Vinculación longitudinal

de los Censos Económicos 1994-2014 de México

The Longitudinal Linkage

of Mexico's Economic Census 1994-2014

Matías Busso,* Óscar Eduardo Fentanes Téllez y Santiago Levy Algazi*****

* Banco Interamericano de Desarrollo (BID), RES, MBUSSO@iadb.org
** BID, VPS, o.fentanes.t@gmail.com
*** BID, VPS, santiagolevy4@gmail.com

Nota: agradecemos al Instituto Nacional de Estadística y Geografía (INEGI) por el acceso a los datos utilizados para este enlace; el enlace presentado en este documento es un trabajo llevado a cabo por personal del Banco Interamericano de Desarrollo y no se considera parte de los registros oficiales del INEGI; los identificadores creados y los datos mencionados pueden ser consultados en el Laboratorio de Microdatos bajo aprobación del INEGI; se agradece, también, a Jesica Torres Coronado y Natalia Volkow por sus valiosos comentarios y sugerencias; las opiniones expresadas en esta publicación son de los autores y no necesariamente reflejan el punto de vista del BID, de su Directorio Ejecutivo ni de los países que representa.

Esta nota técnica describe la metodología para construir una base de datos longitudinal a partir de los Censos Económicos de 1994 hasta 2014. El proceso se basa en un algoritmo que enlaza establecimientos con idéntica o similar ubicación, entidad legal y clase de actividad. Puesto que ya existe un conjunto de identificadores longitudinales para los de 2009 y 2014, estos son utilizados para validar nuestros resultados, obteniendo 90% de precisión. Enlazamos 0.92 millones de establecimientos para los Censos 1994-1999, 1.44 millones para 1999-2004, 1.52 millones para 2004-2009 y 2.15 millones para 2009-2014.

Palabras clave: datos longitudinales; Censos Económicos; INEGI.

Códigos JEL: C81, D21

This technical note describes the methodology to construct a longitudinal dataset using the Economic Censuses of Mexico from 1994 to 2014. The procedure is based on an algorithm that links establishments with identical or significantly similar location, legal entity and kind of activity. Since a set of longitudinal identifiers is already available for the 2009 and 2014 Economic Censuses, it is used to validate our results, obtaining 90% of accuracy. This paper links 0.92 million establishments for the period 1994-1999, 1.44 million for 1999-2004, 1.52 million for 2004-2009, and 2.15 million for 2009-2014.

Key words: Longitudinal Database; Economic Census; INEGI.

JEL Classification: C81, D21

1. Introducción

El Instituto Nacional de Estadística y Geografía (INEGI) lleva a cabo en México los Censos Económicos (CE) quinquenalmente desde 1989. Los CE recolectan información de todos los negocios operando en instalaciones fijas y ubicados en localidades urbanas de más de 2 500 habitantes.

Los Censos 2009 y 2014 introdujeron el identificador Clave Única de Identificación Estadística (CLEE),[1] el cual enlaza longitudinalmente establecimientos de ambos levantamientos censales y los subsecuentes. A pesar de que la CLEE ya puede ser utilizada para estudios longitudinales, no está disponible para ediciones anteriores, limitando el potencial de estas bases de datos.

En esta nota técnica, describimos el proceso de vinculación de los Censos Económicos 1994 hasta 2014. A pesar de que en principio se podría incluir también los de1989, existen dificultades para armonizar su codificación industrial y geográfica con otras ediciones.

El trabajo de vinculación de los CE 1999 hasta 2014 ya fue descrito en la nota técnica de Busso, Fentanes y Levy (2018). La presente investigación la extiende a los Censos 1994 e incluye en el *Apéndice B* el funcionamiento de un par de comandos para STATA que permiten corregir inflexiones en cadenas de texto para facilitar los trabajos de vinculación con otros conjuntos de registros.

El resto del documento se estructura como sigue: en la sección 2 describimos los datos; la 3 se enfoca en el algoritmo de vinculación de los Censos; en la 4 presentamos los resultados; en la 5, realizamos algunos ejercicios de validación, incluyendo medidas de flujos de trabajadores y entrada y salida de establecimientos; finalmente, en la sección 6, discutimos algunas consideraciones sobre nuestro procedimiento y en la 7 explicamos cómo solicitar el acceso a los identificadores aquí descritos a través del Laboratorio de Microdatos del INEGI.

2. Los Censos Económicos de México

2.1 Cobertura

Nuestra fuente de datos son los Censos Económicos. La cobertura temporal es 1994, 1999, 2004, 2009 y 2014.[2] Utilizamos todas las clases de actividad (cerca de 800 por año) y todas las localidades urbanas disponibles. Ya que la unidad económica de los CE es el establecimiento, la vinculación que aquí presentamos también es a este nivel. El número de establecimientos por periodo se muestra en el cuadro 1.

La cantidad de unidades económicas de los CE se incrementa, principalmente, por el nacimiento de nuevos establecimientos, pero también por ampliaciones en la cobertura, sobre todo, la adición de nuevas localidades reclasificadas como urbanas.

Cuadro 1

Establecimientos

Censos	Total
1994	2 209 443
1999	2 804 984
2004	3 005 157
2009	3 724 019
2014	4 230 745

2.2 Variables

Para todos los CE, contamos con información detallada que nos permite identificar establecimientos; por ejemplo, entidad legal, nombre de la unidad, claves de localización, año de inicio de actividades, clase de actividad, entre otra. La lista completa se muestra en el cuadro 2.

Los códigos de ubicación de E03 a E07 son codificaciones estandarizadas definidas por el INEGI; las variables E10, E11 y E14 son cadenas de texto capturadas manualmente; E08 es el nombre del establecimiento, por ejemplo, *Minimercado María*;

1 Fue creada mediante una combinación de procedimientos humanos y computacionales de identificación.

2 Cada edición captura información del año previo, por ejemplo, la edición 2009 contiene información del 2008.

Cuadro 2

Variables disponibles

Ubicación	Entidad legal	Industria
E03 estado	E01 NIC	E17 clase de actividad
E04 municipio	E02 NOP	
E05 localidad	E08 nombre del establecimiento	
E06 AGEB[a]	E09 razón social	
E07 manzana	G111 año de inicio de actividades	
E10 calle		
E11 número		
E14 colonia		

[a] AGEB significa Área Geoestadística Básica.

E09 es la entidad legal, por ejemplo, *María SA de CV*; si la unidad no pertenece a una entidad legal, E09 reporta el nombre del(la) propietario(a); E17 es la clase de actividad de acuerdo con el Sistema de Clasificación Industrial de América del Norte (SCIAN) para los Censos 1999 hasta 2014. Para los de 1994 y 1999, las clases de actividad se agrupan de acuerdo con la Clasificación Mexicana de Actividades y Productos (CMAP).[3]

Las variables E01 (NIC)[4] y E02 (NOP)[5] son identificadores disponibles para todos los Censos. Estos pueden ser utilizados para vinculación longitudinal para todos los CE, pero solo para un número limitado de establecimientos, generalmente grandes. Para la mayor parte de las unidades económicas, la combinación de códigos NIC-NOP solo puede ser utilizado como un identificador dentro de un levantamiento, pero no longitudinalmente.

3. Vinculación

Su proceso se puede resumir en cinco pasos, como el modelo presentado en Christen (2012) —el enlace se lleva a cabo del I al IV; el V, validación, será discutido en la sección 5—, que son:

I. Estandarización: se sustituyen o eliminan caracteres especiales, como acentos y signos de puntuación; también, se armonizan descripciones de los establecimientos, como *Abarrotes* o *Tienda*, que representan el mismo tipo de negocio; asimismo, se corrigen errores en el tipo de entidad legal, como *SA d CV* en lugar de *SA de CV*; en general, se eliminan, estandarizan o sustituyen caracteres en todas las variables capturadas manualmente (no claves estandarizadas), que son propensas a errores de captura.
II. Indexación (previnculación): se proponen candidatos para vinculación; por ejemplo, si dos establecimientos tienen la misma ubicación en t y $t+5$, se compara el nombre del(la) propietario(a) o del establecimiento para decidir si es un buen enlace.
III. Comparación: se usan diferentes estrategias para comparar pares de establecimientos indexados; en general, se utilizan procedimientos de STATA para comparar cadenas de texto.
IV. Clasificación de enlaces: se asigna un identificador único a los establecimientos enlazados; después, se etiquetan con el fin de ser excluidos en futuras fases del algoritmo —discutido más adelante—; además, se asigna el número de la fase en la que fue enlazado un establecimiento.
V. Validación: se mide la precisión del algoritmo aplicándolo a los CE 2009 y 2014,

3 Los Censos Económicos 1999 incluyen ambas clasificaciones industriales.
4 Número de Identificación Censal.
5 Número Operativo.

los cuales ya fueron enlazados y validados por el INEGI (ver sección 5).

Para llevar a cabo los pasos I-IV, se define un algoritmo de 10 fases. Todas se basan en las reglas de continuidad definidas por la Organización para la Cooperación y el Desarrollo Económicos (OCDE, 2008). Estas consideran tres factores de continuidad: ubicación, entidad legal y clase de actividad; si alguna unidad económica mantiene al menos dos de tres de un periodo a otro, se considera la misma.

Para ejecutar las 10 fases, se emplea principalmente STATA. En algunas utilizamos el comando *matchit*, escrito por Raffo (2017), el cual compara cadenas de texto y asigna un coeficiente de similitud entre 0 y 1. También, usamos el comando *soundex*, el cual consiste en la primera letra de la cadena de texto seguido de tres dígitos asignados por STATA; estos son los mismos para similares cadenas de consonantes.

Las fases son:

1. Se enlazan los establecimientos con combinación idéntica de NIC y NOP.[6]
2. Se enlazan establecimientos con la misma combinación de estado, municipio, localidad, AGEB, manzana y clase de actividad.
3. Se indexan unidades económicas con la misma combinación de estado, municipio, localidad, AGEB, manzana y número exterior. Luego, se enlazan si tienen un coeficiente de similitud de, al menos, 45% en el nombre del establecimiento y 75% en la entidad legal.[7]
4. Se indexan establecimientos con la misma combinación de estado, municipio, clase de actividad y entidad legal. Luego, se enlazan si tienen un coeficiente de similitud de, al menos, 30% en el nombre de la unidad.
5. Se enlazan establecimientos con la misma combinación de estado, municipio, AGEB y entidad legal.
6. Se indexan los establecimientos con la misma combinación de estado, municipio, localidad, AGEB, manzana y clase de actividad. Luego, se enlazan si tienen el mismo *soundex* en el nombre de la unidad económica y entidad legal.
7. Se enlazan los establecimientos con la misma combinación de estado, municipio, localidad, AGEB, manzana, clase de actividad y año de inicio de actividades.
8. Se enlazan los establecimientos con la misma combinación de estado, municipio, localidad, AGEB, manzana, clase de actividad y número exterior.
9. Se indexan las unidades económicas con la misma combinación de estado, municipio, localidad y AGEB o clase de actividad. Luego, se enlazan si tienen un coeficiente de similitud de, al menos, 65% en el nombre del establecimiento y en entidad legal.
10. Se enlazan los establecimientos con la misma combinación de clase de actividad, nombre del establecimiento y entidad legal.

Siempre que se enlazan unidades económicas de acuerdo con una secuencia de variables se consideran solo aquellas que presenten una combinación única dentro del levantamiento censal; por ejemplo, en la fase 2 enlazamos las que tienen la misma ubicación y clase de actividad en t y $t+5$; sin embargo, si dos reportan la misma ubicación y clase de actividad en t, no será claro cuál de los dos es el que reapareció en $t+5$. Para evitar ambigüedades y minimizar errores de vinculación, excluimos estos casos y se intenta en fases futuras enlazar los establecimientos con diferentes combinaciones de variables.

Los valores de los coeficientes de similitud que se requieren en algunas fases fueron determinados de tal forma que logren predecir correctamente al menos 90% de los enlaces (en los CE 2009 y 2014); se puede ser más restrictivo con estos, pero las ganancias en precisión no necesariamente compensan las pérdidas de buenos enlaces por no cumplir con los nuevos criterios. En la sección 5 se detalla la precisión de cada fase.

6 Algunos NIC-NOP tiene duplicados. En los CE 1999 eran menos de 400; en los Censos 2004 y 2009, menos de 100; y ninguno en la edición 2014.

7 Si el nombre del establecimiento o la entidad legal aparece vacía o reporta *SIN NOMBRE*, no se considera para el enlace.

4. Resultados

Tras llevar a cabo las 10 fases del algoritmo, se obtienen los resultados mostrados en el cuadro 3. Para cualquier par de CE adyacentes, t y $t+5$, enlazamos al menos 50% de los establecimientos en t.

El cuadro 4 desglosa los enlaces totales por fase.[8] Las fases 1 a la 6 son, por mucho, las más importantes, representando al menos 86% de los enlaces para cualquier periodo (y como veremos más adelante, también son las más precisas).

El algoritmo de enlace solo se aplica a pares de CE consecutivos; sin embargo, podemos seguir algunos establecimientos a través de varias ediciones. De acuerdo con los cuadros 3 y 5, de los 2.2 millones de establecimientos registrados en los CE 1994, sobrevivieron 991 mil en los de 1999; de estos, 637 mil reaparecieron en los Censos 2004; para los de 2009, de estos quedaron 433 mil; y, finalmente, en la edición 2014 se capturaron de nuevo 333 mil de ellos. En otras palabras, podemos formar un panel balanceado de 333 mil establecimientos de los levantamientos censales 1994 hasta 2014. Por otra parte, también es posible integrar otro de 1999 hasta 2014 de 675 mil (ver cuadro 4; otras posibilidades se pueden observar en este mismo).

Cuadro 3

Establecimientos enlazados

t a $t+5$	Establecimientos en t	Enlaces	% t
1994-1999	2 209 443	991 230	44.9
1999-2004	2 804 984	1 444 584	51.5
2004-2009	3 005 157	1 522 578	50.7
2009-2014	3 724 019	2 154 410	57.9

8 La fase 1 no fue aplicada para 2009-2014 porque las variables NIC y NOP son redundantes con la CLEE. Mientras que para periodos anteriores, la fase 1 logra alrededor de 7% de los enlaces, para 2009-2014 representaría 100 por ciento.

Cuadro 5

Posibles paneles balanceados

Periodos	CE	Establecimientos
3	1994-1999-2004	637 465
	1999-2004-2009	907 992
	2004-2009-2014	1 081 257
4	1994-1999-2004-2009	433 033
	1999-2004-2009-2014	675 379
5	1994-1999-2004-2009- 2014	333 041

5. Validación

La calidad del enlace depende de su cobertura y precisión; ambas características pueden ser evaluadas respondiendo a las siguientes preguntas:

(i) Cobertura: ¿cuántos establecimientos deben ser enlazados para cada par de CE consecutivos?
(ii) Precisión: ¿cuál es la probabilidad de que dos establecimientos vinculados sean efectivamente el mismo?

Cuadro 4

Porcentaje de enlaces por fase

		Fase										
Periodo	Enlaces	1	2	3	4	5	6	7	8	9	10	Total
1994-1999	991 230	7.7	33.8	25.8	0.8	27.3	3.1	0.1	1.0	0.4	0.0	**100.0**
1999-2004	1 444 584	7.1	38.2	17.2	15.1	6.5	2.1	0.5	12.5	0.8	0.1	**100.0**
2004-2009	1 522 578	7.1	40.0	16.9	11.2	8.9	3.0	1.6	10.4	0.8	0.1	**100.0**
2009-2014	2 154 410	0.0	49.2	14.1	16.8	9.4	2.7	2.8	4.3	0.5	0.1	**100.0**

Para responder ambas, comparamos el enlace realizado por el INEGI a través de la CLEE y el que se logra con nuestro algoritmo.

5.1 Cobertura del enlace

El cuadro 6 muestra que 58% de los establecimientos de los CE 2009 pueden ser enlazados con unidades económicas de los Censos 2014 utilizando la CLEE. Por lo tanto, esperaríamos que el algoritmo enlazara un porcentaje similar; así, éste enlaza 2 154 410, es decir, 57.9% de los establecimientos de los CE 2009. El número de enlaces de ambos métodos es virtualmente el mismo, el algoritmo logra 99.7% del monto total que enlaza la CLEE.

Cuadro 6

Enlaces por método, 2009-2014

Método	Establecimientos en 2009	Enlaces	%
CLEE	3 724 019	2 159 804	58.0
Algoritmo	3 724 019	2 154 410	57.9

El algoritmo alcanza el número esperado de enlaces de los CE 2009-2014; sin embargo, esto no responde por completo a la pregunta (i), necesitamos estimar también cuántos establecimientos debe enlazar este para 1994-1999, 1999-2004 y 2004-2009.

Una forma de responder esta pregunta es utilizando el año de inicio de actividades declarado por informantes. Si un establecimiento reporta una edad mayor o igual a 5 en $t + 5$, podría potencialmente ser observado en t. De acuerdo con el cuadro 7, en los Censos 2009, cerca de 1.8 millones de los 3.7 millones de establecimientos reportaron operaciones en el levantamiento censal 2004 o antes según su edad. Como en los CE 2004 capturaron 3 millones, esperaríamos enlazar 59.7% de ellos con unidades de los de la edición 2009. Aplicando el mismo razonamiento, esperaríamos enlazar 51.5% de los establecimientos de 1999 con unidades de 2004 y 53.3% de 1994 con los de 1999.

Cuadro 7

Establecimientos por edad

Edad	1999	2004	2009	2014
Menos de 5	1 627 248	1 561 466	1 928 674	1 806 638
5 o más	1 177 736	1 443 691	1 795 345	2 424 107
Total	**2 804 984**	**3 005 157**	**3 724 019**	**4 230 745**

Sin embargo, hay dos razones por las que el número de enlaces esperados podría estar sobrestimado si nos basamos solo en el año de inicio de actividades reportado. La primera es que los Censos Económicos expanden su cobertura geográfica en cada edición dado que algunas localidades crecen y comienzan a ser consideradas como urbanas;[9] la segunda, dado que el año de inicio de actividades es reportado por informantes, este puede ser impreciso. Mediante el identificador CLEE, podemos medir la discrepancia entre los que sobreviven de acuerdo con su edad y los que en efecto son enlazados de los CE 2009 hasta 2014.

En los Censos Económicos 2014 se capturaron 4.2 millones de establecimientos, de los cuales, el INEGI enlazó 2.2 millones mediante la CLEE. Al mismo tiempo, en la edición 2014, 2.4 millones reportaron actividades en los CE 2009 según su edad; en otras palabras, el INEGI enlazó 89.1% de los establecimientos que declararon operaciones en los levantamientos 2009 y 2014. Si asumimos que este grado de discrepancia (por cobertura o error de reporte) se mantiene constante en el tiempo, esperaríamos enlazar solo 89.1% de los que reportan actividades en cualesquiera dos CE según su edad.

En el cuadro 8 se observa que 1.1 millones de establecimientos manifestaron actividades en los Censos 1994 y 1999 de acuerdo con su edad. De estos, esperamos enlazar 89.1%, es decir, poco más de 1 millón. Finalmente, se enlazan 0.99 millones, es decir, una cobertura de 94.5 por ciento. Bajo este mismo razonamiento, se logra una co-

9 En 2009, el levantamiento censal se hizo en 2 mil localidades urbanas. Para 2014, se realizó en 3 600; sin embargo, las 1 600 nuevas solo aportaron 12 mil establecimientos adicionales.

bertura de 112.3% para las ediciones 1999-2004, de 95.2% para 2004-2009 y de 99.7% para 2009-2014. Una superior a 100% implica una posible sobrestimación de la supervivencia de las unidades económicas y una inferior a 100%, una sobrestimación de su mortalidad.

5.2 Precisión del enlace

La pregunta (ii) también puede ser respondida tomando como referencia el enlace 2009-2014 hecho por el INEGI mediante la CLEE. Como se mencionó con anterioridad, el algoritmo y la CLEE enlazan de manera virtual el mismo número de establecimientos; sin embargo, esto no significa que ambos métodos enlacen exactamente los mismos. De hecho, se pueden obtener los siguientes cuatro tipos de enlace (o no enlace):

1. Verdadero positivo: enlazado con el algoritmo y la CLEE.
2. Falso positivo: enlazado con el algoritmo, pero no con la CLEE.
3. Verdadero negativo: no enlazado con el algoritmo ni la CLEE.
4. Falso negativo: no enlazado con el algoritmo, pero sí con la CLEE.

El cuadro 9 presenta los porcentajes de establecimientos de los CE 2009 y 2014 según las cuatro posibilidades de enlace descritas. De los enlazados con el algoritmo en el levantamiento 2009, 90% también fue enlazado por la CLEE (89.8% respecto a 2014). En general, el porcentaje de verdaderos positivos puede interpretarse como la precisión del algoritmo, o bien, como la probabilidad de que dos establecimientos enlazados por el algoritmo sean efectivamente el mismo.

Cuadro 8

Enlaces esperados y realizados

Periodo	Edad ≥ 5	Enlaces esperados	Enlaces	%
1994-1999	1 177 736	1 049 363	991 230	94.5
1999-2004	1 443 691	1 286 329	1 444 584	112.3
2004-2009	1 795 345	1 599 652	1 522 578	95.2
2009-2014	2 424 107	2 159 879	2 154 410	99.7

Cuadro 9

Tipos de enlace por año

	No enlazados			Enlazados		
Año	V. neg.	F. neg.	Total	F. pos.	V. pos.	**Total**
2009	86.0	14.0	100.0	10.0	90.0	**100.0**
2014	89.1	10.9	100.0	10.2	89.8	**100.0**

Por otra parte, el porcentaje de falsos positivos fue de 10% para los CE 2009 y 10.2% para la edición 2014. Este es el precio que se paga por obtener un número alto de enlaces. Una forma de disminuirlo es incrementando las restricciones en algunas de las 10 fases del algoritmo; por ejemplo, requiriendo porcentajes de similitud superiores en el nombre del establecimiento o entidad legal. La desventaja al hacer esto es que se incrementará el porcentaje de falsos negativos, dado que algunos establecimientos que antes eran correctamente enlazados ya no cumplirán con los nuevos criterios.

La otra cara de la moneda cuando hablamos de la precisión del algoritmo son los verdaderos y falsos negativos, es decir, los que no fueron enlazados. Dejar fuera a establecimientos que debieron ser enlazados implica una sobrestimación de la salida de unidades económicas del mercado. Según el cuadro 9, 14% de los establecimientos no enlazados en los Censos 2009 sí reaparecieron en el levantamiento 2014. En términos absolutos, no fueron enlazados 220 mil que efectivamente sobrevivieron, pero al mismo tiempo se tiene un número similar de falsos positivos, por lo que la cantidad de enlaces totales se mantiene similar entre la CLEE y el algoritmo (ver cuadro A1 en el *Apéndice*).

Los porcentajes mostrados en el cuadro 9 son agregados, incluyendo establecimientos de todos los tamaños, clases de actividad y estados de México; además, consideran todas las fases del algoritmo y no todas son igualmente precisas; podemos desagregar los porcentajes para saber si

existe una diferencia sistemática en la precisión del algoritmo dado el tamaño, clase de actividad, estado o fase.

5.3 Precisión por tamaño

Los cuadros 10 y 11 reportan que la precisión del algoritmo (porcentaje de verdaderos positivos) incrementa con el tamaño. Esto significa que si el algoritmo predice que dos establecimientos grandes son el mismo es prácticamente cierto que lo son. El riesgo de falsos positivos es mayor para los pequeños. Por otra parte, el porcentaje de falsos negativos es también mayor para los grandes, es decir, es más difícil enlazarlos. La mortalidad de establecimientos grandes podría estar sobrestimada. Note que los porcentajes agregados se asemejan a los de los establecimientos de menos de cinco trabajadores, que forman 95% del total.

5.4 Precisión por sector de actividad

Los cuadros A2 y A4 en el *Apéndice* muestran los porcentajes de tipos de enlace desglosados por sectores a dos dígitos del SCIAN. En 2009, la clase con el menor porcentaje de verdaderos positivos es 55 Servicios Administrativos; sin embargo, está formada por solo 204 establecimientos, por lo que no tiene un gran impacto en la precisión agregada. El sector manufacturero, comúnmente utilizado en la literatura, es muy preciso, con menos de 10% de falsos positivos. Los sectores 11, 21, 22, 23 y 55 muestran niveles altos de falsos negativos, lo que implica una sobrestimación en la mortalidad de dichas unidades; sin embargo, representan solo 1.2% del levantamiento censal, por lo que tienen poco impacto en la precisión agregada (ver cuadros A3 y A5 en el *Apéndice*). El resto de los sectores presentan poca variación en los porcentajes respecto al agregado.

5.5 Precisión por estado

Los cuadros A6 y A7 en el *Apéndice* muestran los porcentajes de tipos de enlace desglosados por estado. El de verdaderos positivos no presenta gran dispersión por entidad, manteniéndose en un rango que va de 87 a 92%; esto significa que la precisión del algoritmo es similar para todas las regiones del país. Por otra parte, los falsos negativos presentan mayor dispersión. En particular, el algoritmo sobrestima la mortalidad de unidades en la Ciudad de México (CDMX). Esto podría ocurrir porque es una zona densamente poblada y establecimientos muy similares se concentran en espacios pequeños, creando ambigüedades difíciles de resolver.

Cuadro 10

Tipo de enlace por tamaño, 2009

Trabajadores	No enlazados			Enlazados		
	V. neg.	F. neg.	Total	F. pos.	V. pos.	Total
[0-10]	86.3	13.7	**100.0**	10.1	89.9	**100.0**
[11-50]	79.6	20.4	**100.0**	7.6	92.4	**100.0**
[50-100]	69.4	30.6	**100.0**	5.3	94.7	**100.0**
> 100	63.3	36.7	**100.0**	5.4	94.6	**100.0**
Total	**86.0**	**14.0**	**100.0**	**10.0**	**90.0**	**100.0**

Cuadro 11

Tipo de enlace por tamaño, 2014

Trabajadores	No enlazados			Enlazados		
	V. neg.	F. neg.	Total	F. pos.	V. pos.	Total
[0-10]	89.4	10.6	**100.0**	10.4	89.6	**100.0**
[11-50]	83.8	16.2	**100.0**	8.2	91.8	**100.0**
[50-100]	74.8	25.2	**100.0**	5.8	94.2	**100.0**
> 100	69.0	31.0	**100.0**	4.9	95.1	**100.0**
Total	**89.1**	**10.9**	**100.0**	**10.2**	**89.8**	**100.0**

5.6 Precisión por fase

Como se anticipó, no todas las fases del algoritmo tienen la misma precisión (porcentaje de verdaderos

positivos). Como se muestra en el cuadro 12, la precisión es superior a 90% de las fases 2 a 6 (para la 1 será de 100% porque la combinación NIC-NOP es redundante con la CLEE). Las últimas cuatro fases tienen niveles inferiores de verdaderos positivos; sin embargo, forman solo 7.8% de los enlaces y su impacto en los porcentajes acumulados es limitado. Dichas fases se incluyen porque completan la cobertura de enlaces y mantienen los criterios de continuidad de la OCDE.

Cabe mencionar que 10% de falsos positivos no es una asignación aleatoria de enlaces; dicho de otra forma, no se enlazará un establecimiento de Walmart con una pequeña ferretería. Si bien el algoritmo enlaza establecimientos que la CLEE no, estos aun así mantienen gran similitud de ubicación, razón social y clase de actividad. Aunque algunos enlaces no correspondan al mismo establecimiento en la realidad, esto puede no ser problemático para efectos estadísticos ya que se trata de unidades muy similares. Desafortunadamente, no se puede saber para periodos anteriores al levantamiento 2009 cuáles son falsos positivos (al menos no por procedimientos computacionales), solo es posible conocer que podrían ser alrededor de 10% de los enlaces como en 2009-2014.

Cuadro 12

Tipo de *match* por fase

Fase	F. pos.	V. pos.	Total	% V. pos.
1	0	0	**0**	-
2	85 527	973 927	**1 059 454**	91.9
3	17 673	287 094	**304 767**	94.2
4	29 680	331 617	**361 297**	91.8
5	16 996	186 513	**203 509**	91.6
6	5 290	51 866	**57 156**	90.7
7	6 358	54 913	**61 271**	89.6
8	45 834	47 076	**92 910**	50.7
9	6 819	4 376	**11 195**	39.1
10	777	2 074	**2 851**	72.7
Total	**214 954**	**1 939 456**	**2 154 410**	**90.0**

6. Medidas de flujo de trabajadores y establecimientos

Otra forma de evaluar la calidad del algoritmo es estimando con este y la CLEE tanto medidas de flujo de trabajadores como entrada y salida de establecimientos. Para estimar estas medidas (anualizadas), se sigue el método de Miranda y Jarmin (2002), y se definen como sigue.

6.1 Índices de creación y destrucción de empleo

$$JCR = \frac{JC}{X}$$

donde:

$$JC = E_{t+5} - E_t$$

donde E denota el empleo en establecimientos que se expanden y recién nacidos y X es el empleo promedio de t y $t + 5$. El índice de destrucción de empleo (JDR) se calcula análogamente, pero E es el empleo de los establecimientos que se contraen y los que salen del mercado.

6.2 Índice de entrada y salida de establecimientos

$$\text{Índice de entrada} = \frac{ENTRY}{AVG}$$

donde $ENTRY$ es el número de establecimientos entrantes en $t + 5$ y AVG, el número de unidades económicas promedio entre t y $t + 5$. El índice de salida es similar, pero se reemplaza el número de entrantes por el de salientes ($EXIT$):

$$\text{Índice de salida} = \frac{EXIT}{AVG} .$$

Las últimas dos filas del cuadro 13 muestran que los índices de entrada y salida tienen, prácticamente, los mismos valores para los dos métodos de enlace; sin embargo, los cálculos a partir de los enlaces del algoritmo sobrestiman ligeramente las ta-

sas de creación (JC) y destrucción de empleo (JD). Vale la pena notar que para periodos de la misma duración, por ejemplo, 1999-2004 y 2004-2009, o bien, 1999-2009 y 2004-2014, los índices mantienen el orden de magnitud.

Cuadro 13

Índices de entrada, salida, creación y destrucción de empleo

Enlace	Periodo	Entrada	Salida	C. de emp.	D. de emp.
Algoritmo	1999-2004	9.2	15.4	9.3	13.6
	1999-2009	6.6	9.1	6.4	8.4
	1999- 2014	4.8	6.6	4.7	6.4
	2004-2009	11.6	12.7	11.3	11.8
	2004-2014	6.9	8.2	6.6	9.8
	2009-2014	9.4	11.2	11.0	13.5
CLEE	2009-2014	9.4	11.1	9.8	12.7

7. Discusión

La primera advertencia sobre este proceso de vinculación es que no toma en cuenta la reorganización de las unidades económicas como fusiones o particiones. Si ocurre un gran número de fusiones de t a $t + 5$, podríamos estar sobrestimando la salida de establecimientos. Por el contrario, si ocurrieron muchas particiones, se sobrestima el nacimiento de unidades. Desafortunadamente, se tiene poca información acerca de estos fenómenos para ser tomados en cuenta en el procedimiento de enlace.

Además, no se realizan ejercicios para tratar de enlazar establecimientos entre CE no consecutivos. Si un establecimiento se reportó en el levantamiento 1999 y reapareció hasta el de 2009 (por inactividad o falta de registro en 2004), este no será enlazado. Este tipo de casos podrían sobrestimar la salida de unidades económicas del mercado.

Por último, no utilizamos tablas de equivalencia para armonizar recodificaciones de clases de actividad y códigos geográficos. A pesar de que los cambios son mínimos entre levantamientos censales consecutivos, podríamos estar clasificando establecimientos como falsos negativos cuando debieron enlazarse.

8. Laboratorio de Microdatos del INEGI

Los Censos Económicos a nivel establecimiento se consideran información confidencial por el INEGI. La única forma de trabajar estos datos es dentro de su Laboratorio de Microdatos, con sede en la Ciudad de México. Si algún investigador está interesado en utilizar los identificadores descritos en este trabajo, tiene que hacer un requerimiento especial y solicitar se incluyan los identificadores del BID en las bases de datos.

Fuentes

Busso, M., O. Fentanes & S. Levy Algazi. *The Longitudinal Linkage of Mexico's Economic Census 1999-2014*. (No. IDB-TN-01477). Inter-American Development Bank, 2018.

Christen, P. "A survey of indexing techniques for scalable record linkage and deduplication", en: *IEEE transactions on knowledge and data engineering.* 24 (9), pp. 1537-1555, 2012.

Jarmin, R. S. & J. Miranda. *The longitudinal business database*. Center for Economic Studies, Working Paper, pp. 2-17, 2002.

Organización para la Cooperación y el Desarrollo Económicos (OCDE). *Eurostat-OECD manual on business demography statistics*. OCDE, 2008.

Raffo, J. *Matchit*: *Stata module to match two datasets based on similar text patterns*. 2017.

Apéndice

A Tablas

Cuadro A1

Tipo de enlace por año

	No enlazados			Enlazados		
Año	V. neg.	F. neg.	Total	F. pos.	V. pos.	Total
2009	1 349 264	220 345	**1 569 609**	214 954	1 939 456	**2 154 410**
2014	1 850 319	226 016	**2 076 335**	220 625	1 933 785	**2 154 410**

Cuadro A2

Tipo de enlace por industria, porcentajes en 2009

Sectores	V. neg.	F. neg.	Total	F. pos.	V. pos.	Total
11 Agricultura	74.6	25.4	**100.0**	1.6	98.4	**100.0**
21 Minería	71.0	29.0	**100.0**	3.5	96.5	**100.0**
22 Utilidades	54.0	46.0	**100.0**	0.3	99.7	**100.0**
23 Construcción	86.8	13.2	**100.0**	4.9	95.1	**100.0**
31 Manufacturas	86.4	13.6	**100.0**	8.1	91.9	**100.0**
32 Manufacturas	87.9	12.1	**100.0**	9.6	90.4	**100.0**
33 Manufacturas	88.2	11.8	**100.0**	8.9	91.1	**100.0**
43 Mayoristas	86.4	13.6	**100.0**	9.5	90.5	**100.0**
46 Minoristas	86.1	13.9	**100.0**	11.1	88.9	**100.0**
48 Transportes	87.8	12.2	**100.0**	11.7	88.3	**100.0**
49 Transportes	92.8	7.2	**100.0**	13.4	86.6	**100.0**
51 Información	89.8	10.2	**100.0**	10.6	89.4	**100.0**
52 Finanzas	92.3	7.7	**100.0**	14.7	85.3	**100.0**
53 Bienes Raíces	93.9	6.1	**100.0**	7.6	92.4	**100.0**
54 Profesionales	91.6	8.4	**100.0**	13.1	86.9	**100.0**
55 Gerencia	64.6	35.4	**100.0**	16.4	83.6	**100.0**
56 Apoyo	92.7	7.3	**100.0**	14.8	85.2	**100.0**
61 Educación	92.0	8.0	**100.0**	7.6	92.4	**100.0**
62 Salud	92.6	7.4	**100.0**	10.7	89.3	**100.0**
71 Entretenimiento	93.3	6.7	**100.0**	9.7	90.3	**100.0**
72 Alimentación	74.4	25.6	**100.0**	4.0	96.0	**100.0**
81 Otros	90.2	9.8	**100.0**	10.1	89.9	**100.0**
Total	**86.0**	**14.0**	**100.0**	**10.0**	**90.0**	**100.0**

Cuadro A3

Tipo de enlace por industria, establecimientos en 2009

Sectores	V. neg.	F. neg.	Total no enlazados	F. pos.	V. pos.	Total enlazados
11 Agricultura	5 994	2 040	**8 034**	184	11 225	**11 409**
21 Minería	1 057	431	**1 488**	51	1 418	**1 469**
22 Utilidades	154	131	**285**	6	2 298	**2 304**
23 Construcción	9 159	1 389	**10 548**	395	7 694	**8 089**
31 Manufacturas	77 899	12 234	**90 133**	11 819	133 410	**145 229**
32 Manufacturas	31 899	4 401	**36 300**	4 664	43 701	**48 365**
33 Manufacturas	44 322	5 949	**50 271**	5 945	60 608	**66 553**
43 Mayoristas	44 323	6 953	**51 276**	6 363	60 389	**66 752**
46 Minoristas	587 281	94 567	**681 848**	117 966	940 708	**1 058 674**
48 Transportes	6 662	926	**7 588**	900	6 776	**7 676**
49 Transportes	1 891	147	**2 038**	54	349	**403**
51 Información	7 254	824	**8 078**	348	2 928	**3 276**
52 Finanzas	8 224	690	**8 914**	1 442	8 350	**9 792**
53 Bienes Raíces	24 875	1 605	**26 480**	2 107	25 601	**27 708**
54 Profesionales	36 727	3 364	**40 091**	5 865	38 739	**44 604**
55 Gerencia	53	29	**82**	20	102	**122**
56 Apoyo	42 372	3 339	**45 711**	5 220	29 991	**35 211**
61 Educación	17 443	1 508	**18 951**	1 860	22 475	**24 335**
62 Salud	48 133	3 873	**52 006**	10 084	84 442	**94 526**
71 Entretenimiento	20 765	1 501	**22 266**	1 893	17 662	**19 555**
72 Alimentación	162 661	56 028	**218 689**	6 944	166 609	**173 553**
81 Otros	170 116	18 416	**188 532**	30 824	273 981	**304 805**
Total	**1 349 264**	**220 345**	**1 569 609**	**214 954**	**1 939 456**	**2 154 410**

Cuadro A4

Tipo de enlace por industria, porcentajes en 2014

Sectores	V. neg.	F. neg.	Total	F. pos.	V. pos.	Total
11 Agricultura	75.9	24.1	**100.0**	1.4	98.6	**100.0**
21 Minería	72.3	27.7	**100.0**	3.6	96.4	**100.0**
22 Utilidades	68.1	31.9	**100.0**	0.3	99.7	**100.0**
23 Construcción	85.4	14.6	**100.0**	4.5	95.5	**100.0**
31 Manufacturas	91.3	8.7	**100.0**	8.5	91.5	**100.0**
32 Manufacturas	89.6	10.4	**100.0**	9.6	90.4	**100.0**
33 Manufacturas	91.7	8.3	**100.0**	9.3	90.7	**100.0**
43 Mayoristas	89.8	10.2	**100.0**	10.1	89.9	**100.0**
46 Minoristas	88.6	11.4	**100.0**	11.3	88.7	**100.0**
48 Transportes	87.8	12.2	**100.0**	11.7	88.3	**100.0**
49 Transportes	85.8	14.2	**100.0**	9.6	90.4	**100.0**
51 Información	92.0	8.0	**100.0**	10.2	89.8	**100.0**
52 Finanzas	94.4	5.6	**100.0**	14.7	85.3	**100.0**
53 Bienes Raíces	94.9	5.1	**100.0**	7.9	92.1	**100.0**
54 Profesionales	93.8	6.2	**100.0**	13.4	86.6	**100.0**
55 Gerencia	83.3	16.7	**100.0**	24.2	75.8	**100.0**
56 Apoyo	94.1	5.9	**100.0**	14.9	85.1	**100.0**
61 Educación	93.6	6.4	**100.0**	7.8	92.2	**100.0**
62 Salud	94.3	5.7	**100.0**	11.1	88.9	**100.0**
71 Entretenimiento	94.5	5.5	**100.0**	9.7	90.3	**100.0**
72 Alimentación	82.3	17.7	**100.0**	4.1	95.9	**100.0**
81 Otros	92.2	7.8	**100.0**	10.6	89.4	**100.0**
Total	**89.1**	**10.9**	100.0	**10.2**	**89.8**	100.0

Cuadro A5

Tipo de enlace por industria, establecimientos en 2014

Sectores	V. neg.	F. neg.	Total no enlazados	F. pos.	V. pos.	Total enlazados
11 Agricultura	6 831	2 168	**8 999**	165	11 243	**11 408**
21 Minería	1 130	432	**1 562**	53	1 417	**1 470**
22 Utilidades	284	133	**417**	8	2 296	**2 304**
23 Construcción	7 667	1 307	**8 974**	361	7 728	**8 089**
31 Manufacturas	117 919	11 241	**129 160**	12 377	132 948	**145 325**
32 Manufacturas	36 285	4 219	**40 504**	4 591	43 265	**47 856**
33 Manufacturas	54 733	4 986	**59 719**	6 212	60 754	**66 966**
43 Mayoristas	60 178	6 857	**67 035**	6 400	56 913	**63 313**
46 Minoristas	753 618	96 562	**850 180**	120 426	941 687	**1 062 113**
48 Transportes	7 721	1 073	**8 794**	903	6 791	**7 694**
49 Transportes	958	158	**1 116**	37	348	**385**
51 Información	5 532	482	**6 014**	339	2 985	**3 324**
52 Finanzas	13 207	788	**13 995**	1 438	8 328	**9 766**
53 Bienes Raíces	33 251	1 793	**35 044**	2 182	25 589	**27 771**
54 Profesionales	42 038	2 770	**44 808**	5 969	38 477	**44 446**
55 Gerencia	194	39	**233**	30	94	**124**
56 Apoyo	53 551	3 340	**56 891**	5 190	29 530	**34 720**
61 Educación	21 629	1 473	**23 102**	1 866	21 914	**23 780**
62 Salud	71 964	4 368	**76 332**	10 485	84 120	**94 605**
71 Entretenimiento	28 564	1 664	**30 228**	1 954	18 210	**20 164**
72 Alimentación	270 039	57 902	**327 941**	7 192	166 315	**173 507**
81 Otros	263 026	22 261	**285 287**	32 447	272 833	**305 280**
Total	**1 850 319**	**226 016**	**2 076 335**	**220 625**	**1 933 785**	**2 154 410**

Cuadro A6

Tipo de enlace por estado, porcentajes en 2009

Estado	V. neg.	F. neg.	Total no enlazados	F. pos.	V. pos.	Total enlazados
Aguascalientes	87.6	12.4	**100.0**	10.7	89.3	**100.0**
Baja California	87.5	12.5	**100.0**	12.8	87.2	**100.0**
Baja California Sur	88.3	11.7	**100.0**	9.5	90.5	**100.0**
Campeche	85.8	14.2	**100.0**	8.5	91.5	**100.0**
Coahuila de Z.	91.5	8.5	**100.0**	10.0	90.0	**100.0**
Colima	85.9	14.1	**100.0**	8.5	91.5	**100.0**
Chiapas	84.5	15.5	**100.0**	11.8	88.2	**100.0**
Chihuahua	89.1	10.9	**100.0**	8.8	91.2	**100.0**
CDMX	79.6	20.4	**100.0**	10.7	89.3	**100.0**
Durango	88.7	11.3	**100.0**	7.6	92.4	**100.0**
Guanajuato	84.9	15.1	**100.0**	10.2	89.8	**100.0**
Guerrero	86.4	13.6	**100.0**	11.5	88.5	**100.0**
Hidalgo	86.4	13.6	**100.0**	10.0	90.0	**100.0**
Jalisco	86.4	13.6	**100.0**	9.8	90.2	**100.0**
México	85.1	14.9	**100.0**	10.2	89.8	**100.0**
Michoacán de O.	87.4	12.6	**100.0**	10.3	89.7	**100.0**
Morelos	88.2	11.8	**100.0**	10.8	89.2	**100.0**
Nayarit	84.7	15.3	**100.0**	7.1	92.9	**100.0**
Nuevo León	90.1	9.9	**100.0**	10.4	89.6	**100.0**
Oaxaca	84.4	15.6	**100.0**	9.2	90.8	**100.0**
Puebla	85.6	14.4	**100.0**	10.2	89.8	**100.0**
Querétaro	85.2	14.8	**100.0**	10.5	89.5	**100.0**
Quintana Roo	86.2	13.8	**100.0**	11.7	88.3	**100.0**
San Luis Potosí	88.3	11.7	**100.0**	9.1	90.9	**100.0**
Sinaloa	87.0	13.0	**100.0**	7.5	92.5	**100.0**
Sonora	88.0	12.0	**100.0**	8.1	91.9	**100.0**
Tabasco	85.4	14.6	**100.0**	11.9	88.1	**100.0**
Tamaulipas	89.5	10.5	**100.0**	9.0	91.0	**100.0**
Tlaxcala	87.1	12.9	**100.0**	10.6	89.4	**100.0**
Veracruz de I. de la Ll.	87.0	13.0	**100.0**	9.3	90.7	**100.0**
Yucatán	84.4	15.6	**100.0**	7.9	92.1	**100.0**
Zacatecas	89.8	10.2	**100.0**	7.8	92.2	**100.0**
Total	**86.0**	**14.0**	100.0	**10.0**	**90.0**	100.0

Cuadro A7

Tipo de enlace por estado, porcentajes en 2014

Estado	V. neg.	F. neg.	Total no enlazados	F. pos.	V. pos.	Total enlazados
Aguascalientes	90.7	9.3	**100.0**	10.9	89.1	**100.0**
Baja California	90.9	9.1	**100.0**	13.1	86.9	**100.0**
Baja California Sur	91.9	8.1	**100.0**	9.7	90.3	**100.0**
Campeche	88.0	12.0	**100.0**	8.6	91.4	**100.0**
Coahuila de Z.	92.0	8.0	**100.0**	10.2	89.8	**100.0**
Colima	88.6	11.4	**100.0**	8.8	91.2	**100.0**
Chiapas	89.0	11.0	**100.0**	12.4	87.6	**100.0**
Chihuahua	90.7	9.3	**100.0**	8.9	91.1	**100.0**
CDMX	82.7	17.3	**100.0**	11.1	88.9	**100.0**
Durango	91.0	9.0	**100.0**	7.9	92.1	**100.0**
Guanajuato	90.2	9.8	**100.0**	10.8	89.2	**100.0**
Guerrero	88.1	11.9	**100.0**	11.5	88.5	**100.0**
Hidalgo	90.6	9.4	**100.0**	10.4	89.6	**100.0**
Jalisco	90.4	9.6	**100.0**	10.1	89.9	**100.0**
México	89.0	11.0	**100.0**	10.6	89.4	**100.0**
Michoacán de O.	90.0	10.0	**100.0**	10.4	89.6	**100.0**
Morelos	89.8	10.2	**100.0**	10.6	89.4	**100.0**
Nayarit	89.4	10.6	**100.0**	7.5	92.5	**100.0**
Nuevo León	91.2	8.8	**100.0**	10.2	89.8	**100.0**
Oaxaca	89.8	10.2	**100.0**	9.6	90.4	**100.0**
Puebla	89.5	10.5	**100.0**	10.6	89.4	**100.0**
Querétaro	90.2	9.8	**100.0**	10.8	89.2	**100.0**
Quintana Roo	89.5	10.5	**100.0**	11.7	88.3	**100.0**
San Luis Potosí	90.8	9.2	**100.0**	9.2	90.8	**100.0**
Sinaloa	90.5	9.5	**100.0**	7.7	92.3	**100.0**
Sonora	89.8	10.2	**100.0**	8.4	91.6	**100.0**
Tabasco	88.5	11.5	**100.0**	12.2	87.8	**100.0**
Tamaulipas	90.2	9.8	**100.0**	8.8	91.2	**100.0**
Tlaxcala	90.6	9.4	**100.0**	11.0	89.0	**100.0**
Veracruz de I. de la Ll.	88.4	11.6	**100.0**	9.5	90.5	**100.0**
Yucatán	88.4	11.6	**100.0**	8.1	91.9	**100.0**
Zacatecas	91.3	8.7	**100.0**	8.0	92.0	**100.0**
Total	**89.1**	**10.9**	100.0	**10.2**	**89.8**	100.0

B Códigos de corrección de inflexiones

Instalación

Los comandos se definen en los archivos *estandariza.ado* y *separa.ado*. Para instalarlos, basta depositar los ADO en el directorio de STATA, donde se alojan los comandos. Para identificar este directorio, se puede ejecutar el comando *sysdir list*; la carpeta está etiquetada como PLUS.

Una vez que se depositan los ADO en el directorio PLUS, es necesario definir una hoja de Excel para el comando a utilizar. Para *estandariza.ado*, se necesitan dos columnas, una con las cadenas de caracteres a estandarizar y una segunda columna con su versión estandarizada. Para el comando *separa.ado* solo se necesita una columna, es decir, una que contenga en cada celda las cadenas de caracteres a identificar y separar.

Sintaxis

Comando *estandariza*

estandariza var, gen() dexcel() sheet() space()

Opciones:

gen( ) Nombre de la variable que contiene la versión estandarizada de var.

dexcel( ) Directorio y nombre de la hoja de Excel. Ejemplo: "C:/…/doc.xlsx".

sheet( ) Hoja dentro del documento de Excel. Ejemplo: "Hoja1".

space( ) Puede tomar los valores ***y*** y ***n***. Si se selecciona *y*, se estandarizan las cadenas de texto solo cuando formen una palabra independiente. Si se selecciona *n*, se estandariza sin importar en qué parte de la cadena de texto aparezca.

Comando *separa*

gen(*) Genera dos variables: *_1 y *_2. *_1 es la parte no especificada en el Excel. *_2 es alguna de las cadenas de texto especificadas en Excel.

dexcel( ) Directorio y nombre de la hoja de Excel. Ejemplo: "C:/…/doc.xlsx".

sheet( ) Hoja dentro del documento de Excel. Ejemplo: "Hoja1".

space( ) Puede tomar los valores ***y*** y ***n***. Si se selecciona *y*, se estandarizan las cadenas de texto solo cuando formen una palabra independiente. Si se selecciona *n*, se estandariza no importa en qué parte de la cadena de texto aparezca.