**La normalización necesaria para hacer frente a las calificaciones ESG divergentes**



Publicado el junio 24, 2022 por [**Editor**](https://www.xbrl.org/news/standardisation-needed-to-tackle-divergent-esg-ratings/)

Existe una divergencia significativa en las calificaciones ambientales, sociales y de gobernanza (ESG) proporcionadas a las empresas por diferentes agencias, según un estudio publicado en Review of Finance. Los autores sugieren que esta divergencia podría tener impactos en el mundo real en las decisiones de inversión y la transición a una economía más sostenible. Esta evidencia refuerza el argumento a favor de una mayor armonización en las divulgaciones ESG y las normas de datos, un caso que recientemente ha sido presentado por el regulador francés *de los mercados financieros, la Autorité des marchés financiers* (AMF), entre otros. En respuesta a la [consulta](https://www.xbrl.org/news/europe-consults-further-on-esg-ratings/) de la Comisión Europea, ha pedido un marco regulador europeo global para los proveedores de datos ASG, calificaciones y servicios relacionados.

Lea más [aquí](https://www.esginvestor.net/esg-ratings-diverge-substantially-with-real-economy-impact/), [aquí](https://academic.oup.com/rof/advance-article/doi/10.1093/rof/rfac033/6590670) y [aquí](https://www.amf-france.org/en/news-publications/news/amf-reiterates-its-call-european-regulation-esg-data-ratings-and-related-services).

[CALIFICACIONES](https://www.xbrl.org/tag/ratings/) [ESG](https://www.xbrl.org/tag/esg/)



**La AMF reitera su llamamiento a favor de una regulación europea de los datos ESG, las calificaciones y los servicios relacionados**

**En su respuesta a la consulta pública de la Comisión Europea sobre las calificaciones ASG, el AMF reitera la necesidad de un marco regulador europeo para los proveedores de datos, calificaciones y servicios relacionados con ESG.**

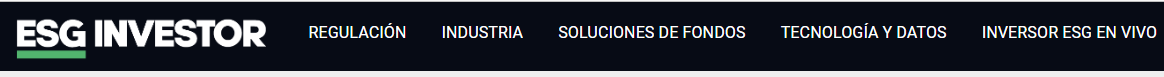
A medida que el mercado de servicios relacionados con la sostenibilidad continúa creciendo a un ritmo rápido y a medida que estos productos desempeñan un papel cada vez más central en los mercados financieros, la provisión de productos relacionados con la sostenibilidad sigue sin estar regulada en gran medida.

Esta situación confirma la necesidad de un marco regulador europeo para el suministro de datos ASG, calificaciones y otros servicios. La respuesta de la AMF a la consulta pública de la Comisión Europea renueva las posiciones desarrolladas en el documento de posición publicado conjuntamente en diciembre de 2020 con su homólogo holandés, la Autoriteit Financiële Markten (AFM).

**Garantizar la protección de los inversores**

La AMF destaca en particular los siguientes puntos principales de atención:

* Este futuro Reglamento debe abarcar toda la gama de datos, calificaciones y servicios ASG y no limitarse a las calificaciones ASG. De hecho, los problemas identificados son comunes a todos estos productos.
* El presente Reglamento debe incluir requisitos de transparencia sobre las metodologías, los datos subyacentes utilizados (fuente y naturaleza) y los objetivos de los productos (en particular, el riesgo o el impacto). También debe incluir requisitos para la gestión de conflictos de intereses, procedimientos de control interno y un diálogo mejorado con las empresas que están sujetas a calificaciones ASG. En esta etapa del desarrollo del mercado, en la que el análisis del rendimiento ESG sigue siendo proteico y evolutivo, este marco regulador no debe conducir a una estandarización de las metodologías, sino que debe garantizar una transparencia suficiente de los actores frente al mercado.
* La supervisión de los agentes debe centralizarse a escala europea, con el objetivo de garantizar una supervisión armonizada y en consonancia con la estructura de este mercado. Confiar esta tarea al regulador europeo de los mercados financieros, la AEVM, permitiría beneficiarse de su experiencia en la supervisión de las agencias de calificación crediticia y sería coherente con el plan de confiarle la supervisión de examinadores externos de bonos verdes europeos.
* Los agentes que deseen proporcionar datos o servicios ASG a los participantes en el mercado de la UE deben operar a través de un establecimiento permanente en la UE y estar registrados en la AEVM.



**Las calificaciones ESG divergen sustancialmente, con un impacto en la economía real**



Por [Susannah Rodgers](https://www.esginvestor.net/author/susannah-rodgers/)

1 de junio de 2022

**El análisis a través de seis agencias de calificación clave encuentra diferencias significativas en la forma en que se miden los indicadores.**

Las calificaciones ESG de diferentes proveedores discrepan sustancialmente, y mucho más que las de las agencias de calificación crediticia, según un estudio publicado en [**Review of Finance**](https://academic.oup.com/rof/advance-article-abstract/doi/10.1093/rof/rfac033/6590670?redirectedFrom=fulltext), la revista de la Asociación Europea de Finanzas.

El estudio, realizado por investigadores del MIT Sloan y la Universidad de Zurich, analizó las calificaciones ESG de KLD, Sustainalytics, Moody's ESG, S&P Global, Refinitiv y MSCI.

La divergencia podría estar teniendo un efecto tangible en la transición de la economía real hacia un futuro más sostenible, dado el impacto potencial de las calificaciones en los precios de los activos y el comportamiento corporativo, según el informe.

"La teoría predice que las preferencias de los inversores por ESG afectan los precios de los activos. En la práctica, sin embargo, las opciones de inversión se guían por las calificaciones ESG, lo que hace que la construcción y el desacuerdo entre las calificaciones ESG sean una preocupación central", dijo el estudio.

Los reguladores se centran cada vez más en las calificaciones de inversión sostenible en un intento por disuadir el lavado verde. Justo este mes, el [**Ministerio de Finanzas del Reino Unido**](https://www.reuters.com/world/uk/britain-decide-this-year-whether-regulate-esg-raters-2022-05-09/) dijo que está considerando si regular los evaluadores ESG, mientras que la [**Comisión de Bolsa y Valores**](https://www.esginvestor.net/us-sec-funds-proposals-aim-to-tackle-greenwashing/) de los Estados Unidos propuso cambios en las reglas para evitar reclamos ESG engañosos por parte de los fondos.

En noviembre pasado, la Organización Internacional de Comisiones de Valores [**publicó**](https://www.iosco.org/library/pubdocs/pdf/IOSCOPD690.pdf) un conjunto de recomendaciones para que los reguladores de valores consideren en su supervisión de las calificaciones ESG y los proveedores de productos de datos.

"La divergencia de la calificación ESG no implica que medir el desempeño ESG sea un ejercicio inútil", dijo el estudio, señalando los beneficios de que los reguladores armonicen la divulgación ESG como una forma de ayudar a proporcionar una base de datos confiables.

"Sin embargo, destaca que medir el desempeño ESG es un desafío, que la atención a los datos subyacentes es esencial y que el uso de calificaciones y métricas ESG debe considerarse cuidadosamente".

**Creciente demanda de calificaciones**

Las calificaciones ESG surgieron por primera vez en la década de 1980 como una forma para que los inversores evalúen a las empresas, y el mercado de calificaciones ha crecido junto con un creciente apetito por la inversión sostenible. Muchos de los primeros proveedores pasaron a ser comprados por compañías establecidas, con MSCI adquiriendo KLD en 2010 y Morningstar tomando una participación del 40% en Sustainalytics en 2017.

Las calificaciones crediticias tienden a divergir menos que sus contrapartes ESG porque la solvencia crediticia puede definirse más claramente y los estándares de información financiera son más maduros, según el estudio. Por el contrario, las agencias de calificación ESG en efecto ofrecen una interpretación de lo que significa el desempeño ESG, en medio de estándares de información competitivos para la divulgación, muchos de los cuales son voluntarios.

La divergencia dificulta la evaluación del desempeño ESG de las empresas, fondos y carteras, disminuye los incentivos de las empresas para mejorar su desempeño ESG e introduce incertidumbre en cualquier decisión tomada en función de las calificaciones ESG, dijeron los investigadores.

"Las empresas reciben señales mixtas de las agencias de calificación sobre qué acciones se esperan y serán valoradas por el mercado", dijo el informe, lo que podría conducir a una inversión insuficiente en actividades de mejora de ESG.

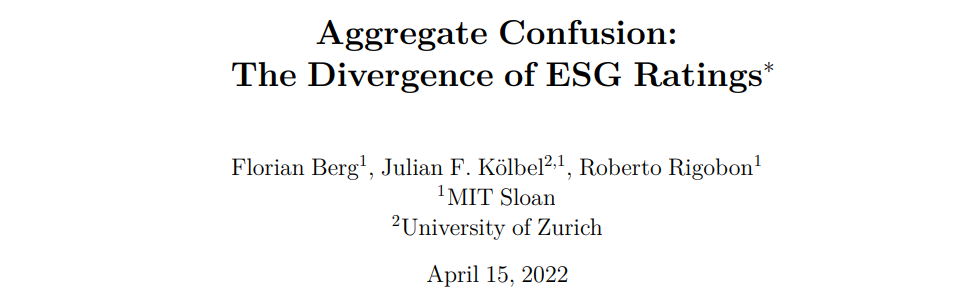
El estudio analizó tres razones principales de la divergencia: el alcance (por ejemplo, una agencia puede incluir actividades de cabildeo mientras que otra no), la medición (las prácticas laborales, por ejemplo, podrían medirse por la rotación de la fuerza laboral o por el número de casos judiciales relacionados con el trabajo tomados contra la empresa) y el peso (por ejemplo, dando mayor importancia a las preocupaciones laborales que al cabildeo).

El estudio encontró que, de los tres, la medición es el factor clave en las diferencias en las calificaciones entre agencias, contribuyendo con el 56% de la divergencia. El alcance representa el 38% y el peso el 6%.

En particular, se encontró que una medición del "efecto halo" impulsa el 15% de la divergencia general, lo que significa que una empresa que recibe una puntuación alta en una categoría tiene más probabilidades de recibir puntuaciones altas en otras categorías, ya que la visión general de la agencia de la empresa informa los resultados.

Con las calificaciones y métricas ESG formando una base importante para el campo de las finanzas sostenibles, la divergencia plantea la cuestión de cómo la incertidumbre en las calificaciones ESG afecta los precios de los activos, según el estudio.

"Nuestros resultados plantean la cuestión de cómo las empresas responden a ser calificadas de manera diferente por diferentes evaluadores, lo que informará los efectos de las finanzas sostenibles en la economía real".



Confusión agregada:

La divergencia de las calificaciones ESG∗

Florian Berg1, Julian F. Kölbel2,1, Roberto Rigobon1

1MIT Sloan

2Universidad de Zúrich

abril 15, 2022

**Extracto**

Este documento investiga la divergencia de las calificaciones ambientales, sociales y de gobernanza (ESG) basadas en datos de seis agencias de calificación ESG prominentes: KLD, Sustainalytics, Moody's ESG (Vigeo-Eiris), S&P Global (RobecoSAM), Refinitiv (Asset4) y MSCI. Documentamos la divergencia de calificación y mapeamos las diferentes metodologías en una taxonomía común de categorías. Usando esta taxonomía, descomponemos la divergencia en contribuciones de alcance, medición y peso. La medición aporta el 56% de la divergencia, el alcance el 38% y el peso el 6%. Analizando más a fondo las razones de la divergencia de medición, detectamos un efecto evaluador donde la visión general de un evaluador de una empresa influye en la medición de categorías específicas. Los resultados requieren una mayor atención a cómo se generan los datos subyacentes a las calificaciones ESG.

Palabras clave: Responsabilidad Social Corporativa, Sostenibilidad Corporativa, Calificaciones ESG, Divergencia

1. Introducción

Los proveedores de calificación ambiental, social y de gobernanza (ESG)1 se han convertido en instituciones influyentes. Un total de 3038 inversores que representan más de 100 billones de dólares en activos combinados han firmado un compromiso para integrar la información ESG en sus decisiones de inversión (PRI, 2020). La inversión sostenible está creciendo rápidamente, y los fondos mutuos que invierten de acuerdo con las calificaciones ESG experimentan entradas considerables (Hartzmark y Sussman, 2019). Debido a estas tendencias, cada vez más inversores confían en las calificaciones ESG para obtener una evaluación de terceros del rendimiento ESG de las corporaciones. Un número creciente de estudios académicos se basan en las calificaciones ESG para su análisis empírico (véase, por ejemplo, Servaes y Tamayo, 2013; Flammer, 2015; Liang y Renneboog, 2017; Lins et al., 2017; Albuquerque et al., 2018). Como resultado, las calificaciones ESG influyen cada vez más en las decisiones, con efectos potencialmente de gran alcance en los precios de los activos y las políticas corporativas.

Sin embargo, las calificaciones ESG de diferentes proveedores no están de acuerdo sustancialmente, como se mostró anteriormente en Chatterji et al. (2016). Confirmamos este hallazgo en nuestro conjunto de datos, donde las correlaciones entre las calificaciones ESG oscilan entre 0,38 y 0,71. Esto se basa en las calificaciones ESG de seis evaluadores diferentes: KLD, Sustainalytics, Moody's ESG (anteriormente Vigeo-Eiris), S&P Global (anteriormente RobecoSAM), Refinitiv (anteriormente Asset4) y MSCI. Este desacuerdo tiene varias consecuencias importantes. En primer lugar, dificulta la evaluación del rendimiento ESG de las empresas, los fondos y las carteras, que es el objetivo principal de las calificaciones ESG. En segundo lugar, la divergencia de calificación ESG disminuye los incentivos de las empresas para mejorar su desempeño ESG. Las empresas reciben señales mixtas de las agencias de calificación sobre qué acciones se esperan y serán valoradas por el mercado. Esto podría conducir a una inversión insuficiente en actividades de mejora de ESG ex ante. En tercer lugar, es menos probable que los mercados valoren el rendimiento ESG ex post de las empresas. El rendimiento ESG puede ser fundamentalmente relevante para el valor o afectar los precios de los activos a través de los gustos de los inversores (Heinkel et al., 2001). Sin embargo, en ambos casos, la divergencia de las calificaciones dispersa el efecto del rendimiento ESG en los precios de los activos. En cuarto lugar, el desacuerdo muestra que es difícil vincular la compensación del CEO con el desempeño ESG. Es probable que los contratos estén incompletos, y los CEOs pueden optimizar para una calificación en particular mientras tienen un rendimiento inferior en otros temas IMPORTANTES de ESG, es decir, los CEOs pueden alcanzar el objetivo establecido por la calificación, pero perder el punto de mejorar el desempeño ESG de la empresa de manera más amplia. Finalmente, la divergencia de calificaciones plantea un desafío para la investigación empírica, ya que el uso de un evaluador frente a otro puede alterar los resultados y conclusiones de un estudio. La divergencia de las calificaciones ESG introduce incertidumbre en cualquier decisión tomada sobre la base de las calificaciones ESG y, por lo tanto, representa un desafío para una amplia gama de tomadores de decisiones.

Este artículo investiga qué impulsa la divergencia de las calificaciones de sostenibilidad. Chatterji et al. (2016) han dado un primer paso importante en este sentido, proporcionando dos razones para la divergencia: lo que los evaluadores ESG eligen medir y si se mide de manera consistente, lo que los autores denominan "teorización" y "conmensurabilidad". En su análisis empírico, los autores muestran que tanto las diferencias en la teorización como la baja conmensurabilidad juegan un papel importante. Sin embargo, su análisis deja abierto hasta qué punto cada uno de estos componentes impulsa la divergencia. Como resultado, sigue sin estar claro si una mejor articulación de lo que se mide podría resolver la divergencia, o si la medición en sí misma es el problema central. Una razón clave para esta brecha restante es que Chatterji et al. (2016) se basan en un conjunto de datos que contiene solo un pequeño subconjunto de los indicadores subyacentes que componen las diferentes calificaciones ESG. Para avanzar en este frente, este documento proporciona una descomposición cuantitativa de la divergencia de calificación ESG, basándose en seis calificaciones ESG junto con el conjunto completo de 709 indicadores subyacentes.

Identificamos tres fuentes distintas de divergencia. La divergencia de alcance se refiere a la situación en la que las calificaciones se basan en diferentes conjuntos de atributos. Una agencia de calificación puede incluir actividades de cabildeo, mientras que otra podría no hacerlo, lo que hace que las dos calificaciones diverjan. La divergencia de medición se refiere a una situación en la que las agencias de calificación miden el mismo atributo utilizando diferentes indicadores. Por ejemplo, las prácticas laborales de una empresa podrían evaluarse sobre la base de la rotación de la fuerza laboral o por el número de casos judiciales relacionados con el trabajo tomados contra la empresa. Finalmente, la divergencia de peso surge cuando las agencias de calificación adoptan diferentes puntos de vista sobre la importancia relativa de los atributos. Por ejemplo, el indicador de prácticas laborales puede entrar en la calificación final con mayor peso que el indicador de cabildeo. Las contribuciones de alcance, medición y divergencia de peso están entrelazadas, lo que dificulta la interpretación de la diferencia entre dos calificaciones ESG.

Abordamos el problema en tres pasos. Primero, categorizamos todos los 709 indicadores proporcionados por los diferentes proveedores de datos en una taxonomía común de 64 categorías. Esta categorización es un paso crítico en nuestra metodología, ya que nos permite observar el alcance de las categorías cubiertas por cada calificación y contrastar las mediciones de diferentes evaluadores dentro de la misma categoría. Creamos una categoría siempre que al menos dos indicadores de diferentes agencias de calificación pertenezcan al mismo atributo. Sobre la base de la taxonomía, calculamos las puntuaciones de categorías específicas del evaluador promediando los indicadores que se asignaron a la misma categoría. En segundo lugar, retrocedemos la calificación original en esos puntajes de categoría. Los modelos de regresión producen versiones ajustadas de las calificaciones originales, y podemos comparar estas calificaciones ajustadas entre sí. En tercer lugar, descomponemos la divergencia en las contribuciones de alcance, medición y peso.

Nuestro estudio arroja tres resultados. En primer lugar, mostramos que es posible estimar la regla de agregación implícita utilizada por las agencias de calificación con una precisión del 79 al 99% en función de nuestra taxonomía común. Esto demuestra que, aunque las calificaciones ESG tienen estructuras incompatibles, es posible encajarlas en un marco coherente que revele en detalle cuánto y por qué razón difieren las calificaciones. En segundo lugar, encontramos que la divergencia de medición es el principal impulsor de la divergencia de calificación, contribuyendo con el 56% de la divergencia. La divergencia de alcance también es importante, ya que contribuye con un 38%, mientras que la divergencia de peso contribuye con un mero 6%. En tercer lugar, encontramos que la divergencia de medición es en parte impulsada por un efecto evaluador. Esto también se conoce como el "efecto halo", lo que significa que una empresa que recibe una puntuación alta en una categoría tiene más probabilidades de recibir puntuaciones altas en todas las demás categorías de ese mismo evaluador. El efecto evaluador es sustancial. Al controlar para qué empresa se califica y en qué categoría se califica a la empresa, el efecto evaluador explica el 15% de la variación de las puntuaciones de categoría.

Realizamos varias comprobaciones de robustez. En primer lugar, evaluamos la sensibilidad de los resultados a nuestra taxonomía. La taxonomía es una aproximación porque la mayoría de los evaluadores no comparten sus datos brutos, lo que hace imposible hacer coincidir los indicadores con las mismas unidades. Sin embargo, restringir el análisis a indicadores perfectamente idénticos daría como resultado que la divergencia se debe enteramente al alcance, es decir, a que no hay un terreno común entre los evaluadores ESG, lo que no refleja la situación real. Por lo tanto, utilizamos una taxonomía que coincide con los indicadores por atributo. Para descartar que nuestro juicio subjetivo impulse los resultados, clasificamos los indicadores de acuerdo con una taxonomía alternativa proporcionada por la Junta de Normas de Contabilidad de Sostenibilidad (SASB).3 Los resultados basados en esta taxonomía alternativa son consistentes con los basados en nuestra propia taxonomía. En segundo lugar, nuestra regla de agregación lineal no es específica de la industria, mientras que la mayoría de las agencias de calificación ESG utilizan reglas de agregación específicas de la industria. Sin embargo, esta aproximación parece ser relativamente inocua porque una regla lineal simple logra una calidad de ajuste muy alta. Los estimadores no lineales más sofisticados, como las redes neuronales, no producen mejores resultados. En tercer lugar, nuestro análisis principal se basa en el año 2014, que maximiza el tamaño de nuestra muestra e incluye KLD como una de las calificaciones que se ha utilizado en la academia con mayor frecuencia hasta ahora. Sin embargo, replicar el análisis para el año 2017 sin KLD arroja resultados muy similares. En cuarto lugar, presentamos un método de descomposición basado en la regresión como una metodología alternativa, que también respalda nuestros resultados.

Ampliamos la investigación existente que ha investigado la divergencia de las calificaciones ESG (Chatterji et al., 2016; Gibson Brandon et al., 2021; Christensen et al., 2021). Nuestra primera contribución a esta literatura es cuantificar los impulsores de la divergencia. Nuestros resultados muestran que la divergencia de calificación ESG está impulsada principalmente por la divergencia de medición, y es, por esa razón, difícil de resolver. El problema más fácil de abordar es la divergencia de peso. Dos calificaciones podrían hacerse consistentes alineando sus esquemas de ponderación. Sin embargo, debido a que la divergencia de peso contribuye solo un 6% a la divergencia total, el ajuste de las ponderaciones logrará poco. La divergencia de alcance es mucho más importante pero más difícil de abordar. La divergencia de alcance implica que las categorías son medidas exclusivamente por un evaluador. Por lo tanto, solo se puede lograr un mayor acuerdo concentrando la evaluación ASG en un conjunto reducido de categorías comunes. Sin embargo, la divergencia de medición sigue siendo el impulsor más relevante de la divergencia, incluso dentro de este conjunto más pequeño. Por lo tanto, abordar la divergencia de calificación ESG requiere que uno entienda cómo se generan los datos que sustentan las calificaciones ESG. Nuestra segunda contribución es una metodología que facilita el tratamiento de la divergencia de calificación ESG. A nivel de empresa, permite rastrear la divergencia a categorías individuales. A nivel agregado, permite identificar las categorías en las que la divergencia de medición es más consecuente, proporcionando áreas prioritarias para futuras investigaciones. Nuestra tercera contribución es el descubrimiento de un efecto evaluador. Esto sugiere que la divergencia de medición no es ruido distribuido aleatoriamente, sino que sigue patrones específicos del evaluador y de la empresa. Estos patrones sugieren razones estructurales para la divergencia de medición, como la forma en que las agencias de calificación organizan su trabajo.

Estos resultados tienen implicaciones importantes para futuras investigaciones en finanzas sostenibles. Las calificaciones y métricas ESG son una base importante para el campo de las finanzas sostenibles. La teoría predice que las preferencias de los inversores por ESG afectan a los precios de los activos. En la práctica, sin embargo, las opciones de inversión se guían por las calificaciones ESG, lo que hace que la construcción y el desacuerdo entre las calificaciones ESG sean una preocupación central. Por lo tanto, la investigación futura debe intentar mejorar la base empírica de las finanzas sostenibles. Los investigadores deben examinar cuidadosamente a los proveedores de datos y evitar depender demasiado de un solo evaluador como comunidad. Sin embargo, no es suficiente considerar múltiples calificaciones. El desacuerdo se extiende a categorías específicas de medio ambiente, sociales y de gobernanza, lo que significa que la medición ruidosa también plantea un desafío para la investigación sobre métricas ESG como las emisiones de carbono o la igualdad de género. Para abordar esto, los investigadores deben invertir en el desarrollo de sus propias métricas específicas de la categoría e idealmente ponerlas a disposición de otros. Además, el efecto evaluador plantea preguntas sobre la economía del mercado de calificaciones ESG. Las razones estructurales o los incentivos que deben entenderse mejor pueden influir en la forma en que se califican ciertas empresas o categorías. Más allá de mejorar la medición, la propia inmersión plantea la cuestión de cómo la incertidumbre en las calificaciones ESG afecta los precios de los activos, un tema que está ganando atención en la literatura (Avramov et al., 2021; Gibson Brandon et al., 2021). Finalmente, nuestros resultados plantean la cuestión de cómo las empresas responden a ser calificadas de manera diferente por diferentes evaluadores (ver también Chatterji et al., 2009), lo que informará los efectos de las finanzas sostenibles en la economía real.

La divergencia de calificación ESG no implica que medir el rendimiento ESG sea un ejercicio inútil. Sin embargo, destaca que medir el rendimiento ESG es un desafío, que la atención a los datos subyacentes es esencial y que el uso de calificaciones y métricas ESG debe considerarse cuidadosamente para cada aplicación. Los inversores pueden utilizar nuestra metodología para conciliar las calificaciones divergentes y centrar su investigación en aquellas categorías en las que las calificaciones no están de acuerdo. Para los reguladores, nuestro estudio señala los beneficios potenciales de armonizar la divulgación de ESG y establecer una taxonomía de categorías ESG. La armonización de la divulgación DE ESG ayudaría a proporcionar una base de datos confiables. Una taxonomía de categorías ESG facilitaría el contraste y la comparación de calificaciones.

El documento está organizado de la siguiente manera: la sección 2 describe los datos; La sección 3 documenta la divergencia de la calificación ESG; La Sección 4 explica la taxonomía y cómo estimamos los procedimientos de agregación. En la Sección 5, descomponemos la divergencia general en las contribuciones de alcance, medición y peso, y la Sección 6 explora el efecto evaluador. Concluimos en la Sección 7 y destacamos las implicaciones de nuestros hallazgos.

2. Datos

Las calificaciones ESG surgieron por primera vez en la década de 1980 como una forma para que los inversores evalúen a las empresas sobre el desempeño ambiental, social y de gobierno corporativo. La primera agencia de calificación ESG, Eiris (fusionada con Vigeo en 2015), se estableció en 1983 en Francia, y 7 años más tarde, Kinder, Lydenberg & Domini (KLD) se estableció en los Estados Unidos. Si bien inicialmente atendió a una clientela inversora altamente especializada, incluidas las organizaciones religiosas, el mercado de calificaciones ESG se ha ampliado drásticamente, especialmente en la última década. Debido a que las calificaciones ESG son una base esencial para la mayoría de los tipos de inversión sostenible, el mercado de las calificaciones ESG creció en paralelo a la inversión sostenible. A medida que la inversión sostenible pasó de nicho a la corriente principal, muchos de los primeros proveedores de calificación ESG fueron adquiridos por proveedores de datos financieros establecidos. Por ejemplo, MSCI compró KLD en 2010, Morningstar adquirió el 40% de Sustainalytics en 2017, Moody's compró Vigeo-Eiris en 2019 y S&P Global compró RobecoSAM en 2019.

Las agencias de calificación ESG permiten a los inversores evaluar el rendimiento DE LAS EMPRESAS, al igual que las calificaciones crediticias permiten a los inversores evaluar la solvencia crediticia de las empresas. Sin embargo, existen al menos tres diferencias importantes entre las calificaciones ESG y las calificaciones crediticias. En primer lugar, si bien la solvencia crediticia se define relativamente claramente como la probabilidad de incumplimiento, la definición de rendimiento ESG es menos clara. Es un concepto basado en valores diversos y en evolución. Por lo tanto, una parte importante del servicio que ofrecen las agencias de calificación ESG es una interpretación de lo que significa el desempeño ESG. En segundo lugar, si bien los estándares de información financiera han madurado y convergido durante el siglo pasado, los informes ESG están en su infancia. Existen estándares de información competitivos para la divulgación de ESG, muchos de los cuales son voluntarios o se limitan a jurisdicciones individuales, lo que otorga a las corporaciones una amplia discreción con respecto a si y qué informar. Por lo tanto, las calificaciones ESG proporcionan un servicio a los inversores mediante la recopilación y agregación de información de un espectro de fuentes y estándares de informes. Estas dos diferencias explican por qué la divergencia entre las calificaciones ESG es mucho más pronunciada que la divergencia entre las calificaciones crediticias, esta última correlacionada en el 99%.4 En tercer lugar, los evaluadores ESG son pagados por los inversores que utilizan las calificaciones, no por las empresas que son calificadas, como es el caso de los evaluadores de crédito. Como resultado, el problema de la compra de calificaciones, que se ha discutido como una posible razón para la divergencia de las calificaciones crediticias (véase, por ejemplo, Bongaerts et al., 2012), no se aplica a los proveedores de calificaciones ESG.

Utilizamos datos de seis proveedores de calificación ESG diferentes: KLD5, Sustainalytics, Moody's ESG, Refinitiv, MSCI y S&P Global. Incluimos KLD porque es el conjunto de datos que se ha utilizado con mayor frecuencia en los estudios académicos. Nuestra selección de los otros evaluadores se guió por la relevancia del mercado. Todos los proveedores de nuestra muestra son ampliamente reconocidos y utilizados por profesionales de finanzas sostenibles.6 Nos acercamos a cada proveedor y solicitamos acceso a las calificaciones, los indicadores subyacentes y la documentación sobre las reglas de agregación y los protocolos de medición de los indicadores.

En la Tabla 1 se presentan estadísticas descriptivas de las calificaciones agregadas7 y sus características muestrales. El año de referencia para nuestro análisis es 2014, que es el año con la muestra común más grande cuando también se incluye KLD. Debido a que la mayor parte de la literatura académica hasta la fecha se basa en datos de KLD, es importante incluirlos en nuestro estudio. También confirmamos nuestros resultados para el año 2017 sin KLD. El Panel A muestra la muestra completa, donde el número de empresas oscila entre 1.665 y 9.662. El Panel B muestra la muestra común de 924 empresas. Las calificaciones ESG medias y medianas son más altas en la muestra común para todos los proveedores, lo que indica que la muestra equilibrada tiende a dejar caer a las empresas de menor rendimiento. Para nuestro análisis posterior, normalizamos la muestra común para que tenga una media cero y una varianza unitaria en la sección transversal.

3. Discrepancia

Nuestro punto de partida es que las calificaciones ESG divergen. En esta sección, establecemos que la divergencia en nuestra muestra es sustancial y consistente con estudios previos. Primero, calculamos el alfa de Krippendorff (Krippendorff, 2004). La ventaja de esta medida es que expresa la fiabilidad general de la evaluación para cualquier número de evaluadores en una estadística. Para los seis evaluadores y 924 empresas de nuestra muestra, obtenemos un valor de 0,55. En general, se prefieren valores por encima de 0,8, y los valores por encima de 0,667 se consideran un mínimo para llegar a conclusiones tentativas sobre el valor verdadero basado en las evaluaciones de los evaluadores (Krippendorff, 2004, p. 204). En otras palabras, el desacuerdo es sustancial.

La Tabla 2 muestra las correlaciones de Pearson por pares entre las calificaciones ESG agregadas y entre sus dimensiones ambiental (E), social (S) y de gobernanza (G). Las correlaciones a nivel ESG son en promedio 0.54 y oscilan entre 0.38 y 0.71. Sustainalytics y Moody's ESG tienen el mayor nivel de acuerdo entre sí, con una correlación de 0,71. La dimensión ambiental tiene la mayor correlación de las tres dimensiones, con un promedio de 0,53. La dimensión social tiene una correlación promedio de 0.42, y la dimensión de gobernanza tiene la correlación más baja, con un promedio de 0.30. KLD y MSCI exhiben las correlaciones más bajas con otros evaluadores, tanto para la calificación ESG agregada como para las dimensiones individuales. Teniendo en cuenta las diferencias de muestreo, estos resultados son consistentes con las correlaciones de calificación ESG informadas por Chatterji et al. (2016) y Gibson Brandon et al. (2021).

Ilustramos la divergencia de calificación en la Figura 1. Para no exagerar el problema, utilizamos la calificación Sustainalytics, que tiene las correlaciones más altas con las otras cinco calificaciones, como punto de referencia. Trazamos los valores de los otros evaluadores contra esta calificación de referencia. La Figura 1 ilustra que las calificaciones ESG están correlacionadas positivamente. A medida que aumenta la calificación de referencia, las otras calificaciones también tienden a aumentar. Sin embargo, la cifra también muestra divergencias sustanciales. Para cualquier nivel de la calificación de referencia, hay una amplia gama de valores dados por las otras calificaciones. Supongamos que una empresa recibe una puntuación de calificación Sustainalytics de +1.5, lo que la coloca entre el 10% superior de las empresas calificadas por Sustainalytics. Sin embargo, otras calificaciones en x = 1.5 califican a la compañía por debajo de cero, colocando a la compañía por debajo del promedio de la muestra. En otras palabras, el alcance de la divergencia es tal que es difícil distinguir a un líder de un ejecutante promedio. Este problema se vuelve aún más pronunciado cuando se usan otras calificaciones como punto de referencia o cuando se observan las clasificaciones. Para obtener una ilustración, consulte la Figura A.1 del Apéndice de Internet.

El propósito de las calificaciones ESG es evaluar el desempeño ESG de una empresa. Sin embargo, las calificaciones ESG no están de acuerdo hasta un punto que deja a los observadores con una incertidumbre considerable en cuanto a qué tan bueno es el desempeño ESG de la empresa. Es natural que aquellos interesados en el desempeño ESG se pregunten qué hace que las calificaciones ESG estén en desacuerdo tan ampliamente. Esto es lo que investigamos a continuación.

4. Alcance, medida y pesos

En esta sección, explicamos cómo especificamos las calificaciones ESG en términos de alcance, medición y peso en función de una taxonomía común. Este marco, ilustrado en la Figura 2, nos permite explicar por qué las calificaciones divergen.

4.1 Ámbito de aplicación

La descomposición de la divergencia de calificación ESG no es trivial porque a nivel granular, las estructuras de las diferentes calificaciones ESG son incompatibles. Cada evaluador elige dividir el concepto de rendimiento ESG en diferentes indicadores y los organiza en diferentes jerarquías. Por ejemplo, en el primer nivel de desagregación, Moody's ESG, S&P Global, MSCI y Sustainalytics tienen tres dimensiones (E, S y G), Refinitiv tiene cuatro y KLD tiene siete. Por debajo de estas dimensiones de primer nivel, hay entre uno y tres niveles de subcategorías más granulares, dependiendo del evaluador. En el nivel más bajo, nuestro conjunto de datos contiene entre 38 y 282 indicadores por evaluador, que a menudo, pero no siempre, se relacionan con atributos subyacentes similares. Estas estructuras incompatibles hacen que sea difícil entender cómo y por qué diferentes evaluadores evalúan a la misma empresa de diferentes maneras.

Imponemos nuestra propia taxonomía sobre los datos para realizar una comparación significativa de estos diferentes sistemas de calificación. Desarrollamos esta taxonomía utilizando un enfoque de abajo hacia arriba. Primero, creamos una larga lista de todos los indicadores disponibles y sus descripciones detalladas. En los casos en que las descripciones no estaban disponibles (o eran insuficientes), se entrevistó a los proveedores de datos para su aclaración. En total, la lista contiene 709 indicadores. En segundo lugar, agrupamos indicadores que describen el mismo atributo en la misma categoría. Por ejemplo, agrupamos todos los indicadores relacionados con la categoría Agua, como se muestra en la Tabla 3. En tercer lugar, refinamos iterativamente la taxonomía, siguiendo dos reglas: (a) cada indicador se asigna a una sola categoría, y (b) se establece una nueva categoría cuando al menos dos indicadores de diferentes evaluadores describen un atributo que aún no está cubierto por las categorías existentes. Por ejemplo, los indicadores relacionados con los bosques se sacaron de la categoría más amplia de Biodiversidad para formar su propia categoría. La clasificación se basa puramente en el atributo que los indicadores pretenden medir, independientemente del método o fuente de datos utilizado. Los indicadores que son exclusivos de un evaluador y no se pueden agrupar con indicadores de otros evaluadores se etiquetaron como "no clasificados" y a cada uno se le dio su propia categoría específica de evaluador.

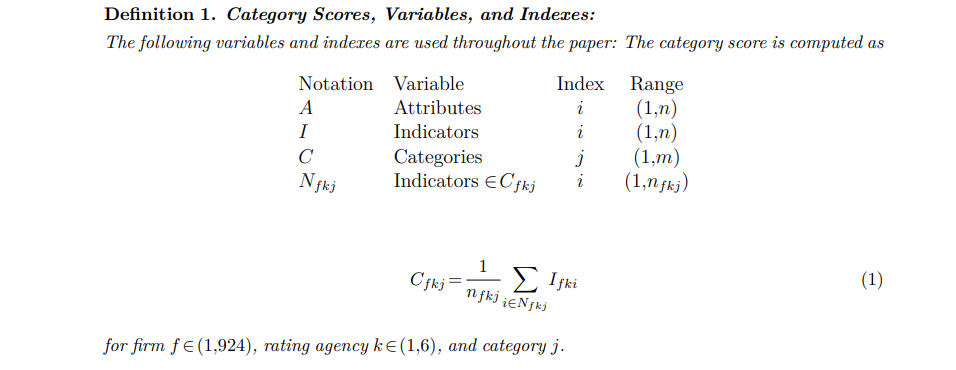
La taxonomía resultante, que se muestra en la Tabla 4, asigna los 709 indicadores a un total de 64 categorías distintas. Refinitiv tiene la mayoría de los indicadores individuales con 282, seguido de Sustainalytics con 163. KLD, S&P Global y MSCI tienen 78, 80 y 68, respectivamente, y Moody's ESG tiene 38. Algunas categorías (Bosques, por ejemplo) contienen sólo un indicador de dos evaluadores. Otros, como Supply Chain, contienen varios indicadores de todos los evaluadores. Podría decirse que los bosques son una categoría mucho más estrecha que la cadena de suministro. La razón de esta diferencia en la amplitud es que no había indicadores en la cadena de suministro que en conjunto representaran una categoría común más estrecha. Por lo tanto, la comparación de la cadena de suministro se encuentra en un nivel más general, y puede parecer obvio que los diferentes evaluadores tienen una visión diferente de esta categoría. Sin embargo, esta amplia comparación representa el nivel más específico posible dados los datos.

La Tabla 4 muestra cuántos indicadores proporciona cada evaluador por categoría. Por un lado, algunas categorías son consideradas por los seis evaluadores, lo que indica que se trata de cuestiones ESG básicas comúnmente aceptadas. Estos son Biodiversidad, Desarrollo de Empleados, Energía, Productos Verdes, Salud y Seguridad, Prácticas Laborales, Seguridad de Productos, Remuneración, Cadena de Suministro y Agua. Por otro lado, muchas celdas vacías muestran que lejos de todas las categorías están cubiertas por todas las calificaciones. Existen lagunas tanto para las categorías que podrían describirse como especializadas, como los campos electromagnéticos, como para la categoría de Impuestos, que podría considerarse una preocupación fundamental en el contexto de las cuestiones ASG. Además, el considerable número de indicadores no clasificados muestra que muchos aspectos ESG solo son medidos por uno de cada seis evaluadores. Refinitiv tiene, con 42, los indicadores más no clasificados, casi todos los cuales provienen de la dimensión económica de Refinitiv. Esta dimensión contiene indicadores, como el crecimiento de los ingresos netos o los gastos de capital, que otras agencias de calificación no consideran. MSCI tiene 34 indicadores no clasificados; esto es lo que MSCI denomina "puntajes de exposición". Junto a los puntajes que evalúan qué tan bien una empresa maneja un problema, MSCI tiene puntajes que miden qué tan relevante es el problema para la empresa específica. Ninguno de los otros evaluadores tiene indicadores que midan explícitamente esto.

La taxonomía impone una estructura a los datos que permite una comparación sistemática. Obviamente, los resultados pueden ser sensibles a la forma particular en que lo construimos. Para asegurarnos de que nuestros resultados no estén impulsados por una clasificación en particular, creamos una taxonomía alternativa como una verificación de robustez. En lugar de construir las categorías de abajo hacia arriba, producimos una taxonomía de arriba hacia abajo basada en SASB. SASB ha identificado 26 categorías de temas generales basadas en un proceso integral de consulta con las partes interesadas. Como tal, estas categorías representan el consenso de una amplia gama de inversores y reguladores sobre el alcance de las categorías ESG relevantes. Mapeamos todos los indicadores contra estas 26 categorías de problemas generales, lo que nuevamente requiere que cada indicador solo se pueda asignar a una categoría. Esta taxonomía alternativa, junto con los resultados que se basan en ella, se proporciona en el Apéndice de Internet. Todos nuestros resultados también son válidos para esta taxonomía alternativa.

4.2 Medición

Podemos estudiar la divergencia de medición utilizando nuestra taxonomía comparando las evaluaciones de diferentes evaluadores a nivel de categorías. Creamos puntajes de categoría (C) para cada categoría, empresa y evaluador. Las puntuaciones de categoría se calculan tomando el promedio de los valores de los indicadores asignados a la categoría. Definamos las notaciones:



Los puntajes de categoría representan la evaluación de una agencia de calificación de una determinada categoría ESG. Se basan en diferentes conjuntos de indicadores que se basan en diferentes protocolos de medición. Se deduce que esas diferencias entre los puntajes de categoría se derivan de las diferencias en la forma en que las agencias de calificación eligen medir, en lugar de lo que eligen medir. Por lo tanto, las diferencias entre las mismas categorías de diferentes evaluadores pueden interpretarse como divergencia de medición. Algunas agencias de calificación emplean diferentes conjuntos de indicadores para diferentes industrias. Tales consideraciones específicas de la industria sobre la medición también se reflejan en los puntajes de categoría porque esos puntajes toman el promedio de todos los valores de indicadores disponibles.

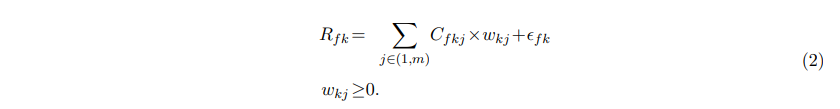
La Tabla 5 muestra las correlaciones entre las categorías. Las correlaciones se calculan sobre la base de observaciones completas por pares por categoría y par de evaluadores. La tabla ofrece dos ideas. En primer lugar, los niveles de correlación son heterogéneos. La política ambiental, por ejemplo, tiene un nivel de correlación promedio de 0,55. Esto indica al menos cierto nivel de acuerdo con respecto a la existencia y la calidad de la política ambiental de las empresas. Sin embargo, incluso las categorías que miden hechos sencillos que se obtienen fácilmente de los registros públicos no todas tienen altos niveles de correlación. Por ejemplo, la membresía del Pacto Mundial de las Naciones Unidas y la separación CEO/Presidente deben ser inequívocas, pero mostrar correlaciones de 0,92 y 0,59, respectivamente. También hay varias correlaciones negativas, como el cabildeo entre Sustainalytics y Moody's ESG o los derechos indígenas entre Sustainalytics y Refinitiv. En estos casos, el nivel de desacuerdo es tan severo que las agencias de calificación llegan no solo a conclusiones diferentes sino opuestas.

La segunda idea es que las correlaciones tienden a aumentar con la granularidad. Por ejemplo, las correlaciones de las categorías Agua y Energía son en promedio 0.36 y 0.38, respectivamente. Esto es sustancialmente menor que la correlación de la dimensión ambiental, con un promedio de 0,53 reportado en la Tabla 2. Esto implica que las divergencias se compensan entre sí hasta cierto punto durante la agregación. Varias razones potenciales pueden explicar esta observación. Una razón podría ser que las puntuaciones de categoría se comportan como medidas ruidosas de una calidad latente subyacente, de modo que el desacuerdo de medición en categorías individuales se cancela durante la agregación. También puede darse el caso de que las agencias de calificación evalúen a una empresa de manera relativamente estricta en una categoría y relativamente indulgente en otra. Una preocupación podría ser que las bajas correlaciones a nivel de categoría son el resultado de una clasificación errónea en nuestra taxonomía, en el sentido de que los indicadores altamente correlacionados se clasificaron en diferentes categorías. Si bien no podemos descartar esto por completo, la taxonomía alternativa basada en los criterios de SASB mitiga esta preocupación. Es una clasificación mucho menos granular, que, por lo tanto, debería disminuir la influencia de cualquier clasificación errónea. Sin embargo, la correlación promedio por par de evaluadores cambia solo un poco y no sistemáticamente cuando se usa esta taxonomía alternativa. Esto proporciona la seguridad de que los niveles de correlación observados no son un artefacto de clasificación errónea en nuestra taxonomía. Las correlaciones con la taxonomía basada en los criterios sasb pueden verse en el cuadro A.3 del Apéndice de Internet.

4.3 Peso

Podemos proceder con un análisis de la divergencia de peso en función de las puntuaciones de categoría. Para ello, estimamos la regla de agregación que transforma las puntuaciones de categoría Cfkj en la calificación Rfk para cada evaluador k. Las puntuaciones de categoría, tal como se definen en la Sección 4.2, sirven como variables independientes. Cuando no hay valores de indicador disponibles para calcular la puntuación de categoría para una empresa determinada, la puntuación se establece en cero. Esto es necesario para ejecutar regresiones sin eliminar todas las categorías con valores faltantes, que son numerosos. Por supuesto, esto implica la suposición de que los datos faltantes indican un rendimiento deficiente. Se eliminan las categorías para las que no se dispone de valores para ninguna empresa de la muestra común. Después de este tratamiento, las puntuaciones de categoría se normalizan a cero media y varianza unitaria, correspondiente a las calificaciones normalizadas. Cada indicador no clasificado se trata como una categoría específica del evaluador separada.

Realizamos una regresión de mínimos cuadrados no negativos, que incluye la restricción de que los coeficientes no pueden ser negativos. Esto se debe a que conocemos a priori la direccionalidad de todos los indicadores y, por lo tanto, podemos descartar pesos negativos en una función lineal. Por lo tanto, estimamos los pesos (wkj) con la siguiente especificación:



Debido a que todos los datos se han normalizado, excluimos el término constante. Debido a la restricción de no negatividad, calculamos los errores estándar por Bootstrap. Nos centramos en el R2 como una medida de la calidad del ajuste.

Los resultados se muestran en la Tabla 6. MSCI tiene el R2 más bajo, con 0.79. Sustainalytics es el segundo más bajo, con 0.90. Las regresiones para KLD, Moody's ESG, Refinitiv y S&P Global tienen valores de R2 de 0.99, 0.96, 0.92 y 0.98, respectivamente. Estos altos valores de R2 indican que un modelo lineal basado en nuestra taxonomía es capaz de replicar las calificaciones originales con bastante precisión.

Los coeficientes de regresión pueden interpretarse como ponderaciones de categoría. Debido a que todas las variables se han normalizado, la magnitud de los coeficientes es comparable e indica la importancia relativa de una categoría. La mayoría de los coeficientes son muy significativos. Algunos coeficientes no son significativos en el umbral del 5%, lo que significa que nuestro peso estimado es incierto. Sin embargo, esos coeficientes son mucho más pequeños en magnitud que los coeficientes significativos; la mayoría de ellos están cerca de cero y, por lo tanto, no parecen tener una influencia importante en la calificación ESG agregada.

Hay diferencias sustanciales en los pesos para los diferentes evaluadores. Por ejemplo, las tres categorías más importantes para KLD son la gestión de riesgos climáticos, la seguridad de los productos y la remuneración. Para Moody's ESG, los tres primeros son Diversidad, Política Ambiental y Prácticas Laborales. Esto significa que no hay superposición en las tres categorías más importantes para estos dos evaluadores. Solo la eficiencia de los recursos y la gestión del riesgo climático se encuentran entre las tres categorías más importantes para más de un evaluador. Al mismo tiempo, algunas categorías tienen un peso cero para todos los evaluadores, como los ensayos clínicos y las multas ambientales, los OGM y los gases que agotan la capa de ozono. Estas observaciones resaltan que los diferentes evaluadores tienen puntos de vista sustancialmente diferentes sobre las categorías más importantes. En otras palabras, hay divergencia de peso entre los evaluadores.

Primero, ejecutamos una regresión ordinaria de mínimos cuadrados para relajar la restricción de no negatividad. Hacerlo solo conduce a pequeños cambios y no mejora la calidad del ajuste para ningún evaluador. En segundo lugar, ejecutamos redes neuronales para permitir una forma no lineal y flexible de la función de agregación. Como las redes neuronales son propensas al sobreajuste, informamos el ajuste fuera de la muestra. Asignamos aleatoriamente el 10% de las empresas a un conjunto de pruebas y el resto a un conjunto de capacitación. Para ofrecer una comparación adecuada, comparamos su rendimiento con el R2 fuera de muestra equivalente para el procedimiento de mínimos cuadrados no negativos. Ejecutamos una red neuronal de una capa oculta con una función de activación lineal y otra con una función de activación relu. Ambos funcionan notablemente mejor para MSCI, pero no para ninguno de los otros evaluadores. Esto implica que la regla de agregación de la calificación MSCI es, hasta cierto punto, no lineal. La explicación relativamente simple parece ser los pesos específicos de la industria. En pruebas no reportadas, confirmamos que la calidad de ajuste para MSCI está muy por encima de 0.90 en submuestras de la industria, incluso para una regresión lineal. En tercer lugar, implementamos un estimador de bosque aleatorio como una técnica alternativa no lineal. Sin embargo, este enfoque produce valores de R2 sustancialmente más bajos para la mayoría de los evaluadores.

También comprobamos si la taxonomía que impusimos a los indicadores originales influye en la calidad del ajuste. Con este fin, replicamos la estimación de mínimos cuadrados no negativos de la regla de agregación utilizando la taxonomía SASB (Ver Tabla A.4 en el Apéndice de Internet). La calidad del ajuste es prácticamente idéntica. Finalmente, ejecutamos una regresión ordinaria de mínimos cuadrados sin ninguna taxonomía, retrocediendo los indicadores originales de cada evaluador en las calificaciones. La calidad del ajuste también es muy similar; el cambio más notable es un pequeño aumento de 0.03 para la calificación MSCI. Finalmente, realizamos la regresión utilizando datos del año 2017 (sin KLD) en lugar de 2014 (ver Tabla A.8 en el Apéndice de Internet). En este caso, la calidad del ajuste es ligeramente inferior para MSCI y Refinitiv, lo que indica que sus metodologías han cambiado con el tiempo. En resumen, concluimos que el modelo de mínimos cuadrados negativos logra una alta calidad de ajuste, y los resultados de la estimación son robustos.

5. Descomposición

Hasta ahora, hemos demostrado que existe una divergencia de alcance, medición y peso. En esta sección, descomponemos la divergencia general de las calificaciones en las contribuciones de alcance, medición y divergencia de peso. Realizamos una descomposición aritmética que se basa en la taxonomía, las puntuaciones de categoría y los pesos de agregación estimados en la Sección 4. La divergencia de alcance se parcializa considerando solo las categorías que están contenidas exclusivamente en una de las dos calificaciones. La divergencia de medición se aísla calculando ambas clasificaciones con ponderaciones idénticas, de modo que las diferencias solo pueden provenir de diferencias en la medición. La divergencia de peso es lo que queda de la diferencia total.

5.1 Resultados de la descomposición

La Figura 3 proporciona un ejemplo específico de la empresa de la descomposición. La figura muestra en detalle cómo descomponemos la diferencia de calificación entre Refinitiv y KLD para Barrick Gold Corporation. Ilustra cómo nuestra descomposición desglosa completamente la diferencia entre dos calificaciones ESG en contribuciones específicas de la categoría de alcance, medición y peso.

Los resultados transversales de la descomposición se presentan en la Tabla 8. En promedio, en todos los pares de evaluadores, la divergencia de medición hace la mayor contribución con el 56%, seguida de la divergencia de alcance con el 38% y la divergencia de peso con el 6%. Más de la mitad de la divergencia de calificación ESG puede atribuirse a agencias de calificación ESG que miden diferentes valores diferentes para la misma categoría.

Los resultados para los pares de evaluadores individuales se alinean muy bien con las expectativas. Por ejemplo, para el par KLD-MSCI, la divergencia de medición contribuye solo con el 17%, mientras que el alcance contribuye con el 81%. Este resultado refleja que esas dos calificaciones provienen del mismo proveedor, probablemente se basan en datos subyacentes muy similares, pero cubren un alcance diferente de atributos. El par Sustainalytics-Refinitiv, con un 22%, tiene la mayor contribución de peso y al mismo tiempo, con un 12%, la menor contribución de alcance. Sustainalytics y Refinitiv tienen muchos indicadores, y la mayoría de las categorías están cubiertas por ambos evaluadores. En este caso, la divergencia de alcance juega un papel menor; en cambio, hay más categorías para las cuales los pesos pueden diferir. El par Moody's ESG-Refinitiv, con un 78%, tiene la mayor contribución de divergencia de medición. Esto sugiere que esos dos evaluadores tienen puntos de vista muy similares sobre de qué se trata ESG. Sin embargo, las diferencias permanecen, y estas se deben principalmente a la divergencia de medición.

El Panel B destaca las diferencias entre los evaluadores. MSCI se destaca como el único evaluador donde el alcance en lugar de la medición contribuye más a la divergencia. Este resultado es impulsado por los puntajes de exposición de MSCI. Como se describe en la Sección 4, estos puntajes esencialmente establecen ponderaciones específicas de la empresa para cada categoría. Como estos puntajes no tienen equivalente en los otros métodos de calificación, aumentan la divergencia de alcance de MSCI con respecto a todos los demás evaluadores. Al mismo tiempo, la contribución del peso es negativa para MSCI debido a una covarianza negativa entre el alcance y la divergencia de peso. En otras palabras, los efectos de la divergencia de alcance y la divergencia de peso tienden a compensarse entre sí en el caso de MSCI. Para todos los demás evaluadores, excepto MSCI, la contribución disminuye de la medición al alcance y al peso.

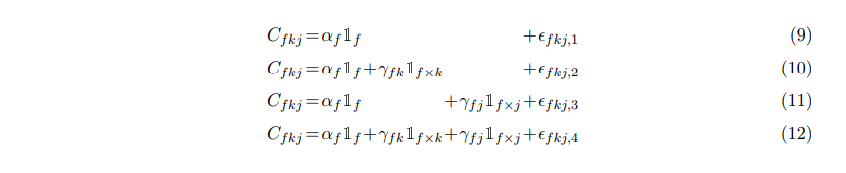
Nuestro análisis también nos permite identificar las categorías que son más importantes para la divergencia de medición en todos los evaluadores. Con este fin, promediamos la divergencia absoluta de medición por categoría. Resulta que algunas categorías para las que existe un pronunciado desacuerdo de medición en última instancia no importan mucho para la divergencia de calificación porque tienden a tener un pequeño peso en las calificaciones agregadas. Estos incluyen multas ambientales, ensayos clínicos, rotación de empleados, programas de VIH y emisiones atmosféricas no GEI. En el otro extremo del espectro se encuentran las categorías donde la divergencia de medición es muy importante para la divergencia general, a saber, Climate Risk Mgmt., Product Safety, Corporate Governance, Corruption, and Environmental Mgmt. System. Estas últimas categorías son objetivos prioritarios en términos de abordar la divergencia de medición.

6. Efecto de pérdida

Para investigar más a fondo las razones subyacentes de la divergencia de medición, esta sección prueba la presencia de un efecto evaluador. El efecto evaluador describe un sesgo, donde el rendimiento en una categoría influye en el rendimiento percibido en otras categorías. Este fenómeno también se llama el "efecto halo", y los sesgos relacionados se han estudiado ampliamente en sociología, gestión y psicología, especialmente en la evaluación del desempeño (ver Shrout y Fleiss, 1979). El proceso de evaluación de los atributos ESG de las empresas parece propenso a un efecto evaluador. La evaluación del desempeño de las empresas en las categorías de Derechos Humanos, Comunidad y Sociedad, y Prácticas Laborales requiere que las agencias de calificación utilicen cierto grado de juicio. El efecto evaluador implica que cuando el juicio de una empresa es positivo para un indicador en particular, también es probable que sea positivo para otro indicador. Evaluamos el efecto evaluador mediante dos procedimientos. En primer lugar, estimamos las regresiones de efectos fijos comparando categorías, empresas y evaluadores. En segundo lugar, ejecutamos regresiones LASSO específicas del evaluador para evaluar la contribución marginal de cada categoría.

6.1 Efectos fijos de la tasa

El primer procedimiento se basa en regresiones simples de efectos fijos. Los puntajes de categoría de una empresa dependen de la propia empresa, de la agencia de calificación y de la categoría que se califica. Examinamos en qué medida esos efectos fijos aumentan el poder explicativo en el siguiente conjunto de regresiones:



donde 1f son dummies para cada empresa, 1f×k es un término de interacción entre los efectos fijos de la empresa y el evaluador, y 1f×j es un término de interacción entre los efectos fijos de la empresa y la categoría. El vector Cfkj apila las puntuaciones de todas las categorías comunes en todos los evaluadores y empresas. Eliminamos los efectos fijos de categoría pura y evaluador debido a la normalización en el nivel de calificación y puntuación de categoría. Solo utilizamos la intersección de categorías de todos los evaluadores y la muestra común de empresas para reducir el sesgo de la muestra.

6.2 UN LAZO

Enfoque del efecto Evaluador Exploramos el efecto evaluador utilizando un procedimiento alternativo. Aquí, nos concentramos exclusivamente en la variación dentro del evaluador. Una agencia de calificación sin efecto evaluador es aquella en la que las correlaciones entre categorías son relativamente pequeñas; una agencia de calificación con un fuerte efecto evaluador implica que las correlaciones son altas. Sin embargo, estas correlaciones no se pueden resumir con precisión mediante comparaciones por pares. En su lugar, podemos probar las correlaciones entre categorías utilizando regresiones LASSO. La idea es que un fuerte efecto evaluador implica que el poder explicativo marginal de cada categoría dentro de un evaluador está disminuyendo cuando las categorías se agregan una tras otra. Esto implica que uno podría replicar una calificación general con menos del conjunto completo de categorías.

El efecto evaluador sugiere que la divergencia de medición no es solo ruido distribuido aleatoriamente. En cambio, una parte de la divergencia sigue un patrón que sugiere razones estructurales. Una posible explicación para el efecto evaluador es que las agencias de calificación están organizadas de modo que los analistas se especializan en empresas en lugar de indicadores. Una empresa que se percibe como buena en general puede ser vista a través de una lente positiva y recibir mejores puntajes de indicadores que una empresa que se percibe como mala en general. En conversaciones con S&P Global, aprendimos sobre otra causa potencial para tal efecto evaluador. Algunos evaluadores hacen imposible que las empresas reciban una buena puntuación indicadora si no dan una respuesta a la pregunta correspondiente en el cuestionario. Esto sucede independientemente del rendimiento real del indicador. La medida en que las empresas responden a preguntas específicas puede estar correlacionada entre indicadores. Por lo tanto, el efecto evaluador también podría deberse a supuestos específicos del evaluador que afectan sistemáticamente a las evaluaciones. También podría haber incentivos económicos que afecten la medición. Por ejemplo, Cornaggia et al. (2017) sugieren que los evaluadores de crédito pueden tener incentivos para inflar ciertas calificaciones. Una vía interesante para futuras investigaciones es si los evaluadores ESG tienen incentivos similares para ajustar sus calificaciones.

7. Conclusiones

La contribución de este artículo es una descomposición de la divergencia de calificaciones ESG. Chatterji et al. (2016) han dado un primer paso importante al distinguir dos aspectos que importan, en primer lugar, cómo los evaluadores ESG definen lo que pretenden medir, y en segundo lugar, cómo lo miden. Sin embargo, su análisis deja abierto hasta qué punto estos dos aspectos impulsan la divergencia. Como resultado, la diferencia entre dos calificaciones cualesquiera sigue siendo difícil de interpretar. En este artículo, descomponemos la divergencia en los elementos de alcance, peso y medición. El alcance y el peso reflejan lo que una calificación ESG pretende medir, mientras que la medición refleja cómo se mide. Mostramos que la divergencia de medición es el principal impulsor de la divergencia de calificación ESG. Nuestros hallazgos demuestran que la divergencia de calificación ESG no es simplemente una cuestión de definiciones variables, sino un desacuerdo fundamental sobre los datos subyacentes. Es legítimo que los diferentes evaluadores adopten diferentes puntos de vista sobre qué categorías son más importantes en la evaluación ESG. Una variedad de opiniones puede ser deseable dado que los usuarios de las calificaciones ESG también tienen preferencias heterogéneas por el alcance y el peso. Sin embargo, la divergencia de medición es problemática si se acepta la opinión de que las calificaciones ESG deben basarse en última instancia en observaciones objetivas que puedan determinarse.

La segunda contribución de este trabajo es en el frente metodológico. Este es el primer documento que compara varias calificaciones ESG basadas en el conjunto completo de indicadores subyacentes. Demostramos que es posible volver a estimar las calificaciones ESG basadas en una taxonomía común impuesta a los datos. A nivel de empresa, esto explica por qué dos métodos de calificación ESG diferentes producen evaluaciones diferentes. En conjunto, permite identificar las categorías que son más importantes para impulsar la divergencia de calificación ESG. La divergencia de medición es más influyente en las categorías Climate Risk Mgmt., Product Safety, Corporate Governance, Corruption, y Environmental Mgmt. Estas categorías son puntos de partida naturales para futuras investigaciones sobre la mejora de los enfoques de medición en las calificaciones ESG.

En tercer lugar, documentamos un efecto evaluador. Las evaluaciones de los evaluadores se correlacionan entre categorías, de modo que cuando una agencia de calificación le da a una empresa una buena puntuación en una categoría, tiende a darle a esa empresa buenas puntuaciones en otras categorías, también. El efecto evaluador sugiere que la divergencia de medición no es simplemente ruido, sino que los patrones influyen en la forma en que se evalúan las empresas. Aunque no identificamos de manera concluyente la causa del efecto evaluador, una posible explicación es que las agencias de calificación ESG dividen el trabajo de los analistas por empresa y no por categoría, de modo que la visión general de un analista de una empresa podría propagarse en las evaluaciones en diferentes categorías. Una vía prometedora para futuras investigaciones es investigar razones adicionales por las que las calificaciones ESG podrían desviarse sistemáticamente en su evaluación, por ejemplo, si existen incentivos económicos para ajustar las calificaciones.

Nuestros resultados tienen implicaciones importantes para investigadores, inversores, empresas, agencias de calificación y reguladores. Los investigadores deben elegir cuidadosamente los datos que subyacen a los futuros estudios ESG. Los resultados obtenidos sobre la base de una calificación ESG podrían no replicarse con las calificaciones ESG de otra agencia de calificación. En particular, nuestros resultados indican que la divergencia es muy pronunciada para KLD, los datos en los que se ha basado hasta ahora la mayoría de la investigación académica existente sobre ESG. Los investigadores tienen tres opciones cuando se trata de lidiar con la divergencia de las calificaciones ESG. Una es incluir varias calificaciones ESG en el análisis (véase, por ejemplo, Liang y Renneboog, 2017). Esto es razonable cuando la intención es medir el "rendimiento ESG de consenso" tal como lo perciben los mercados financieros en los que se utilizan varias calificaciones. En segundo lugar, los investigadores pueden usar una calificación ESG en particular para medir una característica específica de la empresa. En este caso, se debe explicar cuidadosamente por qué la metodología de calificación específica es la más adecuada para el estudio. En tercer lugar, los investigadores pueden construir hipótesis en torno a subcategorías más específicas del rendimiento ESG, como las emisiones de GEI o las prácticas laborales. Esto evita los problemas de divergencia de peso y alcance, pero el riesgo de divergencia de medición permanece. Por lo tanto, los investigadores idealmente deberían trabajar con datos sin procesar que puedan verificarse de forma independiente. Si eso no es factible, los investigadores deben examinar cuidadosamente cómo se generan los datos y permanecer escépticos de los datos donde el proceso de generación de datos no es completamente transparente. Cuando no se dispone de datos de alta calidad, los investigadores también deben considerar la recopilación de datos ESG por sí mismos y compartir el conjunto de datos. En resumen, dada la divergencia de la calificación ESG, cualquier investigación que utilice calificaciones o métricas ESG debe prestar especial atención a la validez de los datos utilizados.

En cuanto a los inversores, nuestra metodología les permite comprender por qué una empresa ha recibido diferentes calificaciones de diferentes agencias de calificación. El ejemplo de la Figura 3 ilustra cómo los inversores pueden desenredar las diversas fuentes de divergencia y rastrear un resultado a categorías específicas. Por ejemplo, los inversores podrían reducir la discrepancia entre las calificaciones obteniendo datos a nivel de indicador de varios evaluadores y luego imponiendo su propio alcance y peso. La divergencia de medición restante podría rastrearse hasta los indicadores que están impulsando la discrepancia, guiando la investigación adicional de un inversor. Promediar indicadores de diferentes proveedores también es una manera fácil de eliminar la divergencia de medición. Sin embargo, el efecto evaluador sugiere que este enfoque puede ser problemático porque las discrepancias no se distribuyen al azar. Alternativamente, los inversores pueden confiar en una agencia de calificación después de convencerse de que el alcance, la medición y el peso están alineados con sus objetivos.

Para las empresas, nuestros resultados ponen de relieve el desacuerdo sustancial sobre su desempeño ESG. Esta divergencia se produce no solo a nivel agregado, sino que en realidad es aún más pronunciada en subcategorías específicas de rendimiento ESG. Esto crea incertidumbre sobre cómo formular objetivos ESG concretos. Mejorar los puntajes con un proveedor de calificación no necesariamente resultará en mejores puntajes en otro. Especialmente cuando las empresas vinculan la compensación de los ejecutivos o las condiciones de endeudamiento a métricas ESG específicas, existe un riesgo significativo de que las mejoras en estas métricas no se reflejen en las calificaciones ESG que utilizan otras métricas. Por lo tanto, las empresas deben asegurarse de que las métricas utilizadas para sus propios fines respalden sus objetivos subyacentes y que el logro de esos objetivos también sea reconocido por los evaluadores. Para lograr eso, las empresas deben trabajar con las agencias de calificación para establecer métricas apropiadas y garantizar que los datos que ellas mismas divulguen sean de acceso público.

En cuanto a las agencias de calificación, nuestros resultados exigen una mayor transparencia. En primer lugar, las agencias de calificación ASG deben comunicar claramente su definición de rendimiento ASG en términos de alcance de atributos y reglas de agregación. En segundo lugar, las agencias de calificación deberían ser mucho más transparentes con respecto a sus prácticas y metodologías de medición. Una mayor transparencia de los métodos permitiría a los inversores y otras partes interesadas, como empresas calificadas, ONG y académicos, evaluar y verificar las mediciones de las agencias. Además, las agencias de calificación deben tratar de comprender qué impulsa el efecto evaluador para evitar posibles sesgos.

Por último, los reguladores podrían abordar la cuestión de la divergencia de calificación ESG. En primer lugar, la armonización de la divulgación DE ESG por parte de las empresas proporcionaría una base de datos confiables y de libre acceso para todas las calificaciones ESG. En segundo lugar, los reguladores podrían ayudar a hacer que la divergencia de la calificación ESG sea más inteligible y fomentar la competencia en la calidad de la medición. Como ha demostrado nuestra taxonomía, hacer coincidir los indicadores con categorías consistentes es un ejercicio difícil. Sin embargo, alguna forma de categorización es esencial para comprender por qué y dónde las metodologías de calificación ESG difieren entre sí. Exigir a las agencias de calificación ESG que asignen sus datos a una taxonomía común haría que tal comparación fuera mucho más simple. Hacerlo también puede estimular la competencia porque los inversores podrían complementar o reemplazar más fácilmente la medición de una categoría específica con datos de un proveedor alternativo. Tal enfoque dejaría a los evaluadores la libertad de mantener metodologías patentadas e innovadoras al tiempo que mejoraría la comparabilidad de las calificaciones ESG.

Referencias

Albuquerque, R., Koskinen, Y., y Zhang, C. (2018) Corporate social responsibility and firm risk: Theory and empirical evidence, Management Science 65, 4451–4469.

Avramov, D., Cheng, S., Lioui, A. y Tarelli, A. (2021) Inversión sostenible con incertidumbre de calificación ESG, Journal of Financial Economics. Próximo.

Bongaerts, D., Cremers, K. J. M. y Goetzmann, W. N. (2012) Desempate: certificación y calificaciones crediticias múltiples, Journal of Finance 67, 113–152.

Chatterji, A. K., Durand, R., Levine, D. I. y Touboul, S. (2016) ¿Convergen las calificaciones de las empresas? Implicaciones para gerentes, inversores e investigadores de estrategia, Strategic Management Journal 37, 1597-1614.

Chatterji, A. K., Levine, D. I. y Toffel, M. W. (2009) How well do social ratings really measure corporate social responsibility?, Journal of Economics & Management Strategy 18, 125–169.

Christensen, D. M., Serafeim, G. y Sikochi, A. (2021) ¿Por qué la virtud corporativa está en el ojo del espectador? El caso de las calificaciones ESG, The Accounting Review. Próximo.

Cornaggia, J. N., Cornaggia, K. J. y Hund, J. E. (2017) Calificaciones crediticias entre clases de activos: una perspectiva a largo plazo, Review of Finance 21, 465–509.

Flammer, C. (2015) ¿La responsabilidad social corporativa conduce a un rendimiento financiero superior? Un enfoque de discontinuidad de regresión, Management Science 61, 2549–2568.

Gibson Brandon, R., Krueger, P. y Schmidt, P. S. (2021) ESG rating disagreement and stock returns, Financial Analysts Journal 77, 104–127.

Hartzmark, S. M. y Sussman, A. B. (2019) ¿Valoran los inversores la sostenibilidad? Un experimento natural que examina la clasificación y los flujos de fondos, Journal of Finance 74, 2789–2837.

Heinkel, R., Kraus, A. y Zechner, J. (2001) The effect of green investment on corporate behavior, Journal of Financial and Quantitative Analysis 36, 431–449.

Jewell, J. y Livingston, M. (1998) Split ratings, bond yields, and underwriter spreads, Journal of Financial Research 21, 185–204.

Krippendorff, K. H. (2004) Content Analysis: An Introduction to Its Methodology Sage Publications Thousand Oaks, USA segunda edición.

Liang, H. y Renneboog, L. (2017) Sobre los fundamentos de la responsabilidad social corporativa, Journal of Finance 72, 853–910.

Lins, K. V., Servaes, H. y Tamayo, A. M. (2017) Capital social, confianza y desempeño de la empresa: el valor de la responsabilidad social corporativa durante la crisis financiera, Journal of Finance 72, 1785–1824.

PRI (2020) Informe Anual del PRI.

URL: <https://www.unpri.org/annual-report-2020/>

Servaes, H. y Tamayo, A. (2013) The impact of corporate social responsibility on firm value: The role of customer awareness, Management Science 59, 1045–1061.

Shrout, P. E. y Fleiss, J. L. (1979) Intraclass correlations: Uses in assessing rater reliability, Psychological Bulletin 86, 420–428.