



Universidad de
SanAndrés

PROPUESTA DE INVESTIGACIÓN

PROFESORA: NOELIA ROMERO

**Un empalme de clasificadores
industriales para el mercado laboral
uruguayo**

Alexander Pacheco, Rafael Paganini, Alan Starobinski

Fecha: 3/12/2023

1. Introducción

En un mundo en que la integración económica y mayores consensos en materia de políticas públicas han inducido procesos de convergencia institucional entre economías con pasados muy disímiles, las instituciones laborales siguen siendo una fuente clave de heterogeneidad a nivel global. Aunque estas instituciones están intrínsecamente vinculadas a preguntas fundamentales en economía y son relevantes en debates contemporáneos, la investigación cuantitativa se ha centrado mayormente en el mercado laboral estadounidense, aislando efectos de cambios institucionales específicos, como el salario mínimo, y enfrentando desafíos para aislar variabilidad exógena en la legislación laboral, particularmente en economías emergentes (Dube et al., 2020; Bertrand et al., 2015; Cengiz et al., 2019, 2022; Dube and Lindner, 2021). La limitada disponibilidad de datos en economías emergentes ha restringido la contribución cuantitativa de la literatura económica a debates vigentes sobre políticas salariales y laborales fuera de los contextos más desarrollados.

Esta propuesta de investigación busca contribuir a cerrar la brecha de datos entre países desarrollados y emergentes, utilizando técnicas de *web-scraping* y *machine learning* para construir datos claves, y hasta ahora no disponibles, para el estudio de las dinámicas del mercado laboral uruguayo. El estudio del mercado laboral, así como de la coyuntura macroeconómica, en Uruguay enfrenta un obstáculo importante: una falta de datos sistematizados respecto de los ajustes salariales fijados por mecanismos institucionales. Los salarios en el sector formal uruguayo tienen ajustes mínimos, dispuestos por convenios laborales a nivel sectorial firmados en negociaciones tripartitas en el marco de los Consejos de Salarios. Pese a la relevancia de estos acuerdos para el pronóstico de salarios, empleo e inflación, no existen bases de datos sobre los ajustes dispuestos en el pasado para cada rama económica.

2. Literatura Previa

En el cambio de siglo, las discusiones sobre las instituciones del mercado laboral se volvieron centrales en las economías avanzadas, desplazando la atención de la política macroeconómica. Los análisis de las décadas pasadas comparaban arreglos neo-corporativos con centralización salarial y descentralización, concluyendo que ambos extremos tenían desempeños más favorables (Freeman, 2007; Calmfors and Driffill, 1988).

Estas discusiones, centradas en países desarrollados como Estados Unidos, Japón y Europa, pasaron por alto a América Latina, con escasa atención debido a la falta de ejemplos de neo-corporativismo en la región. Con el cambio de milenio, el neo-corporativismo perdió relevancia en Europa Occidental, mientras que características tradicionales emergieron en países periféricos latinoamericanos ([Etchemendy, 2019](#)).

En contextos de países en desarrollo, la escasez y falta de conexión de datos plantea desafíos para el análisis. En un trabajo inicial, ([Perazzo, 2012](#)) propuso una herramienta de evaluación que vincula los convenios colectivos del Consejo de Salarios (CS) con la Encuesta Continua de Hogares (ECH) para analizar la implementación efectiva de acuerdos y los impactos distributivos de los cambios salariales. Sin embargo, se enfrentó a desafíos técnicos, ya que los criterios de clasificación entre las fuentes de información mostraron notables discrepancias.

([Cabrera and Cárpena, 2013](#)) ofrecieron mejoras, destacando el uso de salarios por hora para comparar laudos pagados y recibidos. A pesar de ello, señalan la falta de concordancia entre los grupos y subgrupos de los Consejos de Salarios (CS) y la Clasificación Industrial Internacional Uniforme (CIIU), evidenciando superposiciones y discrepancias entre subgrupos y ramas de actividad.

Nuestro trabajo viene a contribuir en mejorar la metodología de empalme de los subgrupos del CS y las categorías del CIIU.

3. Entorno institucional

Los Consejos de Salarios surgieron en Uruguay en 1943 con la implementación de la Ley 10.449, estableciendo ámbitos de negociación salarial a nivel de rama de actividad mediante una estructura tripartita. Tras interrupciones en su aplicación durante el gobierno de facto (1973-1985) y en el período 1992-2004, los Consejos de Salarios volvieron a aplicarse a partir de 2005.

En el marco de los Consejos de Salarios se firman convenios colectivos, que establecen, además de pautas de condiciones laborales, aumentos porcentuales a aplicar sobre los salarios brutos, junto a correctivos por inflación que aseguran el cumplimiento de metas para el salario real. En 2005 se crearon 23 grupos de actividad económica y se clasificaron las empresas privadas en un total de 186 subgrupos ([Perazzo, 2012](#)), una vez que se tiene en cuenta a la inclusión de trabajadores rurales. Aunque la cantidad de grupos ha permanecido constante (excepto por la inclusión del trabajo doméstico en 2008), la cantidad de subgrupos ha aumentado ligeramente desde 2005. La mayoría de los convenios aplica al nivel de subgrupo, aunque existen casos de convenios negociados a nivel de grupo o capítulo (una categoría más desagregada que la de subgrupo).

Las agencias encargadas de publicar datos en Uruguay, principalmente el INE y el BCU, utilizan la Clasificación Internacional Industrial Uniforme (CIIU) para informar sobre datos desagregados por rama económica. Sin embargo, esta práctica presenta desafíos en la asignación entre categorías de negociación salarial y los sectores económicos afectados por estas negociaciones. [Perazzo \(2012\)](#) y [Cabrera and Cárpena \(2013\)](#) han llevado a cabo el ejercicio de empalmar ambas clasificaciones, pero la correspondencia se establece de manera cualitativa, lo que no permite en la evaluación y contraste de supuestos utilizados en la correspondencia sin replicar el estudio realizado. Además, no proporciona información detallada sobre la incidencia de cada subgrupo de los Consejos de Salarios en cada categoría CIIU en casos donde los subgrupos asignados a una categoría son diversos.

4. Datos y fuentes

Las dos partes del ejercicio propuesto se nutren de fuentes de información diferentes. En primer lugar, se plantea construir series de ajustes salariales obligatorios en el pasado, de acuerdo con todos los convenios sectoriales firmados, una base de datos que no existe de manera sistematizada y públicamente disponible. Para esto, se

utilizará información que sí está al alcance del público general: las actas de ajuste. La página *web* del Ministerio de Trabajo y Seguridad Social (MTSS)¹ permite acceder no solamente a los convenios colectivos firmados hasta la fecha, sino también a las llamadas “actas de ajuste”. Éstas últimas constituyen documentos de ratificación de lo acordado previamente, firmadas en ocasión de cada ajuste salarial dispuesto por convenio, y son algo redundantes desde un punto de vista legal pero útiles a efectos de desambiguar implicancias numéricas de los ajustes dispuestos en los convenios, de manera de facilitar su aplicación el sector económico correspondiente. Los ajustes por inflación, junto con algunos ejemplos especiales de ajustes que tuvieron en cuenta la evolución del PIB (u otros indicadores de actividad) a nivel nacional o sectorial, están sujetos a la evolución de los indicadores correspondientes y por tanto no se explicitan en formato numérico en el documento de convenio elaborado *ex ante*. En las actas de ajuste, en cambio, aparece el porcentaje final de ajuste salarial que aplicará a todas las empresas del sector, en cada fecha en que un ajuste corresponde.

Una descripción más detallada de la estructura del código HTML de la página, que será utilizado para la aplicación de *web scraping*, será dada en la sección de **Metodología de *web scraping***.

En segundo lugar, se utilizará una base de datos administrativos del Banco de Previsión Social (BPS), ente público a cargo del sistema de seguridad social en Uruguay. La base de datos contiene información mensual a nivel de individuo, desde enero de 1996, para trabajadores formales. Más allá de sexo, nacionalidad y fecha de nacimiento, todas las características individuales son referidas a la relación laboral. La base cuenta con una variable identificadora a nivel de empresa, que permite considerar a todos los individuos para una empresa dada. Los otros datos a nivel de empresa disponibles son la cantidad de dependientes y la categoría industrial a 5 dígitos según la CIIU (Revisión 4). Los datos laborales para el individuo incluyen el tipo de aportación, días y horas de trabajo, la fecha de ingreso a la empresa, el vínculo funcional y diferentes tipos de remuneración, distinguiendo (entre otras cosas) salario bruto de salario neto de impuestos.

¹<https://www.gub.uy/ministerio-trabajo-seguridad-social/tematica/consejos-salarios-negociacion-colectiva>

Para el ejercicio de asignación de empresas a subgrupos o capítulos de los Consejos de Salarios, se realizarán algunas transformaciones a la base y luego se construirán series mensuales desde enero de 2005 (primer año de aplicabilidad del mecanismo de fijación salarial tripartita en su versión actual). Primeramente, se eliminarán los individuos cuyo tipo de aportación es Monotributo o Monotributo Social Mides (Ministerio de Desarrollo Social), así como los que tienen un vínculo funcional con la empresa que no es de dependencia. Se eliminarán asimismo a las empresas que tengan datos faltantes para algunos trabajadores, en caso de haberlos, a efectos de evitar sesgar el cálculo de los incrementos salariales a nivel de empresa. Si estos datos superan el 1% de la muestra, se repetirá el procedimiento descrito en la sección de **Metodología de empalme** eliminando solamente a los trabajadores que presenten algún dato faltante en la variable de remuneración, a efectos de comparar con los resultados obtenidos inicialmente, como chequeo de robustez.

Luego de esto, se calculará, para cada trabajador de la muestra, la variación porcentual de su salario bruto (sobre el que aplica lo dispuesto en los convenios colectivos). Con esto, se construirán series de media, a nivel de empresa, de incrementos salariales porcentuales, excluyendo los trabajadores que perciben el Salario Mínimo Nacional (SMN) en cada período (construyendo manualmente una serie de los SMN a partir de los decretos de actualización, disponibles en la página del gobierno nacional ². La razón de esta exclusión es que los incrementos del SMN típicamente son superiores a los laudados en el marco de los Consejos de Salarios, volviendo inactivos los ajustes salariales por convenio, que no se acumulan con los del SMN. También se construirá una serie con la variación porcentual de la cantidad total de puestos de trabajo activos en el resto de las empresas de la misma categoría CIIU a 5 dígitos, a efectos de poder luego controlar por dinámicas de mercado a nivel sectorial que puedan afectar la propensión a ajustar por encima de los laudos fijados en los convenios colectivos. A efectos de evitar anomalías asociadas a empresas de corta existencia, se eliminarán los datos correspondientes a empresas que figuran por menos de un año en la base de datos.

²A modo de ejemplo, el siguiente enlace permite ver el decreto más reciente de actualización del SMN: <https://www.gub.uy/ministerio-trabajo-seguridad-social/comunicacion/noticias/salario-minimo-nacional-21107-desde-1o-enero>

5. Metodología de *web scraping*

La construcción de las series de ajuste fijados por convenio colectivo se hará utilizando las técnicas de *web scraping* vistas en el curso, y utilizando *Jupyter Notebooks*. La página del MTSS permite acceder, en primera instancia, a cinco URL diferentes, correspondientes a “Grupos de Industria, Comercio y actividades en general”, “Trabajo doméstico” y “Trabajo rural”, que serán las URL de interés, junto a “Consejo Superior Tripartito” y “Sector público”. Una inspección del código HTML de la página permite ver que asocia diferentes URL a la clase “Box-name” con la etiqueta “h3”, y cada una está enlazada mediante la etiqueta “a” con un texto que indica la descripción de la URL en cuestión (por ejemplo “Grupos de Industria, Comercio y actividades en general”). Esta estructura se repite, pero con un elemento “li” con la clase vacía en lugar de la etiqueta “h3” asociada a “Box-name”, para acceder a las URL correspondientes a cada grupo. El acceso a las URL de los subgrupos y, dentro de estos, a los documentos asociados a cada subgrupo (o a los capítulos, y luego a los documentos, en los casos correspondientes) está nuevamente dispuesto con la estructura con “Box-name” y “h3”. Dentro de estos, los PDF están asociados a elementos “div” con clase “Download-title”.

La estructura anterior permite construir un código de *Python* que, haciendo uso de la librería “BeautifulSoup”, acceda iterativamente a los PDF asociados al nivel más desagregado posible dentro de las tres URL de conjuntos de grupos de interés (correspondientes a trabajo rural, doméstico y general del sector privado, respectivamente). Los PDF de interés serán, en primera instancia, los que corresponden a actas de ajuste, como se discutió en la sección de **Datos y fuentes**. Estas actas tienen un formato sumamente estandarizado, y por lo general contienen un fragmento de texto similar al siguiente: “**SEGUNDO**: Que en virtud de lo establecido, el ajuste a ser aplicado el 1° de julio de 2023, sobre los salarios vigentes al 30 de junio de 2023 es de 4,58%”. Esto permitirá utilizar la biblioteca “request” para extraer la fecha y el monto exacto de ajuste en cada caso. Haciendo los ajustes necesarios al código para contemplar casos anómalos, se construirán así series de tiempo, almacenadas en un *dataframe* de la librería *Pandas*, inicializado con series de tiempo del período considerado con

entradas iguales a cero. Las series resultantes recabarán los ajustes obligatorios para cada mes, tomando el valor 0 en los meses en que no aplica ninguno.

6. Metodología de empalme

Con las series de ajustes salariales obligatorios construidas de acuerdo al proceso descrito en la sección anterior, junto con las series construidas según lo dispuesto en la sección de **Datos y fuentes**, se procederá a realizar regresiones de MCO con regularización LASSO. Cada regresión tendrá como variable dependiente a la media de los ajustes salariales. Las variables explicativas serán todas las series de ajustes obligatorios por subgrupo, la variación de empleos en el resto de la categoría (en casos en que la empresa sea la única de su categoría en alguna parte de la muestra, esta variable se omitirá) y un rezago de la variable dependiente, a efectos de limpiar la regresión de problemas de autocorrelación. Cuando la serie tenga menos datos que regresores, se aplicará un análisis de componentes principales para reducir la dimensionalidad. Se espera que una alta correlación entre series de ajustes salariales que no estén desfasadas (con aumentos mensuales en las mismas fechas), y un aporte explicativo bajo de las series desfasadas, que induciría a LASSO a llevar a prácticamente todos los coeficientes a 0. A su vez, se espera un claro poder explicativo conjunto de las series: en Uruguay, la práctica más habitual es aumentar salarios en la cantidad exacta requerida y en la fecha de aumento dispuesta, sin adelantar incrementos como para suavizar la estructura de costos (al punto de que demoras en negociaciones de nuevos convenios tienen impactos visibles en el Índice Medio de Salarios). Por otra parte, efectos de unos segmentos del mercado laboral sobre otros deberían ocurrir con una temporalidad muy diferente a la que exhiben las series de ajustes (de manera más paulatina, y sin una estacionalidad tan clara), por lo que no se esperaría que induzcan a LASSO a conservar predictores asociados a subgrupos diferentes al de pertenencia de la empresa. Con esto, la expectativa es de que la gran mayoría de las regresiones conserve a un solo predictor, y que este se corresponda con el subgrupo al que verdaderamente pertenece la empresa. El hiperparámetro de regularización λ será elegido por

10-Fold cross-validation, entre un conjunto de valores que oscilarán entre 10^{-10} y 10^{10} . La “predicción” para períodos pasados que ellos implicaría es algo contraintuitiva, pero puede realizarse sin problemas, dado el formato de regresión planteado (no se están incorporando componente ARMA, que inducirían más complicaciones que el simple rezago de la variable dependiente).

Un problema potencial de LASSO es que, en caso de alta correlación entre las series, puede elegir arbitrariamente a una de las series, a razón de la posible no unicidad de la solución inducida por la convexidad no estricta de este tipo de penalización. El supuesto en que descansa la metodología es que incluso pequeñas discrepancias entre la serie de ajustes *correctaz* cualquier otra inducirán a la regularización LASSO a elegir al subgrupo indicado, debido a que esa pequeña discrepancia estará asociada a una clara (aunque leve) mejora en la capacidad predictiva, en un contexto en que se espera escasa variabilidad de los salarios por fuera de los ajustes dispuestos por los convenios colectivos. Esto último se debe a las razones antes explicadas respecto de las prácticas habituales en el mercado laboral uruguayo, junto al hecho de que la variable dependiente construida elimina la variabilidad asociada a efectos de composición (se trata de aumentos porcentuales para trabajadores que estaban en la empresa desde al menos el mes anterior), y no se esperan grandes alteraciones asociadas a aumentos negociados de manera individualizada o promociones.

Dado que la expectativa de conservación de un único predictor, asociada al verdadero subgrupo de pertenencia de una empresa, no está asegurada, en un principio la interpretación de los resultados será la siguiente: el conjunto de subgrupos (o capítulos, en caso de corresponder) asociados a los predictores seleccionados para cada empresa constituye el conjunto de subgrupos relevantes para la determinación de sus salarios. Esta “relevancia” no es necesariamente una indicación de correspondencia entre empresa y conjunto de subgrupos, por varias razones. Dado que cada empresa está asignada solamente a un subgrupo, un conjunto de subgrupos seleccionado compuesto por más de uno no permite la interpretación al estilo “esta empresa pertenece a este subgrupo”. En segundo lugar, incluso en caso de que solamente un subgrupo sea seleccionado, es posible que LASSO haya elegido a uno al que la empresa no pertenece

verdaderamente, por razones de imprecisión de la regresión o de efectos mayores (o de más rápido impacto) de unos segmentos laborales sobre otros.

Por esta razón, se realizarán contrastes de consecuencias observacionales de una asignación correcta, o incorrecta, de empresas a subgrupos. En primera instancia, las empresas para las que se eligió a más de un predictor constituyen casos en que la metodología no cumplió con su cometido original. Se corroborará, entonces, que la proporción de empresas para las que esto efectivamente sucedió es menor al 1 %.

En caso en que el criterio anterior se verifique, se procederá a explorar formas en que la única asignación de predictor puede ser "incorrecta". Si las razones, antes esbozadas, por las que la regularización LASSO tendería a elegir la serie correcta no apliquen, se esperaría que que otras metodologías arrojen resultados sumamente disímiles. En particular, podría darse que, al utilizar una regresión de *Random Forests*, la reducción de impureza en cada nodo con el criterio de Gini, utilizada para medir la importancia de variables explicativas, arroje una importancia relativa de predictores muy diferente a la de la regularización LASSO. Con esta consideración en mente, se llevará adelante una regresión análoga a la mencionada, pero utilizando *Random Forests*, para cada empresa. La medida de importancia de la variable elegida por LASSO, según la regresión de con *Random Forests*, será dividida por la máxima medida de importancia de entre las demás series de ajustes. El promedio simple de estos coeficientes, tomado entre todas las empresas consideradas, debería ser cercano a 0 para favorecer a la interpretación de asignación correcta de empresas a subgrupos³.

Por último, se verificará el porcentaje de empresas para las que el predictor elegido se condice con el subgrupo (o uno de los subgrupos) que le corresponden a la categoría CIIU de la empresa de acuerdo con el empalme de [Cabrera and Cárpene \(2013\)](#) (se

³En principio, sería razonable utilizar regularización *Ridge* para este chequeo de robustez, en vistas de que se comporta mejor que LASSO en situaciones de correlación entre predictores. Sin embargo, la presencia de más de 180 subgrupos, en un período de tiempo de 213 meses si se toma desde febrero de 2005 hasta octubre de 2023, puede inducir una alta presencia de empresas para las que se tienen menos observaciones que predictores, lo que obliga a trabajar con metodologías como *Random Forests* o LASSO (con análisis de componentes principales) que sean aplicables en estos contextos.

tomará la categoría a 4 dígitos, utilizada en ese trabajo). Se interpretará como un ejercicio exitoso un resultado en que más del 95 % de las regresiones, dentro del 99 % o más que preservan solamente un predictor, arrojan un resultado compatible con la asignación cualitativa.

7. Productos finales

En caso en que se obtengan resultados compatibles con la hipótesis de correcta asignación de empresas a subgrupos, se computarán porcentajes de subgrupos asociados para cada categoría de la CIIU (eliminando, primeramente, a las empresas para las que se conservó a más de un predictor de ajuste salarial por convenio). Para esto se computará, para cada categoría, la proporción de trabajadores que pertenecen a empresas que han sido asignadas a cada subgrupo. Una correspondencia análoga pero en dirección contraria, asignando porcentajes de categorías CIIU (con la ventaja de la granularidad de la desagregación a 5 dígitos), se construirá para cada subgrupo de los Consejos de Salarios. Esto se considerará el producto final.

Si alguno de los chequeos descritos arroja resultados no compatibles con la interpretación buscada, se construirán *clusters* de series de ajustes sumamente correlacionadas, junto con series agregadas de variación salarial a nivel de categoría CIIU (eliminando a las que aún así tengan insuficiente observaciones) y se llevará a cabo una regresión con regularización *Ridge*. Luego, se computarán porcentajes de variación explicada por cada serie de ajustes (dentro del total explicado por todas ellas). Finalmente, se llevará adelante un procedimiento análogo al mencionado en el párrafo anterior, construyendo un promedio ponderado (por cantidad de empleados) de estos porcentajes de poder explicativo a nivel de subgrupos, para cada categoría CIIU. Este producto no será interpretado como un empalme, sino como un insumo potencialmente útil para la predicción de los salarios privados uruguayos. La correspondencia en el sentido contrario (categorías CIIU para cada subgrupo de Consejos de Salarios) no sería construible en este caso.

Referencias

- Bertrand, M., Kamenica, E., and Pan, J. (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics*, 130(2):571–614.
- Cabrera, V. and Cárpena, C. (2013). Metodología para empalmar la ech y los convenios colectivos. *Serie Documentos de Trabajo/FCEA-IE; DT09/13*.
- Calmfors, L. and Driffill, J. (1988). Bargaining structure, corporatism and macroeconomic performance. *Economic Policy*, 3(6):14–61.
- Carnevale Fernández, M. and Gesto Giannattasio, N. (2009). Cobertura de los consejos de salarios.
- Cengiz, D., Dube, A., Lindner, A., and Zentler-Munro, D. (2022). Seeing beyond the trees: Using machine learning to estimate the impact of minimum wages on labor market outcomes. *Journal of Labor Economics*, 40(S1):S203–S247.
- Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, 134(3):1405–1454.
- David, M. A., Lambert, F., and Toscani, M. F. G. (2019). *More work to do? taking stock of Latin American labor markets*. International Monetary Fund.
- Dube, A., Jacobs, J., Naidu, S., and Suri, S. (2020). Monopsony in online labor markets. *American Economic Review: Insights*, 2(1):33–46.
- Dube, A. and Lindner, A. (2021). City limits: What do local-area minimum wages do? *Journal of Economic Perspectives*, 35(1):27–50.
- Etchemendy, S. (2019). The rise of segmented neo-corporatism in south america: Wage coordination in argentina and uruguay (2005-2015). *Comparative Political Studies*, 52(10):1427–1465.
- Freeman, R. (2007). *Labor Market Institutions Around the World*, page Chapter 34. Sage, London. Alternate chapter title: National Economic Performance NBER WP # 13,242.

- Mazzuchi, G. (2009). Labour relations in uruguay: 2005-2008. Technical report, International Labour Organization.
- Perazzo, I. (2012). La negociación salarial en uruguay: un modelo para analizar sus efectos. *Serie Documentos de Trabajo/FCEA-IE; DT07/12*.
- Wang, H., Li, G., and Tsai, C.-L. (2007). Regression coefficient and autoregressive order shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 69(1):63–78.