

Response to FCA Discussion Paper DP21/2 on Diversity and Inclusion in the Financial Sector
Alex Edmans, Professor of Finance, London Business School; Mercers School Memorial
Professor of Business, Gresham College
www.alexedmans.com, aedmans@london.edu

30 September, 2021

Thank you for publishing this discussion paper and for inviting responses. I am a Professor of Finance at London Business School and Academic Director of the Centre for Corporate Governance. Over the past 15 years, I have published in all the top academic journals on corporate governance and CSR. This includes the seminal paper showing that being a Great Place to Work (which, in turn, depends on credibility, fairness, respect, and pride/camaraderie – factors closely linked to diversity and inclusion) leads to long-term stock return outperformance. I am a minority in both ethnicity and age, and am a strong supporter of diversity in general. I took over as Managing Editor of the *Review of Finance*, the #1 academic finance journal in Europe, in 2017 and, at the first available opportunity, appointed the first female Editors onto the board in our 21-year history.

However, it is crucial to base any policy on the highest-quality evidence, rather than our personal views on the matter. The phrase “research shows” is used in the DP, but the fact that “research shows” something is meaningless. It is almost always possible to find “research” that supports what one would like to show. There is substantial variation in the quality of research, with many studies making basic methodological errors and misrepresenting their results. This goes far beyond the simple principle that “correlation is not causation” – many papers don’t even document correlation to begin with. The following thoughts are based on research published in the very top peer-reviewed journals. Note that peer review is not simply a rubber stamp; the top journals reject approximately 95% of papers submitted to them.

Below are my responses to the DP, focusing on the four questions that are closest to my expertise.

Q1: What are your views on the terms we have used, how we have defined them, and whether they are sufficiently broad and useful, now and in the future?

I fully agree with 1.13, which refers to diversity as “diversity of thought”. This is a very useful definition because it is multifaceted and nuanced. Diversity on these issues may well improve firm performance and financial stability.

However, this broad definition of diversity is forgotten in many parts of the DP. For example, Section 4 on reporting focuses almost exclusively on demographic characteristics, even though 1.13 refers to “different perspectives, abilities, knowledge, attitudes, information styles, and demographic characteristics” – i.e. many other aspects of diversity. Similarly, the research quoted only studies demographic characteristics. Indeed, in practice, diversity ends up being narrowly focused on demographic diversity (through explicit quotas, reporting, or investor/stakeholder action), which is why we must be very careful with approaches to increase diversity. No matter how broad one’s definition of diversity is at the outset, the definition often becomes much narrower because demographic characteristics are easiest to measure.

Q1 falls within Chapter 1, which is entitled “Overview”. There are two important big-picture issues relevant to the DP which I did not see any questions about, which are “Does diversity improve firm performance?” and “If it does, what is the regulator’s role?” These questions are fundamental to the issues in the DP, yet are never asked – it seems to be taken for granted that diversity is a good thing and the regulator should try to increase it. Since one aspect of “diversity of thought” is to question taken-for-granted assumptions, I am taking the liberty of addressing these questions. Since there is no specific question on these issues, I am doing so within Q1 since this falls under “Overview”.

Does Diversity Improve Firm Performance?

Throughout the DP, there is the assumption that diversity improves firm performance – indeed, this is the motivation behind regulatory action. The second paragraph of the Foreword makes this claim, and the claim is made even more explicitly elsewhere (e.g. Section 1.31 says “There is *substantial evidence* that makes clear the benefits that different backgrounds and perspectives can bring to an organisation. For example, there is a *strong consensus* within academic research that a more balanced Board improves performance on corporate governance metrics” (emphasis added).

I am familiar with all of the top-tier research on diversity and was surprised to see these claims, which are strongly contradictory to my understanding of the literature. I also had not heard of any of the papers cited in the DP. I thus went to the separate literature review¹, which says “The overarching finding is that the evidence on the benefits of D&I in the workplace is mixed”. This is a much more accurate representation of the literature, but in quite strong contrast to the actual DP. The DP suggests the evidence is much stronger than it actually is and selectively quotes weak papers purporting to have demonstrated benefits of diversity, ignoring rigorous papers showing the opposite.

¹ “Review of research literature that provides evidence of the impact of diversity and inclusion in the workplace” (FCA, July 2021).

Moreover, while the literature review is more accurate than the DP, it still remains misleading since it simply counts how many papers find a positive link vs. a negative link or no link, without taking any account of quality. The authors state that they have “not added our interpretation or opinions on the research we have identified”, which means they have taken every paper at face value. This approach is flawed, given the huge variety in the quality of research available.

As an example of how basic and fundamental the flaws are in frequently-cited diversity research, the cited McKinsey Report on “Diversity Wins” claims that “the business case for gender and ethnic diversity in top teams is stronger than ever”. However, this study has been shown to be irreproducible even with their chosen performance measure (EBIT) and preferred methodology.² Moreover, there is no link between diversity and other performance measures – gross margin, return on assets, return on equity, sales growth, or total shareholder return – or when using more established methodologies (e.g. considering all the data, rather only the top and bottom quarter of diversity). Note that this study is on ethnic diversity, so I have a strong personal interest in its results being true, but they are not.

A second example is the FRC-commissioned report on “Board Diversity and Effectiveness in FTSE 350 Companies”, which was published after the DP and thus not cited. It claims that **“gender-diverse boards are more effective than those without women”** (emphasis in original). However, out of the 90 regressions that it runs which attempt to link diversity to the EBITDA margin (Tables C7-C9), not a single one (0 out of 90) is significant. Even ignoring the multiple flaws in its methodology, the actual results are in stark contrast to the claims. This highlights the danger in taking research at face value. Given that this report was commissioned by a fellow regulator, it may have significant impact on any action the FCA takes. Thus, Appendix A reviews this report and explains why it is fundamentally unreliable and thus no weight should be placed on it.

The best way to find academic consensus on an issue is to survey the literature, taking into account the quality of each paper rather than simply counting the number of papers that find a result. This is why it is particularly important for leading academics to conduct such surveys. Katherine Klein, the Edward H. Bowman Professor of Management at Wharton and the Vice Dean of the Wharton Social Impact Initiative, summarises the academic consensus in a non-technical article at (<http://knowledge.wharton.upenn.edu/article/will-gender-diversity-boards-really-boost-company-performance/>). She writes:

Research conducted by consulting firms and financial institutions is not as rigorous as peer-reviewed academic research. Here, I dig into the findings of rigorous, peer-reviewed studies of the relationship between board gender diversity and company performance. Spoiler alert: Rigorous, peer-reviewed studies suggest that companies do not perform better when they have women on the board. Nor do they perform worse. Depending on which meta-analysis you read, board gender diversity either has a very weak relationship with board performance or no relationship at all.

² “Diversity matters/delivers/wins revisited in S&P 500 firms” (Jeremiah Green and John Hand, 2021). Note that this study is not published in an academic journal, because the McKinsey study it critiques was not published in an academic journal to believe in. Since the original study is not considered to be scientific research to begin with, academic journals would not be interested in publishing a critique.

A recent survey by Jesse Fried, the Dane Professor of Law at Harvard Law School, entitled “Will Nasdaq’s Diversity Rules Harm Investors?” concludes:

While Nasdaq claims these rules will benefit investors, the empirical evidence provides little support for the claim that gender or ethnic diversity in the boardroom increases shareholder value. In fact, rigorous scholarship—much of it by leading female economists—suggests that increasing board diversity can actually lead to lower share prices. Adoption of Nasdaq’s proposed rules would thus generate substantial risks for investors.

If the evidence is so weak, why might it be that claims of a “business case for diversity” are so widespread? As Alice Eagly, the James Padilla Professor of Arts and Sciences at Northwestern University explained in her Presidential Address to the Society for the Psychological Study of Social Issues, entitled “When Passionate Advocates Meet Research on Diversity, Does the Honest Broker Stand a Chance?”:

From advocacy and policy perspectives, there is an obvious appeal in simple, straightforward claims that diversity in groups and organizations produces performance gains. Given this appeal, simplistic renditions of scientific findings on diversity continue to find favor among diversity’s advocates and the legions of practitioners and consultants engaged in helping organizations meet their diversity goals. Presented as if they were evidence-based findings, broad claims about the advantages of diversity for group and organizational performance appear regularly in promotional materials of consultants and advocates.

Thus, when focusing on top-tier research, the evidence for the business case of diversity is even weaker than in the literature review. Moreover, the DP selectively quotes from the literature review and thus misrepresents it. For example, it says “*Wilson and Altanlar, in a wide-ranging study of mainly SMEs, found compelling evidence that more gender diverse Boards reduce insolvency risk*”, and indeed Wilson and Altanlar is cited twice in the DP. However, this is an unpublished working paper from 2009. While a very recent working paper may be reliable but just not yet have gone through the peer review process, a paper that was released 12 years ago and is still not published has likely failed peer review multiple times. While the DP cites unpublished papers and papers published in minor journals, it makes no reference to the Adams and Ferreira (2009) and Ahern and Dittmar (2013) papers cited in the literature review, even though both were published in top-tier journals and have been collectively cited over 6,500 times. They are seminal papers in this field.

Note that the absence of a business case for diversity does not mean that we should take no action on diversity. Instead, it highlights the danger in diversity initiatives based on basic demographic characteristics (which regulatory approaches will inevitably focus on, since they are easiest to measure). As I wrote in a Telegraph op-ed³, included in Appendix B of this submission:

Importantly, the invalidity of popular diversity research does not invalidate diversity initiatives. Perhaps there is a business case for diversity, but existing studies haven’t found one due to blunt classifications based on only gender and ethnicity. Diversity comprises a myriad of different dimensions, such as socioeconomic, educational, regional, or experiential background, as well as the practices a company puts into place to foster inclusivity.

³ “No, boardroom diversity does not mean higher profits” (Alex Edmans, 20 August 2021).

And even if there is no clear business case for diversity, there are strong moral and ethical cases. Many people – myself included – believe that companies have a responsibility to contribute to a diverse and inclusive society. Perhaps doing so may not maximise profits but many shareholders and stakeholders are willing to accept that trade-off – just as consumers buy organic food, despite its greater cost, due to non-financial considerations.

Moreover, study-based arguments for diversity are problematic because they relegate dimensions of diversity for which no study exists. There is no rigorous evidence on the financial benefits of inclusion based on disability or socioeconomic background, but it is desirable for other reasons.

Should Regulators Regulate Diversity?

Let us assume that the evidence for the business case for diversity is overwhelming. Even if so, it is still not clear whether regulators should be involved.

Certainly, regulators should promote financial stability. However, there are many factors that promote financial stability, such as risk management practices, investment decisions, and capital allocation. Even focusing on human capital, financial stability depends on attributes such as competence. Regulators do not police whether those appointed to risk management roles within banks have “hard” risk management expertise, and “soft” skills such as the willingness to challenge, even though these are critical competencies. Thus, it is not clear why the regulator should isolate diversity, above all of the other factors that determine financial stability.

Indeed, companies have strong incentives to promote financial stability themselves. As long as executives are appropriately incentivised (e.g. with long-term equity that they must continue to hold after their departure, and – for financial institutions – potentially also with debt⁴), they have strong incentives to take all actions to ensure financial stability through all means – effective risk management, investment, and capital allocation, as well as hiring the best people which includes both competence and diversity.

I do not believe in “leaving everything to the market” and believe that markets fail; I’ve written extensively on the need for regulatory intervention to correct market failure.⁵ However, before deciding to intervene, we need to identify what the market failure is, *and* whether regulators are best placed to deal with market failure. I am very open to believing that financial institutions don’t make perfect, or even close to perfect, hiring decisions; nor do they create perfect cultures. However, it is not clear why regulators are better placed to make these decisions, given the difficulties in measuring the multifaceted dimensions of diversity highlighted in 1.13. Regulators are not involved in most aspects of hiring (besides issues such as the “fit and proper test”, which only focuses on senior management, and is much less blunt than demographic diversity requirements which certain people have no way of meeting).

⁴ “Inside Debt” (Alex Edmans and Qi Liu, *Review of Finance*, 2011).

⁵ For example, Chapter 10 in “Grow the Pie: How Great Companies Deliver Both Purpose and Profit.”

Q3: Do you agree that collecting and monitoring of diversity and inclusion of data will help drive improvements in diversity and inclusion in the sector? What particular benefits or drawbacks do you see?

As a Professor of Finance, who has conducted numerous studies involving data, I understand the power of data. However, data also has limitations. In particular, it focuses only on quantitative aspects and ignores the qualitative. Not only does it provide an incomplete picture *ex post*, but *ex ante* creates incentives to focus on only the aspects that can be reported. For example, not only do quarterly earnings present an incomplete view of the true quality of a company *ex post*, but the need to report quarterly earnings encourages companies to take short-term actions to boost earnings *ex ante*. These concerns exist with non-financial as well as financial numbers. Test scores not only provide an incomplete picture of a true quality of a school *ex post*, but encourage teaching-to-the-test *ex ante*.

Of course, all data is incomplete, yet we still collect and monitor many types of data. Thus, the desirability of collection and monitoring depends on how incomplete the data is. For diversity, paragraph 1.13 highlights how multifaceted diversity is, involving “different perspectives, abilities, knowledge, attitudes, information styles, and demographic characteristics”. It is difficult to report objective, aggregate data on anything other than demographic characteristics, which can lead to substantial incompleteness. Inclusion is even harder to report. Mandatory reporting runs the huge risk of box-ticking approaches to diversity, e.g. diversity being a more important driver of recruitment, retention, and promotion than ability/competence (since the latter cannot be reported), or companies only focusing in the measures of diversity that can be reported (e.g. I might be chosen over a white male who did not go to university).

My caution on data-driven approaches to diversity does not mean that we should do nothing. Instead, narrative reporting on the policies and processes that companies to ensure diversity and inclusion across all dimensions – both demographic and non-demographic – would be more informative. Companies should still feel free to report statistics if they believe they are relevant and informative (see my answer to Q13 below). However, this should not be mandatory, and investors and stakeholders should be careful about how to interpret these statistics rather than thinking that more is always better.

It is important to recognise that there is a huge amount of discrimination on many non-demographic dimensions. For example, top-tier research shows that beauty has a strong effect on hiring, labour market success, and obtaining access to finance; indeed, evolution makes us naturally more drawn towards good-looking people. Similarly, people immediately make inferences when seeing a tattooed person. This is not meant to be a facetious comment, but to highlight many aspects of discrimination that are hidden and would not be captured in any data. I am an angel investor in a start-up founded by a tattooed CEO but many investors would discriminate against such an entrepreneur (unless he operated in a pigeon-holed sector).

Paragraph 4.6 argues that mandatory reporting of diversity data will help academic research: “academics ... are all increasingly interested in firm diversity data.” However, there has already been copious academic research on diversity published in the top academic journals. There is no need for even more research on diversity; what we need is the findings of existing research to be taken seriously rather than ignored.

Q12: What are your views on linking remuneration to diversity and inclusion metrics as part of non-financial performance assessment? Do you think this could be an effective way of driving progress?

I am very strongly against such an approach. Q3 highlights the incompleteness of diversity data as a true measure of diversity and inclusion; moreover, D&I is only one aspect of human capital, and human capital is only one determinant of financial stability and performance. There is no justification for paying according to this one, highly incomplete, measure. Doing so will simply lead to companies “hitting the target but missing the point”, as has been documented through decades of research on the problems of target-based incentives (which applies to non-financial as well as financial targets). For further details, please see my Wall Street Journal article “Why Companies Shouldn’t Tie Pay to ESG Metrics” (included as Appendix C of this submission).

Q13: What are your views about whether all firms should have and publish a diversity and inclusion policy?

I strongly support this. Such a policy would be much broader than just isolated quantitative metrics and capture the full range of activities that a company is doing to promote diversity and inclusion. These include actively encouraging dissenting viewpoints and creating a psychologically safe place to work.

If the regulator is concerned with financial stability rather than social policy, diversity should involve openness to a variety of viewpoints not only inside the organisation, but externally as well. The company should also discuss how it is taking action to seek the viewpoints of major shareholders and other stakeholders. The FCA has proposed allowing dual class share structures, but they are a way of limiting diversity of thought by restricting shareholder voice.

Appendix A: Evaluation of the Report “Board Diversity and Effectiveness in FTSE 350 Companies”

The Financial Reporting Council commissioned the LBS Leadership Institute and the consultancy SQW to write a report on “Board Diversity and Effectiveness in FTSE 350 Companies”. The FRC should be commended for commissioning research on such an important issue. Given this report was commissioned by a fellow regulator, and its findings have become widely known and cited, it is likely to have significant impact on the FCA’s decisions. As a result, I thought it would be helpful to evaluate the report and the extent to which we can rely on it.

The report claims that to have found a strong link between diversity and firm performance. Moreover, the authors suggest that this link is causal – in addition to using the phrase “effect(s)/impact(s) of (gender) diversity” 38 times⁶, they contrast their report with previous research that could not document causality or used less rigorous methodology. Thus, according to the authors, if a company increases board diversity, its financial performance will increase on average. The following are examples of such claims:

- “There has been a good deal of research about the business case for diversity. Often a correlation is found, but not necessarily full causality ... I am pleased to see the analysis from this research builds the case for diversity across the board.” (Foreword by CEO of FRC).
- “The academic rigour with which data was collected and analysed yields new insights on the impact of diversity and how to make diversity work. We all stand to learn from the authors’ methodology and findings.” (Foreword by Dean of LBS).
- “The Leadership Institute at London Business School treated this opportunity accordingly, with the rigour and care of our best scholarly research.” (Foreword by LBS Leadership Institute).
- “The research design and analysis used in this report is both innovative and rigorous ... The study finds positive relationships between having more women on FTSE 350 boards and future financial performance.” (Foreword by SQW).
- “Higher levels of gender diversity of FTSE 350 boards positively correlate with better future financial performance (as measured by EBITDA margin). ... These results are significant because, for decades, researchers have largely failed to confirm any causal link” (Executive Summary).
- **“These results suggest that gender-diverse boards are more effective than those without women”** (main paper, emphasis in original).

With such strong claims, made so prominently, it is not surprising that many articles about the report believe it finds that board diversity causes better performance. Examples follow below:

- “Diverse boards lead to better corporate culture and performance” (title of [FRC’s](#) press release announcing its report)
- “Boardroom diversity improves financial performance” (title of [Minerva Analytics](#) article)
- “The effort to diversity boards pays benefits in terms of boardroom culture and performance” ([Linklaters](#))

⁶ There are many other causal claims using different wording, e.g. “having at least one woman on the board has a statistically significant positive effect on one-year stock returns.”

- “The case for diverse boards was given further clout after new research published by the FRC supported the thesis that it leads to better corporate culture and performance” (opening sentence of [ICAEW](#) article)

Unfortunately, the actual analysis in the report does not come close to justifying such claims. There are two serious concerns. First, even taking the methodology at face value, the results do not support the claims. Second, the methodology itself is flawed.

The Results

The Executive Summary claims that “Higher levels of gender diversity of FTSE 350 boards positively correlate with better future financial performance (as measured by EBITDA margin)”. The analysis linking gender diversity to EBITDA margin is in Tables C.7, C.8 and C.9 (p85-86). These tables run 90 regressions linking 6 different measures of diversity to EBITDA margin over the next 5 years in 3 different indices ($6 \times 5 \times 3 = 90$). Not a single one (0 out of 90) is statistically significant at the 5% level. The claim is strongly contradicted by the authors’ own analysis.

The main body claims that “**These results suggest that gender-diverse boards are more effective than those without women**” (p34). This claim follows the discussion of the analysis linking gender diversity to EBITDA margin and stock returns. The prior paragraph has discussed the EBITDA margin results. The stock return results are in Tables C.12-C.15 (p92-95). Out of the 90 regressions, they find that diversity is significantly positively correlated with stock returns in 7, and significantly *negatively* related in 2.

Combining “these results” together (for both EBITDA margin and stock returns), out of 180 regressions, only 7 show positive results and 2 show negative results. If you run 180 regressions, 5% (9) will be significant in either direction even if there is no relationship, which is exactly the number of significant results found by the authors. Thus, the claim in bold is strongly contradicted by the authors’ own analysis.

The Methodology

The questions studied by the authors can be addressed with simple, standard methodologies, but the authors either use unnecessarily complicated methodologies, or use standard methodologies but make them seem much more sophisticated than they actually are.⁷ An unusual amount of space is devoted to over-explaining what should be standard regressions. This is an example of the “Delusion of Rigorous Research” – a seemingly sophisticated methodology giving the impression that the research is rigorous, even if there are basic flaws.⁸

One basic flaw is the stock return results. Stock returns are comprised of capital gains and dividends, but dividends are excluded from the analysis. I am not aware of any paper published

⁷ For example, the authors argue “Following an approach to addressing reverse causality when using panel data discussed in Leszczensky and Wolbring (2019), we formulated the following model in which EBITDA margin is explained by lagged values of gender diversity of the board and contemporaneous values of control variables”. This makes it seem that they have used a sophisticated methodology to rule out reverse causality and thus show causation rather than just correlation. However, it is very well known that using lagged (i.e. past) values does not address the issue – this is the *post hoc ergo propter hoc* fallacy. That A precedes B does not mean that A caused B. Opening an umbrella does not cause it to rain.

⁸ Rosenzweig, Phil (2014): “The Halo Effect ... and the Eight Other Business Delusions That Deceive Managers.”

in a top academic journal which does this. The authors give three reasons on p90, but none are satisfactory:

- *Dividends are paid out when companies make profit and we analysed EBITDA margin, a measure of profitability.* This argument does not make sense, since increases in profitability lead not only to increases in dividends but also increases in stock prices. The same argument could be used to justify excluding capital gains.
- *According to the generalised dividend model, stock prices contain information about expected future dividends.* This is true, but the more important consideration is stock prices are *reduced* by the payment of a dividend. Dividends are a positive signal of future prospects. If firms A and B make the same profits today, and A expects more profits in the future, it will pay a higher dividend, causing its stock price to fall. Under the authors' methodology of excluding dividends, A performs worse even though it's actually performing better.
- *Sample size considerations – our data set contained less data about dividend payments than about other financial characteristics, thus using this variable would reduce the sample size.* This argument is unsatisfactory. Any standard data set contains dividends, particularly for the FTSE 350 studied by the authors. The authors should have used an appropriate data set.

In addition to EBITDA margin and shareholder returns, the third measure of firm performance is shareholder dissent. This is an invalid measure of performance. The argument throughout the paper is that current system is non-diverse, and thus dominated by “old white men” (I am only using this term as shorthand) with a narrow view of the world. If so, and if diverse boards leads to a company being more innovative, then it should be encountering *more* shareholder dissent, because shareholders are old white men who cannot see the value of the innovation. For example, when The Weir Group innovatively decided to restructure its executive pay structure away from complex bonuses towards simple, long-term equity, it was initially voted down by shareholders in 2016, before being approved in 2018. If avoiding dissent is the goal, this is best achieved by filling the board with old white men who will preserve the status quo.

Overall, one measure of performance (EBITDA margin) is correlated with gender diversity in 0 out of 90 regressions, and the two others (shareholder returns and shareholder dissent) are flawed. Unfortunately, a sophisticated methodology cannot help when the underlying inputs are flawed.

A separate issue is the use of a 10% threshold for statistical significance. What this means is the following. Even if there is no link between diversity and firm performance, you might still find a link in the data due to chance (just like, even if a coin is fair, it still might land 6 or more heads out of 10). 10% statistical significance means that, if there is less than a 10% probability that the link is due to luck, you accept it as significant.

The standard threshold for significance is 5%. The authors argue that 10% is sufficient “given the complexity of the relationship between board diversity and effectiveness”. There is no reason why a more complex relationship should lead to a laxer threshold. Indeed, if the relationship is more complex, the researchers have greater latitude to “data-mine” – to choose the specification that gives the results they want. This should lead to a stricter threshold for statistical significance. For

example, the seminal article “Why Most Published Research Findings Are False”⁹ states, in Corollary 4: “The greater the flexibility in designs, definitions, outcomes, and analytical modes in a scientific field, the less likely the research findings are to be true.” (Incidentally, Corollary 5 highlights the danger of confirmation bias: “The greater the financial and other interests and prejudices in a scientific field, the less likely the research findings are to be true.”)

(Note that, *even if* you accept a 10% significance level, only 2 of the 90 EBITDA and 9 of the stock return regressions become significant. The number of stock return regressions that are *negative* and significant doubles from 2 to 4).

To the authors’ credit, they do include caveats in the paper that not all their findings are statistically significant, and that some results are correlations rather than causation. However, buying gifts for your spouse on some days does not make up for mistreatment on other days. The claims at the start of this Appendix are made in prominent places, such as the Foreword and Executive Summary, or in bold. Not every reader will have time to read the entire report and see the caveats, which is why articles by serious institutions took the claim at face value.

Please note that the goal of this Appendix is to be entirely constructive. Diversity is such an important topic, and regulation has such potential to be either very effective or very ineffective, that it is important to base it on the highest-quality evidence. My response to the FRC’s consultation on the Corporate Governance Code focused on the McKinsey study, since that was the most prominent study at the time.¹⁰ It is to my own chagrin that the report was co-authored by an institute at my own employer, but the scientific process is about being equally discerning about all research, regardless of who wrote it. Just as different views by Eugene Fama and Richard Thaler at the University of Chicago have significantly illuminated the debate on whether markets are efficient or irrational, I hope that different views from London Business School may help the FCA on this important topic.

⁹ Ioannidis, John P.A. (2005): “Why Most Published Research Findings Are False.” *PLOS Medicine*

¹⁰ <https://alexedmans.com/wp-content/uploads/2019/02/Diversity.pdf>

No, boardroom diversity does not mean higher profits

Beware of the temptation to accept research just because it supports a view we would like to be true

Alex Edmans, 20 August 2021

It has become a truth universally acknowledged that diverse companies perform better. A much heralded McKinsey report, entitled *Diversity Wins*, argued that “the business case for gender and ethnic diversity in top teams is stronger than ever”.

Last month, the Financial Reporting Council (FRC) released a study, which concluded that “gender-diverse boards are more effective than those without women”. The evidence is supposedly so compelling that one chairman claimed in the study that “there have been enough reports ... statistics and ... evidence-based research to stop talking about it and get on with it”.

Another declared that “I don’t want to see any men. I don’t care if they’re Jesus Christ. I don’t want to see them.”

Accordingly, regulators are taking action. The US’s Nasdaq has proposed a requirement for boards of listed companies to contain at least two minority directors. In the UK, the Financial Conduct Authority is consulting on whether to introduce “comply-or-explain” targets for diversity.

Both proposals are based on the belief that diversity improves financial performance. As an ethnic minority and strong supporter of responsible business, I would really like this to be true.

But the business case for diversity is far weaker than commonly claimed. The McKinsey study has been shown to be irreproducible even with their chosen performance measure (EBIT) and preferred methodology. Moreover, there is no link between diversity and other performance measures – gross margin, return on assets, return on equity, sales growth, or total shareholder return – or when using more established methodologies.

The study commissioned by the FRC runs 90 regressions investigating the link between diversity and profitability. Eighty-eight find no relationship, and two find a weak relationship that fails the standard threshold for statistical significance. Yet many articles have been written based on the study’s headline claim, without looking at its actual results.

The foreword to the study (not written by the FRC) proclaims that “the academic rigour with which data was collected and analysed yields new insights ... we all stand to learn from the authors’ methodology”.

But making claims unsupported by the evidence is not rigorous, nor a methodology that should be learned from. Beyond these two studies, rigorous academic research, published in the most stringent peer-reviewed journals, either finds no or a negative link between diversity and performance.

Why are we so quick to believe such weak evidence? Because of confirmation bias – the temptation to accept any study that supports what we’d like to be true. We want to live in a world in which diverse companies outperform. But our acceptance of a study should depend on its scientific rigour, not whether we like its findings. This point applies beyond diversity to any research – on business, politics, or health advice.

Importantly, the invalidity of popular diversity research does not invalidate diversity initiatives. Perhaps there is a business case for diversity, but existing studies haven’t found one due to blunt classifications based on only gender and ethnicity. Diversity comprises a myriad of different dimensions, such as socioeconomic, educational, regional, or experiential background, as well as the practices a company puts into place to foster inclusivity.

And even if there is no clear business case for diversity, there are strong moral and ethical cases. Many people – myself included – believe that companies have a responsibility to contribute to a diverse and inclusive society. Perhaps doing so may not maximise profits but many shareholders and stakeholders are willing to accept that trade-off – just as consumers buy organic food, despite its greater cost, due to non-financial considerations.

When I took over as managing editor of the *Review of Finance*, the leading academic finance journal in Europe, I appointed the first female editors in our 21-year history. I didn’t do so because of evidence that diversity instrumentally improves journal performance, but because diversity is intrinsically desirable and because these candidates were excellent in their own right.

Moreover, study-based arguments for diversity are problematic because they relegate dimensions of diversity for which no study exists. There is no rigorous evidence on the financial benefits of inclusion based on disability or socioeconomic background, but it is desirable for other reasons. This goes beyond diversity to any initiative to serve society.

During the pandemic, many companies paid furloughed workers, gave away free products, or prioritised their most vulnerable customers. They did not do so because studies showed that it would improve their public image and generate more profit in the long term.

Researchers’ mission is to analyse the data in the most careful way possible and let it speak. Sometimes, it will uncover win-wins where social and financial goals coincide. But if there are trade-offs, researchers should be upfront about them. Even if an action sacrifices profits, some companies and investors will be willing to make this sacrifice. But it’s their choice, and the role of evidence is to inform this choice rather than torturing the data to get the result we want.

Why Companies Shouldn't Tie Pay To ESG Metrics

It may sound like a good idea. But it likely won't achieve the results proponents want

Alex Edmans, 27 June, 2021

Executive pay has been the poster child for everything that's wrong with capitalism. Chief executive officers are paid millions, while some of their employees make minimum wage. And to trigger their bonus targets, executives allegedly take short-term actions that mortgage their company's future.

But change is afoot. A rapidly growing trend is for executive pay to be tied not only to financial numbers, but also environmental, social and governance (ESG) targets. Two recent studies found that [51% of large](#) U.S. companies and [45% of leading](#) U.K. firms use ESG metrics in their incentive plans.

The rationale seems obvious. First, many advocates have claimed that good ESG practices boost a company's bottom line, so incentivizing ESG performance also will improve financial performance. This may be why even [private-equity investors](#) have started to mandate ESG targets.

Second, it is commonly believed that rewarding performance is the best way to ensure performance. Conversely, if a company won't pay for an outcome, that is a telltale sign that it doesn't actually care about it. Companies are making grand promises about diversity, decarbonization and resource usage—but these promises are hollow if they don't affect CEO pay.

As an ESG advocate, I should applaud this trend. But as essayist H.L. Mencken is often paraphrased, "Every complex problem has a solution which is simple, direct, plausible—and wrong." And this may be the case with ESG targets.

Let's start with the second argument—that rewarding performance ensures performance. The evidence shows that paying for targets indeed encourages employees to hit those targets. But it doesn't encourage them to improve performance. The crux of the problem is that you can't measure many of the performance dimensions that you care about— "not everything that counts can be counted," as stressed by sociologist William Cameron (and commonly misattributed to Albert Einstein). For example, paying executives to meet earnings benchmarks leads many of them to [cut research and development](#) to do so.

Importantly, the problem of "hitting the target, but missing the point" occurs for any target—whether financial or nonfinancial. Binary thinking often equates "financial" with "short-term" and "nonfinancial" with "long-term." But nonfinancial targets can be short-termist. Paying teachers according to test scores encourages them to teach to the test even at the expense of teaching students to develop critical thinking skills. Rewarding CEOs according to average employee pay may encourage them to outsource or automate low-paid jobs, or focus on salary rather than meaningful work, skills development and working conditions.

These unintended consequences might be even worse for ESG than financial targets. One challenge is that, for financial performance, only a couple of measures might be relevant. But ESG performance is multifaceted. Companies have a responsibility to many stakeholders—employees, customers, suppliers, the environment, communities and taxpayers—and for each stakeholder, many dimensions are relevant, as we've seen for employees. Either the contract includes only a

couple of ESG measures and the CEO ignores others, or it includes most of them and the contract becomes so complex that it loses any motivational effect.

A second problem is measurement. For a financial target such as earnings-per-share, there's consensus on how to measure it. But that isn't the case for an ESG metric. Should ethnic diversity be captured by the number of minorities on the board, in senior management, or in the workforce—or other factors such as the ethnic pay gap, or the proportion of minorities who get promoted from each level? Even ESG-rating agencies [disagree significantly](#) on how to measure ESG performance, so any measure might be perceived as unfair or ignore important dimensions.

Let's turn to the first rationale, which holds that boosting ESG performance will always boost financial performance. The evidence is [far less conclusive](#) than often claimed. In fact, only performance related to material ESG dimensions (those that are relevant to the company's specific business model) ultimately pays off; boosting immaterial factors [doesn't](#). Thus, if ESG targets are to be used, they should be selected carefully based on materiality. Instead, they're often a knee-jerk reaction to the demands of pressure groups or whatever issue happens to be the order of the day.

So what's the solution? The answer is to scrap all bonuses—on both financial and nonfinancial targets—and instead pay CEOs like owners, with long-term shares that they can't sell for five to seven years and must retain beyond their departure. Since material ESG factors ultimately improve the long-term stock price, this holds them accountable for material ESG issues—even if they aren't directly measurable. Indeed, evidence shows that long-term pay plans [improve](#) not only financial performance, but ESG performance as well, and the relationship is causation, not just correlation. Long-term equity is also simple and transparent – there's no need to decide which ESG factors to include and which to leave out, how high to set the targets, and how much extra pay to give for hitting them.

Companies should still set ESG goals and report on whether they are meeting them. A CEO already has strong reputational incentives, and intrinsic motivation, to meet a publicly announced ESG goal—so you don't need pay to ensure a target is hit. But there's a big jump between simply reporting on performance and linking pay to it, as the latter amplifies the risk of manipulation. As Goodhart's Law suggests, when a measure becomes a target, it ceases to be a good measure.

Finally, ESG pay targets might be appropriate in some companies if the above concerns are muted. For example, in an energy company, decarbonization is arguably much more important than any other stakeholder issue, so there is less of a concern about overweighting a single ESG factor. Moreover, there is relatively little disagreement on how to measure direct greenhouse-gas emissions. But ESG targets aren't the ubiquitous panacea often claimed. The best way to ensure that CEOs create long-term value for both shareholders and society is to pay them like long-term owners.