

Un marco metodológico para el trabajo empírico en Economía: el caso de las variables¹

ÓSCAR MILLONES*

JORGE BAZÁN**

RESUMEN

En este artículo se revisa el manejo del concepto de las variables y las unidades de análisis (sujetos) en la investigación empírica en Economía. Se incorporan ideas desde otras áreas cuantitativas, y se ordenan en un marco metodológico. Se presenta una taxonomía de las variables y se discuten aspectos metodológicos que se derivan de ella. Se sugiere que el reflexionar sobre la naturaleza de las variables y sujetos desde el inicio del trabajo empírico es una tarea que debe ser parte de la formulación misma de teorías, hipótesis o ideas involucradas en el trabajo empírico económico. El lenguaje usado es no técnico para captar el interés de audiencias más amplias.

Palabras clave: modelos econométricos, variables, unidades de análisis, investigación empírica y medición en Economía.

ABSTRACT

This article reviews the concept of variables and units of analysis (subjects) in empirical research in economics. It incorporates ideas from other quantitative areas, and are arranged in a conceptual framework. We present a taxonomy for the variables and units and discuss methodological issues that arise from the framework. We suggest that the reflection on the nature of the variables and subjects from the beginning of the research is a task which should be part of the formulation of theories, hypothesis or ideas involved in the empirical economic research. We use non technical language to call the attention from a wider audience.

Keywords: Econometric models, variables, empirical research and measurement in Economics.

¹ Una versión preliminar y extendida de este trabajo aparece en Millones y Bazán (2008).

* Profesor del Departamento de Economía de la Pontificia Universidad Católica del Perú.

** Profesor del Departamento de Ciencias de la Pontificia Universidad Católica del Perú.

1. INTRODUCCIÓN: ¿POR QUÉ UN MARCO CONCEPTUAL METODOLÓGICO?

En los últimos años se han producido nuevos desarrollos en la investigación empírica tanto cuantitativa como cualitativa y especialmente en la denominada metodología mixta (ver por ejemplo Piore 2006; Creswell y Piano 2007; Tashakkori y Teddlie 2008). También se ha producido un importante incremento de métodos, técnicas, modelos y aplicaciones econométricas (en microeconometría y macroeconometría). Entre estas técnicas podemos mencionar el meta-análisis, los modelos dinámicos factoriales, la econometría bayesiana y el análisis de sistemas complejos. Algunas referencias importantes son Cameron y Trivedi 2005; Stock y Watson 2002b y 2006; Zellner 2008; Fair 2009; Colander 2000; Tesfatsion 2006. Estos desarrollos definen un nuevo desafío relacionado con la eficiencia en la utilización de recursos para el aprendizaje y la investigación en economía. En otras palabras, contamos cada vez con más literatura de gran implicancia metodológica empírica, pero disponemos de pocas guías orientadoras para optimizar su uso.

Este reto sugiere la necesidad de desarrollar estrategias que ordenen este panorama sobretodo desde la perspectiva del estudiante de economía de pre-grado, quien enfrenta la tarea de culminar sus estudios con una primera investigación.

El objetivo de este artículo es definir una estrategia o una metodología para la interpretación y el uso de las técnicas a partir de la caracterización de las variables en Economía. Con esto se pretende desarrollar un conjunto de conocimientos y habilidades que nos permiten pasar en forma lógica de los objetivos del investigador, al uso adecuado de métodos de verificación empírica y de la información disponible, de tal manera que le brinden aprendizajes significativos sobre los objetivos de la investigación.

En la sección 2 definimos brevemente el término «variables» y señalamos el rol de su medición. En la sección 3 presentamos el marco conceptual metodológico para las variables complementando con una dimensión sobre las unidades de análisis. Aquí desarrollamos las cuatro dimensiones del marco conceptual, especie de taxonomía para las variables para el trabajo empírico. En la sección 4 presentamos temas adicionales sobre el tratamiento de las variables en el trabajo de investigación en Economía vinculando estos temas al del marco conceptual. Se enfatiza a manera de discusión las implicancias del marco. Finalmente en la sección 5 se presenta las conclusiones del trabajo.

2. VARIABLES Y MEDICIÓN

¿Qué es una variable en la investigación empírica? Consideramos importante dar una definición heurística. Así, variable es una característica, factor o atributo de una unidad de análisis (individuos, hogares, firmas, objetos o sistemas) susceptible de ser medido o conceptualizado. También como su mismo nombre lo sugiere una variable es algo de la unidad de análisis que puede esperarse varíe con el tiempo o con el espacio de medición de referencia.

En este contexto, ¿qué es medir? Nuevamente una aproximación heurística nos permite considerar que medir es asignar números o valores a las unidades de análisis para representar ciertas propiedades². A manera de ejemplo, considérese un tema de investigación donde aparecen diferentes variables.

Ejemplo 1. Se plantea un estudio para entender por qué los estudiantes del bachillerato están demorando más de lo esperado en terminar su tesis. Algunos datos recogidos en las unidades de análisis que se pueden considerar como variables son el número de años dedicados a la especialidad, el promedio de rendimiento obtenido en la carrera, la pertenencia o no al tercio superior, la escala de pensiones de los estudiantes, el hecho de tener empleo o no, percepción sobre la calidad de la enseñanza medida con un test, el tipo de seminario de tesis llevado en el pregrado, etc.

En el ejemplo no todas las variables son del mismo tipo ni tienen igual jerarquía. Para caracterizar estas peculiaridades es necesario contar con un marco para entender mejor la naturaleza de estas variables, lo que se presenta a continuación.

3. EL MARCO METODOLÓGICO

En Economía, podemos proponer cuatro criterios para comprender la naturaleza de las variables, estos son:

- Criterio A: el rol de las variables en el tema en estudio
- Criterio B: la naturaleza de la construcción de las variables
- Criterio C: la escala de medición de las variables
- Criterio D: el diseño muestral para las variables

El *rol de las variables*, lo distinguimos porque nos permite apreciar las variables de acuerdo a su rol funcional en la investigación. La *naturaleza de la construcción de las variables* se refiere al hecho de que las variables consideradas son o no directamente observables y por tanto susceptibles de ser medidas con los datos directamente recolectados o derivados o inferidos a partir de estos. El tercer criterio, el *nivel o escala de medición*, se refiere a un conjunto de propiedades que distinguen las variables medidas de acuerdo a las operaciones matemáticas que pueden ser permitidas con ellas. Finalmente, el efecto *del diseño muestral* sobre la medición de las variables se refiere a si las variables planificadas para ser usadas son obtenidas en la forma muestral esperada por el investigador en cuanto a tres dimensiones: tiempo, espacio y tipo de relación entre las unidades de análisis.

En las siguientes secciones discutiremos, con mayor detalle, los cuatro aspectos mencionados³.

² Por los objetivos de este trabajo, aquí no será necesario hacer referencia a la definición axiomática de números en la medición (ver J. Pfanzagl 1971).

³ Aunque el orden de discusión de estos criterios no es importante, se presentan en este orden por la aparente complejidad de los mismos.

3.1. EL ROL DE LAS VARIABLE

Para el estudiante de Economía el concepto de rol debe ser familiar. Desde nuestras primeras clases de Macroeconomía o Microeconomía, nos hemos encontrado con especificaciones tales como la función consumo, inversión, o ecuaciones de demanda. El rol de las variables es el papel que juega cada variable en cada una de estas especificaciones, de allí que se prefiere hablar de rol funcional, en una visión simple, identificando variables dependientes e independientes.

Una variable es dependiente (o endógena) si es explicada por otra(s). Y es independiente (o exógena) si se incluye en la especificación del modelo pero no es explicada por otras⁴. El siguiente es un ejemplo en el contexto uniecuacional.

Ejemplo 2. Se trata de indagar los determinantes del rendimiento educativo en Matemáticas para el 6^{to} grado de primaria en una escuela urbana. Se puede decir que asociados a este rendimiento hay factores o variables del estudiante (horas por semana dedicadas al estudio, ingreso familiar), del profesor (calificación, experiencia) o de la escuela (material educativo, recursos informáticos). Para este ejemplo, el rendimiento en matemáticas es la variable endógena y las otras son exógenas.

Las variables dependientes en un estudio pueden ser independientes en otros estudios, y viceversa. Igualmente entre las variables independientes hay grados de jerarquías que han sido dadas en la literatura. En las variables del ejemplo 2, horas de estudio cumplen un rol directo en el tema, mientras que otras variables (tratamientos o factores) sociodemográficas como género o estatus económico pueden cumplir un rol de control en el análisis.

¿Siempre hay rol para las variables?: Análisis exploratorio y confirmatorio

Una pregunta importante sobre el rol de las variables es si siempre hay rol para las variables o no.

Decir que siempre hay rol para una variable equivale a subestimar la variedad de paradigmas metodológicos existentes para la investigación en Economía y en Ciencias Sociales. Resumiremos algunas ideas ampliamente aceptadas.

Las técnicas de investigación se pueden clasificar como técnicas exploratorias y confirmatorias. Las técnicas exploratorias se usan frecuentemente cuando poco se sabe sobre un tema de interés, especialmente en los inicios de un tema de investigación. Se puede identificar a las técnicas exploratorias como técnicas descriptivas que pueden aplicarse a poblaciones o muestras. Por lo que el tema de la inferencia no necesariamente es importante.

Por el contrario las técnicas confirmatorias implican dominio del tema a tal punto que el primer paso del análisis confirmatorio es la presentación formal de hipótesis o

⁴ En la literatura se encuentra denominaciones como «variables explicadas» y «explicativas»; «regresando» y «regresores», y otras.

teorías, las que van a ser sometidas a pruebas o contrastes en el trabajo empírico. En este sentido, el concepto de muestra y población es relevante.

Hay una variedad de técnicas exploratorias que incluyen por ejemplo desde el análisis exploratorio mismo (Tukey 1977) y (Benzecri *et al.* 1973; Nishisato 1994), pasando por las técnicas de reducción de información (Hotelling 1933; Anderson 1958), funcional data analysis (Ramsay y Silverman 1997) hasta los llamados ejercicios de Data Mining (Hastie *et al.* 2001). Esta última técnica se caracteriza por asociarse con recolección masiva de datos, el uso intensivo de procesadores (multiprocesadores) e incluir algoritmos en sus procedimientos.

La búsqueda de especificación de modelos puede hacerse incluso con la misma información, algunas veces con técnicas de *remuestreo* –data que luego se usará para futuros contrastes empíricos más confirmatorios, ver por ejemplo Leamer (1978).

La metodología cualitativa y metodología mixta

El trabajo que sirve para explorar y confirmar también pueden combinarse en un mismo estudio en forma secuencial como en Calvo *et al.* (1993), en donde en la primera parte se realiza una reducción de información para luego estimar un modelo recursivo tipo VAR.

En el análisis exploratorio no hay rol para las variables, al menos al inicio de la investigación. Esta característica de no rol para las variables es propicia para metodologías de investigación que son más cualitativas o menos positivistas en el lenguaje epistemológico (ver por ejemplo Greenacre y Blasius 1983). Más recientemente se están considerando en la literatura metodologías mixtas de investigación (ver Tashakkori y Teddlie 2008).

Rol de las variables e historia econométrica

Es interesante observar cómo el énfasis sobre el rol en el tratamiento de variables en Economía se ha movido como un péndulo. Esto se ve más claro en series de tiempo. Observamos el estudio de variables con rol Real Business Cycle (ver los Barómetros de Harvard de los años veinte y treinta en Lange 1964), que luego fueron calificados posteriormente como estudios de no rol (en la práctica los métodos del National Bureau of Economic Research fueron calificados por Koopmans a fines de los años cuarenta como estudios de «medida sin teoría», ver Malinvaud 1983). Luego hubo un tránsito al desarrollo con fuerte componente teórico (con rol) desde los años cincuenta con el pensamiento *probability approach* de la Cowless Commission, pasando luego en los años setenta por tratamientos más exploratorios, ver, por ejemplo, las aplicaciones a la economía en los trabajos iniciales de Box y Jenkins (1974) sobre las funciones de transferencias que solo fueron referidas hasta entonces para aplicaciones en física (Cap. 10), para volver últimamente a enfoques que son menos exploratorios. Dos ejemplos de esto último lo constituyen el concepto de cointegración (ver test en Engle & Granger 1987) y la literatura reciente referida por Stock y Watson (2005, 2006) en *Factor Dynamic Models*.

¿Es mejor la investigación confirmatoria?

Sin entrar a discusiones metodológico-epistemológicas, atendamos a hechos concretos. Desde el llamado nacimiento de la econometría (que incluyen los desarrollos incipientes) hasta nuestros días se ha entablado una disputa, aparentemente innecesaria, entre estrategias exploratorias y confirmatorias. Por ejemplo, en el análisis de series de tiempo, se puede interpretar el «conflicto» de la alternativa de usar VAR no restringido versus Ecuaciones Estructurales como una disyuntiva de este tipo.

Sobre este aspecto, sin embargo, se está reconociendo (aunque implícitamente) en la literatura econométrica que las estrategias exploratorias vs. confirmatorias son complementarias. Como se ha indicado, en desarrollos recientes se está incorporando técnicas más exploratorias para combinarse en el trabajo econométrico (ver Calvo *et al.* 1993; Cukierman, Edwards y Tabellini 1992; Annett 2000). En la literatura reciente de series de tiempo se está dando atención a la estrategia de levantar indicadores para alimentar modelos econométricos como, por ejemplo, el desarrollado para la economía americana, ver Stock y Watson (2006). Esta estrategia de combinar indicadores con variables observables fue sugerida inicialmente por Geweke (1977) y después por Diebold y Rudebusch (1996). Sobre este punto volveremos en la sección siguiente.

Este es un avance metodológico interesante en Economía, en parte debido tal vez a la influencia metodológica de otras áreas de las Ciencias Sociales (notablemente desde Sociología y Antropología), en donde no siempre el investigador plantea su trabajo con hipótesis previas. Para entender mejor y comparar estas estrategias, ver Travers (2001).

3.2. LA NATURALEZA DE LA CONSTRUCCIÓN DE LAS VARIABLES

Desde el punto de vista de cómo se obtienen (construyen) los valores para las variables, estas pueden ser *observables* (medibles directamente a través de los datos) o *constructos* (derivados o inferidos de los datos).

Variables observables

Las variables son observables si el valor obtenido o dato es una buena aproximación (o coincide exactamente) al concepto que se quiere medir. Hay una relación cercana, si no directa, entre el concepto teórico abstracto y su operacionalización en el mundo real (ver Torgerson 1958). Son ejemplos de variables observables, algunas de las que encontramos en la Encuesta Nacional de Hogares, allí tenemos: ingresos por trabajo, pagos mensuales por crédito, tipo de ocupación, tenencia de celular, estado civil, etc. Otros ejemplos de otras fuentes oficiales son: compra de dólares del tesoro público, exportaciones (en valores o cantidades), tipo de cambio intercambiario promedio, etcétera.

Los valores de estas variables (los datos) se obtienen directamente en el sentido de que su elaboración o construcción no crea mayor problema y es relativamente fácil de obtener (aun cuando su determinación sea algo compleja como el ingreso). La facilidad

de su obtención se debe a que no hay diferencia entre lo que se quiere medir (el concepto a medir, edad, por ejemplo) y lo que se obtiene como medición (información en el DNI, edad en una encuesta).

Es cierto que a menudo hay problemas de medición de variables observables, y la exactitud de su medición depende del contexto de su uso. Por ejemplo, en algunos trabajos, si hablamos del ingreso personal o familiar, es probable que nos refiramos al ingreso monetario (no en especie) de la ocupación principal, obtenido en un tipo de ocupación (actividad principal) y en un periodo dado (mes). Pero cualquiera que sea la definición que finalmente escojamos, es posible obtener una medición satisfactoria de lo que se quiere obtener a través de los datos recogidos. En econometría se ha dado énfasis a la diferencia (sesgo) entre el *constructo* de interés y sus indicadores. El tema allí está asociado al problema de los errores de medición y se ha enfatizado más bien las consecuencias de los sesgos en los métodos de estimación de los parámetros; esto se discute en el Anexo 2 de Millones y Bazán, 2008- en adelante referido como documento DDD-264.

Variables constructos⁵

Las variables constructos o latentes son las que no se pueden medir directamente de los datos recolectados. Estas variables latentes se miden a través de sus indicadores. Algunos ejemplos son:

Pobreza, variable que, dada su definición por el investigador, puede ser aproximada por algunos indicadores observables como el ingreso familiar, acceso a servicios básicos, grado de malnutrición, etc., pero como constructo, es diferente de estos indicadores.

Desempeño del estudiante en econometría, Capacidad de análisis económico (a los que podríamos aproximarnos por las notas en sus exámenes, en prácticas calificadas, etc.). Otros ejemplos son tecnología, productividad, expectativas de inflación, inestabilidad política, nivel de preferencias o utilidad, ingreso permanente, competitividad, barreras arancelarias, intangibles, etcétera.

La definición de constructo dada aquí es simple y tal como se ha definido arriba. Esta definición es clásica y se estableció originalmente en otras áreas fuera de la Economía. Concuere con el concepto de «constructo» en Sociología, así como con su uso en Psicología, donde se le ha denominado variable latente⁶.

Es decir, en un constructo o en una variable latente, el concepto se conoce pero es imposible medirla directamente de los datos porque carece (tal vez por ahora) de contrapartida empírica, u *operacionalidad* empírica. Aun cuando existan los recursos (y las intenciones) para medirla, los constructos solo pueden ser derivados o inferidos a través de indicadores, o de las propiedades medibles que envuelve el concepto.

⁵ Aquí se evitará usar el nombre de «No-observable» por las razones explicadas en el Anexo 2 de DDD-264.

⁶ Para mayor discusión revisar Borsboom, Mellenbergh y van Heerden (2003).

Constructos en otras áreas

En otras áreas como en Sociología, Psicología, Análisis Político, Marketing, el tratamiento de las variables constructo ha sido metodológicamente más sistemático y explícito que en el caso de Economía.

En Sociología se ha experimentado una variedad de opciones metodológicas en las últimas décadas, especialmente en el área de investigación cualitativa que va desde una posición interpretativa hasta una positivista sugiriendo estrategias diversas para el trabajo de campo.

Por otro lado, desde una perspectiva cuantitativa, la Psicología es un área en que probablemente casi todas las variables de interés son latentes. Su uso es una tradición, como se puede intuir, por ejemplo, de los temas sobre la inteligencia, su naturaleza, medición, desarrollo. La inteligencia es como un conglomerado de otras variables observables. Ver los tratamientos clásicos como en Spearman (1927), Thurstone (1938) y más recientemente Sternberg, Lautrey y Lubart (2003)⁷.

Con el desarrollo de estos métodos en la Psicología cuantitativa, se han brindado marcos empíricos para otras áreas, como los métodos de reducción de información, ejemplo, los usos de componentes principales, análisis factorial en finanzas y en series de tiempo. Igualmente la evaluación psicométrica de instrumentos de medición se ha derivado a propuestas como la Teoría de Respuesta al Item en Educación; el tratamiento de variables categóricas con escalamiento multidimensional en áreas de la Sociología, Filosofía y Política.

Constructos en Economía

El tratamiento en Economía de las ideas que solo se plasman en conceptos como constructos ha jugado un papel clave en el desarrollo de la Teoría Económica desde sus comienzos; por ejemplo, en el tratamiento del concepto de utilidad o bienestar y en otros conceptos en la teoría de la demanda y oferta y en particular en el concepto de escasez. Así un tema clásico en econometría, desde los escritos de Wright (1929) y Robbins (1932), fue el problema de distinguir empíricamente entre la demanda y cantidad tranzada. Sin embargo el reconocimiento formal como distinción entre constructo y observable ha sido explícito solo desde los años setenta y ochenta⁸.

⁷ Es interesante notar que hay otro grupo de variables latentes que han sido consideradas de mayor dificultad respecto a su medición como las actitudes, valores, ansiedad (nivel de), depresión, autoritarismo, modernismo, liberalismo, etcétera.

⁸ La literatura econométrica ya ha puesto explícito el tratamiento de variables latentes desde fines de los años setenta por Joreskog y Goldeberger (ver Aigner y Goldberger 1977) en el marco de modelos estructurales lineales (LISREL). Esto ha sido complementado a nivel microeconómico con los trabajos de Heckman y MacFadden en los modelos de elección discreta. Ver http://nobelprize.org/nobel_prizes/economics/laureates/2000/

Más recientemente en la literatura económica, por ejemplo en temas de organización industrial, se definen categorías latentes, como contexto político, controles verticales, ventajas y desventajas del costo de aprendizaje, cambios institucionales, flexibilidad del sistema regulatorio, barreras comerciales y otros, conceptos de gran interés.

Por último, y en un contexto más general en el marco del desarrollo de teorías con variables constructos y observables, Blalock (1984) ha mencionado que «para aplicar (contrastar) dichas teorías con datos empíricos se debe, de alguna manera, enlazar los conceptos más abstractos con varios indicadores u operadores de investigación». Para un tratamiento *operacionalizado* de esta idea ver Muthén y Muthén 2007. Evidentemente detrás de este desarrollo está el tema clásico sobre identificación discutidos en la literatura econométrica desde la época de los forjadores como Phillips G. Wright y posteriormente por Jean Tinbergen entre otros (ver excelente recopilación en Hendry y Morgan 1995).

3.3. LA ESCALA DE MEDICIÓN

En cursos introductorios de estadística para los economistas se describe una distinción simple de variables: variables cuantitativas y cualitativas. El desarrollo de métodos y técnicas señaladas en la sección 1 hace que esta clasificación en Economía sea insuficiente. Aunque hay una variedad de alternativas para clasificar las escalas de medición, nosotros empleamos aquí la escala de Stevens.

Según Stevens (1946), existen cuatro tipos de escalas o tipos de medición en las variables (o maneras de interpretar estos números), que son la escala nominal, ordinal, de intervalo y de ratio. La columna (3) de la Tabla 1 ilustra estos tipos de medición⁹:

Tabla 1. Propiedades y ejemplos de escalas de medición de las variables

Escala (1)	Propiedad (2)			Ejemplo 1 (3)
	Orden	Distancia	Origen	
NOMINAL	NO	NO	NO	Tipo de seminario llevado en el pre-grado
ORDINAL	SÍ	NO	NO	Escala de pensiones del alumno
INTERVALO	SÍ	SÍ	NO	Calidad de la enseñanza
RATIO	SÍ	SÍ	SÍ	Años en la especialidad

El tipo de escala de medición (columna 1) se determina por las propiedades operacionales (columna 2) que tienen los números asignados en la medición. Estas propiedades son: orden, distancia y origen. Para una mayor explicación de estas propiedades remitimos al trabajo DDD 264.

⁹ Los lectores no familiarizados con esta clasificación pueden revisar el anexo 1 del documento DDD-264.

3.4. EL DISEÑO MUESTRAL PARA LAS VARIABLES: LAS UNIDADES DE ANÁLISIS

Este criterio, a diferencia de los anteriores, se refiere más a los sujetos sobre quienes se miden las variables, que a las variables mismas. De allí que su discusión tenga otra dimensión, las unidades de análisis, como se explica seguidamente.

El diseño muestral para las unidades de análisis se refiere al modo de selección y forma de análisis de las unidades de estudio. Se distinguen tres aspectos:

- (a) La ocasión utilizada en la producción de sus valores
- (b) El espacio en que se definen las unidades de análisis
- (c) El tipo de relación (jerarquía o interacción) entre las unidades de análisis.

Es importante notar que en Econometría se ha dado énfasis a las dimensiones (a) y (b). Así, los aspectos en (a) sugieren el uso de los modelos de series de tiempo (la macroeconometría) y los aspectos en (b) nos conducen a los modelos de corte transversal (microeconometría). Por otro lado una combinación de estos aspectos desemboca en los modelos de panel y más recientemente en modelos de panel dinámico (ver Arellano 2003).

Por otro lado se está prestando mayor atención al estudio cuantitativo de las relaciones sociales entre las unidades de análisis señaladas en el criterio (c). Estos desarrollos son los modelos basados en agentes y se interesan más en las conductas de los sujetos y sus relaciones jerárquicas, que en las variables mismas asociadas de estos. Una colección interesante de temas y aspectos operativos de *software* se puede encontrar en Judd y Tesfatsion (2006). Ampliaremos este tema del tipo (c) con el concepto de «dualidad de la información» en la siguiente sección.

4. IMPLICANCIAS DEL MARCO CONCEPTUAL

La dualidad de la información

En conexión con el diseño muestral de las variables hay aspectos que es importante enfatizar más de lo que ya se ha expuesto en los estudios de panel y en el uso de variables nominales (*dummy*) en la literatura econométrica. Se trata del carácter dual de los datos. Por dualidad de los datos se entiende aquí la peculiaridad de analizar los datos por el lado de las variables y también por el lado de los sujetos.

Para visualizar mejor esta idea conviene considerar el aspecto físico de la información, es decir, un archivo de columnas y filas donde las columnas son las variables y las filas los sujetos.

Si miramos a los sujetos (filas), como especie de «variables transversales» podríamos interpretar los datos en filas como «variables especiales» generadas por el lado de los sujetos. Con esta óptica, podemos concentrarnos en el análisis estadístico para los sujetos, así como acostumbramos a pensar en el análisis estadístico para las variables. En otras áreas

de investigación cuantitativa, por ejemplo en Análisis Factorial, esta dualidad es conocida como análisis R (para las variables) y análisis Q (para los sujetos), etiquetas dadas por la estrategias de basarse en el estudio de la matriz de correlaciones en el primer caso, y en el estudio de la equivalente matriz por el lado de los sujetos, en el segundo caso. En algunas ramas de las ciencias administrativas e informáticas, esta idea del análisis de los sujetos está reforzándose con el nuevo desarrollo de *softwares*. Por ejemplo, en el marco del análisis de sistemas dinámicos para eventos discretos, ver Borschev y Filipov (2004) y en el marco del análisis de sistemas complejos para el estudio de sistemas ecosociales, ver el «comportamiento cooperativo» de Sornette y modelos de simulación (Earls 2008).

Dos dimensiones en los datos: antecedentes

La dualidad de los datos ha sido destacada en la literatura psicométrica por Nishisato (1980, 1994) en una síntesis de cinco décadas de desarrollos en el análisis multivariado¹⁰.

La discusión de la naturaleza de los patrones que se pueden identificar desde las filas (sujetos) es tratada en la literatura econométrica en forma progresiva. Una razón para la dificultad de este desarrollo se puede plantear considerando las filas, en esta matriz idealizada de sujetos y variables. Una fila es un vector de observaciones que implica tomar en cuenta parte o todas las variables de las columnas, resultando en el análisis de una heterogeneidad de variables del tipo de las presentadas por los criterios A, B y C. Esto implica considerar métodos de manejo de variables que sean flexibles y que enfrenten a una gama amplia en el tipo de variables. Para algunos casos, por ejemplo si todas las variables son de ratio o de nivel intervalo, métodos como componentes principales utilizados en Economía (Calvo *et al.* 1993; Alexander 2001; Stock y Watson 2006) están siendo implementados exitosamente. Sin embargo, cuando estas variables son parcial o totalmente nominales, los métodos tradicionales en econometría aún son limitados.

Hay desarrollos en los trabajos de Weitzman (1992) o Nehring y Puppe (2002), que se ocupan de este lado de la información (sujetos) y que han sido discutidos en temas de conservación del medio ambiente, sin embargo todavía están siendo planteados a nivel de proposición de teorías. Cabe destacar, en todo caso el giro hacia el análisis basado en agentes y la disponibilidad de *softwares* señalados en la literatura informática, ver Tesfatsion 2006 y Schieritz y Milling 2003.

En otros estudios empíricos como por ejemplo, los temas de distribución de ingresos y convergencia espacial, ha predominado un enfoque agregado y por lo tanto perdiendo información valiosa individual de los sujetos (ver críticas en Quah 1996). Sin embargo, por la propiedad de dualidad de los datos (Nishisato 1980) es posible estudiar comportamientos en forma desagregada y agregada simultáneamente para una variedad amplia de tipos de variables, sean medidas en escalas del nivel nominal y/o otras escalas.

¹⁰ Para una breve introducción a estos métodos con ejemplos e ilustraciones ver Nishisato (1994a).

Con la disponibilidad reciente de software (y hardware de mayor velocidad y economía), y con la creciente disponibilidad de información al nivel nominal o categórico, la idea de descomponer la varianza de las filas y/o columnas, por algún método de estimación, *singular value decomposition*, promedios recíprocos, máximo-verosímil, bayesiano u otro, se hace cada vez más factible un análisis con flexibilidad del comportamiento de los sujetos o grupos de ellos.

La incorporación de estos métodos en el desarrollo de temas de Economía es algo que puede sugerirse desde la literatura econométrica. Por ahora estas aplicaciones pueden inspirarse en el análisis exploratorio hecho ya algún tiempo en otras áreas como en Lebart, Morineau y Warwick 1984; Bourdieu 1979, en sociología y filosofía; y Mockus y Corzo 2003, en política.

5. CONCLUSIONES

El propósito de este trabajo fue proponer un marco metodológico conceptual para enfrentar el reto de digerir el vertiginoso desarrollo reciente de metodologías y métodos tanto cuantitativos como cualitativos.

Ha habido desarrollos desde otras áreas que son muy útiles al trabajo empírico en Economía. Lo que indica que también en el trabajo empírico es fundamental el trabajo interdisciplinario.

A manera de conclusión nosotros delineamos algunas implicancias que se desprenden de este marco metodológico y que son importantes de resaltar.

- La proliferación de más métodos y técnicas econométricas y otras metodologías hace necesario delinear una estrategia que ordene el panorama para el uso óptimo de esta tecnología disponible.
- La consideración de la naturaleza de las variables y las relaciones entre las unidades de análisis (sujetos) es útil y podría decirse necesaria para formular un plan para el trabajo empírico. Debiera ser parte (si no el primer paso) en la formulación de los objetivos en los trabajos de investigación. Elegir una estrategia empírica incluye tanto ver la naturaleza de las variables como la información por el lado de las conductas observadas entre los sujetos.
- La visión tradicional que clasifica las variables en cuantitativas (continuas y discretas) y cualitativas está resultando estrecha para destacar las alternativas de análisis de datos que existe en la literatura actual.
- Académicamente sería importante introducir y reforzar temas metodológicos de una manera más formal (sea como un curso o capítulo especial en libros de texto o más agresivamente en forma transversal a los cursos del currículo en economía), introducir una instancia de tratamiento de métodos o metodología que propicie el desarrollo de estos temas y así tender un puente entre teoría y verificación empírica.

- Ha existido un desbalance metodológico en el tratamiento empírico de variables en Economía. Relativamente más uso de variables de ratio y en un esquema predominantemente confirmatorio. La Econometría, en parte, ha contribuido a propiciar este desbalance en cuanto al conjunto de técnicas que ha puesto disponibles en los textos econométricos.
- El desequilibrio mencionado ha favorecido al enfoque específico con atención de solo algunas características de las variables, por ejemplo, las que se relacionan con su función de distribución *paramétrica*. La consideración de la escala, el rol y la naturaleza de la construcción de las variables podrían aminorar estos desbalances o potenciales sesgos.
- El proceder operativo que induce el uso del *software* convencional debe ser reconsiderado. Es necesario un nuevo rol para el investigador que sea más interpretativo en el uso de la información y del *software*. Esto puede lograrse con un marco más flexible donde la comprensión de la información o dato y del método de análisis sean tomados en cuenta.
- El «papel» del rol de las variables nos invita a considerar opciones más exploratorias en el trabajo de investigación y repensar al trabajo empírico como un proceso. En las etapas iniciales de la investigación es de esperar que no hay rol para las variables. Esto es un factor que debe estar ponderando mucho en los estudiantes del nivel de bachillerato.
- El desaliento por parte del investigador en el desarrollo de hipótesis o teorías que produce el no contar con información, podría ser explicado por la desconexión metodológica entre constructos y observables. En su extremo, esto lleva algunas veces al abandono de una estrategia empírica en la investigación.
- Existe literatura cuantitativa y cualitativa bastante flexible como para corregir la tendencia de sobre manipular la información para adaptarla a modelos econométricos, antes que seleccionar el modelo o metodología empírica que se adapte a la información disponible.
- Sobre este último punto, el análisis exploratorio de datos, la aproximación bayesiana y metodologías cualitativas, por ejemplo de la antropología y otras estrategias de investigación, juegan un papel fundamental para llenar estos vacíos.

Finalmente, en síntesis, en cuanto al trabajo empírico en Economía, delineando estrategias que aclaren los criterios básicos para seleccionar y hacer uso de la literatura econométrica y cuantitativa actual, podríamos maximizar los alcances y las potencialidades que nos brinda la información disponible.

REFERENCIAS

AIGNER, D.J. y A.S. GOLDBERGER (ed.)

1977 *Latent Variables in Socioeconomic Models*. Amsterdam: North-Holland.

ALEXANDER, C.

2001 *Market Models. A Guide to Financial Analysis*. New York: John Wiley & Sons, LTD.

ANDERSON, T. W.

1958 *An Introduction to Multivariate Statistical Analysis. A Wiley Publication in Mathematical Statistics*. New York: John Wiley.

ANDERSON, L. y D.A. KRATHWOHL

2001 *Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

ANNETT, A.

2000 «Social Fractionalization, Political Instability, and the size of Government». IMF Working Paper, No. 00/82. International Monetary Fund.

ARELLANO, M.

2003 *Panel Data Econometrics*. New York: Oxford University Press.

BANERJEE, A. y M. MARCELLINO

2008 «Factor-Augmented Error Correction Models». EUI Working Papers ECO 2008/15. European University Institute, Department Of Economics

BENZECRI, J. P. *et al.*

1973 *L'Analyse Des Données: II. L'Analyse des Correspondences*. París: Dunod.

BLALOCK, H. M. Jr.

1984 *Construcción de teorías en ciencias sociales. De las formulaciones verbales a las matemáticas*. México: Trillas.

BORSBOOM, D., G.J. MELLENBERGH y J. VAN HEERDEN

2003 «The theoretical status of latent variables». *Psychological Review*, 110, pp. 203-219.

BORSCHEV, A. y A. FILIPOV

2004 «From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools». *The 22nd International Conference of the System Dynamics Society*, July 25-29, Oxford, England.

BOURDIEU, P.

1979 *La distinction: critique sociale du jugement*. París: Minuit.

BOX, G.E.P. y G.M. JENKINS

1974 *Time Series Analysis. Forecasting and Control*. San Francisco: Holden Day.

CALVO G., L. LEIDERMAN y C. REINHART

1993 «Capital inflows and real exchange rate appreciation in Latin America. Section III: Econometric Analysis». *International Monetary Fund. Staff papers*, Vol. 40, N° 1.

- CAMERON, A. C. y P.K. TRIVEDI
2005 *Microeconometrics Methods and Applications*. Cambridge: Cambridge University Press.
- COLANDER, D. (ed.)
2000 *The Complexity Vision and the Teaching of Economics, Northampton, Mass.* Northampton, MA: Edward Elgar Publishing.
- CRESWELL, J. W., y V. L. PIANO CLARK
2007 *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.
- CUKIERMAN, A. S. EDWARDS y G. TABELLINI.
1992 «Seigniorage and Political Instability». *American Economic Review*. Vol. 82, N° 3, pp. 537-555.
- DIEBOLD, F. y G. D. RUDEBUSCH
1996 «Measuring Business Cycles: A Modern Perspective». *Review of Economics and Statistics* 78, February, pp. 67-7.
- EARLS, J.
2008 *Introducción a la Teoría de Sistemas Complejos*. Lima: Instituto de Estudios Ambientales. IDEA-PUCP.
- ENGLE R. F. y C. W. J. GRANGER
1987 «Co-integration and error correction representation, estimation, and testing». *Econometrica* 55, pp. 251-276.
- FAIR, R. C.
2009 *Using a Macroeconometric Model to Analyze the 2008-2009 Recession and Thoughts on Macroeconomic Forecastability*. New Haven, Connecticut: Yale University.
- FAN, J. y Q. YAO
2005 *Nonlinear Time Series*. Nueva York: Springer.
- GEWEKE, J.
1977 «The Dynamic Factor Analysis of Economic Time-Series Models». En D.J. Aigner y A.S. Goldberger (ed.). *Latent Variables in Socioeconomic Models*. Amsterdam: North-Holland, pp: 26-383.
- GREENACRE, M. y J. BLASIUS (eds.)
1983 *Correspondence analysis in the Social Sciences. Recent developments and applications*. London: Academic Press.
- GUILFORD, J.P.
1977 *La naturaleza de la inteligencia humana*. Buenos Aires: Paidós.
- HASTIE, T., R. TIBSHIRANI y J. FRIEDMAN
2001 *The elements of statistical Learning*. New York: Springer-Verlag.
- HENDRY, D. y M. MORGAN
1995 *The Foundation of Econometric Analysis*. Cambridge: Cambridge University Press.
- HOTELLING, H.
1933 «Analysis of a complex of statistical variables into Principal Components». *Journal of Educational Psychology*, 24, pp. 417-441, pp. 498-520.

JUDD, K. y L. TESHATSION (eds.)

2006 *Handbook of Computational Economics. Vol. 2: Agent-Based Computational Economics*. North-Holland.

LANGE, O.

1964 «Capítulo 1». *Introducción a la Econometría*. México D.F.: Fondo de Cultura Económica.

LEAMER, E. E.

1978 *Specification searches: ad hoc inference with nonexperimental data*. New York: J. Wiley.

LEBART, L., A. MORINEAU y K. WARWICK

1984 *Multivariate descriptive statistical analysis*. New York: John Wiley and Sons.

MALINVAUD, E.

1983 «Econometric Methodology at the Cowles Commission: Rise and Maturity». Abstracted from the Cowles Fiftieth Anniversary Volume (conference held at Yale University on June 3-4, 1983 to commemorate the Cowles research group's fiftieth anniversary). Ver <http://cowles.econ.yale.edu/archive/reprints/50th-malinvaud.htm>.

MILLONES, O. y J. L. BAZÁN

2008 «El tratamiento de las Variables en Economía: Una revisión y un marco conceptual». Documento de Discusión Centro de Investigaciones Sociales, Económicas, Políticas y Antropológicas CISEPA. DDD264. Pontificia Universidad Católica del Perú.

MOCKUS, A. y J. CORZO

2003 «Dos caras de la convivencia. Cumplir acuerdos y normas y no usar ni sufrir violencia». *Análisis Político*, N° 48, enero/abril.

MUTHÉN, L.K. y B. O. MUTHÉN

2007 *Mplus: Statistical Analysis with Latent Variables: User's Guide*. Los Ángeles CA: Muthén & Muthén.

NEHRING, K. y C. PUPPE

2002 «A theory of diversity». *Econometrica*, 70, pp. 1175-1178.

NISHISATO, S.

1980 *Analysis of categorical data: Dual Scaling and its applications*. Toronto: University of Toronto Press.

1994a *Dual Scaling in a Nutshell*. Toronto: MicroStats.

1994b *Elements of Dual Scaling. An introduction to practical Data Analysis*. New Jersey: LEA Publishers.

PFANZAGL, J.

1971 *Theory of Measurement*. Segunda edición. Würzburg: Physica Verlag.

PIORE, M

2006 «Qualitative Research: Does it fit in economics?». *European Management Review*, Vol. 3 (1), pp. 17-23, 2006. En Curren, S. y E. Peregman (ed.) *Handbook for Social Science: Field Research Essays & Bibliographic Sources on Research Design and Methods*. Thousand Oaks, CA: Sage.

QUAH, D.

1996 «Regional Convergence Cluster across Europe». *European Economic Review*, 40, pp. 951-958.

RAMSAY, J.O. y B. W. SILVERMAN

1997 *Functional data analysis*. New York: Springer.

ROBBINS, L.

1932 *An essay on the nature and significance of economic science*. London: Macmillan.

SCHIERITZ, N. y A. GROSSLER

2003 «Emergent Structures in Supply Chains - A Study Integrating Agent-Based and System Dynamics Modeling». *The 36th Annual Hawaii International Conference on System Sciences*. Washington.

SCHIERITZ, N. y P. MILLING

2003 «Modeling the Forest or Modeling the Trees - A Comparison of System Dynamics and Agent-Based Simulation». *The 21st International Conference of the System Dynamics Society*. New York.

SPEARMAN, C.

1927 *The abilities of man. Their nature and measurement*. Toronto: The MacMillan Company of Canada.

STERNBERG, R.J., J. LAUTREY y T.I. LUBART

2003 «Where are we in the field of intelligence, how did we get here and where are we going». En Sternberg et al. (ed.). *Models of Intelligence. International perspectives*. Washington DC: American Psychological Association.

STEVENS, S. S.

1946 «On the theory of scales of measurement». *Science*, Vol. 10, 103, pp. 677-680.

STOCK, J., M. WATSON

2002a «Forecasting using principal components from a large number of predictors». *Journal of the American Statistical Association*, 97, pp. 1167-1179.

2002b «Macroeconomic forecasting using diffusion indexes». *Journal of Business and Economic Statistics*, 20, pp. 147-162.

2005 «Implication of dynamic factor models for VAR analysis». Working Paper 11467, National Bureau of Economic Research, Inc.

2006 «Forecasting with many predictors». En Elliott, G., C.W.J. Granger, A. Timmermann (ed.). *Handbook of economic forecasting*, Vol. 1, pp. 515-554.

TASHAKKORI, A. y Ch. TEDDLIE

2008 *Foundations of Mixed Methods Research. Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. Ciudad: Louisiana State University, Baton Rouge Florida International University

TESFATSION, L.

2006 «Agent-Based Computational Economics: A Constructive Approach To Economic Theory». En K. Judd y L. Tesfatsion (ed.). *Handbook of Computational Economics, Vol. 2: Agent-Based Computational Economics*. North-Holland.

THURSTONE, L. L.

1938 *Primary Mental Abilities*. Chicago: University of Chicago Press.

TORGERSON, W.

1958 *Theory and Methods of Scaling*. New York: Wiley.

TRAVERS, M.

2001 *Qualitative Research Through Case Studies*. London: Sage Publications.

TUKEY, J. W.

1977 *Exploratory Data Analysis*. Massachusetts: Addison-Wesley

WEITZMAN, M.

1992 «On diversity». *Quarterly Journal of Economics*, 107, pp. 363-405.

WRIGHT, P. G.

1929 «Review of Schultz, Statistical Laws of Demand and Supply». *Journal of American Statistical Association*, Vol. 24, pp. 207-215.

ZELLNER, A.

2008 *Bayesian Econometrics: Past, Present and Future*. Paper 0803, Graduate School of Business, University of Chicago.