



BANDERAS ROJAS Y CORRUPCIÓN: UNA EVALUACIÓN EMPÍRICA



Tomás Cortés Pacheco y Eduardo Engel

Introducción

Las compras públicas son un pilar clave en la administración del Estado, toda vez que su correcta ejecución impacta directamente tanto en la disponibilidad como en la calidad de los servicios públicos que se proveen. En adición, las compras públicas contribuyen a la confianza que la ciudadanía deposita en el gobierno, puesto que son una fracción importante del presupuesto público financiado por los contribuyentes. Por ende, una gestión eficiente de estos recursos es fundamental para el éxito del Estado.

La magnitud de los volúmenes de dinero transados en gastos de compras públicas⁸ representan un desafío no solo administrativo, sino también relacionado con la probidad y la transparencia. En efecto, los montos en juego, en conjunto con la interacción directa entre funcionarios públicos y empresas privadas, sumado a la complejidad de los procesos, hacen de las compras públicas tierra fértil para la corrupción. Esta noción de riesgo es compartida por la ciudadanía. En Chile, según una encuesta realizada por la Contraloría General de la República (CGR), las compras públicas son el área del Estado donde se cree que existe la mayor cantidad de corrupción.

Dada la relevancia del problema, existe una fuerte demanda por instrumentos que permitan detectar procedimientos de compra potencialmente irregulares. En este contexto, una de las alternativas más populares son las denominadas banderas rojas (red flags), indicadores de posibles irregularidades presentes en el proceso de compra. Este método es activamente promovido por una serie de organismos multilaterales, tales como el Banco Mundial (World Bank, 2013), la OECD (2020) y Transparencia Internacional (2014), entre otros. Hasta la fecha, sin embargo, hay escasa evidencia sistemática sobre su efectividad para detectar la corrupción.

La investigación en torno a este tema es compleja fundamentalmente por dos razones. En primer lugar, medir la corrupción es un desafío empírico en sí mis-

mo, de modo que la falta de bases de datos que incluyan procedimientos de compra irregulares ha sido clave en la obstaculización de esta agenda. En segundo lugar, está el problema metodológico de cómo evaluar la efectividad de las banderas rojas. La premisa de las banderas rojas es que permitirían detectar procedimientos irregulares, de modo que la métrica para evaluar si cumplen con su objetivo debiese estar ligada a su poder predictivo para clasificar licitaciones riesgosas.

En este informe se usa una novedosa base de datos con información de licitaciones a las que la CGR detectó alguna irregularidad, junto con datos provenientes de ChileCompra. Se evaluó la efectividad de las banderas rojas según su poder predictivo para detectar licitaciones irregulares usando diversos algoritmos de Machine Learning. En particular, la estrategia empírica construye clasificadores usando Random Forests y Gradient Boosting Trees, cuyo objetivo es clasificar licitaciones irregulares usando como insumo los indicadores de banderas rojas para evaluar su desempeño en tareas de predicción.

Una pieza clave en el análisis son las bases de datos creadas para este informe. El punto de partida son los informes de auditoría de Contraloría. Se descargaron todos los informes relacionados con compras públicas desde 2012 con el objeto de singularizar, mediante técnicas de extracción de texto, aquellas licitaciones en las que se detectaron irregularidades. Para crear el grupo de licitaciones sin irregularidades, se optó por dos alternativas. La primera consiste en extraer las licitaciones que se mencionan en los informes, pero que no aparecen en las conclusiones del texto, de modo que se pueden interpretar como licitaciones investigadas, pero no sancionadas. La segunda opción consiste en extraer licitaciones al azar de ChileCompra que sean “comparables” con el grupo de procedimientos irregulares. Luego de unir el grupo de licitaciones irregulares con cada una de las alternativas para el grupo sin irregularidades, surgie-

8. El gasto promedio en compras públicas para los países pertenecientes a la OCDE representa aproximadamente el 30% del gasto total del gobierno.

ron las dos muestras con las que se desarrollaron los análisis de este trabajo.

Cada muestra está diseñada para responder preguntas diferentes. Por un lado, la resultante de combinar las licitaciones irregulares junto con aquellas investigadas mas no sancionadas sirve para estudiar si, cuando un ente regulador enfrenta un pool de licitaciones que investigar, las banderas rojas se pueden usar para sugerir cuáles procedimientos tienen una mayor probabilidad de ser riesgosos. Por otro lado, la segunda muestra entrega luces respecto de si las banderas rojas sirven para mejorar la forma de generar el pool de licitaciones que investigará el organismo fiscalizador.

Dado que el objetivo es evaluar el poder predictivo de las banderas rojas, se adoptó el instrumental metodológico desarrollado por la literatura de Machine Learning como base del análisis. En efecto, así como el interés de la microeconometría está centrado en la inferencia causal, la atención de los algoritmos de Machine Learning está puesta en lograr predicciones precisas. Una serie de algoritmos persigue ese objetivo, tales como neural nets, k-nearest neighbors, support vector machines, Multivariate Adaptive Regression Splines, trees, etc. En este informe, se usan los algoritmos de trees (árboles de ahora en adelante) por su flexibilidad para adaptarse a una buena cantidad de escenarios. En concreto, el siguiente párrafo de Elements of Statistical Learning resume las propiedades de los algoritmos de árboles.

Of all the well-known learning methods, decision trees come closest to meeting the requirements for serving as an off-the-shelf procedure for data mining. They are relatively fast to construct and they produce interpretable models (if the trees are small). (...) they naturally incorporate mixtures of numeric and categorical predictor variables and missing values. They are invariant under (strictly monotone) transformations of the individual predictors. As a result, scaling and/or more general transformations are not an issue, and they are immune to the effects of predictor outliers. They perform internal feature selection as an integral part of the procedure. They are thereby resistant, if not completely immune, to the inclusion of many irrelevant predictor varia-

bles. These properties of decision trees are largely the reason that they have emerged as the most popular learning method for data mining (Friedman, Hastie y Robert, 2001, p. 352).

A pesar de sus propiedades, los árboles de decisión por sí solos no permiten generar predicciones precisas. Para estos efectos, se pueden agregar varios árboles a través de métodos de ensamblaje que mejoran drásticamente la precisión de estos algoritmos. En particular, se usaron las dos aplicaciones más conocidas: Random Forests y Gradient Boosting Trees.

Además de las banderas rojas por licitación, existen banderas rojas para la institución, las cuales caracterizan el riesgo de corrupción resumiendo el comportamiento agregado de una entidad compradora. Con el objetivo de incorporarlas al análisis, se estimó la relación entre las faltas que encontró la Contraloría en un organismo dado con las banderas rojas en el año previo mediante un modelo de efectos fijos, aprovechando así la estructura de datos de panel observada.

Los resultados indican que las banderas rojas carecen de poder predictivo para clasificar las licitaciones irregulares, resultado que se mantiene tanto para Random Forests como para Gradient Boosting Trees a través de ambas muestras. En particular, las predicciones generadas por los algoritmos son siempre inferiores o iguales a la tasa base de licitaciones irregulares en la muestra, por lo que los indicadores parecen no aportar información. En concreto, no es posible mejorar la forma en que la Contraloría genera el pool de licitaciones para investigar, ni tampoco usar las banderas rojas de este informe para predecir, dentro del pool, cuáles son los procedimientos más riesgosos. Respecto de las banderas rojas para cada institución, se observa que prácticamente ningún indicador alcanza la significancia estadística, y que las banderas rojas no son útiles para capturar la variación en las faltas que encuentra la Contraloría a través de los años.

Los hallazgos de este trabajo se relacionan con una incipiente literatura que busca evaluar los indicadores de banderas rojas. Decarolis y Giorgiantonio (2020) también recurren a herramientas de Machine Learning para estudiar la efectividad de las ban-

deras rojas en predecir corrupción. Encuentran que aquellos indicadores que no son “evidentes” —y que, por lo tanto, los agentes corruptos no pueden manipular— efectivamente poseen poder predictivo para clasificar procedimientos riesgosos. En particular, señalan que el único algoritmo con buen desempeño predictivo es Random Forests⁹. Ferwerda, Deleanu y Unger (2016) también analizan la efectividad de las banderas rojas usando instrumental econométrico clásico, y encuentran que solo un grupo tiene poder predictivo para detectar anomalías¹⁰.

A continuación se describen las fuentes de datos y cómo se construyen las muestras con las cuales posteriormente se realizan los análisis. Luego, en la sección de metodología, se ofrece una breve introducción a los algoritmos de árboles y a los métodos de ensamblaje que constituyen la base de la estrategia empírica. En la sección de resultados se discuten los hallazgos de utilizar la estrategia empírica a los datos y finalmente se presentan las conclusiones.

Los datos

Fuente de los datos

Los datos que se usan para el análisis empírico de este informe provienen de los informes de auditoría de la Contraloría General de la república (CGR). En particular, mediante web scraping se recogieron todos los informes de auditoría emanados por la CGR desde 2012 en adelante que incluyesen las palabras clave “contratación pública”, lo cual generó un resultado de 8.041 informes. A partir de esta búsqueda, se generaron las dos variables clave de irregularidades que se usan posteriormente en este informe. Primero se verificó si en las conclusiones del informe la CGR había detectado irregularidades en los procesos de compras públicas, en cuyo caso se codificó como que existió una irregularidad en dicho organismo comprador. Luego se comprobó si en las conclusiones del informe se mencionaban irregularidades en alguna

licitación en particular, en cuyo caso se codificó como que la CGR detectó irregularidades en dicha licitación¹¹. El resultado de este proceso entrega dos variables de irregularidades, una por institución y otra por licitación.

La segunda fuente de datos proviene de los servidores de ChileCompra, institución que dirige la plataforma transaccional a través de la cual los compradores públicos demandan productos y servicios. Estos servidores contienen la información de todas las transacciones que realizan los compradores del Estado, desde su preparación hasta su adjudicación y las órdenes de compra que se devengan de dichos procesos. Sin embargo, no tiene información de los acontecimientos postransacción, lo cual limita la cantidad de datos que se podría utilizar. En adición, es necesario tener en cuenta que varios de los inputs de la plataforma son datos que introducen de forma manual los servidores públicos, con lo cual existen potenciales errores en la información disponible. Además, representa eventuales espacios para irregularidades o corrupción, puesto que se podría ingresar información errada a la plataforma con el objeto de enmascarar procedimientos ilícitos. A pesar de estas limitaciones, la información del servidor es lo suficientemente abundante para generar los indicadores de banderas rojas que serán parte esencial del análisis empírico que sigue en la próxima sección.

La evaluación empírica de las banderas rojas, que es el centro de este informe, tiene dos aristas: las licitaciones y las municipalidades¹². En las subsecciones a continuación se describe cómo se juntan las fuentes de datos previamente descritas y que son la base de ambos análisis.

Licitaciones¹³

Con el objetivo de evaluar empíricamente si los indicadores de banderas rojas pueden predecir aquellas licitaciones riesgosas, es necesario contar con un

9. Los autores evalúan el poder de predicción para el algoritmo de Random Forests usando el out-of-bag estimation error, que no siempre es la mejor alternativa para estimar el error de predicción, toda vez que es muy ruidosa. En efecto, a medida que el tamaño de la muestra tiende a infinito, el OOB estimation error converge al Leave-One-Out Cross Validation error, conocido por tener una varianza más grande que procedimientos como k-fold Cross-Validation.

10. Es preciso notar que con poder predictivo se refieren a predicción dentro de la muestra en una regresión lineal, contexto que no es el más indicado para evaluar predicción.

11. Al realizar este proceso para un año en particular, prácticamente siempre que una licitación aparecía en la conclusión de un informe, era debido a que se había encontrado una irregularidad. Por lo mismo, para el resto de los años todas las licitaciones que aparecían en los informes de Contraloría se clasificaron como irregulares. Para verificar que el método funciona, se seleccionaron 100 licitaciones al azar, de las cuales 96 presentaron irregularidades.

12. Para más información respecto de las banderas rojas para las instituciones, revisar Jorquera (2019).

13. En el Anexo 1 se presenta un esquema con el detalle de cómo se crearon ambas muestras.

grupo de licitaciones que no hayan sido detectadas como irregulares. En este sentido, desde el punto de vista del econometrista, lo ideal es observar el registro completo de licitaciones que investigó la CGR, incluyendo aquellas que finalmente no fueron irregulares, y usar esa muestra en el análisis. Dado que no se dispone de dicha información, el grupo de control debe ser generado artificialmente. A continuación se presentan dos formas de generar dicho grupo de licitaciones, las cuales dan origen a las muestras que se usarán en los análisis empíricos.

Licitaciones #1

La primera alternativa para generar el grupo de control se basa en encontrar licitaciones “similares” a las irregulares que fueron detectadas por la CGR. Para ello se extrajeron todas las licitaciones registradas en ChileCompra desde 2012 hasta la actualidad. Luego, se seleccionaron aquellas que pertenecían a organismos donde la CGR ya hubiese detectado licitaciones irregulares, para finalmente extraer licitaciones al azar dentro de ese grupo. Es importante notar que en el último punto se extrajeron las licitaciones de forma

tal de replicar los momentos de la distribución de los montos adjudicados en las licitaciones irregulares. El objetivo fue encontrar licitaciones que tuviesen un perfil de riesgo similar al de las licitaciones investigadas y sancionadas por la CGR. A la muestra resultante de juntar estas licitaciones con las irregulares detectadas por Contraloría se la denominó “Licitaciones #1”.

En la Tabla 12 se presenta la estadística descriptiva de Licitaciones #1. Se observa que, efectivamente, los montos adjudicados entre ambos grupos son relativamente similares, lo cual se refleja no tan solo en que sus medias son parecidas, sino también en que la diferencia de medias no es estadísticamente significativa. Además, en concordancia con la Tabla 54, las licitaciones irregulares tienen consistentemente un menor cumplimiento de la normativa básica de documentos obligatorios tales como las bases de licitación, las actas de evaluación y las resoluciones de adjudicaciones. También es importante destacar que las licitaciones irregulares tienden a adjudicar consistentemente un mayor valor que el presupuesto estimado, no así el otro grupo, en que generalmente se adjudica por menos del monto estimado.

Tabla 12: Estadística descriptiva Licitaciones #1

	Irregularidad		
	Sí (1)	No (2)	Diferencia (1) - (2)
Monto adjudicado	4841 (47906)	3808 (176676)	1033 (8543)
Contrato renovable	0,03 (0,17)	0,03 (0,17)	0,00 (0,01)
Licitación con varias adjudicaciones	6,13 (38,09)	1,68 (2,42)	4,45*** (0,76)
Diferencia entre monto estimado y adjudicado	1,64 (12,39)	0,50 (0,46)	1,14*** (0,27)
Número de participantes	13,80 (93,35)	5,43 (6,48)	8,37*** (1,86)
Licitación tuvo un único oferente	0,25 (0,44)	0,23 (0,42)	0,02 (0,02)
Licitación sin bases de licitación	0,92 (0,28)	0,67 (0,47)	0,25*** (0,02)
Licitación sin acta de evaluación	0,73 (0,44)	0,65 (0,48)	0,08*** (0,02)
Licitación sin resolución de adjudicación	0,88 (0,33)	0,71 (0,45)	0,17*** (0,02)
Primer Componente Principal	0,26 (3,32)	-0,05 (0,28)	0,31*** (0,07)
Observaciones	478	2.574	

Fuente: Elaboración propia.

Notas: Estadística descriptiva para la muestra "Licitaciones #1". Las licitaciones se agrupan entre aquellas donde la CGR detectó irregularidades y aquellas que se obtuvieron mediante el proceso detallado en la sección "Licitaciones".

La última columna muestra la diferencia de medias entre los grupos detallados anteriormente, donde la significancia estadística de la diferencia al 1%, 5% y 10% se denota por ***, ** y * respectivamente.

Licitaciones #2

La segunda forma de generar el grupo de control es mediante las licitaciones mencionadas en los informes emanados de Contraloría que no aparecen en las conclusiones de los mismos. Para extraer dichas licitaciones, se usaron técnicas de reconocimiento de texto. A la muestra resultante de juntar estas licitaciones con las irregulares se la denominó “Licitaciones #2”.

En la Tabla 13 se presenta la estadística descriptiva de Licitaciones #2. Las licitaciones clasificadas como

“no irregulares” tienen un monto adjudicado promedio aproximadamente seis veces menor que aquellas irregulares, diferencia que es estadísticamente significativa.

Sin embargo, la diferencia entre los montos estimados y adjudicados son similares entre ambos grupos, lo cual se ratifica al observar que su diferencia no es estadísticamente significativa. Por otro lado, se destaca que las licitaciones sin irregularidades tienen un número de participantes promedio tres veces menor que aquellas irregulares, diferencia que nuevamente es estadísticamente significativa.

Tabla 13: Estadística descriptiva Licitaciones #1

	Irregularidad		
	Sí (1)	No (2)	Diferencia (1) - (2)
Monto adjudicado	4841 (47906)	818 (9976)	4023*** (994)
Contrato renovable	0,03 (0,17)	0,03 (0,16)	0,00 (0,01)
Licitación con varias adjudicaciones	6,13 (38,09)	2,52 (16,09)	3,61*** (0,99)
Diferencia entre monto estimado y adjudicado	1,64 (12,39)	1,73 (32,11)	-0,09 (1,59)
Número de participantes	13,80 (93,35)	5,60 (25,90)	8,20*** (2,02)
Licitación tuvo un único oferente	0,25 (0,44)	0,24 (0,43)	0,01 (0,02)
Licitación sin bases de licitación	0,92 (0,28)	0,86 (0,35)	0,06*** (0,02)
Licitación sin acta de evaluación	0,73 (0,44)	0,70 (0,46)	0,03* (0,02)
Licitación sin resolución de adjudicación	0,88 (0,33)	0,87 (0,33)	0,01 (0,02)
Observaciones	478	3.412	

Fuente: Elaboración propia.

Nota: Estadística descriptiva para la muestra “Licitaciones #2”. Las licitaciones se agrupan entre aquellas donde la CGR detectó irregularidades y aquellas que se obtuvieron mediante el proceso detallado en la sección “Licitaciones”.

La última columna muestra la diferencia de medias entre los grupos detallados anteriormente, donde la significancia estadística de la diferencia al 1%, 5% y 10% se denota por ***, ** y * respectivamente.

El diseño de cada una de las muestras de licitaciones impacta de forma directa en cómo se deben entender los resultados de los análisis a continuación. Este punto se discute en detalle en la sección de resultados.

Municipalidades

La muestra para el análisis por municipalidades considera todas estas instituciones entre 2012 y 2019. Para cada combinación de municipalidad y año se aplicó una serie de indicadores de banderas rojas, en conjunto con una variable indicador que se activa si la CGR detectó un proceso irregular de compras públicas en esa municipalidad y año. En la Tabla 54 se presenta la estadística descriptiva de dicha muestra.

La descripción de cada una de las banderas rojas se encuentra en el Anexo 2. En la Tabla 14 se aprecia

que para varios indicadores no existe una diferencia estadísticamente significativa entre aquellas municipalidades en que se detectaron irregularidades y en las que no. Para aquellas banderas rojas donde la diferencia sí es significativa, no todas las diferencias van en la dirección intuitiva. Por ejemplo, el número de participantes distintos es aproximadamente un 40% mayor en aquellas municipalidades donde se detectaron irregularidades.

Algunos de los indicadores cuya diferencia sí va en el sentido esperado son aquellos que se refieren a la transparencia de los procesos licitatorios, como el porcentaje de cumplimiento de los documentos que por ley deben estar disponibles en cualquier licitación. En estas banderas rojas se observa que, consistentemente, son aquellos municipios con irregularidades los que menos cumplen con tener disponibles las bases de licitación, las actas de evaluación y las resoluciones de adjudicación.

Tabla 14: Estadística descriptiva banderas rojas por municipalidad

	Irregularidad		
	Sí (1)	No (2)	Diferencia (1) - (2)
% concentración 4 empresas con más contratos	0,19 (0,07)	0,19 (0,08)	0,00 (0,00)
% concentración 4 empresas con más OC	0,37 (0,16)	0,37 (0,16)	0,00 (0,01)
Índice ID (cantidad)	0,18 (0,14)	0,17 (0,14)	0,01 (0,01)
Índice ID (monto)	0,35 (0,24)	0,35 (0,24)	0,00 (0,01)
Índice HHI (cantidad)	0,02 (0,02)	0,02 (0,02)	0,00 (0,00)
Índice HHI (monto)	0,07 (0,07)	0,07 (0,07)	0,00 (0,00)
Índice participación	0,05 (0,02)	0,04 (0,02)	0,01*** (0,00)
% licitaciones cuyo plazo fue menor a 20 días	0,26 (0,24)	0,30 (0,26)	-0,04** (0,01)
Número de participantes distintos	768,68 (530,64)	536,12 (382,38)	232,56*** (23,19)
% licitaciones con un único oferente	0,21 (0,09)	0,24 (0,09)	-0,03*** (0,01)
% del monto total adjudicada por TD o LP	0,14 (0,11)	0,14 (0,10)	0,00 (0,01)
% de contratos total adjudicada por TD o LP	0,26 (0,17)	0,28 (0,16)	-0,02** (0,01)
Número de participantes promedio	5,65 (1,78)	5,11 (1,78)	0,54*** (0,10)
% de cumplimiento archivos obligatorios	0,38 (0,14)	0,45 (0,14)	-0,07*** (0,01)
% de licitaciones sin bases de licitación	0,57 (0,32)	0,41 (0,32)	0,16*** (0,02)
% de licitaciones sin acta	0,46 (0,25)	0,38 (0,23)	0,08*** (0,01)
% de licitaciones sin resolución de adjudicación	0,47 (0,24)	0,40 (0,23)	0,07*** (0,01)
Primer Componente Principal	-0,60 (2,18)	0,09 (1,97)	-0,69*** (0,12)

Fuente: Elaboración propia.

Notas: Estadística descriptiva para la muestra de banderas rojas por municipalidades entre 2013 y 2020, separando aquellas municipalidades en las que Contraloría determinó la existencia de irregularidades de las que no.

La última columna muestra la diferencia de medias entre los grupos detallados anteriormente, donde la significancia estadística de la diferencia al 1%, 5% y 10% se denota por ***, ** y * respectivamente.

Metodología

El objetivo de este informe es proveer una evaluación empírica de la efectividad de los indicadores de banderas rojas. La premisa fundamental es que permitirían identificar los riesgos presentes en el proceso de compra, con lo cual se podrían detectar transacciones potencialmente fraudulentas o corruptas que un organismo de control externo podría auditar y sancionar. Por ende, para evaluar su efectividad es natural estudiar en qué medida las banderas rojas son capaces de predecir qué licitaciones u organismos han sido sancionados por los órganos de control del Estado.

Con este fin, se presentan dos evaluaciones. En primer lugar, se prueba la capacidad de las banderas rojas de predecir irregularidades detectadas por la Contraloría General de la República en procedimientos de licitación. En segundo lugar, se estudia si las banderas rojas de los indicadores para el nivel institucional del periodo anterior correlacionan con la presencia de irregularidades en el periodo siguiente. A continuación se describen los métodos que subyacen a los análisis de este trabajo.

Árboles

En esencia, determinar si los indicadores de banderas rojas son capaces de predecir procedimientos irregulares es un problema de clasificación. En otras palabras, lo que se busca es evaluar en qué medida los indicadores permiten clasificar licitaciones entre irregulares y no irregulares. Si bien este tipo de problemas se puede abordar mediante métodos estadísticos estándar¹⁴ como modelos de probabilidad lineal (MPL) o logit, estos presentan la desventaja de requerir supuestos sobre el proceso generador de datos (DGP) que generalmente se desconocen. Entonces, se recurre a modelos no paramétricos que sean agnósticos respecto del DGP subyacente.

En esta línea, la literatura de Machine Learning ha desarrollado algoritmos precisamente para la clasificación, y en este informe, en particular, se usan algoritmos de árboles, caracterizados por su flexibilidad para abordar interacciones y no linealidades, y por su rapidez computacional¹⁵. En adición, estos algo-

ritmos tienen la ventaja de que fueron desarrollados con el objetivo de predicción en mente, que es exactamente la finalidad en este caso.

Para introducir a literatura de árboles, a continuación se presentan los algoritmos usados en este informe. Primero se detalla el algoritmo básico y luego se describen los procedimientos de ensamblaje que combinan varios árboles para mejorar su desempeño.

CART. En abstracto, los algoritmos de árboles se basan en estratificar o segmentar el espacio generado por los predictores en varias regiones. Luego, para hacer una predicción dada una observación, típicamente se toma el promedio simple de las observaciones a la región que pertenece. Como el conjunto de las reglas de decisión que se usan para segmentar el espacio generado por los predictores se puede representar a través un árbol de decisión, es que estos procedimientos toman su nombre.

En concreto, el algoritmo consiste en hacer crecer un árbol a partir de un nodo inicial, de modo que cada nodo sucesivo se divide en dos a partir de una regla de decisión. Cada bifurcación involucra a una variable X_j y un valor s , de manera que se generan dos grupos, uno con las observaciones con $X_j < s$ y otro con aquellas $X_j \geq s$. Luego, cada nodo terminal agrupa aquellas observaciones que cumplen con cada regla de decisión de los nodos previos, y es con esas observaciones que se calcula el valor predicho de la variable objetivo, tomando el promedio simple de dichas observaciones. Debido a la simpleza de este algoritmo su representación a través del árbol de decisión (dendograma) que genera es intuitivo y fácil de interpretar.

Una desventaja de los algoritmos de árboles es que, por sí solos, tienden a tener un mal desempeño en tareas de predicción. Esto se debe a que las particiones generadas por las reglas de decisión se basan en minimizar el error cuadrático medio de la predicción dentro de la muestra. Por lo tanto, el algoritmo *aprende demasiado bien las características* de la muestra particular que está observando, pero luego, al probar la predicción en una muestra distinta, esta tiende a ser imprecisa. En la literatura de Machine Learning a este fenómeno se le conoce como *sobreajuste*, lo cual refleja que, con el objetivo de minimizar

14. Esto es especialmente cierto cuando la variable que se debe predecir es binaria, como en este caso.

15. Para una evaluación completa de las propiedades de los modelos de árboles, revisar Friedman, Hastie y Tibshirani, (2001, p. 352).

el error de predicción, el algoritmo asigna todos sus esfuerzos a disminuir el sesgo del procedimiento sin compensar por el aumento en la varianza que eso implica¹⁶.

Para equilibrar sesgo y varianza en los algoritmos de Machine Learning existen los métodos de ensamblaje. La idea principal consiste en construir un modelo de predicción combinando las fortalezas de una colección de modelos más simples, por ejemplo, árboles. En este informe, se usan dos métodos de ensamblaje aplicados a árboles: Random Forests y Boosting Trees, los cuales se detallan a continuación.

Random Forests (*bagging*). La esencia del algoritmo consiste en generar un “comité” de árboles, donde cada cual cuenta con un voto para clasificar a una observación dada, y donde la clasificación final se obtiene por mayoría simple. En concreto, para este método se requiere generar B muestras con reemplazo de la muestra original, lo que se conoce como muestras *bootstrap*, con las cuales se estiman B árboles para luego tomar el promedio simple de las predicciones resultantes, que corresponde a la predicción final. En el fondo, tomar el promedio de los árboles permite reducir su varianza, que era el problema inicial de los métodos basados en árboles. A este proceso de agregación de modelos se le conoce como *bagging*, cuyo nombre proviene de combinar las palabras *bootstrap* con *aggregation*. Random Forests es similar a *bagging*, con la salvedad de que, al momento de construir los árboles, una selección de m predictores (en lugar del total de p predictores) son escogidos como candidatos a ser las variables de bifurcación en cada nodo. El objetivo de este procedimiento es reducir la correlación entre los árboles que se construyen, de modo que al momento de tomar el promedio entre estos su varianza sea aún menor. Una propiedad deseable de este procedimiento es que como subproducto del mismo se tiene una medida para evaluar la efectividad del algoritmo para predecir a la variable objetivo. En efecto, dado que los datos se muestrean con reemplazo, no se usan todas las observaciones en cada árbol que se estima, con lo que se puede construir una predicción fuera de muestra para cada observación que no se utiliza. Al error de estimación resultante del proceso anteriormente descrito se le conoce como *out-of-bag prediction error*, y es fundamental para escoger el valor de m óptimo. En efecto,

se calculó el error de predicción OOB para todos los valores posibles de m y luego se escogió aquel valor de m que entregó el menor error de predicción OOB.

Gradient boosting trees. Mientras que *bagging* estima cada árbol de forma independiente, el algoritmo de *boosting* estima cada árbol de forma secuencial usando una versión modificada de la muestra original, donde cada árbol está limitado a un tamaño d . A diferencia de *bagging*, la versión modificada en este caso considera todas las observaciones de la muestra, pero en cada iteración se pone mayor énfasis en aquellas observaciones donde el clasificador enfrenta dificultades para categorizarlas correctamente. El algoritmo detecta las observaciones donde tiene dificultades a través del método del gradiente descendente, de ahí el nombre de este procedimiento. En definitiva, el algoritmo de *boosting* también genera un “comité” de árboles para categorizar los datos, de modo que la decisión final se toma mediante un promedio ponderado cada clasificador, asignando un mayor peso a aquellos árboles con mayor precisión.

Respecto del algoritmo de *bagging*, *boosting* requiere que el usuario “ajuste” una mayor cantidad de parámetros con la finalidad de obtener un buen resultado. En este sentido, *bagging* es más amigable por cuanto el mismo algoritmo se puede usar para distintos tipos de problemas, mientras que *boosting* requiere que para cada problema el usuario sepa ajustar los parámetros del modelo para lograr un buen desempeño. Por ejemplo, *gradient boosting trees* requiere que el usuario especifique una función de pérdida para identificar las observaciones donde el algoritmo tiene problemas para predecir la variable dependiente. En este caso, se especificó una función de pérdida Bernoulli, que es estándar para problemas de clasificación binarios. Luego, el algoritmo también requiere especificar el tamaño máximo permitido para cada árbol d . Cuando el algoritmo se aplica a árboles, el valor de d indica la cantidad de interacciones posibles: 1 estima un modelo aditivo (sin interacciones), 2 permite interacciones de primer orden, y así sucesivamente. Para escoger tanto el valor del parámetro d como la cantidad de árboles, se usó una fracción de la muestra para estimar el modelo y calcular el error de predicción usando *k-fold Cross-Validation*, para luego ver cómo el modelo resultante se desempeña en las observaciones que no se usaron para la estimación.

16. En términos técnicos se dice que el algoritmo se ubica en el lado de la varianza dentro del bias-variance trade-off.

La desventaja de los modelos de ensamblaje vistos previamente es que se pierde la interpretación intuitiva que tiene el algoritmo de árbol más simple. Con el fin de evaluar cuáles son las variables más importantes para predecir las sanciones de la Contraloría, se introdujeron gráficos de importancia relativa y de dependencia parcial. El primero da una noción de la relevancia de cada predictor, mientras que el segundo presenta el efecto marginal asociado a cada predictor. En particular, se presentan dos medidas de importancia relativa: *Mean Decrease Accuracy* (MDA) y *Mean Decrease Gini* (MDG). MDA entrega una medida de la pérdida de precisión promedio en predicción cuando se excluye a una variable del modelo, mientras que MDG mide el decrecimiento en el *Gini Impurity Index* promedio a través de los árboles, de manera que este índice entrega una medida de la efectividad de cada bifurcación. La interpretación directa de ambos indicadores no es trivial, pero sus magnitudes son útiles para notar cuáles son las variables que el algoritmo considera más importantes.

A continuación se señala cómo se escogieron los parámetros de los modelos y cómo se evaluó su desempeño. En primer lugar, se aprecia que todos los métodos para escoger la parametrización del modelo se basan en los errores de predicción fuera de muestra. En efecto, en el caso de *random forests* se usó el *out-of-bag* prediction error, mientras que en *boosting* se aplicó el método de *k-fold Cross-Validation* para generar errores de predicción fuera de muestra y sobre esa base parametrizar el modelo. Esta precisión es relevante puesto que es simple generar un modelo cuya predicción dentro de la muestra sea todo lo precisa que se desee, pero que su desempeño fuera de muestra sea deficiente.

En segundo lugar, para medir la efectividad de las banderas rojas es que tanto en *random forests* como en *boosting* se guardó una fracción de la muestra para validar el modelo. Esto es, se dividieron de forma aleatoria los datos en una muestra de entrenamiento y otra de validación, usando la primera para la estimación (y parametrización) de los modelos y la segunda para calcular su desempeño, como es habitual en la literatura de Machine Learning. A través de los métodos recién descritos se evaluó el desempeño de las banderas rojas para predecir licitaciones irregulares. En concreto, se aplicaron los algoritmos de

árboles a cada una de las muestras detalladas en la sección “Licitaciones”.

Estimador de efectos fijos

Con el objeto de evaluar la relación entre los indicadores de banderas rojas para las instituciones con la presencia de irregularidades detectadas en el periodo siguiente, en el segundo análisis de este informe se aplicó un estimador de efectos fijos. En otras palabras, se estimó la siguiente ecuación:

$$\text{sanción}_{it} = \delta_i + \text{RF}_{it-1} \beta + \varepsilon_{it}$$

Donde sanción_{it} es una variable binaria que se activa cuando el organismo i en el periodo t cuenta con al menos una irregularidad detectada por la CGR, δ_i es un efecto fijo para la institución, RF_{it-1} denota el vector de banderas rojas descritas en la Tabla 12 para el organismo i en el periodo $t-1$ y, finalmente, ε_{it} representa el término de error.

En primer lugar se muestran los resultados asociados a los modelos de árboles. En particular, se presenta el algoritmo de CART para el caso de un solo predictor, con el fin de obtener una interpretación intuitiva de la versión más simple del algoritmo. Luego se exhiben los resultados de *Random Forests* y *Gradient Boosting Trees*, con foco en la *misclassification rate* de cada uno de estos algoritmos.

Además, se enumeran medidas de importancia relativa de cada uno de los predictores junto con gráficos de dependencia parcial. Junto con ello, se presenta el resultado del modelo de efecto fijo en el contexto de las banderas rojas institucionales para estudiar su relación con las irregularidades que detecta la Contraloría en el periodo siguiente.

Resultados

A continuación se presentan los resultados de las metodologías detalladas en la sección anterior. Para los algoritmos de árboles cabe señalar que, debido a la forma en que se construyeron ambas muestras, estas están diseñadas para responder preguntas distintas. Por un lado, Licitaciones #1 permite estudiar si un or-

ganismo fiscalizador —como la CGR— podría usar las banderas rojas para identificar qué licitaciones investigar. Por el otro lado, Licitaciones #2 permite analizar si en caso de que el organismo fiscalizador enfrentara un pool de procedimientos licitatorios, las banderas rojas servirían para conducirlo a las licitaciones más riesgosas.

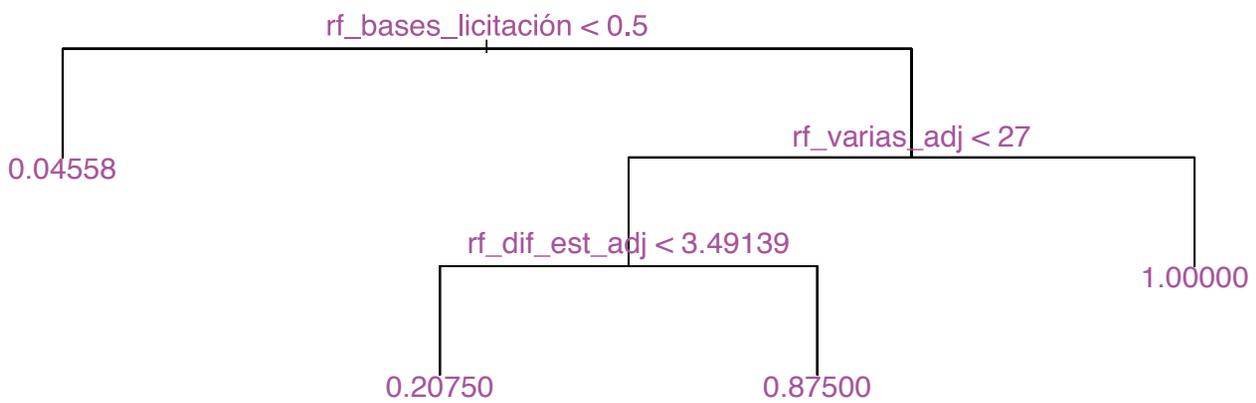
Licitaciones #1

En primer lugar, se estimó un modelo simple de árbol para entender el objeto que será la base de los análisis

posterior. En la Figura 18 se observa la representación gráfica de aplicar el algoritmo CART para predecir la variable de irregularidad usando como predictores los indicadores de banderas rojas descritos en la Tabla 52.

A esta representación visual, conocida como dendograma, sirve para aterrizar el algoritmo descrito en la sección anterior. En cada bifurcación el algoritmo escoge una variable y un umbral, de modo de minimizar el error de predicción, y así sucesivamente hasta que los datos hayan sido ajustados a cierto nivel de tolerancia.

Figura 18: Dendograma Licitaciones #1



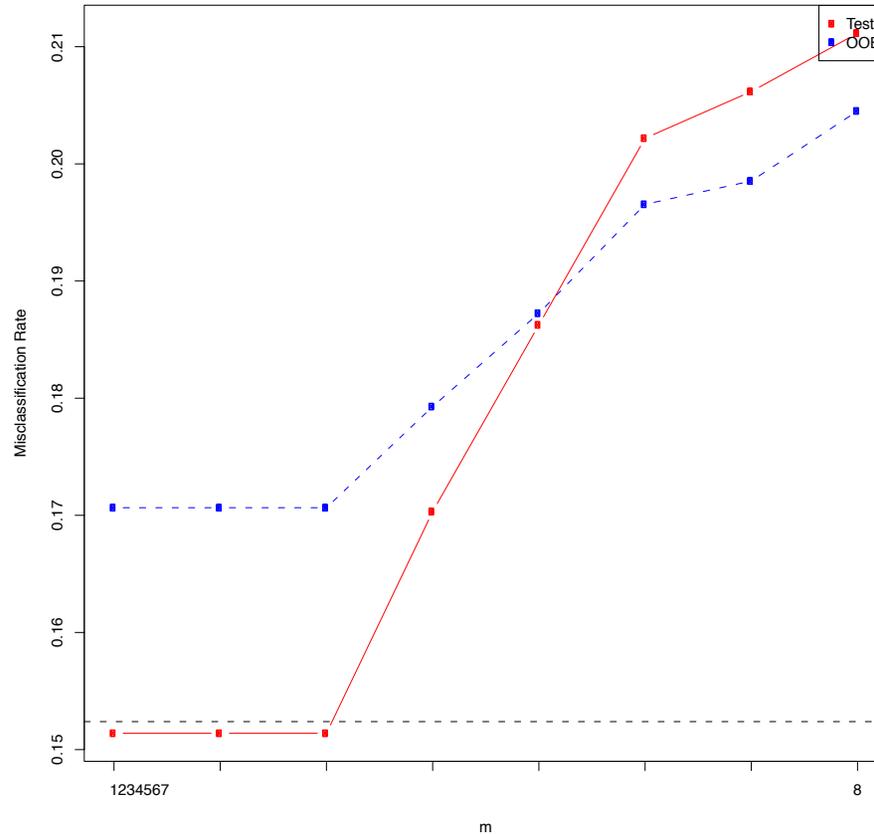
Fuente: Elaboración propia.

A pesar de que se incluyeron 8 predictores, el algoritmo solamente usó 3 para realizar la clasificación. El árbol considera que la variable más relevante para predecir licitaciones irregulares es si se publicaron o no sus bases de licitación. En efecto, si están publicadas, entonces la probabilidad de que sea irregular es del 4,6% según el clasificador. Si no están publicadas, entonces la clasificación depende del valor que tome el indicador de varias adjudicaciones. Si dicho valor es mayor o igual a 27, entonces la licitación se clasifica como irregular. Si es menor a 27, entonces depende del valor que toma la diferencia entre el valor estimado y el adjudicado (desviación). Si esta es

menor a 3,5, entonces la licitación no clasifica como irregular, mientras que en caso contrario sí clasifica como irregular.

Dada la tendencia al sobreajuste de los datos del algoritmo de árbol, en la Figura 18 en realidad se describen los datos particulares que se usaron para estimar el algoritmo. Para analizar el valor predictivo de las banderas rojas se continuó con *Random Forests*. En el Gráfico 43 se presentan tanto los errores de predicción *out-of-bag*¹⁷ como en la muestra de validación para cada valor posible del parámetro *m*.

17. Calculados con la muestra de entrenamiento.

Gráfico 8: *Out-of-bag Prediction Error: Licitaciones*

Fuente: Elaboración propia.

Del Gráfico 8 se desprende que ambos errores son crecientes en el parámetro m ¹⁸. El valor de m óptimo en este caso es 1, el cual entrega un error de clasificación del 15,1% en la muestra de validación. Para poner esta cifra en contexto, la línea horizontal del Gráfico 43 presenta la fracción de licitaciones irregulares en la muestra, correspondiente al 15,2%. Esto quiere decir que el algoritmo lo hace ligeramente mejor en

términos de predicción que la estrategia base de clasificar todas las licitaciones como sin irregularidades, estrategia que llevaría a una tasa de error 0,1 puntos porcentuales mayor. Para ahondar en el desempeño en términos de predicción del algoritmo, en la Tabla 55 se presenta la matriz de confusión, que describe para cada tipo de licitación la predicción que el algoritmo hizo sobre su estado (irregular o no irregular).

18. Notar que $m=8$ corresponde a bagging, ya que se dispone de 8 predictores.

Tabla 15: Matriz de confusión para Licitaciones #1

Tipo de licitación	Predicción		
	Sin irregularidades	Irregular	Error de clasificación
Sin irregularidades	848	3	0,35%
Irregular	149	4	97,39%

Fuente: Elaboración propia.

En la Tabla 15 se observa que el error de clasificación para las licitaciones sin irregularidades es del 0,35%, error que aumenta al 97,39% al momento de predecir las licitaciones irregulares. El complemento de esta última tasa se conoce como sensibilidad del clasificador, y describe qué fracción de los verdaderos positivos son capturados por el modelo. En este caso, la sensibilidad es del 2,61%. En definitiva, las banderas rojas en este ejercicio parecen no tener poder predictivo alguno para predecir aquellos procedimientos irregulares. Antes de revisar los resultados para el algoritmo de boosting, se verá qué banderas rojas son aquellas que el algoritmo considera más importantes para predecir irregularidades.

En la Tabla 16 se presentan dos medidas de importancia para cada indicador de banderas rojas. La columna Mean Decrease Accuracy indica que la variable más relevante en el modelo es la bandera roja de bases de licitación. Esta variable es también la más relevante para predecir la irregularidad de una licitación, lo cual se aprecia en la tercera columna. Para predecir cuándo una licitación no presenta irregularidades, la bandera roja más importante es la diferencia entre el monto estimado y el adjudicado, lo cual se ve en la segunda columna. Dicho indicador es el segundo más importante según la MDA y la más relevante según el Mean Decrease Gini.

Tabla 16: Importancia de las variables

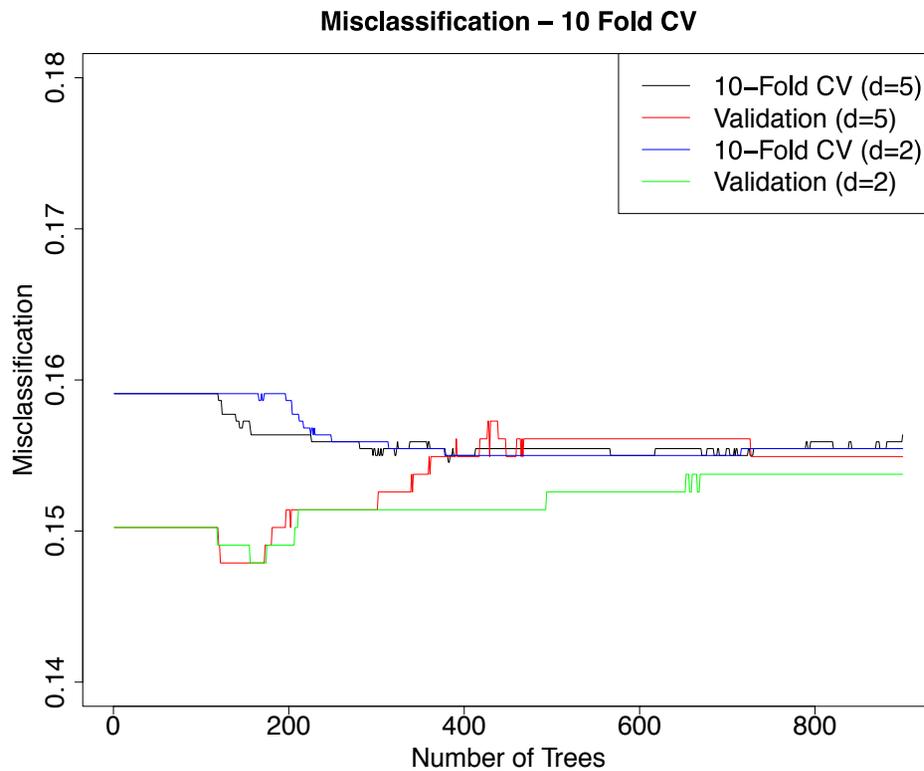
	Irregular		Medidas de importancia	
	No	Sí	Mean Decrease Accuracy	Mean Decrease Gini
rf renew	8,14	-4,61	6,05	4,12
rf varias adj	30,54	-14,77	31,1	15,69
rf dif est adj	31,97	10,12	34,69	103,93
rf numpar	27,06	3,44	28,77	35,29
rf 1 of	-3,95	14,55	5,15	4,43
rf bases licitacion	28,3	26,66	37,68	16,41
rf acta eval	5,89	6,22	8,23	7,65
rf resol adj	15,06	-8,88	10,26	6,85

Fuente: Elaboración propia.

Para finalizar, se muestra la implementación del algoritmo de Gradient Boosting Trees para el problema de predicción. En la Figura 19 se presenta el proceso de Cross-Validation que se usó para determinar qué

parametrización del modelo entregaba la menor tasa de error, medida como la cantidad de veces que el algoritmo se equivoca al momento de predecir el valor de la variable objetivo.

Gráfico 9: Cross-Validation: d y M



Fuente: Elaboración propia.

De este ejercicio se concluye que el modelo óptimo es aquel con parámetro $d=5$ y $M=382$, lo cual indica que el modelo permite interacciones de hasta quinto orden entre los predictores y que se utilizan 382 árboles agregados. Entonces, la tasa de error en la muestra de validación es del 15,6%, mientras que en la misma muestra las licitaciones irregulares repre-

sentaban el 15% aproximadamente, por lo cual nuevamente los indicadores de banderas rojas parecen no tener valor predictivo sobre las licitaciones irregulares. Finalmente, con el objetivo de determinar cuáles banderas rojas individuales son útiles, se presenta el análisis de importancia relativa y de efectos marginales para el algoritmo de *boosting*.

Tabla 17: Importancia de las variables

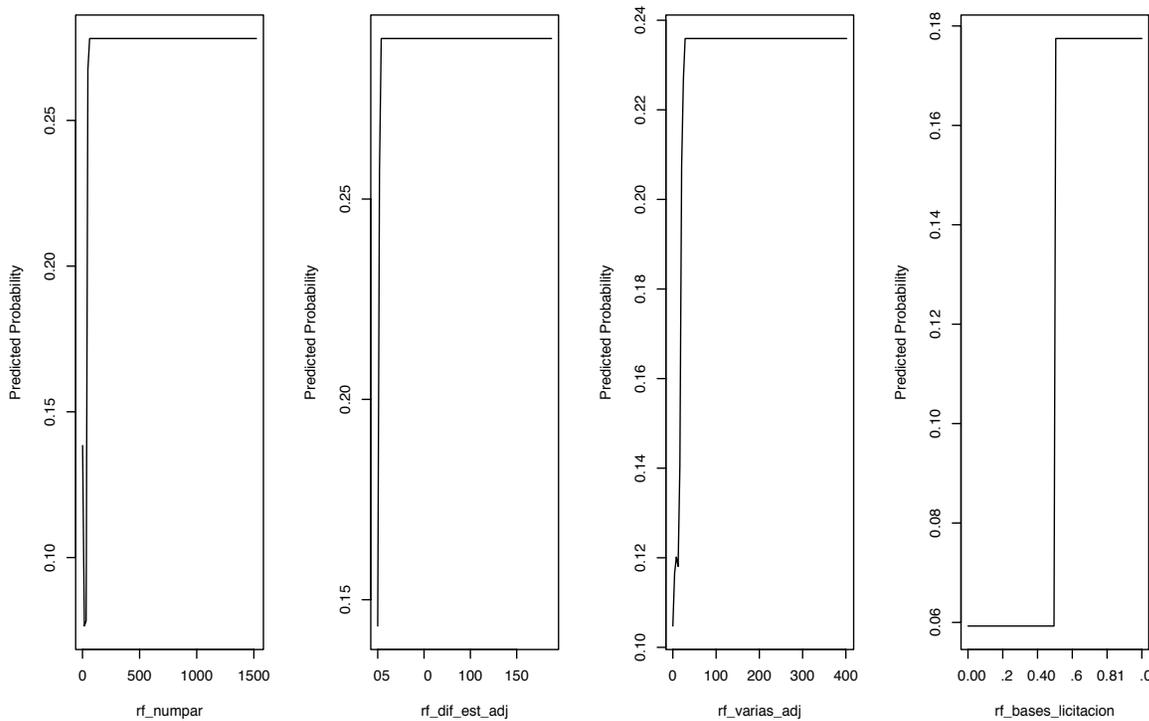
	Mean Decrease Accuracy
rf numpar	34,27
rf dif est adj	22,15
rf varias adj	19,01
rf bases licitacion	13,36
rf resol adj	7,35
rf acta eval	3,6
rf renew	0,22
rf 1 of	0,04

Fuente: Elaboración propia.

En la Tabla 17 se observa que la variable más importante para determinar el estatus de la variable dependiente es el número de participantes en el procedimiento. Luego siguen variables como la diferencia

entre el monto adjudicado y el estimado, si el contrato tuvo varias adjudicaciones y si la licitación no presenta sus bases de evaluación.

Figura 19: Dependencia parcial



Fuente: Elaboración propia.

Por último, en la Figura 19 se presenta el efecto marginal en la probabilidad predicha de cada una de las 4 banderas rojas más importantes según la Tabla 15. En primer lugar, se aprecia que al aumentar marginalmente el número de competidores se reduce la probabilidad de que la licitación sea clasificada como irregular, sin embargo, y de forma contraintuitiva, esta probabilidad empieza a aumentar luego de cierto umbral de competidores. Algo similar ocurre con contratos que tienen varias adjudicaciones: aumentos en esta variable se traducen en un alza en la probabilidad de irregularidad. En adición, se observa que no tener las bases de la licitación también incrementa la probabilidad de que el clasificador catalogue a la licitación como irregular.

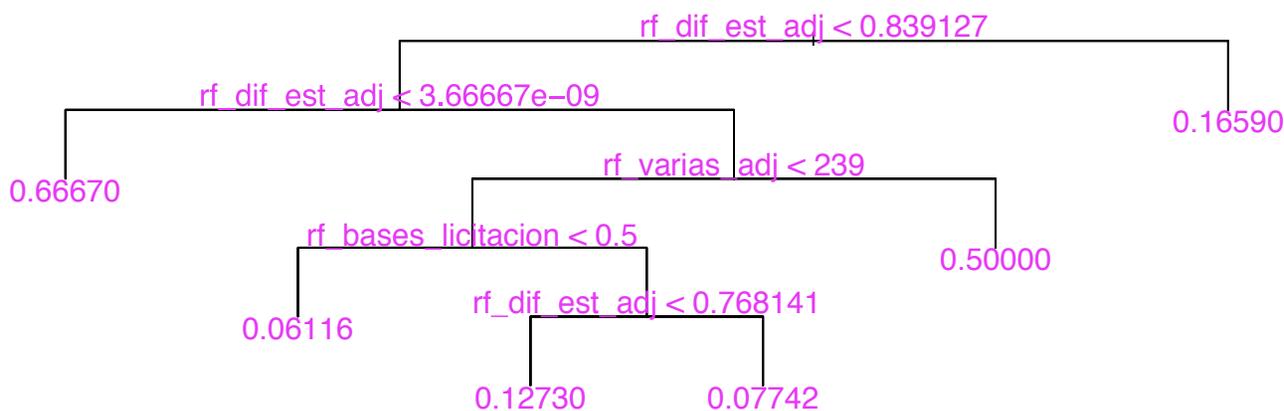
Los resultados previos sugieren que las banderas rojas no son efectivas para predecir las licitaciones irregulares de la muestra Licitaciones #1. Dada la forma en que se construyó esta muestra, es posible interpretar que estos resultados indican que, al menos con las banderas rojas usadas en este ejercicio, no se puede mejorar la matriz de riesgo que utiliza la CGR

para escoger qué procedimientos investigar. A continuación se presentan los mismos análisis aplicados a la segunda muestra.

Licitaciones #2

En la Figura 20 se presenta el dendograma para el algoritmo CART aplicado a la muestra Licitaciones #2. Como se mencionó en la subsección anterior, en estos casos el algoritmo de árbol da una buena descripción de los datos que se usaron para estimarlo (dada su tendencia a sobreajustar) más que una fórmula para predicción. En este sentido, se aprecia que el algoritmo usa tres banderas rojas solamente: la diferencia entre el monto adjudicado y el monto estimado, el número de adjudicaciones en el contrato, y si el procedimiento publicaba o no sus bases de licitación. De los nodos terminales se extrae que en solo uno de seis el algoritmo clasifica al grupo de licitaciones resultantes como irregulares, mientras que en otro le asigna una probabilidad del 50%.

Figura 20: Dendograma Licitaciones #2

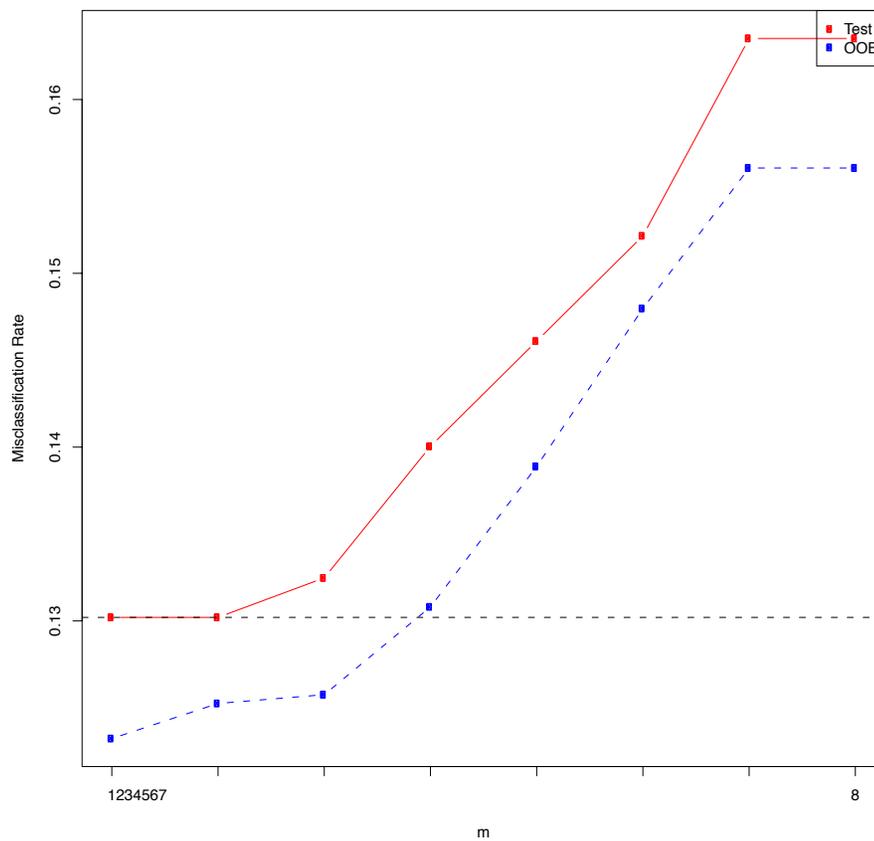


Fuente: Elaboración propia.

Para conocer el desempeño en predicción, se revisaron los resultados para el algoritmo de *Random Forests*. En el Gráfico 9 se presentan los errores de predicción OOB usados para parametrizar el modelo, junto con los errores de predicción en la muestra de validación. Se observa que, nuevamente, $m=1$ es

aquel valor que minimiza el error OOB, y que entrega una *misclassification rate* en la muestra de validación idéntica a la fracción de licitaciones irregulares en la muestra de aproximadamente 13%, la cual está graficada en la línea horizontal.

Gráfico 10: *Out-of-bag Prediction Error: Licitaciones #2*



Fuente: Elaboración propia.

Para ahondar en cómo el algoritmo efectúa las clasificaciones, en la Tabla 18 se desglosan los aciertos y discrepancias entre la clasificación del modelo contra el tipo verdadero de cada licitación. Se puede ver que *Random Forests* no clasifica ninguna observa-

ción como irregular, lo cual explica por qué su tasa de error era idéntica a la tasa base descrita en el párrafo anterior. En este contexto, la sensibilidad del clasificador es del 0%, con lo que, ciertamente, no tiene sentido usar este modelo para fines de predicción.

Tabla 18: Importancia de las variables

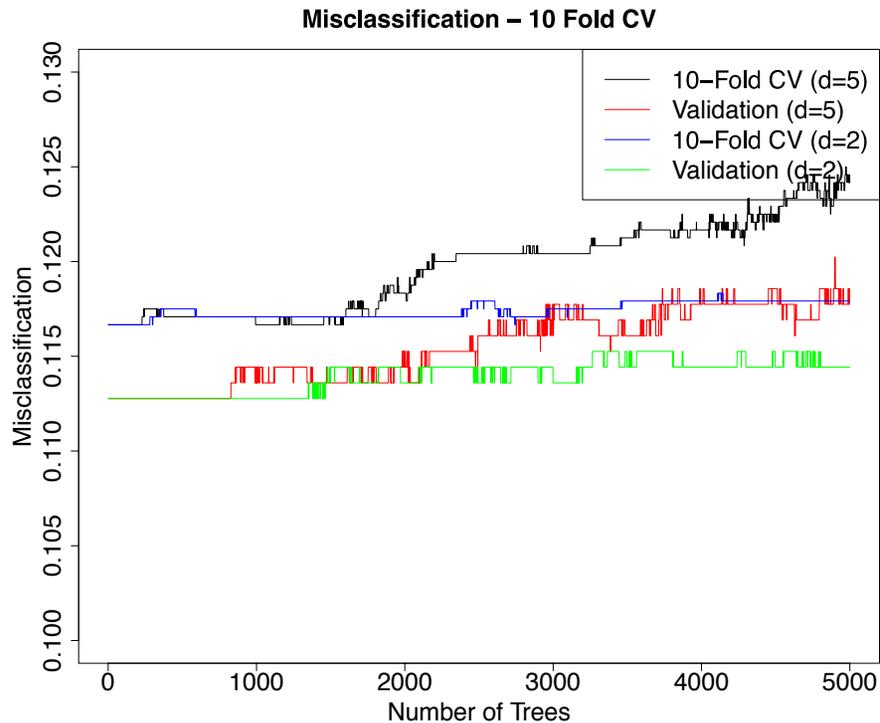
Tipo de licitación	Predicción		
	Sin irregularidades	Irregular	Error de clasificación
Sin irregularidades	848	3	0,35%
Irregular	149	4	97,39%

Fuente: Elaboración propia.

Tampoco tiene sentido mostrar las medidas de relevancia de las variables, por lo cual se revisan directamente los resultados usando el algoritmo de boosting. En efecto, en la Figura 21 se presenta la

misclassification rate calculada mediante 10-fold Cross-Validation para la muestra de entrenamiento como aquella calculada sobre la muestra de validación.

Figura 21: Cross-Validation: d y M



Fuente: Elaboración propia.

De acuerdo con la Figura 22, la parametrización óptima del modelo es aquella con $d=2$ y con $M=1$. Al igual que con *Random Forests*, *boosting* entrega una tasa de clasificación errónea igual a la tasa base de licitaciones irregulares en la muestra, ya que no predice ninguna licitación irregular. Nuevamente, los resultados indican que las banderas rojas son incapaces de clasificar correctamente licitaciones irregulares dentro de la muestra Licitaciones #2. Dado el desempeño del modelo, se omitió la presentación de las medidas de importancia relativa.

Tomando en cuenta cómo se construyó esta muestra, los resultados previos sugieren que en los procedimientos que revisa la CGR, las banderas rojas aquí presentadas no pueden mejorar el proceso de detección de cuáles terminarán siendo sancionadas. Junto con lo señalado en la sección anterior para la muestra Licitaciones #1, los ejercicios de este infor-

me indican que las banderas rojas consideradas parecen no tener poder predictivo para clasificar los procedimientos licitatorios. Sin embargo, se debe tener en cuenta que estos resultados son específicos tanto para las banderas rojas consideradas como para las muestras generadas. Por ende, los hallazgos de este trabajo no pretenden ser conclusivos y pueden cambiar al considerar otras banderas rojas u otros datos de mejor calidad.

Municipalidades

En el caso de las municipalidades, el interés radicó en estudiar la relación entre los indicadores de banderas rojas en las instituciones durante el periodo anterior y las irregularidades en el periodo siguiente. El resultado del modelo presentado en la sección anterior se resume en la Tabla 19.

Tabla 19: Estimación de efectos fijos

Variable	OLS	FE
Índice concentración 4 empresas más grandes (cantidad)	0,493* (0,270)	0,352 (0,386)
Índice concentración 4 empresas más grandes (monto)	0,093 (0,096)	0,069 (0,105)
Índice ID (cantidad)	0,046 (0,079)	0,058 (0,093)
Índice ID (monto)	-0,034 (0,045)	-0,016 (0,046)
Índice HHI (cantidad)	-1,401 (1,605)	-1,346 (1,710)
Índice HHI (monto)	0,075 (0,231)	-0,018 (0,243)
Índice de participación	0,350 (0,348)	-0,110 (0,432)
% de licitaciones cuyo plazo de oferta fue acelerado	-0,030 (0,023)	0,002 (0,029)
# de participantes distintos entre todos los procedimientos	0,000*** (0,000)	0,000*** (0,000)
% Licitaciones con único oferente	0,108 (0,096)	0,017 (0,122)
% de los montos adjudicados mediante trato directo	0,032 (0,079)	-0,059 (0,108)
% de los procedimientos adjudicados mediante trato directo	0,013 (0,049)	0,133 (0,090)
Promedio de participantes por procedimiento	0,006 (0,005)	0,004 (0,010)
% de procedimientos que publican todos los doc. obligatorios	-1,011 (1,104)	-0,720 (1,274)
% de procedimientos sin bases de licitación	-0,201 (0,274)	-0,085 (0,317)
% de procedimientos sin acta de evaluación	-0,178 (0,279)	-0,092 (0,313)
% de procedimientos sin resolución de adjudicación	-0,221 (0,276)	-0,124 (0,322)
cons	0,559 (0,823)	0,316 (0,957)
Observations	2.748	2.748
R-squared within	0,059	0,047
R-squared between		0,128
R-squared overall		0,052

Fuente: Giorgiantonio y Decarolis, 2020.

En la Tabla 59 se aprecia que, cuando no se tiene en cuenta la estructura de datos de panel y se estima la ecuación mediante mínimos cuadrados ordinarios, se obtiene que aumentos tanto en el índice de concentración (cantidad) como en el número de participantes distintos tienden a aumentar la probabilidad de que se detecten irregularidades en el año siguiente, y que ambos aumentos son estadísticamente significativos (al 10% y 1% respectivamente). La dirección de este efecto hace sentido para el índice de concentración, pero no para el número de participantes distintos, puesto que se esperaría que aumentos en el número de participantes distintos en los procedimientos estuviesen correlacionados con licitaciones más competitivas, y, por ende, menos irregulares. Sin embargo, este efecto está precisamente estimado en 0, por lo cual no es problemático.

La última columna de la Tabla 59 refleja los resultados estimados mediante efectos fijos al nivel de municipalidad. Ambos modelos resultan ser similares, y la única diferencia en términos de las variables es que el índice de concentración (cantidad) deja de ser estadísticamente significativo. Por otro lado, el número de participantes distintos en los procedimientos sigue siendo precisamente estimado en 0.

Estos modelos lineales permiten estimar la fracción de la variación en la variable dependiente que es atribuible a variaciones en las variables independientes, a través del r -cuadrado. Más aún, en el modelo de efectos fijos es posible descomponer el R^2 en el R^2 intra y entre municipalidades. En efecto, el R^2 intramunicipalidades indica cuánta de la variación en la variable dependiente al interior de las municipalidades es capturada por el modelo, mientras que el R^2 entre municipalidades señala la fracción de la variación en la variable dependiente entre municipalidades que es atribuible al modelo. De la tabla 59 se desprende que el modelo captura en torno al 12,8% de la variación entre municipalidades, sin embargo, apenas captura el 4,7% de la variación al interior de las municipalidades. En otras palabras, si se colapsa la dimensión temporal de los datos y se observa la media, el modelo captura en torno a un décimo de la variación de irregularidades promedio a través de las municipalidades. Sin embargo, si se analiza la muestra detrended, entonces el modelo solo captura en torno a 1/20 de la variación total. Es decir, buena par-

te de la variación que el modelo captura no se debe a las banderas rojas propiamente tal, sino a diferencias previas entre municipalidades. En particular, que el R^2 -intra sea tan bajo es un indicador de que la variación en el tiempo de las banderas rojas correlaciona poco con las variaciones en la presencia de irregularidades detectadas.

Conclusión

Las recomendaciones de política para combatir la corrupción en las compras públicas incluyen generalmente alguna versión de indicadores de banderas rojas. A pesar de su popularidad, no existe un cuerpo de literatura robusto que evalúe su efectividad en la práctica. En este trabajo se pone a prueba el poder predictivo de las banderas rojas para clasificar licitaciones irregulares usando algoritmos de Machine Learning especializados en predicción. Para esto se combinaron auditorías realizadas por la Contraloría General de la República junto con información detallada de cada procedimiento licitatorio registrado en ChileCompra. Los resultados indican que las banderas rojas carecen de poder predictivo para clasificar las licitaciones riesgosas. En el contexto de este informe, esta ausencia de poder predictivo tiene dos implicancias. En primer lugar, si un ente fiscalizador quisiera escoger un pool de procedimientos licitatorios para auditar, las banderas rojas usadas en este trabajo no podrían ayudarle a escoger dicho conjunto eficientemente. En segundo lugar, si un ente fiscalizador se enfrenta a un pool dado de licitaciones para auditar, entonces nuevamente las banderas rojas no podrían guiarle con el objeto de darle prioridad a las que tienen una mayor probabilidad de ser riesgosas.

Es importante notar que, si bien estos hallazgos apuntan en la dirección de que las banderas rojas no son efectivas para predecir procedimientos de compra riesgosos, deben interpretarse de forma matizada. En primer lugar, los resultados indican que el grupo particular de banderas rojas presentado carece de poder predictivo, por lo que existe la posibilidad de que “mejores” banderas rojas sí tengan poder predictivo. En este sentido, los métodos propuestos en este informe pueden ser útiles para poner a prueba nuevas banderas rojas. En segundo lugar, los resultados pueden ser sensibles a las muestras con las que se lleva-

ron a cabo los análisis. En efecto, dada la naturaleza de los actos de corrupción, obtener los datos que se desearía observar para hacer los análisis no son factibles de obtener, por lo cual deben ser aproximados de alguna forma. Si bien para tales efectos se presentaron dos formas alternativas de aproximar los datos, estas están sujetas a error. En último lugar, existe literatura —aunque incipiente— que sí encuentra que las banderas rojas tienen poder predictivo, por lo que

los esfuerzos en esta área debiesen estar orientados a entender por qué son efectivas en ciertos contextos y no en otros.

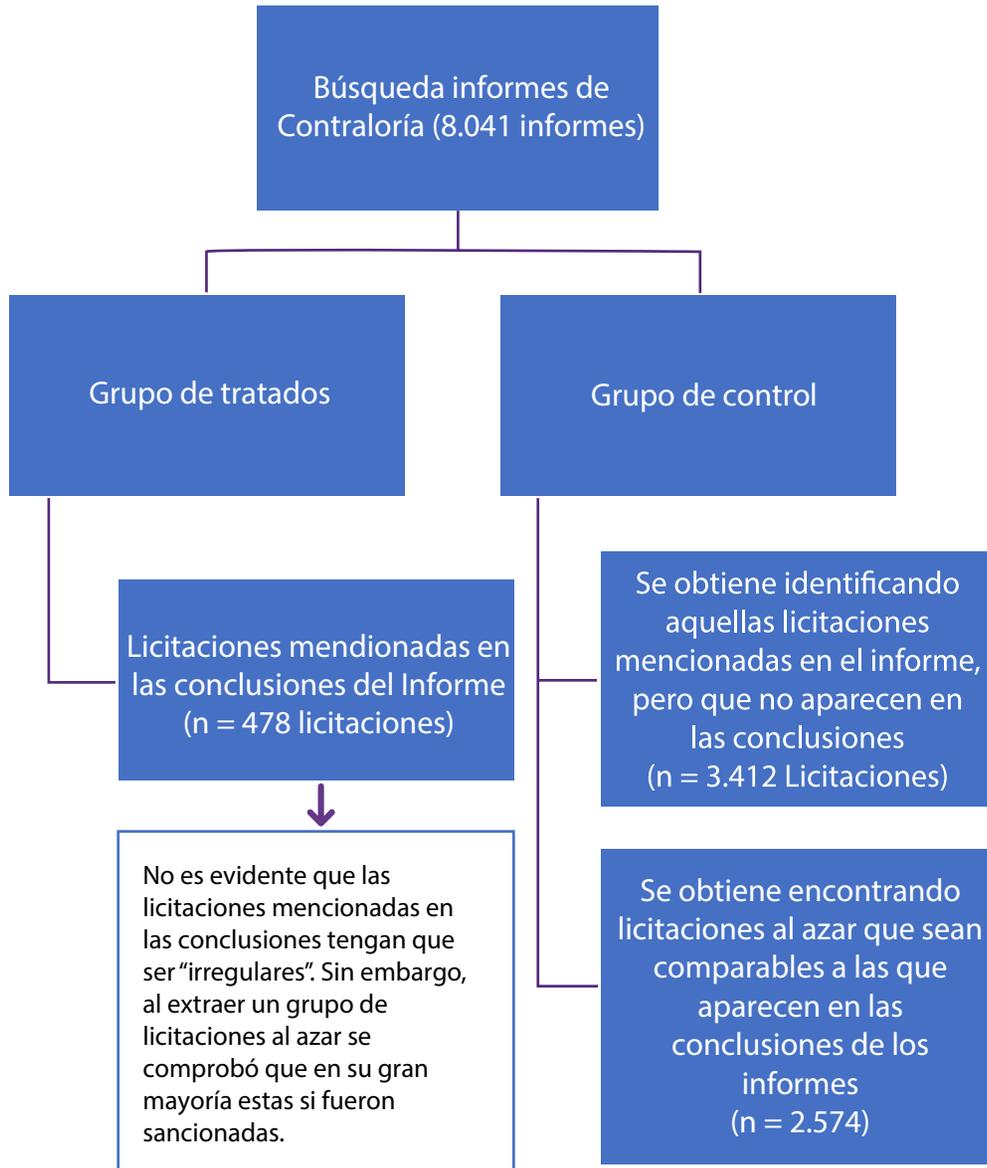
En definitiva, los resultados contribuyen a expandir el cuerpo de literatura existente en esta materia, pero es necesaria más investigación para entender cuándo son efectivos estos indicadores.

Bibliografía

- Ferwerda, J., Deleanu, I., y Unger, B. (2016). Corruption in Public Procurement: Finding the Right Indicators. *European Journal on Criminal Policy and Research*, 245-267.
- Friedman, J., Hastie, T., y Robert, T. (2001). *The elements of statistical learning*. Nueva York: Springer.
- Giordiantonio, C., y Decarolis, F. (2020). Corruption Red Flags in Public Procurement: New Evidence from Italian Calls for Tenders. *Questioni di Economia e Finanza*.
- Jorquera, M. (2019). Compras públicas y Big Data: Investigación en Chile sobre índice de riesgo de corrupción. *Espacio Público*. Recuperado de www.sistemaspublicos.cl/publicacion/compras-publicas-y-big-data-investigacion-en-chile-sobre-indice-de-riesgo-de-corrupcion
- OCDE (2020). Recommendation of the Council on Public Procurement. Recuperado de www.oecd.org/gov/public-procurement/OECD-Recommendation-on-Public-Procurement.pdf
- Transparency International (2014). *Curbing Corruption in Public Procurement: A Practical Guide*. Recuperado de www.transparency.org/en/publications/curbing-corruption-in-public-procurement-a-practical-guide
- World Bank (2013). *Fraud and Corruption Awareness Handbook*. Recuperado de <https://openknowledge.worldbank.org/handle/10986/18153>

Anexos

Anexo 1



Anexo 2. Banderas rojas usadas

Licitaciones

Contrato renovable: Indicador binario que se activa si el contrato tiene una cláusula de renovación.

Licitación con varias adjudicaciones: Indicador que se construye tomando el número de adjudicaciones para cada licitación.

Diferencia entre monto estimado y adjudicado: Para cada licitación, este indicador se construye tomando la división entre el monto adjudicado con el estimado. Luego, valores positivos indican que el monto adjudicado fue mayor que el estimado y viceversa.

Número de participantes: Indicador que se construye tomando la cantidad de participantes en una licitación dada.

Licitación con un único oferente: Indicador binario que se activa cuando la licitación recibió una única oferta.

Licitación sin bases de licitación: Indicador binario que se activa cuando la licitación no tiene disponibles sus bases de licitación.

Licitación sin acta de evaluación: Indicador binario que se activa cuando la licitación no tiene disponible su acta de evaluación.

Licitación sin resolución de adjudicación: Indicador binario que se activa cuando la licitación no tiene disponible su resolución de adjudicación.

Municipalidades¹⁹

% de licitaciones con un único oferente: Dividir el número de licitaciones públicas y privadas con un único oferente por el total de licitaciones.

Promedio de participantes promedio: Contar el número de oferentes en cada licitación pública y privada y promediar ese número entre todas las licitaciones.

% de contratos total adjudicados por TD o LP: Dividir el número de procedimientos asociados a trato directo y licitación privada por el potencial número de contratos que debiese haber firmados.

% del monto total adjudicada por TD o LP: Dividir el monto adjudicado en los procedimientos asociados a trato directo y licitación privada por el monto adjudicado entre todos los tipos de contratación.

Índice HHI (cantidad): Calcular la participación de mercado respecto del número potencial de contratos que debiese haber firmados por cada proveedor entre todos los tipos de contratación, para luego sumar cada uno de estos valores al cuadrado.

Índice HHI (monto): Calcular la participación de mercado respecto al monto adjudicado de cada proveedor entre todos los tipos de contratación, para luego sumar cada uno de estos valores al cuadrado.

19. Para mayor información, consultar las banderas rojas descritas en Jorquera (2019).

Índice ID (cantidad): Calcular la participación de mercado respecto del número potencial de contratos que debiese haber firmados por cada proveedor entre todos los tipos de contratación y elevar esos valores al cuadrado. Luego, dividir estos valores por el valor del indicador de IHH por número de contratos, para posteriormente sumar cada uno de esos términos al cuadrado.

Índice ID (monto): Calcular la participación de mercado respecto del monto adjudicado de cada proveedor entre todos los tipos de contratación y elevar esos valores al cuadrado. Luego, dividir estos valores por el IHH, para posteriormente sumar cada uno de esos términos al cuadrado.

Índice de participación: Calcular para el proveedor con mayor número de propuestas presentadas qué porcentaje representa su total de propuestas entre el total de propuesta recibidas por el organismo público.

Número de participantes distintos: Contar el número de participantes distintos dividido por el número total de licitaciones públicas y privadas.

% de concentración 4 empresas con más contratos: Calcular el porcentaje de mercado que representan las 4 empresas con mayor número de contratos potenciales.

% de concentración 4 empresas con más OC: Calcular el porcentaje de mercado que representan las 4 empresas con mayor monto adjudicado entre todos los tipos de contratación.

% de cumplimiento archivos obligatorios: Calcular el porcentaje de incumplimiento respecto de la publicación de las bases de licitación, el acta de evaluación, la resolución de adjudicación y los contratos para cada licitación pública y privada, y luego promediar estos valores entre todas las licitaciones públicas y privadas.

% de licitaciones cuyo plazo fue menor a 20 días: Contar las licitaciones públicas y privadas mayores o iguales a 1.000 UTM y menores a 5.000 UTM, que tuvieron un plazo entre 10 y 20 días en consideración a que eran bienes y servicios de simple y objetiva especificación, y dividir por el total de licitaciones de este tipo.

% de licitaciones sin bases de licitación: Contar las licitaciones públicas y privadas sin algún documento publicado referido a las bases de la licitación y dividir por el total de licitaciones.

% de procedimientos sin acta de evaluación: Contar las licitaciones públicas y privadas sin algún documento publicado referido al acta de evaluación y dividir por el total de licitaciones.

% de procedimientos sin resolución de adjudicación: Contar las licitaciones públicas y privadas sin algún documento publicado referido a la resolución de adjudicación y dividir por el total de licitaciones.

% de procedimientos sin contrato: Contar las licitaciones públicas y privadas que no disponen de todos los contratos asociados a los proveedores adjudicados y dividir por el total de licitaciones.