



Munich Personal RePEc Archive

A model SARIMA to predict chilean unemployment

Idrovo Aguirre, Byron and Contreras, Javier

Cámara Chilena de la Construcción

22 June 2009

Online at <https://mpra.ub.uni-muenchen.de/19369/>
MPRA Paper No. 19369, posted 16 Dec 2009 14:21 UTC

Un Modelo SARIMA para Predecir la Tasa de Desempleo de Chile

Javier Contreras*

Byron Idrovo**

6 de septiembre de 2009

Resumen

Anticipar la evolución de la tasa de desempleo es de suma relevancia para la conducción de la política económica, ya que permite al Gobierno (*Policy Makers*) adoptar oportunamente medidas conducentes a minimizar las fluctuaciones del producto respecto de su nivel potencial de crecimiento. Con tal propósito, el presente documento utiliza un modelo de series de tiempo SARIMA (Seasonal Autorregresive Integrated Movable Average) que además de ajustar los componentes cíclico y estacional propios de una serie económica como es la tasa de desempleo, se caracteriza por ser un modelo fácil de estimar e interpretar –esto porque no requiere de otras variables ni de estructuras complejas de la econometría tradicional. Alternativamente, se ajustó un modelo ARFIMA (Autorregresive Fractionary Integrated Movable Average), debido a los signos de persistencia que muestra el indicador de desempleo en su comportamiento; sin embargo, a partir de los métodos de estimación de Reisen (1994), Geweke *et al* (1983), y Whittle (1962) se obtuvieron parámetros de integración mayor que 0.5, lo que empíricamente sustenta el tratamiento de la tasa de desempleo como una serie no estacionaria. Por último, basándonos en Albagli *et al* (2003) se compara la raíz del error cuadrático medio (RECM) de predicciones móviles fuera de muestra de tres modelos anidados: AR(1), ARIMA(1,1,2) y el modelo propuesto SARIMA(1,1,2) \times (0,1,1)₁₂, todos estimados por Máxima Verosimilitud (ML). Los modelos presentan ciertas limitaciones para capturar las discontinuidades (o saltos bruscos) de la serie histórica del indicador de desempleo. Sin embargo, el modelo SARIMA además de presentar la menor media y volatilidad del RECM de las predicciones móviles fuera de muestra, captura los puntos de giro del ciclo de la variable de interés.

KEY WORDS: Tasa de desempleo, Modelo SARIMA, Estacionalidad. **JEL Clasification:** C1, J6.

*jcontreras@dim.uchile.cl, Centro de Modelamiento Matemático (CMM), Universidad de Chile, Santiago, Chile.

**bidrovo@cchc.cl, Gerencia de Estudios, Cámara Chilena de la Construcción (CChC), Santiago, Chile.

1. Introducción

La tasa de desempleo mensual es un indicador de alerta temprana de la actividad económica respecto de las Cuentas Nacionales publicadas trimestralmente por el Banco Central. A esto se agrega que la dinámica del empleo es considerado como uno de los factores más incidentes en el desempeño económico del país (Bergoing *et al*, 2005). En este contexto, es de suma relevancia para la conducción de la política económica anticipar la evolución de la tasa de desempleo entre otras variables, ya que permitirá al Gobierno (*Policy Makers*) adoptar medidas conducentes a minimizar las fluctuaciones del producto respecto de su nivel potencial de crecimiento. Con tal propósito, el presente documento utiliza modelos de series de tiempo del tipo SARIMA (Seasonal Autorregresive Integrated Movile Average) para predecir el comportamiento de corto y mediano plazo de la tasa de desempleo de Chile. Para ello se utilizó información disponible desde febrero de 1986 hasta abril de 2009 (279 observaciones). Alternativamente, se ajustó un modelo ARFIMA (Autorregresive Fractionary Integrated Movile Average), debido a que el desempleo muestra cierta persistencia en su comportamiento —coherente con el lento ajuste del salario real frente a los distintos estados del ciclo económico. Sin embargo, a partir de los métodos de estimación de Reisen (1994), Geweke *et al* (1983), y Whittle (2002) se obtuvieron parámetros de integración mayor que 0.5, lo que empíricamente sustenta el tratamiento de la tasa de desempleo como una serie no estacionaria.

El proceso de selección del mejor modelo predictivo está basado en los criterios de información de Akaike y BIC (Criterio Bayesiano de Información), con estimaciones *in-sample* y *out-sample* para un horizonte de predicción de uno y ocho meses.¹ Una vez encontrado el mejor modelo SARIMA, se comparan gráficamente la tasa de desempleo experimentada durante la crisis asiática (1998-1999) versus la crisis actual -que incluye el período observado y proyectado de los años 2008 y 2009, respectivamente. Las similitudes encontradas entre las trayectorias de los niveles de desempleo de ambos períodos, pone de manifiesto la significativa inercia del mercado laboral, al punto que los shocks materializados durante la crisis asiática puedan ser replicados en la actualidad.

Los datos de desempleo son publicados mensualmente a partir de febrero de 1986 por el Instituto Nacional de Estadísticas (INE)²; y su elaboración consiste en dividir el número total de personas desocupadas en el trimestre móvil de cada mes t , para la suma total de personas ocupadas y desocupadas (fuerza laboral) en el trimestre móvil del mismo mes t . En otras palabras, la tasa de desempleo mide la proporción de la fuerza laboral que carece de trabajo en el mercado formal. Por ejemplo, el número de desocupados al cierre estadístico del presente estudio (trimestre móvil febrero-abril de 2009) es 716.33 miles

¹Los ocho meses corresponden al período faltante para cerrar el año 2009, considerando que el cierre estadístico del presente trabajo fue marzo de 2009.

²Fuente oficial de Estadísticas de Chile.

de personas, la fuerza laboral en igual período es 7,316.25 miles de personas, por lo tanto, la tasa de desempleo publicada por el INE en abril del presente año es 9.8%, cifra que resulta de dividir el número de desocupados para la fuerza laboral.

Por último, siguiendo de cerca a Albagli *et al* (2003), se compara la raíz del error cuadrático medio (RECM) de predicciones móviles fuera de muestra de tres modelos anidados: AR(1), ARIMA(1, 1, 2) y el modelo propuesto ARIMA(1, 1, 2) \times (0, 1, 1)₁₂, todos estimados por Máxima Verosimilitud (ML). Todos los modelos presentan ciertas limitaciones para capturar las discontinuidades (o bruscos saltos) en la trayectoria histórica del indicador de desempleo. Sin embargo, el modelo SARIMA además de presentar la menor media y volatilidad del RECM de las predicciones móviles fuera de muestra, captura los puntos de giro del ciclo de nuestra variable de interés.

El presente documento está organizado en cuatro secciones de la siguiente forma: La primera presenta la introducción; la segunda sección discute la metodología; la tercera analiza los principales resultados obtenidos; y la cuarta sección concluye.

2. Metodología

Desde febrero de 1986 –primer registro de desempleo publicado por el INE– la tasa de desempleo nacional ha mostrado un comportamiento cíclico y estacional muy característico. Es decir, en épocas de auge (parte más alta del ciclo económico) el desempleo disminuye considerablemente hasta ubicarse bajo su promedio histórico. Por el contrario, frente a escenarios de crisis dicha variable tiende a expandirse significativamente respecto de su medida de tendencia central. Con relación al componente estacional, la tasa de desempleo alcanza su mayor nivel en los meses de invierno -siendo julio y agosto los meses más predominantes de cada año. Esto se debe en parte, a la menor actividad del sector agrícola (período estacional de baja cosecha) y a la postergación de las obras de infraestructura y edificación -altamente intensivas en la utilización de mano de obra- ante condiciones climáticas poco propicias para el normal desarrollo de tales actividades.

En este contexto, es posible modelar la tasa de desempleo como un proceso del tipo ARIMA $(p, d, q) \times (P, D, Q)_s$. Este modelo captura la particularidad estacional y cíclica de nuestra serie de interés (la parte cíclica es la que se asocia al componente autorregresivo y de media móvil).

$$\Phi(B)\Phi_s(B)\nabla_{12}^D\nabla_1^dX_t = \Theta(B)\Theta_s(B)\epsilon_t$$

$$(1 - \sum_{k=1}^p \phi_k B^k)(1 - \sum_{k=1}^P \phi_k^s B^k)(1 - B^s)^D(1 - B)^d X_t = (1 - \sum_{k=1}^q \theta_k B^k)(1 - \sum_{k=1}^q \theta_k^s B^k) \epsilon_t$$

donde, $\Phi(B)$ es un polinomio de coeficientes, con operador de rezagos B del componente autorregresivo AR; y $\Phi_s(B)$ es análogo al caso anterior pero aplicado al proceso autorregresivo de la parte estacional SAR. Finalmente, $\Theta(B)$ y $\Theta_s(B)$ son los polinomios de coeficientes del proceso de media móvil MA y su símil estacional SMA. Por otro lado, se sospecha que la tasa de desempleo es no estacionaria, debido principalmente a los significativos cambios de media y varianza durante las fases de boom y recesión económica. De igual forma, se presume que el tramo ARIMA del componente periódico (o estacional) también es no estacionario, ya que existen factores (propios del efecto calendario) que indirectamente podrían acentuar la estacionalidad de la serie como la composición de los días de la semana. En resumen, el procedimiento utilizado para modelar la serie de la tasa de desempleo es:

1. Se estimó la función de autocorrelación (ACF) de la serie y se observó que ésta converge hiperbólicamente a cero. Al descartar la posibilidad de que sea estacionaria con integración fraccionaria –análisis pormenorizado en la sección 3.1–, se optó por transformar la serie tomando una diferenciación.
2. Se obtuvo nuevamente la ACF y se encontró un patrón estacional con un comportamiento sinusoidal de larga memoria, lo cual es típico en variables que muestran alta parsistencia en su trayectoria. Por lo tanto, se optó por diferenciar en doce meses la serie obtenida en el paso inmediatamente anterior. Sin embargo, como futura investigación proponemos testear la posibilidad de modelar un proceso SARFIMA, pero con integración fraccionaria en el parte estacional.
3. Finalmente, basado en los criterio de Akaike y BIC se seleccionó el mejor modelo SARIMA estimado mediante Máxima Verosimilitud (ML).³ Finalmente, los residuos del modelo resultaron ruido blanco –según el test de Box-Ljung.

3. Resultados

En el período 1986-1997 la tasa de desempleo mantuvo una trayectoria con una fuerte tendencia a la baja –con un promedio histórico de 8.1%. Este fenómeno es coherente con el nivel potencial de crecimiento económico observado durante dicho período –en torno a 7.8% (Chumacero, 1996). Dada la característica de la serie, cualquier predicción de desempleo, en base a estos datos, difícilmente habría

³En el apéndice de este documento se presenta una tabla que contiene los distintos modelos que se evaluaron con sus respectivos estadígrafos de ajuste.

dado cuenta del fuerte cambio de tendencia como la observada durante la crisis asiática (1998-1999). Para entonces, el desempleo aumentó en forma persistente hasta anotar su máximo nivel de 11.9% en agosto de 1999, para luego mantenerse oscilante en torno a 9.2% en los siguientes ocho años. En igual periodo (post-crisis asiática), la actividad económica se caracterizó por su bajo desempeño –con un crecimiento potencial de sólo 5% (Johnson, 2001)– relativo a la pronta recuperación experimentada por los países del sudeste asiático. Estos sucesos dieron origen al debate sobre la aparente persistencia de la actividad económica y por ende del empleo. Así, algunos autores sostienen que la dinámica del empleo es uno de los factores altamente incidentes en el prolongado deterioro de la economía (Bergoeing *et al*, 2005). Particularmente, la incertidumbre que se derivó de los cambios al código del trabajo en 2001, explicarían la prolongada caída en el empleo post-crisis asiática. En cambio, otros autores (Cowan *et al* 2003) encuentran que las fluctuaciones del empleo son absolutamente cíclicas –sin cambios estructurales; es decir, la rigidez del mercado laboral se debería a la baja respuesta de los salarios reales.

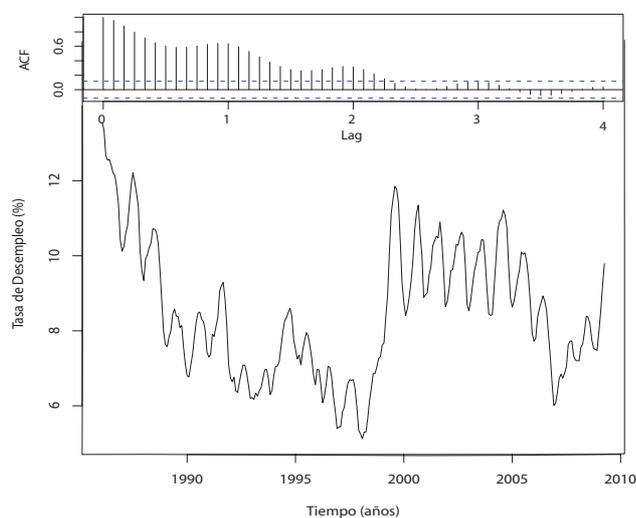


Figura 1: Tasa de desempleo entre los años 1986 y 2009 junto a su correlograma.

A partir del segundo semestre de 2008, tras los efectos de la crisis financiera internacional, la tasa de desempleo nacional ha tendido nuevamente al alza; incluso, hasta alcanzar niveles similares a los observados durante crisis asiática. Esto refleja en parte, un cierto grado de homogeneidad en la reacción del mercado laboral frente a shocks externos, al menos, durante la recesión de 1999 y la más reciente aún en proceso. En este contexto y sumando las características de no estacionariedad de la tasa de desempleo, proponemos efectuar un análisis basado en un modelo de series de tiempo del tipo SARIMA para predecir la tasa de desempleo mensual en el corto plazo. Basado en el análisis de densidad espectral, las discontinuidades de la serie de desempleo son puramente cíclicas. La figura 2 destaca un peak en torno a 0.5 ($2\pi/0.5 \approx 12$ meses), lo que refleja la predominancia del componente cíclico de la serie, validando empíricamente la

especificación de nuestro modelo. Esto a su vez permite que el shock de la crisis asiática sea un argumento válido para simular la crisis actual. Es más, las similitudes encontradas entre las trayectorias de los niveles de desempleo de la crisis asiática versus la crisis actual ponen de manifiesto la significativa inercia del mercado laboral, al punto de que los shocks materializados durante la crisis asiática puedan ser replicados en la actualidad.

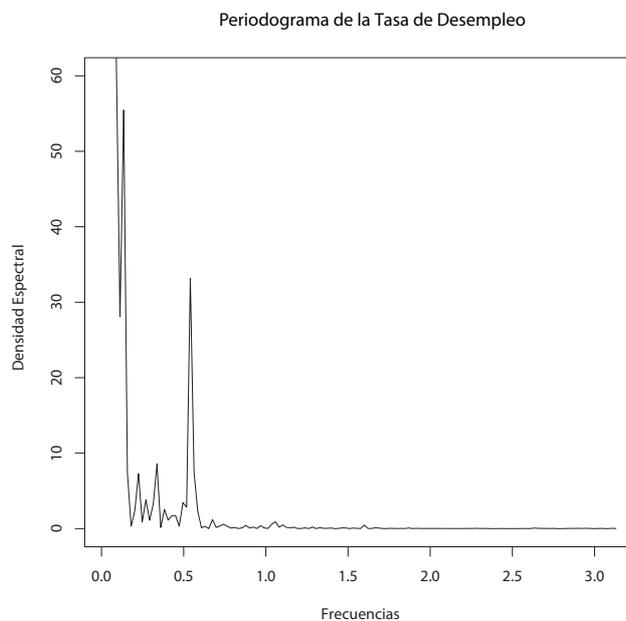


Figura 2: Periodograma de la Tasa de Desempleo obtenido a través de la Transformación Rápida Discreta de Fourier (Singleton 1979).

Por último, a diferencia de otros estudios en los que se modela una estructura económica para ajustar la tasa de desempleo nacional, el presente estudio permite que los propios datos temporales de la variable desempleo nos revelen las características de la estructura probabilística subyacente. En este sentido, el presente documento no indaga sobre los factores que están detrás de la posible rigidez del mercado laboral, sino que más bien, el objetivo es modelar y proyectar la tasa de desempleo en base a su propia historia, es decir, valiéndonos únicamente de su significativa correlación con el pasado. Por otra parte, la elevada persistencia y volatilidad de la serie de desempleo hizo que cualquier intento de estabilización mediante la aplicación de funciones de transformación de la variable de tipo Box Cox, resultara infructuosa. Además, la función de autocorrelación (ACF) muestra un prolongado decaimiento a medida que aumentan los rezagos de tiempo (Ver figura 1), por lo que en primera instancia sospechamos que podría tratarse de un proceso de larga memoria del tipo ARFIMA (Autoregressive Fractionary Integrated Movable Average), lo cual analizamos en la siguiente sección.

3.1. Ajuste ARFIMA

El término de memoria larga utilizado en el estudio de series de tiempo, suele asociarse con la persistencia que muestran algunas series estacionarias en sus funciones de autocorrelación, que convergen paulatinamente hacia cero. Este comportamiento no es compatible con el modelo SARIMA aquí propuesto, que considera una medida extrema de persistencia de la tasa de desempleo. En este sentido, además del modelo SARIMA, también se ajustó un proceso de larga memoria ARFIMA (Autorregresive Fractionary Integrated Movile Average) para modelar el comportamiento del desempleo, ya que éste muestra cierta persistencia —coherente con el lento ajuste del salario real frente a los distintos estados del ciclo económico. Formalmente se tiene:

$$\Phi(B)X_t = \Theta(B)\nabla_1^{-d}\epsilon_t \Rightarrow (1 - \sum_{k=1}^p \phi_k B^k) = (1 - \sum_{k=1}^q \theta_k B^k)(1 - B)^{-d}\epsilon_t$$

donde, $\Phi(B)$ es un polinomio de coeficientes con operador de rezagos B del componente autorregresivo AR; $\Theta(B)$ es el polinomio de coeficientes del proceso de media móvil MA y d corresponde a la diferenciación fraccionaria del modelo. A continuación se presentan tres estimadores para el parámetro d de la tasa de desempleo:

Método	Estimación \hat{d}	S.E
Reisen	0.829	0.068
Geweke and Porter-Hudak	0.968	0.208
Whittle	1.002	0.033

1. Reisen: Está basado en la ecuación de regresión usando la función de periodograma alisado (smoothed) como una estimación de la densidad espectral.
2. Geweke and Porter-Hudak: Está basado en la ecuación de regresión usando la función de periodograma como una estimación de la densidad espectral.
3. Whittle: Es el estimador local exacto de Whittle a través de métodos cuasi-máximo-verosímiles (QMLE).

Notemos que el estimador de Whittle es el que posee menor error estándar y que su estimación es un valor cercano a 1, por lo que finalmente ajusta un modelo autoregresivo integrado de primer orden, es decir, la tasa de desempleo es no estacionaria. Sin embargo, al descartar la posibilidad de que sea estacionaria con integración fraccionaria, se concluyó que la serie posee una tendencia estocástica, lo que valida nuestro modelo autorregresivo y de media móvil aplicado a la variable de desempleo diferenciada.

3.2. Análisis de tendencia y estacionalidad del ajuste SARIMA

Tras descartar la posibilidad de que la tasa de desempleo sea estacionaria con integración fraccionaria, se optó por la primera diferenciación. En otras palabras, para eliminar la tendencia estocástica, se procedió a diferenciar la serie una vez. El siguiente gráfico muestra la pérdida de tendencia, aunque la variabilidad producida por la estacionalidad de la serie persiste al observar la función de autocorrelación de la serie ajustada. Dado que eurísticamente estacionalidad es sinónimo de no estacionariedad, se realizó el respectivo ajuste estacional de la serie.

La siguiente figura muestra las ACF y la función de autocorrelación parcial (PACF) de la tasa de desempleo después de controlar por su tendencia estocástica y estacionalidad. A partir de la ACF se determinó que existen los siguientes peaks: uno en la parte MA estacional (esto es cada 12 meses por ser la serie de frecuencia mensual) y cuatro en la parte MA de los parámetros ARIMA de la serie. En el gráfico PACF es posible notar que existen los siguientes saltos: dos en la parte AR estacional (períodos 12 y 24, respectivamente) y tres significativos en la parte AR de los parámetros ARIMA de la tasa de desempleo. De esta forma determinamos que nuestro modelo preeliminar es un $ARIMA(3, 1, 4) \times (2, 1, 1)_{12}$, estimado por el método de máxima verosimilitud (MV). No obstante, al estimar los parámetros de este modelo no se encontró significancia para el componente AR estacional según el estadístico *t-student*. De esta forma y tomando en consideración el principio de parsimonia, se decidió estimar el modelo $ARIMA(1, 1, 2) \times (2, 1, 1)_{12}$, el cual obtuvo valores AIC y BIC menores que el inicialmente propuesto (-38.05 contra -37.06 y -25.52 contra -16.16, respectivamente).

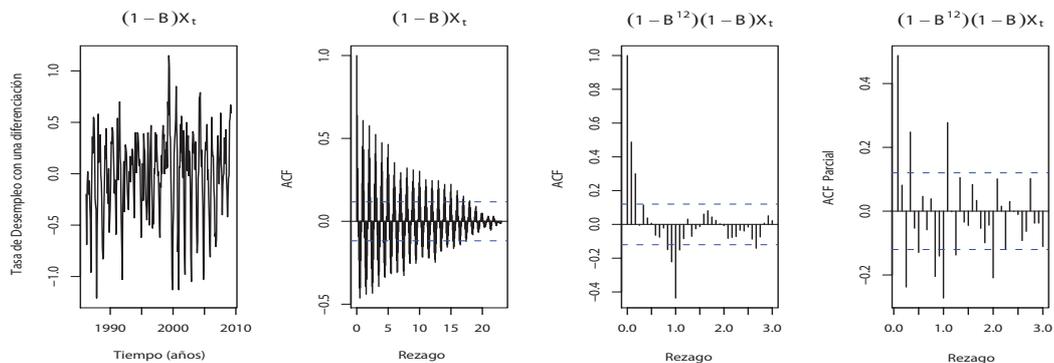


Figura 3: De izquierda a derecha: Seria diferenciada una vez y su ACF, ACF y PACF de la serie diferenciada 12 veces.

Todos los parámetros del modelo $ARIMA(1, 1, 2) \times (0, 1, 1)$ resultaron significativos a un nivel de confianza de 95 % y errores ruido blanco distribuidos normalmente (residuos independientes). La siguiente tabla y gráficos dan cuenta de ello:

Parámetro	Estimación	s.e	Test-t
ϕ_1	-0.2433	0.0751	3.238
θ_1	0.9032	0.0483	18.719
θ_2	0.7726	0.0575	13.434
θ_1^s	-0.8776	0.0566	15.498

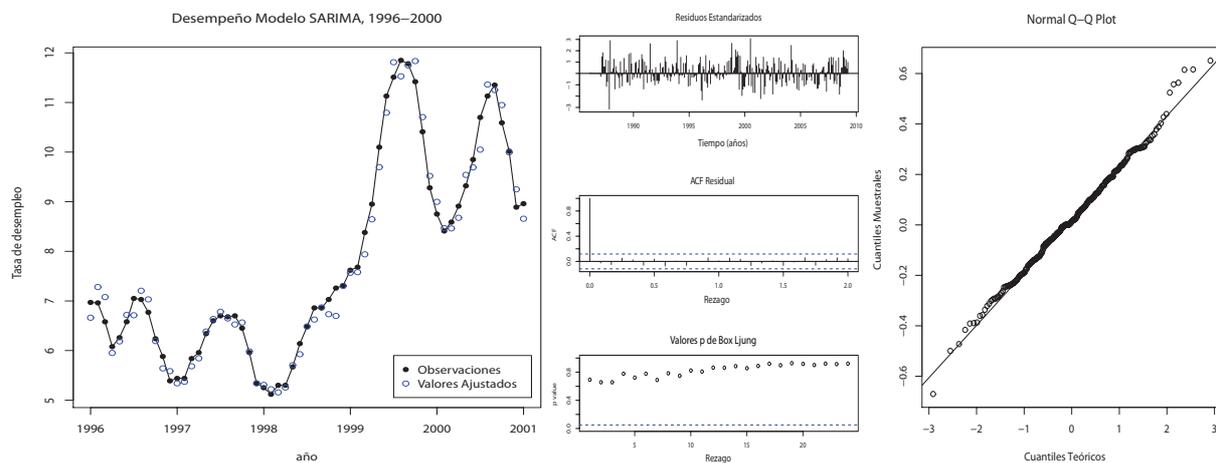


Figura 4: A la izquierda se presenta el gráfico de los valores ajustados por el modelo SARIMA para los años 1996-2000. Al centro los Residuos estandarizados, ACF residual y valores p de Box-Ljung. A la derecha el gráfico Normal Q-Qplot residual.

En la figura 4 se muestra que en la mayoría de los casos, la serie estimada pasa por los puntos observados; sin embargo, aún existen observaciones de carácter estacional que no logran ser capturados por el modelo ajustado. Por lo que, probablemente, los resultados de la estimación estén sugiriendo que un modelo del tipo SARFIMA (Seasonal Autorregresive Fractionary Integrated Movile Average) podría ser el más apropiado para este caso. Por otro lado, en los gráficos del centro se puede apreciar que el test de Box-Ljung muestra que los valores p asociados a los residuos del modelo son mayores que el error de confianza $\alpha = 0.05$ para un rezago equivalente a 24 meses, las autocorrelaciones residuales se concentran dentro de las bandas de Bartlett, por lo que concluimos que los residuos no están correlacionados. La figura del extremo izquierdo muestra que los residuos en gran medida se concentran en la recta normal y el test de normalidad Shapiro-Wilk arroja un valor p de 0.433 ($W = 0.994$) lo cual no rechaza la hipótesis nula de normalidad residual. De esta forma concluimos que los residuos son ruido blanco con distribución normal.

Por otra parte, la estructura de los modelos de series de tiempo que aquí analizamos permiten aproximar el período de permanencia de una perturbación en la tasa de desempleo. Para ello, obtenemos la función de *impulso-respuesta* a partir del modelo:

$$(1 + \phi B)(1 - B)(1 - B^{12})y_t = (1 + \theta_1 B + \theta_2 B^2)(1 - \theta_{12} B^{12})\epsilon_t$$

y definiendo

$$x_t = (1 - B)(1 - B^{12})y_t$$

se obtiene la siguiente expresión:

$$x_t = \frac{(1 + \theta_1 B + \theta_2 B^2)(1 - \theta_{12} B^{12}) \epsilon_t}{(1 + \phi B)}$$

Realizando la descomposición de Wald para el denominador se tiene que:

$$x_t = \left(\sum_{j=0}^{+\infty} (-\phi)^j B^j \right) (1 + \theta_1 B + \theta_2 B^2 - \theta_{12} B^{12} - \theta_{12} \theta_1 B^{13} - \theta_{12} \theta_2 B^{14}) \epsilon_t$$

Definiendo una perturbación $\epsilon_t = 1$, la función de *impulso-respuesta* del modelo está dada por:

$$\omega(j) = \frac{\partial x_t}{\partial \epsilon_{t+j}} = \begin{cases} 1, & j = 0 \\ \theta_1 - \phi, & j = 1 \\ (-\phi)^j - \theta_1(-\phi)^{j-1} + \theta_2(-\phi)^{j-2}, & 2 < j < 12 \\ 2(\phi^{12} - \theta_1 \phi^{11} + \theta_2 \phi^{10} - 0.5\theta_{12}), & j = 12 \\ 2(-\phi^{13} + \theta_1 \phi^{12} - \theta_2 \phi^{11} - 0.5(\theta_1 - \phi)\theta_{12}), & j = 13 \\ 2[-\phi^j + \theta_1 \phi^{j-1} - \theta_2 \phi^{j-2} - 0.5((-\phi)^{j-12} + \theta_1(-\phi)^{j-13} + \theta_2(-\phi)^{j-14})\theta_{12}], & j > 13 \end{cases}$$

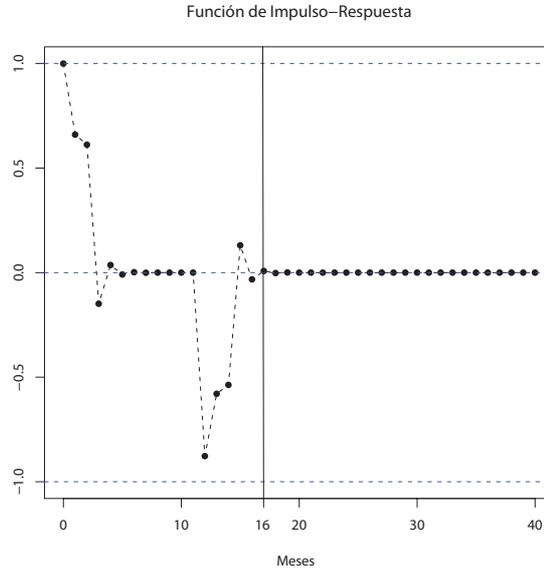


Figura 5: Gráfico de la Función de Impulso Respuesta $\omega(j)$ (la periodicidad de los rezagos es mensual).

Es directo notar del gráfico anterior la elevada persistencia de una perturbación en la serie. Según el modelo propuesto, la persistencia en la dinámica de la tasa de desempleo podría prolongarse por más de un año (esto es hasta 16 meses).

3.3. Predicciones In - Out sample

En esta sección procedemos a evaluar la capacidad predictiva de nuestro modelo mediante la raíz del error cuadrático medio (RECM) tanto de las estimaciones dentro de muestra (*in-sample*) como de las predicciones fuera de muestra (*out-sample*). Para el caso de las predicciones *out-sample*, se generó arbitrariamente una submuestra que representa el 70 % de las observaciones disponibles y se proyectó la tasa de desempleo para un horizonte de uno, tres y doce meses, respectivamente –en términos cronológicos las predicciones comprenden el período mayo 2002 a abril 2003. Por su parte, las predicciones *in-sample* se obtienen a partir de modelo SARIMA estimado con el 100 % de las observaciones disponibles y se calcula el RECM para el mismo período del caso anterior. En términos formales, la raíz del error cuadrático medio (o raíz de la media aritmética del error de predicción) es:

$$RECM(\hat{y}_i) = \sqrt{\frac{1}{p} \sum_{i=1}^p (y_i - \hat{y}_i)^2}$$

donde p es el horizonte de predicción, \hat{y}_i es la i -ésima predicción e y_i es el valor real de la serie.

Mes/año	Predicción a 12 meses Out-Sample	Predicción a 3 meses Out-Sample	Predicción a 1 mes Out-Sample	Predicción In-Sample	Observación
05/2002	9.983 (± 0.220)	9.983 (± 0.220)	9.983 (± 0.220)	9.981	9.93
06/2002	10.282 (± 0.422)	10.282 (± 0.422)	10.282 (± 0.219)	10.192	10.30
07/2002	10.610 (± 0.644)	10.610 (± 0.644)	10.608 (± 0.218)	10.667	10.28
08/2002	10.719 (± 0.788)	10.718 (± 0.214)	10.716 (± 0.218)	10.209	10.50
09/2002	10.727 (± 0.913)	10.724 (± 0.411)	10.720 (± 0.217)	10.446	10.63
10/2002	10.403 (± 1.023)	10.401 (± 0.628)	10.399 (± 0.216)	10.666	10.52
11/2002	9.877 (± 1.122)	9.876 (± 0.214)	9.874 (± 0.216)	9.950	9.75
12/2002	9.220 (± 1.212)	9.222 (± 0.411)	9.222 (± 0.215)	8.885	8.70
01/2003	9.191 (± 1.297)	9.191 (± 0.628)	9.190 (± 0.214)	8.471	8.53
02/2003	9.206 (± 1.376)	9.204 (± 0.214)	9.204 (± 0.214)	8.511	8.80
03/2003	9.480 (± 1.451)	9.474 (± 0.411)	9.473 (± 0.213)	9.319	9.21
04/2003	9.542 (± 1.522)	9.528 (± 0.628)	9.527 (± 0.213)	9.364	9.59
RECM	0.309	0.309	0.308	0.207	

De la figura 6 y la tabla anterior se desprende que las predicciones fuera de muestra resultan bastante exitosas en el corto plazo. Esto se constata en la inexistente diferencia entre la serie real y las predicciones dentro y fuera de muestra para un horizonte de dos meses. De igual forma, para un horizonte de predicción de mayor alcance, el modelo SARIMA es capaz de capturar los puntos de inflexión (o puntos de giro) del ciclo de la tasa de desempleo, pese a sus limitaciones para ajustarse a la variabilidad absoluta de los datos.

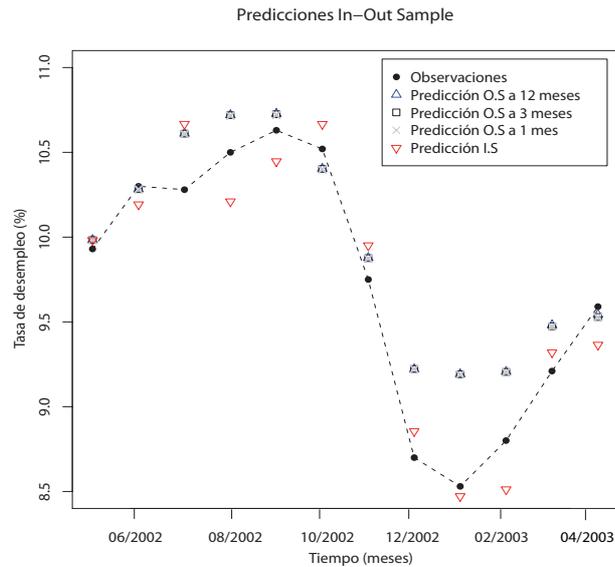


Figura 6: Predicciones fuera de muestra (*in-sample*) y dentro de muestra (*out sample*) de la tasa de desempleo nacional a 1, 3 y 12 meses para el período comprendido entre 05/2002 a 04/2003.

3.4. Perspectivas del error cuadrático medio de predicciones out-sample

En esta sección, nuestro objetivo es realizar comparaciones entre el modelo $ARIMA(1, 1, 2) \times (0, 1, 1)_{12}$, el modelo estándar $AR(1)$ y el modelo no estacional $ARIMA(1, 1, 2)$ (con los mismos componentes no estacionales del modelo $SARIMA$). Para las comparaciones se recurrió al RECM, al criterio tradicional AIC (Criterio de Información de Akaike) y al Criterio BIC para los cuales, se realizaron los siguientes pasos:

1. Se seleccionaron las primeras observaciones correspondientes al 70 % de la muestra (195 observaciones).
2. De esa muestra, se realizó una predicción de un horizonte de 12 meses del cual se extrajo solamente la primera estimación para luego incorporarla a la base original de 195 observaciones.
3. Se realizó esta operación 73 veces hasta agotar la muestra. Es decir, hasta que la posición de la última predicción y del último horizonte estimado coincidiera con el último dato que completa el 100 % de las observaciones.
4. Dado que se han obtenido 73 series con 84 valores estimados a partir de 73 modelos, se realizó en cada paso el cálculo del RECM, AIC y BIC, respectivamente.
5. A partir de estos resultados, se procedió a calcular los estadígrafos de media y desviación estándar como forma de sintetizar los resultados.

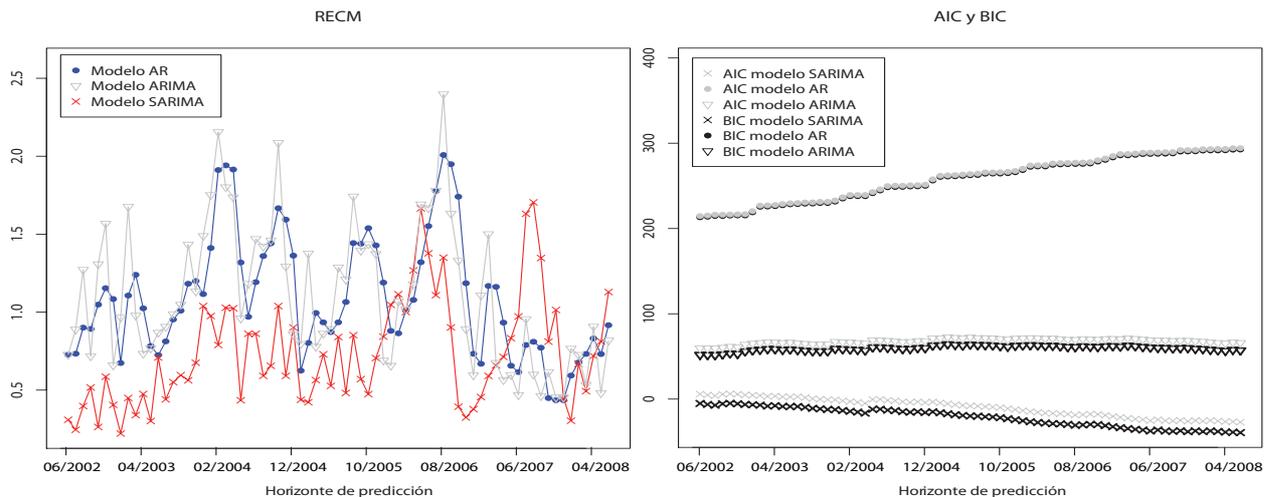


Figura 7: Gráfico de RECM y AIC.

Note que los valores AIC y BIC del modelo SARIMA son inferiores respecto de sus símiles de los modelos AR(1) y ARIMA, cuyo resultado es corroborado por el gráfico de RECM. Los promedios RECM, muestran una pequeña diferencia entre el modelo estacional y no estacional, aunque el estacional presenta menor volatilidad. No obstante, cabe mencionar que pese a que existen episodios en que el modelo SARIMA no logra capturar la variabilidad absoluta de los datos sobrestimando de esta forma el valor real, es capaz de alinearse a los puntos de inflexión o puntos de giro del ciclo de la tasa de desempleo –tal como se corroboró en la sección anterior. La figura 7 muestra que los modelos en promedio presentan menor RECM, mostrando episodios con mayor error en relación al modelo estándar AR(1), tal es el caso de los meses 01/2006 y 06/2007 en donde coinciden con el brusco decaimiento de la tasa de desempleo en esos años.

4. Comentarios Finales

Según el análisis espectral, encontramos que los cambios de la tasa de desempleo son netamente cíclicas. Por otra parte, la función de autocorrelación y los métodos de estimación de diferenciación fraccionaria de un modelo ARFIMA permitieron descartar que la variable de interés se comporte como un proceso estacionario. En este contexto, se decidió modelar la tasa de desempleo como un proceso $ARIMA(1, 1, 2) \times (0, 1, 1)_{12}$, según el criterio de información de Akaike, BIC y función de autocorrelación. La validación de este modelo se sustentó en los siguientes métodos: test residual de Box-Ljung y la raíz cuadrática media del error (RECM) de las predicciones *out-sample*. El RECM del modelo seleccionado fue menor con relación a los obtenidos de los procesos estándares AR(1) y ARIMA(1, 1, 2). No obstante, cabe mencionar que pese a que existen episodios en que el modelo SARIMA no logra capturar la variabilidad ab-

solita de los datos sobrestimando de esta forma el valor real, es capaz de alinearse a los puntos de inflexión o puntos de giro del ciclo de la tasa de desempleo

Por otra parte, la estructura de los modelos de series de tiempo que aquí analizamos, permiten aproximar el período de permanencia de una perturbación en la tasa de desempleo. Según nuestro modelo, la persistencia en la dinámica de la tasa de desempleo podría prolongarse por más de un año (esto es hasta 16 meses).

Tras los efectos de la crisis internacional, la tasa de desempleo ha tendido nuevamente al alza e inclusive, hasta alcanzar niveles similares a los observados durante crisis asiática. Esto refleja en parte, un cierto grado de homogeneidad en la reacción del mercado laboral frente a shocks externos, al menos durante la recesión de 1999 y la más reciente aún en proceso según se observa en la figura 8.

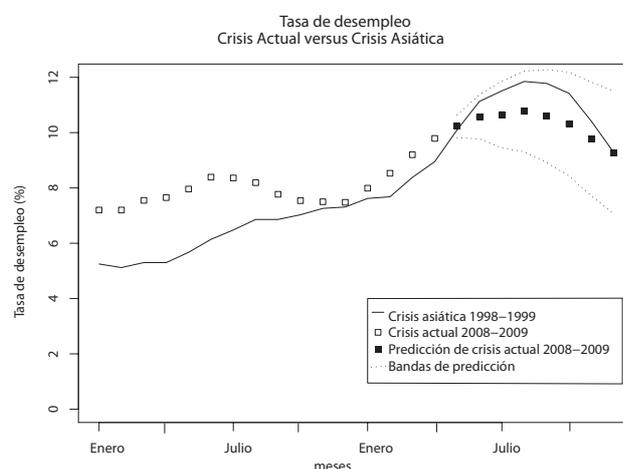


Figura 8: Gráfico de la Tasa de desempleo de los años 1998-1999, 2008 y predicciones desde Junio a Diciembre de 2009

Lo anterior a su vez confirma que el shock de la crisis asiática podría ser un argumento válido para simular la crisis actual, poniendo de manifiesto la significativa inercia del mercado laboral al punto de que los shocks materializados en el primer caso puedan ser replicados en la actualidad. Finalmente, si bien en la mayoría de los casos la serie estimada replica la trayectoria de la tasa de desempleo reportada mensualmente por el INE, aún existen observaciones de carácter estacional que no logran ser ajustadas por el modelo propuesto. Por lo que probablemente los resultados obtenidos estén sugiriendo que un modelo del tipo SARFIMA (Seasonal Autorregresive Fractional Integrated Movile Average) sea el más apropiado para este caso, lo que dejamos propuesto para futuras investigaciones.

Referencias

- [1] Albagli, E; G Contreras; P. García; I. Magendzo; R. Valdés. *Errores de Predicción en Perspectiva. Banco Central de Chile*. Documento de Trabajo No. 199. Enero 2003.
- [2] Bergoing, R.; F. Mornadé; F. Piguillem. *Labor Market Distortions, Employment, and Growth: The Recent Chilean Experience, General Equilibrium Models for The Chilean Economy*, R. Chumacero y K. Schmidt-Hebbel, editores, Banking Development Series, Central Bank of Chile, 395-414, 2005.
- [3] Cowan, K., A. Micco, A. Mizala, C. Pages y P. Romaguera. (2003) *Un Diagnóstico del Desempleo en Chile*, Mimeo, IADB.
- [4] Christian Johnson. (2001) *Un Modelo de Switching para el Crecimiento en Chile*, Cuadernos de Economía, Vol. 38, N° 115, pp. 291-319.
- [5] Chumacero, R. y J. Quiroz. (1996) *La Tasa Natural de Crecimiento de la Economía Chilena: 1985-1996*, Cuadernos de Economía 100, 453-72.
- [6] Geweke, J. and Porter-Hudak, S. (1983) *The Estimation and Application of Long Memory Time Series Model*. Journal of Time Series Analysis, 4, 221-238.
- [7] Jonathan D. Cryer; Kung-Sik Chan. (2008) *Time Series Analysis With Applications in R*, Second Edition, Springer.
- [8] R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [9] Reisen, V.A. (1994) *Estimation of the Fractional Difference Parameter in the ARFIMA(p,d,q) model using the Smoothed Periodogram*. Journal of Time Series Analysis, 15, 335-350.
- [10] Singleton, R. C. (1979) *Mixed Radix Fast Fourier Transforms, in Programs for Digital Signal Processing*. IEEE Digital Signal Processing Committee eds. IEEE Press.
- [11] Whittle, P. (1962) *Gaussian estimation in stationary time series*. Bulletin of the International Institute of Statistics, 39, 105-129.

5. Apéndice

5.1. A1: Ajuste del modelo SARIMA

La siguiente tabla muestra distintas especificaciones de modelos de series de tiempo con sus respectivos estadígrafos de ajuste: Criterio de Información de Akaike (AIC) y Criterio de Información Bayesiana (BIC). De acuerdo a estos resultados, se eligió el modelo $ARIMA(1, 1, 2) \times (0, 1, 1)_{12}$ que además de ser parsimonioso exhibió un AIC y BIC relativamente altos en valor absoluto.

Modelo	BIC	AIC
$ARIMA(1, 1, 1) \times (0, 1, 1)_{12}$	27.613	18.719
$ARIMA(1, 1, 2) \times (0, 1, 1)_{12}$	-25.527	-38.052
$ARIMA(1, 1, 3) \times (0, 1, 1)_{12}$	-20.295	-36.451
$ARIMA(1, 1, 4) \times (0, 1, 1)_{12}$	-15.451	-35.238
$ARIMA(2, 1, 1) \times (0, 1, 1)_{12}$	14.478	1.953
$ARIMA(2, 1, 2) \times (0, 1, 1)_{12}$	-20.722	-36.878
$ARIMA(2, 1, 3) \times (0, 1, 1)_{12}$	-15.661	-35.449
$ARIMA(2, 1, 4) \times (0, 1, 1)_{12}$	-13.453	-36.872
$ARIMA(3, 1, 1) \times (0, 1, 1)_{12}$	0.193	-15.962
$ARIMA(3, 1, 2) \times (0, 1, 1)_{12}$	-17.306	-37.093
$ARIMA(3, 1, 3) \times (0, 1, 1)_{12}$	-11.869	-35.287
$ARIMA(3, 1, 4) \times (0, 1, 1)_{12}$	-11.775	-38.825
$ARIMA(4, 1, 1) \times (0, 1, 1)_{12}$	-1.009	-20.797
$ARIMA(4, 1, 2) \times (0, 1, 1)_{12}$	-12.107	-35.526
$ARIMA(4, 1, 3) \times (0, 1, 1)_{12}$	-6.045	-33.095
$ARIMA(4, 1, 4) \times (0, 1, 1)_{12}$	-8.987	-39.667