podrá lograr unos mejores resultados si en alguna medida logra reducir la incertidumbre sobre los sucesos situados en el futuro, o que el decisor desconoce. Las técnicas de predicción van dirigidas, precisamente, a reducir la incertidumbre sobre el futuro y, por lo tanto, reducir el riesgo a la hora de tomar decisiones.

Al predecir se trata de calcular algún hecho futuro, en general, como resultado de un análisis racional o de un estudio de los datos existentes. Para que la predicción sea útil en el proceso de planificación es necesario abandonar el planteamiento simplista de una predicción única y permanentemente mantenida. Es conveniente realizar predicciones múltiples y someterlas a un proceso continuo de revisión y perfeccionamiento. Cuando se toma una acción determinada, la predicción ha de ajustarse de manera que refleje el impacto de dicha acción. Si la predicción no se revisa, puede inducir a error si se usa como base para tomar nuevas decisiones. Además, ya no sería posible evaluar la precisión de la predicción porque ya no refleja las circunstancias (o supuestos) que existían cuando se llevó a cabo.

El campo de aplicación de la predicción es muy amplio en el área de la economía.

A nivel de *economías domésticas* podemos considerar los siguientes ejemplos:

- predicciones sobre la evolución futura del salario y otras rentas para decidir qué parte de su tiempo dedican al ocio y qué parte a actividad laboral.
- predicciones sobre la rentabilidad que se espera recibir en el futuro para decidir entre consumo y ahorro deberán tener alguna idea.
- predicciones sobre el nivel de renta futura y la evolución de los precios para adquirir bienes duraderos.

En una *Empresa* es preciso utilizar predicciones para poder llevar a cabo la planificación del día a día y para poder controlar las operaciones a largo plazo de la compañía. El éxito de una empresa depende en parte de la habilidad de sus gestores de prever el futuro. Ejemplos de la necesidad de previsión en el ámbito de la empresa:

a) *Los Departamentos de Marketing* realizarán predicciones de los niveles de demanda para poder planificar sus estrategias de venta. Por ejemplo, les puede interesar conocer la demanda en diversas zonas de un mercado o entre distintos grupos de consumidores para poder llevar a cabo estrategias publicitarias efectivas. También les interesa saber cuales son las preferencias de los consumidores en las características del producto, color, etc. Cual va a ser la política de los competidores y cómo va a afectar a la participación de la empresa.

b) *El Departamento de Finanzas* querrá predecir los tipos de interés para poder planificar y financiar las próximas adquisiciones de capital. Cuales van a ser a largo plazo las necesidades de capital. La predicción de los ingresos y los gastos es también fundamental para poder mantener la liquidez de la compañía..

c) *El Departamento de Personal* necesitará conocer el número de trabajadores requeridos en cada área para llevar a cabo una política de contratación o reciclaje razonable.

d) Para *organizar la producción* puede ser necesario contar con predicciones de la demanda de cada línea de producto. Con estas predicciones la empresa puede planificar la producción y el mantenimiento de inventarios. Las predicciones de la demanda de cada producto se pueden traducir en predicciones de las materias primas necesarias y planificar así su compra (que también dependerá de predicciones sobre precios y recursos disponibles). Se necesitará conocer la evolución de los precios de esas materias primas. Cuales van a ser las necesidades de mano de obra, su especialización y la evolución salarial. Los cambios tecnológicos en su campo, etc.

e) Los *gerentes* precisarán de predicciones sobre las condiciones económicas generales, precios y costes, factores productivos, inventarios, progresos tecnológicos, cuota de mercado, etc. para diseñar el funcionamiento a largo plazo de la empresa.

Las empresas están afectadas por tres areas de actividad: la macroeconómica, la sectorial y su propia empresa. En general, una empresa solo puede controlar su propio funcionamiento, por lo que para poder predecir, necesita estar informado de lo que ocurre a nivel macroeconómico y sectorial, para lo que necesita también predicciones.

Las *Predicciones macroeconómicas* son necesarias tanto para el gobierno, como para la industria y los agentes financieros. Los gobiernos necesitan predicciones para tomar decisiones de política económica:

a) Para hacer el presupuesto el gobierno necesita predicciones de ingresos y gastos sociales (desempleo, pensiones, administraciones locales, ..). Algunos de los gastos dependen del gobierno pero para planificar otros necesita tener una idea general de como va la economía: paro, inflación, tipo de interés, crecimiento, ..

b) Cuando el gobierno decide cual va a ser su política de financiación del deficit, la hace tras prever cuál va a ser la evolución que van a seguir los tipos de interés así como el efecto que se puede derivar de la adopción de dicha política.

c) Para diseñar la política monetaria en un momento dado, se tendrán en cuenta los pronósticos hechos sobre la evolución de variables como la inflación o el tipo de cambio.

d) El sector internacional también es importante ya que las exportaciones y las importaciones afectan a la Renta Nacional directamente y además junto con la política monetaria y fiscal, determinan el valor del tipo de cambio.

e) Previsión de la infraestructura necesaria para el país. El gobierno ha de realizar los Planes Nacionales de Carreteras, Electricidad, Hidrólogico, etc. Todos estos planes precisan de grandes inversiones y llevan mucho tiempo, por lo que es fundamental tener conocimiento sobre las necesidades futuras del país: niveles futuros de tráfico, consumo de electricidad, etc.

f) A los gobiernos locales les puede interesar predecir, por ejemplo, la demanda de plazas de Educación Primaria y Secundaria en los próximos años, para poder redistribuir sus recursos educativos racionalmente

En los *Mercados Financieros* la actividad se centra en la interacción entre los valores presentes y futuros de la cartera de valores. Las interrelaciones entre los tipos de interés a corto y largo plazo, los tipo de cambio presentes y futuros y la política monetaria y fiscal afectan las perspectivas que tienen los agentes del mercado. Las predicciones macroeconómicas tratan de predecir todas esas variables y por lo tanto, pueden ser un factor importante a la hora de tomar decisiones.

Tipos de predicciones

Existen diferentes tipos de predicciones dependiendo de qué es lo que se predice y que requerirán enfoques y técnicas distintas.

1. *Predicción de los efectos de un suceso*

En este caso, se sabe que un suceso se va a producir en un futuro con certeza y se quiere determinar cuáles serán sus efectos. Por ejemplo, saber quién ganará las próximas elecciones, o qué efectos tendrá una ley que se promulgará próximamente o cuáles serán las ventas futuras de una nueva marca que sale al mercado, etc.

El problema que se plantea es que el suceso puede ser único, por lo que o bien no se dispondrá de información relevante o será muy difícil de adquirir. El mejor enfoque posible en estas situaciones es buscar o generar datos relevantes, es decir, es preciso aumentar el conjunto de información relevante de alguna forma. Por ejemplo, para predecir el resultado de unas elecciones, se puede realizar un sondeo entre los electores, o para predecir el éxito de la introducción de una nueva marca en el mercado se pueden realizar pruebas de mercado.

2. *Predicción del tiempo en que se produzca un suceso*

Esta clase de predicciones se cuestionan cuando y si es que, se va a producir un determinado suceso, es decir, cuándo serán las próximas elecciones o cuándo se producirá la recuperación de la economía o cuándo sacarán mis competidores un nuevo producto al mercado.

En algunos de estos ejemplos puede existir una secuencia de hechos similares en el pasado, por ejemplo, las fechas de las elecciones. En este caso, observando el patrón de los tiempos entre sucesos, se podría predecir cuando ocurrirá el próximo. Sin embargo, la manera habitual de trabajar es buscar *indicadores adelantados*, que son sucesos susceptibles de ocurrir antes del que estamos tratando de predecir. Este enfoque se utiliza mucho para predecir puntos de cambio en la evolución de la economía. Por ejemplo, antes de lanzar un producto al mercado se puede observar que una empresa ha reservado mucho tiempo de anuncios televisivos.

3. *Predicción de series temporales*

Una serie temporal es un conjunto de observaciones recogidas a intervalos regulares de tiempo. Por ejemplo, las temperaturas horarias, los precios diarios de acciones al cierre de la bolsa, la tasa de desempleo mensual o la Renta Nacional anual.

Supongamos que se observa una serie temporal denotada por Y_t , para $t = 0$ hasta $t = T$, donde T es el momento final de observación de la variable. La predicción de series temporales supone decir algo sobre el valor que tomará la serie en momentos futuros, $T+\ell$, donde ℓ representa el número de periodos hacia adelante que estamos considerando. Si $\ell = 1$, entonces se denomina predicción un periodo hacia adelante. Por ejemplo, supongamos que la serie Y_t representa las ventas mensuales de automóviles observadas durante 100 meses consecutivos, $t = 1, 2, \ldots, 100$, de Enero de 2000 hasta Abril de 2008 y que queremos predecir el valor tomado por esta serie en Mayo de 2008 ($\ell = 1$), y el valor en Diciembre de 2008 ($\ell = 8$). Como las causas para que los automóviles se vendan más o menos son muy complejas, la serie Y_t se puede considerar como una sucesión de variables aleatorias. Si estamos en el momento T y queremos predecir lo que ocurrirá en el momento $T + \ell$, no tenemos ninguna razón para pensar que vamos a acertar, a no ser por pura suerte.

Limitaciones de la predicción

La mayoría de las críticas que se hacen a las predicciones (desarrollos inesperados, hechos predichos que no suceden nunca, grandes errores de predicción, errores en el momento, intensidad de los cambios predichos, ...) están bien fundadas. Sin embargo, los agentes decisores que utilizan las predicciones tienen también cierta culpa de lo que pasa en el sentido de que sus expectativas suelen ser muy altas.

1.- ¿Qué se puede predecir y qué no?

No se deben tener expectativas poco realistas sobre la precisión y certeza con que se puede predecir el futuro. Un prerrequisito necesario para poder predecir, sea con el método que sea, es que exista un patrón de comportamiento en el fenómeno que estamos estudiando. Si no existe un patrón de comportamiento no es posible predecir aunque, a veces, se puedan dar opiniones subjetivas basadas en situaciones pasadas semejantes.

En economía la predecibilidad de un fenómeno varía desde ser casi nula (precio diario de una acción) hasta ser excelente (patrones estacionales basados fundamentalmente en razones climatológicas). El problema es que en economía los patrones y las relaciones se mezclan con componentes aleatorios y pueden cambiar en el tiempo de forma impredecible. Dos causas, entre otras, de estos cambios en los patrones o relaciones son:

- Aleatoriedad del comportamiento humano
- Capacidad de la gente de influenciar el futuro con sus propias acciones.

2.- Factores que influyen en la predecibilidad.

- *Cantidad de información.* A la hora de extraer el patrón de conducta seguida en el pasado y poder hacer conjeturas sobre el futuro un punto muy importante es la cantidad de información con la que contemos.

  Centrándonos en las series temporales, consideremos en primer lugar una serie temporal constituida por observaciones diarias sobre la hora de la salida del sol durante cincuenta años. El problema sería predecir a qué hora va a salir el sol mañana. Con una serie de datos de este tipo es muy fácil realizar esta predicción. Bien es cierto que, en lugar de utilizar las observaciones del pasado, se puede realizar también la predicción en base al conocimiento de las leyes sobre el movimiento de los astros. En cualquier caso, y es lo que interesa recalcar aquí, en un fenómeno de este tipo a partir de la observación del pasado se puede hacer una buena previsión de cual va a ser el futuro. ¿Por qué sucede esto? Sencillamente porque la serie contiene mucha información siendo los valores pasados de gran utilidad para predecir el futuro.

  Como segundo ejemplo, puede considerarse una serie temporal en la que se ha registrado el primer premio de la lotería de los sorteos que se han celebrado en los últimos cincuenta años. Con esta información disponible se presenta el problema de predecir cual va a ser el primer premio en el próximo sorteo de lotería. Aún siendo series del mismo tamaño (supongamos que los sorteos son diarios) y recogidas formalmente de la misma manera, está claro que la serie del primer premio de lotería es una serie que contiene poca información. En realidad, esta información no sirve absolutamente de nada para predecir, suponiendo que se cumplen todas las normas y principios que debe seguir un sorteo aleatorio.

  Estos ejemplo son dos casos extremos de series temporales. En el primer caso, la variable observada tiene un esquema o patrón de comportamiento fijo, mientras que en segundo caso no existe ningún patrón de comportamiento.

- *Horizonte de predicción.* Cuanto mayor es el horizonte de predicción mayor es la posibilidad de cambio en patrones o relaciones porque:

  - el comportamiento o actitudes de la gente puede cambiar.

  - hay más tiempo para utilizar las predicciones para modificar el futuro con el fin de alcanzar los beneficios deseados.

  - pueden producirse cambios fundamentales en el entorno, por ejemplo, cambios tecnológicos.

  Las predicciones a *corto plazo* son predicciones a un plazo menor de tres meses por lo que hay dos aspectos a tener en cuenta. Por una parte, los cambios en los patrones y relaciones económicos se pueden producir y de hecho se producen. Pero debido a la gran inercia que presentan la mayoría de los fenómenos económicos, cuando cambia alguna relación, el resultado de este cambio no es inmediato. Este concepto de inercia es muy importante en el campo económico. Por ejemplo, hizo falta casi un año desde la denominada "crisis del petróleo" para que las economías occidentales entraran en recesión. Debido a esta inercia y a los retardos en la respuesta, el estado corriente de muchas variables es un buen predictor de su valor en un futuro próximo. Es decir, se pueden extrapolar los patrones establecidos a corto plazo con cierto grado de precisión.

Las predicciones a *medio plazo* son las predicciones que cubren el periodo que va de los tres meses a los dos años y, en general, se derivan de las predicciones a largo plazo, o se construyen acumulando las predicciones a corto plazo. Estas predicciones no suelen ser muy precisas y, en general, suele ser difícil predecir los puntos de cambio en los ciclos económicos, ni las recesiones ni las épocas de expansión. Pero estas predicciones son también necesarias para tomar decisiones sobre presupuestos o asignación de recursos, por lo tanto, los planificadores deben aceptar sus limitaciones para predecir recesiones y booms en la economía y desarrollar planes flexibles que sean capaces de ajustarse a los cambios cíclicos.

Las predicciones a *largo plazo*. Son las predicciones que cubren un periodo de dos años en adelante. Las conclusiones que se recogen en la literatura sobre la precisión de estas predicciones son pesimistas: es difícil señalar de antemano, el tamaño del error de predicción, se pueden producir cambios imprevistos en la tendencia, discontinuidades, nuevos acontecimientos, etc.

¿Quien podía predecir a comienzos de los 70 que se iba a producir un embargo de petróleo, precios tan altos como 39 dólares el barril, escasez de materias primas, estanflación, colapso del mercado de valores y dos recesiones en menos de cinco años? Aún más incluso menos podía haber predicho el giro que iba a tomar la economía con una caída de los precios del petróleo hasta menos de 10 dólares en 1986, un mercado de valores efervescente , una caída en los precios de las materias, la inflación, los tipos de interés.

En resumen, las predicciones a largo plazo, tienden a ser imprecisas, pero son necesarias para las planificaciones estratégicas y presupuestarias. Por lo tanto, todos los problemas creados por la incertidumbre de estas predicciones deber ser estudiados y no ignorados.

1.2. Técnicas de Predicción

Con las técnicas de predicción se trata de hacer pronósticos lo más acertadamente posible sobre sucesos que todavía no han tenido lugar. Para hacer estas predicciones, una pitonisa utiliza como herramienta una bola de cristal, aunque implícitamente puede tener en cuenta la historia pasada. Por el contrario, las predicciones que realiza el económetra o el estadístico están basadas en un análisis explícito de la información proporcionada por los sucesos ocurridos en un pasado más o menos inmediato.

La predicción en general y la económica en particular, no tienen una base científica unitaria. En ella se conjugan enfoques metodológicos provenientes de la estadística o la economía, junto con desarrollos puntuales de la física, la ingeniería, la psicología e incluso la biología.

Desde un punto de vista metodológico, los métodos de predicción se pueden agrupar en dos grandes bloques: métodos cualitativos (o tecnológicos o subjetivos) y métodos cuantitativos. Los métodos cuantitativos se utilizan cuando contamos con datos históricos, mientras que las técnicas de predicción cualitativas se utilizan cuando los datos son escasos o no están disponibles. Los métodos cuantitativos se utilizan cuando se espera que el patrón de comportamiento de los datos persista en el tiempo, y las técnicas cualitativas se usan para predecir cuando el patrón existente en los datos pueda cambiar.

Métodos Cualitativos

Básicamente los métodos cualitativos se utilizan en aquellos casos en los que el pasado no proporciona una información directa sobre el fenómeno considerado, por lo que se les denomina también métodos sin historia. Por ejemplo, en el caso de la aparición de nuevos productos en el mercado no se puede recurrir a la información sobre el volumen de ventas en el pasado, ya que este producto no existía todavía en el mercado. En este caso, para poder predecir, la empresa habrá de basarse en la opinión de sus expertos de los departamentos de ventas y marketing. También los métodos cualitativos pueden ser aplicados a estudios sobre resultados de aplicación de nuevas técnicas o a investigaciones de tipo político o sociológico sobre posibles cambios en los patrones sociales históricos.

En la predicción de carácter cualitativo o tecnológico, los métodos estadísticos juegan un papel relativamente secundario. En estos casos lo más importante es contar con un grupo de expertos, dotados de intuición y sagacidad, que tengan un buen conocimiento tecnológico o especializado acerca del fenómeno cuya proyección en el futuro se trata de analizar. En este contexto, los métodos estadísticos serían utilizables en la organización y sistematización de las opiniones de los expertos.

En el método cualitativos denominado el *brainstorming* (literalmente tormenta de cerebros), la previsión se efectúa a partir de la discusión entre un grupo de expertos donde se crea un determinado ambiente para facilitar que afloren nuevas ideas. La aplicación de este método no requiere prácticamente la utilización de elementos estadísticos.

Otro método de predicción cualitativo muy conocido es el *Método Delphi* desarrollado por la Corporación RAND. Esta técnica se basa en reunir a un conjunto de expertos para que hagan predicciones sobre alguna cuestión específica como, por ejemplo, si se va a producir algún desarrollo novedoso en algún campo en particular. Este método supone que la combinación del conocimiento de todo el conjunto de expertos producirá predicciones al menos tan buenas como cada uno de los miembros. Reunir a todo el conjunto de expertos en torno a una mesa a discutir presenta problemas ya que, por una parte, siempre va a existir algún individuo o grupo de individuos que van a dominar la discusión, y, por otra, este tipo de decisiones conjuntas pueden estar influidas por diversas presiones sociales. Para evitar esto, el método Delphi mantiene a cada miembro del conjunto físicamente separado. Cada participante rellena una serie de cuestionarios que son revisados por un coordinador del equipo. Este elabora un documento de síntesis donde se destacan las coincidencias y las discrepancias, enviando a cada experto este documento. Posteriormente han de rellenar otros cuestionarios que ya van acompañados por información referente a las opiniones del grupo en su conjunto. Cada participante puede revisar sus opiniones teniendo en cuenta las opiniones del conjunto. Es de esperar que después de varias vueltas se llegue a algún tipo de consenso que se pueda utilizar como predicción. De todas formas, este método no busca el consenso, sino que permite que haya diferencias de opinión justificadas.

Métodos Cuantitativos

En las predicciones de carácter cuantitativo, se parte del supuesto de que se dispone de información sobre el pasado del fenómeno que se quiere estudiar. Generalmente la información sobre el pasado aparece en forma de series temporales.

Para predecir, el profesional debe analizar los datos del pasado y debe basar la predicción en los resultados de este análisis. La información se suele utilizar de la forma siguiente. En primer lugar, se analizan los datos para poder identificar el patrón que se puede utilizar para describirlos. Después se extrapola este patrón de comportamiento en el futuro para realizar la predicción. Esta estrategia que se utiliza en la mayoría de las técnicas cuantitativas de predicción descansa en el supuesto de que el patrón que hemos identificado en los datos pasados va a continuar en el futuro. Es de esperar que ninguna técnica de predicción va a dar buenos resultados si este supuesto no es válido.

Una técnica de predicción cuantitativa va a consistir en dos fases fundamentales:

a) Construcción del modelo: un modelo de predicción se diseña en función de los datos pasados y la teoría disponible. En algunos casos esta teoría nos puede sugerir modelos, en otros casos, tal teoría no existe o es incompleta y hemos de basar la especificación de nuestros modelos únicamente en los datos históricos.

b) Fase de predicción: el modelo obtenido en la etapa anterior se utiliza para predecir. Como estas predicciones van a depender del modelo especificado en primer lugar nos tenemos que asegurar de que el modelo y sus parámetros permanecen constantes durante el periodo de predicción.

En los métodos cuantitativos la misión del profesional consiste en extraer toda la información posible contenida en los datos y, en base al patrón de conducta seguida en el pasado, realizar conjeturas sobre el futuro.

Con los métodos de predicción cuantitativos se pretende conocer los patrones de comportamiento subyacentes de una serie con objeto de realizar previsiones del futuro. El instrumento de análisis que se suele utilizar es un **modelo** que permita reproducir el comportamiento de la variable de interés.

Los modelos cuantitativos se pueden agrupar, en principio, en dos grandes bloques: Análisis de Series Temporales y Análisis Causal.

1. Análisis de Series Temporales.

   En el análisis de series temporales se trata de hacer predicciones de valores futuros de una o varias variables, utilizando como información únicamente la contenida en los valores pasados de la serie temporal que mide la evolución de la variables objeto de estudio. Es decir, se trata de identificar los patrones históricos y después, bajo el supuesto de que se mantienen en el futuro predecir extrapolandolos.

   Ningún tipo de decisiones que adopte el decisor van a cambiar las predicciones generadas por un modelo de series temporales. Este tipo de modelos es útil para predecir cuando pensamos que las condiciones no van a cambiar. En este sentido, el modelo no será muy útil para predecir el impacto de ciertas decisiones que pueda tomar el agente decisor. Los métodos de predicción basados en el análisis de series temporales tratan al sistema como una caja negra y no intentan descubrir los factores que afectan su funcionamiento. A pesar de esta desventaja, existen importantes razones para utilizar los métodos de predicción basados en el análisis de series temporales:

   - Es posible que no conozcamos el comportamiento del sistema, o que aunque lo conozcamos, sea muy difícil o imposible medir las relaciones que supuestamente gobiernan su comportamiento.
   - Puede ser que nuestro interés sea sólo predecir que es lo que va a pasar y no el por qué pasa.
   - Puede interesarnos de igual manera conocer el comportamiento del sistema como el predecirlo, pero mientras estudiar los factores que determinan el funcionamiento del sistema puede ser muy costoso, el predecir su evolución por medio de un análisis de series temporales no.

   Los modelos de series temporales puros no contienen variables explicativas aparte de las que son únicamente funciones del tiempo. Las predicciones de las futuras observaciones se llevan a cabo extrapolando la información pasada. En este sentido se suele hablar de predicciones naive. Sin embargo, son en muchas ocasiones muy efectivas y nos proporcionan un punto de referencia con el que comparar el funcionamiento de otros modelos más sofisticados.

2. Análisis Causal.

   Se denomina así porque en la explicación de la variable o variables objeto de estudio intervienen factores explicativos externos. Quizás el término de causalidad, por las connotaciones de todo tipo que pueda tener, sea demasiado fuerte para describir la forma en que inciden los factores externos en la explicación de una variable. Sin embargo, este término ha adquirido su carta de naturaleza especialmente en el análisis estadístico de variables relacionadas temporalmente. Básicamente los métodos causales suponen que el valor de cierta variable es función de una o más variables.

   Estos métodos causales o explicativos tratan de identificar las relaciones que produjeron (o causaron) los resultados observados en el pasado. Después se puede predecir aplicando estas relaciones al futuro, supuestos conocidos los inputs del sistema en momentos futuros.

   La fuerza o ventaja real de un método causal de predicción es que se puedan desarrollar una gama de predicciones correspondientes a una gama de valores para las diferentes variables explicativas. Estos modelos causales permiten evaluar el impacto de políticas o decisiones alternativas. Sin embargo, una desventaja de estos métodos es que requieren información sobre varias variables. Además como incluyen varios factores son más sensibles a los cambios en las relaciones subyacentes que un modelo de series temporales. Por otra parte, requieren la estimación de los valores futuros de los factores antes de predecir.

No tiene sentido, sin embargo, decir que los métodos econométricos tradicionales y los métodos de series temporales son contrapuestos, ya que son complementarios.

Simplificando un poco, la diferencia entre el enfoque económico tradicional y el enfoque de series temporales, acerca de la elaboración de modelos causales sería el siguiente: en el primero, el énfasis se sitúa en el modelo de partida, efectuándose contrastes sobre la adecuación entre modelo y datos; por el contrario, en el enfoque de series temporales el punto de partida son los datos y quizás alguna idea general sobre el fenómeno que se trata de modelizar, siendo el modelo el resultado final de la investigación.

El uso de variables explicativas en un modelo requiere más esfuerzo y más información que predecir a partir de un modelo de series temporales puro. La desventaja de un modelo de series temporales puro es que puede que no sea estable a lo largo del tiempo y, como consecuencia, las predicciones pueden ser altamente imprecisas. La introducción de variables explicativas nos puede conducir a un modelo en el que los parámetros permanezcan constantes en el tiempo. Sin embargo, no hay garantía de que tal modelo exista. Lo que es más un modelo con variables explicativas mal especificado nos puede llevar a peores predicciones que un modelo de series temporales naive.

Ahora bien, en una investigación concreta puede surgir la siguiente cuestión: ¿qué línea es conveniente seguir, un modelo de series temporales o un modelo causal? Cuando se dispone de un modelo teórico apropiado, de información estadística suficiente y de una previsión exacta de los factores externos o variables explicativas es preferible utilizar un modelo causal como instrumento de predicción. No obstante, en muchas ocasiones, no se dan simultáneamente esas circunstancias siendo necesario recurrir a un análisis de carácter univariante. En cualquier caso siempre es útil como una primera aproximación a un fenómeno, efectuar un análisis univariante de las variables involucradas en el modelo econométrico.

Al elegir una técnica de predicción, el profesional ha de tener en cuenta los siguientes factores:

a) El tipo de predicción que deseamos. En ocasiones la predicción por punto puede ser suficiente, pero puede ser que nos interese una predicción por intervalo. Esto condiciona la elección del método de predicción, ya que algunos de ellos no nos proporcionan predicciones por intervalo.

b) El horizonte temporal. El horizonte temporal lo solemos catalogar en corto plazo, medio plazo y largo plazo. Normalmente conforme el horizonte temporal es más largo, la precisión de las predicciones es menor y en este caso las técnicas cualitativas se vuelven más útiles conforme el horizonte temporal crece.

c) El patrón de los datos. Si estamos utilizando datos históricos para predecir es fundamental tener en cuenta que tipo de estructura presentan los datos (si tienen tendencias, estacionalidades, etc.) y así adecuar el método elegido a los datos.

d) El coste de la predicción. A la hora de elegir una técnica de predicción hemos de tener en cuenta los siguientes costes. En primer lugar, el coste de desarrollar el método de predicción. Este coste cambiará de técnica a técnica dependiendo de su complejidad. En segundo lugar, el coste de almacenar los datos necesarios para predecir. Mientras que algunos métodos de predicción requieren guardar muy pocos datos, en otros hay que acumular gran cantidad de información. Por último, el coste de utilizar la técnica de predicción. En algunos casos el método de predicción es operativamente muy simple, pero en otros puede llegar a ser muy complejo.

e) La precisión deseada.

f) La disponibilidad de los datos. Los métodos cuantitativos solo se podrán utilizar si contamos con los datos necesarios. Los distintos métodos necesitan distinta cantidad de datos y, en muchas ocasiones, el que podamos utilizar unas técnicas u otras va a depender de los datos disponibles. Sin o disponemos de los datos históricos que necesitamos habremos de llevar a cabo métodos especiales de recogida de datos que pueden suponer un coste.

g) La facilidad operacional y de comprensión. La facilidad de utilización y compresión de una técnica de predicción es fundamental. Los agentes económicos que han de tomar las decisiones y que muchas veces han de basarlas en predicciones, deben ser capaces de entender estas técnicas, para poder tener confianza en ellas.

La elección del método de predicción a usar en una situación determinada conlleva encontrar una técnica que responda satisfactoriamente a las cuestiones planteadas. El "mejor" método de predicción no es siempre el más "preciso". El método de predicción que debería ser utilizado es aquel que cubre nuestras necesidades con el menor coste e incoveniencia.

El profesional debe tratar de construir modelos simples, fáciles de entender y, por lo tanto, de explicar. Un modelo elaborado puede generar predicciones más precisas pero puede ser muy costoso y difícil de implementar. El principio de parsimonia nos dice, que al elegir entre modelos competidores, si todo lo demás es igual, hemos de elegir el más simple.

A veces solo necesitamos predicciones muy burdas, en otros casos, la precisión es esencial. En algunas aplicaciones la precisión puede ser muy costosa, por ejemplo, una predicción imprecisa de un indicador económico puede llevar al Banco Central Europeo a subir los tipos de interés erróneamente con todas las consecuencias que esto tendría. Por otro lado, aumentar la precisión suele incrementar mucho los costes tanto de adquisición de datos, como de personal o de uso de ordenador. Si una pérdida pequeña de precisión no es muy importante, y baja el coste sustancialmente, puede que prefiramos el modelo simple menos preciso, que el complejo.

1.3. Evaluación de las predicciones

Cuando se hace una predicción siempre se va a cometer un error de predicción, que se define como la diferencia entre el verdadero valor y la predicción:

$$e_T(\ell) = Y_{T+\ell} - Y_T(\ell)$$

Las fuentes o causas de los errores de predicción son numerosas. En la ciencia económica abundan los errores de medida en las variables, no es posible la experimentación, etc. Debido a la complejidad de las situaciones económicas, la inconsistencia del comportamiento humano, los retardos entre acciones y resultados, etc. los errores de predicción son mucho mayores que en otras ciencias. Su tamaño y persistencia dependen de:

a) Identificación errónea de los patrones y relaciones:

- Podemos identificar un patrón ilusorio, no existente
- Un modelo estadístico basado en pocos datos puede identificar un patrón que no se mantiene durante mucho tiempo. Hay que tener en cuenta que la serie de datos que nosotros observamos no es más que una parte de un proceso más largo que comenzó en algún momento del pasado y se extiende en el futuro. Si nuestra serie es muy corta podemos estar enfocando una parte relativamente poco relevante de los patrones subyacentes del proceso.

- La relación entre dos variables puede ser espúrea, debido a un tercer factor común a ambas.
- Podemos identificar patrones de forma incorrecta por falta de información o porque la realidad es muy compleja como para ser modelada con un número limitado de variables.

La identificación ilusoria o inapropiada puede causar errores de predicción muy serios y no aleatorios porque el futuro puede resultar ser muy diferente de la relación o patrón erróneo que se postuló.

b) Patrones inexactos o relaciones imprecisas.

En las ciencias sociales por sus propias características, los patronos son inexactos e imprecisos. Aunque se pueda identificar un patrón o relación en promedio, las fluctuaciones en torno a esa media existen en la mayoría de los casos. El propósito de la modelización estadística o econométrica es identificar los patronos y las relaciones de manera que en el pasado las fluctuaciones respecto a la media sean lo más pequeñas y lo más aleatorias posible. Si es o no es una buena estrategia, se puede discutir, pero aunque sea apropiada no garantiza que los errores futuros sean aleatorios o simétricos o que no excedan cierta magnitud.

c) Cambios en los patrones o relaciones.

Para entender las ventajas y limitaciones de la predicción es muy importante reconocer que todos los tipos y formas técnicas de predicción son extrapolativas por naturaleza. A la hora de hablar de precisión de las predicciones y de qué podemos esperar de ellas, hay que tener en cuenta muchos factores:

- Si estamos haciendo análisis de series temporales, estamos prediciendo únicamente extrapolando los valores pasados de la serie
- Si seguimos un enfoque causal, hemos de saber que las relaciones entre los diferentes factores no permanecen constantes y nuestra habilidad para entender fenómenos complejos es limitada. Además la existencia de relación no implica causalidad. Pero, por otra parte, si identificamos los factores que afectan los cambios y la dirección de la causalidad, nos puede ser muy útil a la hora de planificar y tomar decisiones.

Tenemos que reconocer que las técnicas cuantitativas de predicción no tienen una manera simple de predecir con cierta confianza cuando los patrones o las relaciones establecidas cambian. Debido a que los métodos cuantitativos, tanto causales como de series temporales, basan sus predicciones o extrapolaciones en patrones de comportamiento o interrelaciones pasadas, solo funciona bien si el futuro es similar al pasado. Sin embargo, en la mayoría de los ambientes tanto económicos como empresariales, tales patrones y relaciones cambian continuamente, y además la tasa de cambio varía. Cuanto mayor sea la tasa de cambio menos precisa será la predicción.

Esta fluidez de las relaciones se acentúa además por la acción y por los objetivos de los agentes decisores que influyen en el curso de los futuros acontecimientos. Así, en muchas ocasiones, los objetivos fijados cambian el entorno de la toma de decisiones y así afectan la precisión de los métodos de predicción.

Cuando las técnicas cuantitativas no funcionan bien, la única forma de predecir el impacto de los cambios es la razón humana con cierta ayuda. Este juicio humano también depende de las tendencias y patrones de comportamiento observados en el pasado, pero presenta la ventaja de que pueden identificar los cambios sistemáticos más rápidamente e interpretar mejor los efectos de tales cambios en el futuro. En contra tiene que el juicio humano puede estar influido por los intereses en juego y por nuestro deseo de que se produzca un determinado resultado.

En las ciencias sociales, los patrones o relaciones están cambiando constantemente en el tiempo de forma no predecible en la mayoría de los casos. Estos cambios de las relaciones pueden causar errores persistentes cuya magnitud no podemos conocer de antemano. La precisión de la predicción está muy influida por los cambios en patrones o relaciones que se pueden clasificar según su carácter, duración o si son aleatorios o sistemáticos.

"Aleatorio" indica que no se puede predecir, mientras que sistemático sugiere que se puede, aunque el conocimiento o el entendimiento necesario para hacer tales predicciones no esté todavía disponible.

Para poder elegir entre varios métodos de predicción convendría contar con un criterio que nos dijera cual es el mejor. Uno de los criterios que más se tienen en cuenta a la hora de evaluar la bondad de las distintas técnicas de predicción es la precisión.

Cuando se predice siempre se comete un error de predicción lo que supone un coste porque el decisor no va a tomar la decisión óptima. En general cuanto mayor sea el error de predicción mayor será el coste. Lo ideal sería que cada decisor conociera su función de coste (qué le cuesta por unidad de error de predicción negativa, y por unidad de error de predicción positiva), y entonces elegir la técnica de predicción que de acuerdo con esta función de costes tiene un coste menor. Pero como el decisor en general, no va a proporcionar la función de costes lo que se suele hacer es elegir, por conveniencia, una función de costes que sea una buena aproximación a la verdadera función de costes.

Dada una función de coste, $C(e)$, el mejor método de predicción dentro de un conjunto de alternativas es aquel que nos proporcione el mejor coste (promedio) esperado.

Supongamos que $f(1)$ y $f(2)$ son dos métodos de predicción con errores de predicción dados por $e_T^1(\ell)$ y $e_T^2(\ell)$, $\ell = 1, \ldots, L$, respectivamente. Si utilizamos una función de coste general $C(e)$, el criterio se expresa de la siguiente manera:

$$\frac{1}{L} \sum_{\ell=1}^{L} C(e_T^1(\ell)) < \frac{1}{L} \sum_{\ell=1}^{L} C(e_T^2(\ell))$$

Si ambos lados de esta desigualdad se multiplican por una cantidad constante positiva, la desigualdad se mantiene (se podría prescindir del término $\frac{1}{L}$).

Supongamos que podemos considerar todos los posibles métodos de predicción que utilizan un conjunto de información particular I_T. El método que nos proporcione el coste promedio mínimo se denomina la predicción óptima basada en el conjunto de información I_T (y dada una función de coste). Es muy difícil, en general, considerar todas las posibles predicciones basadas en I_T, aunque este conjunto de información contenga solo datos numéricos. Los predictores que consideremos usualmente son los que tratan los datos de I_T de forma lineal. Por ejemplo, si solo utilizamos una serie temporal, $I_T : Y_{T-j} \geq 0$, entonces la predicción de $Y_{T+\ell}$ se puede expresar como:

$$Y_T(\ell) = \sum_{j=0}^{m} \gamma_j \, Y_{T-j}$$

y se denomina predicción lineal. Si elegimos m y γ_j de forma que la predicción es la mejor de acuerdo con nuestro criterio, entonces tenemos la predicción óptima, linear, univariante.

Para decidir qué método de predicción es el mejor es necesario tener un punto de referencia con el comparar la precisión de la predicción. Precisamente una de las dificultades al manejar el criterio de precisión es la ausencia de una medida de precisión que sea aceptada universalmente. Las medidas más utilizadas en la práctica son las siguientes:

a) La desviación media absoluta o error medio absoluto (EAM): es una medida de la precisión general que da una idea del grado de dispersión y en la que todos los errores reciben el mismo peso:

$$EAM = \frac{1}{L} \sum_{\ell=1}^{L} |e_T(\ell)|$$

donde $|e_T(\ell)|$ es el error de predicción en valor absoluto, y L es el número de predicciones.

b) El error cuadrático medio (ECM) es también una medida de precisión general que proporciona una indicación del grado de dispersión, pero donde se da más peso a los errores más grandes. Es la medida de precisión más comúnmente utilizada:

$$ECM = \frac{1}{L} \sum_{\ell=1}^{L} (e_T(\ell))^2$$

Un problema con esta función de coste es que es simétrica y las funciones de coste reales suelen ser asimétricas. Sería posible que el decisor tuviera en cuenta el hecho de que se elige un función de coste incorrecta, pero para hacer esto debería contar con una estimación de la verdadera función de coste y esto no se da en la práctica. Elegir como función de coste el error cuadrático se justifica sólo en términos prácticos, pero no está claro como se pueden reducir los costes extras que esta elección produce al decisor. La ventaja de esta función de coste es que corresponde a los criterios de mínimos cuadrados con lo que nos permite utilizar teoría estadística estándar muy conocida.

Ejemplo 1.1. Una comparación de las implicaciones de elegir el EAM o el ECM como medida de precisión se puede ver en la siguiente tabla:

		Met. Pred. A			Met. Pred. B		
ℓ	Obser.	Pred.	Error	$Error^2$	Pred.	Error	$Error^2$
1	25	23	2	4	22	3	9
2	28	29	-1	1	31	-3	9
3	23	24	-1	1	20	3	9
4	30	20	10	100	27	3	9
5	27	26	1	1	31	-4	16

Dada la tabla es fácil calcular las medidas de precisión de la predicción:

$$EAM_A = 3{,}0 \qquad ECM_A = 21{,}4 \qquad EAM_B = 3{,}2 \qquad ECM_B = 10{,}4$$

Examinando los resultados se puede observar que el método de predicción A produce predicciones más próximas al valor real que el método B en cuatro de los cinco casos, la mayoría de las veces es más preciso. Sin embargo, en el periodo 4 el error de predicción es grande. El método B, por otro lado, aunque no es tan preciso, salvo en el periodo 4, no tiene errores grandes. Estos hechos se reflejan en las diferentes medidas. El método A tiene un EAM de 3 que es menor que el del método B, mientras que el ECM del método A es casi dos veces mayor que el del método B.

Muchas veces se utiliza como medida la raíz cuadrada del ECM, el RECM, porque en este caso, la gravedad del error de predicción se recoge en las mismas dimensiones en que están el valor real y el predicho.

c) Error Medio Absoluto Porcentual es la medida relativa que corresponde con el EAM y viene dada por:

$$EMAP = \frac{1}{L}\sum_{\ell=1}^{L} \frac{|e_T(\ell)|}{Y_{T+\ell}} \times 100$$

Esta medida no recibe demasiada atención en los libros de texto más habituales.

d) Raíz del Error Cuadrático Medio Porcentual es la medida relativa que se corresponde con el ECM:

$$ECMP = \sqrt{\frac{1}{L}\sum_{\ell=1}^{L}\left[\frac{e_T(\ell)}{Y_{T+\ell}}\right]^2} \times 100$$

Esta medida toma el error como un porcentaje del valor actual.

e) U de Theil

$$U = \sqrt{\frac{\frac{1}{L}\sum_{\ell=1}^{L} e_T^2(\ell)}{\frac{1}{L}\sum_{\ell=1}^{L} Y_{T+\ell}^2}}$$

En ocasiones se evalúa esta medida para las fluctuaciones en Y_t:

$$U_\Delta = \sqrt{\frac{\frac{1}{L}\sum_{\ell=1}^{L}(\Delta Y_T(\ell) - \Delta Y_{T+\ell})^2}{\frac{1}{L}\sum_{\ell=1}^{L} \Delta Y_{T+\ell}^2}}$$

donde ΔY_T puede denotar bien la variación absoluta en Y_t o su tasa de variación porcentual. El estadístico U_Δ evalúa la bondad del modelo predictivo para anticipar puntos de giro en Y_t.

1.4. Predicción con series temporales

Cuando buscamos datos para estudiar el comportamiento de una variable económica y su relación con otras a lo largo del tiempo, estos datos se presentan frecuentemente en forma de series temporales.

Una *serie temporal* es una secuencia ordenada de observaciones cada una de las cuales está asociada a un momento de tiempo.

Ejemplos de series temporales las podemos encontrar en cualquier campo de la ciencia. En Economía podemos pensar en series como los precios diarios de las acciones, las exportaciones mensuales, el consumo mensual, los beneficios trimestrales, etc. En Metereología, tenemos series temporales de temperatura, cantidad de lluvia caída en una región, velocidad del viento, etc. En Marketing son de gran interés las series de ventas mensuales o semanales. En Demografía se estudian las series de Población Total, tasas de natalidad, etc. En Medicina, los electrocardiogramas o electroencefalogramas. En Astronomía, la actividad solar, o en Sociología, datos como el número de crímenes, etc.

El conjunto de técnicas de estudio de series de observaciones dependientes ordenadas en el tiempo se denomina Análisis de Series Temporales. El instrumento de análisis que se suele utilizar es un *modelo* que permita reproducir el comportamiento de la variable de interés.

Los Modelos de Series Temporales pueden ser:

- Univariantes: sólo se analiza una serie temporal en función de su propio pasado
- Multivariantes: se analizan varias series temporales a la vez. Un ejemplo muy popular en la literatura son las series de número de pieles de visón y rata almizclera capturadas en Cánada. Se sabe que existe una relación víctima-depredador entre ambos animales lo que se supone que afecta a la dinámica de ambas series. La forma de reflejar estas interacciones dinámicas entre ambas series es construir un modelo multivariante. Cuando se construye un modelo multivariante, para casos como éste, suponemos que hay cierta dependencia o relación entre los pasados de las diversas series.

  Estas interacciones dinámicas también aparecen cuando construimos modelos multivariantes para variables económicas, tales como la renta, consumo e inversión que, cómo es bien sabido, influyen las unas en las otras.

En este texto se van a considerar únicamente los modelos de series temporales univariantes, por lo que se va a centrar en el análisis de las series temporales univariantes.

Una serie temporal univariante consiste en un conjunto de observaciones de una variable Y. Si hay T observaciones, se denota por

$$Y_t, \quad t \in \Im \qquad\qquad Y_t, \quad t = 1, \ldots, T$$

El subíndice t indica el tiempo en que se observa el dato Y_t. Los datos u observaciones se suelen recoger a intervalos iguales de tiempo, es decir, equidistantes los unos de los otros; es el caso de series mensuales, trimestrales, etc.

Cuando las observaciones se recogen solo en momentos determinados de tiempo, generalmente a intervalos iguales, nos referimos a una *serie temporal discreta*. Puede darse el caso de que los datos se generen de forma continua y se observen de forma continua, como, por ejemplo, la temperatura, que se observa de forma continua en el tiempo por medio de aparatos físicos. En este caso denotamos la serie temporal por

$$Y_t, \qquad t \in R$$

y contamos con un número infinito de observaciones. En este caso nos referiremos a una *serie temporal continua*. Sin embargo, la mayoría de las series disponibles, en particular, en las ciencias sociales, se observan en tiempo discreto a intervalos iguales, aunque se puedan suponer generadas por algún proceso en tiempo continuo. Por lo tanto, nos vamos a centrar en el estudio de variables discretas tomadas a intervalos regulares dejando de lado las variables continuas y las variables discretas tomadas a intervalos irregulares.

Cada uno de los datos, Y_t, puede representar o una acumulación sobre un intervalo de tiempo de alguna cantidad subyacente, por ejemplo, lluvia diaria, o bien un valor tomado en un momento dado. Las *variables flujo* son aquellas que se miden respecto a un intervalo de tiempo, por ejemplo, el consumo mensual de petróleo. Las *variables stock* son aquellas que se miden en un momento determinado de tiempo, por ejemplo, la temperatura, las cotizaciones de bolsa, etc. Los métodos de análisis de series temporales son, en general, los mismos para ambos tipos de variables, pero puede haber ocasiones en que la distinción sea de interés debido a las distintas características que tienen las observaciones.

La mayoría de los métodos estadísticos elementales suponen que las observaciones individuales que forman un conjunto de datos son realizaciones de variables aleatorias mutuamente independientes. En general, este supuesto de independencia mutua se justifica por la atención prestada a diversos aspectos del experimento, incluyendo la extracción aleatoria de la muestra de una población más grande, la asignación aleatoria del tratamiento a cada unidad experimental, etc. Además en este tipo de datos (tomamos una muestra aleatoria simple de una población más grande) el orden de las observaciones no tiene mayor importancia. En el caso del análisis de series temporales, hemos de tener en cuenta, sin embargo, que:

- el orden es fundamental: tenemos un conjunto de datos ordenado
- el supuesto de independencia no se sostiene

Debido a estas características especificas de los datos de series temporales no se pueden utilizar para analizarlas algunas de las técnicas ya estudiadas en las asignaturas de Estadística, sino que se han de desarrollar modelos específicos que recojan y aprovechen la dependencia entre las observaciones ordenadas de una serie temporal.

Estudiando algunos gráficos de series económicas y sociales podemos notar que las observaciones no son independientes, sino que su evolución parece seguir un cierto patrón de comportamiento. Además, la naturaleza y estructura de esta dependencia es de gran interés en sí misma. Por otra parte, si se analizan con cuidado algunos de estos gráficos se puede observar que, tendencias y estacionalidades son rasgos relevantes de estos datos. Parece deseable que un analista de series temporales sea capaz de identificar estos rasgos característicos de las series, analizarlos y estudiarlos.

El análisis univariante resulta poco costoso en términos de información: sólo una variable debe ser observada. Por la misma razón, el objetivo de dicho análisis no puede ser muy ambicioso. En general, se pueden abordar *dos objetivos básicos*:

a) *Describir* las características de la serie, en términos de sus componentes de interés. Por ejemplo, podemos desear examinar la tendencia para ver cuales han sido los principales movimientos de la serie. Por otro lado, también el comportamiento estacional es de interés, y para algunos propósitos, nos puede interesar extraerlo de la serie: desestacionalizar.

   Esta descripción puede consistir en algunos estadísticos resumen (media, varianza, etc.) pero es más probable que incluya una o más representaciones gráficas de los datos. La complejidad de una serie temporal (opuesta a una m.a.s.) es tal que a menudo requiere una función, más que un simple número, para señalar las características fundamentales de las series; por ejemplo, una función μ_t en vez de un número μ para recoger el valor medio de la serie.

b) *Predecir* futuros valores de las variables. Un modelo de series temporales univariante se formula únicamente en términos de los valores pasados de Y_t, y/o de su posición con respecto al tiempo (ningún modelo univariante puede ser tomado seriamente como un mecanismo que describe la manera en que la serie es generada: no es un proceso generador de datos). Las predicciones obtenidas a partir de un modelo univariante no son por lo tanto más que extrapolaciones de los datos observados hasta el momento T. En este sentido se dice que son naïve; sin embargo, son en muchas ocasiones muy efectivas y nos proporcionan un punto de referencia con el que comprar el funcionamiento de otros modelos más sofisticados.

   Cuando las observaciones sucesivas son dependientes, los valores futuros pueden ser predichos a partir de las observaciones pasadas. Si una serie temporal se puede predecir exactamente, entonces se diría que es una serie determinista. Pero la mayoría de las series son estocásticas en que el futuro sólo se puede determinar parcialmente por sus valores pasados, por lo que las predicciones exactas son imposibles y deben ser reemplazadas por la idea de que los valores futuros tienen una distribución de probabilidad que está condicionada al conocimiento de los valores pasados.

Ambos objetivos pueden conseguirse a muy distintos niveles. Es evidente que, dada la muestra, calcular la media y la desviación típica de las observaciones supone describir características de la serie. De la misma manera podemos predecir que los valores futuros de la variable van a ser iguales al último valor observado. Sin embargo, en ninguno de los dos casos se usa la información muestral de una manera sistemática. El uso sistemático de la información muestral pasa normalmente por la formulación de modelos que pueden describir la evolución de la serie.

Los modelos utilizados para describir el comportamiento de las variables económicas de interés, siempre responden a la misma estructura:

$$Y_t = PS_t + a_t$$

donde: PS_t = Parte sistemática o comportamiento regular de la variable y a_t es la parte aleatoria, también denominada *innovación*.

En los modelos de series temporales univariantes la PS_t se determina únicamente en función de la información disponible en el pasado de la serie:

$$PS_t = f(Y_t, Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots)$$

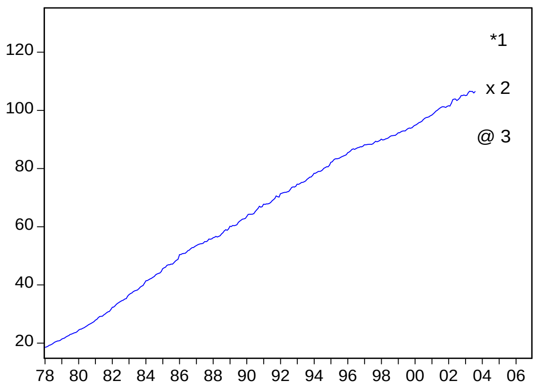
El análisis de series temporales se basa en dos nociones fundamentales:

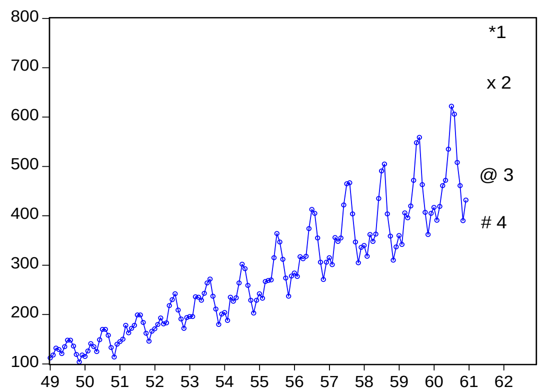
Componentes no observados. Se basa en la idea de que una serie temporal puede ser considerada como la superposición de varios componentes elementales no observables: tendencia, estacionalidad y ciclo. La estimación de tendencias y el ajuste estacional atraen mucho la atención debido a su importancia práctica para el análisis de series económicas.

Modelos ARIMA. Son modelos paramétricos que tratan de obtener la representación de la serie en términos de la interrelación temporal de sus elementos. El instrumento fundamental a la hora de analizar las propiedades de una serie temporal en términos de la interrelación temporal de sus observaciones es el denominado *coeficiente de autocorrelación* que mide la correlación, es decir, el grado de asociación lineal que existe entre observaciones separadas k periodos.

Estos coeficientes de autocorrelación proporcionan mucha información sobre como están relacionadas entre sí las distintas observaciones de una serie temporal, lo que ayudará a construir el modelo apropiado para los datos. Por otro lado, proporcionan también información para predecir. Supongamos que se calcula el coeficiente de autocorrelación de orden 1, r_1, que refleja la estructura de correlación entre observaciones separadas un periodo, es decir, observaciones consecutivas. Si obtenemos un valor de r_1 muy grande próximo a $+1$, y se ha observado un valor de Y_t por encima de la media, esperamos que el próximo valor esté también por encima de la media. De forma análoga, si r_1 es muy negativo, próximo a -1, esperamos que el próximo valor esté por debajo de la media (si el valor de Y_t está por encima). Por lo tanto, intuitivamente, parece claro que los coeficientes de autocorrelación tienen un papel muy importante para la predicción de series temporales.

Ejercicio 1.1. ¿Qué predicción propondrías en los siguientes gráficos? Discute en cada caso por qué rechazas o aceptas cada uno de los puntos posibles.





Capítulo 2

Modelos de Componentes No Observados

Los modelos de series temporales de Componentes No Observados (CNO) tratan de recoger las regularidades en el comportamiento de una variable a lo largo del tiempo. La idea de los componentes no observados es básicamente una manera de mirar a los datos: se observa un determinado fenómeno, se toma nota de sus regularidades o patrones y se formulan modelos basados en estas regularidades observadas. Se basan en la idea de que una serie temporal se puede descomponer en diferentes elementos que no se observan pero que se cree que forman parte de la serie y que pueden explicar su evolución. Estos modelos de CNO no pretenden representar el proceso generador de los datos sino describir las principales características de las series en función de elementos que son de interés en sí mismos.

La noción de que una serie temporal está formada por la superposición de diversos componentes no observados fue ya muy útil para las mediciones realizadas por los astrónomos en el siglo XVII. Dentro del campo de la economía esta idea se hizo común a mediados del siglo XIX. El uso de variables no observables en economía ha sido aceptado ampliamente como un enfoque muy fructífero para describir los fenómenos económicos. Los primeros modelos tratan a la estacionalidad o los errores de media como componentes no observados que deben ser extraídos. Otros modelos consideran el ciclo económico como una variable no observable que determina indirectamente el comportamiento de la serie observada. Una aplicación muy conocida también es la de la renta permanente, que aunque no se puede medir, explica las regularidades observadas en los datos. Algunos modelos macroeconómicos recientes tratan con variables como expectativas, la tasa de interés real o la tasa natural de desempleo que no se pueden observar pero que presumiblemente ayudan a explicar nuestros datos.

El punto de partida para la construcción de estos modelos es, por lo tanto, la existencia de dichos componentes no observados o regularidades. En el campo del análisis de series temporales es práctica común clasificar los tipos de movimientos que caracterizan una serie temporal como tendencia, estacionalidad, ciclo e irregular, entre otros.

- Tendencia: comportamiento a largo plazo de la serie. Son los movimientos a largo plazo en el nivel promedio de la serie una vez eliminados los movimientos cíclicos y el irregular.

En muchas series económicas se puede observar un crecimiento continuado a lo largo de este siglo. Existen muchas razones para la existencia de estas tendencias crecientes, incluyendo incrementos de la población y la inflación que se observa en muchos países. Debido a esto muchas variables se consideran en términos "per capita" para remover los efectos del incremento de la población o en términos "reales" para tener en cuenta los cambios en los precios. Otras series económicas presentan tendencias debido a cambios tecnológicos y a incrementos en el nivel de vida, por ejemplo, el PIB, etc. Otras tendencias son debidas a cambios en actitudes o en la estructura de la sociedad, como la disminución de la natalidad, o el incremento de la delincuencia.

- Estacionalidad: comportamiento cíclico de periodo exactamente el año y que se agota en sí mismo.

  Suele ser frecuente su presencia en series que se observan varias veces al año, por ejemplo, series mensuales o trimestrales, aunque no se tiene por qué dar siempre. Ejemplos de series con un componente estacional muy fuerte son: producción, ventas, ingresos y gastos del gobierno, el desempleo suele ser más alto en invierno y más bajo en verano, etc. Las principales razones de la existencia de estos ciclos estacionales son las fiestas laborales, como Navidades, Semana Santa o verano, el pago de dividendos, y, por supuesto, los efectos climatológicos.

- Ciclo: movimiento cíclico a medio plazo de periodo superior a un año.

  Es posible notarlo si se cuenta con series anuales durante un largo periodo de tiempo. Así, hay teóricos que argumentan que la economía presenta los denominados "ciclos de negocios" de periodo entre cinco y siete años.

- Irregular: movimientos no sistemáticos de la serie que no se pueden predecir y que suponemos que en promedio son cero.

La especificación de un Modelo de Componentes No Observados se basa en el supuesto de que la serie observada Y_t se forma como una composición de todos (o algunos de estos) elementos. Las figuras del gráfico 2.1 representan la idea subyacente en los modelos de CNO. En este ejemplo, la serie observada resulta de la superposición de un componente irregular al movimiento a largo plazo o tendencia.

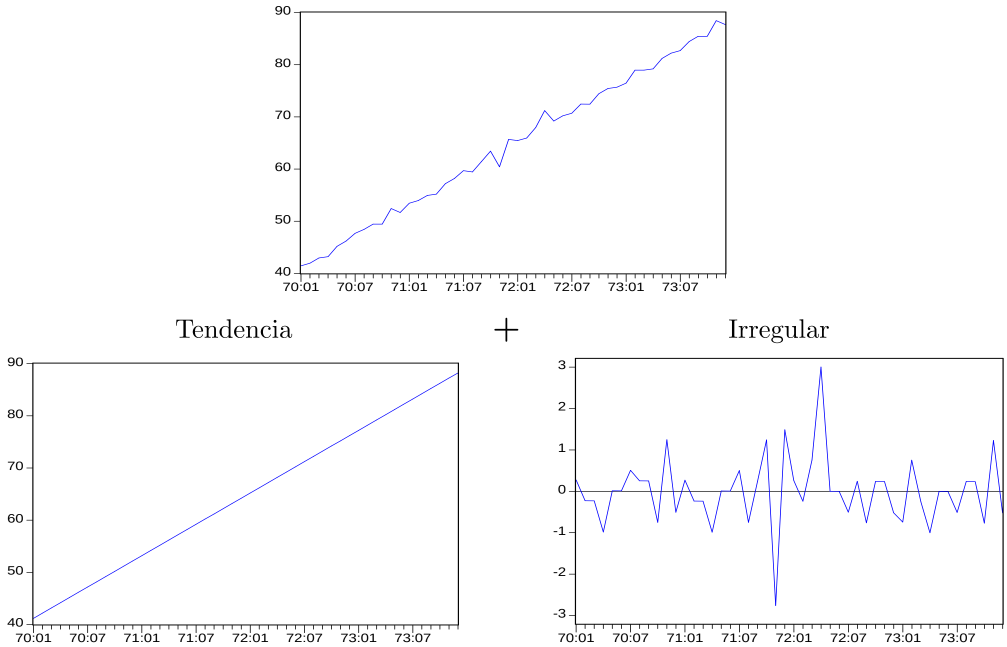
En el caso de las series económicas el punto de partida más general para formular el modelo de CNO es el siguiente:

$$\text{Serie Observada} = f(\text{ Tendencia, Estacionalidad, Ciclo, Irregular })$$

$$\text{Serie Observada} = f(T_t, S_t, C_t, I_t)^1$$

[1]Estacionalidad = S_t, nomenclatura que proviene del inglés *Seasonality*.

Gráfico 2.1: Serie observada = Tendencia + Irregular



Tanto la definición de los componentes como cuáles deben formar parte del modelo es relativamente arbitraria. Los componentes que forman parte de la serie que se analiza dependen de la longitud, del periodo de medida de la misma y de sus características especificas.

La periodicidad de los datos define lo que debe considerarse como tendencia. En series largas, es posible identificar el componente de ciclo; en series no tan largas, sin embargo, los componentes de tendencia y ciclo se confunden. Los movimientos cíclicos raramente tienen un tratamiento independiente, por lo que es bastante habitual que un solo componente recoja los movimientos de más larga duración. A este componente se le suele denominar *Tendencia-Ciclo*, o simplemente Tendencia.

Por otro lado, en las series cuyo periodo de medición es anual no se puede identificar un componente estacional. Para ello, es preciso contar con varias observaciones a lo largo del año, es decir, series trimestrales (periodo estacional, 4) o series mensuales (periodo estacional, 12). Esto no significa que toda serie mensual o trimestral, tenga un comportamiento estacional.

Por último, el componente irregular siempre está presente y forma el residuo o término de error, que puede servir para analizar la propia bondad de la aplicación realizada.

El analista de series temporales debe identificar los principales rasgos del fenómeno que desea analizar e incorporarlos de forma explícita en el modelo. En este sentido, la descomposición anterior no es arbitraria. La simple inspección visual del gráfico de una serie temporal revela frecuntemente que tendencias y estacionalidades son rasgos relevantes de los datos y, por lo tanto, parece deseable modelar estas características explícitamente. Los componentes no observables tratan en definitiva de representar aquellas características de la serie bajo estudio que, por ser de interés en sí mismas para los economistas, conviene especificar en el modelo.

Los modelos de CNO más utilizados en el análisis de series temporales de frecuencia superior o igual a la mensual son del tipo:

$$\text{Serie Observada} = f(T_t, S_t, I_t)$$

El problema que se plantea es determinar cuál es la relación que une a los componentes no observados de una serie, es decir, qué es f. Existen múltiples opciones, pero los modelos más comúnmente utilizados son:

Modelo Aditivo: $Y_t = T_t + S_t + I_t$

Modelo Multiplicativo: $Y_t = T_t \times S_t \times I_t$

Como muestran las figuras del gráfico 2.2, en el modelo aditivo, el valor de la serie consta de un valor dado de la tendencia en el momento t, T_t, al que se le suma el valor de la estacionalidad del mes correspondiente al momento t, S_t, y el del irregular, I_t. Los índices estacionales son positivos, si ese mes está por encima de la media, o negativos, si está por debajo de la media. Como el componente estacional se agota en sí mismo, es decir, no debe añadir nada al comportamiento promedio de la serie recogido por la tendencia, la media de los factores estacionales ha de ser cero.

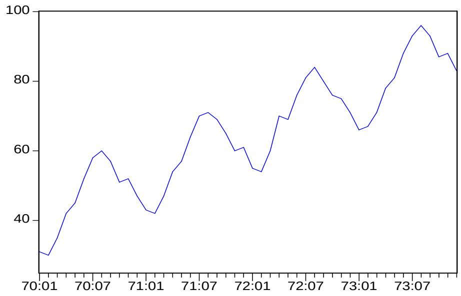
El modelo multiplicativo se puede interpretar como sigue: dado un valor de la tendencia, T_t, su producto por un factor de estacionalidad, S_t, y añadiéndose el irregular, I_t, se obtiene una valoración adecuada de los valores de la serie original (veáse el gráfico 2.2). En este caso, los índices estacionales son mayores que 1 si ese mes está por encima de la media, o menores que 1, si está por debajo de la media. Como el componente estacional se agota en sí mismo, se ha de cumplir que la media de los factores estacionales a lo largo del año ha de ser 1 (ó 100 en índice).

El gráfico de los datos es, precisamente, quien mejor puede indicar cuál es el modelo apropiado si el aditivo o el multiplicativo. Si se observa que la estacionalidad es constante respecto a la tendencia, el modelo apropiado sería el aditivo. Si el gráfico muestra una estacionalidad cuya amplitud aumenta con la tendencia, el modelo apropiado sería el multiplicativo.

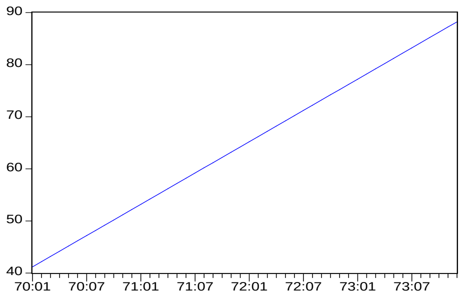
Los pasos que se siguen al analizar series temporales mediante los modelos de componentes no observados son los siguientes:

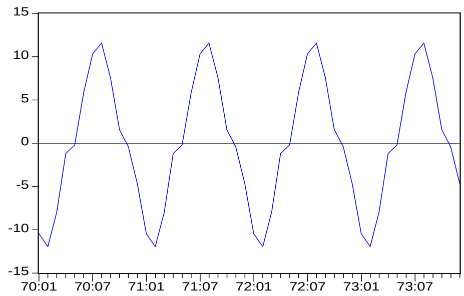
Gráfico 2.2: Serie observada = f(Tendencia, Estacionalidad, Irregular)

Modelo ADITIVO

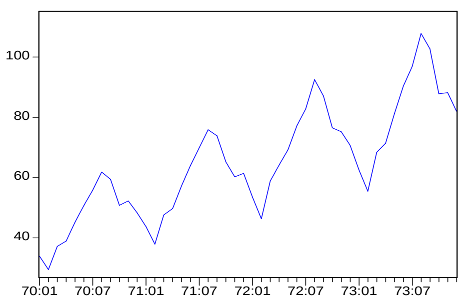


Tendencia + Estacionalidad

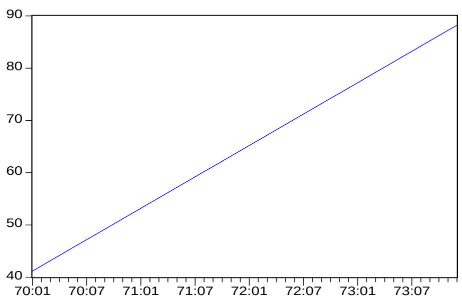


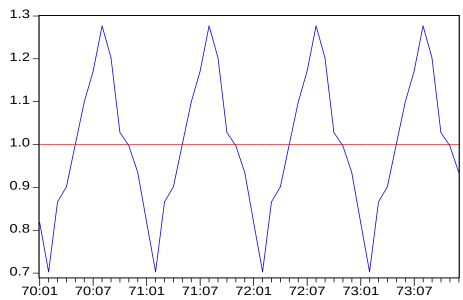


Modelo MULTIPLICATIVO



Tendencia **x** Estacionalidad





a) Especificar el modelo apropiado para la serie, lo que implica, seleccionar los componentes que forman parte de la misma y la relación que existe entre ellos. Además, la descomposición de la serie en componentes no es única y, generalmente, se incluye en este tipo de análisis algún tipo de modelización de los componentes tanto explícita como implícita, es decir, qué modelo es apropiado para la tendencia, el ciclo, etc.

   A veces, aunque el modelo apropiado sea el modelo multiplicativo, no resulta sencillo trabajar directamente con él. En este caso, si tomamos logaritmos al modelo multiplicativo, se convierte en un modelo aditivo,

$$Y_t = T_t \times S_t \times I_t \quad \longrightarrow \quad \log Y_t = \log T_t + \log S_t + \log I_t \quad \longrightarrow \quad \log Y_t = T_t + S_t + I_t$$

b) Estimar los componentes no observados a partir de la serie original:

$$\hat{T}_t, \;\; t = 1, 2, \ldots, T \qquad\qquad \hat{S}_t, \;\; t = 1, 2, \ldots, T$$

c) Con los componentes estimados se puede:

1. Describir la evolución temporal de la serie. Una vez estimada la tendencia para cada momento t, $\hat{T}_t$, podemos describir sus características (crece, decrece, a qué ritmo, sufre cambios, etc.) y lo mismo con la estacionalidad estimada S_t (qué meses son más relevantes, etc.).

Ahora bien, también nos puede interesar utilizar estos componentes estimados para eliminarlos de la serie. Así, en muchas aplicaciones económicas, se prefiere utilizar las series libres del componente estacional para, por ejemplo, poder observar mejor el comportamiento a largo plazo. Se denomina *serie desestacionalizada* a la serie de la que se ha eliminado el componente estacional.

Modelo CNO aditivo: Serie sin tendencia: $Y_t^{st} = Y_t - \hat{T}_t$

Serie desestacionalizada : $Y_t^d = Y_t - \hat{S}_t$

Modelo CNO multiplicativo: Serie sin tendencia: $Y_t^{st} = \dfrac{Y_t}{\hat{T}_t}$

Serie desestacionalizada : $Y_t^d = \dfrac{Y_t}{\hat{S}_t}$

2. Predecir. Vamos a distinguir entre los modelos aditivos y multiplicativos.

<u>Modelo CNO Aditivo:</u>

$$Y_t = T_t + S_t + I_t, \;\; t = 1, 2, \ldots, T$$

Para predecir ℓ periodos hacia adelante, el verdadero valor de la serie viene dado por el modelo:

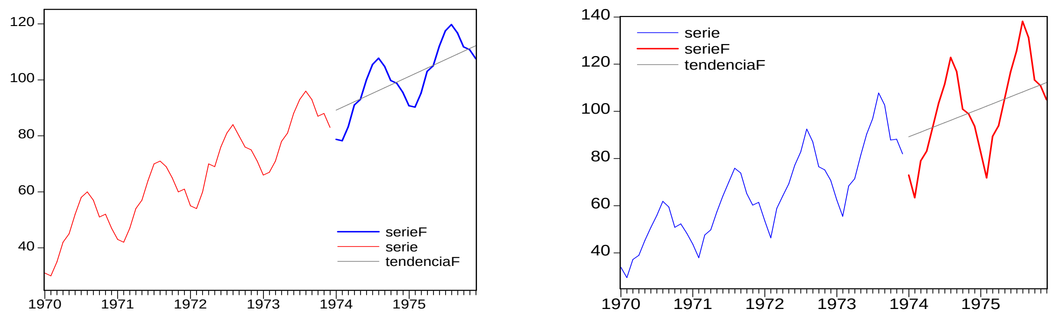
$$Y_{T+\ell} = T_{T+\ell} + S_{T+\ell} + I_{T+\ell}, \quad \ell = 1, 2, \ldots, L$$

La predicción viene dada por[2]:

$$Y_T(\ell) = T_T(\ell) + S_T(\ell)$$

La figura de la izquierda del gráfico 2.3 muestra como la predicción de los valores de la serie se obtiene proyectando la tendencia a la que se suma el componente estacional estimado, de forma que la amplitud del componente estacional permanece constante.

Gráfico 2.3: Predicción por punto.



Modelo Aditivo — Modelo Multiplicativo

Modelo CNO Multiplicativo:

$$Y_t = T_t \times S_t \times I_t, \quad t = 1, 2, \ldots, T$$

Para predecir ℓ periodos hacia adelante, el verdadero valor de la serie viene dado:

$$Y_{T+\ell} = T_{T+\ell} \times S_{T+\ell} \times I_{T+\ell}, \quad \ell = 1, 2, \ldots, L$$

La predicción viene dada por: $\quad Y_T(\ell) = T_T(\ell) \times S_T(\ell)$

Es decir, la predicción se obtiene multiplicando el valor estimado de la tendencia en el momento $T + \ell$ por los factores estacionales estimados. Como la tendencia es creciente, se puede observar en la figura derecha del gráfico 2.3 que en las predicciones la amplitud del ciclo estacional crece en el tiempo.

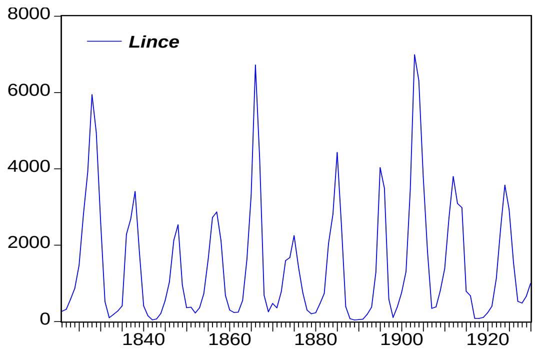
[2] $T_T(\ell) = \hat{T}_{T+\ell}$, es decir, la estimación de la tendencia en el momento $T + \ell$, obtenida a través del modelo especificado para la tendencia.
$S_T(\ell) = \hat{S}_{T+\ell}$, es decir, la estimación de la estacionalidad en el momento $T + \ell$, obtenida a través del modelo especificado para la estacionalidad.

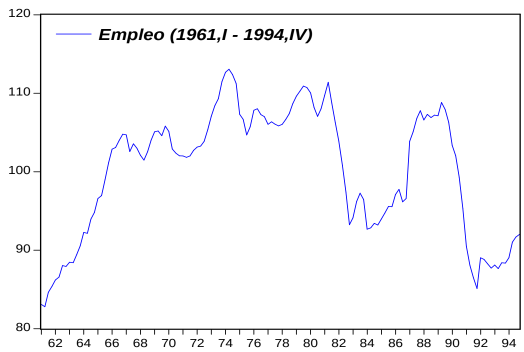
Los componentes no observados se pueden especificar de forma determinística como funciones del tiempo. En este caso los modelos de CNO se denominan modelos globales deterministas. Estos modelos imponen a la serie la restricción de seguir un patrón fijo a lo largo del tiempo lo que no es muy satisfactorio cuando se trabaja con series económicas. Además, en estos modelos a la hora de predecir todas las observaciones a lo largo del tiempo reciben el mismo peso. Este tipo de modelos serán desarrollados en el capítulo 3 para series con tendencia y en el capítulo 4 para series con tendencia y estacionalidad.

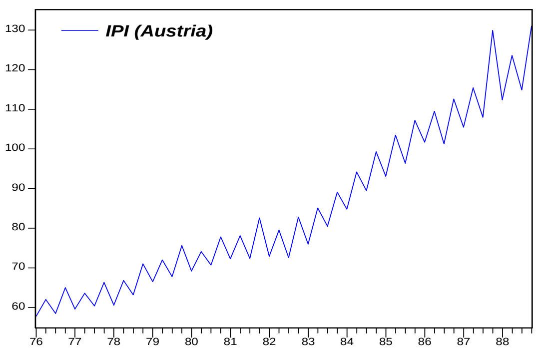
Se pueden formular también modelos más flexibles en los que ni la tendencia ni el componente estacional son funciones deterministas del tiempo sino que son capaces de variar suavemente a lo largo del tiempo recogiendo los posibles cambios de estructura de la serie. Esto se traduce, a la hora de hacer predicciones, en poner más peso en las observaciones más recientes. Los modelos de alisado exponencial con tendencia se estudian en el capítulo 3) y los modelos de alisado exponencial con estacionalidad en el capítulo 4.

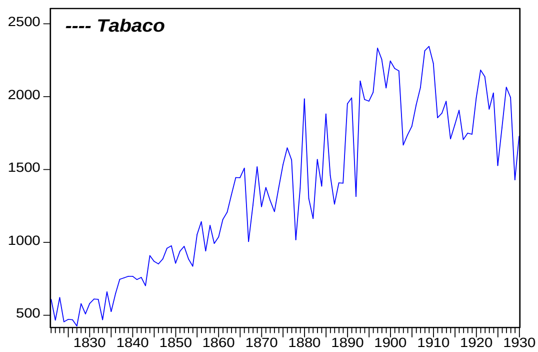
Otra posibilidad para especificar modelos flexibles se basa en especificar explícitamente los componentes no observados de tendencia, estacionalidad, etc. de forma estocástica de manera que se permite en el modelo una lenta evolución de los mismos. Este tipo de modelos se denominan Modelos Estructurales de Series Temporales (Harvey, 1989) y se van a desarrollar en el capítulo 5. Estos modelos estructurales son modelos locales y estocásticos que incluyen a los modelos globales deterministas como caso límite. Esto quiere decir que se pueden interpretar como modelos de regresión en los que las variables explicativas son funciones del tiempo y sus parámetros cambian con el tiempo.

Ejercicio 2.1. Especifica los modelos de CNO apropiados para las siguientes series.









Capítulo 3

Análisis de una serie con tendencia

La definición más común de tendencia señala que es el movimiento a largo plazo de las series, una vez eliminadas las variaciones irregulares y los ciclos si los hubiera. Ahora bien, es bastante común denominar con la palabra *tendencia* dos ideas diferentes. Por un lado, el nivel medio de la serie en el momento t viene dado por T_t y se le denomina *tendencia*. Por otro, si la serie presenta un comportamiento creciente o decreciente, a la pendiente se le suele denotar también tendencia. En este caso la tendencia recogería el cambio en el nivel de la serie por unidad de tiempo. De todas formas, suele ser fácil, por el contexto, saber a qué nos estamos refiriendo en cada caso.

En lo que al componente de tendencia se refiere, es interesante identificarla y estimarla porque para predecir suele ser importante conocer el comportamiento sistemático de la serie. Existen diferentes métodos para analizar y estimar tendencias que dividiremos en dos grandes grupos: métodos globales y métodos locales. Por otro lado, en ocasiones, también puede ser de interés eliminar la tendencia de la serie, por ejemplo, porque se quiera aislar la estacionalidad. Para llevarlo a cabo, se puede o bien eliminar la tendencia ya estimada de la serie o bien tomar diferencias a la serie original.

3.1. Modelos globales: ajuste de funciones matemáticas

Los modelos globales se basan en que la evolución del largo plazo de la serie, o tendencia, se puede recoger simplemente como una función del tiempo. Se pueden utilizar infinidad de funciones del tiempo para representar la tendencia pero, en general, las de más interés son las funciones más sencillas que no dependen de muchos parámetros.

Funciones polinómicas del tiempo.

1.a) Función lineal:

$$T_t = a + b\,t \quad t = 1, 2, \ldots, T$$

donde a es el intercepto y b es la pendiente. Con esta especificación suponemos un comportamiento lineal de la tendencia, es decir, independientemente del nivel de la serie o del momento de tiempo el crecimiento ($b > 0$) o decrecimiento ($b < 0$) de la tendencia es siempre constante:

$$\frac{\partial T_t}{\partial t} = b$$

1.b) Función cuadrática:

$$T_t = a + b\,t + c\,t^2 \quad t = 1, 2, \dots, T$$

La tendencia no es lineal, es decir, el crecimiento de la tendencia depende del tiempo:

$$\frac{\partial T_t}{\partial t} = b + 2\,c\,t \quad t = 1, 2, \dots, T$$

1.c) En general se puede ajustar cualquier función polinómica:

$$T_t = a_0 + a_1\,t + a_2\,t^2 + \ldots + a_n\,t^n \quad t = 1, 2, \dots, T$$

Ahora bien, no se trata de ajustar una curva que pase por todos los puntos de la serie temporal[1], sino de estimar el nivel promedio de la serie, su evolución a largo plazo. Se intentará siempre especificar el modelo más sencillo posible que responda a la evolución a largo plazo de la serie.

La estimación de la tendencia con los datos disponibles es muy sencilla en los modelos globales con tendencias polinómicas. El modelo de CNO queda, en general, como sigue:

$$\left.\begin{array}{l} Y_t = T_t + I_t \\ T_t = f(t) \end{array}\right\} \quad t = 1, 2, \dots, T$$

En el caso particular de la tendencia lineal:

$$\left.\begin{array}{l} Y_t = T_t + I_t \\ T_t = a + b\,t \end{array}\right\} \quad t = 1, 2, \dots, T$$

Sobre el componente irregular, I_t, sólo se supone que es aleatorio, impredecible y que en promedio es cero. Sustituyendo el modelo para la tendencia, se obtiene:

$$Y_t = a + b\,t + I_t \quad t = 1, 2, \dots, T \tag{3.1}$$

que es un modelo de regresión lineal donde Y_t es la variable endógena y t, el tiempo, es la variable explicativa. Por lo tanto, los parámetros del modelo, a, b, se pueden estimar

[1] Como es sabido si se dispone de una serie temporal de T observaciones y se estima una función polinómica de order $T - 1$, el ajuste sería perfecto ya que dicha función pasaría por los T puntos de la serie temporal.

por Mínimos Cuadrados Ordinarios (MCO), es decir, mediante el criterio de minimizar la suma de los errores al cuadrado (veáse apéndice A):

$$S = \sum_{t=1}^{T} (Y_t - T_t)^2 = \sum_{t=1}^{T} (Y_t - a - bt)^2$$

lo que resultaría en:

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = (X'X)^{-1} X'Y$$

donde $Y = [Y_1, Y_2, \ldots, Y_T]'$ es el vector de datos de la serie observada y X es la matriz de datos de las variables explicativas.

La variable *tiempo* no es una variable explicativa, en el sentido de que se pueda tomar una muestra de sus valores. De hecho, el *tiempo* se puede medir de muchas maneras: su valor en sí mismo no es importante, lo que tiene que recoger esta variable es la definición del origen a partir del cual se comienza a contar el tiempo y luego las unidades en que se mide. Por ejemplo, si la variable endógena toma valores anuales desde 1960 a 2001, la variable tiempo se puede medir, entre otras opciones, como sigue:

$$\begin{array}{lcl} \text{Opción A:} \quad t & = & 1960, 1961, 1962, \ldots, 2001 \\ \text{Opción B:} \quad t & = & 1, 2, 3, \ldots, 42 \\ \text{Opción C:} \quad t & = & -21, -20, -19, \ldots, 20 \end{array}$$

Por lo tanto, la matriz X para el modelo (3.1) puede tomar muchas formas:

$$X^A = \begin{bmatrix} 1 & 1960 \\ 1 & 1961 \\ 1 & 1962 \\ \vdots & \vdots \\ 1 & 2002 \end{bmatrix} \qquad X^B = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ \vdots & \vdots \\ 1 & 43 \end{bmatrix} \qquad X^C = \begin{bmatrix} 1 & -21 \\ 1 & -20 \\ 1 & -19 \\ \vdots & \vdots \\ 1 & 21 \end{bmatrix}$$

De cualquiera de estas maneras la variable *tiempo* mide lo mismo, a saber, el número de periodos para los que tenemos datos de la serie temporal.

Una vez estimados los parámetros del modelo (3.1), la tendencia estimada para cada t es:

$$\hat{T}_t = \hat{a} + \hat{b}\, t, \quad t = 1, 2, \ldots, T$$

La figura superior izquierda del gráfico 3.1 representa la *Tasa de actividad de la mujer* en la Comunidad Autónoma de Euskadi desde el tercer trimestre de 1988 al segundo de 2003. Se puede observar un crecimiento continuado y regular de la misma por lo que el modelo (3.1) parece, en principio, apropiado para reproducir el comportamiento de la serie. Los resultados de su estimación por MCO son (véase la tabla del gráfico 3.1):

$$\hat{Y}_t = \underset{(0,33)}{25,97} + \underset{(0,0044)}{0,17}\ t \qquad R^2 = 0,965 \quad \bar{R}^2 = 0,9645 \quad AIC = 1,901 \quad SIC = 1,969$$

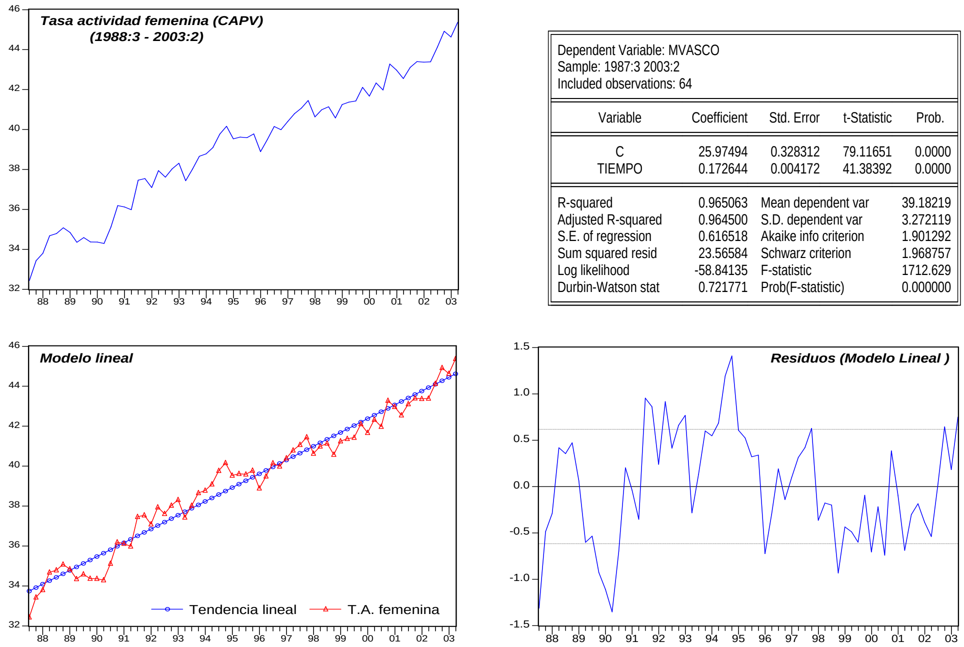
siendo la tendencia estimada:

$$\hat{T}_t = 25,97 + 0,17\ t$$

Para predecir las futuras observaciones de la serie temporal, $Y_{T+\ell}, \ell = 1, 2, \ldots$, su verdadero valor viene dado por el modelo, en este caso, el modelo (3.1):

$$Y_{T+\ell} \ = \ T_{T+\ell} + I_{T+\ell} \ = \ a + b\,(T + \ell) + I_{T+\ell}, \quad \ell \ = \ 1, 2, \ldots$$

Gráfico 3.1: Tasa de actividad de la mujer.



Dependent Variable: MVASCO
Sample: 1987:3 2003:2
Included observations: 64

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	25.97494	0.328312	79.11651	0.0000
TIEMPO	0.172644	0.004172	41.38392	0.0000

R-squared	0.965063	Mean dependent var	39.18219
Adjusted R-squared	0.964500	S.D. dependent var	3.272119
S.E. of regression	0.616518	Akaike info criterion	1.901292
Sum squared resid	23.56584	Schwarz criterion	1.968757
Log likelihood	-58.84135	F-statistic	1712.629
Durbin-Watson stat	0.721771	Prob(F-statistic)	0.000000

Como estamos considerando que la serie temporal Y_t depende de una tendencia determinista y de un componente irregular aletorio, el valor que se quiere predecir, $Y_{T+\ell}$, es también una variable aleatoria. Para predecir una variable aleatoria, se debería predecir su función de distribución y así poder hacer afirmaciones tales como: $prob(163 < Y_{T+\ell} \leq 190) = 0,42$. Ahora bien, en general, va a ser muy difícil determinar completamente la forma de la función de densidad sin hacer supuestos muy fuertes y poco realistas sobre la forma de esta función. Un objetivo menos ambicioso sería diseñar unos intervalos de confianza alrededor del valor $Y_{T+\ell}$, que nos permitan decir que $prob(B < Y_{T+\ell} \leq A) = 0,95$. Estos valores A y B permiten poner unos límites al valor que se quiere predecir con un grado de confianza de estar en lo cierto suficientemente alto.

Es interesante distinguir entre *predicción por intervalo* que conlleva la construcción de estos intervalos de predicción y la *predicción por punto*, que implica simplemente asignar un valor a $Y_{T+\ell}$ que de alguna manera represente a toda la distribución de valores. Por ejemplo, una predicción por punto sería afirmar que "se predice un número de ocupados en el sector industrial el primer trimestre de 2010 de 120.000". Un ejemplo de la predicción por intervalo, sería decir: "se predice un número de ocupados en el sector industrial el primer trimestre de 2010 entre 100.000 y 140.000 con una probabilidad de 0,95". Si el método de predicción es bueno y si se pudiera hacer un secuencia de predicciones de este tipo, es de esperar que el número real de ocupados estaría fuera de los intervalos construidos solo en un 5 % de los casos.

Como el componente irregular no es sistemático y se supone que en promedio es cero, las *predicciones por punto* para el modelo (3.1) son:

$$\begin{aligned} Y_T(1) &= T_T(1) = \hat{T}_{T+1} = \hat{a} + \hat{b}\,(T+1) \\ Y_T(2) &= T_T(2) = \hat{T}_{T+2} = \hat{a} + \hat{b}\,(T+2) \\ Y_T(3) &= T_T(3) = \hat{T}_{T+3} = \hat{a} + \hat{b}\,(T+3) \\ \vdots \quad & \quad \vdots \end{aligned}$$

La función de predicción es una función de ℓ que nos permite obtener todas las predicciones por punto para cada ℓ. En nuestro ejemplo,

$$Y_T(\ell) = T_T(\ell) = \hat{T}_{T+\ell} = \hat{a} + \hat{b}\,(T+\ell) \qquad \ell = 1, 2, 3, \ldots$$

Al calcular las predicciones por punto hay que tener en cuenta cómo se ha medido la variable t. En el ejemplo anterior, la predicción un periodo hacia adelante, es decir, del valor de la serie en el año 2003 sería:

$$\begin{array}{ll} \text{Opción A:} & Y_{2002}(1) = T_{2002}(1) = \hat{T}_{2003} = \hat{a} + \hat{b}\ 2003 \\ \text{Opción B:} & Y_{43}(1) = T_{43}(1) = \hat{T}_{44} = \hat{a} + \hat{b}\ 44 \\ \text{Opción C:} & Y_{21}(1) = T_{21}(1) = \hat{T}_{22} = \hat{a} + \hat{b}\ 22 \end{array}$$

Si suponemos que el componente irregular, I_t sigue la distribución Normal, la *predicción por intervalo* de un $(1-\alpha)$ % de confianza es (veáse apéndice A):

$$P\left[Y_T(\ell) - N_{\alpha/2}\sqrt{V(e_T(\ell))} \;\le\; Y_{T+\ell} \;\le\; y_T(\ell) + N_{\alpha/2}\sqrt{V(e_T(\ell))}\right] = (1-\alpha)\,\%$$

donde $Y_T(\ell)$ es la predicción por punto y $V(e_T(\ell))$ es la varianza del error de predicción correspondiente. Esta varianza no se suele conocer por lo que hay que estimarla y el intervalo de predicción del $(1-\alpha)$ % se obtiene como sigue:

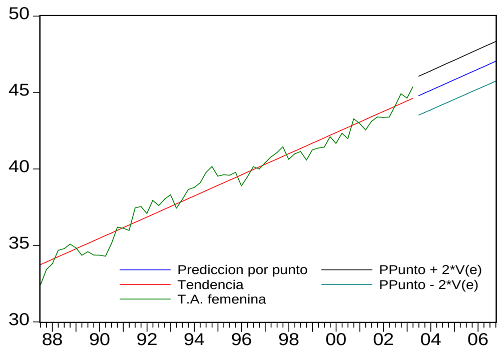
$$P\left[Y_T(\ell) - t_{\alpha/2}(T-k)\sqrt{\hat{V}(e_T(\ell))} \;\le\; Y_{T+\ell} \;\le\; Y_T(\ell) + t_{\alpha/2}(T-k)\sqrt{\hat{V}(e_T(\ell))}\right]$$

donde, $\hat{V}(e_T(\ell)) = \hat{\sigma}_I^2\, [1 + X_{T+\ell}\,(X'X)^{-1}\,X'_{T+\ell}], \quad X'_{T+\ell} = [1 \quad (T+\ell)]$

La *interpretación* del intervalo de predicción es la siguiente: si contáramos con muchas series temporales de la misma variables y fuéramos capaces de construir repetidos intervalos el $(1-\alpha)\,\%$ de ellos contendrían el verdadero valor de $Y_{T+\ell}$. Como solo se suele tener una serie temporal y solo construimos un intervalo no podemos estar seguros de que contenga el verdadero valor.

El gráfico 3.2 muestra las predicciones por punto y por intervalo para la *Tasa de Actividad femenina* para el periodo de 2003:3 a 2006:4. Como se puede observar, la función de predicción es simplemente la proyección de la tendencia lineal.

Gráfico 3.2: Tasa de actividad femenina. Predicción.



En ocasiones se pueden ajustar diferentes funciones de tendencia a una misma serie. Por ejemplo, si ajustamos a la serie de *Tasa de Actividad femenina* un modelo con tendencia cuadrática:

$$Y_t \;=\; T_t + I_t \;=\; a + b\,t + c\,t^2 + I_t \qquad t = 1, 2, \ldots, T \tag{3.2}$$

se obtienen los siguientes resultados presentados en la tabla y la figura del gráfico 3.3.

Ejercicio 3.1. ¿Qué forma toma la matriz X para la función de tendencia cuadrática? ¿Cuál sería la función de predicción para este modelo?

El problema que se plantea ahora es cómo seleccionar el modelo que mejor reproduce el comportamiento de la serie Tasa de actividad de la mujer.

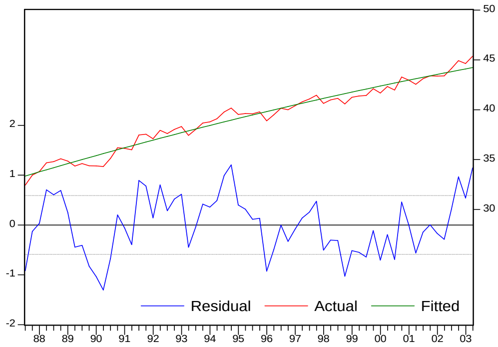
En el análisis de regresión contamos con el coeficiente de determinación, R^2, que es una medida de la bondad de ajuste del modelo que indica el porcentaje de la variabilidad muestral de la variable endógena que es explicada por las variables exógenas, en este caso la variable tiempo. Pero este coeficiente presenta el problema de que nunca decrece al añadir más variables explicativas, por lo que el coeficiente de determinación del modelo

Gráfico 3.3: Tasa actividad femenina. Tendencia cuadrática.

Dependent Variable: MVASCO
Sample: 1987:3 2003:2
Included observations: 64

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	22.62872	1.372992	16.48132	0.0000
TIEMPO	0.265544	0.037315	7.116181	0.0000
TIEMPO^2	-0.000607	0.000242	-2.504045	0.0150

R-squared	0.968320	Mean dependent var	39.18219
Adjusted R-squared	0.967281	S.D. dependent var	3.272119
S.E. of regression	0.591875	Akaike info criterion	1.834698
Sum squared resid	21.36928	Schwarz criterion	1.935896
Log likelihood	-55.71034	F-statistic	932.2394
Durbin-Watson stat	0.795348	Prob(F-statistic)	0.000000



(3.2) nunca va a ser menor que el del modelo (3.1). Para paliar este problema, se han diseñado medidas que tienen en cuenta la inclusión de más variables y penalizan el tener que estimar más parámetros en el modelo:

$\bar{R}^2$, denominado coeficiente de determinación corregido:

$$\bar{R}^2 = 1 - \frac{\sum_{t=1}^{T} \hat{I}_t^2/(T-k)}{\sum_{t=1}^{T} (y_t - \hat{y}_t)^2/(T-1)}$$

AIC, o Criterio de Información de Akaike:

$$AIC = e^{2k/T} \, \frac{\sum_{t=1}^{T} \hat{I}_t^2}{T}$$

SIC, o Criterio de Información de Schwarz

$$SIC = T^{k/T} \, \frac{\sum_{t=1}^{T} \hat{I}_t^2}{T}$$

Ejercicio 3.2. ¿Qué modelo seleccionarías para la tasa de actividad: el modelo con tendencia lineal o con tendencia cuadrática?

Funciones exponenciales. Un comportamiento no lineal a largo plazo se puede representar también mediante funciones de tendencia exponenciales,

$$T_t = e^{a+b\,t}$$

de forma que

$$Y_t = T_t\ I_t \tag{3.3}$$

La figura izquierda del gráfico 3.4 muestra el volumen mensual de acciones negociado en la bolsa de New York en un determinado periodo[2]. Los ajustes de tendencias lineales y cuadráticas no son capaces de recoger el comportamiento de la serie (veánse las figuras del gráfico 3.5).

Gráfico 3.4: Serie: Volumen de acciones en la bolsa de New York.

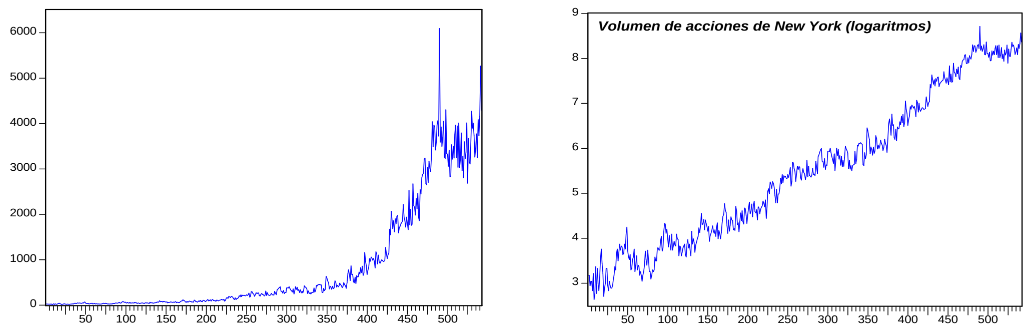
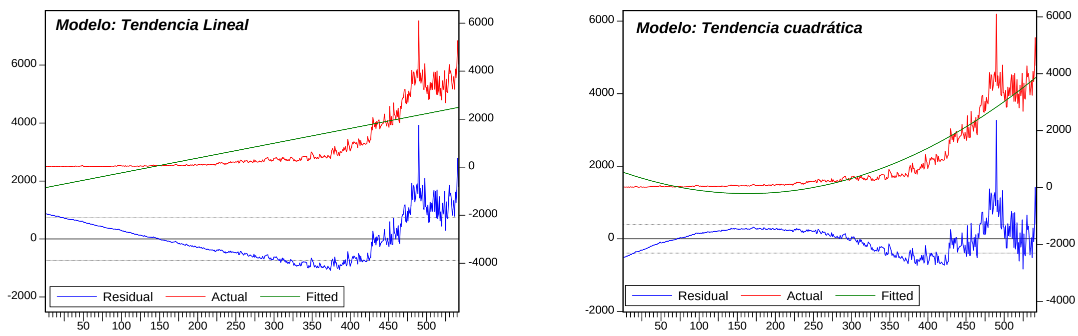


Gráfico 3.5: Bolsa New York. Tendencia lineal y cuadrática.



Sin embargo, si tomamos logaritmos a la serie, su evolución es lineal, como muestra la figura de la derecha del gráfico 3.4. Esto implica la presencia de una tendencia exponencial en los datos originales, ya que tomando logaritmos al modelo (3.3), se obtiene:

$$\log(Y_t) = a + b\,t + I_t \tag{3.4}$$

[2]Veáse Francis X. Diebold (1998). *Elements of forecasting*, Ed. South-Western.

La estimación de estas tendencias exponenciales se puede realizar, estimando por MCO el modelo (3.4) o estimando directamente el modelo (3.3).

Curvas de crecimiento. Las más utilizadas son las siguientes:

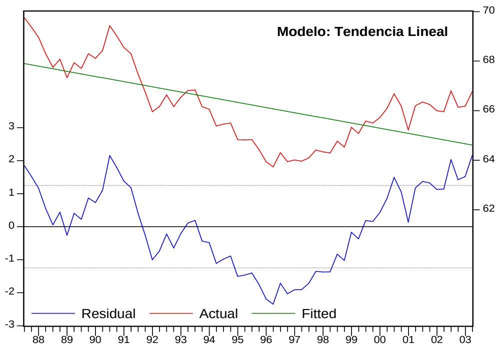
- La curva Gompertz $\log T_t = a - br^t \qquad 0 < r < 1.$
- La curva logística: $T_t = \dfrac{1}{a + br^t} \qquad 0 < r < 1.$

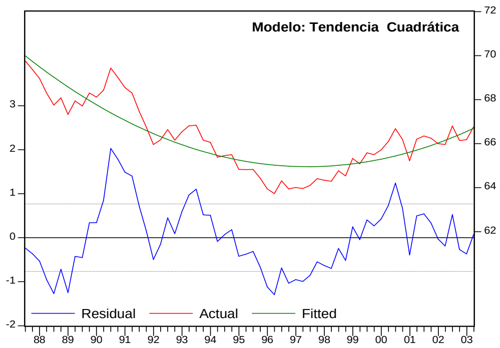
Estas curvas tienen la forma de S y toman cierto valor límite cuando $t \to \infty$. Este valor es e^a para la curva Gompertz y $1/a$ para la logística. Por lo tanto, estas curvas solo son apropiadas para variables que tengan un límite superior alcanzable además en un futuro no muy lejano, como por ejemplo, el porcentaje de hogares de un país con agua corriente.

Los parámetros desconocidos de estos componentes de tendencia (a, b) se pueden estimar por MCO trás linealizar los modelos realizando las transformaciones correspondientes.

Como se puede observar estimar tendencias mediante el ajuste de funciones del tiempo es bastante sencillo. El inconveniente de estos procedimientos es que son esquemas muy rígidos ya que exigen que el modelo se ajuste a todas las observaciones. Tanto la tendencia más sencilla de la ecuación lineal como las otras funciones propuestas, más complejas, son funciones deterministas del tiempo y suelen recibir el nombre de *tendencias lineales globales*. Como este supuesto es, en general, poco realista en la actualidad se tiende a pensar en esquemas más flexibles que permitan a los parámetros a y b variar en el tiempo, de forma que la tendencia vaya evolucionando con el tiempo, en lo que se denomina *tendencia local lineal*.

Gráfico 3.6: Tasa actividad masculina. Tendencia lineal y cuadrática.





Esta cuestión se observa claramente analizando la serie de Tasa de actividad del hombre en la CAE. Las figuras del gráfico 3.6 presentan los resultados del ajuste de una tendencia lineal y una cuadrática. Se observa que esta variable tiene una tendencia que cambia de pendiente durante este periodo. Ni el ajuste lineal ni el ajuste cuadrático resultan satisfactorios. Convendría o bien ajustar dos funciones de tiempo diferentes para los dos tramos de comportamiento diferenciado presentes en la serie o utilizar modelos más flexibles que permitan estimar tendencias que vayan cambiando según la evolución temporal de la serie.

3.2. Métodos de alisado

Estos métodos son muy útiles si se desea obtener predicciones a corto plazo de un gran número de variables porque se pueden utilizar muy fácilmente y proporcionan predicciones razonablemente buenas.

La noción básica de las técnicas de alisado es que existe algún patrón subyacente en la evolución de las variables que se tienen que predecir y que los datos históricos de las mismas representan tanto el patrón subyacente como la variabilidad aleatoria. El objetivo de estos métodos de predicción es distinguir entre las variaciones aleatorias y el patrón subyacente por medio del alisado, del promedio de datos. Esto supone eliminar la aleatoriedad de los datos y basar la predicción en el patrón alisado de los datos.

Por alisado o suavizamiento entendemos la descomposición de una serie temporal en un componente suave (alisado) y uno irregular. Al alisar una serie pretendemos eliminar los movimientos no sistemáticos con el fin de resaltar los principales aspectos de la misma; todo ello sin formular de forma explícita modelos o hipótesis. Dentro de este marco, se puede considerar que los métodos de ajuste de funciones polinómicas son métodos de alisado con un modelo previo y que proporcionan tendencias globales.

Supongamos que se cuenta con la serie de datos $Y_1, Y_2, \ldots, Y_T$. Los métodos de predicción más sencillos, que se suelen denominar ingenuos o naïve, son aquellos procedimientos de predicción que repiten de forma mecánica un comportamiento pasado. Por ejemplo:

$$\begin{aligned} Y_T(1) &= Y_T \\ Y_T(1) - Y_T &= Y_T - Y_{T-1} \end{aligned}$$

En el primer caso consideramos que la predicción es igual al último dato observado y, en el segundo, creemos que los incrementos son iguales a lo largo del tiempo.

Una variante un poco más elaborada es utilizar como predicción el valor medio a lo largo de un periodo. ¿Cómo se puede obtener este valor? Calcular la media aritmética es algo trivial: basta con sumar todas las observaciones y dividirlas por el número total de ellas, T. La media aritmética de una serie Y_t es una media global, se ajusta a todas las observaciones, dándole además el mismo peso a todas ellas. Este esquema de predicción sólo tendría sentido en una serie sin tendencia y con oscilaciones aleatorias alrededor de la media. En este caso, el valor más probable de predicción sería precisamente la media. Sin embargo, si la serie, por ejemplo, crece sistemáticamente con el tiempo y queremos tomar alguna decisión con respecto a esta media, no parece buena idea tomar los mismos valores al final que al principio de la serie. Sería más flexible trabajar con medias móviles, es decir, con medias de un número preestablecido de datos, en que se va añadiendo sucesivamente un dato nuevo y quitando al mismo tiempo, el más antiguo de los incluidos en la media anterior.

Los métodos de alisado se basan en las medias móviles. En estos métodos lo que se hace es utilizar los datos disponibles para obtener un valor alisado de la serie. Después este valor alisado se extrapola y se convierte en nuestra predicción para un valor futuro de la serie. En este capítulo se van a desarrollar dos clases de métodos de alisado: las Medias Móviles y las técnicas de Alisado Exponencial.

3.2.1. Medias móviles

La serie correspondiente a la media móvil, por un lado, alisa las variaciones estacionales, cíclicas y/o erráticas de la serie original, y, por otro, si existe algún dato anómalo que se desvía mucho de los demás, al hacer medias móviles ya no se desviará tanto; de ahí que se incluya entre las denominadas técnicas de alisado. Así, la serie alisada aunque tenga oscilaciones siempre será más regular que la original[3].

Cuanto mayor sea el número de términos de la media móvil, la denominada banda de alisado, más suavizada quedará la serie, pero también serán más los momentos de tiempo para los que no podrá calcularse la media móvil, por falta de datos anteriores, es decir, perderemos más observaciones al principio y al final de la muestra (si se calculan medias móviles centradas). El caso más extremo es la media muestral global que se puede reinterpretar como una media móvil de orden $s = T$, con lo que solo queda un valor alisado. No existen resultados de carácter general que permitan aconsejar sobre la amplitud de la banda de alisado, es decir, sobre el orden de la media móvil . Ahora bien, podemos observar que las medias móviles suponen un suavizado de la serie eliminando precisamente las oscilaciones de periodo s, es decir, de periodo igual a la banda de alisado. Se podría escoger como periodo s, el periodo de las oscilaciones más importantes de la serie.

La idea subyacente en este método, es que como el componente irregular es impredecible y en media suponemos que es cero, al hacer promedios de la serie se espera eliminar aproximadamente el componente irregular y quedarnos así solo con el componente de largo plazo o tendencia. Si además la serie presentará algún tipo de comportamiento cíclico, tomando medias móviles de igual orden que el periodo del ciclo este componente también se eliminaría de la serie y quedaría sólo la tendencia. Así, una vez obtenidas las medias móviles de orden $s = 2r + 1$, la tendencia será la línea quebrada que las una:

$$\hat{T}_t = M_t, \quad t = r + 1, r + 2, \ldots, T - r$$

[3]La definición de media móvil y algunos ejemplos se puede encontrar en el apéndice B.

La tendencia estimada mediante medias móviles no es una tendencia global sino que es una *tendencia local*, porque va cambiando y ajustándose a la evolución temporal de la serie. Con el método de las medias móviles se obtiene una tendencia variable a lo largo de la serie, aunque tanto más estable cuanto mayor sea el número de términos de la media móvil o "memoria". Cuanto más ancha sea la banda de alisado más tiempo permanece la información de Y_t en la serie de medias móviles M_t del proceso de alisado.

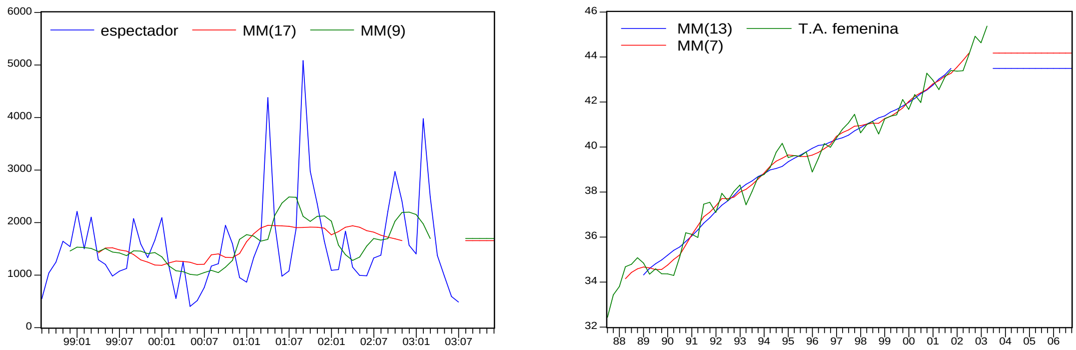
Ahora bien, a efectos de predicción, se toma la última media móvil calculada:

$$Y_T(1) = M_{T-r} \quad Y_T(2) = M_{T-r} \quad Y_T(3) = M_{T-r}, \ \ldots$$

$$\text{Función de predicción} \qquad Y_T(\ell) = M_{T-r} \quad \ell = 1, 2, 3, \ldots$$

Como la función de predicción obtenida es horizontal, solo es razonable para series que no presenten tendencia ni estacionalidad. La serie de medias móviles, es decir, la tendencia estimada puede tener comportamiento creciente o decreciente en el tiempo, pero no es inmediato incorporarlo en la función de predicción. Por lo tanto, aunque este método puede ser útil para analizar y estimar tendencias, no lo será tanto para predecir valores futuros de las series.

Gráfico 3.7: Predicción mediante Medias Móviles.



Consideréndese los siguientes casos. La figura izquierda del gráfico 3.7, estima la tendencia de la serie *Espectadores de cine español* mediante medias móviles y luego la proyecta al futuro. Los resultados pueden ser razonables, ya que esta serie oscila en torno a una tendencia bastante estable. Sin embargo, la figura derecha del gráfico 3.7, estima la tendencia por medias móviles de la serie *Tasa de Actividad de la mujer* que presenta una tendencia creciente. Las medias móviles son capaces de estimar la tendencia de la serie (al ser un procedimiento local se ajusta a los cambios de evolución en la tendencia de la serie) pero, sin embargo no pueden incorporar este comportamiento en la función de predicción.

Las predicciones obtenidas por medias móviles se pueden revisar fácilmente. Una vez calculada la predicción para el periodo $T+1$, se puede obtener la del periodo $T+2$

cuando llegue la nueva observación Y_{T+1} como sigue:

$$\hat{Y}_{T+2} = \frac{Y_{T+1}}{s} - \frac{Y_{T-s}}{s} + \hat{Y}_{T+1}$$

Escrito de esta forma podemos ver como cada nueva predicción basada en medias móviles es simplemente un ajuste de la anterior predicción.

3.2.2. Métodos de Alisado Exponencial

Al menos se pueden encontrar dos limitaciones importantes en el método de medias móviles, lo que ha llevado a los usuarios de técnicas de predicción a aplicar los métodos del alisado exponencial en su lugar:

a) Para calcular una predicción es necesario almacenar s valores observados

b) El método de las medias móviles da el mismo peso a cada una de las s observaciones y ninguno en absoluto al resto de los $T - s$ datos

Un argumento de bastante peso específico es que ya que las observaciones más recientes contienen la información más actual sobre lo que va a suceder en el futuro, deberían recibir más peso en la función de predicción que las observaciones más antiguas. Por lo tanto, sería preferible un esquema de ponderaciones que aplicar más peso a las observaciones más recientes y ponderaciones decrecientes a los valores más viejos. Los métodos de alisamiento exponencial se basan en esta idea y además eliminan la necesidad de almacenar datos de la variable.

La noción de descontar observaciones pasadas da idea de tendencia local en vez de tendencia global, es decir, de una tendencia que puede cambiar de dirección durante la muestra y que no es una función determinista del tiempo que se "ha de cumplirse" en todos los puntos a lo largo de la muestra, como son las tendencias globales ajustadas por Mínimos Cuadrados Ordinarios. Las técnicas de alisado exponencial proporcionan funciones de predicción basadas en el descuento de observaciones pasadas. Estos métodos tienen la ventaja de que permiten que la función de predicción se actualice muy fácilmente al llegar una nueva observación.

Los métodos de alisado exponencial son métodos de predicción que extrapolan los patrones históricos de los datos tales como la tendencia y la estacionalidad en el futuro. Las predicciones se computan promediando los datos para aislar los patrones de comportamiento de la fluctuación puramente aleatoria. Estos métodos son bastante populares, porque son fáciles de utilizar y bastante efectivos. Sin embargo, son métodos ad hoc y no se basan en ningún modelo estadístico propiamente definido.

AE1: Alisado Exponencial Simple

El método de alisado exponencial simple contempla a la serie temporal como compuesta de un nivel y un irregular de media cero que no se puede predecir:

$$Y_t = T_t + I_t = m + I_t, \quad t = 1, 2, \ldots, T$$

El único elemento predecible de la serie es el nivel, m, y, por lo tanto, ésta va a ser la base de las predicciones. El nivel de la serie se supone constante al menos localmente ya que se le permite variar lentamente en el tiempo.

¿Cómo se puede estimar el nivel de la serie con toda la información de la muestra? Si los datos fueran una muestra aleatoria, lo obvio sería tomar la media muestral:

$$\hat{m} = \bar{Y} = \sum_{t=1}^{T} \frac{1}{T} \, Y_t = \sum_{j=0}^{T-1} \frac{1}{T} \, Y_{T-j}$$

Sin embargo, si el objetivo al estimar el nivel, es usarlo como base para predecir futuras observaciones, resulta más razonable poner más peso en las observaciones más recientes. Así, la estimación del nivel de la serie en el momento T, m_T, es decir, con todo el conjunto de información, se podría obtener mediante una media móvil ponderada:

$$m_T = \sum_{j=0}^{T-1} w_j \, Y_{T-j} \qquad \sum_{j=0}^{T-1} w_j = 1$$

Una manera de dar más peso en las observaciones más recientes, es especificar los pesos w_j de forma que decrezcan exponencialmente:

$$m_T = \alpha \sum_{j=0}^{T-1} (1-\alpha)^j \, Y_{T-j} \tag{3.5}$$

La media móvil (3.5) es una media móvil con ponderaciones decrecientes en forma de progresión geométrica, donde $0 < \alpha \leq 1$ es la constante de alisamiento. Si el tamaño muestral T es suficientemente grande:

$$\lim_{T\to\infty} \sum_{j=0}^{T-1} w_j = \lim_{T\to\infty} \alpha \sum_{j=0}^{T-1} (1-\alpha)^j = \frac{\alpha}{1-(1-\alpha)} = 1$$

por lo que la condición de que los pesos sumen la unidad se satisface aproximadamente.

Esta estimación del nivel m_T, que es la estimación de la tendencia de la serie obtenida en base a todos los datos pero descontando las observaciones pasadas, es la predicción para las futuras observaciones. La *función de predicción* viene dada por:

$$Y_T(\ell) = m_T \qquad \ell = 1, 2, \ldots \tag{3.6}$$

por lo tanto, la predicción en los métodos de alisado puede interpretarse como una media ponderada de los valores pasados y presentes. Es interesante observar que la función de predicción es una línea recta horizontal.

Aplicando la expresión (3.5), se puede estimar el nivel de la serie, m_t, para cualquier valor de $t = 1, 2, \ldots, T$. Si se divide esta formula en dos partes, se obtiene lo siguiente:

$$
\begin{aligned}
m_t &= \alpha \sum_{j=0}^{t-1} (1-\alpha)^j \, Y_{t-j} \qquad\qquad \forall t = 1, 2, \ldots, T \\
m_t &= \alpha \, Y_t \, + \, \alpha \sum_{j=1}^{t-1} (1-\alpha)^j \, Y_{t-j} = \alpha \, Y_t \, + \, \alpha(1-\alpha) \sum_{j=0}^{t-2} (1-\alpha)^j \, Y_{t-1-j}
\end{aligned}
$$

$$
m_t = \alpha \, Y_t \, + \, (1-\alpha) \, m_{t-1} \qquad\qquad t = 1, 2, \ldots, T \tag{3.7}
$$

De donde se deduce que para estimar el nivel o tendencia de la serie en el momento t, m_t, no es preciso procesar todas las observaciones sino que basta con la última observación disponible, Y_t, y la estimación del nivel en el momento anterior, m_{t-1}, que lleva la información relevante sobre el nivel contenida en todas las observaciones anteriores.

Como (3.7) es una formula recursiva, se ha de contar con valores iniciales para poder comenzar, es decir, se ha de dar un valor a m_0. Si no se dispone de alguna información a priori sobre este valor se suele suponer que $m_0 = 0$. Alguna idea para un valor apropiado de m_0, se puede obtener también de los datos; por ejemplo, se pueden tomar las k primeras observaciones para estimar un valor inicial:

$$
m_0 \, = \, \bar{Y}_k \, = \, \frac{\sum_{i=1}^{k} Y_t}{k}
$$

y comenzar el proceso recursivo (3.7) a partir de la observación para $t = k + 1$. Otro enfoque satisfactorio es suponer que $m_1 = Y_1$ y comenzar la recursión en m_2:

$$
\begin{aligned}
m_2 &= \alpha \, Y_2 \, + \, (1-\alpha) \, m_1 \, = \, \alpha \, Y_2 \, + \, (1-\alpha) \, Y_1 \\
m_3 &= \alpha \, Y_3 \, + \, (1-\alpha) \, m_2 \, = \, \alpha \, Y_3 \, + \, \alpha(1-\alpha) \, Y_2 + (1-\alpha)^2 \, Y_1 \\
m_4 &= \alpha \, Y_4 \, + \, (1-\alpha) \, m_3 \, = \, \alpha \, Y_4 \, + \, \alpha(1-\alpha) \, Y_3 \, + \, (1-\alpha)^2 \, Y_2 \, + \, (1-\alpha)^3 \, Y_1 \\
\vdots \; & \; \vdots \; \; \vdots
\end{aligned}
$$

Estas recursiones producen un resultado algo diferente que la formula (3.5) porque Y_1 recibe un peso de $(1-\alpha)^{T-1}$ en vez de $\alpha(1-\alpha)^{T-1}$. Sin embargo, tiene la propiedad de que los pesos suman 1 incluso en muestras pequeñas:

$$
\alpha \quad \alpha(1-\alpha) \quad \alpha(1-\alpha)^2 \quad \alpha(1-\alpha)^3 \quad \ldots \quad (1-\alpha)^{T-1}
$$

$$S = \frac{\alpha - \alpha(1-\alpha)^{T-1}}{1-(1-\alpha)} + (1-\alpha)^{T-1} = 1$$

La elección de los valores iniciales es arbitraria, lo que es un problema porque es importante. Si la constante de alisamiento α es pequeña, el valor inicial fijado influirá en los resultados durante muchos periodos. La elección de $m_0 = \bar{Y}_k$, funciona bien si la media de la serie cambia poco. Sin embargo, es mejor elegir $m_1 = Y_1$ si la media cambia bastante.

Como según la función de predicción (3.6), m_t es la predicción apropiada para Y_{t+1}, la recursión dada por la expresión (3.7) se suele escribir como:

$$Y_t(1) = (1-\alpha)\, Y_{t-1}(1) + \alpha\, Y_t, \quad t = 1, 2, \ldots \tag{3.8}$$

Así, con información hasta t, la predicción para el periodo siguiente es una media móvil de la observación actual, Y_t, y de la predicción de dicha observación hecha en el periodo anterior, $Y_{t-1}(1)$.

Esta recursión se conoce con el nombre de *medias móviles exponencialmente ponderadas* (Exponential Weigthed Moving Average), y el método de construir la función de predicción como *Alisamiento Exponencial Simple* (Simple Exponential Smoothing).

De forma alternativa, la predicción (3.8) se puede escribir

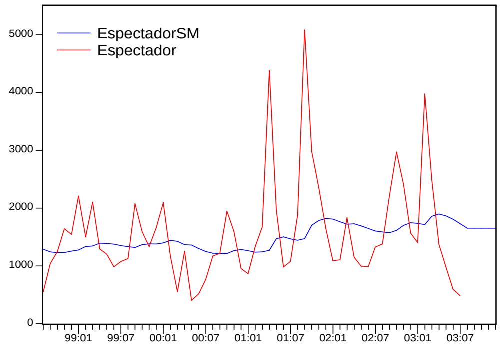
$$\begin{aligned} Y_t(1) &= Y_{t-1}(1) + \alpha\ (Y_t - Y_{t-1}(1)) \\ Y_t(1) &= Y_{t-1}(1) + \alpha\ e_{t-1}(1) \end{aligned} \tag{3.9}$$

lo que indica que la actualización de las predicciones una vez que llega nueva información se realiza de acuerdo con el error de predicción que hayamos cometido el periodo precedente. Cada vez que observamos un nuevo dato de la serie, calculamos el error de predicción. Si el error es positivo (la última predicción fue muy baja), incrementamos la predicción. Si el error es negativo (la última predicción fue muy alta), disminuimos la predicción. De esta forma, los errores de predicción cometidos nos ayudan a dirigir nuestras predicciones hacia su objetivo: el verdadero nivel de la serie. La recursión (3.9) explica también por que las predicciones para $\ell > 1$ serán sucesivamente iguales: al no disponer de los correspondientes errores de predicción los suponemos nulos. La función de predicción es, por lo tanto, una línea horizontal.

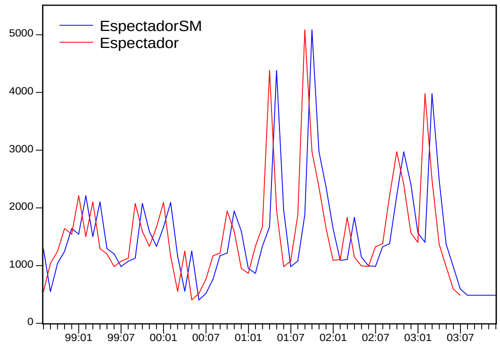
La actualización de la predicción depende también de α. Si $\alpha \cong 1$, la nueva predicción incluirá un ajuste muy fuerte en base al error cometido en la anterior predicción. Si $\alpha \cong 0$, la nueva predicción no sufrirá mucho ajuste. Hay que tener en cuenta que del valor de α depende hasta qué punto se descuentan las observaciones pasadas. Estos resultados se pueden observar en las figuras del gráfico 3.8. En la figura superior se ha utilizado un valor de α pequeño y la serie se ha alisado mucho. En la figura inferior del gráfico se ha utiliza $\alpha = 1$ por lo que la predicción viene dada por la última observación.

Gráfico 3.8: Espectadores de cine español: Alisado Exponencial Simple.

Sample: 1998:08 2003:07	
Included observations: 60	
Method: Single Exponential	
Original Series: Espectador	
Forecast Series: EspectadorSM	
Parameters: Alpha	0.0640
Sum of Squared Residuals	47925518
Root Mean Squared Error	893.7330
End of Period Levels: Mean	1650.469



Sample: 1998:08 2003:07	
Included observations: 60	
Method: Single Exponential	
Original Series: Espectador	
Forecast Series: EspectadorSM	
Parameters: Alpha	1.0000
Sum of Squared Residuals	53143107
Root Mean Squared Error	941.1262
End of Period Levels: Mean	485.8100



La elección de α debe acomodarse, por lo tanto, a las características particulares de cada serie para lo que es interesante conocer previamente las implicaciones de tomar distintos valores del parámetro α. En general, se considera que un α alto es indicativo de fuertes oscilaciones o de la existencia de tendencia en la serie, lo que conlleva un reducido alisado para ajustarse mejor a estos cambios continuos. Por el contrario, una serie con pequeñas oscilaciones irregulares aconsejará un α bajo (habitualmente comprendido entre 0,01 y 0,3) que supone un fuerte alisado de la serie al considerar un elevado número de valores de la serie. Para comprobar estos comportamientos basta con analizar la aplicación del método de Alisado Exponencial Simple a la serie de *Espectadores de cine español* con el valor $\alpha = 0,064$ y compararla con lo obtenido para la serie de *Tasa de Actividad de la mujer* en el gráfico 3.9 con un valor de $\alpha 00,824$.

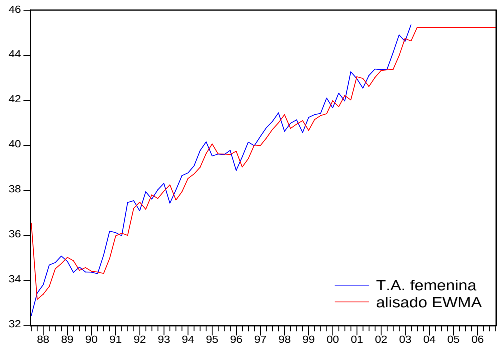
Los factores que interaccionan a la hora de determinar el mejor parámetro de alisado α son fundamentalmente dos:

- Cantidad de ruido en la serie. Cuanto más ruido tenga, menor será el parámetro, para evitar reacciones exageradas al ruido.

- Estabilidad de la media de la serie. Si la media es relativamente constante, el parámetro será pequeño. Si la media cambia, el parámetro será grande para ser capaz de

Gráfico 3.9: Tasa de actividad de la mujer: Alisado Exponencial Simple.

Sample: 1987:3 2003:2	
Included observations: 64	
Method: Single Exponential	
Original Series: TAmujer	
Forecast Series: TAmujersm	
Parameters: Alpha	0.8240
Sum of Squared Residuals	35.46528
Root Mean Squared Error	0.744409
End of Period Levels: Mean	45.24373



adaptarse a estos cambios. En el límite, el parámetro llegará a la unidad, que significa que la nueva predicción es igual al último dato

Mirando las figuras del gráfico 3.8 se puede observar cómo la elección de α afecta la estabilidad de las predicciones. Con $\alpha = 0{,}064$, las predicciones son estables, que es lo que requiere esta serie temporal de *Espectadores de cine español* porque su media es relativamente constante, pero con $\alpha = 1$, las predicciones fluctúan exageradamente y los errores son mucho mayores.

Para poder aplicar el método de Alisado Exponencial Simple (AES) es preciso especificar un valor para la constante de alisamiento α por lo que conviene hacer algunas consideraciones sobre este tema. El parámetro α se puede establecer a priori, teniendo en cuenta las consideraciones anteriores. Como una regla general, la constante de alisamiento para un modelo con un nivel constante debería estar entre 0,01 y 0,4. Una técnica que se utiliza mucho es llevar a cabo una serie de experimentos con la serie utilizando diferentes valores de α y después seleccionar aquel valor de α que cumple algún criterio de efectividad. Un criterio que se suele utilizar es escoger aquella constante de alisamiento que hubiera predicho mejor los datos que tenemos, que se puede materializar en elegir la α que nos proporciona los mejores errores de predicción un periodo hacia adelante, es decir, que minimiza el Error Cuadrático Medio (ECM) de Predicción:

$$ECM = \sum_{t=m}^{T} (Y_t - Y_{t-1}(1))^2$$

siendo m lo suficientemente grande para que el efecto de la elección de los valores iniciales desaparezca.

Algunos autores defienden que si el proceso de estimación de α nos lleva a una constante de alisamiento mayor que 0,33, entonces se debe plantear la válidez del modelo: puede ser que la serie presente cambios de nivel sistemáticos o ciclos que deberían recogerse en un modelo más complicado. Sin embargo, no existe suficiente evidencia para mantener este supuesto. En bastantes trabajos empíricos se han estimado valores de $\alpha > 0{,}3$, lo que sugiere que debería considerarse un intervalo mayor para α. En conclusión, los estudios realizados hasta el momento demuestran que es peligroso hacer supuestos sobre los valores de la constante de alisamiento, que, en general, debe ser estimada a partir de los datos.

Para usar la técnica del Alisado Exponencial Simple, solo se necesita el valor observado más reciente, la última predicción y un valor para α. La utilización del AES es fácil y barata porque los programas pueden calcular el mejor α automáticamente. Además la experiencia práctica es que AES es un método preciso, efectivo y confiable.

El Alisado Exponencial Simple deja abiertas, sin embargo, cierto número de cuestiones:

a) La media muestral no se puede obtener como un caso especial, estableciendo $\alpha = 0$.

b) No parece que haya una solución ideal para encontrar valores iniciales de las recursiones AES. Ambas cuestiones se solucionan proporcionando un marco estadístico al problema.

c) Este procedimiento, como el de medias móviles, sólo se debe aplicar cuando los patrones de comportamiento históricos de los datos se puedan considerar estables. Sin embargo, esta técnica no será efectiva, en general, para manejar tendencias o patrones estacionales (veáse el gráfico 3.9).

Las predicciones del AES se realizan sin hacer referencia a ningún modelo en particular. Ahora bien, este método presenta varias ventajas:

- Ecuaciones de actualización: hace que sea muy fácil computar nuevas predicciones.
- Es un método completamente automático una vez que se fije α. Esto es también una desventaja ya que todas las series se tratan de la misma manera.

AE2: Método de Alisado Exponencial con Tendencia

El método del Alisado Exponencial Simple es apropiado teóricamente cuando las series contienen un patrón horizontal (es decir, no hay pendiente en la tendencia). Si el AES se utiliza en series que presentan una pendiente persistente, las predicciones siempre van a ir un paso por detrás de esta tendencia. El modelo de Alisado Exponencial con tendencia (AET) evita este problema reconociendo de forma explícita la presencia de esta tendencia e incluyéndola en la función de predicción. Este método se denomina también Alisamiento exponencial lineal con doble parámetro o Alisado Exponencial Holt-Winters.

La función de predicción del AES era una línea recta. Ahora bien para series con tendencia aproximadamente lineal parece interesante introducir de alguna forma un componente de pendiente en la función de predicción, para eliminar el sesgo en la predicción:

$$Y_T(\ell) = m_T + b_T\, \ell \qquad\qquad \ell = 1, 2, \ldots$$

Un modelo de alisamiento para una tendencia lineal es una extensión de un modelo de regresión con el tiempo como variable explicativa. La ecuación de regresión sería:

$$Y_t = m + b\, t$$

donde m es el intercepto y b es la pendiente. En el modelo de regresión, tanto m como b se promedian con todos los datos y se tratan como constantes. En un método de alisamiento exponencial la diferencia está en que m y b no son constantes sino que se actualizan cada periodo dando mayor peso a los datos más recientes.

Sean m_{t-1} y b_{t-1}, las estimaciones del nivel y la pendiente en el momento $t-1$[4]. La predicción un periodo hacia adelante vendría dada por:

$$Y_{t-1}(1) = m_{t-1} + b_{t-1}$$

Una vez que se cuenta con la nueva observación, Y_t, al igual que en el método del AES, se procede a la actualización del nivel promedio y de la pendiente con información hasta t, m_t y b_t. Holt(1957) y Winters(1960) introdujeron un esquema para actualizar m_t y b_t en el que las observaciones pasadas se descuentan por medio de dos constantes de alisamiento α y β, tal que $0 < \alpha\ ,\ \beta < 1$.

La estimación actualizada del nivel m_t es, como en el AES, una combinación de la predicción anterior, $Y_{t-1}(1)$, y la nueva observación, Y_t:

$$m_t = \alpha\, Y_t + (1-\alpha)\, Y_t(1) = \alpha\, Y_t + (1-\alpha)\, [m_{t-1} + b_{t-1}]$$

Con esta estimación actualizada del nivel, podemos construir una estimación actualizada de la pendiente como diferencia entre niveles, $m_t - m_{t-1}$. Esto sugiere que la estimación actualizada de la pendiente, b_t, se puede obtener como combinación de $m_t - m_{t-1}$ y la estimación anterior b_{t-1}:

$$b_t = \beta\, (m_t - m_{t-1}) + (1-\beta)\, b_{t-1}$$

Las denominadas recursiones de Holt-Winters son, por lo tanto:

$$\begin{aligned} Y_t(1) &= m_t + b_t \quad t = 1, 2, \ldots \\ m_t &= \alpha\, Y_t + (1-\alpha)\, [m_{t-1} + b_{t-1}] \\ b_t &= \beta\, (m_t - m_{t-1}) + (1-\beta)\, b_{t-1} \end{aligned}$$

La idea que hay bajo estas recursiones es la siguiente: en el AES se promedian los datos, “se alisan” para eliminar la aleatoriedad, en el AET, se promedia para alisar la aleatoriedad

[4]Se puede demostrar que ambas se pueden expresar en forma de medias móviles exponencialmente ponderadas, ver Pulido (1989) pag. 93.

como en el AES y además se corrige este alisamiento para poder ajustarse a la tendencia existente en la serie.

Estas recursiones pueden ser manipuladas de manera que se obtenga lo siguiente:

$$
\begin{aligned}
m_t &= m_{t-1} + b_{t-1} + \alpha\, e_t(1) \\
b_t &= b_{t-1} + \alpha\, \beta\, e_t(1)
\end{aligned}
$$

que nos indica que la actualización del nivel y de la pendiente dependen del error de predicción y de las constantes de alisamiento. Cuanto más se acerca α a cero, menos descontamos las observaciones pasadas al formar la estimación actual del nivel. De la misma manera, cuanto más cerca esté β de cero, menos se descuenta al estimar la pendiente.

Al igual que para el método de AES, los valores iniciales para ambas recursiones se pueden construir a través de las observaciones iniciales. Existen varias alternativas:

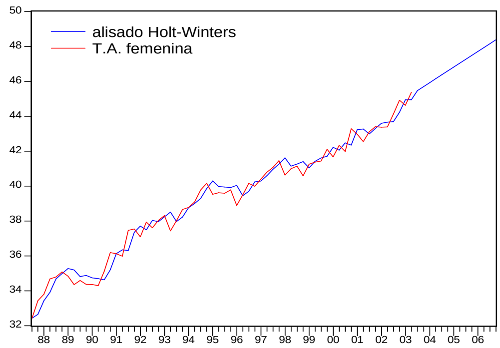
a) $m_2 = Y_2$ $\qquad b_2 = Y_2 - Y_1$

b) $m_1 = Y_1$ $\qquad b_1 = 0$

c) Obtener un punto inicial para el nivel y la tendencia de la serie, realizando una regresión con el tiempo como variable explicativa.

Gráfico 3.10: Tasa de actividad femenina. Alisado Exponencial Simple.

Sample: 1987:3 2003:2		
Included observations: 64		
Method: Holt-Winters No Seasonal		
Original Series: TAmujer		
Forecast Series: TAmujersm		
Parameters:	Alpha	0.7100
	Beta	0.0000
Sum of Squared Residuals		15.69614
Root Mean Squared Error		0.495230
End of Period Levels:	Mean	45.24658
	Trend	0.223750



En cuanto a las constantes de alisamiento, se pueden fijar a priori o estimar minimizando el Error Cuadrático Medio de predicción. En general, se suele obtener como resultado que el mejor β es menor que el mejor α. La razón para este resultado es que la cantidad de pendiente en cada periodo es muy pequeña en comparación con la cantidad de nivel.

El alisamiento exponencial de Holt-Winters se puede emplear en series con tendencia pero no con estacionalidad (veáse el gráfico 3.10). Una solución operativa es partir de la serie desestacionalizada y afectar posteriormente las predicciones por los factores de estacionalidad calculados.

Métodos con tendencia no lineal. El modelo con tendencia lineal se puede modificar para acomodar tendencias no lineales. La idea se basa en añadir un nuevo parámetro que controle la tasa de crecimiento de las predicciones, ϕ:

$$\begin{aligned} Y_t(1) &= m_t + \phi\, b_t \\ m_t &= \alpha\, Y_t + (1-\alpha)\,[m_{t-1} + \phi\, b_{t-1}] \\ b_t &= \beta\,(m_t - m_{t-1}) + (1-\beta)\,\phi\, b_{t-1} \end{aligned}$$

De la misma forma, estas recursiones se pueden manipular de forma que:

$$\begin{aligned} m_t &= m_{t-1} + \phi\, b_{t-1} + \alpha\, e_t(1) \\ b_t &= \phi\, b_{t-1} + \alpha\,\beta\, e_t(1) \end{aligned}$$

y la función de predicción sería de la forma:

$$Y_t(\ell) = m_t + \sum_{i=1}^{\ell} \phi^i\, b_t \qquad\qquad \ell = 1, 2, \ldots$$

Como se puede observar la única diferencia con el modelo de tendencia lineal de Holt-Winters es que la pendiente va multiplicada por el parámetro ϕ:

a) Si $\phi > 1$, se tendría una tendencia exponencial, es decir, la tasa de crecimiento de las predicciones sería mayor cada periodo.

b) Si $\phi < 1$, la tendencia sería amortiguada con una tasa de crecimiento de las predicciones menor cada periodo.

Estos métodos con tendencia no lineal son más generales que los métodos con tendencia lineal, a los que incluyen como caso particular:

a) Si $\phi = 1$: Método con tendencia local lineal

b) Si $\phi = 0$: Método con nivel constante

3.3. Diferenciación

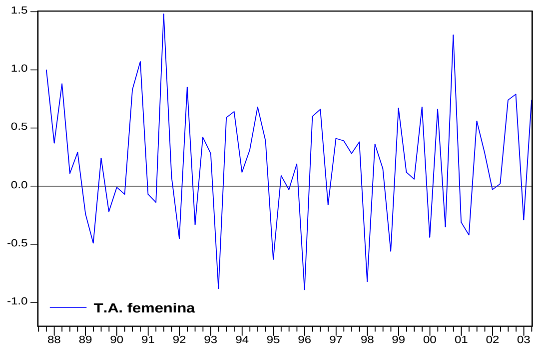
El objetivo es estudiar como se puede eliminar la tendencia de una serie de forma que se pueda predecir lo que queda por diferentes métodos. Una clase de filtros que es útil para eliminar la tendencia, es simplemente la diferenciación de los datos originales de las series. Para series no estacionales, generalmente basta con tomar la primera diferencia. Así la serie Y_t se convierte en la serie diferenciada W_t:

$$W_t \;=\; \Delta\, Y_t \;=\; (1-L)\, Y_t \;=\; Y_t - Y_{t-1}$$

donde L es el operador de retardos (veáse apéndice C).

El gráfico 3.11 muestra como la tendencia creciente de la serie *Tasa de Actividad femenina* desaparece con la primera diferencia.

Gráfico 3.11: Tasa de actividad femenina. Diferencia de orden 1.



De forma que si contamos con algún método para predecir W_t, podemos obtener predicciones para Y_t teniendo en cuenta que:

$$Y_{T+\ell} \;=\; W_{T+\ell} + Y_{T+\ell-1}$$

por lo que las predicciones serán:

$$Y_T(\ell) \;=\; W_T(\ell) \;+\; Y_T(\ell-1) \quad \ell = 1, 2, \ldots$$

Supongamos que Y_t consta de una tendencia lineal y término estocástico:

$$Y_t \;=\; a + b\; t + \varepsilon_t$$

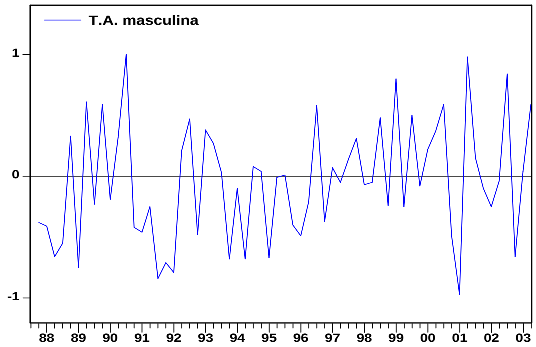
se tiene que tomando diferencias de orden 1:

$$W_t \;=\; Y_t - Y_{t-1} \;=\; b + \varepsilon_t - \varepsilon_{t-1}$$

por lo que mientras que Y_t contiene una tendencia, su diferencia, W_t, no la tiene. De esta forma, además, se puede comprobar cómo la idea de diferenciar está íntimamente relacionada con la de eliminar una tendencia de tipo polinomial.

Si la tendencia no es lineal, entonces diferenciar la serie una vez no conseguirá eliminar totalmente la tendencia. Existen diferentes métodos para solucionar este problema. Algunas tendencias se llaman localmente lineales, y se pueden aproximar adecuadamente mediante

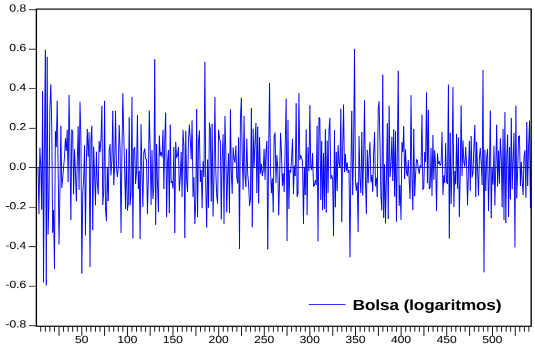
Gráfico 3.12: Tasa de actividad del hombre. Diferencia de orden 1.



líneas rectas a trozos. En este caso podremos eliminar la tendencia diferenciando (véase el gráfico 3.12), pero predecir va a ser muy complicado.

Por otro lado, si la tendencia es exponencial, tomando logaritmos a los datos se volvería lineal y podríamos diferenciar, como se ha realizado en el gráfico 3.13 para la serie *Volumen de acciones de la bolsa de New York*.

Gráfico 3.13: Acciones de la Bolsa de New York. Diferencia de orden 1.



Una de las transformaciones más frecuentes que se suele hacer a las variables económicas es tomar las primeras diferencias de los logaritmos:

$$\Delta \log Y_t = \log Y_t - \log Y_{t-1} = \log \frac{Y_t}{Y_{t-1}} \simeq \frac{Y_t}{Y_{t-1}} - 1 = \frac{Y_t - Y_{t-1}}{Y_{t-1}}$$

Por lo tanto, tomar las primeras diferencias de logaritmos es equivalente a trabajar con tasas de crecimiento.

Por último, si la serie y_t es una función determinista cuadrática del tiempo de la forma:

$$Y_t = a + b\,t + c\,t^2$$

Diferenciando dos veces se elimina la tendencia:

$$W_t = (1-L)^2\,Y_t = \Delta^2\,Y_t = \Delta\,Y_t - \Delta\,Y_{t-1} = b + c + 2tc - (b + c + 2(t-1)c) = 2c$$

Capítulo 4

Análisis de una serie con estacionalidad

La estacionalidad es un movimiento cíclico de periodo un año que se repite sistemáticamente. Si la serie temporal presenta comportamiento estacional, este puede relacionarse con la tendencia de distintas maneras:

a) Si la amplitud del componente estacional es estable a lo largo de la serie, entonces la variación estacional se sumaría a la tendencia en un modelo aditivo

$$Y_t = T_t + S_t + I_t \quad t = 1, 2, \ldots, T \tag{4.1}$$

La estacionalidad cumple la condición de que su media es cero a lo largo del año.

b) Si la amplitud del componente estacional va creciendo según crece la tendencia, la estacionalidad se relacionaría de forma multiplicativa con la tendencia:

$$Y_t = T_t \; S_t \; I_t \quad t = 1, 2, \ldots, T \tag{4.2}$$

En este caso, la media anual de los factores o índices estacionales S_t ha de ser uno.

Se suele trabajar bajo el supuesto de que los efectos estacionales son prácticamente invariantes en el tiempo, es decir, $S_t \simeq S_{t-s}$, donde s es el periodo estacional, a saber, $s = 4$ para series trimestrales y $s = 12$ para series mensuales.

4.1. Modelos globales. Variables ficticias estacionales.

Supongamos que el modelo apropiado para la serie temporal Y_t es el siguiente:

$$Y_t = T_t + S_t + I_t \quad t = 1, 2, \ldots, T$$

En el capítulo 3 se explicó como, dentro del marco de los modelos globales, la tendencia se puede modelar mediante funciones del tiempo: polinómicas, exponenciales, etc. En lo que se refiere al componente estacional para modelar el comportamiento diferenciado de cada mes respecto de la tendencia se pueden utilizar variables ficticias del tipo:

$$D_{jt} = \begin{cases} 1 & \text{si t = j, 2j, } \dots \\ 0 & \text{en otro caso} \end{cases}$$

de forma que el componente estacional se especificaría como sigue:

$$S_t = \sum_{j=1}^{s} s_j \, D_{jt} \qquad s.a. \quad \sum_{j=1}^{s} s_j = 0$$

donde s_j es el factor estacional correspondiente al mes j-ésimo. De esta forma, para cada momento de tiempo t, la función del tiempo, S_t, asigna a cada observación de la serie temporal el factor estacional correspondiente a su mes.[1]

Suponiendo, por ejemplo, que la serie es mensual y que la tendencia sigue una función

[1]Dentro del marco de los modelos globales, el componente estacional también se puede representar como una combinación de ondas seno-coseno a las frecuencias estacionales. Un modelo simple que puede representar el comportamiento cíclico estacional determinista es:

$$f(t) = A\cos(\lambda t - \theta)$$

donde A es la amplitud del ciclo; θ es la fase; λ es la frecuencia y $2\pi/\lambda$ es el periodo del ciclo.

Aplicando resultados trigonométricos conocidos, lo podemos escribir como una combinación lineal de funciones seno y coseno:

$$f(t) = \alpha\cos\lambda t + \beta\sin\lambda t$$

donde $\alpha = A\cos\theta$ y $\beta = A\sin\theta$.

Sin embargo, no existe ninguna razón para esperar que un componente cíclico, como un ciclo anual, por ejemplo, sea representable por una sola onda coseno.

Una función periódica $f(t)$ de periodo p, cumple que $f(t + kp) = f(t)\forall k$.

Se puede demostrar que toda función periódica se puede representar por la suma :

$$f(t) = \sum_{k=0}^{[p/2]} (a_k \cos\lambda kt + b_k \sin\lambda kt)$$

donde $\lambda = 2\pi/p$, dado que p es un múltiplo del periodo muestral.

Esta no es más que la *representación de Fourier* de una serie periódica. La frecuencia λ es conocida como la frecuencia fundamental, y las frecuencias λk son los armónicos de λ.

En el caso del componente estacional, la frecuencia fundamental es $\lambda = 2\pi/s$, y sus armónicos son $2\pi j/s,\ j = 1, \dots, s/2$:

$$S_t = \sum_{j=1}^{[s/2]} \left[\beta_j \sin\frac{2\pi j}{s} + \gamma_j \cos\frac{2\pi j}{s}\right]$$

Para una serie dada $Y_1, \dots, Y_T$, los estimadores de los parámetros β_j, γ_j , se obtienen por los métodos estándar de regresión MCO.

lineal, el modelo de componentes no observados global para la serie Y_t es el siguiente:

$$Y_t = a + b\,t + \sum_{j=1}^{12} s_j\ D_{jt} + I_t \qquad s.a. \quad \sum_{j=1}^{12} s_j = 0 \tag{4.3}$$

Este es un modelo de regresión lineal sujeto a una restricción, por lo que sus parámetros, $(a,\, b)$ para la tendencia y $s_j, j = 1, 2, \ldots, 12$ para la estacionalidad, se pueden estimar por Mínimos Cuadrados Restringidos. Este método se basa en minimizar la suma de los errores al cuadrado teniendo en cuenta que los parámetros han de cumplir la restricción impuesta.

Una forma de estimar los parámetros por Mínimos Cuadrados Restringidos, se reduce a aplicar Mínimos Cuadrados Ordinarios al modelo en el que se ha incluido la restricción. En primer lugar, se reescribe la restricción de forma operativa, es decir,

$$s_1 = -s_2 - s_3 - \ldots - s_{12} \tag{4.4}$$

y se incluye en el modelo (4.3) obteniéndose:

$$Y_t = a + b\,t + (-s_2 - s_3 - \ldots - s_{12})\, D_{1t} + s_2\, D_{2t} + s_3\, D_{3t} + \ldots + s_{12}\, D_{12,t} + I_t$$

Y, haciendo factor común a los factores estacionales queda:

$$Y_t = a + b\,t\ +\ s_2\,(D_{2t} - D_{1t}) +\ s_3\,(D_{3t} - D_{1t}) + \ldots +\ s_{12}\,(D_{12,t} - D_{1t}) + I_t \tag{4.5}$$

Estimando los parámetros del modelo (4.5) por MCO, obtenemos la tendencia estimada:

$$\hat{T}_t = \hat{a} + \hat{b}\,t, \quad t = 1, 2, \ldots, T$$

y la estimación de 11 factores estacionales:

$$\hat{s}_2 \quad \hat{s}_3 \quad \hat{s}_4 \quad \ldots \quad \hat{s}_{12}$$

El factor estacional del primer mes lo estimamos a partir de la restricción (4.4):

$$\hat{s}_1 = -\hat{s}_2 - \hat{s}_3 - \ldots - \hat{s}_{12}$$

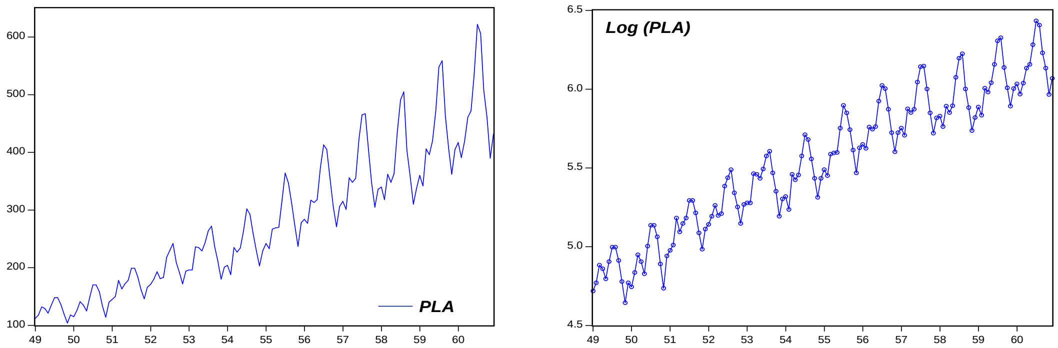
Por lo tanto, se han estimado los dos patrones de comportamiento sistemático de la serie, la tendencia y la estacionalidad.

Para predecir las futuras observaciones, $Y_{T+\ell}, \ell = 1, 2, \ldots$, la función de predicción se basaría en la proyección de la tendencia estimada a la que se suma el componente estacional estimado correspondiente:

$$\begin{aligned} Y_T(\ell) &= T_T(\ell) + S_T(\ell) \;=\; \hat{T}_{T+\ell} + \hat{S}_{T+\ell} \\ Y_T(\ell) &= \hat{a} + \hat{b}\,(T + \ell) + \hat{s}_j \quad \ell = 1, 2, \ldots \end{aligned}$$

Ejemplo 4.1. Si el modelo apropiado para la serie fuera el modelo multiplicativo, ¿cómo se especificaría el modelo global con tendencia y estacionalidad?

Gráfico 4.1: Pasajeros de líneas aéreas



En la figura izquierda del gráfico 4.1 está representada la serie de *Pasajeros de Líneas Aéreas* de enero de 1949 a diciembre de 1960. Esta serie muestra una tendencia creciente y un componente estacional cuya amplitud crece conforme aumenta el nivel de la tendencia, por lo que el modelo de Componentes No Observados apropiado sería el multiplicativo (4.2). Para aplicar los métodos globales de estimación de tendencias y estacionalidades es más sencillo trabajar con el modelo aditivo, por lo que se toman logaritmos a la serie, de forma que:

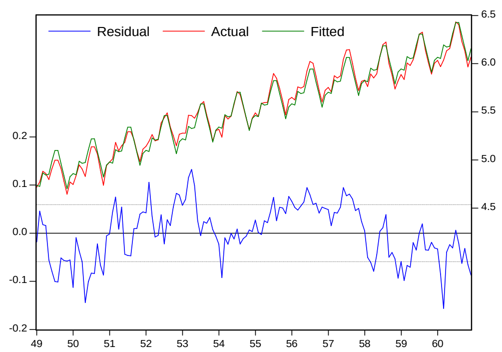
$$\log Y_t = T_t + S_t + I_t \quad t = 1, 2, \ldots, T$$

Gráfico 4.2: Pasajeros de líneas aéreas. Ajuste global

Dependent Variable: LOG(pasajero)
Method: Least Squares
Sample: 1949:01 1960:12
Included observations: 144

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.812188	0.009962	483.0757	0.0000
TIEMPO	0.010069	0.000119	84.39895	0.0000
FEB-ENE	-0.107462	0.016399	-6.552794	0.0000
MAR-ENE	0.022765	0.016396	1.388455	0.1674
ABR-ENE	-0.008504	0.016393	-0.518735	0.6048
MAY-ENE	-0.010876	0.016392	-0.663537	0.5082
JUN-ENE	0.111270	0.016391	6.788572	0.0000
JUL-ENE	0.215212	0.016391	13.13010	0.0000
AGO-ENE	0.205917	0.016392	12.56235	0.0000
SEP-ENE	0.061283	0.016393	3.738262	0.0003
OCT-ENE	-0.076876	0.016396	-4.688693	0.0000
NOV-ENE	-0.220593	0.016399	-13.45128	0.0000
DIC-ENE	-0.106728	0.016404	-6.506330	0.0000

R-squared	0.983468	Mean dependent var	5.542176
Adjusted R-squared	0.981954	S.D. dependent var	0.441456
S.E. of regression	0.059304	Akaike info criterion	-2.726355
Sum squared resid	0.460715	Schwarz criterion	-2.458247
Log likelihood	209.2976	F-statistic	649.4254
Durbin-Watson stat	0.425184	Prob(F-statistic)	0.000000



Se estima, por lo tanto, el modelo (4.3) para la serie de Pasajeros en logaritmos con los resultados recogidos en la tabla del gráfico 4.2:

$$\begin{aligned}\hat{Y}_t &= 4,81 + 0,01\,t - 0,083\,D_{1t} - 0,107\,D_{2t} + 0,022\,D_{3t} - 0,009\,D_{4t} - 0,011\,D_{5t} + \\ &+ 0,11\,D_{6t} + 0,215\,D_{7t} + 0,206\,D_{8t} + 0,06\,D_{9t} - 0,077\,D_{10,t} - 0,22\,D_{11,t} - 0,107\,D_{12,t}\end{aligned}$$

La figura derecha del gráfico 4.2 recoge el ajuste del modelo a la muestra y los residuos.

Gráfico 4.3: Pasajeros de líneas aéreas. Componentes.

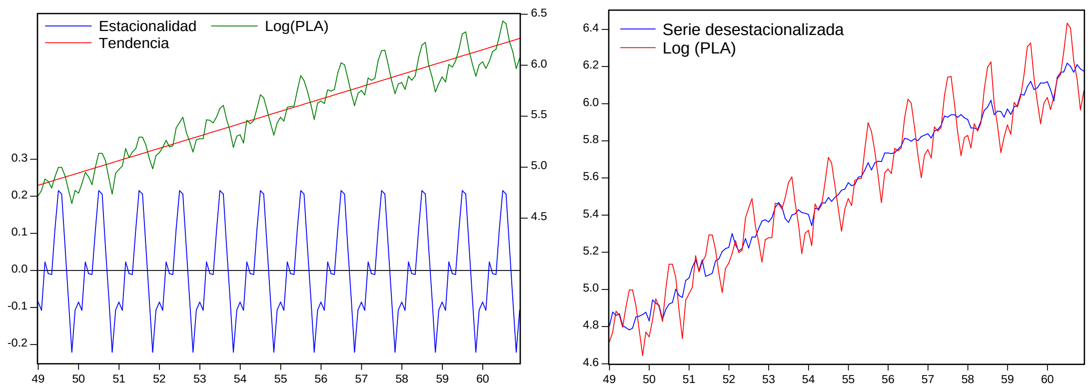
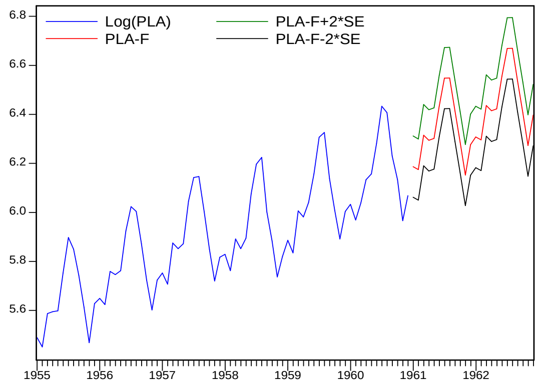


Gráfico 4.4: Pasajeros de líneas aéreas. Predicción



Como se han estimado la tendencia y la estacionalidad es posible, por un lado, descomponer la serie en estos componentes y, por otro, si así interesara calcular la serie desestacionalizada restando a la serie en logaritmos el componente estacional estimado (veánse las figuras del gráfico 4.3). Como se puede observar comparando ambas figuras, la serie desestacionalizada es mucho más ruidosa que la serie de tendencia. Esto no se debe solamente a que hayamos impuesto un modelo muy suave para la tendencia, sino que hay que recordar que la serie desestacionalizada es la serie de la que se ha eliminado el componente estacional, pero aún sigue presente no sólo la tendencia sino también el componente irregular.

Por último, el gráfico 4.4 muestra la predicción para un periodo de dos años tanto por punto como por intervalo.

4.2. Métodos de alisado

4.2.1. Método de Relación a la Media Móvil

Una técnica muy utilizada para *desestacionalizar* las series temporales ha sido la derivada en el Bureau de Census de los Estados Unidos siguiendo las propuestas de Shiskin de finales de los años 50 y que se concretó en 1965 en un método muy elaborado, aún ampliamente utilizado a nivel internacional, el Filtro X-11. Partiendo de este método se han realizado muy diversas adaptaciones y variantes, la última de las cuales ha resultado en le Filtro X-12 ARIMA de amplio uso por los analistas de coyuntura de todo el mundo. Estos métodos de desestacionalización son mecánicos, en el sentido de que tratan a todas las series por igual independientemente de sus características específicas. Además estos métodos son de naturaleza heurística surgiendo de la práctica y no de ningún concepto teórico ni modelo estocástico.

La idea original del método de la relación a la media móvil se puede resumir en las tres fases siguientes, tomando como base para la serie temporal el modelo de tendencia y estacionalidad multiplicativo (4.2):

a) Estimación de la tendencia mediante medias móviles. Se trata, en realidad, de estimar una tendencia local que puede recoger, al menos parcialmente, ciertos movimientos cíclicos. Con datos mensuales, por ejemplo, se sugiere una media móvil de orden 12 como valor de la tendencia para cada periodo:

$$\frac{Y_{t-6} + \ldots + Y_{t-1} + Y_t + Y_{t+1} + \ldots + Y_{t+5}}{12}$$

Dado que la media móvil de orden par no se puede centrar directamente, se promedia con la media siguiente:

$$\frac{Y_{t-5} + \ldots + Y_{t-1} + Y_t + Y_{t+1} + \ldots + Y_{t+6}}{12}$$

para conseguir tal efecto, de forma que la tendencia estimada viene dada por:

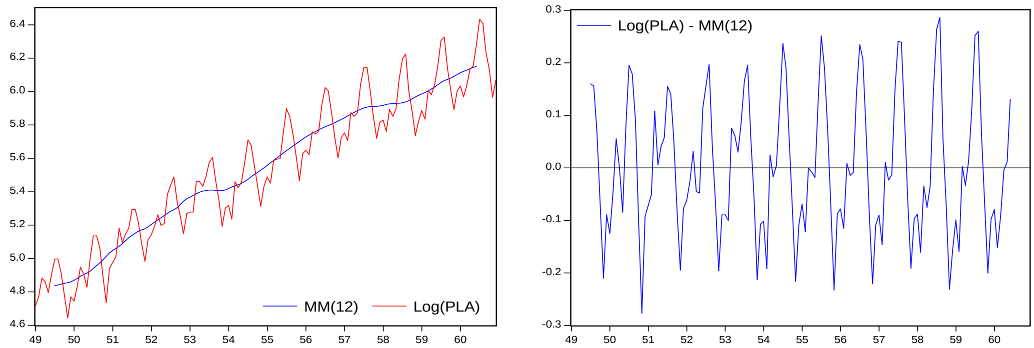
$$M_t^{12} = \frac{1}{2}\left[\frac{Y_{t-6} + 2\,Y_{t-5} + \ldots + 2Y_t + \cdots + 2\,Y_{t+5} + Y_{t+6}}{12}\right] \qquad t = 7, 8, \ldots, T-6$$

b) Cálculo de los índices de los datos originales respecto a las medias móviles centradas correspondientes:

$$\frac{Y_t}{M_t^{12}} \qquad t = 7, 8, \ldots, T-6$$

En las figuras del gráfico 4.5 se representan para la serie *Pasajeros de líneas aéreas*, por un lado, la tendencia estimada por medias móviles de orden 12 centradas y, por otro, el resultado de eliminar de la serie la tendencia.

Gráfico 4.5: Pasajeros de líneas aéreas. Componentes estimados.



c) Obtención de la media de estos índices para cada mes (de todos los eneros, todos los febreros, etc.) que se tomarán como coeficientes de estacionalidad del esquema multiplicativo:

$$s'_i = \frac{1}{n}\left[\frac{Y_{1i}}{M_{1i}^{12}} + \frac{Y_{2i}}{M_{2i}^{12}} + \ldots + \frac{Y_{ni}}{M_{ni}^{12}}\right] \quad \text{donde } t = ji \begin{cases} \text{j} = 1,\ 2,\ \ldots,\ \text{n} \\ \text{i} = 1,\ 2,\ \ldots,\ 12 \end{cases}$$

en el que el primer subíndice de Y o M, j, hace referencia al año (n en total) y el segundo i, al mes de referencia (12 en total).

Los factores estacionales han de valer 1 en promedio:

$$\frac{1}{12}\sum_{i=1}^{12} s_i = 1$$

Pero esta condición no se tiene por qué cumplir para los s'_i que se han calculado hasta el momento, porque no se ha impuesto explícitamente esta restricción. Hará falta, por lo tanto, corregirlos como sigue para que la cumplan:

$$s_i = \frac{s'_i}{\sum_{i=1}^{12} s'_i/12}$$

El proceso de estimación de los factores estacionales no se acaba aquí. De hecho, los anteriormente calculados se consideran aún como una primera estimación los coeficientes de estacionalidad. Los perfeccionamientos propuestos por Shiskin añaden tres nuevas etapas:

a) Cálculo de los valores desestacionalizados de la serie original:

$$Y_{ti}^{D} = \frac{Y_{ti}}{s_i} \qquad t = 1, 2, \ldots, T$$

b) Obtener una segunda estimación de la tendencia de la serie mediante medias móviles centradas sobre los datos desestacionalizados. El orden de la media móvil dependerá de la variabilidad de los datos. Por experiencia Shiskin propone una media simple de orden 9, 13 o 23.

c) Con estos nuevos datos de tendencia se repite el cálculo de índices para cada mes y, finalmente, se determinan los coeficientes estacionales definitivos para cada mes.

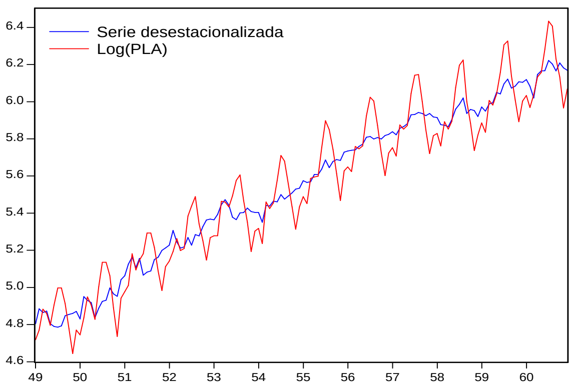
El resultado final es, en cualquier caso, la obtención de unos coeficientes estacionales, constantes en el tiempo para cada mes. Una vez estimados estos estacionales, se puede analizar el componente estacional o se puede desestacionalizar la serie.

Gráfico 4.6: Pasajeros de líneas aéreas. Relación a la media móvil.

Sample: 1949:01 1960:12
Included observations: 144
Difference from Moving Average
Original Series: PasajeroL
Adjusted Series: PasajeroLSA

Scaling Factors:

1	-0.085815
2	-0.114413
3	0.018113
4	-0.013046
5	-0.008966
6	0.115393
7	0.210816
8	0.204512
9	0.064836
10	-0.075271
11	-0.215846
12	-0.100315



En el gráfico 4.6 se pueden observar los resultados de aplicar este método a la serie de *Pasajeros de líneas aéreas* en logaritmos. Dado que la serie está en logaritmos, el modelo apropiado para esta serie sería el modelo aditivo. En la tabla se muestran los factores estacionales estimados y en el gráfico la serie desestacionalizada.

Aunque el filtro X-11 es un método muy utilizado en la práctica no está libre de críticas. Muchas provienen del hecho de que es un método que carece de toda base estadística o teórica lo que impide hacer una interpretación clara de los resultados que se obtienen.

Ejercicio 4.1. Deriva el método de relación a la media móvil para el modelo aditivo.

4.2.2. Alisado exponencial de Holt-Winters con estacionalidad

La función de predicción del método de alisado de Holt-Winters con tendencia era una línea recta con pendiente. Ahora bien, si la serie temporal presenta comportamiento estacional de la forma siguiente:

$$Y_t = T_t + S_t + I_t, \quad t = 1, 2, \ldots, T$$

sería deseable que la función de predicción incluyera también un componente estacional, de forma que quedara:

$$Y_T(\ell) = (m_T + b_T\,\ell) + S_T(\ell) \qquad \ell = 1, 2, \ldots$$

El método de predicción de Holt-Winters para la tendencia local lineal se puede extender fácilmente para introducir la estacionalidad.

El método de alisado exponencial de Holt-Winters con estacionalidad, H-W, consiste en tres ecuaciones, cada una de las cuales suaviza un factor asociado con cada uno de los componentes de la serie: aleatoriedad, tendencia y estacionalidad, por lo que ha de utilizar tres constantes de alisamiento, α, β, γ, todas comprendidas entre 0 y 1. Por esto, a este método se le conoce también como método de alisamiento exponencial con triple parámetro. Existen dos versiones de este procedimiento dependiendo de si los componentes de tendencia y estacionalidad se combinan aditiva o multiplicativamente.

Consideremos el modelo aditivo anterior. Las recursiones para el nivel, la pendiente y la estacionalidad, m_t, b_t y s_t, son respectivamente:

$$\begin{aligned} m_t &= \alpha\,(Y_t - S_{t-s}) + (1-\alpha)\,[m_{t-1} + b_{t-1}] \\ b_t &= \beta\,(m_t - m_{t-1}) + (1-\beta)\,b_{t-1} \\ S_t &= \gamma\,(Y_t - m_t) + (1-\gamma)\,S_{t-s} \end{aligned}$$

La función de predicción con toda la información disponible hasta t es de la forma:

$$Y_t(\ell) = m_t + b_t\,\ell + S_t(\ell) \qquad \ell = 1, 2, \ldots$$

donde $S_t(\ell)$ es el factor estacional apropiado para la observación que se predice, $Y_{t+\ell}$.

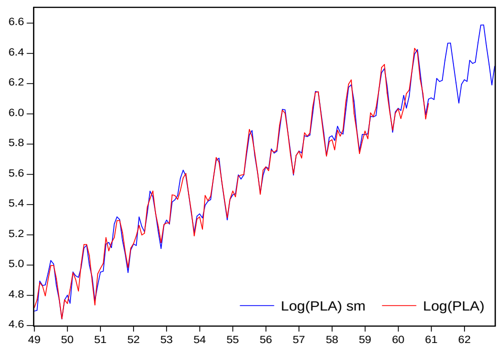
El método de alisado de Holt-Winters con estacionalidad es intuitivamente muy atractivo. La ecuación de actualización del factor estacional S_t es comparable a un índice estacional. Este índice se encuentra como la resta entre el valor actual de la serie Y_t y el valor suavizado de la serie m_t. Si Y_t es mayor que m_t el resto será positivo, y si es más pequeño, el resto será menor que cero. Para entender el papel del índice estacional, S_t, es importante tener en cuenta que m_t es un valor alisado de la serie que incluye tendencia pero no estacionalidad. Los valores de Y_t, por el contrario, contienen estacionalidad. Por lo tanto, la diferencia $Y_t - m_t$, da idea sobre la parte de estacionalidad en el dato observado. Sin embargo, la estacionalidad en cada periodo no es perfecta, sino que contiene aleatoriedad. Para eliminar esta aleatoriedad hay que promediar o alisar. Para alisar la estacionalidad, la ecuación para S_t pondera el nuevo factor estacional $Y_t - m_t$ con γ y el factor estacional más reciente correspondiente a la misma estación S_{t-s} con $(1-\gamma)$. Las dos primeras recursiones son las mismas que para el modelo con tendencia lineal, pero con la Y_t corregida también por el efecto estacional.

Una debilidad obvia de este método es que cada componente estacional se actualiza únicamente cada s periodos y la desestacionalización de la serie en la primera recursión utiliza, por lo tanto, una estimación del factor estacional que está s periodos "atrasada". Sin embargo, estos métodos parece que funcionan muy bien en la práctica.

Gráfico 4.7: Pasajeros de líneas aéreas. Holt-Winters con estacionalidad.

Sample: 1949:01 1960:12
Included observations: 144
Method: Holt-Winters Additive Seasonal
Original Series: pasajeroL
Forecast Series: AIRELSM

Parameters:	Alpha		0.7500
	Beta		0.0000
	Gamma		0.0000
Sum of Squared Residuals			0.186013
Root Mean Squared Error			0.035941
End of Period Levels:	Mean		6.181905
	Trend		0.009985
	Seasonals:	1960:01	-0.085867
		1960:02	-0.107839
		1960:03	0.022472
		1960:04	-0.008713
		1960:05	-0.011002
		1960:06	0.111228
		1960:07	0.215254
		1960:08	0.206043
		1960:09	0.061492
		1960:10	-0.076583
		1960:11	-0.220217
		1960:12	-0.106268



Como es un método recursivo son necesarios los valores iniciales. Para obtenerlos, se puede estimar por MCO con las k primeras observaciones un modelo de regresión lineal con variables ficticias para los factores estacionales y empezar las recursiones para $t = k + 1$.

El gráfico 4.7 muestra los resultados de predecir la serie *Pasajeros de líneas aéreas* en logaritmos por el método de Holt-Winters con estacionalidad. La tabla proporciona las constantes de alisamiento estimadas por el criterio de minimizar el error cuadrático medio de predicción un periodo hacia adelante y, además, las estimaciones del nivel, la pendiente y los factores estacionales obtenidas con toda la muestra. En la figura están representadas la serie original y la serie suavizada (Log(PLA)sm). Esta última contiene las predicciones un periodo hacia adelante para $t = 1, \ldots, T$, y las predicciones por punto con información hasta T para $t = T + \ell \;\; \ell = 1, 2, \ldots$.

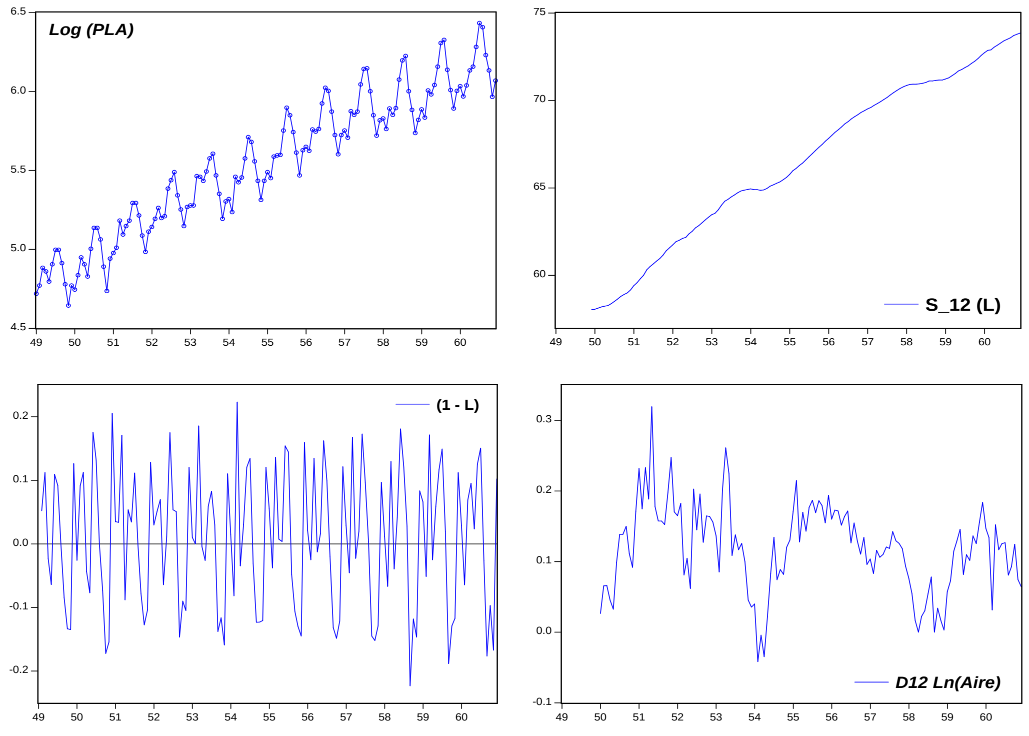
Ejercicio 4.3. Deriva las recursiones del método H-W para el modelo multiplicativo.

4.3. Diferenciación

El componente estacional se puede eliminar de la serie aplicando los denominados operadores de diferencias estacionales, que relacionan un mes con el precedente. Por ejemplo, si los datos son mensuales, se toman diferencias de orden 12, o lo que es lo mismo se aplica el operador Δ_{12} donde:

$$\Delta_{12}\, Y_t \;=\; Y_t - Y_{t-12}$$

Gráfico 4.8: Pasajeros de líneas aéreas(logs). Diferenciación.



Si suponemos que la serie presenta un componente estacional regular cada 12 meses:

$$Y_t \;=\; S(t) + x_t$$

donde Y_t es la serie observada, $S(t)$ es una estacionalidad perfectamente regular y x_t es el componente no estacional de la serie. Tomando diferencias de orden 12 se obtiene:

$$W_t \, = \, Y_t - Y_{t-12} \, = \, S(t) - S(t-12) + x_t - x_{t-12} \, = \, x_t - x_{t-12}$$

de forma que la serie diferenciada ya no presenta comportamiento estacional.

Para entender mejor cómo funciona el operador de diferencias estacional, hay que tener en cuenta que lo podemos descomponer como sigue:

$$\Delta_{12} \, = \, (1 - L^{12}) \, = \, (1 + L + L^2 + \ldots + L^{11}) \, (1 - L) \, = \, S_{12}(L) \, (1 - L)$$

Luego, el operador Δ_{12} incluye el operador suma de todas las observaciones a lo largo del año, $S_{12}(L)$, por lo que elimina la estacionalidad, y el operador de diferencias regular, $(1 - L)$, por lo que se elimina la tendencia creciente.

El gráfico 4.8 muestra los resultados de aplicar diferentes operadores de diferencias a la serie *Pasajeros de líneas aéreas* en logaritmos. En las cuatro figuras de dicho gráfico se puede apreciar con claridad los distintos efectos que la aplicación de cada uno de los operadores tiene sobre la serie original. Así, si se toma una diferencia de orden 1 $(1 - L)$, se elimina la tendencia creciente, pero queda la estacionalidad y el irregular. La aplicación del operador $S_{12}(L)$ elimina la estacionalidad dejando la tendencia creciente y el irregular, mientras que el operador de diferencias estacional elimina ambos componentes.

Capítulo 5

Modelos Estructurales de Series Temporales

5.1. Especificación del modelo estructural

Los modelos estructurales de series temporales se formulan directamente en función de componentes no observables pero que tienen una interpretación directa:

$$Y_t = T_t + S_t + I_t \qquad t = 1, 2, \ldots, T \tag{5.1}$$

donde Y_t representa las observaciones (a veces en logaritmos), T_t la tendencia, S_t la estacionalidad e I_t el componente irregular. El modelo (5.1) se puede completar con otros componentes que recojan ciclos, efectos diarios, festivos, etc. (Harvey 1989).

Una de las principales características de los modelos estructurales es que cada uno de sus componentes se modela explícitamente. Además, los modelos estructurales entran dentro de la categoría de modelos estocásticos de series temporales porque los componentes no se especifican de forma determinista, sino que se les permite evolucionar a lo largo del tiempo, introduciendo en su modelización un elemento aleatorio.

Tendencia

La tendencia se puede especificar de maneras muy diferentes aún centrándonos en modelos lineales. La forma más sencilla de modelar la tendencia es como una función determinista del tiempo (véase capítulo 3):

$$T_t = a + b\ t \tag{5.2}$$

Este modelo proyecta una tendencia determinista global, en particular una línea recta que se obtiene dando el mismo peso a todas las observaciones en la función de predicción. Si partimos de la base de que los diversos componentes de la serie evolucionan a lo largo del tiempo, es razonable pensar que la información más importante sobre los valores corrientes y futuros de la tendencia estará contenida en las observaciones más recientes de la serie. Por lo tanto, intuitivamente parece mejor proyectar la tendencia, así como otros componentes de la serie, poniendo mayor peso en las observaciones más recientes.

La tendencia local se puede especificar mediante un modelo estocástico. Para derivar este modelo se puede partir del modelo determinista (5.2) que escrito de forma recursiva queda como sigue:

$$\begin{aligned} T_t &= T_{t-1} + \beta_{t-1} \\ \beta_t &= \beta_{t-1} \end{aligned}$$

donde β_t es la pendiente de la tendencia en el momento t y siendo $T_0 = a$ y $\beta_0 = b$.

Las perturbaciones estocásticas se introducen en el modelo fácilmente:

$$\begin{aligned} T_t &= T_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t &= \beta_{t-1} + \zeta_t \end{aligned}$$

donde η_t y ζ_t son perturbaciones aleatorias con media cero y matriz de covarianzas Q. Con el fin de asegurar la identificabilidad de los componentes han de imponerse algunas restricciones en la matriz Q. Por simplificar, vamos a suponer que la matrix Q es diagonal:

$$Q = \begin{bmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\zeta^2 \end{bmatrix}$$

La introducción de las perturbaciones aleatorias en el modelo permite que la tendencia evolucione con el tiempo: la perturbación η_t hace que el nivel de la tendencia pueda subir o bajar y ζ_t permite que la pendiente cambie. Si $\sigma_\zeta^2 = 0$, el proceso que genera la tendencia se convierte en:

$$T_t = T_{t-1} + \beta + \eta_t$$

Este modelo permite que el nivel de la tendencia pueda evolucionar en el tiempo alrededor de un punto fijo, pero no así la pendiente. Por otro lado, si $\sigma_\eta^2 = \sigma_\zeta^2 = 0$, la tendencia estocástica se transforma en una tendencia determinista. Por lo tanto, los modelos de tendencia estocástica (o tendencia local) engloban como un caso particular los modelos de tendencia global. Cuanto mayores sean las varianzas de las perturbaciones estocásticas, mayores serán los movimientos aleatorios de la tendencia.

Estacionalidad

Se puede esperar la presencia de un comportamiento estacional acusado si trabajamos con series mensuales y trimestrales ó, en general, con series formadas por un número de observaciones s, tomadas regularmente cada año ($s = 4$ para las series trimestrales y $s = 12$ para las series mensuales).

Los procedimientos para incorporar un componente estacional estocástico en los modelos estructurales son muy variados. Partiendo del componente estacional determinista especificado por medio de variables ficticias estacionales (véase capítulo 4), se puede especificar la estacionalidad estocástica como sigue. El efecto estacional determinístico cumple para cada momento t la restricción de que la suma de los efectos estacionales a lo largo del año ha de ser cero:

$$\sum_{j=0}^{s-1} s_{t-j} = 0 \tag{5.3}$$

Como esta modelización es muy restrictiva, en ocasiones, interesa un esquema más flexible basado en permitir que el componente estacional evolucione en el tiempo sujeto a la restricción de que la suma esperada de los efectos estacionales a lo largo del año sea cero. Esto se puede especificar introduciendo una perturbación aleatoria en la restricción (5.3):

$$\sum_{j=0}^{s-1} s_{t-j} = \omega_t$$

donde ω_t sigue una distribución normal con media cero y varianza σ_ω^2, lo que implica que la suma de los efectos estacionales evoluciona aleatoriamente en torno a cero. Este modelo permite que el patrón estacional varíe en el tiempo mediante un mecanismo que asegura que la suma de los efectos estacionales cada s periodos consecutivos tiene un valor esperado cero y una varianza constante en el tiempo.

Irregular

El componente irregular representa aquellas influencias aleatorias en la serie que tienen sólo efectos momentáneos y es estacionario por definición. Se suele especificar como un proceso ruido blanco con media cero y varianza σ_I^2.

5.1.1. Principales Modelos Estructurales

En esta sección se describen los principales modelos estructurales de series temporales en los que los distintos componentes que hemos especificado, tendencia, estacionalidad e irregular, se van a combinar de forma aditiva.

Modelo de paseo aleatorio con ruido

El modelo estructural más sencillo está formado por un componente de tendencia más una perturbación aleatoria. Este modelo se va a denominar paseo aleatorio con ruido, debido a que la tendencia T_t se especifica como un proceso de paseo aleatorio, es decir, depende únicamente de su pasado en el momento $t-1$ y de una perturbación aleatoria. Este modelo fue introducido originalmente por Muth (1960):

$$\begin{aligned} Y_t &= T_t + \varepsilon_t \\ T_t &= T_{t-1} + \eta_t \end{aligned} \tag{5.4}$$

donde ε_t y η_t son perturbaciones aleatorias incorreladas entre sí y que se distribuyen conjuntamente como sigue:

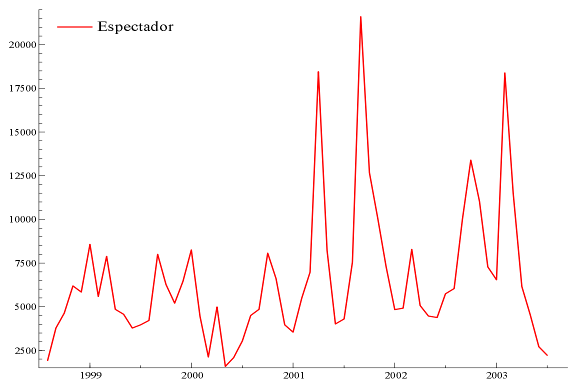
$$\left[\begin{array}{c} \varepsilon_t \\ \eta_t \end{array}\right] \sim N\left(0, \left[\begin{array}{cc} \sigma_\varepsilon^2 & 0 \\ 0 & \sigma_\eta^2 \end{array}\right]\right)$$

Una extensión muy utilizada de este modelo es el modelo denominado paseo aleatorio con rumbo en el que la tendencia incluye una pendiente fija:

$$\begin{aligned} Y_t &= T_t + \varepsilon_t \\ T_t &= T_{t-1} + \beta + \eta_t \end{aligned}$$

Ejemplo 5.1. Serie *Espectadores de cine.*

Gráfico 5.1: Espectadores de cine



El gráfico 5.1 muestra la evolución de la serie espectadores de cine desde agosto de 1998 hasta julio de 2003. Los componentes que se observan en esta serie son tendencia e irregular. Dado que la tendencia no presenta un crecimiento (o decrecimiento) continuado, sino que oscila en torno a un nivel promedio más o menos constante el modelo de paseo aleatorio con ruido (5.4) sería apropiado para describir el comportamiento de esta serie.

Modelo de tendencia local lineal

El modelo de tendencia local lineal es una extensión del modelo (5.4) en la que tanto el nivel como la pendiente de la tendencia siguen procesos de paseo aleatorio. El modelo fue introducido por Theil y Wage (1964) y se especifica como sigue:

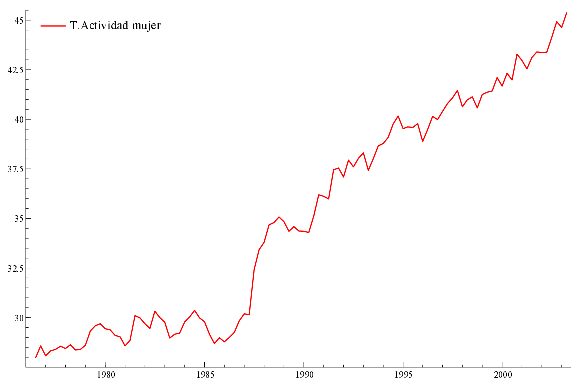
$$\begin{aligned} Y_t &= T_t + \varepsilon_t \\ T_t &= T_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t &= \beta_{t-1} + \zeta_t \end{aligned} \tag{5.5}$$

donde ε_t, η_t y ζ_t son perturbaciones aleatorias incorreladas entre sí y que se distribuyen conjuntamente como sigue:

$$\begin{bmatrix} \varepsilon_t \\ \eta_t \\ \zeta_t \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma_\varepsilon^2 & 0 & 0 \\ 0 & \sigma_\eta^2 & 0 \\ 0 & 0 & \sigma_\zeta^2 \end{bmatrix}\right)$$

Ejemplo 5.2. Serie *Tasa de actividad de la mujer.*

Gráfico 5.2: Tasa de actividad de la mujer



El gráfico 5.2 muestra la evolución de la serie de tasa de actividad de la mujer en Bizkaia desde el tercer trimestre de 1976 hasta el segundo trimestre de 2003. Los componentes que se observan en esta serie son tendencia e irregular. Dado que la tendencia presenta un crecimiento continuado el modelo apropiado para describir el comportamiento de esta serie seria el modelo de tendencia local lineal (5.5).

Modelo estructural básico

El modelo adecuado, en principio, para series mensuales o trimestrales con claro comportamiento estacional es el modelo estructural básico (MEB) en el que la serie observada se descompone en tendencia, estacionalidad e irregular:

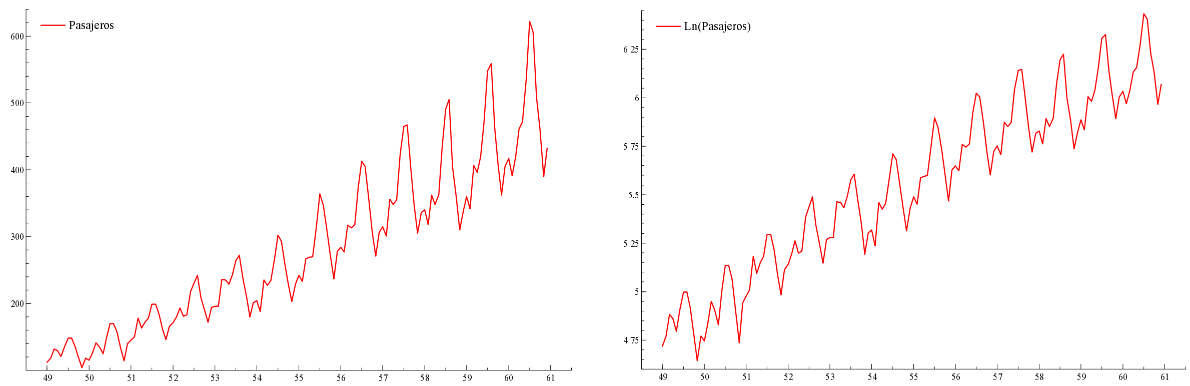
$$
\begin{aligned}
Y_t &= T_t + S_t + \varepsilon_t \\
T_t &= T_{t-1} + \beta_{t-1} + \eta_t \\
\beta_t &= \beta_{t-1} + \zeta_t \\
s_t &= -\sum_{j=1}^{s-1} s_{t-j} + \omega_t
\end{aligned} \tag{5.6}
$$

donde los términos de perturbación $\varepsilon_t, \eta_t, \zeta_t$ y ω_t están incorrelacionados entre sí, tienen media cero y y que se distribuyen conjuntamente como sigue:

$$
\begin{bmatrix} \varepsilon_t \\ \eta_t \\ \zeta_t \\ \omega_t \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma_\varepsilon^2 & 0 & 0 & 0 \\ 0 & \sigma_\eta^2 & 0 & 0 \\ 0 & 0 & \sigma_\zeta^2 & 0 \\ 0 & 0 & 0 & \sigma_\omega^2 \end{bmatrix}\right)
$$

Ejemplo 5.3. Serie *Pasajeros de líneas aéreas.*

Gráfico 5.3: Pasajeros de líneas aéreas



La figura de la izquierda del gráfico 5.3 representa la serie mensual de pasajeros de líneas aéreas desde el mes de enero de 1949 hasta el mes de diciembre de 1960. Como se puede observar esta serie presenta comportamiento estacional además de tendencia y componente irregular. Además, la amplitud del ciclo estacional aumenta conforme crece la tendencia por lo que la relación entre el componente de tendencia y el componente estacional es multiplicativa.

Para poder especificar un modelo aditivo, se toman logaritmos a la serie. La figura derecha del gráfico 5.3 muestra que la amplitud del componente estacional de la serie $Ln(Pasajeros)$ es más homogénea y ya no crece con la tendencia. Por lo tanto, el modelo adecuado para reproducir la evolución de la serie $Ln(Pasajeros)$ es el Modelo Estructural Básico (5.6).

5.2. Estimación máximo verosímil

Los parámetros desconocidos de los modelos estructurales de series temporales (MEST) se pueden estimar por máxima verosimilitud construyendo la función de verosimilitud mediante el método de la descomposición del error de predicción. Este procedimiento de estimación requiere la representación de los MEST en el espacio de los estados. Los modelos escritos en el espacio de los estados presentan la ventaja de que permiten utilizar una serie de algoritmos basados en el filtro de Kalman. Mediante este filtro se puede calcular la función de verosimilitud para modelos complejos. En lo que se refiere a la estimación de los MEST, una vez expresados en el espacio de los estados y bajo el supuesto de que las perturbaciones siguen una distribución normal, se puede estimar los parámetros construyendo la función de verosimilitud a partir de los residuos del filtro de Kalman mediante la descomposición del error de predicción.

Una vez estimados los parámetros del modelo se pueden obtener estimaciones óptimas de los componentes no observados de interés como tendencia, estacionalidad, etc. mediante un algoritmo basado también en el filtro de Kalman.

En primer lugar vamos a explicar brevemente qué son los modelos en el espacio de los estados y después se presentará el filtro de Kalman.

5.2.1. Modelos en el Espacio de los Estados

Consideremos el caso más general en el que contamos con una serie temporal multivariante, Y_t de N elementos. El *modelo general en el espacio de los estados*(MEE), relaciona estas variables observadas Y_t con un vector $(m \times 1)$, α_t , denominado *vector estado* a través de la llamada *ecuación de medida*:

$$Y_t = Z_t \, \alpha_t + d_t + \epsilon_t \qquad\qquad t = 1, \ldots, T \tag{5.7}$$

donde Z_t es una matriz de orden $(N \times m)$, d_t es un vector $(N \times 1)$ y ϵ_t es un vector $(N \times 1)$ de perturbaciones aleatorias serialmente incorreladas con

$$E(\epsilon_t) = 0 \qquad y \qquad Var(\epsilon_t) = H_t \qquad t = 1, \ldots, T$$

e incorrelado con el vector estado, es decir, $Cov(\epsilon_t, \alpha_{t-k}) = 0 \;\; k \geq 0$.

En general, las variables α_t no son observables pero se suele suponer que están generadas por un proceso markoviano de primer orden. La ecuación que nos describe la evolución en el tiempo del vector estado α_t se denomina *ecuación de transición*:

$$\alpha_t = C_t \; \alpha_{t-1} + e_t + R_t \eta_t \qquad\qquad t = 1, \ldots, T \tag{5.8}$$

donde C_t es una matriz de orden $(m \times m)$, e_t es un vector $(m \times 1)$, R_t es una matriz $(m \times g)$, y η_t es un vector $(g \times 1)$ de perturbaciones serialmente incorreladas con

$$E(\eta_t) = 0 \qquad y \qquad Var(\eta_t) = Q_t$$

La inclusión de la matriz R_t junto con el vector η_t es hasta cierto punto arbitraria. El término de pertubación siempre se podría redefinir de forma que su matriz de covarianzas fuera $R_t Q_t R_t'$. De todas maneras, la especificación dada por (5.8) es más adecuada cuando η_t se identifica con un conjunto particular de perturbaciones del modelo.

La especificación completa de un modelo en el espacio de los estados suele incluir dos supuestos adicionales:

a) El vector estado inicial α_0, tiene media a_0 y matriz de covarianzas P_0:

$$E(\alpha_0) = a_0 \qquad y \qquad Var(\alpha_0) = P_0$$

b) Los términos de perturbación ϵ_t y η_t están incorrelados entre sí y con el vector estado inicial:

$$\begin{array}{lcl} E(\epsilon_t \eta_s') = 0 & & \forall s, t = 1, \ldots, T \\ E(\epsilon_t \alpha_0') = 0 & E(\eta_t \alpha_0') = 0 & t = 1, \ldots, T \end{array}$$

En general vamos a considerar que las matrices $Z_t, d_t, H_t, C_t, e_t, R_t, Q_t$ no son estocásticas[1] de forma que el MEE es un modelo lineal[2]. Si además imponemos la restricción de que no cambien a lo largo del tiempo, diremos que el MEE correspondiente es invariante en el tiempo:

$$\begin{aligned} Y_t &= Z\,\alpha_t + d + \epsilon_t \\ \alpha_t &= C\alpha_{t-1} + e + R\eta_t \end{aligned} \tag{5.9}$$

Para un modelo univariante e invariante en el tiempo, la representación en el espacio de los estados sería:

$$\begin{aligned} Y_t &= z\,\alpha_t + d + \epsilon_t \\ \alpha_t &= C\alpha_{t-1} + e + R_t\eta_t \end{aligned} \tag{5.10}$$

donde Y_t es un escalar y z es un vector $(1 \times m)$ y $Var(\varepsilon) = H = \sigma^2$.

Los modelos estructurales univariantes, objeto de estudio de este texto, están escritos en el espacio de los estados directamente, donde el vector estado α_t contiene los componentes no observados de tendencia, estacionalidad, etc.

[1] Si se les permite cambiar en el tiempo es de forma prederminada.

[2] El modelo es lineal en el sentido de que para cualquier valor de t, y_t se puede expresar como una combinación lineal de valores pasados y presentes de η_t, β_t, X_t, ϵ_t y el vector estado inicial α_0.

Modelo de Paseo Aleatorio con ruido. Este modelo está directamente escrito en el espacio de los estados con:

$$\alpha_t = \mu_t \qquad C = z = R = 1 \qquad d = e = 0 \qquad Q = \sigma^2_\eta$$

El vector estado recoge el único componente no observable de la serie que viene representado por el nivel de la tendencia μ_t.

Es importante tener en cuenta que la definición del vector estado α_t para cualquier modelo estadístico viene dada por construcción. Los elementos de α_t pueden estar o no identificados con componentes susceptibles de una interpretación interesante, por ejemplo, en términos de tendencia y estacionalidad.

Modelo de tendencia local lineal. Supongamos que el vector estado recoge los componentes de nivel y pendiente de la tendencia, μ_t, β_t :

$$\begin{aligned} y_t &= [1\ \ 0]\, \alpha_t + \epsilon_t \\ \alpha_t &= \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \zeta_t \end{bmatrix} \end{aligned}$$

con $Q = diag(\sigma^2_\eta,\ \sigma^2_\zeta)$.

Modelo estructural básico. Consideremos el caso particular de un modelo estructural básico con un componente estacional de periodo $s = 4$ y definido en función de variables ficticias. La representación en el espacio de los estados:

$$\begin{aligned} y_t &= [1\ 0\ 1\ 0\ 0]\, \alpha_t + \epsilon_t \\ \alpha_t &= \begin{bmatrix} \mu_t \\ \beta_t \\ \gamma_t \\ \gamma_{t-1} \\ \gamma_{t-2} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \\ \gamma_{t-1} \\ \gamma_{t-2} \\ \gamma_{t-3} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \zeta_t \\ \omega_t \\ 0 \\ 0 \end{bmatrix} \end{aligned}$$

donde con $Q = diag(\sigma^2_\eta,\ \sigma^2_\zeta,\ \sigma^2_\omega,\ 0,\ 0)$.

Ejercicio 5.1. Escribe en el espacio de los estados el modelo estructural básico para una serie mensual.

5.2.2. El Filtro de Kalman

Una vez que el modelo está escrito en la forma del Espacio de los Estados, se pueden aplicar un número importante de algoritmos que tienen como centro el *Filtro de Kalman* (FK). Para simplificar, se van a considerar los modelos univariantes y invariantes en el tiempo.

El FK es un procedimiento recursivo cuyo objetivo es calcular *estimadores óptimos* del vector estado en un momento dado t, basados en toda la información disponible hasta ese momento t. Este conjunto de información consistiría en las observaciones de la serie Y_t hasta e incluyendo el momento t. Tanto z, C, R, Q, y σ^2, como los valores iniciales a_0 y P_0 se suponen conocidos en todos los momentos de tiempo y, por lo tanto, no es necesario que estén explicitamente incluidos en el conjunto de información.

Este método o algoritmo recursivo se puede dividir en dos grupos de ecuaciones:

- Ecuaciones de Predicción, que obtienen un predictor óptimo del vector estado y de la próxima observación con toda la información hasta t:

  dados a_t, P_t se obtienen $a_{t+1/t}$, $Y_{t+1/t}$, $P_{t+1/t}$

- Ecuaciones de Actualización que incorporan la nueva observación Y_{t+1} en la estimación del vector estado:

  dada la nueva observación Y_{t+1} se obtienen a_{t+1}, P_{t+1}

Una de las ventajas del FK es que permite actualizar la estimación del vector estado continuamente según vamos obteniendo nuevas observaciones. Esta aplicación es de gran utilidad, en general, en ingeniería, donde se producen nuevas observaciones muy frecuentemente. No parece, sin embargo, que tenga tanta importancia en Economía donde las observaciones aparecen a intervalos más largos, por lo que cabe la pregunta ¿cuál es la utilidad del FK en el campo de la economía?

- Produce predicciones óptimas (en el sentido de ECM mínimo) tanto de las observaciones futuras como del vector estado junto con sus ECM de predicción.

- En cuanto al vector se refiere, no siempre tiene interpretación económica, pero cuando la tiene (tendencia, estacionalidad, etc.) interesa estimar su valor en cada momento t, a_t , utilizando toda la información de la muestra y no sólo parte de ella, es decir, la disponible hasta t. Esto es lo que se conoce como *extracción de señales*. Existen algoritmos relacionados con el FK que solucionan este problema.

- Otra razón fundamental es que cuando las perturbaciones y el vector estado inicial se distribuyen normalmente, se puede construir la función de verosimilitud via lo que se conoce como la descomposición del error de predicción. Este resultado proporciona una base para estimar los parámetros desconocidos del modelo y realizar contrastes de diagnósticos y especificación.

Forma general del Filtro

Consideremos el modelo en el espacio de los estados invariante en el tiempo siguiente:

$$
\begin{aligned}
Y_t &= z\,\alpha_t + d + \epsilon_t \\
\alpha_t &= C\alpha_{t-1} + e + R\eta_t
\end{aligned}
$$

donde:

$$
\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim NID\left[0, \begin{pmatrix} \sigma^2 & 0 \\ 0 & Q \end{pmatrix}\right]
$$

$$
Cov(\epsilon_t, \alpha_{t-k}) = 0 \quad \forall k \geq 0 \qquad\qquad \alpha_0 \sim N(a_0,\, P_0)
$$

$$
E(\epsilon_t\, \eta_s') = 0 \qquad E(\epsilon_t\, \alpha_0') = 0 \qquad E(\eta_t\, \alpha_0') = 0 \qquad \forall t, s
$$

El conjunto de ecuaciones que forman el FK se dividen en *ecuaciones de predicción* que proporcionan el predictor óptimo para el vector estado α_{t+1} con toda la información hasta t, y las *ecuaciones de actualización* que nos permiten incorporar la nueva información disponible dentro del estimador del vector estado.

Sea a_{t-1} el estimador óptimo [3] de α_{t-1}, basado en todas las observaciones disponibles hasta e incluyendo Y_{t-1}. Denotemos por P_{t-1} la matriz de covarianzas del error de estimación del vector estado:

$$
P_{t-1} = E\left[(\alpha_{t-1} - a_{t-1})\,(\alpha_{t-1} - a_{t-1})'\right]
$$

Dados a_{t-1} y P_{t-1}, el estimador óptimo de α_t junto con su ECM viene dado por las *ecuaciones de predicción* :

$$
a_{t/t-1} = C\, a_{t-1} + e \tag{5.11}
$$

$$
P_{t/t-1} = C\, P_t\, C' + R\, Q\, R' \tag{5.12}
$$

Dada esta estimación del vector estado en el momento t con información hasta el momento t-1, se puede obtener la predicción del valor futuro Y_t:

$$
Y_{t/t-1} = z\, a_{t/t-1} + d
$$

El error de predicción correspondiente es:

$$
\nu_t = Y_t - Y_{t/t-1} = Y_t - z\, a_{t/t-1} - d = z\,(\alpha_t - a_{t/t-1}) + \epsilon_t
$$

Estos errores de predicción se denominan *innovaciones* ya que representan la nueva información contenida en la última observación.

[3] Se dice que a_t es un estimador óptimo de α_t cuando, $\forall a_t$ dentro de la clase considerada:

$$
E\left[(\alpha_t - a_t)\,(\alpha_t - a_t)'\right] < E\left[(\alpha_t - \tilde{a}_t)\,(\alpha_t - \tilde{a}_t)'\right]
$$

donde $<$ indica que la matriz de la izquierda tiene elementos menores en la diagonal principal.

Cuando llega la nueva observación, Y_t, el estimador de α_t, $a_{t/t-1}$, junto con su matriz de ECM, se puede actualizar, mediante las *ecuaciones de actualización*:

$$
\begin{aligned}
a_t &= a_{t/t-1} + P_{t/t-1} z' \, f_t^{-1} \, (Y_t - z \, a_{t/t-1} - d) &(5.13)\\
P_t &= P_{t/t-1} \; - \; P_{t/t-1} \, z' \, f_t^{-1} z \, P_{t/t-1} &(5.14)
\end{aligned}
$$

donde:

$$
f_t = z \, P_{t/t-1} \, z' \; + \sigma^2
$$

Se puede observar que las innovaciones tienen un papel fundamental en la actualización del vector estado. Cuanto más alejado este el error ν_t del vector nulo, mayor será la corrección en el vector estado. Además podemos observar tambien que f_t no es más que la varianza del error de predicción.

Las ecuaciones de predicción junto con las de actualización conforman el FK. Este se puede especificar tambien como un solo conjunto de recursiones:

$$
\begin{aligned}
a_{t+1/t} &= C \, [a_{t/t-1} + \; P_{t/t-1} z' \, f_t^{-1} \, (Y_t - z' \, a_{t/t-1} - d)] + e &(5.15)\\
P_{t+1/t} &= C \, [P_{t/t-1} \; - \; P_{t/t-1} \, z' \, f_t^{-1} z \, P'_{t/t-1}] \, C' + R \, Q \, R' &(5.16)
\end{aligned}
$$

Operando:

$$
\begin{aligned}
a_{t+1/t} &= C \, a_{t/t-1} + \; k_t (Y_t - z' \, a_{t/t-1} - d)] + e &(5.17)\\
P_{t+1/t} &= C \, [P_{t/t-1} \; - \; P_{t/t-1} \, z' \, f_t^{-1} z \, P'_{t/t-1}] \, C' + R \, Q \, R' &(5.18)
\end{aligned}
$$

donde:

$$
k_t = C \, P_{t/t-1} z' \, f_t^{-1}
$$

es conocida como la *ganancia de Kalman*. La ecuación de recursión para el vector estado se puede escribir como una combinación lineal de la nueva observación y de la última estimación:

$$
a_{t+1/t} = (C - k \, z') \, a_{t/t-1} + k_t \, Y_t
$$

La recursión para la matriz de covarianzas del error de estimación (6) se conoce como la *ecuación de Ricatti*.

Los valores iniciales del FK se pueden especificar en términos de a_0 y P_0 y/o de $a_{1/0}$ y $P_{1/0}$. Dadas las condiciones iniciales, el FK nos proporciona el estimador óptimo del vector estado según va llegando cada nueva observación. Cuando se han procesado las T observaciones, el filtro nos proporciona el estimador óptimo del vector estado en el momento T, a_T, y en el $T + 1$ basado en el conjunto de información completo, $a_{T+1/T}$. Este estimador contiene toda la información necesaria para llevar a cabo predicciones óptimas de los valores futuros tanto de Y_t como del vector estado.

Interpretación del Filtro de Kalman

Bajo los supuestos de normalidad siguientes:

$$\alpha_0 \sim N(a_0,\, P_0)$$

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left[0, \begin{pmatrix} \sigma^2 & 0 \\ 0 & Q \end{pmatrix}\right]$$

podemos derivar el FK y darle una interpretación determinada. Para $t = 1$, tenemos que:

$$\alpha_1 = C\alpha_0 + R\eta_1$$

donde α_1 sigue una distribución normal multivariante porque es combinación lineal de variables normales. Además, *condicionado* a Y_0, tenemos que la media de α_1 es:

$$a_{1/0} = Ca_0$$

y su matriz de covarianzas:

$$P_{1/0} = CP_0C' + RQR$$

Con lo que la distribución condicionada es:

$$\alpha_1/Y_0 \sim N(a_{1/0},\, P_{1/0})$$

donde $a_{1/0}$ es la media de α_1 condicionada a la información disponible en el momento $t = 0$.

Si se conociera la distribución de α_1 en $t = 0$, entonces los resultados anteriores son redundantes y las condiciones iniciales vendrían dadas en términos de $a_{1/0}$ y $P_{1/0}$.

Para obtener la distribución de α_1 condicionda a Y_1, se hace lo siguiente:

$$\begin{aligned} \alpha_1 &= a_{1/0} + (\alpha_1 - a_{1/0}) \\ Y_1 &= z'\, a_{1/0} + Z'\,(\alpha_1 - a_{1/0}) + \epsilon_1 \end{aligned}$$

El vector $(\alpha_1',\, Y_1)$ tiene una distribución normal con media $(a_{1/0}'\; z'\, a_{1/0})$ y matriz de covarianzas:

$$\begin{bmatrix} P_{1/0} & P_{1/0}Z \\ z'P_{1/0} & z'P_{1/0}z + \sigma^2 \end{bmatrix}$$

Aplicando el siguiente lema:

$$\begin{aligned} \begin{bmatrix} X \\ Y \end{bmatrix} &\sim \left[\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix}\right] \\ X/Y &\sim N\left[\mu_{X/Y},\, \Sigma_{XX/Y}\right] \end{aligned}$$

donde:

$$\begin{aligned} \mu_{X/Y} &= \mu_X + \Sigma_{XY}\,\Sigma_{YY}^{-1}\,(Y - \mu_Y) \\ \Sigma_{XX/Y} &= \Sigma_{XX} - \Sigma_{XY}\,\Sigma_{YY}^{-1}\,\Sigma_{YX} \end{aligned}$$

se ve que la distribución de α_1 condicionada a Y_1 es:

$$\alpha_1/Y_1 \sim N(a_1, P_1)$$

donde:

$$\begin{aligned} a_1 &= a_{1/0} + P_{1/0}Z\,\left(Z'P_{1/0}Z + H\right)^{-1}\,\left(Y_1 - Z'a_{1/0}\right) \\ P_1 &= P_{1/0} - P_{1/0}Z\,\left(Z'P_{1/0}Z + H\right)^{-1} Z'\,P_{1/0} \end{aligned}$$

Repitiendo los pasos sucesivamente, tenemos las ecuaciones de predicción y actualización del FK.

La derivación del FK sólo permite interpretar a_t y P_t como la media y la matriz de covarianzas de la distribución condicionada de α_t. Pero al presentar el FK, a_t ha sido descrito como el estimador óptimo de α_t basado en toda la información disponible en el momento t, mientras que P_t se presentaba como la matriz de covarianzas del error de estimación.

¿Por qué bajo el supuesto de Normalidad, el FK proporciona estimadores óptimos?

Se puede demostrar que, bajo los supuestos de Normalidad, el FK proporciona la media y la matriz de covarianzas de la distribución de α_t condicionada a toda la información disponible hasta el momento t:

$$\begin{aligned} a_t &= E[\alpha_t|Y_1, Y_2, \ldots, Y_t] = E_t(\alpha_t) \\ P_t &= E_t\left\{(\alpha_t - E_t(\alpha_t))\,(\alpha_t - E_t(\alpha_t))'\right\} \end{aligned}$$

Análogamente se tiene que:

$$\begin{aligned} a_{t/t-1} &= E[\alpha_t|Y_1, Y_2, \ldots, Y_{t-1}] = E_{t-1}(\alpha_t) \\ P_{t/t-1} &= E_{t-1}\left\{(\alpha_t - E_{t-1}(\alpha_t))\,(\alpha_t - E_{t-1}(\alpha_t))'\right\} \end{aligned}$$

Consideremos la media de la distribución de α_t condicionada a la información disponible en el momento t, $E_t(\alpha_t)$. La media condicionada es una *estimación* de α_t de Error Cuadrático Medio Mínimo, en el sentido de que cualquier otra estimación tiene una matriz de ECM que se puede escribir como:

$$V = P_t + A$$

donde A es una matriz semidefinida positiva.

Esta media condicionada se puede considerar tambien como un *estimador* de α_t. La diferencia entre una estimación y un estimador es que la primera es un número mientras el segundo es una regla. Es decir, la media condicionada como una estimación es una colección de números asociadas a una realización particular de observaciones; mientras que como estimador es una expresión que se aplica a cualquier conjunto de observaciones. Con este cambio de interpretación la media condicionada se convierte en un vector de variables aleatorias. Ahora bien, se puede demostrar que este estimador minimiza el ECM cuando el valor medio se toma sobre todas las variables del conjunto de información y no sobre un conjunto particular de valores. En este sentido, $E_t(\alpha_t)$ es un estimador de ECM mínimo de α_t.

Como, en general, la cantidad que va a ser estimada, o sea el vector estado, es aleatorio, no podemos hablar de matriz de covarianzas del estimador, ni de insesgadez del estimador, etc. De la misma forma que cuando hablamos de insesgadez nos referimos a que el valor medio del error de estimación es cero, tambien hablamos de matriz de covarianzas del error de estimación que tambien denominamos matriz de ECM del estimador.

Cuando las perturbaciones del modelo en el espacio de los estados no se distribuyen normalmente ya no es cierto que, en general, el FK proporcione la media condicionada del vector estado. ¿Qué propiedades tiene el estimador que proporciona el FK, a_t? Restringiéndose a los estimadores que son *lineales* en las observaciones, entonces a_t es el estimador que minimiza el ECM. Es decir, a_t es el estimador lineal de ECM mínimo de α_t.

Todo lo dicho anteriormente se aplica a $a_{t/t-1}$ y $P_{t/t-1}$, y a las predicciones de Y_t con información hasta el momento $t-1$. Dada la estimación del vector estado $a_{t/t-1}$, la predicción del valor futuro, Y_t es:

$$Y_{t/t-1} = z\, a_{t/t-1}$$

que son las predicciones de ECM mínimo si el modelo es normal, y si no lo es, $Y_{t/t-1}$ es el predictor lineal de ECM mínimo de y_t. Los errores de predicción son:

$$\nu_t = Y_t - Y_{t/t-1} = z\,(\alpha_t - a_{t/t-1}) + \epsilon_t \qquad t = 1, \ldots, T$$

y se conocen como *innovaciones*, ya que representan la nueva información contenida en la última observación.

En un modelo gaussiano el valor medio del error ν_t es cero, y en este sentido el predictor es insesgado, y su varianza (o ECM de predicción un periodo hacia adelante) viene dada por:

$$V(\nu_t) = E\left[z\,(\alpha_t - a_{t/t-1}) + \epsilon_t\right]\,\left[z\,(\alpha_t - a_{t/t-1}) + \epsilon_t\right]' = f_t$$

Se puede demostrar que estos errores de predicción están incorrelados entre sí para diferentes momentos de tiempo[4]:

$$E(\nu_t\,\nu_s') = E\left[z\,(\alpha_t - a_{t/t-1}) + \epsilon_t\right]\,\left[z\,(\alpha_s - a_{s/s-1}) + \epsilon_s\right]' = 0 \qquad t \neq s$$

[4]

$$E\left[\epsilon_t\,(\alpha_t - a_{t/t-1})\right] = 0$$

de forma que:

$$\nu_t \sim NID(0, f_t)$$

Si no hacemos el supuesto de normalidad en el modelo, la media del vector de innovaciones sigue siendo cero, mientras que su matriz de covarianzas sigue siendo f_t.

Es interesante hacer hincapié, en que los resultados sobre la distribución de las innovaciones solo son exactos si las matrices del sistema son fijas y conocidas. No son, en general, ciertos si estas matrices contienen algunos parámetros desconocidos que hay que sustituir por estimaciones.

Los valores iniciales necesarios para comenzar las recursiones del filtro de Kalman se puede especificar en función de a_0 y P_0 o de $a_{1/0}$ y $P_{1/0}$. Si existe información a priori de este tipo, es decir, si la media y la matriz de covarianzas del vector estado inicial son conocidas, entonces el FK produce estimaciones optimas del vector estado así como predicciones optimas de los valores de las futuras observaciones, basadas en toda la información disponible incluida la observación actual.

Condiciones Iniciales

Considerénse las condiciones iniciales para un modelo en el espacio de los estados invariante en el tiempo. En principio, los valores iniciales vendrían dados por la media y la matriz de varianzas de la distribución no condicionada del vector estado. Las recursiones comienzan en a_0, P_0 o $a_{1/0}, P_{1/0}$. Dado que $a_{1/0}$ es la predicción de α_1 basada en ninguna información sobre Y o X. Por lo tanto, es simplemente la esperanza no condicionada de α_1:

$$a_{1/0} = E(\alpha_1)$$

y su ECM asociado:

$$P_{1/0} = E\left[\alpha_1 - E(\alpha_1)\right]\,\left[\alpha_1 - E(\alpha_1)\right]'$$

El vector estado es estacionario si $|\lambda_i(C)| > 1$, es decir, si los valores propios de la matriz C son todos menores que la unidad en valor absoluto. En este caso la media no condicionada de α_1 se puede obtener a partir del propio modelo tomando esperanzas no condicionadas a ambos lados:

$$E(\alpha_1) = C\,E(\alpha_{t-1})$$

Como α_t es estacionario en covarianza:

$$(I_m - C)\,E(\alpha_1) = 0$$

porque ϵ_t está incorrelado con α_t por definición, y además está tambien incorrelado con $a_{t/t-1}$ porque es una función lineal de Y_{t-1}.

y como ninguno de los valores propios de C es la unidad, porque el vector estado es estacioanrio, la matriz $(I_m - C)$ no es singular y la única solución para este sistema es:

$$E(\alpha_1) = 0$$

La matriz de covarianzas no condicionada de α_1 , P_1, se puede obtener tambien a trav,s del modelo, simplemente multiplicandolo por su traspuesta y tomando esperanzas:

$$\begin{aligned}
E(\alpha_t)(\alpha_t)' &= E\left[C\,\alpha_{t-1} + R\eta_t\right]\left[C\,\alpha_{t-1} + R\eta_t\right]' = \\
&= C E(\alpha_{t-1})(\alpha_{t-1})' C' + E(R\eta_t\eta_t' R') \\
\\
\Sigma &= C\,\Sigma\,C' + RQR'
\end{aligned}$$

cuya solución general es:

$$vec(\Sigma) = [I - C \bigotimes C]^{-1}\, vec(RQR')$$

Por lo que la matriz de covarianzas no condicionada de α_1, se obtiene resolviendo la ecuación de Ricatti. Es decir, en general, si el vector estado es estacionario, y como la distribución no condicionada de α_1 es la misma que la distribución no condicionada de α_0, las iteraciones del FK pueden comenzar por:

$$\begin{aligned}
a_0 = 0 &\qquad P_0 = \Sigma \\
a_{1/0} = 0 &\qquad P_{1/0} = \Sigma
\end{aligned}$$

Cuando el vector estado no es estacionario, su distribución no condicionada no está definida. A no ser que tengamos información a priori, la distribución inicial de α_0 se puede especificar como sigue:

a) A Priori Difuso: $P_0 = \kappa I, \quad \kappa > 0$

   El 'a priori' difuso se obtiene cuando $\kappa \to \infty$, que corresponde con P_0^{-1}. En este caso tenemos una distribución impropia en el sentido de que no integra a 1. Hay que señalar que tambien se puede aplicar el 'a priori' difuso a α_1, con lo que $P_0 = \kappa I$.

b) Obtener las condiciones iniciales a partir de las primeras observaciones. Si $d \leq m$ elementos del vector estado no son estacionarios, entonces utilizaremos las d primeras observaciones para construir las condiciones iniciales y comenzaremos las recursiones a partir de la observación $d + 1$.

Para modelos univariantes, se puede demostrar que "el uso de un a priori difuso es equivalente a construir las condiciones iniciales a partir de las m primeras observacioenes dado que el modelo es observable".

Ejemplo 5.4. Consideremos el modelo de paseo aleatorio con ruido siguiente:

$$\begin{aligned} Y_t &= \mu_t + \epsilon_t & V(\epsilon_t) = \sigma^2 \\ \mu_t &= \mu_{t-1} + \eta_t & V(\eta_t) = \sigma^2_\eta \end{aligned}$$

Un aspecto que podemos tener en cuenta cuando trabajamos con filtros en modelos univariantes es que a veces puede ser interesante construir el modelo en el espacio de los estados de forma que la varianza de las perturbaciones sea proporcional a un escalar positivo. De esta forma:

$$V(\epsilon_t) = \sigma^2_* h \qquad V(\eta_t) = \sigma^2_* Q$$

Si la matriz de covarianzas inicial se especifica tambien con un factor de proporcionalidad, σ^2_* entondes el FK se puede correr independientemente de σ^2_* con las cantidades con asterisco:

$$P_{t+1/t} = \sigma^2_* P^*_{t+1/t} \qquad f_t = \sigma^2_* f^*_t$$

apareciendo en las recursiones.

Consideremos el modelo de paseo aleatorio con ruido escrito como sigue:

$$\begin{aligned} Y_t &= \mu_t + \epsilon_t & V(\epsilon_t) = \sigma^2 \times 1 \\ \mu_t &= \mu_{t-1} + \eta_t & V(\eta_t) = \sigma^2_\eta = \sigma^2 q = \sigma^2 \frac{\sigma^2_\eta}{\sigma^2} \end{aligned}$$

Ecuaciones de predicción

$$\begin{aligned} m_{t/t-1} &= m_{t-1} \\ p_{t/t-1} &= p_{t-1} + q \end{aligned}$$

Ecuaciones de actualización

$$\begin{aligned} m_t &= m_{t/t-1} + p_{t/t-1} f_t^{-1} (y_t - m_{t/t-1}) \\ p_t &= p_{t/t-1} - p_{t/t-1} f_t^{-1} p_{t/t-1} \\ f_t &= p_{t/t-1} + 1 \end{aligned}$$

de forma que:

$$\begin{aligned} m_t &= m_{t/t-1} + \frac{p_{t/t-1}}{p_{t/t-1} + 1} (y_t - m_{t/t-1}) \\ p_t &= p_{t/t-1} - \frac{p^2_{t/t-1}}{p_{t/t-1} + 1} \end{aligned}$$

Condiciones Iniciales. Supongamos que el proceso de paseo aleatorio comenzó en algún momento del pasado remoto, entonces $P_{1/0}$ sería infinita. Pongamos $P_{1/0} = \kappa, \quad \kappa > 0$ y arbitriamente grande:

$$\begin{aligned} m_{2/1} &= (1-k_1)m_{1/0} + k_1 y_1 \qquad k_1 = \frac{\kappa}{1+\kappa} \\ p_{2/1} &= \kappa - \frac{\kappa^2}{1+\kappa} + q \end{aligned}$$

Cuando:

$$\kappa \to \infty \quad \Rightarrow \quad k_1 = 1 \quad \Rightarrow \quad \begin{array}{c} m_{2/1} = y_1 \\ p_{2/1} = 1 + q \end{array}$$

5.2.3. Estimación por Máxima Verosimilitud

Supongamos que hemos de estimar el vector de parámetros θ, en base a T conjuntos de observaciones $Y_1, Y_2, \ldots, Y_T$ independiente e idénticamente distribuidas. La función de densidad conjunta será:

$$L(Y;\theta) = \prod_{t=1}^{T} p(Y_t;\theta) \tag{5.19}$$

donde $p(Y_t)$ es la función de densidad del t-ésimo conjunto de observaciones. Tomada la muestra, $L(Y,\theta)$ se puede reinterpretar como la función de verosimilitud y el estimador máximo-verosimil (MV) será aquel que maximice la función (5.19) respecto de θ.

Una de las caracteristicas de las series temporales es que las T observaciones no son independientes. Por lo tanto, no se puede utilizar (5.19) y se ha de definir la función MV como sigue: [5]

$$L(Y;\theta) = \prod_{t=1}^{T} p(Y_t/\mathbf{Y}_{t-1}) \tag{5.20}$$

donde $p(Y_t/\mathbf{Y}_{t-1})$ representa la distribución de Y_t condicionada al conjunto de información disponible en el momento $t-1$, es decir, $\mathbf{Y}_{t-1} = (Y_1, \ldots, Y_{t-1})$. Para estimar la función de verosimilitud necesitamos conocer la distribución $p(Y_t/\mathbf{Y}_{t-1})$.

[5] Esta definición de la función de verosimilitud se obtiene a partir del conocido resultado:

$$p(A \bigcap B) = p(A/B)\, p(B)$$

La densidad conjunta de la muestra $\mathbf{Y}_T = (Y_1, Y_2, \ldots, Y_T)$ se puede escribir entonces de la siguiente manera:

$$p(\mathbf{Y}_T) = p(Y_T, \mathbf{Y}_{T-1}) = p(Y_T/\mathbf{Y}_{T-1})\, p(\mathbf{Y}_{T-1})$$

y aplicando sucesivamente el mismo resultado obtenemos la función de verosimilitud (5.19).

Si el modelo está escrito en la forma del espacio de los estados y si las perturbaciones y el vector estado inicial del MEE se distribuyen como Normales Multivariantes, la distribución de Y_t condicionada a $\mathbf{Y}_{t-1}$ será tambien normal. Además, como se ha señalado anteriormente

$$E_{t-1}(Y_t) = z\, a_{t/t-1} = Y_{t/t-1}$$

$$ECM_{t-1}(Y_t) = E_{t-1}\left[(Y_t - E_{t-1}(Y_t))(Y_t - E_{t-1}(Y_t))'\right] = V(\nu_t) = f_t$$

obteniendose ambas mediante el filtro de Kalman.

Por lo tanto, la función de Verosimilitud se puede escribir como:

$$logL = -\frac{T}{2}\log 2\pi - \sum_{t=1}^{T}\log f_t - \frac{1}{2}\sum_{t=1}^{T}\frac{(Y_t - Y_{t/t-1})^2}{f_t}$$

Como $Y_t - Y_{t/t-1} = \nu_t$, que puede ser interpretado como el vector de errores de predicción un periodo hacia adelante, a esta expresión de la función de verosimilitud se le denomina *descomposición del error de predicción*. Los estimadores MV de θ se obtendrán minimizando la función de Verosimilitud mediante algún procedimiento de optimización numérica.

Ejemplo 5.5. *Modelo de paseo aleatorio con ruido.*

La estimación del modelo de paseo aleatorio con ruido (5.4) para la serie Espectadores de cine proporciona los siguientes resultados:

- Parámetros estimados

$$\hat{\sigma}_\varepsilon = 2988,7 \qquad \hat{\sigma}_\mu = 1632,1$$

- Estimación del vector estado al final de la muestra (entre paréntesis el ECM):

$$m_T = 4126,9 \quad (1929,9)$$

  donde m_T es la estimación del nivel de la tendencia con toda la información de la muestra.

Ejemplo 5.6. *Modelo de tendencia local lineal.*

La estimación del modelo de tendencia local lineal (5.5) para la serie Tasa de actividad de la mujer proporciona los siguientes resultados.

- Parámetros estimados

$$\hat{\sigma}_\varepsilon = 0,12223 \qquad \hat{\sigma}_\mu = 0,49467 \qquad \hat{\sigma}_\zeta = 0,00000$$

- Estimación del vector estado al final de la muestra (entre paréntesis el ECM):

$$m_T = 45,34 \quad (0,1188) \qquad b_t = 0,16195 \quad (0,0478)$$

donde m_T, b_T son las estimación del nivel y la pendiente de la tendencia respectivamente, con toda la información de la muestra.

Ejemplo 5.7. *Modelo estructural básico.*
La estimación del modelo estructual báscio (5.6) para la serie Pasajeros de líneas aéreas proporciona los siguientes resultados:

- Parámetros estimados

$$\hat{\sigma}_\varepsilon = 0,01138 \qquad \hat{\sigma}_\mu = 0,026447 \qquad \hat{\sigma}_\zeta = 0,00000 \qquad \hat{\sigma}_\omega = 0,008008$$

- Estimación del vector estado al final de la muestra (entre paréntesis el ECM):

$$\begin{array}{ll}
\hat{\mu}_T = 6,1809 \quad (0,01698) & \hat{\beta} = 0,00937 \quad (0,0022) \\
s_{1,T} = -0,1101 \quad (0,0152) & s_{2,T} = -0,21568 \quad (0,01379) \\
s_{3,T} = -0,06963 \quad (0,0137) & s_{4,T} = 0,040004 \quad (0,01374) \\
s_{5,T} = 0,21936 \quad (0,01374) & s_{6,T} = 0,23184 \quad (0,01374) \\
s_{7,T} = 0,10554 \quad (0,01374) & s_{8,T} = -0,00295 \quad (0,01375) \\
s_{9,T} = -0,100244 \quad (0,01376) & s_{10,T} = -0,014385 \quad (0,01378) \\
s_{11,T} = -0,11648 \quad (0,013786) &
\end{array}$$

donde m_T, b_T son las estimación del nivel y la pendiente de la tendencia respectivamente y $s_{j,T}$ es la estimación del factor estacional del mes j-ésimo, con toda la información de la muestra.

5.3. Predicción

Una de las finalidades del análisis de series temporales es la predicción de futuras observaciones. En el campo de los modelos estructurales de series temporales, una vez especificado el modelo apropiado para la serie temporal, se puede obtener la función de predicción de las observaciones y de los componentes no observables del modelo, junto con sus errores cuadráticos medios, mediante el filtro de Kalman.

En resumen, si conocemos los parámetros del modelo, la predicción de Y_{T+1} junto con su error cuadrático medio vienen dados directamente por las ecuaciones de predicción del filtro de Kalman. En el modelo gaussiano, el Filtro de Kalman proporciona a_T, el estimador de ECM mínimo de α_T basado en todas las observaciones. Además, proporciona tambien la predicción un periodo hacia adelante y su ECMP:

$$\begin{aligned} a_{T+1/T} &= C a_T \\ Y_{T+1/T} &= z\, a_{T+1/T} \\ ECM[Y_{T+1/T}] &= z\, P_{T+1/T}\, z' + \sigma_\varepsilon^2 \end{aligned}$$

Consideremos ahora el problema general de obtener predicciones de valores futuros de $Y_{T+\ell}, \ell = 1, 2, \ldots$ basadas en la información de la muestra, es decir, conocidos a_T y P_T.

En el modelo gaussiano, la predicción óptima, en el sentido de que minimiza el error cuadrático medio de predicción, de $Y_{T+\ell}$, $Y_T(\ell)$, es la esperanza condicionada:

$$Y_T(\ell) = E[Y_{T+\ell}|Y_T, Y_{T-1}, \ldots, Y_1] = z\, a_{T+\ell/T}$$

donde

$$a_{T+\ell/T} = E[a_{T+\ell}|Y_T, Y_{T-1}, \ldots, Y_1] = z\, a_{T+\ell/T}$$

y el Error Cuadrático Medio de Predicción viene dado por:

$$ECM(Y_{T+\ell}) = z\, P_{T+\ell/T}\, z' + \sigma_\varepsilon^2 \qquad \ell = 1, 2, \ldots$$

La mejor forma de obtener $a_{T+\ell/T}$ y $P_{T+\ell/T}$ es sustituir repetidamente en las ecuaciones de predicción del FK saltándose las de actualización (ya que no tenemos nuevas observaciones):

$$\begin{aligned} \alpha_{T+\ell} &= C^\ell\, \alpha_T + \sum_{j=1}^{\ell-1} C^{\ell-j}\, R\, \eta_{T+j} + R\, \eta_{T+\ell} \qquad \ell = 2, 3, \ldots \\ E_T(\alpha_{T+\ell}) &= a_{T+\ell/T} = C^\ell a_T \\ P_{T+\ell/T} &= C^\ell\, P_T\, (C')^\ell + \sum_{j=0}^{\ell-1} C^j\, Q\, (C')^j \end{aligned}$$

Hay que tener en cuenta que las matrices de ECM, $P_{T+\ell/T}$, no tienen en cuenta los errores que provienen de estimar cualquier parámetro desconocido en la matrices del sistema: C, Z_t, etc. Pero normalmente el modelo contiene un conjunto de parámetros desconocidos, θ, por lo que la formula para el $ECM(y_{T+\ell/T})$ subestima el verdadero ECM porque no tiene en cuenta la variación extra debida a la estimación de los parámetros.

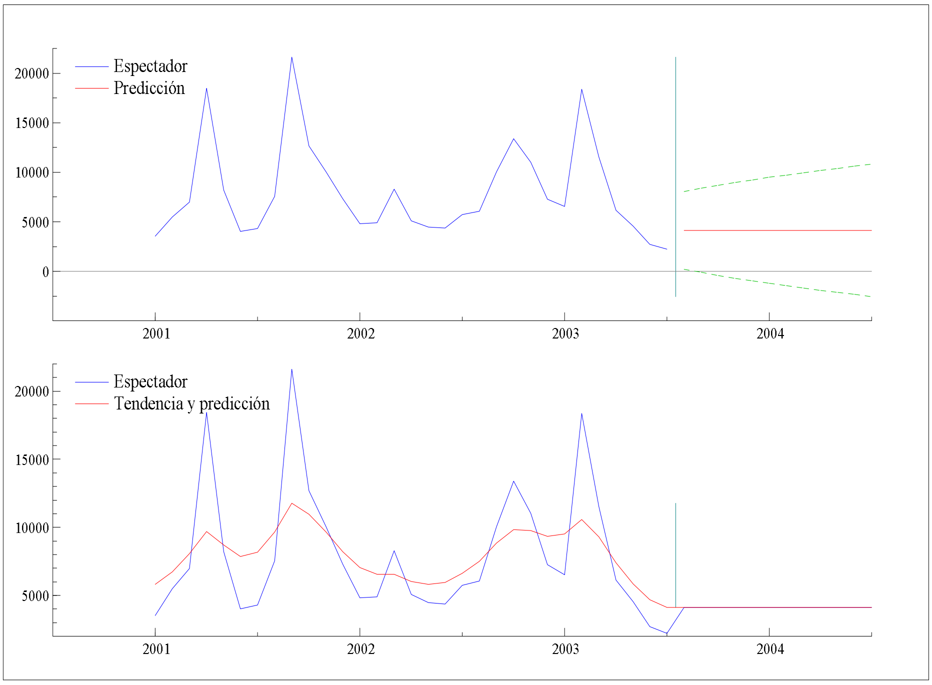
Ejemplo 5.8. *Modelo de paseo aleatorio con ruido.*

Una vez estimado el modelo de paseo aleatorio con ruido (5.4) para la serie Espectadores de cine, se pueden obtener predicciones para el futuro tanto por punto como por intervalo (vease el gráfico 5.4). Nótese que en este caso las predicciones para la serie vienen dadas por la proyeccion del nivel de la tendencia estimado con toda la informacion hasta T:

$$
\begin{aligned}
a_{T+\ell/T} &= C^{\ell} a_T = a_T = m_T \\
Y_T(\ell) &= z\, a_{T+\ell/T} = a_T = m_T \\
ECM[Y_T(\ell)] &= z\,(C^{\ell}\, P_T\,(C')^{\ell} + \sum_{j=0}^{\ell-1} C^j\, Q\,(C')^j)\, z' + \sigma_{\varepsilon}^2 = \\
&= P_T + \ell\, \sigma_{\eta}^2 + \sigma_{\varepsilon}^2
\end{aligned}
$$

donde P_T es el error cuadrático medio de m_T estimación del nivel en el momento T.

Gráfico 5.4: Predicción



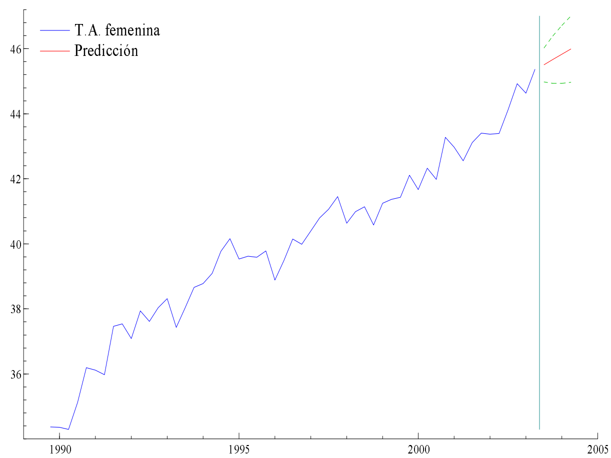
Ejemplo 5.9. *Modelo de tendencia local lineal.*

Una vez estimado el modelo de paseo aleatorio con ruido (5.5) para la serie Tasa de Actividad de la mujer, se pueden obtener predicciones para el futuro tanto por punto como por intervalo (vease el gráfico 5.5).

En este caso las predicciones para la serie vienen dadas por la proyección en el futuro de la tendencia lineal donde tanto el nivel como la pendiente han sido estimados con toda la informacion hasta T:

$$
\begin{aligned}
a_{T+\ell/T} &= C^{\ell} a_T = \begin{bmatrix} 1 & \ell \\ 0 & 1 \end{bmatrix} \begin{bmatrix} m_T \\ b_T \end{bmatrix} \\
Y_T(\ell) &= z\, a_{T+\ell/T} = [1\ \ 0] \begin{bmatrix} 1 & \ell \\ 0 & 1 \end{bmatrix} \begin{bmatrix} m_T \\ b_T \end{bmatrix} = m_T + \ell\, b_T \\
ECM[Y_T(\ell)] &= z\,(C^{\ell}\, P_T\, (C')^{\ell} + \sum_{j=0}^{\ell-1} C^j\, Q\, (C')^j)\, z' + \sigma_{\varepsilon}^2
\end{aligned}
$$

Gráfico 5.5: Predicción



Ejemplo 5.10. *Modelo estructural basico.*

Una vez estimado el modelo de paseo aleatorio con ruido (5.6) para la serie Pasajeros de lineas aereas, se pueden obtener predicciones para el futuro tanto por punto como por intervalo (vease el gráfico 5.6).

En este caso, la predicción del valor futuro de Y es la composicion de la proyección de los dos componentes no observados de la serie, tendencia y estacionalidad (vease el gráfico 5.7) estimados con toda la informacion hasta T:

$$
\begin{aligned}
a_{T+\ell/T} &= C^{\ell} a_T \\
Y_T(\ell) &= z a_{T+\ell/T} = a_T = m_T + \ell\, b_T + s_{j,T} \\
ECM[Y_T(\ell)] &= z\,(C^{\ell}\, P_T\, (C')^{\ell} + \sum_{j=0}^{\ell-1} C^j\, Q\, (C')^j)\, z' + \sigma_{\varepsilon}^2
\end{aligned}
$$

Gráfico 5.6: Predicción de futuras observaciones

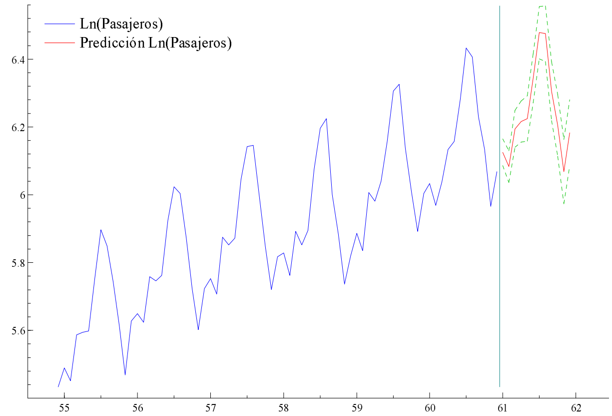
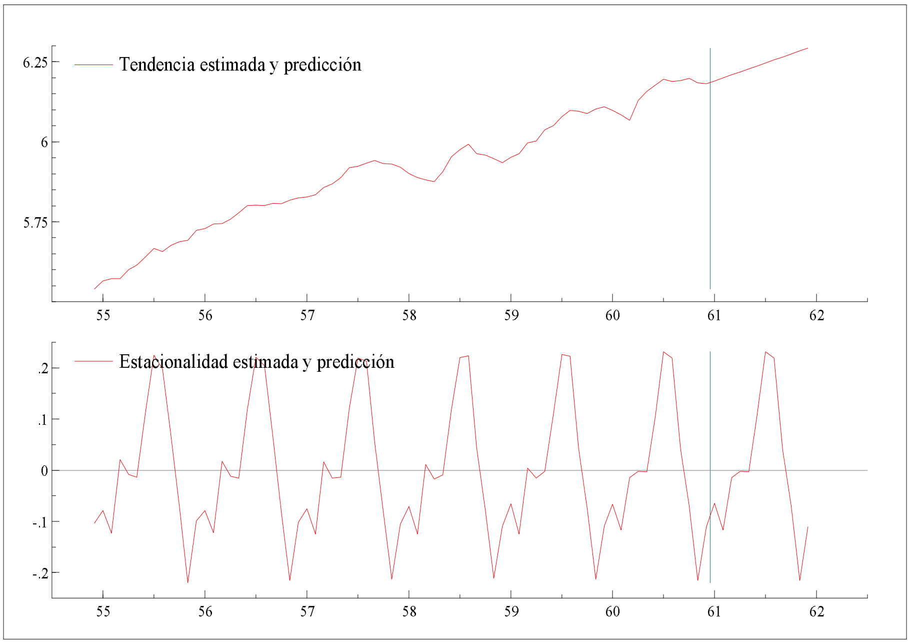


Gráfico 5.7: Predicción de los componentes



5.4. Extracción de señales

Uno de los objetivos principales al analizar una serie temporal suele ser la estimación en cada momento t de ciertos componentes no observables de la serie que tienen interés por sí mismos y que tradicionalmente se denominan tendencia, estacionalidad, etc. lo que se conoce como *extracción de señales*. Por ejemplo, si contamos con series mensuales que presentan comportamiento estacional puede interesar estimar la estacionalidad en cada momento t con el fin de obtener la serie desestacionalizada.

Los modelos estructurales de series temporales están especificados directamente en función de los componentes no observados de tendencia, estacionalidad, etc. Por otro lado, estos modelos se representan muy fácilmente en el espacio de los estados. Como hemos visto, una vez estimados los parámetros desconocidos del modelo por máxima verosimilitud, las recursiones del filtro de Kalman proporcionan estimaciones óptimas de los componentes no observados de la series que están incluidos en el vector estado α_t, basadas en las observaciones disponibles hasta el momento t. Sin embargo, sería posible obtener una estimación más eficiente de los componentes del vector estado α_t si utilizaramos toda la información muestral hasta el momento T, es decir, el objetivo es calcular $E(\alpha_t|Y_1, Y_2, \ldots, Y_T)$. A esta media condicional de α_t se le denomina estimación suavizada y al filtro que la genera le denominaremos algoritmo suavizador.

Es importante hacer la siguiente distinción. El objetivo del filtrado es encontrar el valor esperado del vector estado α_t, condicionado a la información disponible en el momento t, es decir, $E(\alpha_t|\mathbf{Y}_t)$. El objetivo del alisado (extracción de señales) es tener en cuenta tambien la información despuçes del momento t. La media de la distribución de α_t, condicionada a toda la muestra, se puede escribir como $E(\alpha_t|\mathbf{Y}_T)$ y se conoce como estimador alisado. Como este estimador se basa en más información que el estimador de filtrado, tendrá un ECM que será en general menor (nunca puede ser mayor).

Si el MEE es gaussiano:

$$a_{t/T} = E_T(\alpha_t) = E(\alpha_t|\mathbf{Y}_T)$$

es el estimador de ECM mínimo de α_t basado en toda la muestra. Si relajamos el supuesto de normalidad, aún sigue siendo el estimador lineal de ECM mínimo.

Existen varios algoritmos de alisado de un modelo lineal. El más interesante en nuestro caso es el *algoritmo de intervalo fijo*. Este algoritmo consiste en un conjunto de recursiones que comienzan con las cantidades finales, a_T y P_T, dadas por el FK y trabaja hacia atrás. Las ecuaciones son:

$$\begin{aligned} a_{t/T} &= a_t + P_t^*\,(a_{t+1/T} - C a_t) \\ P_{t/T} &= P_t + P_t^*\,(P_{t+1/T} - P_{t+1/t}) P_t^{*\prime} \end{aligned}$$

donde:

$$P_t^* = P_t\, C^{-1}\, P'_{t+1/t} \qquad t = T-1, \ldots, 1$$

con $a_{T/T} = a_T$ y $P_{T/T} = P_T$. Este algoritmo requiere, por lo tanto, que a_t y P_t se guarden para todo t, así como $P_{t+1/t}$.

Con estos algoritmos podemos estimar la senda seguida por parámetros que varían en el tiempo, componentes no observados de las series como tendencias y estacionalidades, etc.

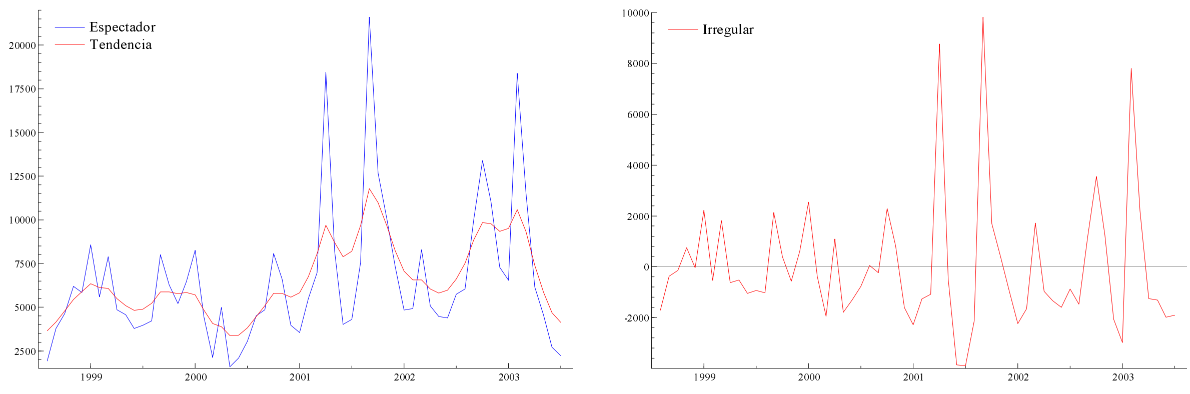
Una de las razones más conocidas por las que nos puede interesar la extracción de señales es para obtener series desestacionalizadas. Desestacionalizar una serie supone remover el componente estacional de la serie original. En un modelo estructural los componentes de una serie están definidos explícitamente, por lo tanto, una vez estimados los parámetros del modelo, podemos obtener el componente estacional mediante un algoritmo suavizador y extraerlo de la serie. La serie desestacionalizada sería:

$$Y_j^D = Y_t - s_{t/T} = Y_t - \sum_{j=1}^{s-1} s_{t-j/T}$$

Ejemplo 5.11. *Modelo de paseo aleatorio con ruido.*

El único componente no observable de este modelo es la tendencia. La figura izquierda del gráfico 5.8 muestra la serie Espectadores de cine y su tendencia estimada con toda la informacion de la muestra. Se puede observar que el componente de tendencia tiene una evolución mucho más suave que la serie. La figura derecha del gráfico 5.8 recoge el componente irregular estimado que, en este caso, es igual a la serie sin tendencia.

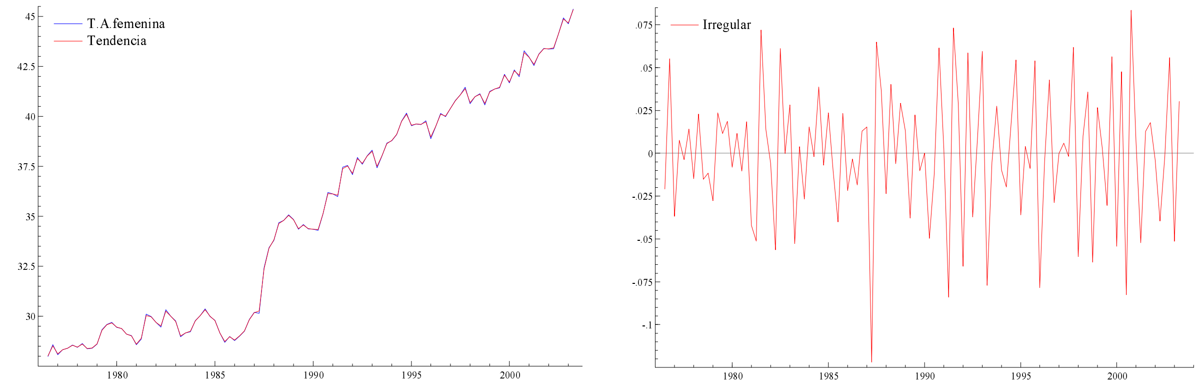
Gráfico 5.8: Componentes suavizados



Ejemplo 5.12. *Modelo de tendencia local lineal.*

El unico componente no observable de este modelo es la tendencia que consta de un nivel y una pendiente. La figura de la izquierda del gráfico 5.9 muestra la serie Tasa de actividad de la mujer y su tendencia estimada con toda la informacion de la muestra. El componente irregular estimado se recoge en el gráfico derecho de la misma figura. Se observar que la tendencia estimada y la serie observada son muy similares, siendo el irregular estimado muy pequeño.

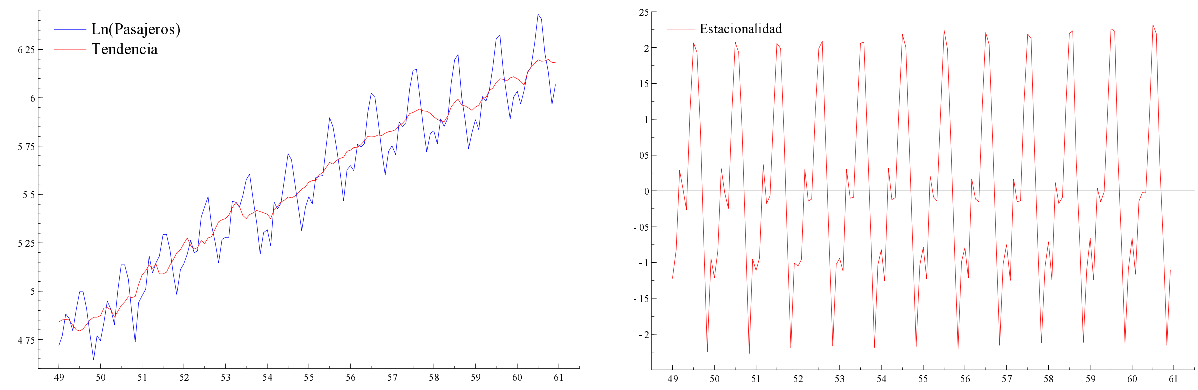
Gráfico 5.9: Componentes suavizados



Ejemplo 5.13. *Modelo estructural básico.*

En este modelo se incluyen dos componentes no observados de interés en sí mismos, la tendencia y la estacionalidad. En las dos figuras del gráfico 5.10 se representan la tendencia y la estacionalidad estimadas para la serie Pasajeros de líneas aéreas en logaritmos con toda la informacion de la muestra.

Gráfico 5.10: Componentes suavizados

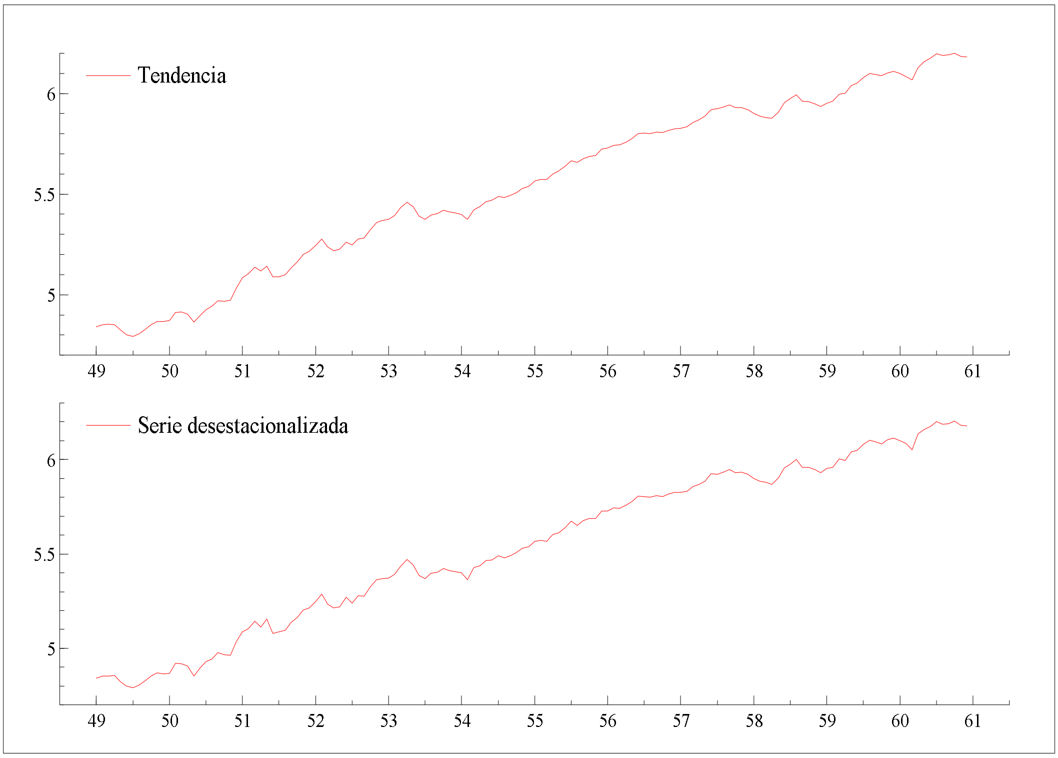


Otro de los objetivos que se plantean cuando se lleva a cabo la extracción de señales en una serie es la eliminación de alguna de estas señales. Así, si se elimina el componente de tendencia se obtiene la serie desestacionalizada:

$$Y_j^D = Y_t - s_{t/T} = Y_t - \sum_{j=1}^{s-1} s_{t-j/T}$$

La figura inferior del gráfico 5.11 muestra la serie Pasajeros de líneas aéreas desestacionalizada. Se puede observar que el ciclo estacional ha desaparecido y presenta una evolución creciente.

Gráfico 5.11: Serie desestacionalidad vs. tendencia suavizada



Cuando se trabaja con una serie con un marcado componente estacional como es el caso de la serie Pasajeros de Líneas aéreas, en ocasiones, interesa no analizar la serie bruta sino alguna señal libre de estacionalidad que permita estudiar mejor el comportamiento a largo plazo. Existen dos señales que se pueden utilizar en este caso: la serie desestacionalizada y la tendencia.

El gráfico 5.11 muestra ambas señales para la serie Pasajeros de líneas aéreas: la tendencia suavizada (figura superior) y la serie desestacionalizada (figura inferior) lo que permite compararlas fácilmente. Obviamente, las dos señales han eliminado la estacionalidad, pero la tendencia siempre va a ser más suave que la serie desestacionalizada ya que esta última incluye tendencia y componente irregular.

Por otro lado, también puede interesar eliminar el comportamiento a largo plazo de la serie. Existen dos señales que se pueden utilizar en este caso:

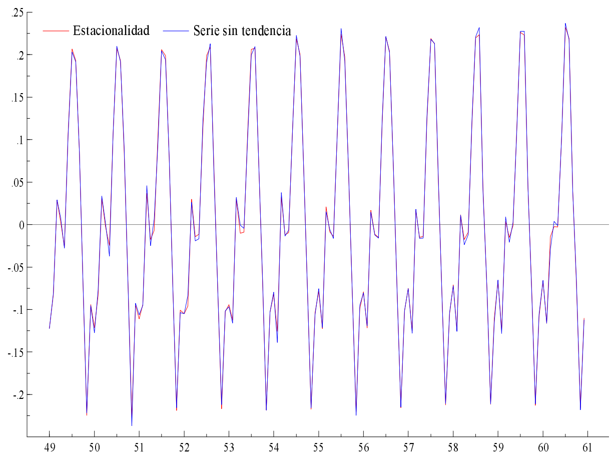
- La serie sin tendencia:

$$Y_j^{ST} = Y_t - T_{t/T} = \sum_{j=1}^{s-1} s_{t-j/T} + I_{t/T}$$

- El componente estacional suavizado

El gráfico 5.11 muestra ambas señales para la serie Pasajeros de líneas aéreas. Se puede observar que ambas señales oscilan en torno a cero y presentan el ciclo estacional. La estacionalidad estimada es más suave que la serie sin tendencia que incluye también el componente irregular.

Gráfico 5.12: Serie sin tendencia vs. estacionalidad suavizada



Capítulo 6

Predicción con modelos ARIMA

Dado que desarrollar el marco teórico de los modelos ARIMA y su metodología de modelización sería muy amplio para la extensión de este documento, se refiere al lector a libros de texto de referencia como Aznar y Trivez (1993) o Uriel (2000) donde encontrará toda la información necesaria sobre estos modelos.

El objetivo de este capítulo es obtener predicciones óptimas de Y_t en algún momento futuro basadas en un conjunto de información

$$I_T = \{Y_T,\, Y_{T-1},\, Y_{T-2},\, Y_{T-2}, \ldots\} \tag{6.1}$$

y bajo el supuesto de la serie Y_t sigue el siguiente modelo $ARIMA(p,d,q)$:

$$\Phi_p(L)\ \Delta^d Y_t \ = \delta \, + \, \Theta_q(L)\, a_t \tag{6.2}$$

donde el polinomio autorregresivo estacionario $\Phi_p(L)$ y el invertible de medias móviles $\Theta_q(L)$ no tienen raíces comunes.

Por predictor óptimo (o predicción óptima) se denomina a áquel que es la mejor en el sentido de que minimiza una determinada función de pérdida. Lo más usual es minimizar el Error Cuadrático Medio de Predicción, por lo que diremos que $Y_T(\ell)$ es un predictor óptimo si minimiza el ECMP, es decir, si cumple que:

$$E[Y_{T+\ell} - Y_T(\ell)]^2 \leq E[Y_{T+\ell} - Y_T^*(\ell)]^2 \qquad \forall\, Y_T^*(\ell)$$

Se puede demostrar que, bajo condiciones de regularidad muy débiles, el predictor por punto óptimo viene dado por la esperanza condicionada al conjunto de información:

$$Y_T(\ell) \ = \ E[Y_{T+\ell}|I_T] \ = \ E[Y_{T+\ell}|Y_T,\, Y_{T-1},\, Y_{T-2},\, Y_{T-3}, \ldots] \ = \ E_T[Y_{T+\ell}]$$

es decir, por el valor esperado de la distribución de $Y_T(\ell)$ condicionada la información disponible.

Nada garantiza que esta esperanza condicionada sea una función lineal del pasado de la serie. Pero si el proceso sigue una distribución normal, se puede demostrar que la esperanza condicionada se puede expresar como una función lineal del conjunto de información,

I_T. Por lo tanto, bajo el supuesto de normalidad, el predictor óptimo en el sentido de minimizar el ECMP es lineal. Si no se cumple este supuesto, la proyección lineal de $Y_{T+\ell}$ en su pasado proporcionaría el predictor óptimo dentro de la clase de predictores lineales.

La predicción óptima por intervalo se construirá a partir de la distribución del error de predicción que, bajo el supuesto de que $a_t \sim RBN(0, \sigma^2)$, es la siguiente:

$$e_T(\ell) \;=\; Y_{T+\ell} \,-\, Y_T(\ell) \,\sim\, N(0,\, V(e_T(\ell)))$$

Tipificando se obtiene:

$$\frac{Y_{T+\ell} \,-\, Y_T(\ell) - 0}{\sqrt{V(e_T(\ell))}} \,\sim\, N(0,\, 1)$$

De forma que el intervalo de predicción de probabilidad $(1-\alpha)\,\%$ es:

$$\left[Y_T(\ell) \,-\, N_{\alpha/2}\,\sqrt{V(e_T(\ell))}, \quad Y_T(\ell) \,+\, N_{\alpha/2}\,\sqrt{V(e_T(\ell))} \right]$$

Para tratar la predicción óptima en el marco de los modelos ARIMA comenzaremos desarrollando la teoría de la predicción para los modelos estacionarios, para después aplicarla a los modelos no estacionarios utilizando como ilustración aquellos modelos ARIMA más sencillos.

6.1. Predicción con modelos estacionarios

Consideremos el modelo lineal general

$$Y_t \;=\; a_t \,+\, \psi_1\, a_{t-1} \,+\, \psi_2\, a_{t-2} \,+\, \psi_3\, a_{t-3} \,+\, \ldots \qquad t = 1, 2, \ldots \tag{6.3}$$

que es un modelo $MA(\infty)$ y el conjunto de información dado por (6.1). La estrategia de predicción se va a basar en escribir el valor que se desea predecir, $Y_{T+\ell}$, tal y como se genera en función del modelo para luego obtener la predicción óptima calculando la esperanza condicionada al conjunto de información. Para la representación medias móviles general, $Y_{T+\ell}$ viene dado por:

$$Y_{T+\ell} \;=\; a_{T+\ell} + \psi_1\, a_{T+\ell-1} + \psi_2\, a_{T+\ell-2} + \ldots + \psi_{\ell-1}\, a_{T+1} + \psi_\ell\, a_T + \psi_{\ell+1}\, a_{T-1} + \psi_{\ell+2}\, a_{T-2} + \ldots$$

Tomando la esperanza condicionada al conjunto de información, se obtiene:

$$Y_T(\ell) \;=\; E_T[Y_{T+\ell}] \;=\; \psi_\ell\, a_T \,+\, \psi_{\ell+1}\, a_{T-1} \,+\, \psi_{\ell+2}\, a_{T-2} \,+\, \psi_{\ell+3}\, a_{T-3} \,+\, \ldots$$

dado que:

$$E_T(a_{T+j}) \;=\; \begin{cases} a_{T+j} & j \leq 0 \\ E(a_{T+j}) = 0 & j > 0 \end{cases}$$

La perturbación a_t es la innovación en el momento t. Si, dado el conjunto de información I_T, se conoce el verdadero valor de Y_t, como la parte sistemática se puede predecir mediante el modelo, la perturbación $a_t = Y_t - \widehat{PS}_t$ está determinada, es fija. Si, dado I_T, no se conoce el verdadero valor de Y_t, entonces la innovación a_t no está determinada por el conjunto de información, con lo que su media condicionada será la misma que su media no condicionada, es decir, cero.

Los errores de predicción son:

$$\begin{aligned}
e_T(1) &= Y_{T+1} - Y_T(1) \;=\; a_{T+1} + \psi_1 a_T + \psi_2 a_{T-1} + \psi_3 a_{T-2} + \ldots - \\
&- (\psi_1 a_T + \psi_2 a_{T-1} + \psi_3 a_{T-3} + \ldots) \;=\; a_{T+1} \\
e_T(2) &= Y_{T+2} - Y_T(2) \;=\; a_{T+2} + \psi_1 a_{T+1} + \psi_2 a_T + \psi_3 a_{T-1} + \ldots - \\
&- (\psi_2 a_T + \psi_3 a_{T-1} + \psi_4 a_{T-2} + \ldots) \;=\; a_{T+2} + \psi_1 a_{T+1} \\
\ldots &= \ldots \\
e_T(\ell) &= Y_{T+\ell} - Y_T(\ell) \;=\; a_{T+\ell} + \psi_1 a_{T+\ell-1} + \psi_2 a_{T+\ell-2} + \ldots + \psi_{\ell-1} a_{T+1} + \\
&+ \psi_\ell a_T + \psi_{\ell+1} a_{T-1} + \psi_{\ell+2} a_{T-2} + \ldots - \\
&- (\psi_\ell a_T + \psi_{\ell+1} a_{T-1} + \psi_{\ell+2} a_{T-2} + \psi_{\ell+3} a_{T-3} + \ldots) = \\
&= a_{T+\ell} + \psi_1 a_{T+\ell-1} + \psi_2 a_{T+\ell-2} + \ldots + \psi_{\ell-1} a_{T+1}
\end{aligned}$$

Los errores de predicción son una combinación lineal de las perturbaciones futuras $a_{T+\ell}, \ell = 1, 2, \ldots$ con valor medio cero:

$$E_T(e_T(\ell)) \;=\; E_T[a_{T+\ell} + \psi_1 a_{T+\ell-1} + \psi_2 a_{T+\ell-2} + \ldots + \psi_{\ell-1} a_{T+1}] \;=\; 0$$

La varianza del error de predicción o Error Cuadrático Medio de Predicción viene dado por:

$$\begin{aligned}
V_T(e_T(1)) &= E_T[e_T(1)]^2 \;=\; E_T[a_{T+1}]^2 \;=\; \sigma^2 \\
V_T(e_T(2)) &= E_T[e_T(2)]^2 \;=\; E_T[a_{T+2} + \psi_1 a_{T+1}]^2 \;=\; (1+\psi_1^2)\,\sigma^2 \\
\ldots &= \ldots \\
V(e_T(\ell)) &= E_T[e_T(\ell)]^2 \;=\; E_T[a_{T+\ell} + \psi_1 a_{T+\ell-1} + \psi_2 a_{T+\ell-2} + \ldots + \psi_{\ell-1} a_{T+1}]^2 \;= \\
&= (1+\psi_1^2+\psi_2^2+\ldots+\psi_{\ell-1}^2)\,\sigma^2 \;=\; \sigma^2 \sum_{i=0}^{\ell-1} \psi_i^2 \qquad (6.4)
\end{aligned}$$

Como se puede observar la varianza del error de predicción va creciendo conforme nos alejamos en el futuro. Ahora bien, si el proceso es estacionario se cumple que:

$$\sum_{i=1}^{\infty} \psi_i^2 < \infty$$

por lo que esta varianza no crece indefinidamente, sino que tiene una cota máxima finita.

Se puede observar que la perturbación o innovación a_t y su varianza σ^2 tienen una nueva interpretación:

- $a_t = Y_t - Y_{t-1}(1)$, es el error de predicción un periodo hacia adelante.
- $V(a_t) = \sigma^2$, es la varianza del error de predicción un periodo hacia adelante.

Si el proceso ruido blanco sigue una distribución normal, se tiene que:

$$e_T(\ell) = Y_{T+\ell} - Y_T(\ell) \;\sim\; N[0,\, V(e_T(\ell))]$$

Por lo que el intervalo de predicción de probabilidad $(1-\alpha)\,\%$ es:

$$\begin{aligned}
Y_{T+1}: &\quad \left[Y_T(1) \,-\, N_{\alpha/2}\sqrt{\sigma^2} \quad ; \quad Y_T(1) \,+\, N_{\alpha/2}\sqrt{\sigma^2}\right] \\
Y_{T+2}: &\quad \left[Y_T(2) \,-\, N_{\alpha/2}\sqrt{\sigma^2\,(1+\psi_1^2)} \quad ; \quad Y_T(2) \,+\, N_{\alpha/2}\sqrt{\sigma^2\,(1+\psi_1^2)}\right] \\
\vdots &\quad \vdots \\
Y_{T+\ell}: &\quad \left[Y_T(\ell) \,-\, N_{\alpha/2}\sqrt{\sigma^2 \sum_{i=0}^{\ell-1}\psi_i^2} \quad ; \quad Y_T(\ell) \,+\, N_{\alpha/2}\sqrt{\sigma^2 \sum_{i=0}^{\ell-1}\psi_i^2}\right]
\end{aligned}$$

6.1.1. Predicción con modelos MA(q).

Comencemos por un modelo de medias móviles sencillo, por ejemplo, el $MA(2)$ de media cero:

$$Y_t \,=\, a_t \,-\, \theta_1\, a_{t-1} \,-\, \theta_2\, a_{t-2} \qquad a_t \sim RBN(0,\sigma^2) \qquad t = 1, 2, \ldots$$

La función de predicción es:

$$\begin{aligned}
Y_{T+1} \;&=\; a_{T+1} \,-\, \theta_1\, a_T \,-\, \theta_2\, a_{T-1} \\
&\qquad Y_T(1) \,=\, E_T[Y_{T+1}] \,=\, E_T[a_{T+1} \,-\, \theta_1\, a_T \,-\, \theta_2\, a_{T-1}] \,=\, -\,\theta_1\, a_T \,-\, \theta_2\, a_{T-1} \\
Y_{T+2} \;&=\; a_{T+2} \,-\, \theta_1\, a_{T+1} \,-\, \theta_2\, a_T \\
&\qquad Y_T(2) \,=\, E_T[Y_{T+2}] \,=\, E_T[a_{T+2} \,-\, \theta_1\, a_{T+1} \,-\, \theta_2\, a_T] \,=\, -\,\theta_2\, a_{T-1}
\end{aligned}$$

$$Y_{T+3} \;=\; a_{T+3} \,-\, \theta_1\, a_{T+2} \,-\, \theta_2\, a_{T+1}$$

$$Y_T(3) \,=\, E_T[Y_{T+3}] \,=\, E_T[a_{T+3} \,-\, \theta_1\, a_{T+2} \,-\, \theta_2\, a_{T+1}] = 0$$

$$Y_T(\ell) \,=\, E_T[Y_{T+\ell}] \,=\, 0 \;\; (= E(Y_t)) \quad \forall\, \ell > 2$$

Por lo tanto, la función de predicción de un $MA(2)$, depende del conjunto de información, I_T , para $\ell = 1, 2$. A partir de $\ell > 2$, la predicción óptima viene dada por la media del proceso.

Estos resultados se pueden generalizar fácilmente para el modelo $MA(q)$:

$$Y_t \,=\, a_t \,-\, \theta_1\, a_{t-1} \,-\, \theta_2\, a_{t-2} \,-\, \ldots \,-\, \theta_q\, a_{t-q} \qquad a_t \sim RBN(0, \sigma^2) \qquad t = 1, 2, \ldots$$

La función de predicción es:

$$Y_T(\ell) \,=\, \begin{cases} Y_T(1) = & -\,\theta_1\, a_T \,-\, \theta_2\, a_{T-1} \,-\, \ldots \,-\, \theta_q\, a_{T+1-q} \\ Y_T(2) = & -\,\theta_2\, a_T \,-\, \theta_3\, a_{T-2} - \, \ldots \,-\, \theta_q\, a_{T+2-q} \\ \cdots & \cdots \\ Y_T(q) = & -\,\theta_q\, a_T \\ Y_T(\ell) = & 0 \quad \forall\, \ell = q+1, q+2, \ldots \end{cases}$$

Como el modelo $MA(2)$ está escrito directamente en forma medias móviles, se obtiene la varianza del error de predicción aplicando la expresión (6.4) con $\psi_i = \theta_i$, $i = 1, 2, \ldots, q$ y $\psi_i = 0$, $\forall i > q$:

$$V(e_T(\ell)) \,=\, \begin{cases} V(e_T(1)) = & \sigma^2 \\ V(e_T(2)) = & (1 + \theta_1^2)\, \sigma^2 \\ \ldots & \ldots \\ V(e_T(q)) = & (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_{q-1}^2)\, \sigma^2 \\ V(e_T(\ell)) = & (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2)\; \sigma^2 \;\; (= V(Y_t)) \quad \ell = q+1, q+2, \ldots \end{cases}$$

Aunque la varianza del error de predicción es una función creciente de ℓ, el horizonte de predicción, tiene una cota máxima que viene dada por la varianza no condicionada del proceso y que se alcanza para $\ell = q$.

Se puede concluir que para un modelo $MA(q)$ las predicciones para los q primeros horizontes de predicción, $\ell = 1, 2, \ldots, q$, dependen del conjunto de información a través de los errores de predicción un periodo hacia adelante $a_T, a_{T-1}, \ldots, a_{T+1-q}$, con lo que se mejora la predicción respecto de la media no condicionada del proceso porque se predice con una varianza del error de predicción menor que la varianza no condicionada del proceso.

A partir de $\ell = q$, el conjunto de información no aporta nada a la predicción porque las predicciones óptimas son la media no condicionada del proceso y la varianza del error de predicción es la varianza no condicionada del proceso. Esto significa que, condicionando al conjunto de información, se obtienen los mismos resultados que sin condicionar, luego a partir de $\ell = q$, I_T ya no es informativo.

La predicción por intervalo viene dada por:

$$\begin{array}{ll}
\ell = 1 & \left[-\theta_1 a_T - \theta_2 a_{T-1} - \ldots - \theta_q a_{T+1-q} \pm N_{\alpha/2}\sqrt{\sigma^2}\right] \\
\ell = 2 & \left[-\theta_2 a_T - \theta_3 a_{T-2} - \ldots - \theta_q a_{T+2-q} \pm N_{\alpha/2}\sqrt{\sigma^2 (1+\theta_1^2)}\right] \\
\ldots & \ldots \\
\ell = q & \left[-\theta_q a_T \pm N_{\alpha/2}\sqrt{\sigma^2 \,(1+\theta_1^2+\ldots+\theta_{q-1}^2)}\right] \\
\ell > q & \left[0 \pm N_{\alpha/2}\sqrt{\sigma^2 \,(1+\theta_1^2+\ldots+\theta_{q-1}^2+\theta_q^2)}\right]
\end{array}$$

La amplitud de los intervalos de predicción va creciendo con ℓ, con el límite impuesto por

$$\pm N_{\alpha/2}\sqrt{\sigma^2 \,(1+\theta_1^2+\ldots+\theta_{q-1}^2+\theta_q^2)} \;=\; \pm N_{\alpha/2}\sqrt{V(Y_t)}$$

6.1.2. Predicción con modelos AR(p)

Consideremos el modelo autorregresivo más sencillo, el $AR(1)$.

$$Y_t = \phi Y_{t-1} + a_t \qquad a_t \sim RBN(0, \sigma^2) \qquad t = 1, 2, \ldots$$

La función de predicción es:

$$\begin{aligned}
Y_{T+1} &= \phi Y_T + a_{T+1} \\
& \quad Y_T(1) = E_T[Y_{T+1}] = E_T[\phi Y_T + a_{T+1}] = \phi Y_T \\
Y_{T+2} &= \phi Y_{T+1} + a_{T+2} \\
& \quad Y_T(2) = E_T[Y_{T+2}] = E_T[\phi Y_{T+1} + a_{T+2}] = \phi E_T[Y_{T+1}] = \phi Y_T(1) \\
Y_{T+3} &= \phi Y_{T+2} + a_{T+3} \\
& \quad Y_T(3) = E_T[Y_{T+3}] = E_T[\phi Y_{T+2} + a_{T+3}] = \phi E_T[Y_{T+2}] = \phi Y_T(2)
\end{aligned}$$

De forma que la función de predicción es:

$$Y_T(\ell) = \phi Y_T(\ell), \quad \ell = 1, 2, 3, \ldots \tag{6.5}$$

dado que:

$$E_T(Y_{T+j}) = \begin{cases} Y_{T+j} & j \leq 0 \\ E(Y_T(j)) & j > 0 \end{cases}$$

La función de predicción (6.5) recoge una regla de cadena para obtener las predicciones de un proceso autorregresivo unas en función de las otras hasta un futuro indefinido. La trayectoria de la función de predicción depende de la estructura de la parte autorregresiva:

$$\begin{aligned} Y_T(1) &= \phi\, Y_T \\ Y_T(2) &= \phi\, Y_T(1) = \phi\,\phi\, Y_T = \phi^2\, Y_T \\ Y_T(3) &= \phi\, Y_T(2) = \phi\,\phi^2\, Y_T = \phi^3\, Y_T \\ Y_T(4) &= \phi\, Y_T(3) = \phi\,\phi^3\, Y_T = \phi^4\, Y_T \\ Y_T(\ell) &= \phi^\ell\, Y_T \quad \ell = 1, 2, 3, \ldots \end{aligned}$$

Como el proceso autorregresivo es estacionario, $|\phi| < 1$, y por lo tanto cuando nos alejamos en el futuro la función de predicción tiende hacia la media no condicionada del proceso:

$$\lim_{\ell \to \infty} Y_T(\ell) = 0 \; (= E(Y_t))$$

Para construir los intervalos de predicción, se ha de obtener la varianza del error de predicción. Para ello es preciso partir del modelo escrito en forma medias móviles. En el caso del $AR(1)$:

$$\begin{aligned} (1 - \phi\, L)\, Y_t = a_t \;\; &\rightarrow \;\; Y_t = \frac{1}{1 - \phi\, L}\, a_t \\ &\rightarrow \;\; Y_t = (1 + \phi\, L + \phi^2\, L^2 + \phi^3\, L^3 + \ldots)\, a_t \\ &\rightarrow \;\; Y_t = a_t + \phi\, a_{t-1} + \phi^2\, a_{t-2} + \phi^3\, a_{t-3} + \ldots \end{aligned}$$

Por lo que la varianza del error de predicción se obtiene aplicando la fórmula general (6.4) con $\psi_1 = \phi^i, \; \forall i:$

$$V(e_T(\ell)) = \begin{cases} V(e_T(1)) = & \sigma^2 \\ V(e_T(2)) = & (1 + \phi^2)\, \sigma^2 \\ V(e_T(3)) = & (1 + \phi^2 + (\phi^2)^2)\, \sigma^2 \\ V(e_T(4)) = & (1 + \phi^2 + (\phi^2)^2 + (\phi^3)^2)\, \sigma^2 \\ V(e_T(\ell)) = & (1 + \phi^2 + (\phi^2)^2 + \ldots + (\phi^{\ell-1})^2)\, \sigma^2 \end{cases}$$

La varianza del error de predicción es monotonamente creciente conforme nos alejamos en el futuro. Como el proceso es estacionario, esta varianza no crece indefinidamente sino que tiene una cota superior dada por la varianza no condicionada del proceso:

$$\lim_{\ell \to \infty} V(e_T(\ell)) = \lim_{\ell \to \infty} \sigma^2 \left[1 + \phi^2 + (\phi^2)^2 + \ldots + (\phi^{\ell-1})^2\right] = \frac{\sigma^2}{1-\phi^2} = V(Y_t)$$

La predicción por intervalo es:

$$\begin{array}{ll}
\ell = 1 & \left[\phi\, Y_T \;\pm\; N_{\alpha/2}\, \sqrt{\sigma^2}\right] \\
\ell = 2 & \left[\phi\, Y_T(1) \;\pm\; N_{\alpha/2}\, \sqrt{\sigma^2\,(1+\phi^2)}\right] \\
\ell = 3 & \left[\phi\, Y_T(2) \;\pm\; N_{\alpha/2}\, \sqrt{\sigma^2\;(1+\phi^2+(\phi^2)^2)}\right] \\
\ldots & \ldots \\
\ell & \left[\phi\, Y_T(\ell-1) \;\pm\; N_{\alpha/2}\, \sqrt{\sigma^2\;(1+\phi^2+(\phi^2)^2+\ldots+(\phi^{\ell-1})^2)}\right]
\end{array}$$

La amplitud de los intervalos de predicción va creciendo con ℓ, con el límite impuesto por

$$\pm N_{\alpha/2} \sqrt{\frac{\sigma^2}{1-\phi^2}} = \pm N_{\alpha/2} \sqrt{V(Y_t)}$$

Los resultados obtenidos para el modelo $AR(1)$ se pueden extender para el modelo $AR(p)$. En general, las funciones de predicción de procesos autorregresivos puros, se obtendrán a partir de reglas de cadena:

$$Y_T(\ell) = \phi_1\, Y_T(\ell-1) + \phi_2\, Y_T(\ell-2) + \phi_3\, Y_T(\ell-3) + \ldots + \phi_p\, Y_T(\ell-p), \quad \ell = 1, 2, 3, \ldots$$

La función de predicción de un proceso $AR(1)$ utiliza la última observación Y_T para obtener la predicción un periodo hacia adelante y luego, a partir de ésta, se obtienen el resto de las predicciones. En el caso de un autorregresivo de orden p autorregresivo de orden p, se utilizarán las p últimas observaciones para obtener las predicciones para $\ell = 1, 2, \ldots, p$, y el resto se obtienen a partir de las p primeras.

6.1.3. Predicción con modelos ARMA(p,q).

Consideremos un modelo $ARMA(p, q)$ sencillo, el $ARMA(1, 2)$:

$$Y_t = \delta + \phi\, Y_{t-1} + a_t - \theta_1\, a_{t-1} - \theta_2\, a_{t-2} \qquad a_t \sim RBN(0, \sigma^2) \qquad t = 1, 2, \ldots \quad (6.6)$$

La media de este proceso no es cero si $\delta \neq 0$:

$$E(Y_t) \,=\, \frac{\delta}{1-\phi}$$

Las predicciones por punto son:

$$\begin{aligned}
Y_{T+1} \;&=\; \delta \,+\, \phi\, Y_T \,+\, a_{T+1} \,-\, \theta_1\, a_T \,-\, \theta_2\, a_{T-1} \\
&\quad Y_T(1) \,=\, E_T[Y_{T+1}] \,=\, E_T[\delta \,+\, \phi\, Y_T \,+\, a_{T+1} \,-\, \theta_1\, a_T \,-\, \theta_2\, a_{T-1}] \,= \\
&\quad = \delta \,+\, \phi\, Y_T \,-\, \theta_1\, a_T \,-\, \theta_2\, a_{T-1} \\
Y_{T+2} \;&=\; \delta \,+\, \phi\, Y_{T+1} \,+\, a_{T+2} \,-\, \theta_1\, a_{T+1} \,-\, \theta_2\, a_T \\
&\quad Y_T(2) \,=\, E_T[Y_{T+2}] \,=\, E_T[\delta \,+\, \phi\, Y_{T+1} \,+\, a_{T+2} \,-\, \theta_1\, a_{T+1} \,-\, \theta_2\, a_T] \,= \\
&\quad = \delta \,+\, \phi\, Y_T(1) \,-\, \theta_2\, a_T \\
Y_{T+3} \;&=\; \delta \,+\, \phi\, Y_{T+2} \,+\, a_{T+3} \,-\, \theta_1\, a_{T+2} \,-\, \theta_2\, a_{T+1} \\
&\quad Y_T(3) \,=\, E_T[Y_{T+3}] \,=\, E_T[\delta \,+\, \phi\, Y_{T+2} \,+\, a_{T+3} \,-\, \theta_1\, a_{T+2} \,-\, \theta_2\, a_{T+1}] \,= \\
&\quad = \,\delta \,+\, \phi\, Y_T(2) \\
\rightarrow &\quad Y_T(\ell) \,=\, E_T[Y_{T+\ell}] \,=\, \delta \,+\, \phi\, Y_T(\ell-1) \quad \forall\, \ell > 2
\end{aligned}$$

La estructura de la función de predicción es la siguiente. Las dos primeras predicciones dependen de la última observación Y_T (parte autorregresiva) y de los últimos errores de predicción un periodo hacia adelante a_T y a_{T-1} (parte medias móviles). Para $\ell > 2$, la parte medias móviles no aparece de forma explícita en la función de predicción, y cada predicción se va obteniendo de las anteriores siguiendo una *regla en cadena* marcada por la parte autorregresiva. Esta función se va acercando a la media del proceso conforme nos alejamos en el futuro:

$$\begin{aligned}
Y_T(3) \;&=\; \delta \,+\, \phi\, Y_T(2) \\
Y_T(4) \;&=\; \delta \,+\, \phi\, Y_T(3) \,=\, \delta \,+\, \phi\,(\delta \,+\, \phi\, Y_T(2)) \,=\, \delta(1+\phi) + \phi^2\, Y_T(2) \\
Y_T(5) \;&=\; \delta \,+\, \phi\, Y_T(4) \,=\, \delta \,+\, \phi\,(\delta(1+\phi) + \phi^2\, Y_T(2)) \,=\, \delta(1+\phi+\phi^2) + \phi^3\, Y_T(2) \\
\ldots \;&\qquad \ldots \\
Y_T(\ell) \;&=\; \delta \,+\, \phi\, Y_T(\ell-1) \,=\, \delta(1+\phi+\phi^2+\ldots+\phi^{\ell-3}) + \phi^{\ell-2}\, Y_T(2)
\end{aligned}$$

de forma que como el modelo $ARMA(2,1)$ es estacionario, $|\phi| < 1$ y:

$$\lim_{\ell\to\infty} Y_T(\ell) \,=\, \delta \sum_{i=0}^{\ell-3} \phi^i \,=\, \frac{\delta}{1-\phi}\,(=\, E(Y_t))$$

Para obtener la varianza del error de predicción y, por lo tanto, las predicciones por intervalo, se deriva la representación medias móviles infinita:

$$(1-\phi L)\,Y_t = (1-\theta_1 L-\theta_2 L^2)\,a_t$$

$$\rightarrow \quad Y_t \;=\; \frac{1-\theta_1 L-\theta_2 L^2}{1-\phi L}\,a_t = \;(1+\psi_1 L+\psi_2 L^2+\psi_3 L^3+\ldots)\,a_t$$

de donde:

$$\frac{1-\theta_1 L-\theta_2 L^2}{1-\phi L} \;=\; 1\,+\,\psi_1\,L\,+\,\psi_2\,L^2\,+\,\psi_3\,L^3\,+\,\ldots$$

$$\rightarrow \quad 1-\theta_1 L-\theta_2 L^2 \;=\; (1-\phi L)(1\,+\,\psi_1\,L\,+\,\psi_2\,L^2\,+\,\psi_3\,L^3\,+\,\ldots)$$

e igualando coeficientes:

$$\begin{array}{lllll}
L & -\theta_1\,L=(\psi_1-\phi)\,L & \Rightarrow \quad -\theta_1=\psi_1-\phi & \Rightarrow & \psi_1=\phi-\theta_1 \\
L^2 & -\theta_2\,L^2=(\psi_2-\phi\psi_1)\,L^2 & \Rightarrow \; -\theta_2=\psi_2-\phi\psi_1 & \Rightarrow & \psi_2=\phi\psi_1-\theta_2=\phi(\phi-\theta_1)-\theta_2 \\
L^3 & 0\,L^3=(\psi_3-\psi_2\,\phi)\,L^3 & \Rightarrow \quad 0=\psi_3-\psi_2\,\phi & \Rightarrow & \psi_3=\phi\,\psi_2 \\
\ldots & \ldots & & &
\end{array}$$

Los pesos de la forma medias móviles infinita son:

$$\psi_i \;=\; \begin{cases} k=0 & \psi_0=1 \\ k=1 & \psi_1=\phi-\theta_1 \\ k=2 & \psi_2=\phi\psi_1-\theta_2=\phi(\phi-\theta_1)-\theta_2 \\ k>2 & \psi_k=\phi\,\psi_{k-1} \end{cases}$$

con estos pesos se pueden construir los intervalos de predicción. Como el proceso $ARMA(1,2)$ es estacionario, la amplitud de los intervalos irá creciendo conforme nos alejamos en el futuro pero con una cota máxima dada por $\left[\pm\,N_{\alpha/2}\,\times\,\sqrt{V(Y_t)}\right]$.

6.1.4. Predicciones con modelos estacionarios estimados

Habitualmente no se conoce el proceso que ha generado la serie temporal Y_t por lo que hay que estimarlo con los datos disponibles, obteniendo:

$$\hat{\phi}_p(L)\,Y_t \;=\; \hat{\theta}_q(L)\,\hat{a}_t \qquad t=1,2,\ldots,T$$

donde $\hat{a}_t$ son los residuos del modelo, pero también una estimación del error de predicción un periodo hacia adelante.

Por ejemplo, en el caso del modelo (6.6), la función de predicción estimada sería:

$$Y_T(\ell) \;=\; \begin{cases} \ell = 1 & Y_T(1) = \hat{\delta} + \hat{\phi}\, Y_T - \hat{\theta}_1\, \hat{a}_T - \hat{\theta}_2\, \hat{a}_{T-1} \\ \ell = 2 & Y_T(2) = \hat{\delta} + \hat{\phi}\, Y_T(1) - \hat{\theta}_2\, \hat{a}_T \\ \ell > 2 & Y_T(\ell) = \hat{\delta} + \hat{\phi}\, Y_T(\ell - 1) \end{cases}$$

Ejemplo 6.1. Modelo $MA(2)$

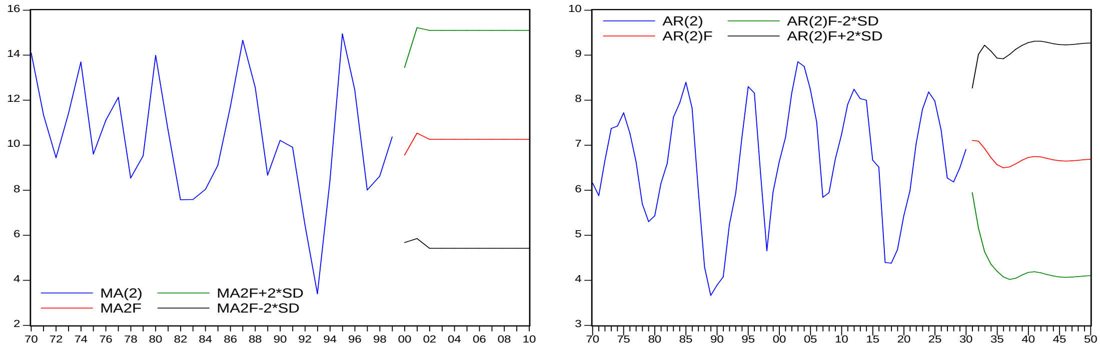
$$Y_t = 10,3 + \hat{a}_t + 0,67\, \hat{a}_{t-1} - 0,31\, \hat{a}_{t-2} \qquad t = 1, 2, \ldots, T$$

Función de predicción:

$$Y_T(\ell) \;=\; \begin{cases} \ell = 1 & Y_T(1) = 10,3 + 0,67\; \hat{a}_T - 0,31\, \hat{a}_{T-1} \\ \ell = 2 & Y_T(2) = 10,3 - 0,31\, \hat{a}_T \\ \ell > 2 & Y_T(\ell) = 10,3\, (= E(\hat{Y}_t)) \end{cases}$$

Todos los procesos medias móviles finitos son estacionarios y, como se puede observar en la figura de la izquierda del gráfico 6.1, a partir de $\ell = 2$ la función de predicción permanece constante en la media del proceso con una amplitud constante del intervalo de predicción.

Gráfico 6.1: Proceso MA(2) versus Proceso AR(2)



Ejemplo 6.2. Modelo $AR(2)$

$$Y_t = 2,4 + 1,33\, Y_{t-1} - 0,69\, Y_{t-2} + \hat{a}_t \qquad t = 1, 2, \ldots, T$$

Función de predicción:

$$Y_T(\ell) \;=\; \begin{cases} \ell = 1 & Y_T(1) = 2,4 + 1,33\, Y_T - 0,69\, Y_{T-1} \\ \ell = 2 & Y_T(2) = 2,4 + 1,33\, Y_T(1) - 0,69\, Y_T \\ \ell = 1, 2, 3, \ldots & Y_T(\ell) = 2,4 + 1,33\, Y_T(\ell - 1) - 0,69\, Y_T(\ell - 2) \end{cases}$$

El proceso $AR(2)$ es estacionario:

$$1 - 1,33\,L + 0,69\,L^2 = 0$$

$$\rightarrow \qquad L_1, L_2 \;=\; \frac{1,33 \pm \sqrt{1,33^2 - 4\,\times\,0,69}}{2\,\times\,0,69} \;=\; \frac{1,33 \pm \sqrt{-0,99}}{1,38} \;=\; \frac{1,33}{1,38} \pm \frac{\sqrt{0,99}}{1,38}\,i \;=\; a \pm b\,i$$

$$|L_1| = |L_1| = \sqrt{a^2\,+\,b^2} = \sqrt{\left(\frac{1,33}{1,38}\right)^2 + \left(\frac{\sqrt{0,99}}{1,38}\right)^2} = \sqrt{1,4486} = 1,2 > 1$$

Por lo que la función de predicción tiende a su media cuando nos alejamos en el futuro:

$$\lim_{\ell\to\infty} Y_T(\ell) \;=\; \hat{E}(Y_t) \;=\; \frac{\hat{\delta}}{1 - \hat{\phi}_1 - \hat{\phi}_2} \;=\; \frac{2,4}{1 - 1,33 + 0,69} \;=\; 6,67$$

Este resultado se puede observar claramente en la figura de la derecha del gráfico 6.1. La serie presenta un comportamiento cíclico que se refleja en la función de predicción que presenta un ciclo amortiguado antes de dirigirse sistemáticamente hacia la media del proceso.

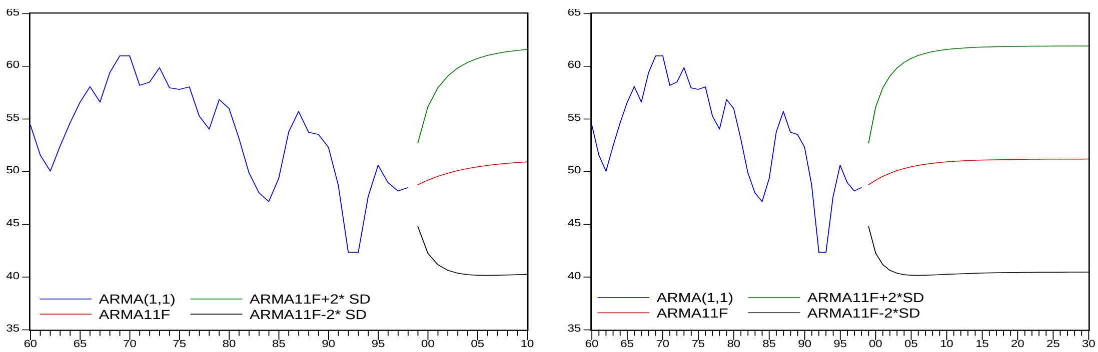
Ejemplo 6.3. Modelo $ARMA(1,1)$

$$Y_t = 9,3 + 0,82Y_{t-1} + \hat{a}_t + 0,62\,\hat{a}_{t-1} \qquad t = 1, 2, \ldots, T$$

La función de predicción es:

$$Y_T(\ell) \;=\; \begin{cases} \ell = 1 & Y_T(1) = 9,3 + 0,82\,Y_T + 0,62\,\hat{a}_T \\ \ell > 2 & Y_T(\ell) = 9,3 + 0,82\,Y_T(\ell - 1) \end{cases}$$

Gráfico 6.2: Función de predicción: ARMA(1,1) estacionario



Como el proceso es estacionario dado que cumple la condición de que la raíz del polinomio autorregresivo es, en valor absoluto, mayor que la unidad,

$$1 - 0,82\,L = 0 \quad \rightarrow \quad L = \frac{1}{0,82} = 1,22 \quad \rightarrow \quad |L| = |1,22| > 1$$

la función de predicción tiende a su media cuando nos alejamos en el futuro:

$$\lim_{\ell \to \infty} Y_T(\ell) \;=\; \hat{E}(Y_t) \;=\; \frac{\hat{\delta}}{1-\hat{\phi}} \;=\; \frac{9,3}{1-0,82} \;=\; 51,67$$

En las figuras del gráfico 6.2 se puede observar el comportamiento de la función de predicción por punto y por intervalo. Las últimas observaciones se encuentran por debajo de la media lo que lleva a que la función de predicción parta de un valor inferior a la media estimada del proceso. De forma que la trayectoria de la función de predicción es creciente al principio para dirigirse hacia la media del proceso donde se va estabilizando. Asimismo, se observa el comportamiento de los intervalos de predicción, de amplitud creciente al principio hasta estabilizarse con la varianza del proceso.

6.2. Predicción con modelos no estacionarios.

La predicción con modelos no estacionarios $ARIMA(p,d,q)$ se lleva a cabo de la misma manera que con los modelos estacionarios $ARMA(p,q)$. El predictor por punto óptimo de $Y_{T+\ell}$ viene dado por la esperanza condicionada al conjunto de información $Y_T(\ell) = E_T[Y_{T+\ell}]$. Para obtener esta esperanza condicionada basta con escribir el modelo en forma de ecuación en diferencias y obtener las esperanzas condicionadas, sabiendo que:

$$E_T[Y_{T+j}] \;=\; \begin{cases} Y_{T+j} & j \leq 0 \\ Y_T(j) & j > 0 \end{cases} \qquad\qquad E_T[a_{T+j}] \;=\; \begin{cases} a_{T+j} & j \leq 0 \\ 0 & j > 0 \end{cases}$$

Para construir los intervalos de predicción,

$$\left[Y_T(\ell) \;\pm\; N_{\alpha/2}\,\sqrt{V(e_T(\ell))} \right] \qquad \text{donde} \quad V(e_T(\ell)) = \sigma \sum_{j=0}^{\ell-1} \psi_j^2$$

el modelo ha de estar escrito en forma $MA(\infty)$ ya que ψ_j son los pesos del modelo $ARIMA$ escrito en forma medias móviles.

Para analizar las características de la función de predicción para modelos no estacionarios $ARIMA(p,d,q)$, consideremos varios ejemplos sencillos.

Ejemplo 6.4. Modelo $ARIMA(0,1,1)$.

Consideremos que la serie Y_t ha sido generada por el siguiente modelo:

$$\begin{aligned}(1-L)\,Y_t &= (1+\theta L)\,a_t \\ Y_t &= Y_{t-1} + a_t + \theta\, a_{t-1}\end{aligned}$$

La función de predicción es:

$$\begin{aligned}Y_T(1) &= E_T[Y_{T+1}] = Y_T + \theta\, a_T \\ Y_T(2) &= E_T[Y_{T+2}] = Y_T(1) = Y_T + \theta\, a_T \\ Y_T(3) &= E_T[Y_{T+3}] = Y_T(2) = Y_T(1) = Y_T + \theta\, a_T \\ \vdots\quad & \quad \vdots \\ Y_T(\ell) &= E_T[Y_{T+\ell}] = Y_T(\ell-1) = Y_T + \theta\, a_T \qquad \ell = 1,2,3,\ldots\end{aligned}$$

Por lo tanto, la función de predicción es una línea horizontal: pasa por la predicción un periodo hacia adelante, $Y_T(1)$, que depende del conjunto de información a través de Y_T y a_T y del parámetro del modelo, θ, y permanece allí conforme ℓ crece.

En el caso del modelo de paseo aleatorio, como $\theta = 0$, la función de predicción es:

$$Y_T(\ell) = E_T[Y_{T+\ell}] = Y_T \qquad \ell = 1,2,3,\ldots$$

Por lo tanto, las predicciones óptimas vienen dadas por la última observación independientemente del horizonte de predicción (veáse la figura izquierda del gráfico 6.3).

Se puede demostrar que si la serie Y_t ha sido generada por un proceso integrado de orden 1, de forma que se puede representar mediante un modelo $ARIMA(p,1,q)$ sin término independiente ($\delta = 0$), la función de predicción cuando $\ell \to \infty$, tiende a una constante:

$$Y_T(\ell) \overset{\ell\to\infty}{\longrightarrow} K^T$$

Hay que tener en cuenta que K^T no es la media del proceso porque como no es estacionario no tiene una media hacia dónde ir, sino que es una constante que depende del conjunto de información y de los parámetros AR y MA del modelo.

Obtengamos la función de predicción para un proceso integrado de orden 1 con constante, por ejemplo, el modelo de paseo aleatorio con deriva (**??**).

$$\begin{aligned}Y_T(1) &= E_T[Y_{T+1}] = Y_T + \delta \\ Y_T(2) &= E_T[Y_{T+2}] = Y_T(1) + \delta = Y_T + 2\,\delta \\ Y_T(3) &= E_T[Y_{T+3}] = Y_T(2) + \delta = Y_T + 3\,\delta \\ \vdots\quad & \quad \vdots \\ Y_T(\ell) &= E_T[Y_{T+\ell}] = Y_T(\ell-1) + \delta = Y_T + \ell\,\delta \qquad \ell = 1,2,3,\ldots\end{aligned}$$

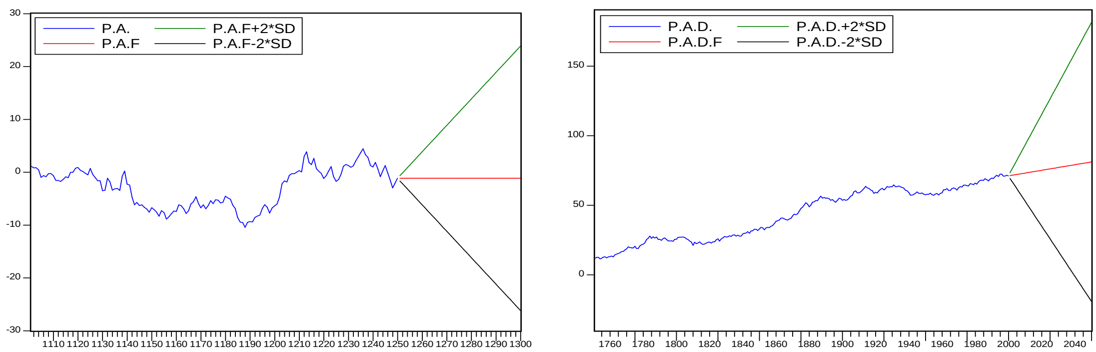
La función de predicción es una linea recta de pendiente δ (veáse la figura derecha del gráfico 6.3).

Se puede demostrar que si la serie Y_t ha sido generada por un proceso integrado de orden 1, de forma que se puede representar mediante un modelo $ARIMA(p,1,q)$ con término independiente ($\delta \neq 0$), la función de predicción tiende a una línea recta cuando $\ell \to \infty$,

$$Y_T(\ell) \quad \overset{\ell\to\infty}{\longrightarrow} \quad K^T + \delta\,\ell$$

La pendiente de la función de predicción viene dada por la constante δ y el intercepto, K^T, depende del conjunto de información y de los parámetros del modelo.

Gráfico 6.3: Modelos ARIMA: Funciones de predicción.



En lo que se refiere a la predicción por intervalo para modelos no estacionarios, las figuras del gráfico 6.3 muestran como la amplitud de los intervalos de predicción para los modelos $ARIMA(p,d,q)$ crece indefinidamente conforme el horizonte de predicción ℓ se hace mayor. Hay que tener en cuenta que cuando el proceso no es estacionario el límite $\lim_{\ell\to\infty} V[e_T(\ell)]$ no existe.

Para calcular la $V[e_T(\ell)]$ de un modelo no estacionario $ARIMA(p,d,q)$ es necesario escribir el modelo en forma $MA(\infty)$. Pongamos como ejemplo el modelo de paseo aleatorio (**??**) que, escrito en forma $MA(\infty)$ queda como sigue:

$$Y_t = a_t + a_{t-1} + a_{t-2} + a_{t-3} + \dots \qquad \text{de forma que} \quad \psi_j = 1 \quad \forall j$$

La varianza del error de predicción es:

$$
\begin{aligned}
V[e_T(1)] &= \sigma^2 \sum_{j=0}^{0} \psi_j^2 \,=\, \sigma^2 \\
V[e_T(2)] &= \sigma^2 \sum_{j=0}^{1} \psi_j^2 \,=\, \sigma^2\,(1+1) \,=\, 2\,\sigma^2 \\
V[e_T(3)] &= \sigma^2 \sum_{j=0}^{2} \psi_j^2 \,=\, \sigma^2\,(1+1+1) \,=\, 3\,\sigma^2 \\
\vdots\quad &\quad \vdots \\
V[e_T(\ell)] &= \sigma^2 \sum_{j=0}^{\ell-1} \psi_j^2 \,=\, \sigma^2\,(1+1+...+1) \,=\, \ell\,\sigma^2 \qquad \ell = 1, 2, 3, \ldots
\end{aligned}
$$

que no tiene límite finito.

Por último, considerando el caso de procesos integrados de orden 2, $ARIMA(p, 2, q)$, su función de predicción final toma la forma:

$$Y_T(\ell) \;\xrightarrow{\ell \to \infty}\; K_1^T \,+\, K_2^T\,\ell$$

Es decir, conforme el horizonte de predicción aumenta y nos alejamos en el futuro la función de predicción tiende a una línea recta donde tanto el intercepto, K_1^T, como la pendiente, K_2^T, dependen del conjunto de información y de los parámetros del modelo. Por lo tanto, aunque la estructura de esta función de predicción es la misma que la de los modelos $ARIMA(p, 1, q)$ con término independiente (por ejemplo, la función de predicción del paseo aleatorio con deriva), esta función de predicción es más flexible.

Capítulo 7

Ejercicios

7.1. Cuestiones

C.1. ¿Cómo se especifica un modelo de componentes no observados? Describe el comportamiento de una serie temporal

- "sin tendencia"
- que presenta tendencia lineal
- que decrece en el tiempo con un componente estacional multiplicativo
- con una tendencia creciente y un componente estacional aditivo

C.2. ¿Qué es una serie desestacionalizada?

- Explica por qué puede ser una buena idea estimar la tendencia de una serie a partir de la serie desestacionalizada cuando trabajamos con series estacionales.
- Si un dato corresponde a un mes con un factor estacional menor que uno, el dato desestacionalizado es mayor que el valor actual de la observación.

C.3. ¿De dónde proviene el nombre de Métodos de Alisado Exponencial?

- ¿Por qué se le denomina también Métodos de Predicción "ad hoc"?
- En el Método de Alisado Exponencial Simple, ¿qué observaciones tienen mayor peso a la hora de estimar el nivel actual de la serie que queremos predecir?
- Para que tipo de series puede ser apropiado utilizar el Método de Alisado Exponencial Simple? ¿Y el Método de Holt-Winters? ¿Y el Método de Holt-Winters con estacionalidad?

- Explica intuitivamente como actualizamos el nivel y la pendiente en el método de Holt-Winters? ¿Y en el método de Holt-Winters con estacionalidad? ¿Y el componente estacional?

C.4. Describe dos procedimientos sencillos para predecir una serie temporal cuyo comportamiento es:

- $Y_t = T_t + I_t$
- $Y_t = T_t + S_t + I_t$
- $Y_t = T_t \times S_t \times I_t$

C.5. Modelos de Componentes No observados globales:

- Si una serie posee una tendencia cuadrática y un componente estacional muy estable aditivo ¿qué modelo de regresión podrías utilizar para predecir valores futuros de la serie?
- Para un modelo con estacionalidad aditiva, explica que recogen los coeficientes s_j que acompañan a las variables ficticias.

C.6. ¿Qué comportamiento esperas de los errores de predicción en los siguientes casos?

- Se ha utilizado un modelo con el componente estacional aditivo para predecir una serie que tiene una tendencia decreciente y un componente estacional multiplicativo.
- Se ha utilizado un modelo con el componente estacional aditivo para predecir una serie que tiene una tendencia decreciente y un componente estacional multiplicativo.

C.7. Explica las ventajas y desventajas de utilizar los modelos globales de regresión frente a los métodos de alisado exponencial y de medias móviles.

C.8. ¿Son ciertas o falsas las siguientes afirmaciones? ¿Por qué? Si fueran falsas modifica su redacción para convertirlas en ciertas.

- Una serie temporal "sin tendencia" fluctúa aleatoriamente en torno a un nivel medio que cambia rápidamente en el tiempo.
- Una serie temporal con tendencia lineal fluctúa aleatoriamente en torno a un nivel medio que cambia de forma lineal o curvilínea en el tiempo.
- Una serie temporal con tendencia cuadrática fluctúa aleatoriamente en torno a un nivel medio que decrece a una tasa creciente.
- Si una serie presenta un componente estacional multiplicativo , la amplitud del ciclo estacional permanece constante a lo largo del tiempo.
- Si una serie presenta una tendencia decreciente y un componente estacional aditivo, la magnitud del ciclo estacional decrece con el tiempo.
- El Método de Alisado Exponencial Simple
  - ▷ se utiliza cuando el patrón de comportamiento de los datos presenta una tendencia lineal.
  - ▷ se puede demostrar que el valor estimado del nivel es una combinación lineal de todas las observaciones pasadas de la serie.
  - ▷ la última observación es la que tiene menor importancia a la hora de estimar el nivel m_T, mientras que el resto de las observaciones tienen más importancia conforme nos alejamos más en el pasado.
  - ▷ valores grandes de la constante de alisamiento, α, hacen que el peso dado a las observaciones pasadas decaiga muy lentamente.
- El Método de alisado exponencial de Holt-Winters con estacionalidad
  - ▷ actualiza el componente de la pendiente combinando la estimación previa de la pendiente con la diferencia entre la estimación de la estacionalidad en el periodo corriente y la estimación de la estacionalidad en el periodo anterior.
  - ▷ el procedimiento apropiado para series con estacionalidad aditiva se puede aplicar a series con estacionalidad multiplicativa reemplazando las divisiones por restas y las sumas por multiplicaciones.

C.9. En un conocido boletín de coyuntura las tasas de crecimiento de las variables se calculan utilizando lo que allí denominan "tasas logarítmicas", ¿a qué se refiere?

7.2. Problemas

Problema 1. Contamos con la siguiente serie de ventas anuales de un conocido producto, medida en cientos de miles de unidades para los años 1987-2003:

Año	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
Ventas	174	154	175	221	200	234	230	249	262	293
Año	1997	1998	1999	2000	2001	2002	2003			
Ventas	270	291	299	327	317	337	336			

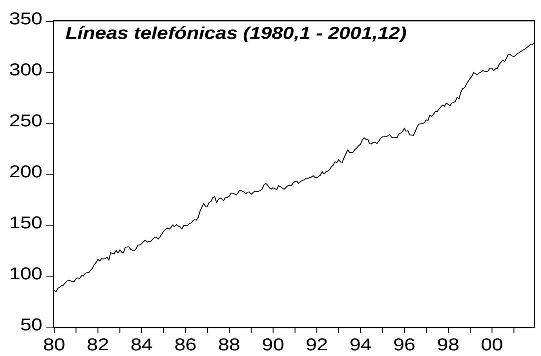
- ¿Qué modelo de componentes no observados es apropiado para la serie? Estima los parámetros del modelo con datos desde 1987 hasta 1996 únicamente. Los siguientes datos te pueden ser de utilidad: $t = 1, \ldots, 10$

  $\sum_{t=1}^{10} t = 55 \qquad \sum_{t=1}^{10} t^2 = 385 \qquad \sum_{t=1}^{10} y_t = 2192 \qquad \sum_{t=1}^{10} t\, y_t = 13185$

- Predice por punto y por intervalo los valores futuros de la serie desde el año 1997 hasta el 2003. Represéntalas gráficamente.

- Dado que conoces los verdaderos valores de las ventas para el periodo 1997-2003, calcula y representa gráficamente los errores de predicción. ¿Qué conclusión obtienes sobre la capacidad predictiva del modelo?

Problema 2. El responsable de una empresa del sector necesita unas previsiones del número de líneas telefónicas contratadas. Se dispone de datos mensuales desestacionalizados, desde enero de 1980 a diciembre de 2001 (en miles de líneas) que están representados en el siguiente gráfico:



- ¿Qué significa que los datos han sido *desestacionalizados*?

- A la vista de la evolución temporal de la serie de datos, ¿qué modelo de componentes no observados propondrías para predecirla? ¿por qué? Interpreta los componentes del modelo propuesto.

- Se han estimado por Mínimos Cuadrados Ordinarios dos modelos diferentes para esta serie con datos desde enero de 1980 a diciembre de 2000. Luego, con cada uno de ellos se han obtenido predicciones de la serie desde enero de 2001 hasta diciembre de 2001. Los resultados han sido los siguientes:

(1) $Y_t = T_t + I_t$ donde $\hat{T}_t = 88,96 + 0,838\ t$

$$R^2 = 0,984 \qquad \sum_{t=1}^{252} \hat{I}_t^2 = 15363,02 \qquad \sum_{\ell=1}^{12} \frac{1}{12}\, e_{00,12}(\ell)^2 = 281,9$$

(2) $Y_t = T_t + I_t$ donde $\hat{T}_t = 96,06 + 0,670\ t + 0,00066\ t^2$

$$R^2 = 0,986 \qquad \sum_{t=1}^{252} \hat{I}_t^2 = 12881,40 \qquad \sum_{\ell=1}^{12} \frac{1}{12}\, e_{00,12}(\ell)^2 = 76,04$$

▷ Interpreta los dos modelos propuestos para estimar los componentes de la serie. ¿Qué tipo de modelos son? ¿Qué tipo de comportamientos generan para la serie temporal?

▷ ¿Cuál de los dos modelos propuestos proporciona mejor ajuste dentro de la muestra?

▷ ¿Cuál es la función de predicción de cada modelo? ¿Qué modelo genera las mejores predicciones?

▷ ¿Qué modelo utilizarías para predecir Y_t? Calcula las predicciones de Y_t para el año 2002, sabiendo que re-estimando los dos modelos con datos de enero de 1980 a diciembre de 2001, se ha obtenido:

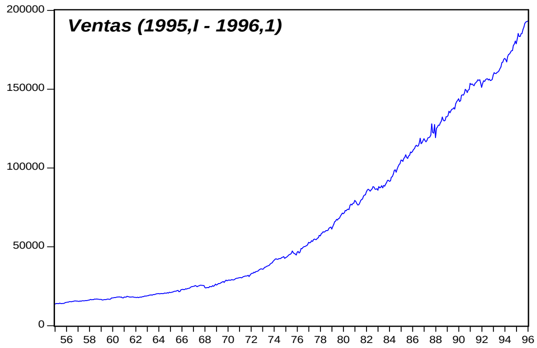
(1) $Y_t = T_t + I_t$ donde $\hat{T}_t = 87,53 + 0,854\ t$

$$R^2 = 0,984 \qquad \sum_{t=1}^{252} \hat{I}_t^2 = 18173,85$$

(2) $Y_t = T_t + I_t$ donde $\hat{T}_t = 96,06 + 0,670\ t + 0,00066\ t^2$

$$R^2 = 0,988 \qquad \sum_{t=1}^{252} \hat{I}_t^2 = 13482,65$$

Problema 3. El gráfico siguiente representa las ventas al por menor en EE.UU. de América contabilizadas en dólares corrientes. Los datos son mensuales, desde enero de 1955 a diciembre de 1994 y han sido previamente desestacionalizados[1].



- ¿Qué significa que los datos han sido *desestacionalizados*?
- ¿Te parece que las unidades de medida son las más adecuadas? ¿Aplicarías alguna corrección sobre la variable?

Respecto a la tendencia de la serie:

- A la vista del gráfico, ¿te parece correcto especificar una tendencia global determinista? ¿Por qué?
- ¿Qué función (o funciones) del tiempo elegirías para representar el comportamiento de la tendencia? ¿Por qué?

A continuación, puedes encontrar los ajustes por Mínimos Cuadrados Ordinarios de diferentes funciones para la tendencia.

a) Tendencia lineal: $Y_t = T_t + I_t$

$$\hat{T}_t = -16391,25 + 349,77t \qquad \bar{R}^2 = 0,898859 \qquad \sum_{t=1}^{468} \hat{I}_t^2 = 117x10^9$$

b) Tendencia parabólica: $Y_t = T_t + I_t$

$$\hat{T}_t = 18708,7 - 98,3t + 0,9t^2 \qquad \bar{R}^2 = 0,997 \qquad \sum_{t=1}^{468} \hat{I}_t^2 = 3,46x10^9$$

[1] Veáse Francis X. Diebold (1998). *Elements of forecasting*, Ed. South-Western, pp.91-99.

c) Tendencia exponencial: $\log Y_t = T_t + I_t$

$$\hat{T}_t = 9,39 + 0,006t \qquad \bar{R}^2 = 0,987 \qquad \sum_{t=1}^{468} \hat{I}_t^2 = 3,934$$

- Propón algún criterio para elegir entre las funciones ajustadas.
- Según ese criterio, ¿cuál te parece mejor, la tendencia lineal o la parabólica?
- ¿Te parece correcto utilizar la suma de cuadrado de los residuos u otra medida similar para elegir entre la tendencia exponencial y la parabólica? ¿Por qué?

Finalmente se ha estimado la tendencia exponencial directamente sobre la serie y_t, utilizando Mínimos Cuadrados No Lineales en lugar de Mínimos Cuadrados Ordinarios. Los resultados han sido:

Tendencia exponencial: $Y_t = T_t + I_t$

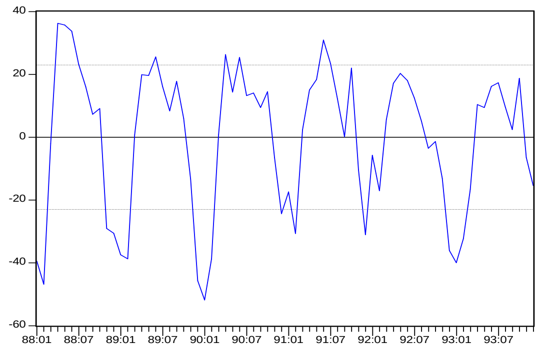
$$\hat{T}_t = 11967,8\, e^{0,00594t} \qquad \bar{R}^2 = 0,989 \qquad \sum_{t=1}^{468} \hat{I}_t^2 = 1,3x10^{10}$$

A la vista de estos resultados, ¿qué tendencia crees que se ajusta mejor, la parábola o la exponencial?

Problema 4. Con datos mensuales de enero de 1988 a diciembre 1993, se ha ajustado por Mínimos Cuadrados Ordinarios un modelo de regresión, obteniéndose:

$$\hat{Y}_t = 6 + 0,2t \qquad \sum_{t=1}^{T}(Y_t - \hat{Y}_t)^2 = \sum_{t=1}^{T} \hat{I}_t^2 = 743,07 \qquad T = 72$$

- ¿Cuáles son los elementos que forman parte de este modelo? Escribe el modelo para Y_t e interpreta los parámetros.
- Calcula la predicción por punto para los doce meses del año 1994.
- El siguiente gráfico representa los valores de los residuos del ajuste mínimo-cuadrático ordinario, $\hat{I}_t$. A la vista del gráfico, ¿crees que el modelo ajustado recoge correctamente el comportamiento de la serie? ¿por qué? Si tu respuesta es negativa, explica que modelo utilizarías.



Problema 5. Considera el siguiente modelo para la serie Y_t:

$$Y_t = a\, e^{bt}\, I_t, \qquad a > 0$$

- Interpreta el modelo. Interpreta el coeficiente b.
- ¿Qué comportamiento esperas si $b > 0$? ¿Y para $b < 0$?
- ¿Cómo se puede estimar el parámetro b?

Problema 6. Un fabricante de marionetas desea predecir las ventas mensuales de un determinado diseño. Este modelo lleva ya en el mercado unos cuantos años y ha observado que su demanda es bastante estable. Los datos disponibles son los siguientes:

	2000	2001	2002
Enero	423	350	330
Febrero	403	400	410
Marzo	474	470	408
Abril	451	311	514
Mayo	465	395	402
Junio	445	333	343
Julio	459	452	438
Agosto	325	414	419
Septiembre	365	310	374
Octubre	331	341	415
Noviembre	376	433	451
Diciembre	331	378	333

- ¿Cuál es el modelo de regresión lineal apropiado para representar la evolución de esta serie?
- Estima los parámetros del modelo por mínimos cuadrados ordinarios utilizando como conjunto de información los años 2000-2001. Interpreta los resultados.
- Predice por punto y por intervalo las ventas para el año 2002. Represéntalas gráficamente.
- Obtén los errores de predicción y comenta los resultados.

Problema 7. La siguiente tabla contiene datos anuales de las series *Tasa mortalidad infantil por mil nacimientos* (A), *Consumo per capita de pescado* (B) y *Producción de energía eléctrica* (C)[2].

TABLA III

Año	A	B	C	Año	A	B	C
1940	47,0	11,0	179	1966	23,7	10,9	1249
1941	45,3	11,2	208	1967	22,4	10,6	1317
1942	40,4	8,70	233	1968	21,8	11,0	1436
1943	40,3	7,9	267	1969	20,9	11,2	1552
1944	39,4	8,7	279	1970	20,0	11,8	1639
1945	38,3	9,9	271	1971	19,1	11,9	1718
1946	33,8	10,8	269	1972	18,5	12,4	1853
1947	32,2	10,3	307	1973	17,7	12,9	1959
1948	32,0	11,1	336	1974	16,7	12,2	1968
1949	31,3	10,9	345	1975	16,1	12,2	2003
1950	29,2	11,8	388	1976	15,2	12,9	2123
1951	28,4	11,2	433	1977	14,1	12,7	2212
1952	28,4	11,2	463	1978	13,8	13,4	2785
1953	27,8	11,4	514	1979	13,1	13,0	2319
1954	26,6	11,2	544	1980	12,6	12,8	2286
1955	26,4	10,5	629	1981	11,9	12,9	2295
1956	26,0	10,4	684	1982	11,5	12,3	2241
1957	26,3	10,2	716	1983	11,2	13,1	2310
1958	27,1	10,6	724	1984	10,8	13,7	2416
1959	26,4	10,9	797	1985	10,6	14,5	2469
1960	26,0	10,3	844				
1961	25,3	10,7	881				
1962	25,3	10,6	946				
1963	25,2	10,7	1011				
1964	24,8	10,5	1083				
1965	24,7	10,9	1157				

- Utilizando los datos de 1940-65 identifica y estima la tendencia de cada una de las series[3].
- Compara tus predicciones para el periodo 1966-1985 con los datos reales. ¿Han sido buenas tus predicciones? Si no, sugiere por qué.

Problema 8. La tabla recoge la serie de *Número de Turistas* llegados a un país. Considerando únicamente la serie para los años 1975-1982 (ambos inclusive),

[2]Los datos han sido tomados de C.W.J. Granger (1995). *Forecasting in business and economics.* Ed. Academic Press.

[3]No hay que fijarse en los datos postmuestrales al elegir la curva de ajuste.

- Especifica un modelo de regresión lineal que represente el comportamiento observado en la serie.
- ¿Cuáles son los factores estacionales estimados?. Calcula serie desestacionalizada. Represéntala gráficamente.
- ¿Cual es la tendencia estimada? Represéntala gráficamente.
- Obtén las predicciones por punto y por intervalo del número de turistas para los años 1983 y 1984. Represéntalas gráficamente.

	1 trim.	2 trim.	3 trim.	4 trim.
1975	224	570	800	512
1976	230	595	835	537
1977	238	622	869	560
1978	249	650	900	585
1979	260	675	932	605
1980	271	702	966	628
1981	280	724	1000	651
1982	291	748	1040	674
1983	300	772	1070	685
1984	312	800	1100	702

Supongamos que conocemos los datos para los años 83-84. Calcula los errores de predicción cometidos. Represéntalos gráficamente y comenta los resultados.

Problema 9. Con los datos de la demanda mensual de automóviles durante cuatro años recogidos en la tabla, obtén la función de predicción para $\ell = 1, 2, \ldots, 12$, mediante un modelo de regresión. Representa gráficamente las predicciones obtenidas.

	1970	1971	1972	1973
Enero	31	43	55	66
Febrero	30	42	54	67
Marzo	35	47	60	71
Abril	42	54	70	78
Mayo	45	57	69	81
Junio	52	64	76	88
Julio	58	70	81	93
Agosto	60	71	84	96
Septiembre	57	69	80	93
Octubre	51	65	76	87
Noviembre	52	60	75	88
Diciembre	47	61	71	83

Problema 10. Sea Y_t una serie temporal de ventas mensuales de un producto que sigue el siguiente comportamiento:

$$\text{Y} = \text{Tendencia} + \text{Estacionalidad} + \text{Irregular}$$

- Especifica un modelo de regresión con las siguientes características: una tendencia lineal y un componente estacional. Interpreta los parámetros del modelo propuesto
- Imagina que las ventas del producto aumentan en Semana Santa ¿Cómo recogerías en el modelo este hecho? (ayuda: considera su semejanza con un factor estacional, algo que se repite anualmente, pero que cada año se sitúa en época distinta)

Problema 11. Considera las dos series temporales siguientes:

Serie: Capturas anuales de bonito en toneladas (Cofradía de San Telmo)

t	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979
	5136	4604	5141	5613	5539	560	5562	5578	4891	4557
t	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
	5981	5744	5140	4798	4886	5321	4198	4517	5073	4821

Serie: Tráfico aéreo mensual en la CAPV

Mes	1	2	3	4	5	6	7	8	9	10	11	12
1985	77,1	92,8	103,2	126,9	105,1	125,5	138,2	146,4	124,8	155,2	160,3	171,5
1986	144,0	169,3	163,2	170,5	150,1	156,5	145,7	144,5	119,0	135,1	143,2	151,8

- ¿Qué modelo de series temporales propones para cada una de ellas?
- Explica detalladamente cómo predecirías los valores futuros de las mismas

Problema 12. Sea Y_t una serie trimestral compuesta por un nivel promedio constante y un componente estacional, es decir:

$$Y_t = a + S_t + I_t$$

Supón que utilizas para predecir la siguiente función de predicción:

$$Y_T(\ell) = \bar{Y}$$

donde $\bar{Y}$ es la media aritmética con las T observaciones disponibles de la serie. ¿Cómo esperarías que se comporten los errores de predicción?

Problema 13. Con los datos de la tabla que corresponden a las ventas trimestrales de un famoso juego en los almacenes El Barato:

- Calcula y representa las medias móviles simples de orden 5 de la serie.
- Calcula los factores estacionales de la serie mediante el método de la relación a la media móvil.
- Estima la tendencia que debemos usar para predecir esta serie.
- Predice las ventas para el año 4.

	1 trim.	2 trim.	3 trim.	4 trim.
1	20	25	35	44
2	28	29	43	48
3	24	37	39	56

Problema 14. El número de toneladas de acero producida por una empresa durante el último mes ascendió a 5250. El tonelaje pronosticado fue de 6000. La compañía utiliza un modelo de alisamiento exponencial simple con una constante de alisamiento de 0,3 para elaborar sus predicciones. ¿Cuál será la predicción de la compañía para el mes siguiente?

Problema 15. El estadístico de una empresa decide utilizar el siguiente modelo sencillo para predecir las ventas trimestrales de su producto, Y_t:

$$\begin{aligned} Y_t &= m_t + S_t + I_t \qquad (1) \\ m_t &= \alpha(Y_t - S_{t-4}) + (1-\alpha)m_{t-1} \\ S_t &= \gamma(Y_t - m_t) + (1-\gamma)S_{t-4} \end{aligned}$$

donde m_t es el nivel de la serie en el momento t, S_t son los factores estacionales (de forma que S_{1+4k} corresponden a los primeros trimestres de cada año, S_{2+4k} a los segundos, etc $k = 0, 1, 2, \ldots$) e I_t es el componente irregular correspondiente al momento t que se supone en promedio cero. Las ventas para los años 1994 y 1995 son:

Trimestre	1	2	3	4
1994	210	118	116	210
1995	214	210	118	212

Los factores estacionales S_t estimados en 1994 son:

Trimestre	1	2	3	4
s_t	8	0	-6	-2

y el valor de m_t estimado para el último trimestre de 1994 fue 116.

- ¿Qué tipo de comportamiento de la serie de ventas está recogiendo el modelo (1)?
- Si $\alpha = 0{,}5$ y $\gamma = 0{,}5$,
  - ▷ Actualiza las estimaciones tanto del nivel como de los factores estacionales de la serie utilizando la información que tienes sobre el año 1975.
  - ▷ Predice las ventas para cada trimestre de los años 1976 y 1977.

Problema 16. Supongamos el siguiente Modelo de Descomposición en Componentes:

$$Y_t = T_t + S_t + I_t$$

- Deriva el método de la "Relación a la media móvil" para estimar los factores estacionales.
- Bajo el supuesto de que la tendencia es lineal, escribe las recursiones para el nivel, la pendiente y el componente estacional para el método de predicción de Holt-Winters con estacionalidad. ¿Cual es la función de predicción?

Supongamos que contamos con los siguientes datos para la serie mensual de consumo de energía eléctrica de una ciudad:

	1997	1998	1999
Enero	31	43	55
Febrero	30	42	54
Marzo	35	47	60
Abril	42	54	70
Mayo	45	57	69
Junio	52	64	76
Julio	58	70	81
Agosto	60	71	84
Septiembre	57	69	80
Octubre	51	65	76
Noviembre	52	60	75
Diciembre	47	61	71

Basándote en el gráfico de la serie, ¿crees que el modelo propuesto es el apropiado para representar en comportamiento del consumo de energía eléctrica? ¿por qué?

Tomando como referencia el modelo propuesto:

- Estima los factores estacionales mediante el método de la relación a la media móvil.
- Estima el nivel y la pendiente de la tendencia lineal mediante un procedimiento adecuado.
- Predice el consumo de energía eléctrica de esta ciudad para cada mes del año 2000.

Apéndice A

Modelo de Regresión Lineal

El Modelo de Regresión Lineal General supone una relación lineal entre la variable endógena Y_t y un conjunto de variables explicativas $x_{j,t}, j = 1, 2, \ldots, k$:

$$Y_t \;=\; \beta_0 \;+\; \beta_1\, x_{1,t} \;+\; \beta_2\, x_{2,t} \;+\; \ldots \;+\; \beta_k\, x_{k,t} \;+\; u_t \qquad t = 1, 2, \ldots, T$$

Este modelo se puede escribir en forma matricial para cada una de las observaciones:

$$Y_t \;=\; \beta'\, X_t' \;+\; u_t \qquad t = 1, 2, \ldots, T$$

donde $\beta' = [\beta_0, \beta_1, \beta_2, \ldots, \beta_k]$ y $X_t = [1 \;\; x_{1,t} \;\; x_{2,t} \;\; \ldots \;\; x_{k,t}]$.

Para todas las observaciones queda de la siguiente manera:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} & \ldots & x_{k,1} \\ 1 & x_{1,2} & x_{2,2} & \ldots & x_{k,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1,T} & x_{2,T} & \ldots & x_{k,T} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix}$$

$$Y = X\,\beta + u$$

El vector de coeficientes β, se puede estimar por Mínimos Cuadrados Ordinarios, es decir, minimizando la suma de los errores al cuadrado:

$$\beta \quad | \quad Min(Y_t - \beta_0 - \beta_1\, x_{1,t} - \beta_2\, x_{2,t} - \ldots - \beta_k\, x_{k,t})^2 \;\equiv\; Min(y - X\,\beta)'(Y - X\,\beta)$$

Las condiciones de primer orden del problema de minimización exigen que la primera derivada de la función objetivo respecto de β se anule:

$$\frac{\partial (Y - X\,\beta)'(Y - X\,\beta)}{\partial \beta} = 2(X'X)\beta - 2X'Y = 0$$

Resolviendo para β:

$$(X'X)\beta = X'Y \qquad \Longrightarrow \qquad \hat{\beta} = (X'X)^{-1}X'Y$$

Una vez estimados los coeficientes, la recta de regresión muestral es:

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1,t} + \hat{\beta}_2 x_{2,t} + \ldots + \hat{\beta}_k x_{k,t} \quad t = 1, 2, \ldots, T$$

La recta de regresión muestral permite estimar el valor promedio de Y_t correspondiente a un conjunto de valores de las variables explicativas.

El objetivo es predecir los valores futuros de la variable endógena $Y_{T+\ell}, \ell = 1, 2, \ldots$. El verdadero valor de la variable endógena viene dado por el modelo:

$$\begin{aligned} Y_{T+\ell} &= \beta_0 + \beta_1 x_{1,T+\ell} + \beta_2 x_{2,T+\ell} + \ldots + \beta_k x_{k,T+\ell} + u_{T+\ell} \\ Y_{T+\ell} &= \beta' X'_{T+\ell} + u_{T+\ell} \quad t = 1, 2, \ldots, T \end{aligned}$$

Suponiendo conocidos los valores futuros de las variables explicativas y bajo el supuesto de que la relación entre y_t y las variables explicativas especificada por el modelo de regresión permanece constante en el futuro, la *predicción por punto* se puede obtener utilizando la recta de regresión muestral:

$$\begin{aligned} Y_T(\ell) &= \hat{\beta}_0 + \hat{\beta}_1 x_{1,T+\ell} + \hat{\beta}_2 x_{2,T+\ell} + \ldots + \hat{\beta}_k x_{k,T+\ell} \\ &= \hat{\beta}' X'_{T+\ell} \quad \ell = 1, 2, \ldots \end{aligned}$$

donde $X_{T+\ell} = [1 \;\; x_{1,T+\ell} \;\; x_{2,T+\ell} \; \ldots \; x_{k,T+\ell}]$ es el vector que recoge los valores de las variables explicativas del modelo en el momento $T + \ell$.

La *predicción por intervalo* se obtiene a partir de la distribución del error de predicción.

El error de predicción viene dado por:

$$e_T(\ell) = Y_{T+\ell} - Y_T(\ell) = \beta' X'_{T+\ell} + u_{T+\ell} - \hat{\beta}' X'_{T+\ell} = (\beta - \hat{\beta})' X'_{T+\ell} + u_{T+\ell}$$

Como la perturbación u_t se supone que sigue una distribución normal, el estimador MCO $\hat{\beta}$ sigue también una distribución normal. El error de predicción es una combinación lineal de variables normales, luego sigue también una distribución normal con media:

$$E[e_T(\ell)] = E[(\beta - \hat{\beta})' X'_{T+\ell} + u_{T+\ell}] = E[(\beta - \hat{\beta})'] X_{T+\ell} + E[u_{T+\ell}] = 0$$

si se cumplen las condiciones para que el estimador MCO sea insesgado, y con varianza:

$$
\begin{aligned}
V(e_T(\ell)) = E[e_T(\ell)]^2 &= E[(\beta - \hat{\beta})'\, X'_{T+\ell} + u_{T+\ell}]'\; [(\beta - \hat{\beta})'\, X'_{T+\ell} + u_{T+\ell}] \\
&= E[X_{T+\ell}\, (\beta - \hat{\beta})\, (\beta - \hat{\beta})'\, X'_{T+\ell}] + u'_{T+\ell} u_{T+\ell} + u'_{T+\ell}\, (\beta - \hat{\beta})'\, X'_{T+\ell} \\
&+ X_{T+\ell}(\beta - \hat{\beta})\, u_{T+\ell}] = X_{T+\ell}\, E[(\beta - \hat{\beta})\, (\beta - \hat{\beta})']\, X'_{T+\ell} + \sigma_u^2 \\
&= X_{T+\ell}\, \sigma_u^2 (X'X)^{-1}\, X'_{T+\ell} + \sigma_u^2 = \sigma_u^2\; [1 + X_{T+\ell}\, (X'X)^{-1}\, X'_{T+\ell}]
\end{aligned}
$$

Por lo tanto, la distribución del error de predicción es:

$$
e_T(\ell) \quad \sim \quad N\,(0,\;\; V(e_T(\ell)))
$$

lo que implica que:

$$
Y_{T+\ell} - Y_T(\ell) \quad \sim \quad N\,(0,\;\; V(e_T(\ell)))
$$

Estandarizando esta distribución:

$$
\frac{Y_{T+\ell} - Y_T(\ell)}{\sqrt{V(e_T(\ell))}} \quad \sim \quad N(0,1)
$$

El intervalo de predicción de probabilidad $(1-\alpha)\,\%$:

$$
P\left[-N_{\alpha/2} \leq \frac{Y_{T+\ell} - Y_T(\ell)}{\sqrt{V(e_T(\ell))}} \leq N_{\alpha/2}\right] = (1-\alpha)\,\%
$$

$$
P\left[Y_T(\ell) - N_{\alpha/2}\,\sqrt{V(e_T(\ell))} \leq Y_{T+\ell} \leq Y_T(\ell) + N_{\alpha/2}\,\sqrt{V(e_T(\ell))}\right] = (1-\alpha)\,\%
$$

Por lo tanto, el intervalo de predicción está centrado en la predicción por punto y su amplitud depende de la varianza del error de predicción: a mayor varianza del error de predicción, mayor amplitud del intervalo.

Usualmente no se conoce la varianza del error de predicción y hay que estimarla[1]:

$$\hat{V}(e_T(\ell)) = \hat{\sigma}_u^2\ [1 + X_{T+\ell}\,(X'X)^{-1}\,X'_{T+\ell}]$$

En este caso, el intervalo de predicción de confianza $(1-\alpha)\,\%$ es:

$$P\left[Y_T(\ell) - t_{\alpha/2}(T-(k+1))\,\sqrt{V(e_T(\ell))} \leq Y_{T+\ell} \leq Y_T(\ell) + t_{\alpha/2}(T-(k+1))\,\sqrt{V(e_T(\ell))}\right]$$

[1] Si no se conocen los valores futuros de las variables explicativas, también habría que estimarlos, por ejemplo, mediante modelos de series temporales univariantes.

Apéndice B

Medias Móviles

Un filtro lineal convierte una serie temporal Y_t en otra Z_t, a través de la operación lineal:

$$Z_t = \sum_{r=-q}^{s} a_r \, Y_{t+r}$$

donde a_r es el conjunto de ponderaciones.

Si la sucesión de pesos es tal que $\sum_{r=-q}^{s} \alpha_r = 1$, el filtro lineal recibe el nombre de *Media Móvil*.

Las medias móviles son, por lo tanto, medias de un número preestablecido de datos, en que se va añadiendo sucesivamente un dato nuevo y quitando al mismo tiempo, el más antiguo de los incluidos en la media anterior.

Así, una media móvil de s términos, denominada media móvil de orden s, se calculará según la expresión:

$$M_t = a_0 \, y_t + a_1 \, Y_{t-1} + \ldots + a_s \, Y_{t-s}, \qquad t = s+1, s+2, \ldots, T$$

donde $\sum_{r=0}^{s} \alpha_r = 1$.

Notése, que la media muestral se puede reinterpretar como una media móvil de orden T (incluye todas las observaciones).

Las medias móviles pueden ser:

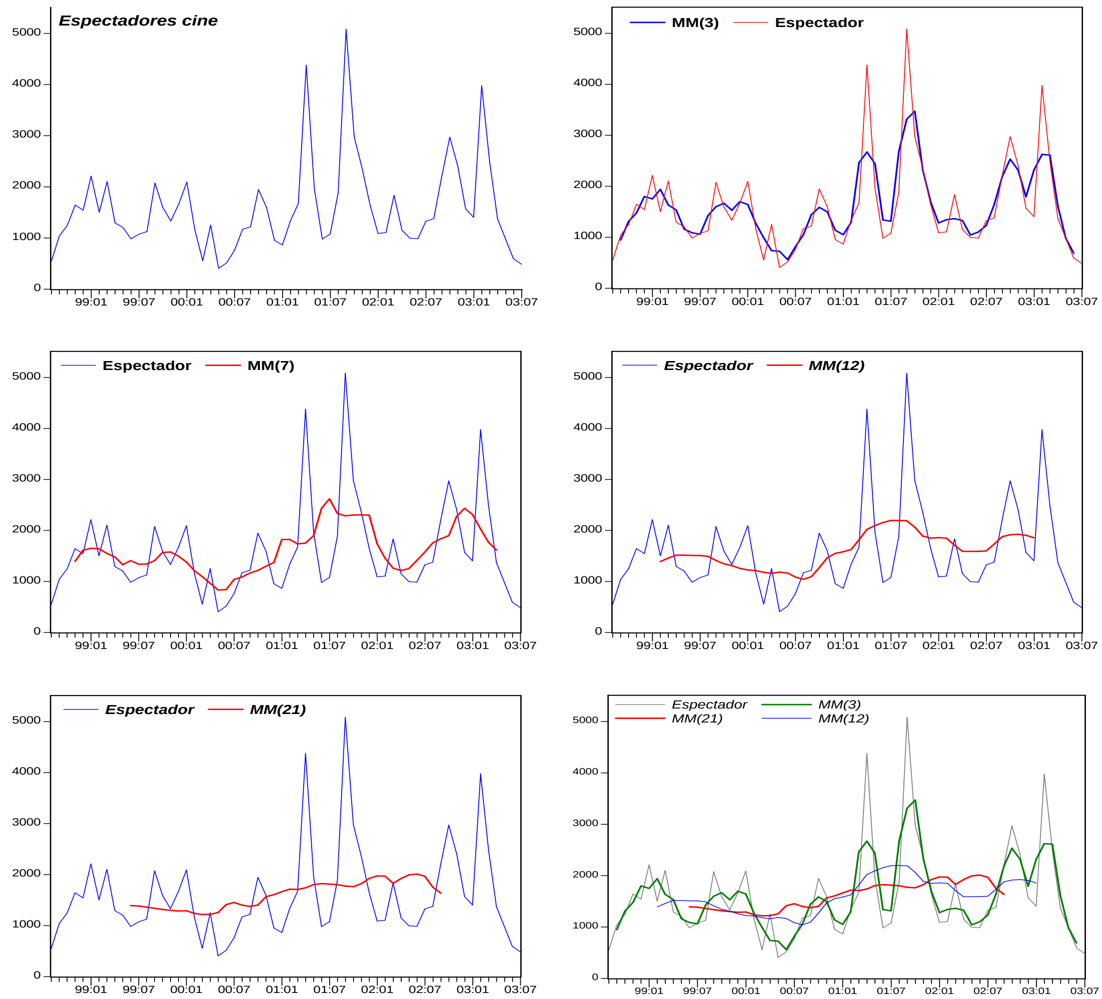
a) *Simples*: Si todos los pesos o ponderaciones son iguales,

$$M_t = \frac{Y_t + Y_{t-1} + \ldots + Y_{t-s}}{s}, \qquad t = s+1, s+2, \ldots, T$$

b) *Ponderadas*: Si los pesos son distintos.

Otro tema relevante es la referencia temporal que debe asignarse a una media móvil, es decir, en una media móvil están involucradas varias observaciones a diferentes momentos de tiempo, ¿a cuál de todos ellos asignamos la media móvil? Las soluciones más habituales son las siguientes:

Gráfico B.1: Serie: Espectadores de cine español (en miles)



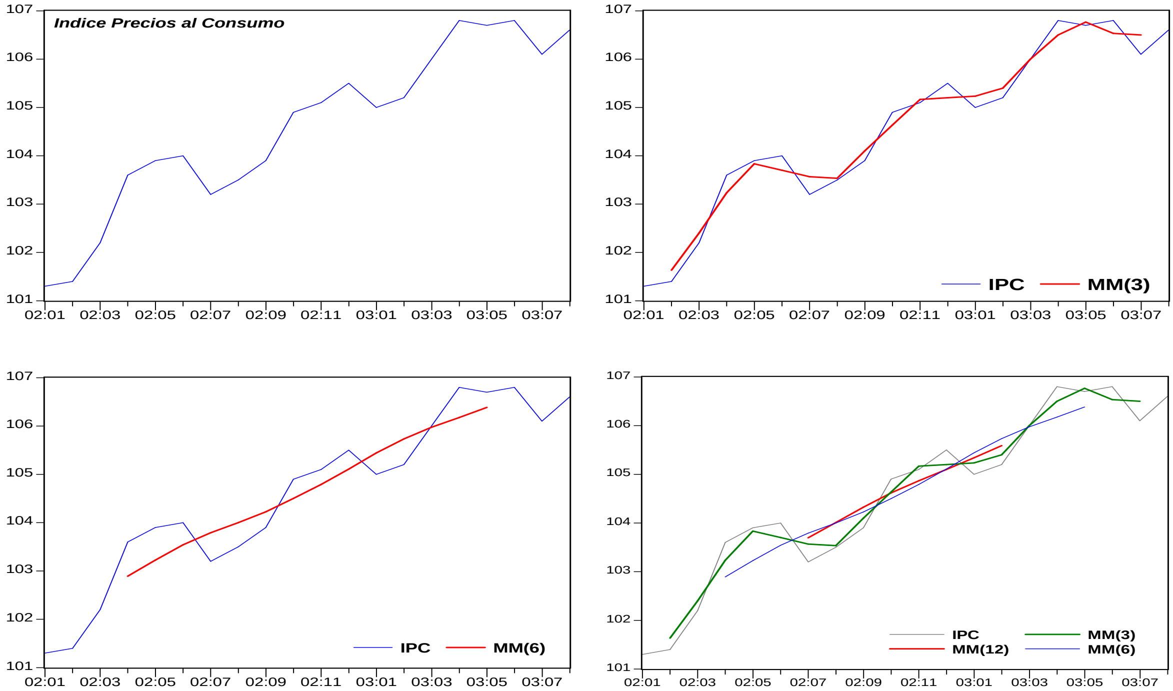
- Pensando en términos de predicción parece inmediato que la media de una serie de valores se asigne a la fecha que corresponde al dato más reciente (al periodo t porque y_t es el valor más cercano a M_t).

- Sin embargo, para algunas aplicaciones de análisis de series temporales, sería más lógico que una media de valores se asignase al periodo medio, con lo que una media móvil debiera definirse basándose en un número similar de datos anteriores y posteriores:

$$M_t = \frac{Y_{t-r} + \ldots + Y_{t-1} + Y_t + Y_{t+1} + \ldots + Y_{t+r}}{2r+1} \qquad t = r, s+2, \ldots, T-r$$

En el caso de que la media se realice con un número par de datos no quedará "centrada" en un momento t en particular y sería entonces necesario calcular una media de medias:

Gráfico B.2: Serie: Indice de Precios al Consumo (base=2001)



Ejemplo:

$$M^1_{t+1} = \frac{Y_t + Y_{t+1}}{2} \qquad M^2_{t+1} = \frac{Y_{t+1} + Y_{t+2}}{2}$$

$$M_{t+1} = \frac{M^1_{t+1} + M^2_{t+1}}{2} = M^1_{t+1} = \frac{Y_t + 2Y_{t+1} + Y_{t+2}}{4}$$

La restricción de que la media móvil esté centrada es artificial, pero se puede imponer para preservar una correspondencia entre la serie Y_t y la serie M_t.

Es importante no confundir una media móvil ponderada en general, con una media móvil simple centrada. Así, por ejemplo, no hay que confundir la media móvil centrada de orden 2 que se ha calculado antes con una media móvil ponderada de orden tres general:

$$M_t = \frac{Y_{t-1} + 2Y_t + 3Y_{t+1}}{3}$$

Cuando se trabaja con un filtro simétrico como suele ser la media móvil surgen problemas en los extremos de la serie ya que M_t se calcula para $t = r+1$ a $t = T-r$. En la práctica, r suele ser mucho más pequeño que T por lo que en algunas situaciones esto no tiene mucha importancia, pero en otras nos puede interesar mucho obtener valores suavizados de la serie hasta el momento $t = T$.

La serie correspondiente a la media móvil alisa las variaciones de la serie original, tanto estacionales como cíclicas o erráticas (si existe algún dato anómalo que se desvía mucho de los demás, al hacer medias móviles ya no se desviará tanto. De ahí que se incluya a las medias móviles entre las denominadas técnicas de alisado.

Apéndice C

Operador de retardos

L: operador de retardos, retrasa 1 periodo a las variables:

$$LY_t = Y_{t-1}$$

$$L^2 Y_t = Y_{t-2}$$

$$L^k Y_t = Y_{t-k}$$

$$L^{-1} Y_t = Y_{t+1}$$

$$L^{-k} Y_t = Y_{t+k}$$

$L^k C = C$, donde C es una constante.

A partir del operador de retardos, se derivan polinomios en el operador de retardos que se utilizan mucho en el Análisis de Series Temporales:

- Operador de diferencias de orden j:

$$\Delta_j = (1 - L^j) \longrightarrow \Delta_j Y_t = (1 - L^j) Y_t = Y_t - Y_{t-j}$$

$$\Delta_2 = (1 - L^2) \longrightarrow \Delta_2 Y_t = (1 - L^2) Y_t = Y_t - Y_{t-2} \quad \text{de orden 2}$$

$$\Delta_7 = (1 - L^7) \longrightarrow \Delta_7 Y_t = (1 - L^7) Y_t = Y_t - Y_{t-7} \quad \text{de orden 7}$$

- Operador de diferencias regular:

$$\Delta = (1 - L) \longrightarrow \Delta Y_t = (1 - L) Y_t = Y_t - Y_{t-1}$$

$$\Delta^2 = (1 - L)^2 \longrightarrow \Delta^2 Y_t = (1 - L)^2 Y_t = (1 - 2L + L^2) Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$$

$$\Delta^d = (1 - L)^d \longrightarrow \Delta^d Y_t = (1 - L)^d Y_t = \underbrace{(1 - L) \dots (1 - L)}_{d\ veces} Y_t$$

- Operador de diferencias estacional. Se utiliza para realizar diferencias anuales, es decir, un mes respecto del mismo mes del año anterior.

Serie trimestral: $\Delta_4 = (1 - L^4) \quad \longrightarrow \quad \Delta_4 Y_t = (1 - L^4) Y_t = Y_t - Y_{t-4}$

Serie mensual: $\Delta_{12} = (1 - L^{12}) \quad \longrightarrow \quad \Delta_{12} Y_t = (1 - L^{12}) Y_t = Y_t - Y_{t-12}$

- Operador suma: $S_m(L)$

$$S_m(L) = 1 + L + L^2 + L^3 + \ldots + L^{m-1}$$

De especial interés es su aplicación con series trimestrales o mensuales para sumar todas las observaciones de un año:

Serie trimestral:
$S_4(L) Y_t = (1 + L + L^2 + L^3) Y_t = Y_t + Y_{t-1} + Y_{t-2} + Y_{t-3}$

Serie mensual:
$S_{12}(L) y_t = (1 + L + L^2 + \ldots + L^{12}) Y_t = Y_t + Y_{t-1} + Y_{t-2} + \ldots + Y_{t-12}$

Apéndice D

Mínimos Cuadrados Descontados

Una tendencia global puede ser estimada por Mínimos Cuadrados Ordinarios (MCO): regresando las observaciones sobre una constante y el tiempo. Un nivel global es, por ejemplo, la media muestral, $\bar{y}$, que es, por supuesto, un estimador MCO.

El Alisado Exponencial Simple se puede obtener también introduciendo un factor de descuento ω, en la suma de cuadrados de la función objetivo. Esto es lo que se conoce como Mínimos Cuadrados Descontados (MCD):

$$S(m,\omega) = \sum_{j=0}^{T-1} \omega^j (Y_{T-j} - m)^2$$

donde $0 \leq \omega \leq 1$. Diferenciando con respecto a m, obtenemos el estimador MCD:

$$m_T = \left[\sum_{j=0}^{T-1} \omega^j \right]^{-1} \sum_{j=0}^{T-1} \omega^j Y_{T-j}$$

que es la media muestral para $\omega = 1$. Para ω menor que la unidad:

$$m_T = \frac{1-\omega}{1-\omega^T} \sum_{j=0}^{T-1} \omega^j Y_{T-j}$$

Si fijamos $\omega = 1 - \alpha$, esta expresión es la misma que la del AES, menos por el divisor $1 - (1 - \alpha^T)$. Cuando $T \to \infty$, la expresión coincide exactamente.

Si queremos introducir una pendiente en la función de predicción, MCD equivale a encontrar los valores del nivel m y de la pendiente b que minimizan:

$$S(m,b,\omega) = \sum_{j=0}^{T-1} \omega^j (Y_{T-j} - m - bj)^2$$

Diferenciando los estimadores son:

$$\begin{bmatrix} m_T \\ b_T \end{bmatrix} = \begin{bmatrix} \sum \omega^j & -\sum j\omega^j \\ -\sum j\omega^j & -\sum j^2\omega^j \end{bmatrix}^{-1} \begin{bmatrix} \sum \omega^j Y_{T-j} \\ -\sum j\omega^j Y_{T-j} \end{bmatrix}$$

Si T es grande, se puede demostrar que:

$$m_T = (1-\omega^2)\sum \omega^j Y_{T-j} - (1-\omega^2)\sum j\omega^j Y_{T-j}$$

Si seguimos manipulando se puede demostrar que estos estimadores MCD se pueden obtener de las recursiones:

$$\begin{aligned} m_t &= m_{t-1} + b_{t-1} + (1-\omega^2)\nu_t \\ b_t &= b_{t-1} + (1-\omega)^2\nu_t \end{aligned}$$

El uso de las recursiones anteriores junto con la función de predicción:

$$\hat{Y}_T(\ell) = m_T + b_T\ell \qquad\qquad \ell = 1, 2, \dots$$

se denomina alisamiento exponencial doble, o alisamiento exponencial lineal con parámetro único.

Generalmente se sugiere que el factor de descuento se fije cercano a la unidad.

Se puede demostrar que el alisamiento exponencial doble es un caso particular del método de H-W con:

$$\alpha_0 = 1-\omega^2 \qquad\qquad \alpha_1 = \frac{1-\omega}{1+\omega}$$

El método de MCD se puede extender hasta cubrir un amplio campo de funciones de tiempo, incluyendo polinomios de alto grado y sinusoides:

$$y_t = \sum_{t=1}^{k} \omega^j f_j(t-j) + e_t$$

Las funciones f_j pueden ser polinomios, exponenciales, funciones seno-coseno para recoger por ejemplo estacionalidades, etc. Estos modelos son sólo locales, de forma que el método de estimación más natural es Mínimos Cuadrados Descontados.

El problema fundamental de MCD es que maneja una sóla constante de alisamiento, lo que le hace menos flexible.

Lecturas recomendadas

a) Aznar, A. y F.J. Trívez (1993). *Métodos de Predicción en Economía*. Volúmenes I y II. Ed. Ariel. Barcelona.

b) Box, G.E.P. y G.M. Jenkins (1970). *Time series analysis: forecasting and control.* San Francisco, Holden Day.

c) Harvey, A.C. (1989). *Forecasting, estructural time series models and the Kalman Filter*. Cambridge University Press. Cambridge.

d) Holt, C.C. (1957). *Forecasting seasonal and trends by exponentially weighted moving average*. Carnegie Institute of Techonology Pittsburgh, ONR Research Memorandum, no. 52.

e) Muth, J.F. (1960). Optimal properties of exponentially weighted forecasts. *Journal of the American Statistical Association*, 55, pp. 299-305.

f) Peña, D. (2005). *Análisis de series temporales.* Alianza editorial. Madrid.

g) Pulido, A. (1989). *Predicción económica y empresarial.* Pirámide. Madrid.

h) Uriel, E. (2000). Introducción al análisis de series temporales. Editorial AC. Madrid.

i) Winters, P.R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6, pp. 324–342.