

Un Modelo Estadístico para la Desagregación Regional del Índice Mensual de Actividad del Construcción*

Facundo Piguillem

Mauricio Tejada

Septiembre 2006

Resumen

Tanto a nivel nacional como regional, el sector construcción es uno de los sectores más importantes de la economía chilena. Esto no sólo por su peso en la producción agregada, sino también porque es muy sensible a los ciclos económicos. Sin embargo, y a pesar de dicha importancia, no se cuenta con información a nivel regional que permita hacer un seguimiento periódico al pulso de la actividad de la construcción en cada región. Así, el objetivo de este trabajo es proveer una herramienta estadística que permita estimar indicadores regionales a partir de la desagregación del Índice Mensual de Actividad de la Construcción (IMACON), utilizando para ello un conjunto de indicadores parciales de actividad que sirven como instrumentos observados en cada región.

Clasificación JEL: C8, N6 y R1

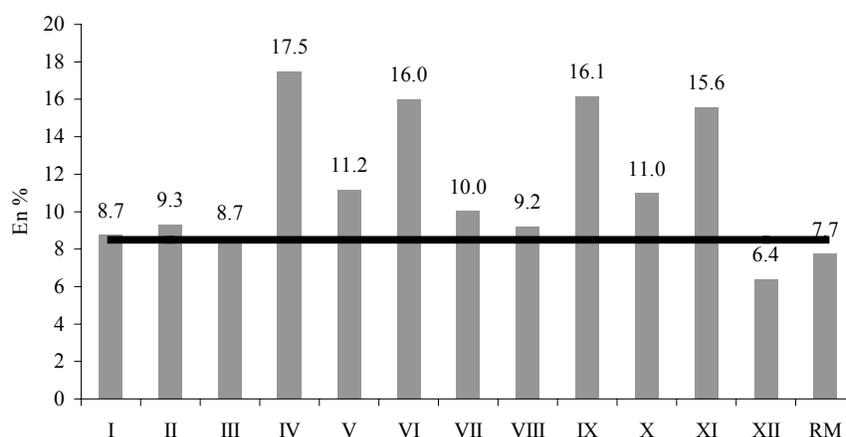
Palabras Clave: Sector Construcción, Desagregación Regional, Variables Instrumentales.

* Comentarios bienvenidos a mtejada@cchc.cl. Un agradecimiento a Byron Idrovo por sus valiosos comentarios. La publicación de los Documentos de Trabajo no está sujeta a la aprobación previa de la Mesa Directiva de la Cámara Chilena de la Construcción. Tanto el contenido de los Documentos de Trabajo, como también el análisis y conclusiones que de ellos se deriven, son exclusiva responsabilidad de su(s) autor(es) y no reflejan necesariamente la opinión de la Cámara Chilena de la Construcción o sus directivos.

1. Introducción

El sector construcción es uno de los sectores más importantes de la economía chilena, no sólo por su peso en la producción agregada, sino también porque es uno de los más sensibles a los ciclos económicos (Piguillem, 2004). Lo anterior también es cierto a nivel regional, donde el sector construcción representa una parte importante del valor agregado de cada región. En efecto, a nivel nacional el sector construcción representa poco más del 8% del valor agregado total de la economía, mientras que a nivel regional para más del 80% de las regiones dicha participación supera a la del promedio nacional (Figura 1).

Figura 1: Participación del PIB de la Construcción en el PIB Agregado Regional (Promedio 1996 - 2003)



Fuente: Banco Central de Chile

Sin embargo, y a pesar de dicha importancia, a nivel regional no se cuenta con información pública que permita hacer un seguimiento periódico a la actividad de la construcción, esto a diferencia del caso de la actividad sectorial agregada.¹ Lo anterior debido a que la última información disponible a nivel regional, que está publicada en el anuario de cuentas nacionales del Banco Central, se encuentra sólo para el período 1996-2003 y con una frecuencia anual. Más aún, a pesar de que el INE cuenta con un indicador de la construcción para el cálculo de los indicadores regionales de actividad (INACER), éste no se encuentra disponible públicamente.

El objetivo de este trabajo es proveer una herramienta estadística que permita hacer un seguimiento al pulso de la actividad de la construcción a nivel regional. En particular, se busca una herramienta que permita la estimación de indicadores regionales a partir de la *desagregación* del Índice Mensual de Actividad de la Construcción (IMACON). Para ello, utilizamos una versión modificada del modelo propuesto originalmente por Chow y Lin (1971) como base metodológica de un modelo de desagregación de series nacionales en series regionales, donde se

¹ ver Piguillem (2004) y Tejada (2006).

busca transformar o *desagregar* un dato nacional en datos regionales utilizando un conjunto de instrumentos observados en cada región.

Se construyeron cuatro modelos, clasificados según los supuestos realizados sobre la matriz de varianzas y covarianzas de los errores de estimación, los cuales van desde un modelo simple donde se imponen los supuestos tradicionales en econometría, hasta un modelo donde se controla por autocorrelación y heteroscedasticidad entre regiones. De la observación de dichos modelos, se encontró que mientras más realista se torna el modelo, es decir, el modelo que controla por autocorrelación y heteroscedasticidad entre regiones, mayores son las ganancias en términos de eficiencia y, además, los indicadores regionales estimados se tornan más volátiles reflejando de alguna forma la heterogeneidad propia de las regiones en Chile. Adicionalmente, se compararon los indicadores regionales estimados y anualizados con el PIB de la construcción, para el período 2001-2003, y se encontró que los modelos que imponen sólo autocorrelación y autocorrelación y heteroscedasticidad, éste último basado en ponderadores estimados por máxima verosimilitud, son los que se asemejan más a la distribución regional del valor agregado observada en el PIB de la construcción.

2. El Modelo

Para la desagregación regional del Índice Mensual de Actividad de la Construcción (IMACON) utilizamos una versión modificada del modelo propuesto originalmente por Chow y Lin (1971). Este modelo fue creado con el objetivo de aumentar la frecuencia de una serie de tiempo, por ejemplo, transformando una serie trimestral en una mensual. En este sentido, se puede decir que esta metodología desagrega una observación temporal en varias observaciones temporales según sea la transformación de frecuencia que sea elegida. Para dicha desagregación, el modelo de estos autores utiliza variables que sirven de instrumentos en la estimación de la variable en mayor frecuencia que es no observada. Estos instrumentos deben ser variables que se espera tenga una alta correlación con la variable que se desea estimar y que sean observados en la frecuencia elegida para la transformación.

Así, se puede pensar en el modelo de Chow y Lin (1971) como base metodológica de un modelo de desagregación de series nacionales en series regionales, donde se busca transformar o desagregar un dato nacional en n datos regionales, utilizando para ello un conjunto de instrumentos observados en cada región. Siguiendo esta idea, en lo que sigue de esta sección se desarrolla el marco metodológico que se utilizará para la desagregación del IMACON en trece indicadores regionales de actividad de la construcción. Lo anterior es posible debido a que este indicador de actividad nacional, al representar el valor agregado latente del sector construcción, tiene interpretación en niveles, lo cual permite dar coherencia a la existencia e interpretación de indicadores desagregados a partir de éste (Tejada, 2006).

Para iniciar la exposición del modelo, supongamos que existen N regiones, de las cuales se cuenta con información para T periodos de tiempo en cada una de ellas. Así, la actividad de cada región puede expresarse de la siguiente manera:

$$y_{t,i} = x_{t,i}\beta + \mu_{t,i} \quad \text{con } i=1,\dots,N \text{ y } t=1,\dots,T \quad (1)$$

Donde $y_{t,i}$ es el indicador que representa la actividad regional y es no observado, $x_{t,i}$ es un vector que contiene P indicadores parciales de la actividad en cada región y β es un vector de dimensión $(1 \times P)$ de parámetros que relacionan los indicadores parciales con el indicador de actividad regional. Finalmente, $\mu_{t,i}$ es un vector de errores aleatorios que sigue una distribución normal. Por el momento no impondremos supuestos adicionales sobre las características de los momentos poblacionales de dichos errores. De esta manera y usando los N indicadores regionales podemos definir la actividad agregada como:²

$$Y_t = \sum_{i=1}^n y_{t,i} \quad \text{con } t=1,\dots,T \quad (2)$$

De la misma forma, se pueden definir las contrapartes agregadas de los indicadores parciales contenidos en el vector $x_{t,i}$ como:

$$X_t^p = \sum_{i=1}^n x_{t,i}^p \quad \text{con } t=1,\dots,T \text{ y } p=1,\dots,P \quad (3)$$

Para fines de la estimación podemos escribir este modelo de una manera más concisa definiendo lo siguiente en notación matricial:

$$y = \begin{bmatrix} y_{1,1} \\ y_{1,2} \\ \cdot \\ \cdot \\ y_{1,N} \\ y_{2,1} \\ y_{2,2} \\ \cdot \\ \cdot \\ y_{T,N-1} \\ y_{T,N} \end{bmatrix}, \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_t \\ \cdot \\ \cdot \\ \cdot \\ Y_{T-1} \\ Y_T \end{bmatrix}, \quad x = \begin{bmatrix} x_{1,1}^1 & \dots & x_{1,1}^P \\ x_{1,2}^1 & \dots & x_{1,2}^P \\ \cdot & & \cdot \\ \cdot & & \cdot \\ x_{1,N}^1 & \dots & x_{1,N}^P \\ x_{2,1}^1 & \dots & x_{2,1}^P \\ x_{2,2}^1 & \dots & x_{2,2}^P \\ \cdot & & \cdot \\ \cdot & & \cdot \\ x_{T,N-1}^1 & \dots & x_{T,N-1}^P \\ x_{T,N}^1 & \dots & x_{T,N}^P \end{bmatrix} \quad \text{y} \quad X = \begin{bmatrix} X_1^1 & \dots & X_1^P \\ X_2^1 & \dots & X_2^P \\ \cdot & & \cdot \\ \cdot & & \cdot \\ X_t^1 & \dots & X_t^P \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ X_{T-1}^1 & \dots & X_{T-1}^P \\ X_T^1 & \dots & X_T^P \end{bmatrix}$$

² A lo largo del artículo utilizaremos la convención de denominar a las variables regionales con letras minúsculas y a las nacionales con sus contrapartes en mayúsculas.

De esta manera podemos expresar la actividad y los indicadores parciales en términos agregados como:

$$Y = Cy \quad y \quad X = Cx \quad (4)$$

Donde y es un vector ($NT \times 1$), Y es un vector ($T \times 1$), x es una matriz ($NT \times P$), X es una matriz ($T \times P$) y C es la matriz de agregación de dimensión ($T \times NT$) con la siguiente estructura:

$$C = \begin{bmatrix} 1 & 1 & . & . & . & 1 & 0 & 0 & . & . & . & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & . & . & 0 \\ . & . & . & . & . & . & . & . & . & . & . & . \\ 0 & 0 & . & . & . & . & . & . & . & . & 1 & 1 \end{bmatrix} = I_T \otimes e_n$$

Donde I_T es la matriz identidad de dimensión T , e_n es un vector de unos de dimensión ($1 \times N$) y \otimes representa el producto de Kronecker. Haciendo uso de estas definiciones podemos escribir la ecuación (1) como:

$$y = x\beta + \mu \quad (5)$$

Donde, en términos generales, la matriz de varianzas y covarianzas de los errores es $\text{cov}(\mu) = \Phi$. Así, utilizando la ecuación (5) y la matriz de agregación C se puede escribir la relación agregada, que es observada en cada momento del tiempo, como:

$$Y = Cy = Cx\beta + C\mu = X\beta + \Sigma \quad (6)$$

Nótese en la ecuación (6) que la variable Y , que representa el indicador agregado de actividad, es una variable observable, en tanto que la matriz y construida sobre la base de los indicadores regionales no lo es. Asimismo, dado que los errores agregados son $\Sigma = C\mu$ y la matriz de varianzas y covarianzas de los errores μ la definimos como $\text{cov}(\mu) = \Phi$, entonces se puede escribir la matriz de varianzas y covarianzas de los errores agregados como $\text{cov}(\Sigma) = C\Phi C'$. En este contexto, el objetivo de la estimación es recuperar el vector y usando x y a partir de la relación existente entre los indicadores agregados dada en la ecuación (6). Siguiendo a Chow y Lin (1971) el estimador por mínimos cuadrados generalizados (MCG) de β esta dado por:

$$\hat{\beta} = \left[X'(C\Phi C')^{-1} X \right]^{-1} \left[X'(C\Phi C')^{-1} Y \right] \quad (7)$$

En tanto que el *estimador óptimo* de y dados los parámetros β hallados mediante la ecuación (7) está dado por:³

$$\hat{y} = x\hat{\beta} + (\Phi C')(C\Phi C')^{-1}\hat{\Sigma} \quad (8)$$

Así, la estimación de los parámetros se realiza corrigiendo el modelo utilizando la matriz de varianzas y covarianzas de los errores y sobre la base de los indicadores agregados, los cuales son observables en todo momento del tiempo. Los indicadores regionales, en tanto, se recuperan mediante la suma entre la parte explicada por los indicadores parciales, cuya relación con el indicador regional es idéntica a la relación agregada (esto es, utilizando los mismos parámetros estimados), y el error regional. Este último es estimado desagregando el error de la relación agregada mediante la información provista por la matriz de varianzas y covarianzas.

Claramente los resultados del modelo dependen crucialmente de los supuestos que se impongan sobre la matriz de varianzas y covarianzas de los errores regionales, esto es Φ asociado a μ en la ecuación (5). En las siguientes tres subsecciones se detallan tres distintas formas en las que tratamos dicha matriz para la estimación de la ecuación mencionada y se presenta en cada caso el procedimiento de estimación, tanto de los parámetros β como de los *indicadores regionales desagregados*.

2.1 Modelo con Errores Homocedásticos y no Autocorrelacionados

Una primera aproximación sería asumir, como en Chow y Lin (1971), que los errores no están correlacionados temporalmente y que la varianza es la misma para todas las regiones. Sin embargo, a diferencia del estudio antes citado, el cual trata principalmente la interpolación de datos dentro de un determinado periodo de tiempo, necesitamos asumir además que las perturbaciones entre regiones no están correlacionadas. Por lo tanto, bajo una primera aproximación adoptamos los siguientes supuestos:

$$\mu_{t,i} \sim N(0, \sigma_{\mu}^2) \quad \forall i = 1, \dots, N \quad (9)$$

$$E(\mu_{t,i}, \mu_{t-1,i}) = 0 \quad \forall i = 1, \dots, N \quad (10)$$

$$E(\mu_{t,i}, \mu_{t,j}) = 0 \quad \forall i \neq j, i = 1, \dots, N \text{ y } j = 1, \dots, N \quad (11)$$

Lo que implica que:

$$E(\mu_{t,i}, \mu_{t-s,j}) = 0 \quad \forall i \neq j, i = 1, \dots, N, j = 1, \dots, N \text{ y } \forall s \quad (12)$$

³ Ver Chow y Lin (1971) para una demostración formal del carácter óptimo de la estimación del vector no observado.

El supuesto (9) esconde en realidad dos supuestos. En primer lugar, implica que la distribución de los errores es normal, y en segundo lugar, que la desviación estándar de los errores es la misma para todas las regiones (homoscedasticidad). El primero de estos supuestos es de uso habitual en estudios econométricos y por lo tanto se impone sin pérdida de generalidad. El segundo sin embargo, es un poco más difícil de justificar en este caso en particular, ya que es altamente probable que los errores en las distintas regiones exhiban diferente variabilidad.

El supuesto (10) implica que para cada región en particular, el error en un período en determinado no está correlacionado con el error en ningún otro período. A priori, este supuesto sería restrictivo ya que en general, cuando se trabaja con series de tiempo, los errores de las relaciones ajustadas muestran una elevada autocorrelación (Greene, 2003). De todas maneras la validez de este supuesto será verificada posteriormente.

Finalmente, el supuesto (11) implica que los residuos en cualquier región determinada no se encuentran relacionados con los residuos en cualquier otra región. Nuevamente, este supuesto puede ser altamente restrictivo, pero posteriormente consideraremos otras alternativas.

Por lo tanto, bajo los supuestos (9)-(12) se obtiene una simple expresión para la matriz de varianzas y covarianzas, la cual está dada por:

$$\Phi = \sigma_{\mu}^2 I_{NT \times NT} \quad (13)$$

Donde $I_{NT \times NT}$ es una matriz identidad de dimensión $(NT \times NT)$. Es evidente de la ecuación (7) que en este caso el estimador de mínimos cuadrados generalizados de (7) está dado por $\hat{\beta} = (X'CC'X)^{-1}(X'CC'Y)$. En tanto que de la ecuación (8) se desprende que el error observado en la relación agregada es asignado en igual proporción a cada región, esto es, una proporción $1/N$ del error agregado es asignado a cada región.

2.2 Modelo con Errores Autocorrelacionados

Un problema común asociado al uso de series de tiempo en el análisis econométrico, es la presencia de correlación serial entre los errores de distintos períodos o autocorrelación. Este problema podría surgir debido a la correlación temporal existentes entre aquellas variables omitidas del modelo y aquellas variables que si son incluidas. Siguiendo a Greene (2003), es bien conocido que la autocorrelación introduce problemas importantes en la estimación de un modelo de regresión como el descrito en la ecuación (6). En particular, la estimación por mínimos cuadrados ordinarios, esto es tomando los supuestos descritos en la subsección anterior, es ineficiente e inconsistente lo que invalida completamente cualquier inferencia basada en esta estimación.

Por esta razón, y en un primer paso, relajamos el supuesto (10) y más bien suponemos que los errores de regresión en la ecuación (5) siguen un proceso autoregresivo de orden uno ($AR(1)$), el mismo que puede ser expresado de la siguiente manera:

$$\mu_{t,i} = \rho\mu_{t-1,i} + \varepsilon_{t,i} \quad \forall i = 1, \dots, N \quad (14)$$

Donde se supone que:

$$\varepsilon_{t,i} \sim N(0, \sigma_\varepsilon^2) \quad \forall i = 1, \dots, N \quad (15)$$

Esto es, ahora $E(\mu_{t,i}, \mu_{t-1,i}) \neq 0$ pero se mantiene el supuesto de homoscedasticidad en los errores, como queda establecido en (15), y por simplicidad se asume que el coeficiente de autocorrelación (ρ) es el mismo para todas las regiones. Para determinar la estructura que tendría la matriz de varianzas y covarianzas de los errores, recordemos que la misma se puede escribir en términos generales como (Greene, 2003):

$$\Phi = E\mu'\mu = \begin{bmatrix} E\mu_{11}\mu_{11} & E\mu_{11}\mu_{12} & \dots & E\mu_{11}\mu_{1,N-1} & E\mu_{11}\mu_{1n} & \dots & E\mu_{11}\mu_{T,N-1} & E\mu_{11}\mu_{T,N} \\ E\mu_{12}\mu_{11} & E\mu_{12}\mu_{12} & \dots & E\mu_{12}\mu_{1,N-1} & E\mu_{12}\mu_{1N} & \dots & E\mu_{12}\mu_{T,N-1} & E\mu_{12}\mu_{T,N} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots \\ E\mu_{t,N-1}\mu_{11} & E\mu_{t,N-1}\mu_{12} & \dots & E\mu_{t,N-1}\mu_{1,N-1} & E\mu_{t,N-1}\mu_{1N} & \dots & E\mu_{t,N-1}\mu_{T,N-1} & E\mu_{t,N-1}\mu_{T,N} \\ E\mu_{t,N}\mu_{11} & E\mu_{t,N}\mu_{12} & \dots & E\mu_{t,N}\mu_{1,N-1} & E\mu_{t,N}\mu_{1N} & \dots & E\mu_{t,N}\mu_{T,N-1} & E\mu_{t,N}\mu_{T,N} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ E\mu_{t+1,1}\mu_{11} & E\mu_{t+1,1}\mu_{12} & \dots & E\mu_{t+1,1}\mu_{1,N-1} & E\mu_{t+1,1}\mu_{1N} & \dots & E\mu_{t+1,1}\mu_{T,N-1} & E\mu_{t+1,1}\mu_{T,N} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ E\mu_{T,N-1}\mu_{11} & E\mu_{T,N-1}\mu_{12} & \dots & E\mu_{T,N-1}\mu_{1,N-1} & E\mu_{T,N-1}\mu_{1N} & \dots & E\mu_{T,N-1}\mu_{T,N-1} & E\mu_{T,N-1}\mu_{T,N} \\ E\mu_{T,N}\mu_{11} & E\mu_{T,N}\mu_{12} & \dots & E\mu_{T,N}\mu_{1,N-1} & E\mu_{T,N}\mu_{1N} & \dots & E\mu_{T,N}\mu_{T,N-1} & E\mu_{T,N}\mu_{T,N} \end{bmatrix}$$

Por lo tanto, de acuerdo con los nuevos supuestos, (14) y (15), y los que se mantienen de la subsección anterior, (9), (11) y (12), se tiene que un elemento representativo de la diagonal principal de la matriz de varianzas y covarianzas Φ puede escribirse de la siguiente manera:

$$E\mu_{t,i}\mu_{t,i} = E[(\rho\mu_{t-1,i} + \varepsilon_{t,i})(\rho\mu_{t-1,i} + \varepsilon_{t,i})] = \rho E\mu_{t-1,i}\mu_{t-1,i} + E\varepsilon_{t,i}^2 = \frac{\sigma_\varepsilon^2}{1 - \rho^2} \quad \forall i, t \quad (16)$$

Adicionalmente, el elemento representativo situado fuera de la diagonal principal puede expresarse como:

$$E\mu_{t,i}\mu_{t,j} = E[(\rho\mu_{t-1,i} + \varepsilon_{t,i})(\rho\mu_{t-1,j} + \varepsilon_{t,j})] = \rho E\mu_{t-1,i}\mu_{t-1,j} + E\varepsilon_{t,i}\varepsilon_{t,j} = 0 \quad \forall i, t \quad (17)$$

Lo anterior debido a que, como se desprende de los supuestos antes mencionados, se cumple que $E\mu_{t,i}\varepsilon_{t,j} = 0, \forall i \neq j$ y $\forall t = 1, \dots, T$. De la misma manera, se satisface $E\varepsilon_{t,i}\varepsilon_{t,j} = 0, \forall i \neq j$ y $\forall t = 1, \dots, T$. Finalmente, la ecuación (17) inmediatamente implica que $E\mu_{t,i}\mu_{t-s,j} = 0, \forall i \neq j$ y $\forall s = 1, \dots, T-1$. Usando esta información, la matriz de varianzas y covarianzas Φ asociada a los errores μ en la ecuación (5) se puede escribir en notación matricial de la siguiente manera:

$$\Phi = \frac{\sigma_{\varepsilon}^2}{1-\rho^2} \begin{bmatrix} 1 & 0 & \dots & 0 & \rho & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & \rho & \dots & 0 & 0 \\ \dots & \dots \\ \dots & \dots \\ 0 & 0 & \dots & \rho^t & 0 & \dots & \dots & \rho^{T-t-1} & 0 \\ 0 & 0 & \dots & 0 & \rho^t & \dots & \dots & 0 & \rho^{T-t-1} \\ \rho^{t+1} & 0 & \dots & 0 & 0 & \dots & \dots & 0 & 0 \\ \dots & \dots \\ 0 & 0 & \dots & \rho^{T-1} & 0 & \dots & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & \rho^{T-1} & \dots & \dots & 0 & 1 \end{bmatrix}$$

O de manera mas concisa:

$$\Phi = \frac{\sigma_{\varepsilon}^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \dots & \rho^{T-2} & \rho^{T-1} \\ \rho & 1 & \rho & \rho^2 & \dots & \dots & \rho^{T-3} & \rho^{T-2} \\ \rho^2 & \rho & 1 & \rho & \dots & \dots & \rho^{T-4} & \rho^{T-3} \\ \rho^3 & \rho^2 & \rho & 1 & \dots & \dots & \rho^{T-5} & \rho^{T-4} \\ \dots & \dots \\ \dots & \dots \\ \rho^{T-2} & \rho^{T-3} & \rho^{T-4} & \rho^{T-5} & \dots & \dots & 1 & \rho \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \rho^{T-4} & \dots & \dots & \rho & 1 \end{bmatrix}_{T \times T} \otimes \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}_{N \times N}$$

Así, si se quiere aplicar las ecuaciones (7) y (8) para estimar el vector desconocido de indicadores regionales, es necesario primero conocer la matriz Φ . Esto, a su vez, implica tener un conocimiento de los parámetros que gobiernan la estructura de dicha matriz, o lo que es lo mismo, los parámetros adicionales ρ y σ_{ε} . Sin embargo, nótese en la ecuación (7) que el parámetro σ_{ε} no es necesario para la estimación de β dado que al ser factorizado como una constante de la matriz Φ , se cancela cuando se calcula la inversa de esta misma matriz. De la misma forma, al inspeccionar la ecuación (8) se observa que σ_{ε} también puede ser cancelado en

la estimación de \hat{y} . No obstante lo anterior, es posible construir un estimador consistente de σ_ε a partir de la matriz de varianzas y covarianzas de los errores agregados (Σ_t). Para ello, nótese que Σ_t es observado y que, dada la estructura de la matriz de agregación C , la varianza de los errores agregados se puede escribir como:

$$\text{var}(\Sigma_t) = \text{var}\left(\frac{1}{N} \sum_{i=1}^n \mu_{t,i}\right) = \frac{1}{N^2} N \frac{\sigma_\varepsilon^2}{1-\rho^2} \quad (18)$$

De la ecuación (18) es notorio que si ρ fuera conocido, σ_ε podría ser recuperado de las observaciones de Σ_t como $\sigma_\varepsilon^2 = n \text{var}(\Sigma_t)(1-\rho^2)$, lo cual reduce todo el problema a la estimación de ρ . Para la estimación del modelo en este caso, se pueden usar dos procedimientos alternativos, los cuales se detallan a continuación.

El primero de ellos se basa en el hecho, ya mencionado, de que el problema se reduce a hallar una estimación del parámetro ρ . En el espíritu de la metodología propuesta por Cochrane y Orcutt (1949),⁴ usamos un método iterativo buscando el valor de ρ que minimiza la siguiente función:⁵

$$L(\sigma_\varepsilon^2, \hat{\beta}; X | \rho) = \frac{1}{2\sigma_\varepsilon^2} \Sigma' (C\Phi C')^{-1} \Sigma \quad (19)$$

El método iterativo funciona de la siguiente manera:

- 1) Se construye una rejilla de puntos (*grid*) para los posibles valores que puede tomar el parámetro ρ , por ejemplo, puntos igualmente espaciados en el intervalo $[-1,1]$ de manera que los errores sean estacionarios, y se define un nivel de tolerancia, por ejemplo, $tol = 0.00001$.
- 2) Se propone un valor inicial del parámetro ρ para empezar la iteración, por ejemplo $\rho_0 = 0$.
- 3) Para cada uno de los valores de ρ dentro la rejilla de puntos (*grid*) se estima los parámetros β , Φ y σ_ε . Luego, usando (19) se calcula el valor de la función $L(\sigma_\varepsilon^2, \hat{\beta}; X | \rho)$ evaluada en los parámetros anterior y para cada uno de los valores de ρ .
- 4) Se escoge ρ que minimiza el valor de la función $L(\sigma_\varepsilon^2, \hat{\beta}; X | \rho)$, llamemos a este valor ρ_1 .
- 5) Si $|\rho_0 - \rho_1| \leq tol$ entonces $\rho = \rho_0$, caso contrario se define $\rho_0 = \rho_1$ y se vuelve al paso 2.

⁴ Citado en Greene (2003)

⁵ Ver Johnston y Dinardo (1997) para la motivación detrás de la función objetivo del problema de optimización.

Este proceso continuará hasta alcanzar convergencia, es decir, cuando el parámetro ρ no difiera sustancialmente (según la tolerancia) del estimado en la iteración anterior. Una segunda alternativa para la estimación de los parámetros del modelo es usar técnicas de máximo verosimilitud, estimándose así todos los parámetros del modelo de manera conjunta, es decir, β , ρ y σ_ε (Venanzoni y Felici, 2002). En este contexto, se buscan ciertos valores de los parámetros que maximicen la siguiente función:

$$\ln L(\beta, \sigma_\varepsilon^2, \rho | X) = -\frac{T}{2} \ln 2\pi - \frac{T}{2} \ln |C\Phi C'| - \frac{1}{2} \Sigma' (C\Phi C')^{-1} \Sigma \quad (20)$$

Donde las matrices C y Φ del modelo ya fueron definidas. Para la resolución del problema de optimización se puede utilizar la función “*fminunc*” provista por *Matlab*, la cual simplifica enormemente los procedimientos. De cualquier forma, y al igual que en el caso de la subsección anterior, una vez estimados los parámetros del modelo se puede utilizar la ecuación (8) para recuperar los indicadores regionales.

3.3 Modelo con Errores Heterocedásticos entre Regiones

Finalmente, una tercera posibilidad para tratar la matriz de varianzas y covarianzas de los errores del modelo (5) sería aquella que busca controlar por la potencial heterogeneidad existente entre las regiones, como pareciera ser el caso de Chile. Este hecho es importante debido a que repercute directamente en las diferencias de volatilidad de los indicadores entre las regiones. En este contexto, claramente el supuesto (11) puede no ser el más apropiado. Un supuesto alternativo podría ser:

$$\varepsilon_{t,i} \sim N(0, \sigma_{\varepsilon_i}^2) \quad \forall i = 1, \dots, N \quad (21)$$

Nótese que el hecho de adoptar el supuesto de la ecuación (21) no altera el supuesto que $E(\varepsilon_{t,i}, \varepsilon_j) = 0$, para $i \neq j$, por lo que todavía es cierto que $E(\mu_{t,i}, \mu_{t-s,j}) = 0$, para $i \neq j$ y para $s = 0, \dots, T-1$. No obstante lo anterior, la matriz de varianzas y covarianzas si requiere de algunas modificaciones, por lo que ahora tendría la siguiente estructura:

$$\Phi = \frac{1}{1-\rho^2} \begin{bmatrix} \sigma_{\varepsilon 1}^2 & 0 & \dots & 0 & \rho\sigma_{\varepsilon 1}^2 & 0 & \dots & 0 & 0 \\ 0 & \sigma_{\varepsilon 2}^2 & \dots & 0 & 0 & \rho\sigma_{\varepsilon 2}^2 & \dots & 0 & 0 \\ \dots & \dots \\ \dots & \dots \\ 0 & 0 & \dots & \rho^t\sigma_{\varepsilon N-1}^2 & 0 & \dots & \dots & \rho^{T-t-1}\sigma_{\varepsilon N-1}^2 & 0 \\ 0 & 0 & \dots & 0 & \rho^t\sigma_{\varepsilon N}^2 & \dots & \dots & 0 & \rho^{T-t-1}\sigma_{\varepsilon N}^2 \\ \rho^{t+1}\sigma_{\varepsilon 1}^2 & 0 & \dots & 0 & 0 & \dots & \dots & 0 & 0 \\ \dots & \dots \\ 0 & 0 & \dots & \rho^{T-1}\sigma_{\varepsilon N-1}^2 & 0 & \dots & \dots & \sigma_{\varepsilon N-1}^2 & 0 \\ 0 & 0 & \dots & 0 & \rho^{T-1}\sigma_{\varepsilon N}^2 & \dots & \dots & 0 & \sigma_{\varepsilon N}^2 \end{bmatrix}$$

O de manera mas concisa:

$$\Phi = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \dots & \rho^{T-2} & \rho^{T-1} \\ \rho & 1 & \rho & \rho^2 & \dots & \dots & \rho^{T-3} & \rho^{T-2} \\ \rho^2 & \rho & 1 & \rho & \dots & \dots & \rho^{T-4} & \rho^{T-3} \\ \rho^3 & \rho^2 & \rho & 1 & \dots & \dots & \rho^{T-5} & \rho^{T-4} \\ \dots & \dots \\ \dots & \dots \\ \rho^{T-2} & \rho^{T-3} & \rho^{T-4} & \rho^{T-5} & \dots & \dots & 1 & \rho \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \rho^{T-4} & \dots & \dots & \rho & 1 \end{bmatrix}_{T \times T} \otimes \begin{bmatrix} \sigma_{\varepsilon 1}^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_{\varepsilon 2}^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma_{\varepsilon 3}^2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \sigma_{\varepsilon N}^2 \end{bmatrix}_{N \times N}$$

Dada esta estructura, las varianzas individuales asociadas a cada región podrían ser estimadas utilizando un proceso iterativo similar al mencionado anteriormente. Sin embargo, dado que ahora tenemos muchos más parámetros, la rejilla de puntos (*grid*) necesarios crecería exponencialmente, esto es, si consideramos 100 puntos para cada parámetro se deberían calcular 100^N combinaciones posibles, lo cual claramente no parece ser muy eficiente si uno requiere precisión. Al mismo tiempo, existe un problema de identificación de las varianzas $\sigma_{\varepsilon i}^2$ y para resolver este problema, al menos en el proceso de estimación, debemos imponer restricciones adicionales. Bajo esta idea nótese que la varianza de los errores Σ_t puede escribirse como:

$$\text{var}(\Sigma_t) = \text{var}\left(\frac{1}{n} \sum_{i=1}^N \mu_{t,i}\right) = \frac{1}{n^2} \frac{\sum_{i=1}^N \sigma_{\varepsilon i}^2}{(1-\rho^2)} \quad (22)$$

Por lo que se obtiene que $\sum_{i=1}^N \sigma_{\varepsilon i}^2 = N^2 \text{var}(\Sigma_t)(1-\rho^2)$. Ahora definiendo los siguientes ponderadores:

$$w_i = \frac{\sigma_{\varepsilon i}^2}{\sum_{i=1}^N \sigma_{\varepsilon i}^2} \quad (23)$$

Así, utilizando (23), podemos escribir la matriz de varianzas y covarianzas como:

$$\Phi = n^2 \text{var}(\Sigma_t) \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \dots & \rho^{T-2} & \rho^{T-1} \\ \rho & 1 & \rho & \rho^2 & \dots & \dots & \rho^{T-3} & \rho^{T-2} \\ \rho^2 & \rho & 1 & \rho & \dots & \dots & \rho^{T-4} & \rho^{T-3} \\ \rho^3 & \rho^2 & \rho & 1 & \dots & \dots & \rho^{T-5} & \rho^{T-4} \\ \cdot & \cdot & \cdot & \cdot & \dots & \dots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \dots & \dots & \cdot & \cdot \\ \rho^{T-2} & \rho^{T-3} & \rho^{T-4} & \rho^{T-5} & \dots & \dots & 1 & \rho \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \rho^{T-4} & \dots & \dots & \rho & 1 \end{bmatrix} \otimes \begin{bmatrix} w_1 & 0 & 0 & \dots & 0 \\ 0 & w_2 & 0 & \dots & 0 \\ 0 & 0 & w_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & w_n \end{bmatrix}$$

La nueva especificación exhibe dos ventajas, primero el término $N^2 \text{var}(\Sigma_t)$, que es el parámetro que determina el *nivel* de la matriz de varianzas y covarianzas, es observable a partir de los residuos agregados y por lo tanto de fácil estimación. En segundo lugar, los ponderadores w_i tienen que cumplir ciertas condiciones que facilitan su identificación, a saber:

$$\sum_{i=1}^N w_i = 1 \quad (24)$$

$$0 < w_i < 1; \quad \forall i = 1, \dots, n \quad (25)$$

Los parámetros w_i pueden ser estimados de dos maneras. La primera, que recurre a un cierto conocimiento sobre las características heterogéneas de cada región, es utilizar información disponible, como por ejemplo el PIB del sector,⁶ y con ella aproximar las ponderaciones respectivas. A esta primera forma de estimación está asociado el problema de que cualquier error de medida en la información disponible o errores en nuestra creencia de la representatividad de la información utilizada, será trasladada directamente a la estimación de la matriz Φ y por tanto a los estimadores regionales. La segunda manera, que es estadística y suple los problemas antes mencionados, estima los ponderadores w_i minimizando (19) sujeto a las restricciones (24) y (25).

⁶ El PIB de la construcción, publicado por el Banco Central de Chile, está disponible en la información de cuentas nacionales para el período 1996-2003 y en frecuencia anual.

A diferencia de la subsección anterior donde se utilizó un *grid* de puntos, que requiere como se comentó antes una enorme cantidad de cálculos, esta vez utilizamos la función “*fmincon*” provista por *Matlab*, lo que simplifica enormemente los procedimientos.

3. Estimaciones y Resultados

A los efectos de estimación, una de las etapas más importantes para una adecuada identificación de los indicadores regionales es escoger apropiadamente la matriz x de indicadores parciales. Sin embargo, en el caso que no ocupa, es decir la actividad regional de la construcción, esta etapa se simplifica enormemente ya que, al menos en lo que se refiere a nuestro conocimiento, sólo existen tres indicadores de actividad de la construcción con disponibilidad regional, estos son: (1) Empleo en el Sector Construcción, disponible en frecuencia mensual desde 1986 (aunque se refiere a trimestres móviles) y provista por el INE; (2) Despachos de Cemento, disponible en frecuencia mensual también pero recién desde enero de 2000 (CChC); y (3) Solicitudes de Permisos de Edificación, disponible en frecuencia mensual desde enero de 1996 (INE).

Por otro lado, se disponen dos medidas de actividad agregada, una es la medición de PIB realizada por el Banco Central, disponible de manera trimestral desde marzo de 1986, y el Indicador Mensual de Actividad de la Construcción (IMACON), disponible de manera mensual desde enero de 1996. Ahora bien, la corta extensión temporal de la serie regional de despachos de cemento, restringe de manera importante las posibilidades de estimación. En primer lugar, resulta prácticamente imposible utilizar el PIB sectorial como medida de actividad regional, ya que al estar disponible sólo trimestralmente habrían disponibles solo 24 datos, lo cual estimando un número razonable de parámetros agota prácticamente los grados de libertad. En segundo lugar, si se va a utilizar IMACON como indicador de actividad agregada se debe tener en cuenta que esta medida se refiere a trimestres móviles, lo que inmediatamente implica que los indicadores parciales debieran ser considerados en trimestres móviles también.

El modelo fue ajustado de manera mensual (aunque se refiere a trimestres móviles) para el período 2001-2006⁷ y usando como indicadores de actividad regional el empleo sectorial, las solicitudes de permisos de edificación y los despachos de cemento. Además, considerando que las solicitudes de permisos de edificación en general no implican actividad inmediata se incluyeron diferentes rezagos de esta variable en la especificación del modelo, en particular, se introdujeron los rezagos 6, 12, 18 y 24 meses.⁸ Finalmente, el modelo fue estimado sin constante

⁷ Si bien se usa una frecuencia mensual para la estimación de los modelos de regresión, esto para aumentar los grados de libertad en el proceso de estimación, los indicadores regionales se calcularán solo trimestralmente, ya que en términos mensuales su volatilidad inherente reduce considerablemente la información útil que puede ser extraída de dichos indicadores.

⁸ Considerando que los permisos de edificación tienen una vigencia de 3 años a partir de su otorgamiento, se probaron además los rezagos 30 y 36 no encontrándose poder explicativo de los mismos en la estimación de la relación agregada.

y sin transformaciones⁹ y utilizando todas las series corregidas por factores estacionales. En lo que sigue utilizaremos la siguiente convención para identificar a los modelos estimados:

- Modelo 1: Modelo Base (Subsección 2.1).
- Modelo 2: Modelo con Errores Autocorrelacionados (Subsección 2.2).
- Modelo 3a: Modelo con Errores Autocorrelacionados y Heterocedásticos entre Regiones (Estimación de ponderadores en base a información del PIB sectorial regional) (Subsección 2.3).
- Modelo 3b: Modelo con Errores Autocorrelacionados y Heterocedásticos entre Regiones (Estimación de ponderadores en base al método de Máxima Verosimilitud) (Subsección 2.3).

Los resultados de la estimación del modelo base, es decir, aquel que asume que los errores no presentan ni autocorrelación ni heteroscedasticidad entre regiones se muestra en la Tabla 1. Como se puede observar en dicha tabla, en general el ajuste del modelo agregado es bueno y que los parámetros estimados son estadísticamente significativos. Nótese, sin embargo, que el estadístico Durbin-Watson¹⁰ tiende a cero, dando una señal de autocorrelación positiva en los errores estimados. De hecho, si se grafica dichos errores, Figura 2, se puede observar que su comportamiento dista mucho de ser ruido blanco, lo que pone en duda la exactitud de este supuesto. Por lo tanto, más adelante se considerará la posibilidad de que $\rho \neq 0$.

Tabla 1: Parámetros Estimados para la Especificación del Modelo 1

Variable	Parámetro	Error Est.
Empleo	0.10544	0.02682
Permisos	1.47E-05	2.66E-06
Cemento	9.13E-05	2.99E-05
Permisos (-6)	1.29E-05	2.88E-06
Permisos (-12)	1.51E-05	2.98E-06
Permisos (-18)	2.25E-06	4.52E-06
Permisos (-24)	-1.41E-06	3.99E-06
R ² ajustado	0.9380	
Akaike	4.4369	
Schwarz	4.6791	
Durbin-Watson	0.5402	

⁹ Nótese que si se usara la transformación en logaritmos, típicamente utilizada cuando se trabaja con series de tiempo, al agregar las series transformadas mediante la suma tendríamos una contradicción respecto a la definición misma de agregación que estamos utilizando.

¹⁰ El estadístico *Durbin-Watson* puede ser utilizado en este contexto debido a que no se incluyen rezagos de la variable dependiente en el lado derecho de la relación agregada.

Utilizando la ecuación (8) y los parámetros estimados en la Tabla 1 se calcularon los indicadores regionales para las trece regiones del país. La Figura 3 presenta dichos indicadores para las regiones Quinta y Metropolitana.¹¹ Como se desprende de estas muestras, el indicador de actividad construido para la región metropolitana sigue muy de cerca el desempeño del IMACON, esto debido a que esta región representa poco menos del 50% de la actividad del sector, mientras que el de la actividad en la Quinta Región muestra un desempeño mucho más variable, exhibiendo profundos altibajos (lo que probablemente está relacionado con la construcción del tren metropolitano).

Figura 2: Errores de la Relación Agregada estimada mediante el Modelo 1

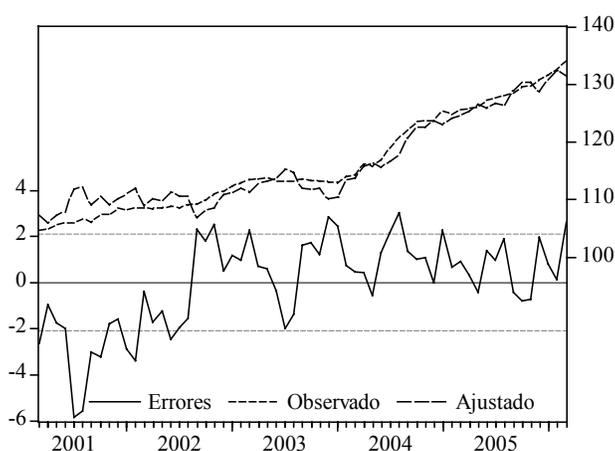
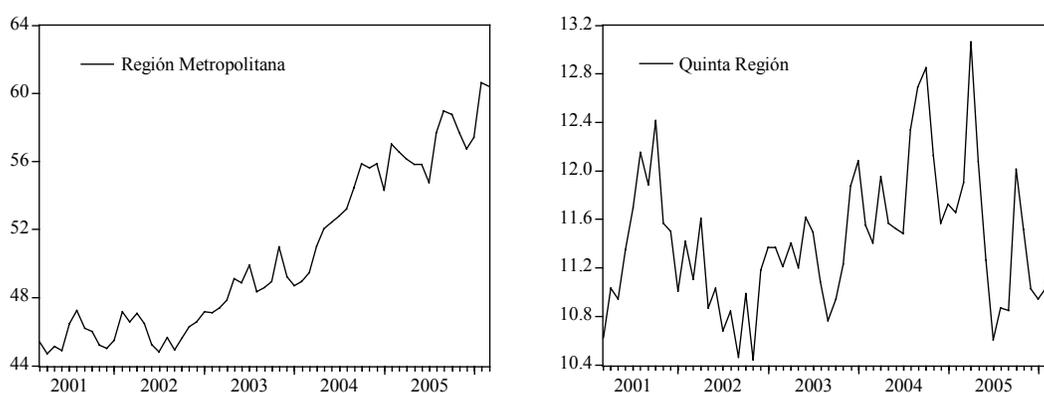


Figura 3: Indicadores Regionales para las Regiones Quinta y Metropolitana estimados mediante el Modelo 1



De la observación de los residuos en la Figura 2 se desprende claramente que el supuesto (9) no parece el más apropiado para este caso en particular. En efecto, el coeficiente de autocorrelación estimado para los residuos de la Figura 2 es 0.73 y es significativamente diferente

¹¹ Sólo se presentan los indicadores para estas dos regiones a modo de ilustración y por restricciones de espacio.

de cero. Dado lo anterior, se estimó el Modelo 2 que corrige por autocorrelación en los errores estimados y supone que dicha autocorrelación es la misma para todas las regiones. Al igual que en el caso de la Tabla 1, se estimó el modelo usando como indicadores parciales de actividad regional el empleo sectorial, las solicitudes de permisos de edificación y los despachos de cemento, todos corregidos por factores estacionales y sin transformaciones. Los parámetros estimados para dicho modelo se presentan en la Tabla 2.

Tabla 2: Parámetros Estimados para la Especificación del Modelo 2

Variable	Parámetro	Error Est.
Empleo	0.14975	0.02002
Permisos	3.06E-06	1.92E-06
Cemento	1.20E-04	2.12E-05
Permisos (-6)	4.59E-06	2.07E-06
Permisos (-12)	4.37E-06	2.06E-06
Permisos (-18)	1.46E-06	2.24E-06
Permisos (-24)	1.23E-06	2.15E-06
Rho	0.94557	0.04774
R ² ajustado		0.9850
Akaike		3.0106
Schwarz		3.2874
Durbin-Watson		1.8521

Se observa un pequeño cambio en los parámetros estimados, de hecho ninguno es negativo ahora, pero en general las magnitudes relativas se mantienen. Por supuesto, los dos parámetros que más nos interesan mejoran sustancialmente. Por una parte la varianza de los residuos disminuye de manera notable, desde 4.44 en el Modelo 1 a 1.03 en el Modelo 2, lo que mejora la calidad y precisión de las estimaciones. De la misma manera, el coeficiente de correlación que antes alcanzaba a 0.73 ahora toma un valor de 0.07 (para los errores corregidos ε), el cual está lejos de generar problemas de no estacionariedad en los residuos del modelo (Figura 4).¹² Finalmente en la Figura 5 se puede observar que los patrones de comportamiento para los indicadores en las dos regiones consideradas lucen similares a los estimados para el modelo 1. Esto es, la diferencia principal entre las dos especificaciones radica en la eficiencia de los estimadores, teniéndose por tanto una importante ganancia de predicción.

¹² Es importante destacar también que el modelo 2 muestra evidencia robusta sobre la validez del supuesto de normalidad en los residuos de la relación agregada (el estadístico Jarque-Bera tenía asociado un valor probabilidad de 0.23). En el modelo 1, por el contrario, la evidencia era más bien en contra de normalidad (el valor probabilidad del estadístico Jarque-Bera en este caso era 0.04). Así, esto brinda algunos puntos adicionales a favor de la segunda especificación.

Figura 4: Errores de la Relación Agregada estimada mediante el Modelo 2

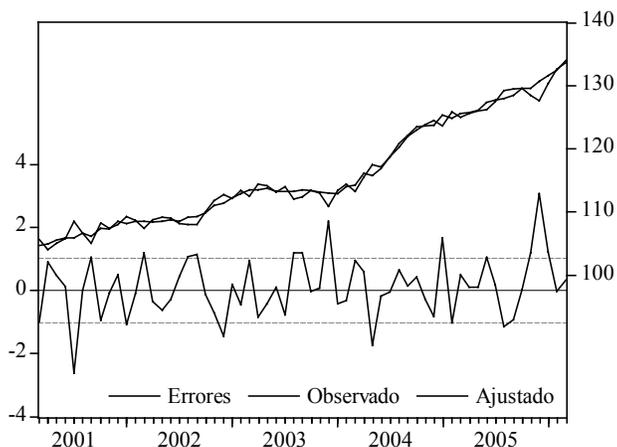
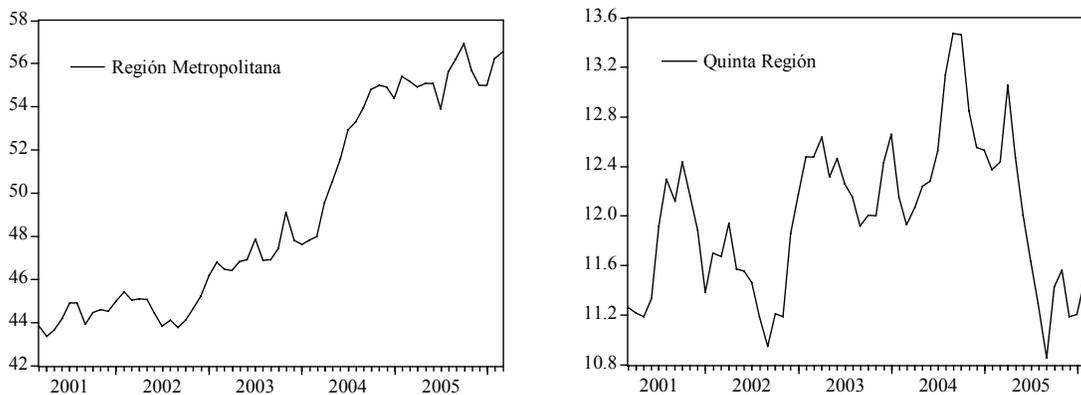


Figura 5: Indicadores Regionales para las Regiones Quinta y Metropolitana estimados mediante el Modelo 2



Como se comentó, la diferencia principal entre el modelo 1 y el modelo 2 está relacionada con las ganancias de eficiencia en la estimación, no obstante, creemos que la estimación se puede mejorar aun más si se toma en cuenta la heterogeneidad propia de las regiones en Chile. En la Tabla 3 se presentan los modelos estimados tomando en cuenta la existencia de heteroscedasticidad entre regiones.

Tabla 3: Parámetros Estimados para las Especificaciones de los Modelos 3a y 3b

Variable	Modelo 3a		Modelo 3b	
	Parámetro	Error Est.	Parámetro	Error Est.
Empleo	0.14975	0.02002	0.14975	0.02002
Permisos	3.06E-06	1.92E-06	3.06E-06	1.92E-06
Cemento	1.20E-04	2.12E-05	1.20E-04	2.12E-05
Permisos (-6)	4.59E-06	2.07E-06	4.59E-06	2.07E-06
Permisos (-12)	4.37E-06	2.06E-06	4.37E-06	2.06E-06
Permisos (-18)	1.46E-06	2.24E-06	1.46E-06	2.24E-06
Permisos (-24)	1.23E-06	2.15E-06	1.23E-06	2.15E-06
Rho	0.94557	0.04774	0.94557	0.04774
w ₁		0.07292		0.08701
w ₂		0.03015		0.13382
w ₃		0.12323		0.15134
w ₄		0.05508		0.00101
w ₅		0.02138		0.08691
w ₆		0.03108		0.10415
w ₇		0.05792		0.05386
w ₈		0.02446		0.09238
w ₉		0.04915		0.05404
w ₁₀		0.04115		0.01478
w ₁₁		0.24969		0.08347
w ₁₂		0.23800		0.03014
w _{RM}		0.00577		0.10773
R ²		0.98700		0.98700
Akaike		3.01060		3.01060
Schwarz		3.28740		3.28740
Durbin-Watson		1.85210		1.85210

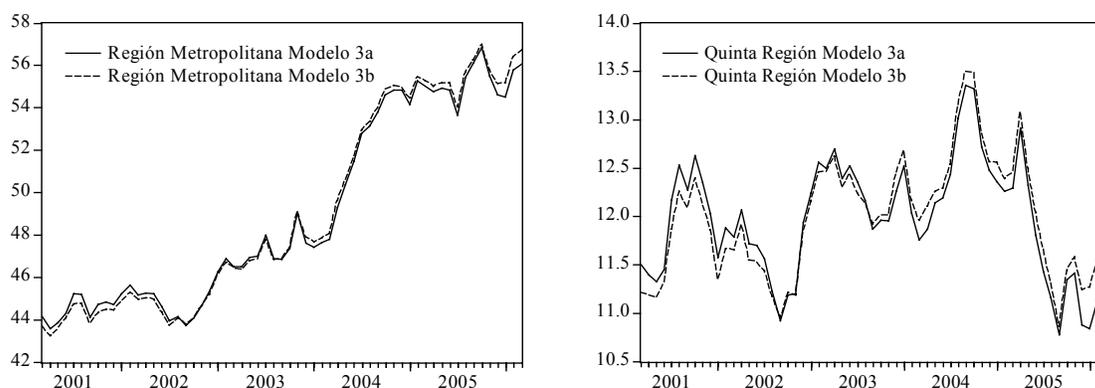
En las columnas primera y segunda se presenta el modelo cuyos ponderadores fueron estimados sobre la base de información del PIB de la construcción regional, el mismo que está disponible para el período 1996-2003. Los ponderadores fueron construidos como el inverso de la participación, promedio para todo el período mencionado, del valor agregado de cada región en el valor agregado total, debidamente escalados para que sumen uno. Lo anterior debido a que, a priori se podría pensar que mientras más pequeño sea el producto en una región determinada, más volátil debería comportarse ésta ante el inicio o término de unos pocos proyectos de construcción

importantes. Nótese también que en el modelo los parámetros asociados a la relación agregada no varían respecto de lo observado en la Tabla 2. Esto se debe a que, como se mencionó en la sección anterior, el problema de estimación se rescribió de tal manera que dicho problema se reduzca sólo a la estimación de los ponderadores y, dado que los mismos suman uno al ser agregados, la estimación de la relación agregada no se ve afectada, ni en sus parámetros ni en sus estadísticos de ajuste.

Las columnas tercera y cuarta de la Tabla 3 presentan el modelo con heteroscedasticidad entre regiones, pero ahora con los ponderadores estimados mediante el método de máximo verosimilitud (Modelo 3b). Nótese que nuevamente los parámetros de la relación agregada y los de ajuste no se ven afectados, lo cual se debe a lo ya comentado antes. La única diferencia respecto del Modelo 3a radica que ahora los ponderadores son estimados estadísticamente. Al compararse los ponderadores de los modelos 3a y 3b es destacable que estos difieren de manera importante tanto en su magnitud como en el ordenamiento relativo de las regiones según la volatilidad en sus errores, hecho que pondría en evidencia las deficiencias en el uso de la información del PIB regional para la construcción de dichos ponderadores.

A pesar de lo anterior, y siguiendo con las ilustraciones utilizadas antes,¹³ la Figura 6 muestra que los indicadores estimados para las regiones Quinta y Metropolitana no difieren sustancialmente entre los modelos 3a y 3b. El problema surge cuando se comparan regiones de pequeño tamaño relativo ya que en este caso la especificación del modelo 3a tiende a exacerbar demasiado las diferencias entre las mismas.

Figura 6: Indicadores Regionales para las Regiones Quinta y Metropolitana estimados mediante los Modelos 3a y 3b



Una vez estimadas todas las especificaciones del modelo de desagregación del IMACON en indicadores regionales queda una pregunta por responder: ¿Cuál es el mejor modelo? o ¿Cuál es el que se debería estimar para el seguimiento de la actividad del sector construcción en cada

¹³ No se presentan los errores estimados para los modelos 3a y 3b debido a que son los mismos que los mostrados en la Figura 3

región?. Estas preguntas son quizá las más difíciles de responder toda vez que no se cuenta con parámetros de comparación lo suficientemente cercanos. Lo que si se conoce es que si bien el modelo 3a es más realista, en el sentido que controla por heterogeneidad entre regiones, éste tiende a generar indicadores extremadamente volátiles para las regiones pequeña,¹⁴ no pudiéndose extraer mucha información útil en este contexto, más aún si se utiliza el indicador para realizar un seguimiento más o menos periódico del actividad. Lo anterior es menos evidente en el caso del modelo 3b que también es más realista que los modelos 1 y 2.

Si bien no se cuenta con parámetros de comparación, un ejercicio útil es comparar los indicadores regionales anualizados con el PIB de la construcción para el período con el que se cuenta de información. La Tabla 4 examina cual de los modelos genera una distribución del valor agregado regional similar la observada en el PIB de la construcción. Como se puede observar en dicha tabla, los modelos 2 y 3b son los que tienen los menores errores cuadráticos medios entre la participación regional promedio para el período 2001-2003 calculada mediante los indicadores regionales y aquella calculada mediante el PIB de la construcción regional.

Tabla 4: Comparación Indicadores Estimados y PIB de la Construcción Regional

Región	Error Cuadrático Medio			
	Modelo 1	Modelo 2	Modelo 3	Modelo 4
I	0.00016	0.00014	0.00014	0.00014
II	0.00284	0.00258	0.00249	0.00269
III	0.00001	0.00001	0.00001	0.00001
IV	0.00020	0.00023	0.00024	0.00027
V	0.00003	0.00007	0.00007	0.00007
VI	0.00111	0.00118	0.00113	0.00121
VII	0.00018	0.00023	0.00024	0.00024
VIII	0.00072	0.00089	0.00094	0.00088
IX	0.00001	0.00001	0.00001	0.00001
X	0.00004	0.00001	0.00002	0.00002
XI	0.00001	0.00002	0.00005	0.00002
XII	0.00002	0.00001	0.00002	0.00001
RM	0.00120	0.00054	0.00061	0.00051
Promedio	0.00050	0.00046	0.00046	0.00047

En resumen, dado el hecho que no existe una medida de comparación para los indicadores estimados, la decisión entre uno de los cuatro modelos para el seguimiento de la actividad de la construcción regional no sólo pasa por buscar el mejor modelo en el sentido estadístico, sino también por buscar aquel modelo que considere la realidades regionales y brinde información

¹⁴ Como lo son los casos de las regiones Onceava y Doceava.

útil en su cálculo periódico. En este sentido, creemos que el modelo 3b es aquel que cumple con estas dos condiciones. Por un lado, es el modelo que mejor ajusta la estimación de la relación agregada y también es uno de los más cercanos a la distribución regional observada en el PIB de la construcción. Por otro lado, si bien las estimaciones del modelo 3b presentan una alta volatilidad, principalmente para las regiones pequeñas lo cual es un hecho inherente a la estimación de indicadores regionales, ésta no se ve exacerbada al punto que se pierda información útil a la hora de hacer seguimiento a la actividad del sector en cada región como sucedía con el modelo 3a.

4. Comentarios Finales

El sector construcción es uno de los sectores más importantes de la economía chilena, no sólo a nivel nacional sino que también a nivel regional. No obstante, y a pesar de dicha importancia, no se cuenta con información pública y actualizada sobre la actividad de la construcción en cada región. Así, este trabajo trata de proveer de una herramienta estadística que permita hacer un seguimiento al pulso de la actividad de la construcción a nivel regional. En particular, se busca un modelo estadístico para la estimación de indicadores regionales a partir de la desagregación del Índice Mensual de Actividad de la Construcción (IMACON).

Aquí utilizamos una versión modificada del modelo propuesto originalmente por Chow y Lin (1971), el cual fue creado con el objetivo de aumentar la frecuencia de una serie de tiempo, por ejemplo, transformando una serie trimestral en una mensual. En este sentido, se puede decir que esta metodología desagrega una observación temporal en varias observaciones temporales según sea la transformación de frecuencia que sea elegida. Para dicha desagregación estos autores utilizan variables que sirven de instrumentos en la estimación de la variable en mayor frecuencia que es no observada. Estos instrumentos deben ser variables que se espera tenga una alta correlación con la variable que se desea estimar y que sean observados en la frecuencia elegida para la transformación. Así y siguiendo este mismo espíritu, el modelo que desarrolla este trabajo busca la desagregación de series nacionales en series regionales, donde se busca transformar o *desagregar* un dato nacional en datos regionales usando para ello un conjunto de instrumentos observados en cada región.

Se construyeron cuatro modelos, clasificados según los supuestos asociados a la matriz de varianzas y covarianzas de los errores de estimación. En el primero se realizaron los supuestos tradicionales en econometría, es decir, homoscedasticidad o varianzas iguales entre regiones y no autocorrelación en los errores. La segunda especificación toma en cuenta el problema de autocorrelación típicamente presente cuando se trabaja con series de tiempo. Las especificaciones tercera y cuarta controlan, además, por la heterogeneidad inherente al uso de variables a nivel regional, y en lo único en que difieren es en la forma que estiman las varianzas asociadas a los errores de cada región.

De la observación de dichos modelos se encuentra que mientras más realista se torna el modelo, es decir, el modelo que controla por autocorrelación y heteroscedasticidad entre regiones, más volátil se vuelve el indicador estimado. Lo anterior es particularmente cierto para el caso en que se usa información del PIB sectorial regional para la construcción de los ponderadores en la matriz de varianzas y covarianzas. Para el caso de regiones pequeñas dicha volatilidad se ve exacerbada a tal punto que es difícil extraer información útil de los indicadores estimados, hecho que claramente como medida periódica de pulso de actividad le resta puntos a ésta especificación. Lo anterior tiende a corregirse cuando los ponderadores son estimados por métodos de máxima verosimilitud. Por otro lado, se comparó los indicadores regionales estimados y anualizados con el PIB de la construcción, para el período 2001-2003, y se encontró que los modelos que imponen sólo autocorrelación y autocorrelación y heteroscedasticidad, éste último basado en ponderadores estimados por máxima verosimilitud, son los que se asemejan más a la distribución regional del valor agregado observada en el PIB de la construcción.

Así, dado el hecho que no existe una medida de comparación lo suficientemente cercana para los indicadores estimados, la decisión entre uno de los cuatro modelos no sólo pasa por buscar el mejor modelo en el sentido estadístico, sino también por buscar aquel modelo que mejor capture la realidad regional y que brinde información útil en su cálculo periódico. En este sentido, creemos que el modelo 3b es aquel que cumple con estas dos condiciones. Por un lado, es el modelo que mejor ajusta la estimación de la relación agregada y también es uno de los más cercanos a la distribución regional observada en el PIB de la construcción. Por otro lado, si bien las estimaciones del modelo 3a presentan una alta volatilidad, siendo en este sentido realista, ésta no se ve exacerbada al punto que se pierda información útil a la hora de hacer seguimiento a la actividad del sector en cada región.

Es claro que esta metodología podría ser mejorada en el futuro en la medida en que se cuente con mayor información sobre el estado coyuntural del sector. Esto no sólo es cierto para mayores observaciones temporales sino también para una mayor cantidad indicadores parciales de actividad sectorial en cada región. En primer lugar, aquí no se tomó en cuenta la posibilidad de que los errores de regresión estén correlacionados entre regiones, al menos en aquellas que sean geográficamente cercanas. Lo anterior requiere de la estimación de un número mayor de parámetros, los elementos fuera de la diagonal principal en la matriz de varianzas y covarianzas, lo cual sin duda requiere de mayores grados de libertad para hacer confiables a las estimaciones obtenidas y, por tanto, mayores observaciones temporales de cada uno de los indicadores. En segundo lugar, debido a limitaciones en la información de indicadores parciales aquí sólo usamos tres de ellos, el empleo sectorial, los permisos de edificación y el consumo de cemento en cada región; no obstante, si se pudieran construir otros indicadores, como por ejemplo aquellos asociados al gasto público en vivienda e infraestructura en cada región, sin duda las estimaciones de los índices mejoraría considerablemente. Estos puntos quedan como agenda de investigación futura.

Referencias

- [1] Chow, Gregory y An-loh Lin (1971): Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series, *The Review of Economics and Statistics*, Vol. 53, No. 4, pp. 372-7.
- [2] Grenne, William (2003): *Econometric Analysis*, Prentice Hall, 5th Edition.
- [3] Johnston, J. y J. Dinardo (1997), *Econometric Methods*, McGraw-Hill, 4th Edition.
- [4] Piguillem, Facundo (2004): Un Indicador Mensual de Actividad de la Construcción, *Cámara Chilena de la Construcción*, Documento de Trabajo No. 20.
- [5] Tejada, Mauricio (2006): Índice Mensual de Actividad de la Construcción: Un Estimador en Frecuencias Mixtas, *Cámara Chilena de la Construcción*, Documento de Trabajo No. 34.
- [6] Venanzoni, Giuseppe y Francesco Felici (2002): Regional Disaggregation of National Data with Instrumental Variables, *Mimeo*.

Apéndice: Indicadores de Actividad de la Construcción Regional en Trimestres - Modelo 3b

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	RM
I-01	2.57	4.41	1.91	4.53	11.22	4.39	6.32	11.97	5.50	6.16	0.51	1.48	43.70
II-01	2.90	4.88	2.46	4.88	11.31	4.73	6.06	10.94	5.47	6.36	0.48	1.37	44.10
III-01	2.54	4.90	2.29	4.80	12.09	5.31	5.81	11.89	4.94	6.02	0.45	1.26	43.83
IV-01	2.65	5.70	2.55	4.71	11.86	5.55	5.73	12.45	4.65	6.37	0.64	1.23	44.46
I-02	2.48	5.05	2.57	5.30	11.66	5.94	5.59	12.57	4.95	5.88	0.56	1.03	44.98
II-02	2.44	4.90	2.35	4.89	11.53	6.19	5.44	14.37	4.85	5.98	0.51	0.95	44.36
III-02	2.78	5.70	2.99	4.93	10.95	6.14	5.22	13.96	4.74	6.17	0.76	1.11	43.78
IV-02	2.58	5.63	2.65	4.88	11.84	6.31	5.83	14.50	4.96	5.57	0.49	0.94	45.19
I-03	2.86	6.19	2.76	4.94	12.47	5.71	5.42	14.58	4.41	6.17	0.52	1.02	46.45
II-03	2.68	6.60	2.44	4.68	12.45	5.71	5.78	13.95	4.05	6.33	0.61	1.06	46.89
III-03	3.02	6.48	2.71	4.77	11.93	5.78	5.56	14.48	3.95	5.99	0.71	1.28	46.93
IV-03	3.17	6.52	2.93	4.51	12.46	5.83	5.51	11.58	4.31	6.29	0.91	1.14	47.91
I-04	3.02	6.78	3.16	4.48	11.96	6.16	6.12	12.04	4.08	6.32	0.88	1.13	48.10
II-04	2.76	6.62	3.10	4.99	12.30	6.36	5.90	11.29	4.17	5.92	0.80	1.06	51.64
III-04	3.07	6.54	2.97	4.77	13.50	6.62	6.66	11.32	4.53	6.20	0.78	1.09	54.02
IV-04	3.11	6.67	2.91	4.75	12.57	7.06	6.59	12.68	4.29	6.34	0.67	1.12	54.98
I-05	3.22	7.31	3.16	4.62	12.46	6.94	6.63	12.74	5.04	6.33	0.82	0.99	55.28
II-05	3.45	6.95	3.58	4.70	12.04	6.72	6.82	13.81	5.68	6.32	0.81	1.22	55.19
III-05	3.01	7.50	3.39	4.90	10.87	6.80	6.39	14.93	5.48	7.14	0.68	1.13	56.26
IV-05	3.20	8.69	4.04	5.32	11.25	7.43	6.89	14.37	5.31	6.77	1.06	1.23	55.17
I-06	3.52	7.41	3.97	5.45	12.04	7.32	7.58	14.20	6.07	7.06	1.21	1.55	56.74