Estado del arte en predicción de series temporales

J.J. Asensio, P.A. Castillo

Departamento de Arquitectura y Tecnología de Computadores
ETSIIT, CITIC-UGR
Universidad de Granada, España
{asensio, pacv}@ugr.es

Resumen bonito abstract test

1. Introducción

2. Métodos convencionales

En esta sección se resumen algunos trabajos relevantes en el área de predicción de series temporales.Los métodos que se comentan a continuación se clasifican según los modelos utilizados.

2.1. Alisado exponencial

Los métodos de alisado exponencial habitualmente se utilizan en la estimación de la demanda de un producto para un periodo dado. Una forma sencilla de realizar la predicción consiste en tomar la media ponderada de los valores previos dando mayor peso a los valores más cercanos en el tiempo. El error en la predicción actual puede ser tenida en cuenta para la siguiente predicción.

Estos métodos tienen su origen en las décadas 50 y 60 con el trabajo de Brown (1959, 1963) [1] [2], Holt (1957, reimpreso en 2004)[3], y Winters (1960) [4]. Pegels (1969) [5] proporcionó una clasificación sencilla pero útil de la tendencia y los patrones estacionales dependiendo de si eran aditivos (lineales) o multiplicativos (no lineales).

Los métodos de alisado exponencial fueron ampliamente utilizados en la industria y los negocios, sin embargo no habían recibido mucha atención por parte de los estadísticos y no tenían un fundamento estadístico bien desarrollado.

A partir de Muth (1960) [6] surge la primera base estadística para el alisado exponencial simple (SES) demostrando que proporcionaba una predicción óptima para caminos aleatorios con ruido.

Posteriormente Box y Jenkins (1970), Roberts (1982)[7], y Abraham y Ledolter (1983, 1986) [8][9], mostraron que algunas predicciones de alisado exponencial resultan ser casos especiales de modelos ARIMA. Snyder (1985) [10] demostró que SES podría ser también una forma de modelo de espacio de estados de innovaciones (con una única fuente de error).

Taylor (2003)[11] proporcionó los únicos métodos de alisado exponencial multiplicativos genuinamente nuevos. Entre los trabajos de aplicación se destacan varios contextos incluyendo componentes de computadores (Gardner, 1993 [12]), pasajeros de vuelo (Grubb & Masa, 2001 [13]), y planificación de producción (Miller & Liberatore, 1993 [14]).

En Hyndman, Koehler, Snyder, y Grose (2002) [15] y en la ampliación de Taylor (2003) [11] se da una clasificación de este tipo de métodos. Cada método se componen de una de cinco tipos de tendencia (ninguna, aditiva, aditiva alisada, multiplicativa y multiplicativa alisada) y uno de tres tipos de estacionalidad (ninguna, aditiva y multiplicativa). Por tanto, hay 15 métodos diferentes, donde entre los más conocidos está SES (sin tendencia ni estacionalidad), el método lineal de Holt (tendencia aditiva, sin estacionalidad), el método aditivo de Holt-Winters (tendencia aditiva y estacionalidad aditiva), y el método no lineal de Holt-Winters (tendencia aditiva y estacionalidad multiplicativa).

2.2. Modelos ARIMA

Los primeros intentos de estudiar series temporales, particularmente en el siglo 19, estaban generalmente caracterizados por la idea de un mundo determinista. Fue la mayor contribución de Yule (1927) [16] la que impulsó la noción de aleatoriedad en series temporales postulando que cada serie temporal puede ser considerada como la realización de un proceso estocástico. Basado en esa sencilla idea, se desarrolló desde entonces un gran número de métodos de series temporales. Slutsky, Walker, Yaglom y Yule formularon el concepto de los modelos auto-regresivos (AR) y media móvil (MA). El teorema de la descomposición de Wold condujo a la formulación y solución del problema de predicción lineal de Kolmogorov (1941) [17]. Desde entonces, una gran cantidad de literatura ha aparecido en el área de series temporales, tratando estimación de parámetros, identificación, comprobación de modelos y predicción, véase Newbold (1983) [18] para un estudio anterior.

La publicación *Time Series Analysis: Forecasting and Control* de Box y Jenkins (1970) [19] integró el conocimiento existente. Además, estos autores desarrollaron una ciclo coherente, versátil e iterativo en tres etapas para la identificación, estimación y verificación de series temporales (conocido como el enfoque Box-Jenkins). El libro ha tenido un enorme impacto en la teoría y práctica moderna de análisis y predicción de series temporales. Con la llegada de los ordenadores, se popularizó el uso de los modelos ARIMA y se extendió a muchas áreas de la ciencia. De hecho, la predicción de series temporales discretas mediante modelos ARIMA univariados, modelos de función de transferencia (regresión dinámica), y modelos ARIMA multivariados han generado bastantes artículos en IJF. A menudo estos estudios son de naturaleza empírica, y usan uno o varios métodos/modelos como punto de referencia para comparar.

Univariado El éxito de la metodología de Box-Jenkins se basa en el hecho de que los modelos pueden imitar el comportamiento de los diversos tipos de series

Tabla 1. Lista de ejemplos de aplicaciones reales

Dataset	Horizonte	Benchmark	Referencia
Univariate ARIMA			
Electricity load (min)	1–30 min	Wiener filter	[20]
Quarterly automobile insurance paid claim costs	8 quarters	Log-linear regression	[21]
Daily federal funds rate	1 day	Random walk	[22]
Quarterly macroeconomic data	1–8 quarters	Wharton model	[23]
Monthly department store sales	1 month	Simple exponential smoothing	[24] [25] [26]
Monthly demand for telephone services	3 years	Univariate state space	[27]
Yearly population totals	20–30 years	Demographic models	[28]
Monthly tourism demand	1–24 months	Univariate state space-multivariate state space	[29]
Dynamic regression/transfer function			
Monthly telecommunications traffic	1 month	Univariate ARIMA	[30]
Weekly sales data	2 years	n.a.	[31]
Daily call volumes	1 week	Holt–Winters	[32]
Monthly employment levels	1–12 months	Univariate ARIMA	[33]
Monthly and quarterly consumption of natural gas	1 month/1 quarter	Univariate ARIMA	[34]
Monthly electricity consumption	1–3 years	Univariate ARIMA	[35]
VARIMA			
Yearly municipal budget data	Yearly (in-sample)	Univariate ARIMA	[36]
Monthly accounting data	1 month	Regression, univariate, ARIMA, transfer function	[37]
Quarterly macroeconomic data	1–10 quarters	Judgmental methods, univariate ARIMA	[38]
Monthly truck sales	1–13 months	Univariate ARIMA, Holt–Winters	[39]
Monthly hospital patient movements	2 years	Univariate ARIMA, Holt–Winters	[40]
Quarterly unemployment rate	1–8 quarters	Transfer function	[41]

y sin requerir muchos parámetros a estimar en la elección final del modelo. Sin embargo, a mediados de los años sesenta, la selección de un modelo era en gran medida una cuestión subjetiva del investigador, no había ningún algoritmo para especificar un modelo único. Desde entonces, muchas técnicas y métodos han surgido para añadir rigor matemático en el proceso de búsqueda de un modelo ARMA, incluyendo el criterio de información Akaike (AIC), el error de predicción final de Akaike (FPE) y el criterio de información bayesiano (BIC). A menudo, estos criterios se reducen a la minimización de errores de predicción a un paso (en la muestra) , con un término de penalización para el sobreajuste. FPE también ha sido generalizado para predicción a varios pasos (véase por ejemplo, Bhansali, 1996 [42] y Bhansali, 1999 [43]), pero esta generalización no ha sido utilizada en el área aplicada. Este también parece ser el caso de criterios basados en principios de validación cruzada y validación por división muestral (véase por ejemplo, West , 1996 [44]), haciendo uso de errores de predicción fuera-de-muestra; véase Peña y Sánchez (2005) [45] para un enfoque relacionado a considerar.

El problema de añadir información externa (prior) en las predicciones de ARIMA univariado fue considerado por Cholette (1982) [46], Guerrero (1991) [47] y de Alba (1993) [48]).

Como alternativa a la metodología de ARIMA univariado, Parzen (1982) [49] propuso la metodología ARARMA. La idea clave es que una serie temporal es transformada de un filtro AR de memoria larga, a un filtro de memoria corta, evitando así un operador de diferenciación más duro. Además, se usa una fase de identificación diferente al convencional de Box-Jenkins. En la competición M (Makridakis et al., 1982 [50]), los modelos ARARMA consiguieron el MAPE más bajo para horizontes de predicción mayores.

El modelado automático de ARIMA univariado, ha mostrado una capacidad de predicción a un paso tan precisa como la de otros métodos competentes (Hill y Fildes, 1984, Liber, 1984, Poulos et al., 1987 y Texter y Ord, 1989). Varios proveedores de software han implementado métodos de predicción automáticos de series temporales (incluso métodos multivariados); véase Geriner y Ord (1991) [51], Tashman y Leach (1991) [52], y Tashman (2000) [53]. A menudo estos métodos actúan como caja negra. La tecnología de sistemas expertos (Mélard & Pasteels, 2000) se puede usar para evitar este problema. Algunas directrices en la elección de un método de predicción automático se da en Chatfield (1988) [54].

Mejor que adoptar un modelo AR para todos los horizontes de predicción, Kang (2003) [55] investigó empíricamente el caso de seleccionar un modelo AR de forma diferenciada para cada horizonte en predicciones multi-paso.

Función de transferencia La identificación de modelos de función de transferencia puede ser difícil cuando hay más de una variable de entrada. Edlund (1984) [56] presentó un método en dos pasos para identificación de una función respuesta impulso cuando varias variables de entrada están correladas. Koreisha (1983) [57] estableció varias relaciones entre funciones de transferencia, implicaciones causales y especificación del modelo econométrico. Gupta (1987) [58] identificó los principales problemas en las comprobaciones de causalidad. Usando análisis de componentes principales, del Moral y Valderrama (1997) [59] sugirieron una representación parsimoniosa de modelo de función de transferencia . Krishnamur, Narayan y Raj (1989) [60] mostraron cómo se pueden obtener estimaciones más precisas del impacto de las intervenciones en modelos de función de transferencia usando una variable de control.

Multivariado El modelo de vector ARIMA (VARIMA) es una generalización multivariada del modelo univariado ARIMA. Las características de población de procesos VARMA parecen haber sido derivados primero por Quenouille (1957) [61], aunque el software para su implementación comenzó estar disponible en los 80 y 90. Riise y Tjostheim (1984) [62] mostraron como los filtros de suavizado pueden construirse dentro de los modelos VARMA. El alisado previene fluctuaciones irregulares en series temporales explicativas al migrar las predicciones de la serie dependiente. Para determinar el máximo horizonte de predicción para procesos VARMA, DeGooijer y Klein (1991) [63] establecieron las propiedades teóricas de las predicciones y el error de predicción acumulado a varios pasos. Lütkepohl (1986) [64] estudió los efectos de la agregación temporal y el muestreo sistemático en predicción, suponiendo que la variable desagregada (estacionaria) sigue un proceso VARMA de orden desconocido. Posteriormente, Bidarkota (1998) [65] consideró el mismo problema pero con las variables observadas integradas en lugar de estacionarias.

Los vectores de auto-regresión (VAR) constituyen un caso especial de una clase más general de modelos VARMA. En esencia, un modelo VAR es una aproximación flexible a la forma reducida de una variedad amplia de modelos econométricos dinámicos. Los modelos VAR se pueden especificar de varias formas. Funke (1990) [66] presentó 5 especificaciones diferentes de VAR y comparó sus rendimientos de predicción usando series mensuales de producción industrial.

En general, los modelos VAR tienden a sufrir sobreajuste con demasiados parámetros libres y no significativos. Como resultado, estos modelos pueden dar pobres predicciones fuera de muestra, incluso aunque el ajuste en las muestras sea bueno; véase Liu, Gerlow y Irwin (1994) [67] y Simkins (1995) [68]. En lugar de restringir algunos parámetros como habitualmente se hace, Litterman (1986) [69] y otros impusieron una distribución preferente en los parámetros, expresando la idea de que muchas variables económicas se comportan como un camino aleatorio.

2.3. Estacionalidad

El enfoque más antiguo para controlar la estacionalidad en series temporales es extraerla usando un procedimiento iterativo de descomposición como el método X-11. Este método y sus variantes (incluyendo versiones más reciente, X-12-ARIMA, Findley, Monsell, Bell, Otto y Chen, 1998 [70], se han estudiado extensivamente durante los últimos años derivando en varios enfoques.

Uno de ellos estudia el uso predicción como parte del método de descomposición estacional. Quenneville, Ladiray y Lefrançois (2003) [71] con otro enfoque observaron las predicciones implícitas mediante los filtros de medias móviles asimétricos del método x-11 y sus variantes. Un tercer enfoque fue ver la efectividad de predicción usando datos ajustados estacionalmente a partir de un método de descomposición estacional.

Además de trabajar en el método X-11 y sus variantes, también se han desarrollado varios métodos nuevos para el ajuste estacional, entre los que destaca el enfoque basado en modelo de TRAMO-SEATS (Gómez y Maravall, 2001 [72] y Kaiser y Maravall 2005 [73]) y el método no paramétrico STL (Cleveland, Cleveland, McRae y Terpenning, 1990 [74]). Otra propuesta ha sido el uso de modelos sinusoidales (Simmons, 1990 [75]).

En la predicción de varias series similares, Withycombe (1989) [76] mostró que puede ser más eficiente estimar una componente estacional combinada para el grupo de series, en lugar de usar sus patrones individuales. Bunn y Vassilopoulos (1993) [77] demostró cómo usar agrupamiento para formar grupos apropiados para esta situación, y Bunn y Vassilopoulus (1999)[78] introdujo algunos estimadores mejorados para los índices de grupo estacional.

Posteriormente hubo un trabajo considerable en el uso e implementación de las pruebas de raíz unitaria estacional incluyendo Hylleberg y Pagan (1997) [79], Taylor (1997)[80] y Franses y Koehler (1998) [81].

Los modelos de series temporales periódicas fueron también exploradas por Wells (1997) [82], Herwartz (1997) [83] y Novales y de Fruto (1997) [84], los cuales encontraron que los modelos periódicos pueden conducir a una mejora en el rendimiento de la predicción comparados con modelos no periódicos bajo algunas condiciones. La predicción de procesos ARMA periódicos multivariados fue considerada por Ullah (1993) [85].

Varios artículos han comparado empíricamente los modelos estacionales. Chen (1997) [86] exploró la robustez de un modelo estructural, un modelo de regresión con estacionalidad, un modelo ARIMA, y el método de Holt-Winters, y encontró que los dos últimos daban predicciones relativamente robustas cuando el modelo no estaba completamente especificado. Noakes, McLeod, y Hipel (1985) [87], Albertson y Aylen (1996) [88], Kulendran y King (1997) [89] y Franses y van Dijk (2005) [90] compararon el rendimiento de predicción de varios modelos estacionales aplicados a datos reales. El modelo con mejor predicción variaba a lo largo del estudio, dependiendo de qué modelo se usara y la naturaleza de los datos. Parece no haber consenso todavía en qué condiciones se puede preferir un modelo u otro.

2.4. Espacio de estado, modelos estructurales y el filtro de Kalman

Los estadísticos empezaron a usar los modelos de espacio de estado para predecir series temporales a principios de los 80, aunque las ideas ya estaban presentes en la literatura de ingeniería desde el innovador trabajo de Kalman (1960) [91]. Los modelos de espacio de estado proporcionan un marco unificado en el que se puede escribir cualquier modelo lineal de serie temporal. La contribución clave para predicción de Kalman (1960) fue dar un algoritmo recursivo (conocido como el filtro de Kalman) para calcular las predicciones. Los estadísticos empezaron a interesarse por los modelos de espacio de estado cuando Schweppe (1965) [92] mostró que el filtro de Kalman era un método eficiente para computar los errores de predicción a un paso y su varianza asociada necesaria para producir la función de probabilidad. Shumway y Stoffer (1982)[93] combinaron el algoritmo EM con el filtro de Kalman para dar un enfoque general para predecir series temporales usando modelos de espacio de estado, incluyendo la posibilidad de tener observaciones inexistentes.

Una clase particular de modelos de espacio de estado, conocidos como *modelos lineales dinámicos* (DLM) fue introducida por Harrison y Stevens (1976) [94], quienes también propusieron un enfoque bayesiano para la estimación. Fildes (1983) [95] comparó las predicciones obtenidas usando el método de Harrison y Stevens con métodos más sencillos como el alisado exponencial concluyendo que la complejidad adicional no conducía a una mejora del rendimiento de predicción. Harvey, 1984 [96] y 1989 [97] amplió esta clase de modelos y siguió un enfoque no bayesiano para la estimación. Harvey (2006) [98] proporciona una revisón comprensible y una introducción para esta clase de modelos incluyendo variaciones no gausianas y en tiempo continuo.

Estos modelos poseen muchas similaridades con los métodos de alisado exponencial, pero tienen múltiples fuentes de error aleatorio. En particular, el *modelo estructural básico* (BSM) es similar a método de Holt-Winters para datos estacionales e incluye componentes de nivel, tendencia y estacionalidad.

Otra clase de modelos de espacio de estado, conocida como *modelos balanceados de espacio de estado*, se ha usado principalmente para predicción de series temporales macro-económicas. Mittnik (1990)[99] proporcionó un estudio de esta clase de modelos, y Vinod y Basu (1995) [100] obtuvo predicciones para el consumo, ingresos y tasas de interés usando estos modelos. Estos sólo tienen una única fuente de error aleatorio y generalizan otros modelos como ARMAX, ARMA y modelos con retardo racional distribuido. Una clase de espacio de estados relacionada son los modelos de *fuente única de error* en los que subyacen los métodos de alisado exponencial.

A parte de estos desarrollos metodológicos, existen varios artículos proponiendo modelos innovadores de espacio de estados para resolver problemas prácticos de predicción. Entre ellos, Coomes (1992) [101] quien usó un modelo para predecir empleos según la industria para regiones locales y Patterson (1995) [102] quien usó un enfoque de espacio de estado para predicción de ingresos personales disponibles.

Los libros de Harvey (1989) [97], West y Harrison (1989) [103]y Durbin y Koopman (2001) [104] han tenido un impacto sustancial en la literatura de series temporales en cuanto a investigación de modelos de espacio de estado, filtrado de Kalman y modelos estructurales discretos y continuos en tiempo.

2.5. Modelos no lineales

La suposición de linealidad es útil y una herramienta potente en muchas áreas. Sin embargo, a finales de los 70 y principios de los 80, los modelos lineales eran insuficientes en muchas aplicaciones reales. Por ejemplo, los ciclos prolongados de tamaño de población animal, los ciclos solares (número anual de manchas solares), flujo de energía y relaciones de amplitud-frecuencia, son ejemplos donde no se adecuan los modelos lineales. Para satisfacer esta demanda, se propusieron modelos no lineales de series temporales útiles durante ese periodo.

Un factor que probablemente ha retrasado la difusión de publicaciones de predicciones no lineales es que hasta entonces no era posible obtener expresiones analíticas cerradas para predicciones multi-paso. Hoy día, las predicciones no lineales se obtienen mediante simulación Monte Carlo o por *bootstrapping* (remuestreo). El último caso se prefiere al no necesitar suposiciones sobre la distribución del proceso de error.

Modelos de cambio de régimen La clase de modelos no lineales autorregresivos con umbral auto-excitado (SETAR) fue promocionada de forma prominente mediante los libros de Tong, 1983 [105] y 1990 [106]. Clements y Smith (1997) [107] compararon varios métodos para obtener predicciones mult-paso para modelos SETAR univariados discretos. Concluyeron que las predicciones mediante simulación Monte Carlo eran satisfactorias en casos donde se sabe que las perturbaciones en el modelo SETAR vienen de una distribución simétrica. En otro caso se prefiere el método de *bootstrapping*.

Una desventaja del modelo SETAR es que la dinámica cambia de forma discontinua de un régimen a otro. En cambio, un modelo de transición suave (STAR) permite una transición más gradual entre los distintos regímenes.

Fok, van Dijk y Franses (2005) [108] examinaron la siguiente cuestión clave: ¿Puede un modelo STAR multinivel de datos de panel para series desagregadas mejorar las predicciones de agregados macroeconómicos como ingresos totales o desempleo total?. El modelo STAR propuesto parece merecer la pena investigarse en más detalle puesto que permite que los parámetros que gobiernan el cambio de régimen sean distintos según los estados. Basándose en simulaciones y hallazgos empíricos, los autores afirman que de hecho se pueden conseguir mejoras en las predicciones a un paso.

Franses, Paap, y Vroomen (2004) [109] propusieron un modelo con umbral AR(1) que permite una inferencia plausible sobre valores específicos de los parámetros. La idea clave es que los valores de los parámetros de AR dependen de una variable indicadora dominante. El modelo resultante mejora el rendimiento respecto a otros modelos no lineales cambiantes en tiempo, incluyendo el modelo de cambio de régimen de Markov, en términos de predicción.

Modelos con coeficiente funcional Un modelo AR con coeficiente funcional (FCAR o FAR) es un modelo AR en el que los coeficientes AR pueden variar como una función suave de otra variable, como un valor retrasado de la propia serie o una variable exógena. El modelo FCAR incluye los modelos TAR y STAR como casos especiales y es análogo al modelo aditivo generalizado de Hastie y Tibshirani (1991) [110].

2.6. Modelos de memoria larga

Cuando el parámetro de integración d de un modelo ARIMA es fraccional y mayor que cero, el proceso exhibe memoria en el sentido de que las observaciones de un periodo de tiempo largo, tiene una dependencia no despreciable. Los modelos estacionarios con memoria larga (0 ¡ d ¡ 0.5), también llamados ARMA diferenciados fraccionalmente (FARMA) o modelos ARMA integrados fraccionalmente (ARFIMA), han sido estudiados en muchos campos; véase Granger y Joyeux (1980) [111] para una introducción. Una motivación para estos estudios es que muchas series temporales empíricas tienen una función de autocorrelación muestral que decae a una tasa menor que en un modelo ARIMA de orden finito y entero d.

El potencial de predicción de modelos ajustados a FARMA/ARFIMA, en contraposición con resultados de predicción obtenidos con otros modelos de series temporales, ha sido tema de varios artículos en IJF y un special issue (2002, 18:2). Ray, 1993a [112] y Ray, 1993b [113] realizaron un estudio comparativo entre los modelos FARMA/ARFIMA estacionales y los modelos estándar (no fraccionarios) ARIMA estacionales. Los resultados muestran que los modelos AR de mayor orden son capaces de predecir bien a largo plazo comparados con los modelos ARFIMA.

Se han explorado muchas extensiones de modelos ARFIMA y comparaciones de su rendimiento de predicción relativo. Por ejemplo, Franses y Ooms (1997) [114] propuso el modelo ARFIMA$(0,d,0)$ periódico donde d puede variar con el parámetro de estacionalidad. Ravishanker y Ray (2002) [115] consideraron la estimación y predicción de modelos ARFIMA multivariados. Baillie y Chung (2002) [116] disertaron sobre el uso de modelos ARFIMA con tendencia estacionaria lineales, mientras que Beran, Feng, Ghosh y Sibbertsen (2002) [117] ampliaron este modelo para permitir tendencias no lineales. Souza y Smith (2002) investigaron el efecto de tasas de muestreo diferentes, tales como mensual versus trimestral, en estimaciones del parámetro de memoria larga d. De forma similar, Souza y Smith (2004) observaron los efectos de la agregación temporal en estimaciones y predicciones de procesos ARFIMA. Dentro del contexto de la calidad de control estadística, Ramjee, Crato y Ray (2002)[118] introdujeron un gráfico de control basado en preducción con una media móvil ponderada hiperbólicamente, diseñada específicamente para modelos ARFIMA no estacionarios.

2.7. Modelos ARCH/GARCH

Una característica clave de las series temporales financieras es que las rentabilidades grandes tienden a continuar rentabilidades grandes y lo mismo con rentabilidades pequeñas, es decir, hay periodos que muestran una alta (o baja) volatilidad. Este fenómeno se conoce como el agrupamiento de volatilidad en econometría y finanzas. La clase de modelos AR condicionados a heteroscedasticidad (ARCH), introducidas por Engle (1982) [119], describen los cambios dinámicos de la varianza condicionada como una función determinista (típicamente cuadrática) de rentabilidades pasadas. Puesto que se conoce la varianza en tiempo -1, las predicciones a un paso están disponibles. A continuación, las predicciones multipaso pueden ser computadas de forma recursiva. Un modelo más parsimonioso que ARCH es el llamado modelo ARCH generalizado (GARCH) (Bollerslev et al., 1994[120] y Taylor, 1987 [121]) donde se permiten dependencias adicionales de los retardos de la varianza condicional. Un modelo GARCH tiene una representación tipo ARMA, de forma que los modelos comparten muchas propiedades.

La familia de modelos GARCH, y muchas de sus extensiones, son ampliamente revisadas en Bollerslev, Chou y Kroner (1992) [122], Bera y Higgins (1993) [123] y Diebold y Lopez (1995) [124]. No es sorprendente que muchos de los trabajos teóricos hayan aparecido en literatura de econometría. Por otra parte, es interesante el hecho de que ni la IJF ni la JoF hayan llegado a ser foros para publicaciones sobre el rendimiento relativo de predicción de los modelos tipo GARCH o de otros varios modelos de volatilidad en general. Como puede verse a continuación, muy pocos trabajos d IJF/JoF han tratado este tema.

Sabbatini y Linton (1998) mostraron que el modelo simple lineal GARCH(1,1) proporciona una buena parametrización para las rentabilidades diarias en el índice de mercado suizo. Sin embargo, la calidad de las predicciones fuera de muestra sugiere que este resultado hay que tomarlo con precaución. Franses y Ghijsels (1999) [125] remarcaron que esta característica puede deberse al despreciar valores atípicos aditivos (AO). Se dieron cuenta de que los modelos GARCH para las rentabilidades corregidas resultaban en unas predicciones mejoradas de la volatilidad del mercado de valores. Brooks (1998) [126] no ve un ganador claro al comparar predicciones de un paso a partir de modelos tipo GARCH estándar (simétricos) con aquellas de otros modelos lineales y ANNs. A nivel de estimación, Brooks, Burke, y Persand (2001) [127], argumentan que los paquetes de software econométrico pueden producir resultados ampliamente variados. Claramente, esto puede tener algún impacto en la precisión de predicción de los modelos GARCH. Esta observación es muy al estilo de la de Newbold et al. (1994) [128] para modelos ARMA univariados. Fuera de la IJF, Karanasos (2001) [129] consideró los efectos en media de las predicciones multipaso de modelos ARMA con GARCH. Su método puede emplearse en la derivación de predicciones multipaso a partir de modelos más complicados, incluyendo GARCH multivariado.

Usando dos series de tasas de cambio diarias, Galbraith y Kisinbay (2005) comparó las funciones de contenido de predicción a partir del modelo GARCH estandar y del modelo GARCH fracionariamente integrado (FIGARCH) (Baillie, Bollerslev, y Mikkelsen, 1996) [130]. Las predicciones de varianzas condicionales parecen tener un contenido de información de aproximadamente 30 días de mercado. Otra conclusión es que las predicciones por proyección autorregresiva sobre volatilidades pasadas da mejores resultados que las predicciones basadas en GARCH, estimadas por probabilidad cuasi máxima y los modelos FIGARCH.

Esto parece confirmar los resultados anteriores de Bollerslev y Wright (2001), por ejemplo. Una crítica que se oye a menudo sobre los modelos FIGARCH y sus generalizaciones es que no razones económicas por las que la predicción financiera de volatilidad tenga memoria larga. Para una crítica más fundamentada del uso de modelos de memoria larga véase Granger (2002) [131].

Empíricamente, las rentabilidades y la varianza condicional de las rentabilidades del siguiente periodo están correladas negativamente. Este fenómeno se denomina volatilidad asimétrica en la literatura (véase Engle y Ng (1993) [132]). Esto motivó a los investigadores para que desarrollaran varios modelos tipo GARCH asimétricos (incluyendo el GARCH con cambio de régimen); véase Hentschel (1995) [133] y Pagan (1996) [134] para resúmenes generales. Awartani y Corradi (2005) investigaron el impacto de la asimetría en la capacidad de predicción fuera de muestra de diferentes modelos GARCH, con varios horizontes.

Paralelamente a GARCH, muchos otros modelos se han propuesto para predicción de volatilidad. Poon y Granger (2003) [135], proporcionan una revisión cuidadosamente realizada de la investigación en esta área durante los últimos 20 años. Compararon los hallazgos de predicción de volatilidad de 93 publicaciones. Dan importantes ideas en cuestiones como evaluación de predicción, el efecto de la frecuencia de los datos en la precisión de la predicción de volatilidad, medidas de *volatilidad actual*, el efecto de confusión de valores extremos, y otros más. El estudio encontró que la volatilidad implícita proporciona mejores predicciones que los modelos de series temporales. Entre los modelos de series temporales (44 estudios) no había ganador claro entre los modelos de histórico de volatilidad (incluyendo camino aleatorio, medias históricas, ARFIMA, y varias formas de alisado exponencial) y los modelos tipo GARCH (incluyendo ARCH y sus varias ampliaciones), pero ambas clases de modelos mejoraban el rendimiento del modelo de volatilidad estadístico; véase Poon y Granger (2005) para una revisión de estos hallazgos.

El estudio de Poon y Granger contiene muchas cuestiones para continuar estudiando. Por ejemplo, los modelos asimétricos GARCH salían bien parados en los concursos de predicción. Sin embargo, no está claro hasta qué punto esto se debe a las asimetrías de la media condicional, de la varianza condicional, y/o de momentos condicionales de mayor orden. Otra cuestión para investigación futura concierne la combinación de predicciones. Los resultados en dos trabajos (Doidge y Wei, 1998 [136] y Kroner et al., 1995 [137]) encuentran de ayuda la combinación, pero en otro estudio (Vasilellis y Meade, 1996) [138] no. Sería también útil examinar el rendimiento de predicción de volatilidad de modelos tipo GARCH multivariados y no lineales, incorporando a ambos dependencias contemporáneas y temporales.

2.8. Predicción de datos de recuento

Los datos de recuento son frecuentes en la industria y los negocios, especialmente en datos de inventario donde son llamados *datos de demanda intermitente*. Consecuentemente, es sorprendente que haya tan poco trabajo realizado en predicción de estos datos. Hay algo sobre métodos a medida para predicción de datos de recuento, pero pocos artículos aparecen sobre predicción de series temporales de datos de recuento que usen modelos estadísticos.

La mayoría de la investigación en predicción de recuentos está basada en Croston (1972) [139] quien propuso usar SES para predecir independientemente los valores no cero de las series y los intervalos de tiempo entre ellos. Willemain, Smart, Shockor y DeSautels (1994) [140] compararon el método de Croston con SES y encontraron que el método de Croston era más robusto, aunque estos resultados se basaban en en MAPEs que a menudo no están definidos para datos de recuento. Las condiciones bajo las cuales el método de Croston mejora al SES fueron discutidas en Johnston y Boylan (1996) [141]. Willemain, Smart y Schwarz (2004) propusieron un procedimiento de remuestreo para datos de demanda intermitente que se encontró que era más preciso que el de SES y el de Croston para las nueve series evaluadas.

La evaluación de predicciones de recuento conlleva dificultades debidas a la presencia de ceros en los datos observados. Syntetos y Boylan (2005) [142] propusieron usar el error absoluto medio relativo,mientras que Willemain et al. (2004) [143] recomendaron usar el método de transformación integral de probabilidad de Diebold, Gunther y Tay (1998) [144].

Grunwald, Hyndman, Tedesco y Tweedie (2000) revisaron muchos modelos estadísticos para series temporales de recuento, usando AR de primer orden como marco unificador para los diferentes enfoques. Un posible modelo, explorado por Brännäs (1995) [145], supone que la serie sigue una distribución de Poisson con media que depende de procesos no observados autocorrelados. Un modelo alternativo MA de valor entero fue usado por Brännäs, Hellström y Nordström (2002) para predecir niveles de ocupación de hoteles suizos.

La distribución de predicción se puede obtener por simulación usando cualquiera de estos modelos estadísticos, pero cómo resumir la distribución no es algo obvio. Freeland y McCabe (2004) propusieron usar la mediana de la distribución de predicción, y dieron un método para computar intervalos de confianza para la distribución completa en el caso de modelos autorregresivos de valores enteros (INAR) de orden 1. McCabe y Martin (2005) ampliaron estas ideas presentando metodologías bayesianas para predicción a partir de la clase de modelos INAR.

2.9. Evaluación de predicción y medidas de precisión

Se ha usado un desconcertante conjunto de medidas de precisión para evaluar el rendimiento de los métodos de predicción. Algunos de los cuales se listan en Mahmoud (1984) [146]. Primero definimos las medidas más comunes.

(poner tabla)

La evolución de las medidas de precisión y evaluación de predicción puede verse a través de las medidas usadas para evaluar los métodos en la mayoría de los estudios comparativos realizados.En la competición original M (Makridakis et al., 1982) [50] las medidas usadas incluían MAPE, MSE, AR, MdAPE y PB. Sin embargo, como apuntaron Chatfield (1988) [54] y Armstrong y Collopy (1992) [147], el MSE no es apropiado para comparaciones entre series puesto

Tabla 2. Medidas normalmente usadas para precisión de predicción

Medida	**Descripción**	**Expresión**
MSE	Mean squared error	$mean(e_t^2)$
RMSE	Root mean squared error	$\sqrt{MSE}$
MAE Mean	Absolute error	$mean(\|e_t\|)$
MdAE	Median absolute error	$median(\|e_t\|)$
MAPE	Mean absolute percentage error	$mean(\|p_t\|)$
MdAPE	Median absolute percentage error	$median(\|p_t\|)$
sMAPE	Symmetric mean absolute percentage error	$mean(2\|Y_t - F_t\|/(Y_t - F_t))$
sMdAPE	Symmetric median absolute percentage error	$median(2\|Y_t - F_t\|/(Y_t - F_t))$
MRAE	Mean relative absolute error	$mean(r_t)$
MdRAE	Median relative absolute error	$median(r_t)$
GMRAE	Geometric mean relative absolute error	$gmean(r_t)$
RelMAE	Relative mean absolute error	MAE/MAE_b
RelRMSE	Relative root mean squared error	$RMSE/RMSE_b$
LMR	Log mean squared error ratio	$log(RelMSE)$
PB	Percentage better	$100mean(/\{r_t < 1\})$
PB(MAE)	Percentage better (MAE)	$100mean(/\{MAE < MAE_b\})$
PB(MSE)	Percentage better (MSE)	$100mean(/\{MSE < MSE_b\})$

que depende de la escala. MAPE también tiene problemas cuando la serie tiene valores cercanos a cero o cero, como apuntaron Makridakis, Wheelwright y Hyndman (1998, p.45) [148]. En la competición M se evitaron MAPEs excesivamente grandes (o infinitos) incluyendo sólo datos positivos. Sin embargo, es una solución artificial imposible de aplicar en todas las situaciones.

en 1992, aparecieron dos artículos de IJF y varios comentarios sobre las medidas de evaluación. Armstrong y Collopy (1992) [147] recomendaron usar errores absolutos relativos, especialmente GMRAE y MdRAE, a pesar del hecho de que los errores relativos tienen varianza infinita y media no definida. Ellos recomendaron *winsorizar* para recortar valores extremos lo que parcialmente resolvía estos problemas, pero lo cual añade complejidad al cálculo y un nivel de arbitrariedad al tener que especificar la cantidad de recorte. Fildes (1992) [149] también prefirió GMRAE aunque lo expresó en una forma equivalente como la raíz cuadrada de la media geométrica de los errores relativos cuadráticos. Esta equivalencia no parece ser advertida por ninguno de los ponentes en los comentarios de Ahlburg et al. (1992) [150].

El estudio de Fildes, Hibon, Makridakis y Meade (1998), el cual se enfocaba en predicción de datos de telecomunicaciones, usó MAPE, MdAPE, PB, AR, GMRAE y MdRAE teniendo en cuenta algunas de las críticas de los métodos usados para la competición M.

La competición M3 (Makridakis y Hibon, 2000) [151] usó tres medidas de precisión: MdRAE, sMAPE y sMdAPE. Las medidas *simétricas* fueron propuestas por Makridakis (1993) [152] en respuesta a la observación de que MAPE y MdAPE tienen la desventaja de que penalizan más errores positivos que negativos. Sin embargo, estas medidas no son tan simétricas como su nombre sugiere. Para el mismo valor de Yt, el valor 2*Abs(Yt-Ft)/(Yt+Ft) tiene más penalización cuando las predicciones son altas en comparación con cuando son bajas. Véase Goodwin y Lawton (1999) [153] y Koehler (2001) [154] para más disertación sobre este punto.

Es notable que ninguno de los estudios comparativos ha usado medidas relativas (a diferencia de medidas de errores relativos) tales como RelMAE o LMR. Esta última fue propuesta por Thompson (1990) [155]quien abogó por su uso basado en sus buenas propiedades estadísticas. Fue aplicada a los datos de la competición M en Thompson (1991) [156].

A parte de Thompson (1990) [155], ha habido poco trabajo teórico en las propiedades estadísticas de estas medidas. Una excepción es Wun y Pearn (1991) quienes estudiaron las propiedades estadísticas de MAE.

Una novedosa medida alternativa de precisión es la *distancia temporal*, que fue considerada por Granger y Jeon, 2003a [157] y Granger y Jeon 2003b [158]. En esta medida se capturan también las propiedades de dirección y desfase de la predicción. De nuevo, esta medida no ha sido usada en ningún estudio comparativo importante.

Una línea de investigación paralela ha examinado tests estadísticos para comparar métodos de predicción. Una contribución precoz fue Flores (1989) [159]. El mejor enfoque conocido para comprobar diferencias entre la precisión de los métodos de predicción es el test de Diebold y Mariano (1995) [124]. Harvey, Leybourne y Newbold (1997) [160] propusieron una modificación de este test con tamaño corregido. McCracken (2004) [161] examinó el efecto de estimación de parámetros en estos tests y proporcionó un nuevo método de ajuste para error de estimación de parámetros.

Otro problema en evaluación de predicción, y más serio que el error de estimación de parámetros, es el *intercambio de datos*, el uso de los mismo datos para diferentes métodos de predicción. Sullivan, Timmermann y White (2003) [162] propusieron un procedimiento de remuestreo diseñado para superar la distorsión resultante de inferencia estadística.

Una línea de investigación independiente ha examinado las propiedades teóricas de predicción de modelos de series temporales. Una contribución importante fue Clements y Hendry (1993) [163] quienes mostraron que el MSE teórico de un modelo de predicción no era invariante a transformaciones lineales con preservación de escala tales como diferenciación de los datos. En su lugar, propusieron el criterio de *momento segundo de error de predicción generalizado* (GFESM), el cual no tiene esa propiedad no deseable. Sin embargo, es difícil aplicar tales medidas empíricamente y la idea parece no haberse extendido ampliamente.

2.10. Combinación de métodos

Durante las últimas 3 décadas se ha estudiado la combinación, mezcla y puesta en común de predicciones obtenidas a partir de diferentes métodos de series temporales y diferentes fuentes de información. Contribuciones prematuras en este área son Bates y Granger (1969) [164], Newbold y Granger (1974) [165] y Winkler y Makridakis (1983) [166]. En una revisión bibliográfica exhaustiva, Clemen (1989) [167] resumió una evidencia irresistible de la eficiencia relativa de predicciones combinadas, normalmente en términos de varianzas de error de predicción.

Se han propuesto numerosos métodos para seleccionar los pesos de combinación. La simple media es el método de combinación más ampliamente usado (véase la revisión de Clemen y Bunn, 1985 [168]), pero el método no usa información pasada respecto a la precisión de las predicciones o la dependencia entre las predicciones. Otro simple método es la mezcla lineal de las predicciones individuales con pesos de combinación determinados por OLS (asumiendo insesgamiento) a partir de la matriz de predicciones pasadas y el vector de observaciones pasadas (Granger y Ramanathan, 1984) [169]. Sin embargo, las estimaciones OLS de los pesos son insuficientes debido a la posible presencia de correlación en serie en los errores de predicción combinados. Aksu y Gunter (1992) [170] y Gunter (1992) [171] investigaron este problema con algo más de detalle. Ellos recomendaron el uso de predicciones combinadas OLS con la restricción de que los pesos sumen uno. Granger (1989) [172] dio varias ampliaciones de la idea original de Bates y Granger (1969) [164] incluyendo combinación de predicciones con horizontes mayores que un período.

Mejor que usar pesos fijos, Deutsch, Granger y Teräsvirta (1994) [173] permitieron que estos cambiaran a lo largo del tiempo usando modelos STAR y de cambio de régimen. Otro esquema de ponderación variante en el tiempo fue propuesto por Fiordaliso (1998) [174], quien usó un sistema difuso para combinar un conjunto de predicciones individuales de forma no lineal. Diebold y Pauly (1990) [175] usaron técnicas de contracción bayesiana para permitir la incorporación a priori de información en la estimación de los pesos de combinación. Zou y Yang (2004) [176] consideraron la combinación de predicciones de modelos muy similares con actualización secuencial de pesos.

La combinación de pesos determinada a partir de los métodos invariantes en tiempo puede llevar a predicciones relativamente pobres si ocurre no estacionariedad entre las diferentes componentes de predicción. Miller, Clemen y Winkler (1992) [177] examinaron el efecto del desplazamiento de ubicación de la no estacionariedad en un conjunto de métodos de combinación. Concluyeron que la simple media batía otros artefactos de combinación complejos, véase también Hendry y Clements (2002) [178] para resultados más recientes. El tema relacionado de combinar predicciones a partir de modelos lineales y otros no lineales, con pesos OLS así como determinados por un método variante en el tiempo fue abordado por Terui y van Dijk (2002)[179].

La forma de la distribución del error de predicciones combinadas y el correspondiente comportamiento estadístico fue estudiado por de Menezes y Bunn (1998) [180] y Taylor y Bunn (1999) [181]. Para distribuciones de error de predicción no normales, la oblicuidad surge como un criterio relevante para especificar el método de combinación. Fang (2003) [182] dio algunas ideas de por qué las predicciones en competición pueden ser fructuosamente combinadas para producir una predicción superior a las predicciones individuales abarcando tests de predicción. Hibon y Evgeniou (2005) [183] propusieron un criterio para seleccionar entre predicciones y sus combinaciones.

2.11. Intervalos de predicción y densidades

El uso de intervalos de predicción, y más recientemente densidades de predicción, ha ido siendo más usual a lo largo los últimos 25 años conforme los profesionales han sido más conscientes de las limitaciones de las predicciones puntuales. Una revisión importante y completa de los intervalos de predicción se da en Chatfield (1993) [184], resumiendo la literatura hasta el momento.

Desafortunadamente, hay algo de confusión en la terminología y algunos autores usan *intervalo de confianza* en lugar de *intervalo de predicción*. El intervalo de confianza se usa para un parámetro del modelo, mientras que el intervalo de predicción se usa para una variable aleatoria. Casi siempre, los pronosticadores, van a querer intervalos de predicción, intervalos que contengan los valores verdaderos de observaciones futuras con su probabilidad específica.

La mayoría de los intervalos de predicción están basados en un modelo estadístico subyacente. Por tanto, ha habido bastantes trabajos para formular modelos estadísticos apropiados para generalizar procedimientos de predicción comunes.

El vínculo entre las fórmulas de intervalos de predicción y el modelo a partir del que se derivan no siempre se ha examinado correctamente. Por ejemplo, el intervalo de predicción apropiado para un modelo de camino aleatorio fue aplicado por Makridakis y Hibon (1987) [185] y Lefrançois (1989) [186] para predicciones obtenidas a partir de muchos otros métodos. Este problema fue advertido por Koehler (1990) [187] y Chatfield y Koehler (1991) [188].

La incertidumbre asociada a la selección del modelo y la estimación de parámetros no se tiene en cuenta con la mayoría de intervalos de predicción basados en modelo. Por tanto, los intervalos son demasiado estrechos. Ha habido investigación considerable en cómo hacer que los intervalos de predicción basados en modelo tengan una cobertura más realista. Una serie de artículos apareció sobre el uso de remuestreo para crear intervalos de predicción basados en modelo para el modelo AR, empezando por Masarotto (1990) [189] e incluyendo a McCullough, 1994 [190], McCullough, 1996 [191], Grigoletto (1998) [192], Clements y Taylor (2001) [193] y Kim (2004b) [194]. Procedimientos similares para otros modelos también se han considerado incluyendo modelos ARIMA (Pascual et al., 2001[195], Pascual et al., 2004 [196], Pascual et al., 2005 [197] y Wall y Stoffer, 2002[198]), VAR (Kim, 1999 [199] y Kim, 2004a [194]), ARCH (Reeves, 2005 [200]) y regresión (Lam & Veal, 2002 [201]). Parece probable que estos métodos de remuestreo se utilicen más conforme las velocidades de cómputo aumente por sus mejores propiedades de cobertura.

Cuando el error de predicción no sigue una distribución normal, hallar la densidad de predicción completa es útil puesto que un intervalo único puede no proporcionar un resumen adecuado del futuro esperado. Un resumen sobre densidad de predicción se da en Tay y Wallis (2000) [202] junto con varios otros artículos del mismo número especial de la JoF. Resumiendo, una densidad de predicción ha sido objeto de interesantes propuestas incluyendo los *gráficos de ventilador* (Wallis, 1999) [203] y las *regiones de mayor densidad* (Hyndman, 1995 [204]). El uso de estos resúmenes gráficos ha crecido rápidamente en los años recientes y las densidades de predicción se utilizan de forma relativamente amplia.

Conforme los intervalos de predicción y las densidades van siendo más usadas, la atención se centra en su evaluación y testeo. Diebold, Gunther y Tay (1998) [144] introdujeron el método notablemente sencillo de *transformación integral de probabilidad*, el cual puede usarse para evaluar una densidad univariada. Este enfoque ha llegado a usarse ampliamente en un periodo de tiempo muy corto y ha sido un avance clave en este área. La idea se extiende para densidades de predicción multivariada en Diebold, Hahn y Tay (1999) [205].

Otros enfoques para evaluar la densidad y los intervalos de predicción se da en Wallis (2003) [206] quien propuso tests chi cuadrado para ambos intervalos y densidades, y Clements y Smith (2002) quienes disertaron sobre algunos tests sencillos pero potentes para evaluar densidades de predicción multivariadas.

2.12. Tendencias en la investigación de los métodos convencionales

En las secciones previas, se ha examinado la historia de las series temporales de la IJF, con la esperanza de arrojar luz al presente. De cara al futuro, es interesante reflejar las propuestas de investigación identificadas.

Chatfield (1988) [54] remarcó la necesidad de investigar métodos multivariados enfatizando en hacerlos más prácticos. Ord (1988) [207] también advirtió que no se había realizado suficiente trabajo en modelos de múltiples series temporales, incluyendo alisado exponencial multivariado. Hasta la fecha la predicción de series temporales multivariadas todavía no es aplicada ampliamente, a pesar de los considerables avances en este área. Se sospecha que existen dos razones para ello: una falta de investigación empírica sobre algoritmos de predicción robustos para modelos multivariados, y la escasez de software fácil de usar. Algunos de los métodos sugeridos (modelos VARIMA) son difíciles de estimar por el gran número de parámetros que incluyen. Otros como el alisado exponencial multivariado, no han adquirido suficiente atención teórica para su aplicación rutinaria. Un enfoque para predicción de series temporales multivariadas consiste en usar modelos de factor dinámico. Estos modelos promete tanto teóricamente (Forni et al., 2005 [208] y Stock y Watson, 2002 [209]) como para su aplicación (Peña y Poncela, 2004 [210]) y se cree que llegarán a usarse más ampliamente en el futuro.

Ord (1988) [207] también indicó la necesidad de investigar más profundamente en los métodos de predicción basados en modelos no lineales. Si bien muchos aspectos de los modelos no lineales se han investigado en la IJF, se debe seguir investigando de forma continuada. Por ejemplo, no existe un claro consenso de que las predicciones de los modelos no lineales superen categóricamente las de los modelos lineales (véase Stock y Watson, 1999 [211]).

Otros temas que sugiere Ord (1988) [207] incluye la necesidad de desarrollar procedimientos de selección de modelo que hagan un uso efectivo tanto de los datos como del conocimiento a priori, a la necesidad de especificar objetivos para las predicciones y desarrollar sistemas de predicción dirigidos hacia estos objetivos. Estas áreas todavía necesitan atención y se espera que aparezcan herramientas para solventar estos problemas.

Dado el frecuente mal uso de métodos basados en modelos lineales con distribuciones de error gaussianas independientes e idénticas, Cogger (1988) [212] abogó por que los nuevos desarrollos del área de métodos estadísticos *robustos* recibieran más atención dentro de la comunidad de la predicción de series temporales. Un procedimiento robusto se espera que funcione bien cuando existan valores atípicos o desplazamientos de ubicación en los datos que sean difíciles de detectar. La estadística robusta puede estar basada tanto en métodos paramétricos como no paramétricos. Un ejemplo de este último, es el concepto de cuantiles de regresión de Kkoenker y Bassett (1978) [213] investigado por Cogger. En predicción, estos puede aplicarse como cuantiles condicionados univariados y multivariados. Un área de aplicación importante es al estimar herramientas de manejo de riesgo como el *valor en riesgo*. Recientemente, Engle y Manganelli (2004) [214] realizaron un comienzo en esta dirección, proponiendo un modelo de valor en riesgo condicional. Se espera que haya muchas más investigación en este área.

Un tema relacionado en donde ha habido bastante actividad investigadora es la densidad de predicción, donde el foco está en la densidad de probabilidad de observaciones futuras más allá de la media o la varianza. Por ejemplo, Yao y Tong (1995) [215] propusieron el concepto de intervalo de predicción de percentil condicional. Su anchura ya no es una constante, como en el caso de los modelos lineales, y puede variar respecto a la posición del espacio de estado a partir del que se realizan las predicciones; véase también De Gooijer y Gannoun (2000) [216] y Polonik y Yao (2000) [217].

Claramente, el área de los intervalos de predicción mejorada requiere mayor investigación. Esto está en concordancia con Armstrong (2001) [218] quen listó 23 principios de gran necesidad de interés incluyendo el elemento 14:13: *Para intervalos de predicción, incorporar la incertidumbre asociada con la predicción de las variables explicativas.*

En los años recientes, las series temporales no gaussianas, han comenzado a recibir una atención considerable y los métodos de predicción van lentamente siendo desarrollados. En particular las series no gaussiana con valores positivos tiene importantes aplicaciones. Dos áreas importantes son la volatilidad y la duración entre transacciones. Algunas contribuciones importantes hasta la fecha han sido el modelo de *duración condicional autorregresivo* de Engle y Russell (1998) [219] y Andersen, Bollerslev, Diebold y Labys (2003) [220]. Dada la importancia de estas aplicaciones, se espera mucho más trabajo en este área.

Si bien, las series temporales no gaussianas con espacio de muestreo continuo han comenzado a recibir mayor atención de investigación, especialmente en el contexto de las finanzas, la predicción de series temporales con un espacio muestral discreto (tales como las series de recuentos) todavía está en su infancia. Tales datos son muy frecuentes en la industria y los negocios y existen muchos problemas asociados a la predicción de recuentos sin resolver teóricamente o de forma práctica; por tanto, se espera una mayor investigación productiva en este área en el futuro cercano.

Otros autores han intentado identificar temas de investigación importantes. Tanto De Gooijer (1990) [221] como Clements (2003) [222] en dos editoriales, y Ord como parte de un artículo de disertación de Dawes, Fildes, Lawrence y Ord (1994) [223] sugirieron más trabajo en predicciones combinadas. Aunque el tema ha recibido una cantidad de atención considerable, todavía hay algunas cuestiones abiertas. Por ejemplo, cuál es el mejor método de combinación para modelos no lineales y lineales y qué intervalo de predicción puede establecerse en torno a la predicción combinada. Un buen punto de partida para mayor investigación en este área es Teräsvirta (2006) [224]; véase también Armstrong (2001, 12.5-12.7) [218]. Recientemente, Stock y Watson (2004) [225] discutieron el *puzle de combinación de predicción*, a saber, el hallazgo empírico repetido de que combinaciones simples como las medias superan combinaciones más sofisticadas en las que la teoría sugiere mejor funcionamiento. Esta importante cuestión práctica sin duda recibirá mayor atención de investigación.

Los cambios en el almacenamiento y recolección de datos también conducirá a nuevas direcciones de investigación. Por ejemplo, en el pasado, los paneles de datos disponibles (llamados datos longitudinales en bioestadística) tenían la dimensión temporal de la serie t pequeña mientras que la dimensión de sección cruzada n era grande. Sin embargo, ahora en muchas áreas aplicadas como marketing, se pueden obtener fácilmente grandes conjuntos de datos con ambos n y t grandes. La extracción de características a partir de lo paneles de datos es el objeto del *análisis de datos funcionales*; véase por ejemplo Ramsay y Silverman (1997) [226]. Sin embargo, el problema de realizar predicciones multi-paso basadas en datos funcionales sigue abierto tanto a investigación teórica como práctica. Dada el predominio creciente de este tipo de datos, se espera que este área sea fructífera en el futuro.

Los conjuntos de datos grandes, también se prestan a métodos de computación intensiva. Mientras las redes neuronales se han usado en predicción durante más de dos décadas, existen muchas cuestiones excepcionales asociadas con su uso e implementación, incluyendo el cuándo van a mejorar probablemente otros métodos. Otros métodos que implican computación pesada (como bagging y boosting) son incluso menos entendidos en el contexto de predicción. Con la disponibilidad de conjuntos de datos muy grandes y potencia alta de computación se espera que ésta sea un área de investigación importante.

3. Machine Learning (ML) en predicción de series temporales

En las últimas dos décadas los modelos de ML se han establecido como serios competidores a los modelos estadísticos convencionales en la comunidad de predicción [227],[228],[229]. Estos modelos, también llamados de caja negra o modelos orientados a datos [230], son ejemplos de modelos lineales no paramétricos en los que sólo se usan datos históricos para aprender la dependencia estadística entre valores pasados y futuros. Werbos encontró que ANN mejoraba los métodos estadísticos clásicos tales como regresión lineal o el enfoque Box-Jenkins [231] [232]. En un estudio similar realizado por Lapedes y Farber [233] se concluye que las ANNs pueden ser usadas satisfactoriamente para modelado y predicción de series temporales no lineales. Posteriormente, otros modelos han aparecido tales como árboles de decisión, máquinas de vectores soporte y regresión por vecinos más cercanos [234] [235]. Además se ha examinado en varias competiciones de predicción la precisión empírica de varios modelos de aprendizaje automático bajo diferentes condiciones de datos (NN3, NN5 y las competiciones ESTSP [236][237][238]) creando interesantes debates en el área de minería de datos y predicción [239][240][236].

A continuación se revisa el rol de las técnicas de aprendizaje automático bajo tres aspectos principalmente: la formalización de problemas de predicción de un paso como tareas de aprendizaje supervisado, el uso de técnicas de aprendizaje local como herramienta para tratar datos temporales y el papel de la estrategia de predicción al pasar de predicciones de un paso a multi-paso.

3.1. Dependencias temporales

El aprendizaje supervisado consiste en el modelado, bajo un conjunto de observaciones finitas, de la relación entre un conjunto de variables de entrada y una o más variables de salida, las cuales se consideran de alguna manera dependientes de la entrada. Una vez que el modelo de esta relación se obtiene, es posible usarlo para predicción a un paso. En predicciones a un paso, los n valores previos de la serie son la entrada para formular un problema genérico de regresión.

El enfoque general para modelar la relación de entrada/salida, con la salida escalar y la entrada vectorial, se basa en la disponibilidad de un conjunto de pares observados conocido como el conjunto de entrenamiento, donde el vector de entrada contiene los n valores previos y el valor de salida es el sucesor a aquellos.

Técnicas de aprendizaje local

Si bien existen multitud de técnicas de aprendizaje supervisado para predicción, las técnicas de aprendizaje local que se describen a continuación están motivadas por las siguientes razones:

- No requieren demasiadas suposiciones: el aprendizaje local no supone un conocimiento a priori del proceso subyacente a los datos. La única información disponible se representa por un conjunto de pares entrada/salida. Esta característica es especialmente importante en conjuntos de datos reales donde puede haber valores perdidos y las relaciones no son estacionarias.
- Capacidad de aprendizaje en línea: el número de muestras de entrenamiento puede incrementarse en el tiempo. En este caso, el aprendizaje local simplemente añade los nuevos puntos al conjunto y no necesita re-entrenamiento al aparecer nuevos datos.
- Modelado de no estacionariedad: los métodos de aprendizaje local pueden tratar configuraciones variables en el tiempo, donde el proceso estadístico subyacente de los datos no es estacionario. En este caso, es suficiente con interpretar la noción de vecindad en un sentido espacio-temporal. Para cada punto, los vecinos son muestras similares obtenidas recientemente. Por tanto la variable temporal es tenida en cuenta para mejorar la precisión de las predicciones.

A continuación se describen dos técnicas de aprendizaje local, a saber, la técnica de vecinos más cercanos y el aprendizaje vago.

- **Vecinos más cercanos**: El método de vecinos más cercanos es el ejemplo más sencillo de enfoque local aplicado al problema de predicción. Este método consiste en buscar en el conjunto de datos los vecinos más cercanos al estado actual y predecir que evolucionará de igual manera que lo hizo su vecindad. Este enfoque fue por primera vez propuesto por Lorenz [241]. Algunas extensiones consideran más vecinos (aproximaciones de mayor orden) [242]. Tong y Lim [243] introdujeron análisis de series temporales por aproximación lineal por partes. Priestley [244] sugirió la importancia de aproximaciones de mayor orden. Farmer y Sidorowich [245] [246] estudiaron el enfoque local y demostraron su efectividad en series temporales sobre varios experimentos numéricos.

- **Aprendizaje vago**: Esta técnica adapta automáticamente el tamaño de la vecindad según un criterio de validación cruzada. El atractivo de esta técnica es su naturaleza tipo *divide y vencerás*, ya que reduce un problema complejo y no lineal a una secuencia de problemas locales manejables. Esto permite explotar el rango completo de técnicas de validación e identificación las cuales son rápidas, fiables y poseen una justificación teórica profunda. El procedimiento consiste básicamente en los siguientes pasos:
  1. Ordenar de forma ascendiente el conjunto de vectores según la distancia al vector sobre el que se quiere predecir el siguiente valor.
  2. Determinar el número óptimo de vecinos.
  3. Calcular, dado el número de vecinos, la predicción mediante un modelo local (constante o lineal).

  Este método ha sido aplicado con éxito a varios problemas de regresión y predicción [247]. En [248] [249] se dan más detalles sobre la técnica de aprendizaje vago y sus aplicaciones.

Predicción para horizontes mayores de un paso

A continuación se consideran tres técnicas para abordar el problema de predicción multi-paso.

1. Mediante recursividad. Primero se entrena un modelo de un paso y se utiliza para una predicción multi-paso. Una desventaja conocida de este método es su sensibilidad al error estimado pues se acumula en predicciones sucesivas.
2. De forma directa. Esta estrategia entrena tantos modelos como pasos del horizonte, y concatena las soluciones de cada uno formando la predicción del horizonte. No está sujeta a acumulación de errores pero requiere mayor complejidad funcional [105] para modelar la dependencia entre valores de la serie distantes [250]. También tiene la desventaja de ser más costosa computacionalmente al tener que aprender tantos modelos como pasos tenga el horizonte. Entre los modelos usados para implementar este método destacan las redes neuronales [251] y árboles de decisión [252].
3. Predicción híbrida. Aquí se combinan las dos arquitecturas anteriores [253]. Para cada horizonte se calcula un modelo pero se añaden los puntos de entrada predichos.

Estrategia de múltiples salidas

El problema de los enfoques mencionados anteriormente es que obvian la existencia de dependencias estadísticas entre los valores de predicción consecutivos. Para remediar esto, se utilizan modelos para aprender la relación entre el vector de valores pasados y otro vector de valores futuros. La estrategía MIMO (*Multi-Input Multi-Output*) [254] [255] evita obviar la dependencia condicional entre los valores futuros aprendiendo un modelo con múltiples salidas. Este método ha sido aplicado a varios problemas reales [254] [255] [256] [257], sin embargo presenta la desventaja de tener que usar la misma estructura del modelo para predecir cualquier horizonte [256]. Una posible variante para superar esta falta de flexibilidad consiste en particionar el horizonte de tamaño H en m bloques de tamaño s y usar MIMO para predecir en cada bloque, lo que se conoce como DIRMO (mezcla entre DIRecta y MIMO). Esto permite calibrar la dependencia de las salidas (sin dependencia en caso de $s = 1$ y máxima dependencia con $s = H$

Aprendizaje local para predicción multipaso
A continuación se discuten algunos trabajos que usan técnicas de aprendizaje local para tratar específicamente el problema de predicción a largo plazo. En [258] [249] se modifica el aprendizaje local para tener en cuenta el comportamiento temporal de la predicción multi-paso y mejorar así la estrategia recursiva. Otra mejora reciente de la estrategia recursiva es RECNOISY [259], la cual introduce perturbaciones en el conjunto de entrenamiento para cada paso del proceso de predicción de forma que pueda manejar de manera más adecuada los valores aproximados.

Dos mejoras de aprendizaje vago fueron presentadas en [260]. La primera se basa en una poda iterativa de las entradas, la segunda realiza búsqueda por fuerza bruta en el conjunto posible de entradas mediante una aproximador de k vecinos más cercanos (K-NN).

En [254] se propone aprendizaje local para predicción MIMO donde se extiende el algoritmo básico de aprendizaje vago para múltiples salidas así como una estrategia de ponderación de los diferentes predictores a largo plazo para mejorar la precisión resultante.

En [261] [257] se presenta la estrategia DIRMO con aprendizaje local. Este método ha sido aplicado con éxito en las competiciones de predicción: ESTSP'07 [261] y NN3[257]. Una revisión más detallada y comparativa de las estrategias para predicción de series temporales multi-paso basadas en el algoritmo de aprendizaje local se puede encontrar en [7].

4. Redes neuronales en predicción

Una red neuronal artificial (ANN) puede ser útil para procesos no lineales que tengan relaciones funcionales desconocidas y por tanto difíciles de ajustar (Darbellay y Slama, 2000 [262]). La idea principal con ANN es que las entradas, o variables dependientes, son filtradas mediante una o más capas ocultas cada una de las cuales consta de nodos antes de alcanzar la variable de salida. La salida intermedia se enlaza a la salida final. Otros modelos son versiones específicas de ANN donde se impone cierta estructura (véase JoF Special Issue 17:5/6 (1998).

Una de las mayores áreas de aplicación de ANN es predicción; véase Zhang, Patuwo y Hu (1998) [229] y Hippert, Pedreira y Souza (2001) [263] para revisiones de la literatura. Sin embargo no hay un claro consenso sobre cuándo es mejor utilizar ANNs para predicción.

Gorr (1994) [264] y Hill, Marquez, OConnor y Remus (1994) [265] sugirieron que la investigación a seguir debería tratar de definir mejor los límites en que ANN mejora las técnicas tradicionales y viceversa.

Un problema general con los modelos no lineales se encuentra en la complejidad de los modelos y su excesiva parametrización. Si se considera realmente importante el principio de parsimonia, es interesante comparar el rendimiento de predicción fuera de muestra de los modelos lineales frente a los no lineales, usando una amplia variedad de criterios de selección de modelo. Esta cuestión fue considerada en bastante profundidad por Swanson y White (1997) [266]. Sus resultados sugirieron que una simple ANN con una única capa oculta *feed-forward*, la cual era muy popular en series temporales econométricas, ofrece una alternativa útil y flexible a modelos lineales de especificación fija, particularmente para horizontes de predicción mayores de un paso. En contraste con Swanson y White, Heravi, Osborn y Birchenhall (2004) [267] encontraron que los modelos lineales producen predicciones más precisas de producción industrial mensual europea sin ajuste estacional que los modelos de ANN. Ghiassi, Saidane, y Zimbra (2005) [268] presentaron una ANN dinámica y compararon su rendimiento de predicción frente a la ANN tradicional y los modelos ARIMA.

Con el tiempo, la importancia del riesgo de la parametrización excesiva y el sobreajuste ha sido reconocida por varios autores; véase Hippert, Bunn y Souza (2005) [269] que usaron una gran ANN (50 entradas, 15 neuronas ocultas y 24 salidas) para predecir perfiles de carga de electricidad diaria. Sin embargo, la cuestión de si la ANN está sobre-parametrizada o no, sigue sin resolverse. Algunas ideas con valor potencial para construir ANNs minimizando el número de parámetros, usando inferencia estadística son las sugeridas en Teräsvirta, van Dijk y Medeiros (2005) [270].

4.1. Selección de retardos

La propia selección de variables de entrada para construir el modelo es en sí mismo un problema de minería de datos. En este caso es necesario encontrar la dimensión de entrada mínima capaz de representar la relación entre los datos históricos. El teorema de Takens [271] establece que una dimensión alta a permite reconstruir correctamente un espacio de estados que garantiza una dinámica topológicamente idéntica a las dinámicas de los espacios de estado del sistema real. Esta elección de la dimensión se conoce como la dimensión activa [**?**] o dimensión intrínseca.

En [**?**] se presenta un método evolutivo inspirado en el teorema de Takens para un modelo iterativo híbrido compuesto por una ANN con un algoritmo genético. El método primero itera sobre el algoritmo genético para obtener una solución con mínimo fitness y determinar así la dimensión activa. Posteriormente la ANN es ajustada.

Lukoseviciute y Ragulskis [**?**] dividen la selección de retardos en dos etapas, encontrando primero la dimensión óptima del espacio de fases mediante el algoritmo de falsos vecinos más cercanos y a continuación buscando un conjunto óptimo de retardos más cercanos mediante un algoritmo evolutivo para un sistema de inferencia difuso.

Otros métodos utilizados para determinar la dimensión intrínseca son el análisis de componentes principales (PCA) el método de Grassberger-Procaccia [272] y el método de Box-Counting.

5. Modelos aplicados a predicción del tráfico

En general las prácticas actuales en la administración y control de estrategias de tráfico están dominadas por el uso emergente de sistemas de transporte inteligente (STI), el rápido desarrollo de los computadores y la existencia de métodos algorítmicos flexibles. En los últimos años un gran número de trabajos han aparecido describiendo diversas especificaciones matemáticas para modelar las características del tráfico y producir predicciones bajo diferentes condiciones, no existiendo un claro consenso en los diferentes requerimientos involucrados durante el modelado. Como ejemplo en [274] se discute el término de fase de tráfico en la reproducción de los patrones espacio-temporales de congestión y la necesidad de una definición más precisa de los términos científicos.

A pesar de las muchas iniciativas que han surgido como respuesta a los retos de una rápida urbanización y una creciente congestión del tráfico, todavía existe una falta de comprensión acerca de la definición apropiada para un STI. En [273] se proporciona un marco metodológico para evaluar este concepto a través de diferentes indicadores aplicables a una ciudad. De esta forma evalúan varias ciudades y obtienen para cada una el grado de *inteligencia* según la puntuación obtenida por tales indicadores en cuanto al transporte se refiere.

Una revisión de los modelos y teorías sobre tráfico fue proporcionada en [275] donde son clasificados según el nivel de detalle con que se describe el flujo de vehículos. Para cada categoría discuten varias cuestiones como precisión del modelo, su aplicabilidad o calibración y validación. En esta línea [276] propone un marco de trabajo para el desarrollo de modelos de predicción de tráfico a corto plazo que consiste en desagregar el proceso de modelado en tres partes: determinar el ámbito, el proceso conceptual de especificación de la salida y el proceso de modelado en sí, que incluye la selección del enfoque metodológico apropiado, el tipo de entradas y salidas usadas y la calidad de los datos.

5.1. Redes neuronales

En la literatura encontramos un uso extensivo de redes neuronales aplicadas a predicción de tráfico con enfoques muy variopintos en cuanto al modelo de red neuronal y los algoritmos de ML utilizados. Entre los trabajos explorados se enumeran a modo de ejemplos los siguientes:

- En [277] se realiza una comparación del rendimiento de cuatro arquitecturas diferentes tipo MIMO, a saber: el perceptrón multi-capa (MLP), una red modular que consiste en varios MLPs conectados y recombinados, una red de dos capas ocultas con PCA a la entrada y un sistema de inferencia neuro-difuso denominado CANFIS.

- En [278] se estudia la capacidad de predicción del tiempo de viaje mediante una MLP demostrando que un modelo relativamente sencillo puede obtener buenos resultados.
- En [279] se utiliza un enfoque recursivo para predecir con éxito las condiciones de tráfico mediante red neuronal, si bien destacan posibles problemas de sensibilidad en los parámetros y la necesidad de llevar a cabo suficientes experimentos.
- Otro modelo de red neuronal consistente en una capa de entrada, una capa competitiva y otra de interpolación fue propuesta en [280] para reducir el tiempo de entrenamiento de backpropagation en un entorno simulado llamado TSIS (Traffic Software Integrated System).
- En [281] se emplea un modelo de red neuronal con retardos temporales (Time Delay Neural Network) evolucionado mediante algoritmos genéticos para predicción de tráfico a corto plazo.

Referencias

1. Brown, R. Statistical Forecasting for Inventory Control (1959) cited By (since 1996)127.
2. Brown, R. Smoothing, Forecasting and Prediction of Discrete Time Series (1963) cited By (since 1996)192.
3. Holt, C.: Forecasting seasonals and trends by exponentially weighted moving averages. International Journal of Forecasting **20**(1) (2004) 5–10 cited By (since 1996)78.
4. Winters, P.: Forecasting sales by exponentially weighted moving averages. Management Science **6**(3) (1960) 324–342 cited By (since 1996)264.
5. Pegels, C.: Exponential forecasting: Some new variations. Management Science **12**(5) (1969) 311–315 cited By (since 1996)35.
6. Muth, J.: Optimal properties of exponentially weighted forecasts. Journal of the American Statistical Association **55**(290) (1960) 299–306 cited By (since 1996)184.
7. Roberts, S.: A general class of holt-winters type forecasting models. Management Science **28**(7) (1982) 808–820 cited By (since 1996)0.
8. Abraham, B., Ledolter, J. Statistical Methods for Forecasting (1983) cited By (since 1996)207.
9. Abraham, B., Ledolter, J.: Forecast functions implied by autoregressive integrated moving average models and other related forecast procedures. International Statistical Review **54** (1986) 51–66 cited By (since 1996)14.
10. Snyder, R.: Recursive estimation of dynamic linear models. Journal of the Royal Statistical Society, Series B **47**(2) (1985) 272–276 cited By (since 1996)40.
11. Taylor, J.: Exponential smoothing with a damped multiplicative trend. International Journal of Forecasting **19**(4) (2003) 715–725 cited By (since 1996)42.
12. Gardner Jr., E.: Forecasting the failure of component parts in computer systems: A case study. International Journal of Forecasting **9**(2) (1993) 245–253 cited By (since 1996)5.
13. Grubb, H., Mason, A.: Long lead-time forecasting of uk air passengers by holt-winters methods with damped trend. International Journal of Forecasting **17**(1) (2001) 71–82 cited By (since 1996)27.

14. Miller, T., Liberatore, M.: Seasonal exponential smoothing with damped trends. an application for production planning. International Journal of Forecasting **9**(4) (1993) 509–515 cited By (since 1996)10.
15. Hyndman, R., Koehler, A., Snyder, R., Grose, S.: A state space framework for automatic forecasting using exponential smoothing methods. International Journal of Forecasting **18**(3) (2002) 439–454 cited By (since 1996)108.
16. Yule, G.: On a method of investigating periodicities in disturbed series with special reference to wolfer's sunspot numbers. Philos. Trans. R. Soc. London **226** (1927) 267–298 cited By (since 1996)218.
17. Kolmogorov, A.: Stationary sequences in hilbert space. Bull. Math. Univ. Moscow **2**(6) (1941) 1–40 cited By (since 1996)72.
18. Newbold, P.: Arima model building and the time series analysis approach to forecasting. Journal of Forecasting **2** (1983) 23–35 cited By (since 1996)5.
19. Box, G., Jenkins, G. Time Series Analysis: Forecasting and Control (1976) cited By (since 1996)7314.
20. Di Caprio, U., Genesio, R., Pozzi, S., Vicino, A.: Short term load forecasting in electric power systems: A comparison of arma models and extended wiener filtering. J. Forecast. (**2**) 59–76 cited By (since 1996)2.
21. Cummins, J., Griepentrog, G.: Forecasting automobile insurance paid claim costs using econometric and arima models. International Journal of Forecasting **1**(3) (1985) 203–215 cited By (since 1996)1.
22. Hein, S., Spudeck, R.: Forecasting the daily federal funds rate. International Journal of Forecasting **4**(4) (1988) 581–591 cited By (since 1996)3.
23. Dhrymes, P., Peristiani, S.: A comparison of the forecasting performance of wefa and arima time series methods. International Journal of Forecasting **4**(1) (1988) 81–101 cited By (since 1996)1.
24. Geurts, M., Patrick Kelly, J.: Forecasting retail sales using alternative models. International Journal of Forecasting **2**(3) (1986) 261–272 cited By (since 1996)9.
25. Geurts, M., Kelly, J.: "in defense of arima modeling", by d.j. pack. International Journal of Forecasting **6**(4) (1990) 497–499 cited By (since 1996)1.
26. Pack, D.: Comments on: "in defense of arima modeling", by m.d. geurts and j.p. kelly. International Journal of Forecasting **6**(4) (1990) 501–502 cited By (since 1996)1.
27. Grambsch, P., Stahel, W.: Forecasting demand for special telephone services. a case study. International Journal of Forecasting **6**(1) (1990) 53–64 cited By (since 1996)13.
28. Pflaumer, P.: Forecasting us population totals with the box-jenkins approach. International Journal of Forecasting **8**(3) (1992) 329–338 cited By (since 1996)16.
29. du Preez, J., Witt, S.: Univariate versus multivariate time series forecasting: An application to international tourism demand. International Journal of Forecasting **19**(3) (2003) 435–451 cited By (since 1996)44.
30. Layton, A., Defris, L., Zehnwirth, B.: An international comparison of economic leading indicators of telecommunications traffic. International Journal of Forecasting **2**(4) (1986) 413–425 cited By (since 1996)6.
31. P. Leone, R.: Forecasting the effect of an environmental change on market performance. an intervention time-series approach. International Journal of Forecasting **3**(3-4) (1987) 463–478 cited By (since 1996)11.
32. Bianchi, L., Jarrett, J., Choudary Hanumara, R.: Improving forecasting for telemarketing centers by arima modeling with intervention. International Journal of Forecasting **14**(4) (1998) 497–504 cited By (since 1996)20.

33. Weller, B.: National indicator series as quantitative predictors of small region monthly employment levels. International Journal of Forecasting **5**(2) (1989) 241–247 cited By (since 1996)4.
34. Liu, L.M., Lin, M.W.: Forecasting residential consumption of natural gas using monthly and quarterly time series. International Journal of Forecasting **7**(1) (1991) 3–16 cited By (since 1996)19.
35. Harris, J., Liu, L.M.: Dynamic structural analysis and forecasting of residential electricity consumption. International Journal of Forecasting **9**(4) (1993) 437–455 cited By (since 1996)31.
36. Downs, G., Rocke, D.: Municipal budget forecasting with multivariate arma models. Journal of Forecasting **2** (1983) 377–387 cited By (since 1996)1.
37. Hillmer, S., Larcker, D., Schroeder, D.: Forecasting accounting data: A multiple time-series analysis. Journal of Forecasting **2** (1983) 389–404 cited By (since 1996)4.
38. Öller, L.E.: How far can changes in general business activity be forecasted? International Journal of Forecasting **1**(2) (1985) 135–141 cited By (since 1996)10.
39. Heuts, R., Bronckers, J.: Forecasting the dutch heavy truck market. a multivariate approach. International Journal of Forecasting **4**(1) (1988) 57–79 cited By (since 1996)4.
40. Lin, W.: Modeling and forecasting hospital patient movements: Univariate and multiple time series approaches. International Journal of Forecasting **5**(2) (1989) 195–208 cited By (since 1996)10.
41. Edlund, P.O., Karlsson, S.: Forecasting the swedish unemployment rate var vs. transfer function modelling. International Journal of Forecasting **9**(1) (1993) 61–76 cited By (since 1996)7.
42. Bhansali, R.: Asymptotically efficient autoregressive model selection for multistep prediction. Annals of the Institute of Statistical Mathematics **48**(3) (1996) 577–602 cited By (since 1996)37.
43. Bhansali, R.: Autoregressive model selection for multistep prediction. Journal of Statistical Planning and Inference **78**(1-2) (1999) 295–305 cited By (since 1996)5.
44. West, K.: Asymptotic inference about predictive ability. Econometrica **68** (1996) 1084–1097 cited By (since 1996)1.
45. Peña, D., Sánchez, I.: Multifold predictive validation in armax time series models. Journal of the American Statistical Association **100**(469) (2005) 135–146 cited By (since 1996)3.
46. Cholette, P.: Prior information and arima forecasting. Journal of Forecasting **1**(4) (1982) 375–383 cited By (since 1996)0.
47. Guerrero, V.: Arima forecasts with restrictions derived from a structural change. International Journal of Forecasting **7**(3) (1991) 339–347 cited By (since 1996)5.
48. de Alba, E.: Constrained forecasting in autoregressive time series models: A bayesian analysis. International Journal of Forecasting **9**(1) (1993) 95–108 cited By (since 1996)6.
49. Parzen, E.: Ararma models for time series analysis and forecasting. Journal of Forecasting **1** (1982) 67–82 cited By (since 1996)27.
50. Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., Winkler, R.: The accuracy of extrapolation (time series) methods: Results of a forecasting competition. Journal of Forecasting **1**(2) (1982) 111–153 cited By (since 1996)0.
51. Geriner, P., Ord, J.: Automatic forecasting using explanatory variables: A comparative study. International Journal of Forecasting **7**(2) (1991) 127–140 cited By (since 1996)7.

52. Tashman, L., Leach, M.: Automatic forecasting software: A survey and evaluation. International Journal of Forecasting **7**(2) (1991) 209–230 cited By (since 1996)0.
53. Tashman, L.: Out-of-sample tests of forecasting accuracy: An analysis and review. International Journal of Forecasting **16**(4) (2000) 437–450 cited By (since 1996)101.
54. Chatfield, C.: What is the 'best' method of forecasting? Journal of Applied Statistics **15**(1) (1988) 19–38 cited By (since 1996)12.
55. Kang, I.B.: Multi-period forecasting using different models for different horizons: An application to u.s. economic time series data. International Journal of Forecasting **19**(3) (2003) 387–400 cited By (since 1996)17.
56. Edlund, P.E.: Identification of the multi-input box-jenkins transfer functions model. Journal of Forecasting **3** (1984) 297–308 cited By (since 1996)5.
57. Koreisha, S.: Causal implications: The linkage between time series and econometric modelling. Journal of Forecasting **2** (1983) 151–168 cited By (since 1996)1.
58. Gupta, S.: Testing causality. some caveats and a suggestion. International Journal of Forecasting **3**(2) (1987) 195–209 cited By (since 1996)1.
59. Del Moral, M., Valderrama, M.: A principal component approach to dynamic regression models. International Journal of Forecasting **13**(2) (1997) 237–244 cited By (since 1996)4.
60. Krishnamurthi, L., Narayan, J., Raj, S.: Intervention analysis using control series and exogenous variables in a transfer function model: A case study. International Journal of Forecasting **5**(1) (1989) 21–27 cited By (since 1996)3.
61. Quenouille, M. The analysis of multiple time-series. 2nd ed. 1968 (1957) cited By (since 1996)1.
62. Riise, T., Tjøstheim, D.: Theory and practice of multivariate arma forecasting. Journal of Forecasting **3** (1984) 309–317 cited By (since 1996)2.
63. De Gooijer, J., Kumar, K.: Some recent developments in non-linear time series modelling, testing, and forecasting. International Journal of Forecasting **8**(2) (1992) 135–156 cited By (since 1996)68.
64. Lütkepohl, H.: Comparison of predictors for temporally and contemporaneously aggregated time series. International Journal of Forecasting **2** (1986) 461–475 cited By (since 1996)8.
65. Bidarkota, P.: The comparative forecast performance of univariate and multivariate models: An application to real interest rate forecasting. International Journal of Forecasting **14**(4) (1998) 457–468 cited By (since 1996)10.
66. Funke, M.: Assessing the forecasting accuracy of monthly vector autoregressive models. the case of five oecd countries. International Journal of Forecasting **6**(3) (1990) 363–378 cited By (since 1996)8.
67. Liu, T.R., Gerlow, M., Irwin, S.: The performance of alternative var models in forecasting exchange rates. International Journal of Forecasting **10**(3) (1994) 419–433 cited By (since 1996)15.
68. Simkins, S.: Forecasting with vector autoregressive (var) models subject to business cycle restrictions. International Journal of Forecasting **11**(4) (1995) 569–583 cited By (since 1996)6.
69. Litterman, R.: Forecasting with bayesian vector autoregressions - five years of experience. Journal of Business and Economic Statistics **4**(1) (1986) 25–38 cited By (since 1996)257.
70. Findley, D., Monsell, B., Bell, W., Otto, M., Chen, B.C.: New capabilities and methods of the x-12-arima seasonal-adjustment program. Journal of Business and Economic Statistics **16**(2) (1998) 127–152 cited By (since 1996)128.

71. Quenneville, B., Ladiray, D., Lefrançois, B.: A note on musgrave asymmetrical trend-cycle filters. International Journal of Forecasting **19**(4) (2003) 727–734 cited By (since 1996)8.
72. Gómez, V., Maravall, A.: Seasonal adjustment and signal extraction in economic time series. Chapter 8 in a course in time series analysis (2001) cited By (since 1996)1.
73. Kaiser, R., Maravall, A.: Combining filter design with model-based filtering (with an application to business-cycle estimation). International Journal of Forecasting **21**(4) (2005) 691–710 cited By (since 1996)11.
74. Cleveland, R., Cleveland, W., McRae, J., Terpenning, I.: Stl: A seasonal-trend decomposition procedure based on loess. Journal of Official Statistics **6**(1) (1990) 3–73 cited By (since 1996)316.
75. Simmons, L.: Time-series decomposition using the sinusoidal model. International Journal of Forecasting **6**(4) (1990) 485–495 cited By (since 1996)8.
76. Withycombe, R.: Forecasting with combined seasonal indices. International Journal of Forecasting **5**(4) (1989) 547–552 cited By (since 1996)11.
77. Bunn, D., Vassilopoulos, A.: Using group seasonal indices in multi-item short-term forecasting. International Journal of Forecasting **9**(4) (1993) 517–526 cited By (since 1996)11.
78. Bunn, D., Vassilopoulos, A.: Comparison of seasonal estimation methods in multi-item short-term forecasting. International Journal of Forecasting **15**(4) (1999) 431–443 cited By (since 1996)23.
79. Hylleberg, S., Pagan, A.: Seasonal integration and the evolving seasonals model. International Journal of Forecasting **13**(3) (1997) 329–340 cited By (since 1996)10.
80. Taylor, A.: On the practical problems of computing seasonal unit root tests. International Journal of Forecasting **13**(3) (1997) 307–318 cited By (since 1996)13.
81. Franses, P., Koehler, A.: A model selection strategy for time series with increasing seasonal variation. International Journal of Forecasting **14**(3) (1998) 405–414 cited By (since 1996)11.
82. Wells, J.: Modelling seasonal patterns and long-run trends in u.s. time series. International Journal of Forecasting **13**(3) (1997) 407–420 cited By (since 1996)6.
83. Herwartz, H.: Performance of periodic error correction models in forecasting consumption data. International Journal of Forecasting **13**(3) (1997) 421–431 cited By (since 1996)7.
84. Novales, A., De Fruto, R.: Forecasting with periodic models: A comparison with time invariant coefficient models. International Journal of Forecasting **13**(3) (1997) 393–405 cited By (since 1996)12.
85. Ula, T.: Forecasting of multivariate periodic autoregressive moving-average processes. J. Time Ser. Anal. **14**(6) (1993) 645–657 cited By (since 1996)14.
86. Chen, C.: Robustness properties of some forecasting methods for seasonal time series: A monte carlo study. International Journal of Forecasting **13**(2) (1997) 269–280 cited By (since 1996)7.
87. Noakes, D., McLeod, A., Hipel, K.: Forecasting monthly riverflow time series. International Journal of Forecasting **1**(2) (1985) 179–190 cited By (since 1996)32.
88. Albertson, K., Aylen, J.: Modelling the great lakes freeze: Forecasting and seasonality in the market for ferrous scrap. International Journal of Forecasting **12**(3) (1996) 345–359 cited By (since 1996)13.
89. Kulendran, N., King, M.: Forecasting international quarterly tourist flows using error-correction and time-series models. International Journal of Forecasting **13**(3) (1997) 319–327 cited By (since 1996)97.

90. Franses, P., van Dijk, D.: The forecasting performance of various models for seasonality and nonlinearity for quarterly industrial production. International Journal of Forecasting **21**(1) (2005) 87–102 cited By (since 1996)16.
91. Kalman, R.: A new approach to linear filtering and prediction problems. Journal of Basic Engineering (1960) 35–45 cited By (since 1996)7883.
92. Schweppe, F.: Evaluation of likelihood functions for gaussian signals. IEEE Transactions on Information Theory **11**(1) (1965) 61–70 cited By (since 1996)144.
93. Shumway, R., Stoffer, D.: An approach to time series smoothing and forecasting using the em algorithm. J. Time Series Anal. **3**(4) (1982) 253–264 cited By (since 1996)428.
94. Harrison, P., Stevens, C.: Bayesian forecasting. Journal of the Royal Statistical Society, Ser B **38** (1976) 205–247 cited By (since 1996)215.
95. Fildes, R.: An evaluation of bayesian forecasting. Journal of Forecasting **2**(2) (1983) 137–150 cited By (since 1996)0.
96. Harvey, A.: A unified view of statistical forecasting procedures. Journal of Forecasting **3**(3) (1984) 245–275 cited By (since 1996)0.
97. Harvey, A. Forecasting, Structural Time Series Models and the Kalman Filter (1989) cited By (since 1996)2173.
98. Harvey, A.: Chapter 7 forecasting with unobserved components time series models. Handbook of Economic Forecasting **1** (2006) 327–412 cited By (since 1996)23.
99. Mittnik, S.: Macroeconomic forecasting experience with balanced state space models. International Journal of Forecasting **6**(3) (1990) 337–348 cited By (since 1996)5.
100. Vinod, H., Basu, P.: Forecasting consumption, income and real interest rates from alternative state space models. International Journal of Forecasting **11**(2) (1995) 217–231 cited By (since 1996)1.
101. Coomes, P.: A kalman filter formulation for noisy regional job data. International Journal of Forecasting **7**(4) (1992) 473–481 cited By (since 1996)6.
102. Patterson, K.: Forecasting the final vintage of real personal disposable income: A state space approach. International Journal of Forecasting **11**(3) (1995) 395–405 cited By (since 1996)6.
103. West, M., Harrison, J. Bayesian Forecasting and Dynamic Models (1997) cited By (since 1996)1041.
104. Durbin, J., Koopman, S. Time Series Analysis by State Space Methods (2001) cited By (since 1996)688.
105. Tong, H.: Threshold models in non-linear time series analysis. Threshold Models in Nonlinear Time Series Analysis (1983) cited By (since 1996)425.
106. Tong, H. Non-linear Time Series: A Dynamical System Approach (1990) cited By (since 1996)1541.
107. Clements, M., Smith, J.: The performance of alternative forecasting methods for setar models. International Journal of Forecasting **13**(4) (1997) 463–475 cited By (since 1996)47.
108. Fok, D., van Dijk, D., Franses, P.: Forecasting aggregates using panels of nonlinear time series. International Journal of Forecasting **21**(4) (2005) 785–794 cited By (since 1996)9.
109. Franses, P., Paap, R., Vroomen, B.: Forecasting unemployment using an autoregression with censored latent effects parameters. International Journal of Forecasting **20**(2) (2004) 255–271 cited By (since 1996)6.
110. Hastie, T., Tibshirani, R. Generalized Additive Models (1990) cited By (since 1996)5324.

111. Granger, C., Joyeux, R.: An introduction to long-memory time series models and fractional differencing. Journal of Time Series Analysis **1**(1) (1980) 15–29 cited By (since 1996)1061.
112. Ray, B.: Long-range forecasting of ibm product revenues using a seasonal fractionally differenced arma model. International Journal of Forecasting **9**(2) (1993) 255–269 cited By (since 1996)49.
113. Ray, B.: Modeling long-memory processes for optimal long-range prediction. Journal of Time Series Analysis **14**(5) (1993) 511–525 cited By (since 1996)34.
114. Franses, P., Ooms, M.: A periodic long-memory model for quarterly uk inflation. International Journal of Forecasting **13**(1) (1997) 117–126 cited By (since 1996)21.
115. Ravishanker, N., Ray, B.: Bayesian prediction for vector arfima processes. International Journal of Forecasting **18**(2) (2002) 207–214 cited By (since 1996)8.
116. Baillie, R., Chung, S.K.: Modeling and forecasting from trend-stationary long memory models with applications to climatology. International Journal of Forecasting **18**(2) (2002) 215–226 cited By (since 1996)19.
117. Beran, J., Feng, Y., Ghosh, S., Sibbertsen, P.: On robust local polynomial estimation with long-memory errors. International Journal of Forecasting **18**(2) (2002) 227–241 cited By (since 1996)7.
118. Ramjee, R., Crato, N., Ray, B.: A note on moving average forecasts of long memory processes with an application to quality control. International Journal of Forecasting **18**(2) (2002) 291–297 cited By (since 1996)4.
119. Engle, R.: Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica **50** (1982) 987–1008 cited By (since 1996)4825.
120. Bollerslev, T., Engle, R., Nelson, D.: Arch models. Handbook of Econometrics **4** (1994) 2959–3038 cited By (since 1996)445.
121. Taylor, S.: Forecasting the volatility of currency exchange rates. International Journal of Forecasting **3**(1) (1987) 159–170 cited By (since 1996)26.
122. Bollerslev, T., Chou, R., Kroner, K.: Arch modeling in finance. a review of the theory and empirical evidence. Journal of Econometrics **52**(1-2) (1992) 5–59 cited By (since 1996)1516.
123. Bera, A., Higgins, M.: Arch models: Properties, estimation and testing. Journal of Economic Surveys **7**(4) (1993) 305–366 cited By (since 1996)288.
124. Diebold, F., Mariano, R.: Comparing predictive accuracy. Journal of Business and Economic Statistics **13**(3) (1995) 253–263 cited By (since 1996)1437.
125. Franses, P., Ghijsels, H.: Additive outliers, garch and forecasting volatility. International Journal of Forecasting **15**(1) (1999) 1–9 cited By (since 1996)64.
126. Brooks, C.: Predicting stock index volatility: Can market volume help? Journal of Forecasting **17**(1) (1998) 59–80 cited By (since 1996)65.
127. Brooks, C., Burke, S., Persand, G.: Benchmarks and the accuracy of garch model estimation. International Journal of Forecasting **17**(1) (2001) 45–56 cited By (since 1996)42.
128. Newbold, P., Agiakloglou, C., Miller, J.: Adventures with arima software. International Journal of Forecasting **10**(4) (1994) 573–581 cited By (since 1996)14.
129. Karanasos, M.: Prediction in arma models with garch in mean effects. Journal of Time Series Analysis **22**(5) (2001) 555–576 cited By (since 1996)8.
130. Baillie, R., Bollerslev, T., Mikkelsen, H.: Fractionally integrated generalized autoregressive conditional heteroskedasticity. Journal of Econometrics **74**(1) (1996) 3–30 cited By (since 1996)588.
131. Granger, C.: Long memory, volatility, risk and distribution. Manuscript (2002) cited By (since 1996)1.

132. Engle, R., Ng, V.: Measuring and testing the impact of news on volatility. Journal of Finance **48**(5) (1993) 1749–1778 cited By (since 1996)1075.
133. Hentschel, L.: All in the family nesting symmetric and asymmetric garch models. Journal of Financial Economics **39**(1) (1995) 71–104 cited By (since 1996)201.
134. Pagan, A.: The econometrics of financial markets. Journal of Empirical Finance **48**(1-3) (1996) 15–102 cited By (since 1996)242.
135. Poon, S.H., Granger, C.: Forecasting volatility in financial markets: A review. Journal of Economic Literature **41**(2) (2003) 478–539 cited By (since 1996)357.
136. Doidge, C., Wei, J.: Volatility forecasting and the efficiency of the toronto 35 index options market. Canadian Journal of Administrative Sciences **15**(1) (1998) 28–38 cited By (since 1996)8.
137. Kroner, K., Kneafsey, K., Claessens, S.: Forecasting volatility in commodity markets. Journal of Forecasting **14** (1995) 77–95 cited By (since 1996)20.
138. Vasilellis, G., Meade, N.: Forecasting volatility for portfolio selection. Journal of Business Finance and Accounting **23**(1) (1996) 125–143 cited By (since 1996)17.
139. Croston, J.: Forecasting and stock control for intermittent demands. Operational Research Quarterly **23**(3) (1972) 289–303 cited By (since 1996)159.
140. Willemain, T., Smart, C., Shockor, J., DeSautels, P.: Forecasting intermittent demand in manufacturing: a comparative evaluation of croston's method. International Journal of Forecasting **10**(4) (1994) 529–538 cited By (since 1996)82.
141. Johnston, F., Boylan, J.: Forecasting intermittent demand: A comparative evaluation of croston's method. comment. International Journal of Forecasting **12**(2) (1996) 297–298 cited By (since 1996)12.
142. Syntetos, A., Boylan, J.: The accuracy of intermittent demand estimates. International Journal of Forecasting **21**(2) (2005) 303–314 cited By (since 1996)79.
143. Willemain, T., Smart, C., Schwarz, H.: A new approach to forecasting intermittent demand for service parts inventories. International Journal of Forecasting **20**(3) (2004) 375–387 cited By (since 1996)89.
144. Diebold, F., Gunther, T., Tay, A.: Evaluating density forecasts with applications to financial risk management. International Economic Review **39**(4) (1998) 863–883 cited By (since 1996)373.
145. Brannas, K., Hellstrom, J., Nordstrom, J.: A new approach to modelling and forecasting monthly guest nights in hotels. International Journal of Forecasting **18**(1) (2002) 19–30 cited By (since 1996)28.
146. Mahmoud, E.: Accuracy in forecasting: A survey. Journal of Forecasting **3**(2) (1984) 139–159 cited By (since 1996)0.
147. Armstrong, J., Collopy, F.: Error measures for generalizing about forecasting methods: Empirical comparisons. International Journal of Forecasting **8**(1) (1992) 69–80 cited By (since 1996)267.
148. Makridakis, S., Wheelwright, S., McGee, V. Forecasting: Methods and Applications (1983) cited By (since 1996)1119.
149. Fildes, R.: The evaluation of extrapolative forecasting methods. International Journal of Forecasting **8**(1) (1992) 81–98 cited By (since 1996)102.
150. Chatfield, C.: A commentary on error measures. International Journal of Forecasting **8**(1) (1992) 100–102 cited By (since 1996)11.
151. Makridakis, S., Hibon, M.: The m3-competition: Results, conclusions and implications. International Journal of Forecasting **16**(4) (2000) 451–476 cited By (since 1996)324.
152. Makridakis, S.: Accuracy measures: theoretical and practical concerns. International Journal of Forecasting **9**(4) (1993) 527–529 cited By (since 1996)97.

153. Goodwin, P., Lawton, R.: On the asymmetry of the symmetric mape. International Journal of Forecasting **15**(4) (1999) 405–408 cited By (since 1996)42.
154. Koehler, A., Snyder, R., Ord, J.: Forecasting models and prediction intervals for the multiplicative holt-winters method. International Journal of Forecasting **17**(2) (2001) 269–286 cited By (since 1996)29.
155. Thompson, P.: An mse statistic for comparing forecast accuracy across series. International Journal of Forecasting **6**(2) (1990) 219–227 cited By (since 1996)11.
156. Thompson, P.: Evaluation of the m-competition forecasts via log mean squared error ratio. International Journal of Forecasting **7**(3) (1991) 331–334 cited By (since 1996)2.
157. Granger, C., Jeon, Y.: A time-distance criterion for evaluating forecasting models. International Journal of Forecasting **19**(2) (2003) 199–215 cited By (since 1996)8.
158. Granger, C., Jeon, Y.: Comparing forecasts of inflation using time distance. International Journal of Forecasting **19**(3) (2003) 339–349 cited By (since 1996)10.
159. Flores, B.: The utilization of the wilcoxon test to compare forecasting methods: A note. International Journal of Forecasting **5**(4) (1989) 529–535 cited By (since 1996)5.
160. Harvey, D., Leybourne, S., Newbold, P.: Testing the equality of prediction mean squared errors. International Journal of Forecasting **13**(2) (1997) 281–291 cited By (since 1996)315.
161. McCracken, M.: Parameter estimation and tests of equal forecast accuracy between non-nested models. International Journal of Forecasting **20**(3) (2004) 503–514 cited By (since 1996)8.
162. Sullivan, R., Timmermann, A., White, H.: Forecast evaluation with shared data sets. International Journal of Forecasting **19**(2) (2003) 217–227 cited By (since 1996)17.
163. Clements, M., Hendry, D.: On the limitations of comparing mean square forecast errors. Journal of Forecasting **12**(8) (1993) 617–637 cited By (since 1996)0.
164. JM, B., CWJ, G.: Combination of forecasts. Operational Research Quarterly **20**(4) (1969) 451–468 cited By (since 1996)700.
165. Newbold, P., Granger, C.: Experience with forecasting univariate time series and the combination of forecasts. Journal of the Royal Statistical Society, Series A **137**(2) (1974) 131–165 cited By (since 1996)157.
166. Winkler, R., Makridakis, S.: The combination of forecasts. Journal of the Royal Statistical Society, Series A **146**(2) (1983) 150–157 cited By (since 1996)66.
167. Clemen, R.: Combining forecasts: A review and annotated bibliography. International Journal of Forecasting **5**(4) (1989) 559–583 cited By (since 1996)605.
168. Bunn, D.: Statistical efficiency in the linear combination of forecasts. International Journal of Forecasting **1**(2) (1985) 151–163 cited By (since 1996)26.
169. Granger, C., Ramanathan, R.: Improved methods of combining forecasts. Journal of Forecasting **3**(2) (1984) 197–204 cited By (since 1996)0.
170. Aksu, C., Gunter, S.: An empirical analysis of the accuracy of sa, ols, erls and nrls combination forecasts. International Journal of Forecasting **8**(1) (1992) 27–43 cited By (since 1996)20.
171. Gunter, S.: Nonnegativity restricted least squares combinations. International Journal of Forecasting **8**(1) (1992) 45–59 cited By (since 1996)9.
172. Granger, C.: Combining forecasts-twenty years later. Journal of Forecasting **8**(3) (1989) 167–173 cited By (since 1996)0.
173. Deutsch, M., Granger, C., Teräsvirta, T.: The combination of forecasts using changing weights. International Journal of Forecasting **10**(1) (1994) 47–57 cited By (since 1996)52.

174. Fiordaliso, A.: A nonlinear forecasts combination method based on takagi-sugeno fuzzy systems. International Journal of Forecasting **14**(3) (1998) 367–379 cited By (since 1996)31.
175. Diebold, F., Pauly, P.: The use of prior information in forecast combination. International Journal of Forecasting **6**(4) (1990) 503–508 cited By (since 1996)37.
176. Zou, H., Yang, Y.: Combining time series models for forecasting. International Journal of Forecasting **20**(1) (2004) 69–84 cited By (since 1996)56.
177. Miller, C., Clemen, R., Winkler, R.: The effect of nonstationarity on combined forecasts. International Journal of Forecasting **7**(4) (1992) 515–529 cited By (since 1996)7.
178. Hendry, D., Clements, M.: Pooling of forecasts. Econometrics Journal **5** (2002) 1–26 cited By (since 1996)31.
179. Terui, N., Van Dijk, H.: Combined forecasts from linear and nonlinear time series models. International Journal of Forecasting **18**(3) (2002) 421–438 cited By (since 1996)66.
180. De Menezes, L., Bunn, D.: The persistence of specification problems in the distribution of combined forecast errors. International Journal of Forecasting **14**(3) (1998) 415–426 cited By (since 1996)4.
181. Taylor, J., Bunn, D.: Investigating improvements in the accuracy of prediction intervals for combinations of forecasts: A simulation study. International Journal of Forecasting **15**(3) (1999) 325–339 cited By (since 1996)13.
182. Fang, Y.: Forecasting combination and encompassing tests. International Journal of Forecasting **19**(1) (2003) 87–94 cited By (since 1996)32.
183. Hibon, M., Evgeniou, T.: To combine or not to combine: Selecting among forecasts and their combinations. International Journal of Forecasting **21**(1) (2005) 15–24 cited By (since 1996)87.
184. Chatfield, C.: Calculating interval forecasts. Journal of Business and Economic Statistics **11**(2) (1993) 121–135 cited By (since 1996)115.
185. Makridakis, S., Hibon, M., Lusk, E., Belhadjali, M.: Confidence intervals. an empirical investigation of the series in the m-competition. International Journal of Forecasting **3**(3-4) (1987) 489–508 cited By (since 1996)8.
186. Lefrançois, P.: Confidence intervals for non-stationary forecast errors. some empirical results for the series in the m-competition. International Journal of Forecasting **5**(4) (1989) 553–557 cited By (since 1996)1.
187. Koehler, A.: An inappropriate prediction interval. International Journal of Forecasting **6**(4) (1990) 557–558 cited By (since 1996)5.
188. Chatfield, C., Koehler, A.: On confusing lead time demand with h-period-ahead forecasts. International Journal of Forecasting **7**(2) (1991) 239–240 cited By (since 1996)3.
189. Masarotto, G.: Bootstrap prediction intervals for autoregressions. International Journal of Forecasting **6**(2) (1990) 229–239 cited By (since 1996)30.
190. McCullough, B.: Bootstrapping forecast intervals: An application to ar(p) models. Journal of Forecasting **13** (1994) 51–66 cited By (since 1996)23.
191. Mccullough, B.: Consistent forecast intervals when the forecast-period exogenous variables are stochastic. Journal of Forecasting **15**(4) (1996) 293–304 cited By (since 1996)6.
192. Grigoletto, M.: Bootstrap prediction intervals for autoregressions: Some alternatives. International Journal of Forecasting **14**(4) (1998) 447–456 cited By (since 1996)23.

193. Clements, M., Taylor, N.: Bootstrapping prediction intervals for autoregressive models. International Journal of Forecasting **17**(2) (2001) 247–267 cited By (since 1996)25.
194. Kim, J.: Bootstrap prediction intervals for autoregression using asymptotically mean-unbiased estimators. International Journal of Forecasting **20**(1) (2004) 85–97 cited By (since 1996)12.
195. Pascual, L., Romo, J., Ruiz, E.: Effects of parameter estimation on prediction densities: A bootstrap approach. International Journal of Forecasting **17**(1) (2001) 83–103 cited By (since 1996)25.
196. Pascual, L., Romo, J., Ruiz, E.: Bootstrap predictive inference for arima processes. Journal of Time Series Analysis **25**(4) (2004) 449–465 cited By (since 1996)20.
197. Pascual, L., Romo, J., Ruiz, E.: Bootstrap prediction intervals for power-transformed time series. International Journal of Forecasting **21**(2) (2005) 219–235 cited By (since 1996)4.
198. Wall, K., Stoffer, D.: A state space approach to bootstrapping conditional forecasts in arma models. Journal of Time Series Analysis **23**(6) (2002) 733–751 cited By (since 1996)14.
199. Kim, J.: Asymptotic and bootstrap prediction regions for vector autoregression. International Journal of Forecasting **15**(4) (1999) 393–403 cited By (since 1996)19.
200. Reeves, J.: Bootstrap prediction intervals for arch models. International Journal of Forecasting **21**(2) (2005) 237–248 cited By (since 1996)19.
201. Lam, J.P., Veall, M.: Bootstrap prediction intervals for single period regression forecasts. International Journal of Forecasting **18**(1) (2002) 125–130 cited By (since 1996)14.
202. Tay, A., Wallis, K.: Density forecasting: A survey. Journal of Forecasting **19**(4) (2000) 235–254 cited By (since 1996)87.
203. Wallis, K.: Asymmetric density forecasts of inflation and the bank of england's fan chart. National Institute Economic Review **167**(1) (1999) 106–112 cited By (since 1996)36.
204. Hyndman, R.: Highest-density forecast regions for non-linear and non-normal time series models. Journal of Forecasting **14** (1995) 431–441 cited By (since 1996)33.
205. Diebold, F., Hahn, J., Tay, A.: Multivariate density forecast evaluation and calibration in financial risk management: High-frequency returns on foreign exchange. Review of Economics and Statistics **81**(4) (1999) 661–673 cited By (since 1996)84.
206. Wallis, K.: Chi-squared tests of interval and density forecasts, and the bank of england's fan charts. International Journal of Forecasting **19**(2) (2003) 165–175 cited By (since 1996)49.
207. Ord, J.: Future developments in forecasting. the time series connexion. International Journal of Forecasting **4**(3) (1988) 389–401 cited By (since 1996)4.
208. Forni, M., Hallin, M., Lippi, M., Reichlin, L.: The generalized dynamic factor model: One-sided estimation and forecasting. Journal of the American Statistical Association **100**(471) (2005) 830–840 cited By (since 1996)146.
209. Stock, J., Watson, M.: Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association **97**(460) (2002) 1167–1179 cited By (since 1996)319.
210. Peña, D., Poncela, P.: Forecasting with nonstationary dynamic factor models. Journal of Econometrics **119**(2) (2004) 291–321 cited By (since 1996)21.
211. Stock, J., Watson, M.: A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. Cointegration, Causality and Forecas-

ting: A Festschrift in Honour of Clive W. J. Granger (1999) 1–44 cited By (since 1996)136.
212. Cogger, K.: Proposals for research in time series forecasting. International Journal of Forecasting **4**(3) (1988) 403–410 cited By (since 1996)1.
213. Koenker, R., Bassett, G.: Regression quantiles. Econometrica **46**(1) (1978) 33–50 cited By (since 1996)2101.
214. Engle, R., Manganelli, S.: Caviar: Conditional autoregressive value at risk by regression quantiles. Journal of Business and Economic Statistics **22**(4) (2004) 367–381 cited By (since 1996)249.
215. Yao, Q., Tong, H.: On initial-condition sensitivity and prediction in nonlinear stochastic systems. Bulletin Int. Statist. Inst. **103** (1995) 395–412 cited By (since 1996)7.
216. De Gooijer, J., Vidiella-i Anguera, A.: Forecasting threshold cointegrated systems. International Journal of Forecasting **20**(2) (2004) 237–253 cited By (since 1996)13.
217. Polonik, W., Yao, Q.: Conditional minimum volume predictive regions for stochastic processes. Journal of the American Statistical Association **95**(450) (2000) 509–519 cited By (since 1996)21.
218. Armstrong, J. Suggestions for further research (2001) cited By (since 1996)2.
219. Engle, R., Russell, J.: Autoregressive conditional duration: A new model for irregularly spaced transaction data. Econometrica **66**(5) (1998) 1127–1162 cited By (since 1996)430.
220. Andersen, T., Bollerslev, T., Diebold, F., Labys, P.: Modeling and forecasting realized volatility. Econometrica **71**(2) (2003) 579–625 cited By (since 1996)621.
221. de Gooijer, J.: The role of time series analysis in forecasting: A personal view. International Journal of Forecasting **6**(4) (1990) 449–451 cited By (since 1996)1.
222. Clements, M.: Some possible directions for future research. International Journal of Forecasting **19**(1) (2003) 1–3 cited By (since 1996)6.
223. Dawes, R., Fildes, R., Lawrence, M., Ord, K.: The past and the future of forecasting research. International Journal of Forecasting **10**(1) (1994) 151–159 cited By (since 1996)9.
224. Teräsvirta, T.: Forecasting economic variables with nonlinear models. Handbook of Economic Forecasting **1** (2006) cited By (since 1996)5.
225. Stock, J., Watson, M.: Combination forecasts of output growth in a seven-country data set. Journal of Forecasting **23**(6) (2004) 405–430 cited By (since 1996)145.
226. Ramsay, J., Silverman, B.: Functional data analysis. Functional Data Analysis (1997) cited By (since 1996)1809.
227. Ahmed, N., Atiya, A., El Gayar, N., El-Shishiny, H.: An empirical comparison of machine learning models for time series forecasting. Econometric Reviews **29**(5) (2010) 594–621 cited By (since 1996)15.
228. Palit, A., Popovic, D.: Computational intelligence in time series forecasting: Theory and engineering applications. Advances in Industrial Control (2005) cited By (since 1996)4.
229. Zhang, G., Eddy Patuwo, B., Y. Hu, M.: Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting **14**(1) (1998) 35–62 cited By (since 1996)946.
230. Mitchell, T. Machine Learning (1997) cited By (since 1996)7537.
231. Werbos, P.: Beyond regression: New tools for prediction and analysis in the behavioral sciences. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences (1974) cited By (since 1996)1191.
232. Werbos, P.: Generalization of backpropagation with application to a recurrent gas market model. Neural Networks **1**(4) (1988) 339–356 cited By (since 1996)146.

233. Lapedes, A., Farber, R.: Nonlinear signal processing using neural networks: Prediction and system modeling. Nonlinear Signal Processing Using Neural Networks: Prediction and System Modeling (1987) cited By (since 1996)273.
234. Hastie, T., Tibshirani, R., Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2001) cited By (since 1996)10663.
235. Alpaydin, E. Introduction to Machine Learning (2004) cited By (since 1996)1082.
236. Crone, S.: Mining the past to determine the future: Comments. International Journal of Forecasting **25**(3) (2009) 456–460 cited By (since 1996)4.
237. Crone, S. (NN5 Forecasting Competition) cited By (since 1996)3.
238. Lendasse, A. ESTSP 2007: Proceedings (2007) (0000) cited By (since 1996)1.
239. Hand, D.: Mining the past to determine the future: Problems and possibilities. International Journal of Forecasting (2008) cited By (since 1996)4.
240. Price, S.: Mining the past to determine the future: Comments. International Journal of Forecasting **25**(3) (2009) 452–455 cited By (since 1996)5.
241. EN, L.: Atmospheric predictability as revealed by naturally occurring analogues. Journal of the Atmospheric Sciences **26**(4) (1969) 636–646 cited By (since 1996)243.
242. Ikeguchi, T., Aihara, K.: Prediction of chaotic time series with noise. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences E78-A (10) (1995) cited By (since 1996)1.
243. Tong, H., Lim, K.: Threshold autoregression, limit cycles and cyclical data. Journal of the Royal Statistical Society B **42**(3) (1980) 245–292 cited By (since 1996)428.
244. Priestley, M. Non-Linear and Non-Stationary Time Series Analysis (1988) cited By (since 1996)596.
245. Farmer, J., Sidorowich, J.: Predicting chaotic time series. Physical Review Letters **59**(8) (1987) 845–848 cited By (since 1996)858.
246. Farmer, J., Sidorowich, J. Exploiting Chaos to Predict the Future and Reduce Noise (1988) cited By (since 1996)14.
247. Bontempi, G.: "local learning techniques for modeling, prediction and control". Local Learning Techniques for Modeling, Prediction and Control (1999) cited By (since 1996)21.
248. Birattari, M., Bontempi, G., Bersini, H.: Lazy learning meets the recursive least squares algorithm. (1999) 375–381 cited By (since 1996)0.
249. Bontempi, G., Birattari, M., Bersini, H.: Local learning for iterated time-series prediction. Machine Learning: Proceedings of the Sixteenth International Conference (1999) 32–38 cited By (since 1996)23.
250. Guo, M., Bai, Z., An, H.: Multi-step prediction for nonlinear autoregressive models based on empirical distributions. Statistica Sinica **9**(2) (1999) 559–570 cited By (since 1996)5.
251. Kline, D.: Methods for multi-step time series forecasting with neural networks. Neural Networks in Business Forecasting (2004) 226–250 cited By (since 1996)12.
252. Tran, V., Yang, B.S., Tan, A.: Multi-step ahead direct prediction for the machine condition prognosis using regression trees and neuro-fuzzy systems. Expert Systems with Applications **36**(5) (2009) 9378–9387 cited By (since 1996)39.
253. Sorjamaa, A., Lendasse, A.: Time series prediction using dirrec strategy. European Symposium on Artificial Neural Networks, ESANN (2006) 143–148 cited By (since 1996)6.
254. Bontempi, G.: Long term time series prediction with multi-input multi-output local learning. Proceedings of the 2nd European Symposium on Time Series Prediction (TSP), ESTSP08 (2008) 145–154 cited By (since 1996)13.

255. Bontempi, G., Ben Taieb, S.: Conditionally dependent strategies for multiple-step-ahead prediction in local learning. International Journal of Forecasting **27**(3) (2011) 689–699 cited By (since 1996)8.
256. Ben Taieb, S., Bontempi, G., Sorjamaa, A., Lendasse, A.: Long-term prediction of time series by combining direct and mimo strategies. Proceedings of the 2009 IEEE International Joint Conference on Neural Networks (2009) 3054–3061 cited By (since 1996)6.
257. Ben Taieb, S., Sorjamaa, A., Bontempi, G.: Multiple-output modeling for multi-step-ahead time series forecasting. Neurocomputing **73**(10-12) (2010) 1950–1957 cited By (since 1996)19.
258. McNames, J.: A nearest trajectory strategy for time series prediction. Proceedings of the International Workshop on Advanced Black-Box Techniques for Nonlinear Modeling (1998) 112–128 cited By (since 1996)38.
259. Ben Taieb, S., Bontempi, G.: Recursive multi-step time series forecasting by perturbing data. Proceedings of IEEE-ICDM 2011 (2011) cited By (since 1996)1.
260. Sorjamaa, A., Lendasse, A., Verleysen, M.: Pruned lazy learning models for time series prediction. European Symposium on Artificial Neural Networks, ESANN 2005 (2005) 509–514 cited By (since 1996)1.
261. Ben Taieb, S., Bontempi, G., Sorjamaa, A., Lendasse, A.: Long-term prediction of time series by combining direct and mimo strategies. International Joint Conference on Neural Networks (2009) cited By (since 1996)2.
262. Darbellay, G., Slama, M.: Forecasting the short-term demand for electricity: Do neural networks stand a better chance? International Journal of Forecasting **16**(1) (2000) 71–83 cited By (since 1996)129.
263. Hippert, H., Pedreira, C., Souza, R.: Neural networks for short-term load forecasting: A review and evaluation. IEEE Transactions on Power Systems **16**(1) (2001) 44–55 cited By (since 1996)647.
264. Gorr, W.: Editorial: Research prospective on neural network forecasting. International Journal of Forecasting **10**(1) (1994) 1–4 cited By (since 1996)41.
265. Hill, T., Marquez, L., O'Connor, M., Remus, W.: Artificial neural network models for forecasting and decision making. International Journal of Forecasting **10**(1) (1994) 5–15 cited By (since 1996)105.
266. Swanson, N., White, H.: Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. International Journal of Forecasting **13**(4) (1997) 439–461 cited By (since 1996)93.
267. Heravi, S., Osborn, D., Birchenhall, C.: Linear versus neural network forecasts for european industrial production series. International Journal of Forecasting **20**(3) (2004) 435–446 cited By (since 1996)42.
268. Ghiassi, M., Saidane, H., Zimbra, D.: A dynamic artificial neural network model for forecasting time series events. International Journal of Forecasting **21**(2) (2005) 341–362 cited By (since 1996)75.
269. Hippert, H., Bunn, D., Souza, R.: Large neural networks for electricity load forecasting: Are they overfitted? International Journal of Forecasting **21**(3) (2005) 425–434 cited By (since 1996)34.
270. Teräsvirta, T., van Dijk, D., Medeiros, M.: Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination. International Journal of Forecasting **21**(4) (2005) 755–774 cited By (since 1996)62.
271. Takens, F.: Detecting strange attractors in turbulence. Lecture Notes in Mathematics **898** (1981) 366–381 cited By (since 1996)4276.

272. Grassberger, P., Procaccia, I.: Measuring the strangeness of strange attractors. In: The Theory of Chaotic Attractors. Springer (2004) 170–189
273. Debnath, A., Chin, H., Haque, M., Yuen, B.: A methodological framework for benchmarking smart transport cities. Cities **37** (2014) 47–56 cited By (since 1996)0.
274. Treiber, M., Kesting, A., Helbing, D.: Three-phase traffic theory and two-phase models with a fundamental diagram in the light of empirical stylized facts. Transportation Research Part B: Methodological **44**(8-9) (2010) 983–1000 cited By (since 1996)41.
275. Hoogendoorn, S., Bovy, P.: State-of-the-art of vehicular traffic flow modelling. Proceedings of the Institution of Mechanical Engineers. Part I: Journal of Systems and Control Engineering **215**(4) (2001) 283–303 cited By (since 1996)152.
276. Vlahogianni, E., Golias, J., Karlaftis, M.: Short-term traffic forecasting: Overview of objectives and methods. Transport Reviews **24**(5) (2004) 533–557 cited By (since 1996)105.
277. Ishak, S., Alecsandru, C.: Optimizing traffic prediction performance of neural networks under various topological, input, and traffic condition settings. Journal of Transportation Engineering **130**(4) (2004) 452–465 cited By (since 1996)26.
278. Innamaa, S.: Short-term prediction of travel time using neural networks on an interurban highway. Transportation **32**(6) (2005) 649–669 cited By (since 1996)27.
279. Zhang, H.: Recursive prediction of traffic conditions with neural network models. Journal of Transportation Engineering **126**(6) (2000) 472–481 cited By (since 1996)37.
280. Dharia, A., Adeli, H.: Neural network model for rapid forecasting of freeway link travel time. Engineering Applications of Artificial Intelligence **16**(7-8) (2003) 607–613 cited By (since 1996)89.
281. Abdulhai, B., Porwal, H., Recker, W.: Short-term traffic flow prediction using neuro-genetic algorithms. ITS Journal-Intelligent Transportation Systems Journal **7**(1) (2002) 3–41