

Ilustración de las técnicas econométricas: Análisis y predicción del sector industrial

Illustration of econometric techniques: Analysis and prediction of the industrial sector

José Salinas Ortiz¹

RESUMEN

Este artículo describe y discute los conceptos fundamentales del Análisis Econométrico usando una aplicación concreta que permite ilustrar el uso adecuado de las técnicas econométricas a los estudiantes de econometría, y en general a las personas interesadas en usar estas técnicas. El periodo de estimación de la aplicación abarca 16 años, en el cual el autor usa las técnicas citadas para analizar el comportamiento del sector manufacturero. Para ilustrar los conceptos y uso de la econometría en la estimación y análisis se usa la técnica de “Pasos Sucesivos” que permite presentar los diferentes estadísticos que indican el grado de éxito alcanzado con cada una de las ecuaciones estimadas, y los principales problemas del análisis de regresión (multicolinealidad, autocorrelación). Además, se presentan criterios para elegir la mejor entre las ecuaciones estimadas. Finalmente, se enfatiza que el investigador debe complementar las técnicas econométricas con sus conocimientos teóricos y su experiencia o entendimientos del fenómeno que busca explicar.

ABSTRACT

This article describes and discusses the fundamental concepts of Econometric Analysis using a concrete application that allows illustrating the proper use of econometric techniques to econometrics students, and in general to people interested in using these techniques. The estimation period of the application covers 16 years, in which the author uses the aforementioned techniques to analyze the behavior of the manufacturing sector. To illustrate the concepts and use of econometrics in estimation and analysis, the “Stepwise” technique is used, which allows presenting the different statistics that indicate the degree of success achieved with each of the estimated equations, and the main problems of the analysis regression (multicollinearity, autocorrelation). In addition, criteria are presented to choose the best among the estimated equations. Finally, it is emphasized that

1 Universidad Nacional de Ingeniería – UNI (Perú)

the researcher must complement the econometric techniques with his theoretical knowledge and his experience or understanding of the phenomenon he seeks to explain.

1. INTRODUCCIÓN

El presente trabajo surgió como un requerimiento de corto plazo, y muy preliminar para establecer tasas de crecimiento factibles para el sector Manufacturero en el corto plazo (1977), dada la coyuntura económica – financiera en 1976, en el contexto de los estudios de la Dirección General de Asuntos Financieros del Ministerio de Economía y Finanzas.

El planteamiento de las hipótesis presentadas en este trabajo no se basó en una investigación cualitativa rigurosa, que permitiera tener un conocimiento preciso del comportamiento del sector en análisis. Sin embargo, las hipótesis aquí planteadas se desprenden de los estudios e ideas bastante generalizadas entre los investigadores del Sector Manufacturero. (GIECO - UNI, 1977)

Posteriormente, se planteó la posibilidad de utilizar los resultados de este trabajo para ilustrar el uso de las técnicas econométricas y poder ser usado por los estudiantes de Econometría de diferentes Universidades en Perú.

A pesar del carácter preliminar de este trabajo, se intenta cubrir todos los pasos de una investigación econométrica, es decir: formulación, estimación y sometimiento a prueba de hipótesis de comportamiento del Producto Nacional de Manufacturas (PNSMR). Es así, que en la sección 2 se hace un análisis del periodo muestral (1960 – 1976), con el fin de tener una descripción somera de las posibles relaciones causales para explicar el comportamiento del PNSMR; seguido de la especificación y explicación de las técnicas de regresión requeridas para nuestro análisis. Las estimaciones y resultados son presentados y discutidos en una sección aparte, a la luz de los estadísticos adecuados. Finalmente, se trata de determinar las posibilidades de crecimiento del PNBMR para el futuro inmediato.

2. PERIODO DE ESTIMACIÓN Y SERIES ESTADÍSTICAS

El periodo de estimación abarca desde 1960 a 1976. A partir de 1960 la Economía Peruana inicia un periodo de transición, pasando de ser agro – minera exportadora para convertirse en una economía industrial sustentada en importaciones y por lo tanto con escasa integración económica. Lo que se aprecia en el cuadro siguiente:

Cuadro N° 1: Participación sectorial en el PBI (En porcentaje)

	1960	1970	1976
Agropecuario	10	15.1	12.7
Minería	10	8.2	6.8
Manufactura	20	23.8	25.9

Esta idea nos llevó a plantear la hipótesis de que el comportamiento de la producción manufacturera estuvo íntimamente relacionada a las importaciones de productos intermedios para la industria en el periodo 1960 – 1976 (lo que se expresa en el coeficiente de correlación simple entre estas variables que fue del 94%). Hipótesis que es corroborada al analizar la evolución del PNB de Manufacturas: entre 1960 y 1966 la producción industrial creció a tasas mayores de 9.0%, lo que determino que se produjese a partir de 1965 un fuerte deterioro de la Balanza de Pagos en Cuenta Corriente. Esta situación se agravó en 1966 y 1967, año en que se devaluó el sol para frenar las importaciones. La restricción de importaciones en 1968 y 1969 determinó, junto con otros factores, la disminución del crecimiento del PNB de Manufacturas a menos del 2% en 1969 (BCR, 1970).

A partir de 1970 y hasta 1974 el sector de manufacturas vuelve a crecer a altas tasas (8% promedio anual), debido a la disponibilidad de divisas para importar dada la cíclica bonanza del mercado internacional, y al tipo de cambio subvaluado tendiente a incentivar las importaciones para la industria.

A partir de 1974 surgen nuevamente problemas de balanza de pagos (una Balanza Comercial negativa), limitándose por lo tanto el crecimiento de las importaciones intermedias para la industria a partir de 1975. Esto junto a otros factores limitó el crecimiento de la Producción Industrial (a 4.7% en 1975 y 4.2% en 1976) pero siempre manteniéndose por encima del crecimiento de PBI global.

Entre los otros posibles factores que explican el crecimiento del sector fabril estarían: el crédito interno al sector, el ingreso por remuneraciones más el ingreso de independientes (factor demanda), los gastos del Gobierno Central en bienes y servicios y en la formación bruta del capital.

Se decidió incluir el Crédito Interno al Sector Industrial, en cuanto éste permitió disponer de los medios financieros para implementar diferentes proyectos de formación, ampliación, modernización, etc. de empresas. Se toma los ingresos por remuneraciones más el de independientes, como factor de demanda; lo interesante hubiera sido poder cuantificar los efectos de la redistribución

de ingresos que se ha producido en los últimos años, sin embargo, se puede decir que durante todo el periodo muestral se dio una continua ampliación de mercado interno para productos manufacturados, debido al proceso de urbanización acelerado.

También se tomó los gastos del Gobierno Central en bienes, servicios y formación bruta de capital, como otro factor de demanda, que dada la política de mayor participación del Estado en la actividad económica, tuvo un fuerte dinamismo en los últimos años del periodo muestral (se triplicó entre 1968 y 1975).

3. MÉTODO DE ESTIMACIÓN

Nuestro interés es examinar cuánto de las variaciones en el Producto Nacional de Manufactura Real (PNBMR) es influencia de las variaciones en las variables elegidas como relevantes para explicarla (ej.: importaciones intermedias para industrias, MIPIR). El coeficiente de correlación de Pearson (o coeficiente de correlación simple), nos da una primera idea de tal influencia (este coeficiente entre el PNBMR y MIPIR es de 94%), pero, como es bien sabido un coeficiente de correlación alto puede ser simplemente el resultado de la influencia común de una tercera variable (ej.: el nivel de ingresos por Remuneraciones más el ingreso de independientes, YRIND). En este caso será necesario examinar la “correlación parcial” entre el PNBMR y las MIPIR, manteniendo constante el valor, y por tanto la influencia del YRIND. Una generalización importante del coeficiente de correlación parcial es el coeficiente de regresión:

Este no solo indica el grado de asociación entre el MIPIR y la PNBMR, manteniendo la influencia de otras variables controladas, sino que nos indica en cuanto la primera influye en la última. Es decir, que los coeficientes en una ecuación de regresión múltiple son medidas con “*ceteris paribus*”, y no indicadores simplemente de la relación causal (Johnson, 1972).

Cuadro N°2: Series estadísticas
(Millones de soles de 1973)

Fuentes:

(1) 1960 – 74: Banco Central de Reserva del Perú.
1975 – 76: Instituto Nacional de Estadística
Estimados Preliminares.

(2) Superintendencia de Banca y seguros. Series de deflactada con índice del PNB. Cubre los créditos del Banco Industrial del Perú, Bancos Comerciales y Financieros.

El carácter preliminar de la conceptualización del Producto Nacional de Manufacturas, adoptado aquí, sólo necesita de la estimación de una ecuación de regresión; olvidándonos del carácter de simultaneidad de las relaciones económicas. Las variables explicativas (regresores) definitivas serán elegidas del conjunto de variables señaladas en la sección anterior, en base a una variedad de criterios teóricos y estadísticos. La ecuación es estimada por el Método de Mínimos Cuadrados Ordinarios (MCO), el cual daría los mejores Estimadores Lineales Insesgados (MELI, es decir de mínima varianza), bajo ciertos supuestos razonables en la naturaleza de la especificación y de los términos perturbadores de nuestra especificación estadística (Johnson, 1972). Pero en nuestro caso muchas de las variables elegidas (ver sección anterior), son más bien endógenas, rompiendo uno de los supuestos claves de MCO, por lo que deberíamos recurrir a otros métodos que consideren el problema de simultaneidad

en nuestro sistema económico. Esto nos obligaría a formular ecuaciones de comportamiento (y/o identidades) para las variables explicativas, pero dado que este trabajo es preliminar y no busca tener una explicación integral del Sistema Económico Nacional es que no usamos uno de los métodos alternativos (ej.: Método de Variables Instrumentales, Bietápicos, Trietápicos, etc.), sin embargo, este punto debe ser considerado para estudios futuros.

Se usa la técnica de “Pasos Sucesivos” (Stepwise), dado que no existe (a este nivel preliminar del estudio), una hipótesis clara del comportamiento del PNB-MR. Esta técnica comprende una serie de regresiones múltiples lineales, con un regresor más en cada paso. La variable añadida es la que hace la mayor reducción en la suma de residuos al cuadrado (s.r.c.), dado que la minimización de esta s.r.c. es el objetivo de la estimación por Mínimos Cuadrados (MC); o equivalentemente podemos decir que entrará la variable con la mayor correlación parcial con la variable dependiente “parcializada” en las variables que ya ingresaron a la ecuación de regresión. Algunas variables adicionales pueden ser forzadas a entrar o salir de una ecuación en base a consideraciones empíricas. Las variables no forzadas son sacadas, o no incluidas en la ecuación de regresión si la reducción proporcional en la s.r.c. como consecuencia de su inclusión no es mayor que un nivel fijado con anterioridad (por ejemplo, si la s.r.c. no disminuye en proporción del 2%).

Luego, los criterios para la inclusión de variables exógenas como regresores en la ecuación son:

- i. El cambio en el poder de explicación de la regresión, indicado por la suma de residuos al cuadrado (s.r.c.), es significativo.
- ii. La presencia de la variable, de su signo y la magnitud de su coeficiente estimado están en concordancia con consideraciones a priori (teóricas y/o empíricas). Desafortunadamente no hay un estudio anterior de este tipo sobre el comportamiento del Sector Manufacturero en el Perú. Sin embargo, de lo expuesto en la sección 2 de este trabajo, podemos decir que el PNBMR está altamente influenciado por la posibilidad de importaciones de bienes intermedios para el sector en estudio (MIPIR) lo que lleva a forzar la inclusión de la variable MIPIR en la ecuación de regresión. Otro factor que refuerza su inclusión es que, en el corto plazo, el cuello de botella para el crecimiento del Sector Manufacturero va a ser la disponibilidad de divisas para atender sus demandas de importaciones, dada la coyuntura actual de la Economía Peruana. Es así que forzamos la inclusión de MIIT, con signo positivo (ver la próxima sección).

iii. Otro criterio usado fue el test de la hipótesis nula (H_0) que el coeficiente de regresión individual era cero, implicando que el regresor respectivo no debería ser incluido. El estadístico relevante para éste, es el estadístico “t”, que bajo la H_0 planteado, resulta ser igual al cociente del coeficiente estimado y su propia desviación estándar, distribuido como la “t” de Student con grados de libertad igual al número de observaciones menos el número de regresores incluidos.

4. ESTIMACIÓN Y ANÁLISIS

Para la estimación usamos la técnica de “Pasos Sucesivos”. Siempre con el fin de ilustrar el uso de las técnicas econométricas vamos a considerar dos conjuntos de estimaciones: el primer conjunto sin forzar a ninguna variable a que entre en nuestra ecuación de regresión, es decir que vamos a usar solamente los criterios (i) y (iii) enunciadas en la sección anterior. En el segundo conjunto, dado lo expuesto en la sección (2) vamos a forzar la entrada de la variable explicativa MIPIR.

Además, con el fin de ilustrar uno de los principales problemas del análisis de regresión (multicolinealidad), se estimará el PNBMR por MCO directos en función de MIPIR y YRIND (que son las variables explicativas que tienen el más alto coeficiente de correlación, ver el cuadro 2).

4.1 REGRESIÓN POR PASOS SUCESIVOS, SIN FORZAR NINGUNA VARIABLE

De una simple observación de la matriz de correlación entre nuestras variables (Cuadro N°2), vemos que el PNBMR está más altamente correlacionado con el YRIND (97.9%), luego con el GCOFK (96.7%), el CRIIR (96.1%) y por último con MIPIR (94.1%). Con esto tenemos una primera idea del grado de relación.

Cuadro N°3: Matriz de correlaciones simples

	PNBMR	MIPIR	YRIND	GCOFK	CRIIR
PNBMR	1.000000	0.940812	0.973861	0.967570	0.960701
MIPIR	0.940812	1.000000	0.965336	0.895234	0.857656
YRIND	0.978861	0.965336	1.000000	0.926101	0.935652
GCOFK	0.967570	0.895264	0.926101	1.000000	0.942837
CRIIR	0.960701	0.857656	0.935652	0.942837	1.000000

Luego, si en nuestra regresión por Pasos Sucesivos, usamos únicamente el criterio (i), vemos que la primera ecuación de regresión obtenida es:

$$(1) \quad \text{PNBMR} = -6312.96 + 0.395 \text{ YRIND}$$

(t= -4.27) (t=18.53)

$$R^2 = 0.955; F(1,15) = 348.6; s=1422,7; d=1.03$$

La ecuación (1), indica que se ha estimado el PNBMR como función, únicamente del YRIND, por MCO. Expresa que el PNBMR es igual a una constante de -6,313.0 más un coeficiente estimado de 0.395 por el YRIND. El análisis por Mínimos Cuadrados produce varios estadísticos, que indican el grado de éxito alcanzado con la ecuación estimada. Así tenemos:

-El Estadístico R², que nos indica la proporción del cambio en la variable dependiente (PNBMR) que es explicada por cambios en las variables explicativas (en este caso una sola, YRIND).

En el caso que existe una explicación perfecta, el R² será 1.0 Generalmente, la ecuación de regresión será mejor cuando R² sea mayor.

En la ecuación (1), el 95.5% de los cambios en PNBMR (en el periodo 1960 – 1976), ha sido explicado por una constante y cambios en YRIND. Hay otros indicadores del grado de éxito del ajuste de Mínimos Cuadrados, los cuales deben ser observados, aparte del R².

-Error Estándar de la Regresión “s”. Este estadístico mide el grado de proximidad que han tenido los “valores estimados”² con respecto a los “valores reales observados” en el periodo muestral (o de análisis). Luego será deseable tener un “s” lo más pequeño posible. También este estadístico puede ser usado para lograr alguna idea sobre la confianza que se puede esperar de las proyecciones.

- El estadístico “t”. - El estadístico t muestra la relevancia de cada variable explicativa en predecir la variable dependiente (Johnston, 1973). Es deseable tener un “t” lo más grande (positivo o negativo) posible para cada variable explicativa. Generalmente, cualquier “t” mayor que +2, o menor que -2 es aceptable³. Si el “t” está entre +2 y -2 para una variable explicativa, eso indica que esa variable no contribuye significativamente a explicar la variable dependiente, y puede ser sacada de la regresión sin disminuir sustancialmente el R² o aumentar el “s”.

2 Los “Valores Estimados” son las estimaciones de la variable dependiente (PNBMR) obtenidos a través de usar los valores históricos de las variables explicativas en la ecuación de regresión

3 Sin embargo, existen variables para estos valores críticos dependiendo de los grados de libertad (número de observaciones menos el número de variables explicativas de la ecuación) y el nivel de confianza requerido. Véase la tabla de la distribución “t Student”, en cualquier libro de estadística o econometría (ej.: J. Johnston, op.cit. pág. 453)

En la ecuación (1) el “t” el YRIND (19.6) y la constante (-4,3) están fuera del intervalo (-2, +2), lo que indica que ambos contribuyen significativamente a explicar el PNBMR.

-El estadístico “F”. - Este estadístico muestra la relevancia del conjunto de variables explicativas, excluyendo la constante, para predecir el comportamiento de la variable dependiente.

Es decir, que mientras los “t” indican la importancia de cada variable en forma individual, el estadístico “F” da una medida de la relevancia de todas las variables incluidas en el lado derecho de nuestra ecuación. Lo adecuado será tener un “F” lo más grande posible, pero a diferencia del estadístico “t” sus valores críticos varían mucho con el número de observaciones y el número de variables explicativas (Wonnacott, 1970) que determinarán los grados de libertad para la distribución “F” – Fisher. Así en nuestra ecuación (1), tenemos: $F(1,15)=343.6$, donde 1 nos indica los grados de libertad del numerador, que es igual al número de variables explicativas excluida la constante, que en este caso es igual a 1; mientras que el 15 nos indica los grados de libertad del denominador (igual al número de observaciones menos el número de variables explicativas incluida la constante)⁴. Si comparamos este valor con el dado en la tabla de distribución “F” – Fisher, vemos que es mucho más grande ($343.6 \geq F=8.68$), con un nivel de confianza del 99%, lo que nos lleva a rechazar la hipótesis nula (H_0) de que la variable “PNBMR no depende de YRIND”.

- El Estadístico de Durbin – Watson “d”. – el estadístico de Durbin – Watson da un test estándar para detectar autocorrelación en los errores o perturbaciones. La autocorrelación es uno de los mayores problemas que se presentan en el análisis de regresiones⁵, dado que uno de los supuestos claves del método de MCO es que las perturbaciones son independientes entre sí. Es decir, que el conocimiento del error en un periodo no debe ayudar a predecir el error en el próximo periodo.

Generalmente, cuando el estadístico “d” está entre 1.5 y 2.5 no existe autocorrelación sería en la ecuación de regresión⁶. Si el estadístico “d” está por debajo de 1.5, es muy probable que existe autocorrelación positiva entre los errores. En este caso, un error positivo en una observación

4 La distribución F – Fisher, es el cociente de dos variables independientes Chi cuadrado modificadas, por los grados de libertad, que se distribuyen independientemente.

5 Otro gran problema es el de colinealidad entre las variables explicativas, conocido como el problema de “Multicolinealidad”. Ver el párrafo 4.8 de este trabajo.

6 Al igual que los estadísticos “t” y “F”, el estadístico “d” varía con el número de observaciones y variables explicativas incluidas en la ecuación.

usualmente implicará un error positivo en la siguiente observación. Si el “d” está por encima de 2.5, indica posible autocorrección negativa, y los errores se alternarán entre positivos y negativos.

En la ecuación (1), el “d” es 1.03 indicando autocorrelación positiva. Si esta ecuación fuera usada para predecir el PNBMR para 1977, el estimado resultante será probablemente bajo, debido a la autocorrelación positiva y la subestimación del PNBMR para 1976 (el error para este año con la ecuación (1), es de 4,888).

En caso de estar seguros que la ecuación (1) es la hipótesis más adecuada para explicar el comportamiento del PNBMR, en el periodo 1960 – 1976, se debe abandonar el método de MCO, y recurrir a otro método que incluyen en la estimación la existencia de autocorrelación en las perturbaciones (así, tenemos el método de Mínimos Cuadrados Generalizados, MCG). Pero, usualmente la presencia de autocorrelación indica que hay una importante parte de la variación de la variable dependiente que no ha sido explicada. La mejor solución para este problema será buscar otra variable explicativa para incluirla en la ecuación de regresión. En nuestro caso, dado que estamos usando la técnica de regresión por Pasos Sucesivos, pasamos al “Segundo Paso”, y vemos cuál de las variables restantes entra en nuestra ecuación, debido a que hace la mayor reducción en la suma de residuos al cuadrado (ya que sólo usamos este criterio), así:

$$(2) \text{ PNBMR} = -624.94 + 0.235 \text{ YRIND} + 1.001 \text{ GCOFK}$$

(6.56) (4.84)

$$R^2 = 0.982; F(2,14) = 440,29; s=900,72; d=2.02$$

Es decir, que la inclusión de GCOFK en la ecuación elimina la autocorrelación en los errores (d=2.02). Además de analizar el resto de indicadores vemos que se ha ganado en explicación (el R² sube de 95.5% a 98.2%, el “s” disminuye de 1,423 a 901) y que las variables resultan todas significativas.

En el segundo Paso se logró reducir la s.r.c. en una proporción del 2.6%. Pero no siempre se puede encontrar variables adicionales para eliminar (o reducir) la autocorrelación, entonces debemos incluir el proceso de autocorrelación en nuestro método de estimación (MCG), logrando proyecciones adecuadas.

Si en el proceso de estimación de Pasos Sucesivos no se impone ninguna restricción para la inclusión de nuevas variables explicativas (número de

pasos y/o un nivel mínimo de reducción proporcional en la s.r.c.) se tendrán las siguientes ecuaciones:

Tercer Paso:

$$(3) \text{ PNBMR} = -114.57 + 0.21 \text{ YRIND} + 0.82 \text{ GCOFK} + 0.85 \text{ CRIIR}$$

(5.19) (3.32) (1.23)

$$R^2 = 0.983; F(3,13) = 307.38; s = 880,97; d = 1.68$$

Cuarto Paso:

$$(4) \text{ PNBMR} = 113.13 + 0.19 \text{ YRIND} + 0.79 \text{ GCOFK} + 0.41 \text{ CRIIR} + 0.15 \text{ MIPIR}$$

(2.07) (2.79) (1.15) (0.29)

$$R^2 = 0.983; F(3,13) = 307.38; s = 880,97; d = 1.68$$

En el tercer paso, el crédito interno para industria (CRIIR) es incluida en nuestra ecuación de regresión, pero resulta ser no significativa (el estadístico “t” de dicha variable es muy pequeño, $1.28 < 2.00$), es decir, que aceptamos la hipótesis nula de que el coeficiente de regresión de la variable CRIIR debe ser cero. Lo que es confirmado a través de observar los otros indicadores de la ecuación (3) y compararlos con la ecuación (2); la proporción de s.r.c. reducida en este paso es de sólo 0.2%; el nivel de explicación, prácticamente no ha aumentado (el R2 ha pasado de 98.2% a 98.3%, el “s” disminuyó de 900.7 a 881). Luego la ecuación (3) no resulta ser mejor que la ecuación (2).

Si no se consideran los resultados de la ecuación (3) y se procede a incluir una variable adicional, MIPIR, se obtiene el paso N°4, que en términos de explicación resulta ser más pobre que la ecuación (3), e incluso la (2). Es decir, que la inclusión de MIPIR, en vez de mejorar el nivel de explicación de la variable dependiente, PNBMR, lo disminuye; el R2 disminuye a 98.1%; la desviación estándar (s) aumenta; y el estadístico “d” disminuye aún más de lo hecho en la ecuación (3). Si analizamos los estadísticos “t”, para ver el grado de significación de cada variable en forma individual vemos que ni CRIIR ni MIPIR, resultan ser significativos.

Luego, la conclusión de este proceso de estimación por Pasos Sucesivos, basados únicamente en los criterios de reducción en la suma de residuos al cuadrado (s.r.c.) y en los estadísticos “t”, llevan a aceptar la ecuación (2) como la más adecuada, para explicar el comportamiento PNBMR, para el periodo 1960 – 1976. Pero, dado lo expuesto en la sección (2), de este trabajo, vemos que un factor determinante, en el periodo en estudio del comportamiento del PNBMR resulta ser las importaciones de bienes

intermedios y materias primas para la industria (MIPIR); y que además dada la coyuntura actual de la Economía Peruana (escasez de divisas), ese factor toma vigencia para explicar el comportamiento inmediato futuro del PNBMR. Es así que se optó por hacer otro proceso de regresiones por Pasos Sucesivos, forzando la inclusión de MIPIR, con signo positivo.

4.2 REGRESIÓN POR PASOS SUCESIVOS, FORZANDO LA INCLUSIÓN DEL MIPIR:

En este proceso se fuerza la inclusión de la variable MIPIR, en base al criterio (ii) (ver sección 3). Es decir, basándose en consideraciones a priori. Además, la forzamos a que entre con signo positivo.

Luego el 1er paso de las regresiones por Pasos Sucesivos será:

$$(5) \quad \text{PNBMR} = -1616.42 + 3.005 \text{ MIPIR}$$

$$\quad \quad \quad (-0.76) \quad \quad (10.75)$$

$$R^2 = 0.877; F(1,15); s = 2357.55; d=1.47$$

Es decir, que la inclusión de la variable MIPIR, como único regresor, si bien es relevante en forma individual ($t=10.75$); logra explicar menos las variaciones del PNBMR, que las ecuaciones anteriores (comparando con la ecuación (1): el R^2 ha disminuido de 95.5% a 87.7%; el “F” de 344 a 116; el “s” aumentó de 1423 a 2357). Sin embargo, esta ecuación lleva a rechazar la hipótesis de existencia de autocorrelación en los errores⁷, luego si se desea mejorar el nivel de explicación de la variable dependiente permitamos el ingreso de otra variable explicativa usando el criterio (i):

Segundo Paso:

$$(6) \quad \text{PNBMR} = 1397.20 + 1.41 \text{ MIPIR} + 1.41 \text{ CRIIR}$$

$$\quad \quad \quad (5.33) \quad \quad (7.02)$$

$$R^2 = 0.971; F(2,14)=268.51; s = 1147.67; d=1.56$$

La inclusión de la variable CRIIR en la ecuación de regresión logra reducir la s.r.c. en una proporción del 8.9%; lo que se refleja en el resto de indicadores del grado de explicación: el R^2 ha aumentado de 87.7% a 97.1%, el s disminuyó de 2357 a 1148; y las variables explicativas MIPIR y CRIIR, son significativas (los estadísticos “t” son mayores que 2). El estadístico “d”, nos indica que no hay autocorrelación⁸ (pues $d=1.56$, $du=1.25$).

⁷ El valor crítico en las tablas del estadístico de Durbin – Watson, son un nivel de significación del 1% para rechazar la existencia de autocorrelación con 17 observaciones ($n=15$) y una variable explicativa ($K=1$) es de 1.10 ($du=1.10$), es decir menor al valor estimado en la ecuación (5): $d=1.47$. Ver J. Johnston, op. cit. Pág. 458.

⁸ Ver pie de página anterior

Si se compara esta ecuación, con los procesos anteriores, vemos que es más explicativa que la ecuación (1) (el R2 en dicha ecuación era de 95.5%) y no presenta el problema de autocorrelación en errores, pero explica menos las variaciones del PNBMR, y tiene estadístico “d” menor que la ecuación (2).

Luego, si se desea mejorar el poder explicativo de la ecuación de regresión, permitamos el ingreso de una nueva variable (siempre usando el criterio (i)).

Tercer Paso:

$$(7) \text{ PNBMR} = 2963.38 + 1.11 \text{ MIPIR} + 0.96 \text{ CRIIR} + 0.666 \text{ COFK}$$

(4.08) (3.44) (2.15)

$$R2 = 0971; F(3,13)=227.04; s = 1022.54; d=1.59$$

Es decir, que la inclusión de la variable GCOFK, logra aumentar el poder explicativo de nuestra ecuación de regresión, pero en un margen muy pequeño, así el R2 aumenta de 97.1% a 97.1% (menos de 1%), el s disminuye de 1148 a 1022. Cosa que se expresa en la reducción de las s.r.c. en una proporción menor al 1% (0.67). Todavía todas las variables explicativas resultan relevantes.

De comparar con la ecuación (2) se aprecia que todavía dicha ecuación tiene mejores resultados; un R2 un poco mayor, un s menor y un “d” más próximo a 2. Si se busca mayor explicación, podemos pasar al cuarto paso de este proceso, que reproducirá la ecuación (4), donde el R2 no es afectado significativamente con respecto a la ecuación (3), e indica que la variable MIPIR y CRIIR ya no son relevantes, en esta nueva especificación (véase la ecuación (4)). Este último resultado se debe al alto grado de colinealidad⁹ entre el YRINP y MIPIR.

De este proceso de regresiones por Pasos Sucesivos, donde se ha usado todos los criterios para la inclusión de variables en la ecuación de regresión, vemos que la ecuación más explicativa del comportamiento del PNBMR, en el periodo 1960 – 1976, es la ecuación (7). Sin embargo, si con el proceso se hubiese impuesto la restricción, para la inclusión de una nueva variable, de que se reduzca por lo menos en 1% la s.r.c., no se hubiese llegado al tercer paso, y la ecuación seleccionada sería la número (6).

Estamos frente a una serie de ecuaciones que explican los cambios en el PNBMR, como escoger la más adecuada. Antes de discutir los criterios para la selección, vamos a ilustrar el problema de Multicolinealidad.

⁹ Véase la próxima sección, para una mayor explicación de este problema.

Usando el criterio (i) para la inclusión de variables explicativas, se vio que la primera variable que entró en nuestra ecuación de regresión fue YRIND; usando el criterio (ii), se debe incluir las MIPIR, por haber sido un factor determinante en el comportamiento del PNBMR¹⁰, con la idea de lograr una ecuación más explicativa podemos hacer la regresión incluyendo como variables explicativas a YRIND y MIPIR, es lo que se hace en la siguiente sección.

4.3 REGRESIÓN POR MÍNIMOS CUADRADOS DIRECTOS.

Buscando una mayor explicación y una mejor especificación (incluir las variables claves) de nuestra ecuación, se hace la regresión del PNBMR, en función del YRIND y MIPIR por M.C.O. directos.

$$(8) \text{ PNBMR} = -6491.08 + 0.42 \text{ YRIND} - 0.19 \text{ MIPIR}$$

(4.97) (-0.29)

$$R^2 = 0.952; F(2,14)=161.3; s = 1468,23; d=1.05$$

Lo primero que llama la atención, es que la relación de causalidad entre el PNBMR y MIPIR sea negativa, pues del análisis histórico esperamos una causalidad positiva. Además, el estadístico “t” de la variable MIPIR, indica que no es relevante. Todo esto indica la existencia del multicolinealidad, es decir la presencia de una alta correlación entre las variables explicativas. Luego, el problema de multicolinealidad es cuando las variables explicativas se interfieren mutuamente, y resulta ser un problema muy grave porque el uso de ecuaciones con multicolinealidad, pueden producir proyecciones inadecuadas.

Como se ve, el problema de multicolinealidad, puede ser reconocida cuando el estadístico “t” de dos variables explicativas consideradas importantes, son pequeños. Incluso, a menudo la existencia de multicolinealidad puede ser causa, de que los estimadores de los coeficientes de regresión de variables explicativas tengan signos opuestos de los lógicamente esperados; como es el caso de la ecuación (8).

La multicolinealidad se confirma cuando las variables son comparadas en la matriz de correlación. Así, de observar dicha matriz (Cuadro N°2), se nota que YRIND y MIPIR están, de hecho, altamente correlacionadas (0.965).

Luego, el tratar de mejorar la ecuación de regresión, combinando criterios, y sin considerar posibles problemas de colinealidad dará un empeoramiento de la explicación de la ecuación de regresión. Siempre tratando

¹⁰ Y dada la coyuntura actual, parece que será el cuello de botella para el crecimiento del PNBMR.

de ilustrar, diremos que hay dos formas de eliminar el problema de colinealidad. Primero, a través de sacar una de las variables altamente correlacionadas de la regresión. Así si sacamos MIPIR, tendremos la ecuación (1), donde el R^2 en vez de disminuir incluso aumenta (de 95.2% a 95.5%) y el s disminuye, aunque no apreciablemente, pero el “ t ” del YRIND mejora drásticamente de 5 a 19 ahora que MIPIR ya no interfiere con dicha variable.

Una segunda forma de evitar el problema de multicolinealidad es cambiar la estructura de la ecuación. Hay diferentes formas de hacer esto; por ejemplo, a través de dividir tanto la variable dependiente como las variables explicativas por algunas series que no alteren la lógica económica, pero si levante la colinealidad. Otra forma es a través de estimar la ecuación en primeras diferencias. Así estimar la variación en el PNBMR en función de las variaciones del YRIND y de MIPIR.

Otra forma de evitar la colinealidad es a través de usar la técnica por “Pasos Sucesivos”, forzando el ingreso de variables claves con un signo determinado, e imponiendo restricciones para el ingreso de nuevas variables no forzadas (Ejemplo: que se reduzca la s.r.c. en por lo menos 1%). Véase la sección 4.2. de este trabajo.

Hasta este punto se ha logrado una serie de regresiones, ¿cuál es la mejor para lograr predicciones?

4.4 ELECCIÓN DE LA MEJOR ECUACIÓN

Con el avance de las computadoras principalmente, el problema de desarrollar ecuaciones de regresión ha cambiado de obtener una ecuación estimada a elegir entre varias ecuaciones que han sido fácilmente estimadas. Para elegir entre ecuaciones alternativas lo primero que debemos hacer es eliminar las que son obviamente malas, para luego elegir entre los que quedan.

Algunas ecuaciones ni siquiera cumplen con criterios estándar aceptables y pueden ser eliminadas fácilmente. Para elegir la mejor entre las que quedan es más difícil, e implica un análisis más detallado de los resultados de cada ecuación. Los siguientes principios han demostrado ser útiles para elegir la apropiada ecuación de proyección:

I. Eliminación de las Malas ecuaciones

1. Tiene sentido la ecuación

Debe haber un sentido de causalidad aceptable entra la variable dependiente y cada una de las variables explicativas. Este criterio inmediatamente elimina a la ecuación (3) donde el problema de colinealidad origina que la variable MIPIR tenga un signo contrario al que razonablemente se esperaba. Otra prueba de sensibilidad implica comparar los datos de la derecha e izquierda de la ecuación de regresión. Así, si la variable dependiente está expresada en unidades constantes, luego las variables explicativas no deben estar en unidades corrientes.

Si la variable dependiente está ajustada estacionalmente, también deben estarlo las variables explicativas. En las ecuaciones que buscan explicar el PNBMR, todas las variables están medidos en soles de 1963 (ver cuadro N°1).

2. Estadístico “t” bajos

Ecuaciones que tienen variables explicativas con “t” bajos deben ser reestimados o eliminados en favor de ecuaciones con todas las variables explicativas significativas. Este test eliminará las ecuaciones donde la colinealidad persiste. Así, la ecuación (4) será eliminada, donde debido al problema de colinealidad entre MIPIR y YRIND, el estadístico “t” es bajo para las variables MIPIR y CRIIR. También eliminaremos la ecuación (3)¹¹ en base a este criterio.

3. Reducción Proporcional en la suma de Residuos al Cuadrado (s.r.c.)

Dado que en la obtención de nuestras ecuaciones estimadas (siete de ocho) hemos usado la técnica de Pasos Sucesivos, este criterio será muy útil, para eliminar la ecuación en la que la última variable en entrar no reduce en una proporción mínima la s.r.c. Si fijamos esta proporción en el 1% (que de hecho es muy bajo), eliminaremos las ecuaciones (3) (ya eliminada con el criterio (2)), y la (7). En esta última ecuación la inclusión de la variable GCOFK, reduce la s.r.c. en solo el 0.67%.

4. R2 bajo

El R2 puede ser usado para ordenar las ecuaciones restantes y se-

¹¹ Analizando la matriz de correlación (Cuadro N°2) se nota que todas las variables están positivamente correlacionadas. Pero existe alta correlación relativa (mayor al 94%) entre el YRIND y MIPIR (96.5%), CRIIR y GCIFK (94.3%).

leccionar los mejores candidatos para la elección final.

Con los criterios anteriores ya eliminamos tres ecuaciones, más la ecuación (1), que fue descartada debido a la existencia de autocorrelación en los errores ($d=1.03$), quedando tres ecuaciones que al ser ordenadas en forma descendente de acuerdo al valor del R^2 se tiene la siguiente secuencia: en primer lugar, la ecuación (2) con R^2 igual a 98.2 %, luego la (6) con $R^2 = 97.1\%$ y por último la ecuación (5) con $R^2 = 87.7\%$.

En la secuencia establecida la ecuación (5) tiene un R^2 relativamente más pequeño que las otras dos ecuaciones. Luego usando este criterio eliminaremos la ecuación (5) como candidato a ser elegido. Es decir, que al usar los criterios anteriores se ha reducido el conjunto de ecuaciones de las que debemos escoger, nuestra ecuación óptima, a solo dos ecuaciones en las cuales la diferencia en la proporción explicada del comportamiento de la variable dependiente es de 1.1%. Esta diferencia puede ser suficiente para elegir la ecuación (2), pero en nuestro caso concreto resulta que las variables explicativas de una ecuación son diferentes a la de la otra. Revisemos los otros criterios, que sugiere Mc Lagan para elegir la Mejor Ecuación.

II. Prueba Ex – Post

Examinando el comportamiento de los valores estimados con respecto a los realizados u observados, tendremos el criterio más fácil para elegir una ecuación dentro de las ecuaciones que no han sido eliminadas todavía.

La ecuación que mejor capta las variaciones en la tendencia histórica (años de estancamiento) y cuya reciente precisión es mejor generalmente proveerá las mejores bases para las proyecciones.

Este criterio, usualmente, será suficiente para elegir a la mejor ecuación.

De analizar el Cuadro N°3, podemos decir que la ecuación (2) supera a la (6), para efecto de este criterio. Así los residuos en los años 1969 y 1976 (recesión) son menores en la ecuación (2); lo mismo si comparamos la precisión en años más recientes vemos que la ecuación (2) resulta superior.

III. Prueba Ex – Ante

Estimando una ecuación en base a una parte del período muestral y

usándola para proyectar el resto del intervalo muestral obtendremos juicios para chequear el grado de exactitud de proyección.

Dado que las proyecciones son hechas a base a un intervalo menor que el usado para estimar la ecuación, las proyecciones se llaman ex – ante. En base a esta prueba, diremos que la ecuación que tiene los valores históricos estimados más próximos a los valores históricos reales es superior. Esta prueba es solo usada cuando la prueba (II) falla en identificar claramente la superioridad de la ecuación a ser elegida.

Cuadro N°4: Valores observados y estimados del PNBMR
(Millones de soles de 1963)

Año	Valores observados	Valores estimados		Residuos	
		Ecuación (2)	Ecuación (6)	Ecuación (2)	Ecuación (6)
1960	10,642	10,591	10,383	51	259
1961	11,694	12,154	12,331	-460	-637
1962	12,912	13,504	13,634	-592	-722
1963	13,839	13,307	15,144	-532	-1305
1964	15,071	15,520	14,933	-440	138
1965	16,330	16,626	16,623	-296	-293
1966	17,935	18,097	17,706	-162	229
1967	18,832	18,711	18,787	121	45
1968	19,284	18,097	17,706	-162	229
1967	18,832	18,711	18,787	121	45
1968	19,284	18,135	18,336	1148	948
1969	19,554	18,995	18,701	559	852
1970	21,685	22,230	19,761	-596	1924
1971	23,550	23,708	24,483	-158	-933
1972	25,269	25,406	25,419	-226	-150
1973	27,164	26,624	28,315	540	-1151
1974	29,310	29,984	29,987	-624	-677
1975	30,687	32,286	31,753	-1599	-1066
1976	31,976	29,766	29,437	2210	2539

IV. El mejor estadístico de Durbin – Watson (d)

Este criterio será usado, si es que con las anteriores pruebas no se ha podido determinar la mejor ecuación. Entre las ecuaciones a ser elegidas, se tomará la ecuación que tenga el estadístico de Durbin – Watson más próximo a 2.0. Usualmente con los criterios anteriores ya se habrá elegido la mejor ecuación, como es en nuestro caso, sin embargo, si se desea usar este test se notará la superioridad de la ecuación (2), dado que tiene “d” igual a 2.02, mientras que en la ecuación (6) el

“d” está más alejado de 2 (1.66).

Es decir, que a través de usar diferentes criterios se puede generalmente, determinar la mejor ecuación que explique los cambios en la variable dependiente y poder ser usada para hacer proyecciones.

5. CONCLUSIONES

Actualmente, en el país, se advierte el uso extendido de la econometría para definir y cuantificar las relaciones entre las diferentes variables que se incluyen en un modelo económico y los datos observados en la vida real. En este sentido, esta investigación permite aprehender las diversas posibilidades de su uso en un caso concreto de estudio. Sin embargo, se hace hincapié en que el investigador no debe dejar todo el trabajo a las técnicas econométricas, sino que las debe complementar en base a sus conocimientos teóricos y su experiencia o conocimiento del fenómeno que busca explicar. Así, del pequeño diagnóstico elaborado en la Sección 2 de este trabajo, vemos que un factor condicionante del crecimiento del sector industrial peruano en el periodo analizado ha sido la disponibilidad de divisas para poder satisfacer la demanda de importaciones de materias primas y bienes intermedios para dicho sector. Además, de observar la coyuntura existente al final del periodo analizado, diremos que el factor restrictivo a corto plazo es la disponibilidad de divisas. Luego, si se desea proyectar el crecimiento de la producción manufacturera a Corto Plazo (1977), deberíamos considerar en nuestra ecuación la variable que exprese la restricción impuesta por la Balanza de Pagos (MIPIR). Es así que para lograr las proyecciones a corto plazo debemos considerar tanto la ecuación (2), como la (6).

REFERENCIAS

- Banco Central de Reserva BCR (1973). *Anuario Estadístico*.
- Christ, Carl F. (1966) *Econometric Models and Methods*. John Wiley & Sons, Inc.
- Grupo de investigaciones Económicas (GIECO) – UNI (1977). *Industrialización y Políticas de Industrialización en el Perú*.
- Johnston, J. (1972). *Métodos de Econometría*.
- McLagan, Donald L. “A Non-econometrician’s Guide to Econometrics”. Business Economics; Vol. VII, N°3, May, 1973.
- Theil, Henry (1971). *Principles of econometrics*. John Wiley & Sons, Inc.
- Wonnacott, T.H. y Wonnacott, R. J. (1970). *Econometrics*. Wiley International Edition,