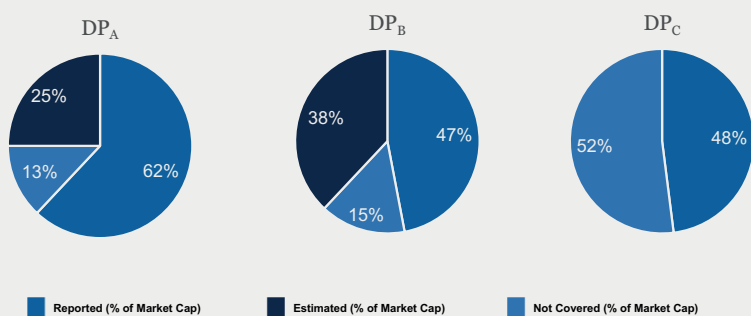# Beware of garbage in, garbage out in carbon emissions data

Investors who seek to mitigate climate change must rely on a combination of reported and estimated data. Our research shows that the estimated data are poor substitutes for reported data.

By *Vitali Kalesnik*

Today, mandatory carbon emissions reporting does not exist on a global scale. Thus, investors who wish to impact climate change must rely on a combination of voluntarily reported data and data estimated by data providers. However, our research in 'Green Data vs. Greenwashing: Do Corporate Carbon Emissions Data Enable Investors to Mitigate Climate Change?' shows that estimated data are poor substitutes for reported data. Estimated emissions data are 2.4 times less effective than reported data in identifying the worst carbon emitters.

The good news is that, on 21 April of this year, the European Commission published the Corporate Sustainability Reporting Directive (CSRD), which extends the scope of the 2014 Nonfinancial Reporting Directive that currently applies in the EU. The CSRD proposes mandatory reporting according to EU sustainability standards and mandatory audits of reported information. The proposed reporting requirements apply to all large EU companies and all companies listed on EU-regulated exchanges. The proposal is a major step toward improving the accuracy of sustainability-related data.

The bad news is that many countries do not have comparable regulations. Until the CSRD is adopted and similar regulations are adopted in other nations, investors remain hindered by imperfect GHG emissions data in successfully fulfilling their climate change mitigation efforts.

The power of accurate data
Investors have two routes to achieving their goal: shift capital from brown to green companies, or invest in brown companies while exerting pressure on corporate management to adopt greener policies.

Investors need accurate data to execute either policy. Without mandatory reporting, such accuracy is compromised. The data providers are not at fault, however. They are doing the best they can with the information available to them. Because greenhouse gas (GHG) emissions are highly concentrated - 71% of GHG emissions since 1988 are linked to just 100 active fossil fuel producers according to the global environmental-reporting nonprofit CDP - when investors have accurate data, their ability to positively impact climate change can be quite meaningful.

How good are the estimates?
We believe GHG emissions-related data should be widely available, comparable between companies and consistent across data providers. Forward-looking information, such as emission reduction targets, should have predictive power and all data should accurately reflect true emissions.

Under the current voluntary reporting framework, we find these five criteria are not being attained.

We analyzed data from three major carbon data providers (anonymously labeled as DP_A, DP_B, and DP_C) for the seven year

---

FIGURE 1: COMPARISON OF MARKET-CAPITALIZATION COVERAGE WITH GHG EMISSIONS, 2010–2016

Note: This figure compares the market-capitalization coverage across the GHG data providers. Reported (% of market cap) represents the fraction of market capitalization that is on average covered with reported GHG data in the respective data set. Estimated (% of market cap) represents the fraction of market capitalization that is on average covered with estimated GHG data in the respective data set. Not covered (% of market cap) represents the fraction of market capitalization that is on average not covered with any GHG data in the respective data set. The numbers reflect the time-series average of cross-sectional means for the period 2010–2016.

Source: Research Affiliates, LLC, and University of Augsburg based on anonymized data from GHG emissions data providers.

period 2010–2016. Their reported carbon emissions data ranged from 47% to 62% of emitting companies based on market-capitalization coverage. They provide roughly the same percentage in carbon emissions coverage by metric gigatons. In both coverage areas, a large gap in reported data must be estimated by the data providers.

Today, no universally accepted reporting standard for GHG emissions exists. The GHG Protocol is the most commonly followed standard, but a number of nations have issued their own reporting guidelines, muddying the waters of data comparability. According to the Task Force on Climate-Related Financial Disclosures (TCFD), in 2018 only one-third of companies reported their GHG emissions following the TCFD's recommendation to use the GHG Protocol. Without a universal reporting standard, data comparability is not achievable. To estimate the missing data of non-reporting companies, data providers turn to proprietary estimation models. The output of these models is inconsistent, complicating the analysis that investors undertake to identify green companies.

Data providers typically position their proprietary estimation models as being quite sophisticated, leading many investors to view the estimated data they produce as similar in quality to reported emissions data. The reality is that the estimates are a noisy proxy of true emissions. They generally rely on broad business metrics and industry affiliations and often very simple financial information, such as net sales or number of employees.

We find that simple correlates, such as industry and size effects, explain the bulk of the variation in the estimated data. Thus, a simple model using a company's net sales, industry affiliation, country of domicile and year as fixed variables is just as accurate as the data estimated by the data providers in our analysis. Failing to capture information beyond simple correlates makes it much more difficult for investors to identify the green companies in brown sectors and can lead to counterproductive investor actions in mitigating climate change.

At the present time, estimated data fail the criterion of reflecting true emissions. We tested the effects of using estimated data to identify the 5% worst emitters in a universe of 10,000 companies. With perfect data (reported or estimated with an R^2 of 100%) an investor would know exactly the 500

worst emitters. With imperfect data, an investor must exclude more companies from their portfolio to safely remove the 500 worst emitters.

In short, when estimating the accuracy of estimated data, we find that by using estimated data the number of stocks excluded from a portfolio is 2.4 times larger (1,190) relative to the 500. Thus, the effectiveness of the investor's desired action to mitigate climate change is reduced by 2.4 times, which we call the *inefficiency of using estimated emissions*.

Perils related to forward-looking data

Data providers also evaluate a set of forward-looking data, which are theoretically just as important as the data on current emissions. Forward-looking information can help investors identify those companies that drive the transition to a green economy and which may perhaps benefit from the decarbonization trend. Conversely, they can inform investors about which companies are not planning to reduce emissions, and are perhaps even planning to increase them. Importantly, these data should have forecasting ability for future changes in emissions.

Our empirical examination of forward-looking emissions estimates found no evidence that these data offer any useful forecasting insights. The lack of predictability is likely driven by the use of non-scientifically verified estimation methods. Also, much of the data used to derive these estimates may just be 'cheap talk' from companies engaged in greenwashing. We argue that forward-looking information should be externally verified before investors rely on it. ◼

Vitali Kalesnik

Partner and Director of Research for Europe, Research Affiliates Global Advisors (Europe) Ltd.

SUMMARY

Investors have a meaningful role to play in helping mitigate climate change, but to do so they need accurate carbon emissions data.

With no current mandatory reporting requirement, nearly half of all data available to investors are estimated.

Estimated data are much less effective than reported data in helping investors identify green companies and thus in successfully meeting their goals of positively impacting climate change.

FIGURE 2: SIMULATED DISTRIBUTION OF TRUE (UNOBSERVABLE) AND ESTIMATED EMISSIONS FOR DIFFERENT LEVELS OF R^2
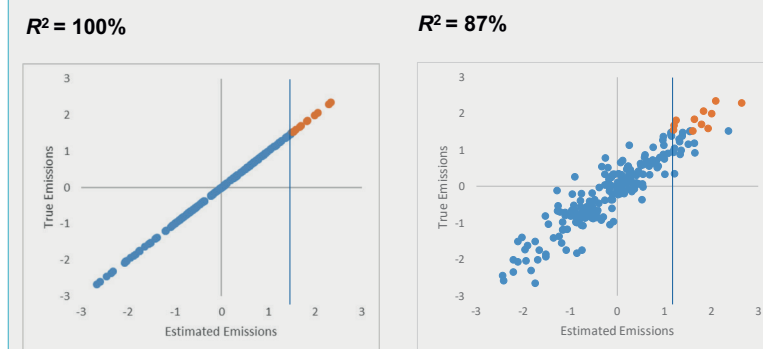
Note: The figure shows the simulated distribution of normally randomly distributed true (unobservable) and estimated (observable) emissions. We mark with crosses the 5% worst emitters based on the true emissions. We also mark with the dashed line the cut-off value of the estimated emissions necessary to select at least 95% of the 5% worst emitters based on the estimated emissions.

Source: Research Affiliates, LLC, and University of Augsburg.