

Working Paper Series

No. 2 of 2022

Costing Analysis with Scarce Data An Example from the Cycle-To-Work Scheme for E-bikes

Akisato Suzuki[†]

December 2022

[†] The author Akisato Suzuki is an Irish Government Economic and Evaluation Service (IGEES) policy analyst / economist in the Parliamentary Budget Office (PBO), Houses of the Oireachtas. PBO Working Papers present primary research in progress intended to elicit comments and encourage debate. The content is subject to review and revision. The analysis and views contained in this paper are those of the author only, and are not necessarily reflective of the position of the PBO or of the Houses of the Oireachtas generally. For queries, contact Akisato.Suzuki@oireachtas.ie.

Costing Analysis with Scarce Data

An Example from the Cycle-To-Work Scheme for E-bikes

Akisato Suzuki
Parliamentary Budget Office
Leinster House, Kildare Street
Dublin 2, D02 XR20, Ireland
akisato.suzuki@oireachtas.ie

5 December 2022

Abstract

This paper showcases a methodological approach to mitigate challenges posed to costing analysis when relevant data are scarce. It uses a statistical simulation technique to "glue" incomplete pieces of information with assumptions used to fill in the missing pieces and the associated uncertainty taken into account. The technique is exemplified by the costing of a tax-expenditure scheme for electric bikes (the Cycle-To-Work scheme). The resulting costing estimate includes significant uncertainty, which is a natural consequence of the data scarcity. As all necessary data for a policy to be costed may not always be available at a high-quality level, a statistical technique such as the one employed in this paper is a useful tool to estimate the cost and understand the uncertainty arising from the data scarcity.

Keywords – data, costing, public policy, statistics, tax expenditure

Disclaimer: Part of the analysis in this paper is based on an actual confidential costing analysis the Parliamentary Budget Office (PBO) conducted (hereafter, "original confidential costing"). However, the policy details and assumptions made are changed for anonymisation and publication purposes. In addition, unlike the original confidential costing, the data used are limited only to those publicly available because of confidentiality issues, and were relevant only when the original confidential analysis was conducted. For these reasons, the analysis and its results should be seen as an example to illustrate the methodology, rather than a specific empirical finding on the particular policy topic used.

Introduction

Policy costing is essential to understand the fiscal implications of a new policy. Analytical challenges arise, when there is no high-quality unified set of data available (such as an administrative dataset including the joint distribution of all relevant information) and, instead, only incomplete pieces of information are available scattered across multiple sources with different degrees of credibility. This paper showcases a methodological approach to mitigate challenges posed to costing analysis by data scarcity. It uses a statistical simulation technique to "glue" incomplete pieces of information with assumptions used to fill in the missing pieces and the associated uncertainty taken into account.

The paper exemplifies the technique by the costing of a tax-expenditure scheme for electric bikes (the Cycle-To-Work scheme). Under the C2W scheme, employers are "not obliged to notify Revenue of the provision of bicycles and/or safety equipment to its employees."¹ Therefore, there is currently no systematic administrative data available on tax forgone, which makes precise costing analysis difficult.² The C2W scheme offers Benefit-In-Kind (BIK) tax exemptions for the bikes and associated safety equipment that employees would obtain and use for commuting.³ More specifically, for electric bikes (e-bikes), the maximum non-taxable value of €1,500 is exempted from the income tax, Employer's and Employee's Pay Related Social Insurance (PRSI), and Universal Social Charge (USC).⁴ This paper estimates the cost of increasing this non-taxable value, hypothetically to €4,000.⁵

The rest of the paper is structured as follows. The first section presents the methodology, divided into two subsections. The first subsection discusses how to address data scarcity by a statistical simulation technique; the second subsection explains how to calculate the cost. The second section presents and discusses the results. The final section is a conclusion.

_

¹ Revenue, "Tax and Duty Manual: Chapter 7 - The Provision of Bikes and Safety Equipment ('Cycle to Work Scheme')," July 2021, https://www.revenue.ie/en/tax-professionals/tdm/income-tax-capital-gains-tax-corporation-tax/part-05/05-01-01g.pdf, p.14.

² Indeed, when a new measure for cargo bikes under the C2W was introduced in the Finance Bill 2022 (and it was not mentioned in Budget 2023), no costing was provided. Parliamentary Budget Office, "Budgetary Issues in the Finance Bill 2022," Publication 27 of 2022, 8 Nov 2022, https://data.oireachtas.ie/ie/oireachtas/parliamentaryBudgetOffice/2022/2022-11-08_budgetary-issues-in-the-finance-bill-2022_en.pdf, p.10.

³ It is considered as a BIK, because the cost of bikes is first covered by employers. It is then employers' choice whether they deduct the amount covered from the relevant employees' salaries. See Revenue, "Taxation of Employer Benefits: Cycle to Work Scheme," 7 February 2022, https://www.revenue.ie/en/jobs-and-pensions/taxation-of-employer-benefits/cycle-to-work-scheme.aspx.

⁴ Revenue, "Taxation of Employer Benefits: Cycle to Work Scheme," 7 February 2022, https://www.revenue.ie/en/jobs-and-pensions/taxation-of-employer-benefits/cycle-to-work-scheme.aspx; Úna Ní Éigeartaigh, Tomás Campbell, and David Crowe, *Spending Review 2021: An Examination of the Cycle to Work Scheme*, November 2021, https://assets.gov.ie/205027/7834435b-7997-4c58-886f-6809e9f3fde0.pdf, p.7.

⁵ The value of €4,000 was chosen simply to make the analysis easier for the reader to follow.

Methodology

Addressing data scarcity by statistical simulation

As mentioned in the introduction, there is no systematic administrative data on the current level of tax forgone under the C2W scheme. However, it was possible to identify several relevant pieces of information across different sources, albeit with different degrees of credibility (e.g., estimates vs. actual observed data).

To use all these pieces under a coherent analytical framework, we develop a statistical simulation model using the Monte Carlo analysis framework. Simulation analysis here creates synthetic data on individuals based on aggregated data and calculates the cost based on these synthetic data. Uncertainty around the factuality of synthetic data is captured by multiple iterations (here, 1,000 times) of the same simulation process with stochastic elements. The full mathematical specification of the data generating process is available in Appendix A.

The Department of Finance estimated that 22,000 people used the Cycle-To-Work scheme in $2020.^7$ According to one source, 40% of the bikes sold in one major shop in Dublin were ebikes,⁸ which conforms to a survey finding that 44.2% of respondents "would likely buy an E-Bikes on their next purchase." We assume that the number of people getting e-bikes to commute to work in 2023 under the Cycle-To-Work scheme would be $8,800 \ (= 22,000 \times 0.4)$. To model the uncertainty around this figure, we allow it to deviate by +/- 20% with a probability of approximately 95%.

To calculate the income tax, Employer's and Employee's PRSI, and USC, we simulate individual-level earnings data, based on the Central Statistics Office (CSO) data on the percentiles of gross weekly earnings from 2020 (the latest available at the time of writing),¹⁰ as done in one of our previous working papers.¹¹ Note that although the impact of the COVID-19 pandemic on the labour market might make the distribution of earnings from 2020 not necessarily representative of the situation in 2023, the 2018 CSO earnings data are not much

⁶ For details on the technique, see Akisato Suzuki, "Simulation Micro Data for Policy and Costing Analysis," PBO Working Paper Series No. 1 of 2022, July 2022, https://data.oireachtas.ie/ie/oireachtas/parliamentaryBudgetOffice/2022/2022-07-12_simulating-micro-data-for-policy-and-costing-analysis en.pdf.

⁷ Department of Finance, "Budget 2022 Report on Tax Expenditures 2021: Incorporating Outcomes of Certain Tax Expenditure & Tax Related Reviews Completed Since October 2020," October 2021, https://assets.gov.ie/201592/f3229427-cdd1-4748-bed3-d089d61ffdd7.pdf, p.123.

⁸ Ronan McGreevy, "E-bikes: The Future of Transport That Has Already Arrived," 10 August 2022, https://www.irishtimes.com/transport/2022/08/10/e-bikes-the-future-of-transport-that-has-already-arrived/.

⁹ Bike To Work Ltd, "The Big Bike to Work 2021 Poll Report," January 2022, https://www.biketowork.ie/blog/latest/the-big-bike-to-work-2021-poll-report, p.16.

¹⁰ Central Statistics Office, "Earnings Analysis using Administrative Data Sources 2020," 21 December 2021, https://www.cso.ie/en/releasesandpublications/ep/p-eaads/earningsanalysisusingadministrativedatasources2020/distribution/.

Akisato Suzuki, "Simulation Micro Data for Policy and Costing Analysis," PBO Working Paper Series No. 1 of 2022, July 2022, https://data.oireachtas.ie/ie/oireachtas/parliamentaryBudgetOffice/2022/2022-07-12_simulating-micro-data-for-policy-and-costing-analysis_en.pdf.

different in terms of the shape of the distribution.¹² We project the 2023 income distribution, by uprating the 2020 income distribution, using the average annual percentage increase of 3.3% in earnings between 2015 and 2021 calculated based on the CSO quarterly average weekly earnings data.¹³

We sample 8,800 individuals from this simulated earnings distribution. We model the likelihood of being sampled to be greater for higher earners, reflecting the CSO data that those who cycle to work earn more. ¹⁴ In the CSO data, the mean earnings of those who cycle to work are 19% greater than the mean earnings of all workers. This proportional difference can be replicated, if we model such that an increase in earnings from the (projected) mean value to the +1 standard deviation increases the likelihood of being sampled by 15 percentage points on average. We accommodate the uncertainty in the likelihood of individuals being sampled as a function of earnings, by allowing an increase in the average likelihood to vary between 9 and 20 percentage points with a probability of approximately 95%.

We assume that the average price of e-bikes purchased would be $\[\in \] 3,000$. According to one Irish source, typical e-bikes cost between $\[\in \] 1,500$ and $\[\in \] 3,000$. Because cargo e-bikes, on average, cost more than standard e-bikes, we consider taking the upper bound of this range to be reasonable. Nonetheless, to capture the uncertainty in this figure, we allow it to deviate by $\[+/\] -20\%$ with a probability of approximately 95%.

We also assume that the prices of e-bikes purchased would be distributed right-skewed. In other words, the majority of e-bikes purchased would be priced below the average, but there would be a few cases where much more expensive e-bikes would be purchased. The assumption of the right-skewed distribution for the prices of e-bikes purchased reflects the fact that the earnings are also distributed right-skewed. We specify the distribution of the prices of e-bikes such that the value range of simulated data generally conform to that of the market prices of e-bikes.¹⁷

¹² Central Statistics Office, "Earnings Analysis using Administrative Data Sources 2018," 15 November 2019, https://www.cso.ie/en/releasesandpublications/ep/p-eaads/earningsanalysisusingadministrativedatasources 2018/distribution/.

¹³ Central Statistics Office, "EHQ03: Average Earnings, Hours Worked, Employment and Labour Costs," 30 August 2022, https://data.cso.ie/table/EHQ03.

¹⁴ Central Statistics Office, "IIA18: Earned Income," 21 September 2020, https://data.cso.ie/table/IIA18.

¹⁵ Ronan McGreevy, "E-bikes: The future of Transport That Has Already Arrived," 10 August 2022, https://www.irishtimes.com/transport/2022/08/10/e-bikes-the-future-of-transport-that-has-already-arrived/.

¹⁶ Note that, after we conducted the original confidential costing (see the disclaimer), a new maximum non-taxable value (€3,000) was recently introduced in the Finance Bill 2022 for cargo bikes including cargo e-bikes. See Department of Transport, "Minister for Transport Welcomes Increase in Support for Cargo Bikes under the Bike to Work Scheme," gov.ie, 20 October 2022, https://www.gov.ie/en/press-release/15246-minister-for-transport-welcomes-increase-in-support-for-cargo-bikes-under-the-bike-to-work-scheme. As noted in the disclaimer, the data used in this paper reflect only the time when the original confidential costing was conducted; therefore, the analysis here does not reflect this latest proposed change. If the analysis were to be updated to reflect this change, it would add another layer of complexity (i.e., it would be necessary to project the proportion of cargo e-bikes to all e-bikes purchased under the C2W scheme).

¹⁷ For example, see "Over 450 Electric Bikes Compared! What Does an Ebike Cost?" eBikesHQ.com, accessed on 23 November 2022, https://ebikeshq.com/cost-of-an-ebike/. To avoid unrealistic values

Because of the lack of data, it is uncertain whether there is a specific relationship between the level of earnings and the price of e-bikes. For instance, higher earners might decide to purchase more expensive e-bikes, as they can afford them. On the other hand, lower earners might live outside cities because of high living costs and purchase more robust but expensive e-bikes to commute a longer distance. In addition, some employers might provide e-bikes without salary deduction, instead of offering a higher wage. We treat the level of earnings and the price of e-bikes as independent from each other, for simplification.

Using these approaches and assumptions, we run the simulation analysis 1,000 times. We report the probability that the additional annual cost due to the increased maximum non-taxable value is equal to, or greater than, a specific threshold cost value, via data visualisation. This is because there is a significant variation in the estimates and, therefore, it is unreasonable to rely only on a single number. While the confidence interval is the standard approach to summarise the uncertainty of estimates, it is inappropriate in the current analysis as the distribution of estimates is heavily skewed.

Calculating the cost of increasing the maximum non-taxable value

Under the current C2W scheme, any excess value beyond €1,500 is added to the gross salary as notional pay, and the combined value is used to calculate income tax, PRSI, and USC liability. The additional annual cost generated by the new maximum non-taxable value of €4,000 is calculated, as the difference between the current scheme and the revised scheme, in terms of the annual revenue that would be collected (i.e., the sum of income tax, Employer's and Employee's PRSI, and USC).

The analysis here generally assumes no behavioural changes but accommodates the following two behavioural changes potentially resulting from an increase in the maximum non-taxable value. First, employees could decide to obtain more expensive e-bikes if the maximum non-taxable value were increased. We then assume that this behavioural change would not make the prices of those more expensive e-bikes exceed the increased maximum non-taxable value. Given this assumption, the calculation of the additional cost due to the increased maximum non-taxable value is equivalent to that under the no-behavioural-change assumption, where the prices of e-bikes would not change. A proof of this is available in Appendix B.

Second, an increase in the maximum non-taxable value could create a new cohort of employees who would obtain e-bikes only under the revised scheme and not under the current scheme. Any e-bikes purchased additionally due to this behavioural change would qualify for the tax exemption. However, the exempted taxes for these additional purchases does not constitute part of the additional cost due to the higher maximum non-taxable value. This is because the additional purchase of e-bikes would not take place in the first place, if the maximum non-taxable value did not increase and, therefore, if there were no behavioural change. In other words, the price of e-bikes, and the tax on these, under the current C2W scheme for these

being simulated, we impose \in 500 as the minimum value of the distribution. It follows that the distribution of the average price of e-bikes also has the minimum value of \in 500 imposed on.

¹⁸ Also note that taxation on BIKs arises in general, if the combined value of an employee's income and taxable value of the BIK exceeds €1,905 in a year. Revenue, "What is Benefit in Kind (BIK)?" 16 March 2022, https://www.revenue.ie/en/employing-people/benefit-in-kind-for-employers/what-is-benefit-in-kind/index.aspx.

individuals are always zero. Combined with the assumption on the first type of behavioural change above, the calculation of the additional cost due to the increased maximum non-taxable value is, again, equivalent to that under the no-behavioural-change assumption. A proof is available in Appendix B.

We calculate the income tax, PRSI, and USC for the simulated data on 8,800 individuals, after taking into account the non-taxable values of e-bikes for these individuals. For analytical simplicity, these taxes are calculated, using the baseline income tax and tax credits, Class A PRSI, and USC, for employees being in a private or public sector, single or under separate treatment for tax purposes, and without any children.

Results

Figure 1 on the next page presents the estimated additional annual cost (in million euro) expected by the increase in the maximum non-taxable value for e-bikes from $\[mathebox{\in} 1,500\]$ to $\[mathebox{\in} 4,000\]$. The figure shows the probability that the additional annual cost will be equal to, or greater than, a specific threshold value (i.e., the "at least" cost value caused by the increase in the maximum non-taxable value). Note that, as the estimates inherently involve uncertainty, it is possible that, if the maximum non-taxable value were indeed increased to $\[mathebox{\in} 4,000\]$, the actual cost could be outside the presented range, for example, because of unexpected external shocks to society and/or the economy.

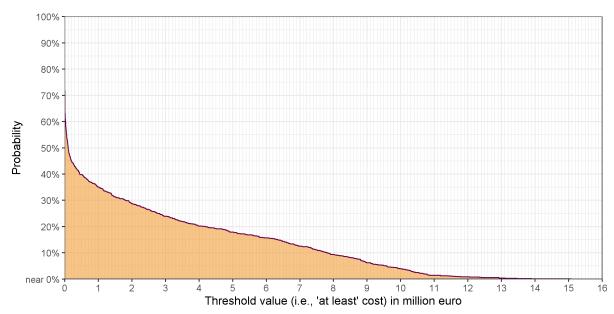
The probabilities are on the y-axis; the thresholds values in million euro are on the x-axis. The figure can be interpreted as follows. For example, the probability of the additional cost being at least $\in 1$ million is approximately 35%. Provided that the increase in the maximum non-taxable does not reduce the number of e-bikes purchased under the C2W scheme, the probability of the additional cost being at least $\in 0$ is, by definition, 100%. However, this does NOT mean that there is 100% probability that the proposed scheme will cost nothing; it means that there is 100% probability that the proposed policy will cost $\in 0$ or any greater amount than this. 19

The probability of the cost falling in a specific range of values can be read from the figure as follows. First, look at the probability for the lower bound of the range of interest, and the probability for the upper bound of the range. Second, subtract the latter probability value from the former probability value. For example, if we want to know the probability of the cost being between &0 and &0.1 million, it is calculated as the probability of the cost being at least &0 (100%) minus the probability of the cost being greater than &0.1 million (approximately 50%), which results in 50% probability that the cost of the proposed policy is between &0 and &0.1 million.

The mean of the distribution is approximately $\in 2$ million. However, as seen on the figure, there is significant uncertainty around the estimates. Therefore, relying on the mean cost estimate may not necessarily be appropriate.

¹⁹ There is the adjective "near" on 0%, because theoretically the upper bound of the additional cost is undefined and, therefore, the exact 0% probability is mathematically incomputable. Meanwhile, since the additional cost due to the revised maximum non-taxable value is ≥ 0 by the setup of the analysis, the probability of the additional cost being equal to or greater than $\epsilon 0$ is 100%.

Figure 1: Probability of the additional annual cost (due to the increased maximum non-taxable value) being equal to or greater than a specific threshold value



Source: PBO's own modelling.

Conclusion

One of the factors to consider in costing analysis is that relevant data are not always available at high-quality levels for a diverse set of policy areas. This paper has demonstrated a methodological approach to mitigate challenges posed to costing analysis when data are scarce. It has utilised a statistical simulation technique using the Monte Carlo analysis framework, to "glue" incomplete pieces of information with assumptions used to fill in the missing pieces and the associated uncertainty taken into account. The paper has shown the approach, by estimating the cost of an increased maximum non-taxable value of the Cycle-To-Work scheme for e-bikes. The results have highlighted significant uncertainty in the cost estimate. The uncertainty is a natural consequence of the data scarcity and, therefore, should be appreciated as a cautionary note for interpreting the costing analysis, rather than being ignored or overlooked. To reduce uncertainty in cost estimates, it will be necessary to obtain more and higher-quality data. This paper, therefore, also points to the importance of recording and keeping data for policy and costing analyses.

Appendix A: Data Generating Process

$$\begin{split} y_i \sim LN(\mu = 6.509005, \sigma = 0.7419373) \\ n \sim N(\mu = 8800, \sigma = 880, \alpha = 0) \\ y_i^* \sim s(y_i, n, p_i = \text{logit}^{-1}(\pi_i y_i^z)) \\ \pi_i \sim N(\mu = \log(1.9), \sigma = \log(1.15)) \\ c_i \sim LN\left(\mu_i = \log(\bar{c}) - \frac{1}{2}(0.2623643^2), \sigma = 0.2623643, \alpha = 500\right) \\ \bar{c} \sim N(\mu = 3000, \sigma = 300, \alpha = 500) \end{split}$$

| Term | Explanation |
|--------------------------------------|---|
| y_i | Earnings for an individual <i>i</i> in the population |
| $LN(\cdot)$ | Log-normal distribution |
| μ | Mean parameter |
| σ | Standard deviation parameter |
| α | Lower bound of the distribution |
| n | Number of employees who will obtain e-bikes under the C2W scheme |
| $N(\cdot)$ | Normal distribution |
| ${\boldsymbol{\mathcal{Y}}_{i}^{*}}$ | Earnings for an individual i in the subset, n |
| $s(\cdot)$ | Function to sample a size n from a vector of y_i with a vector of probability p_i |
| $logit^{-1}(\cdot)$ | Inverse logit function to map log odds onto probability |
| π | Log odds ratio coefficient |
| y_i^z | Statistically standardised y_i |
| c_i | Price of an e-bike for an individual <i>i</i> |
| \bar{c} | Expected value of c_i |

Appendix B: Proofs on the Points about Behavioural Changes

Greater prices of e-bikes

Let p_i^0 be the price of an e-bike that an individual i would obtain under the current scheme, where the superscript, 0, denotes the current scheme. Let p_i^1 be the price of an e-bike that i would obtain, if the maximum non-taxable value were increased, where the superscript, 1, denotes the revised scheme. The behavioural change with respect to the prices of e-bikes obtained implies $p_i^1 \ge p_i^0$ for all i.

Let β^0 and β^1 be the maximum non-taxable value under the current and revised schemes respectively, where $\beta^0 < \beta^1$. Let $f(\cdot)$ be the function to calculate a taxable value on a benefit x: f(x) = x if $x \ge 0$; f(x) = 0 otherwise (i.e., if x < 0).

The taxable value under the current scheme is $v_i^0 = f(p_i^0 - \beta^0)$. The taxable value under the revised scheme is $v_i^1 = f(p_i^1 - \beta^1)$. If there is no behavioural change with respect to the prices of e-bikes, then $p_i^{1^*} \equiv p_i^1 = p_i^0$. The taxable value under this special case is $v_i^{1^*} = f(p_i^{1^*} - \beta^1)$. Note that it follows from the definition of $f(\cdot)$ that $v_i \geq 0$ for all i.

The additional cost per i due to an increased maximum non-taxable value is $v_i^0 - v_i^1$. Under the no-behavioural-change assumption, the additional cost is $v_i^0 - v_i^{1*}$. Relaxing this assumption (i.e., allowing $p_i^1 - p_i^0 > 0$), $v_i^1 = v_i^{1*}$ as long as $p_i^1 - \beta^1 \le 0$.

A greater number of beneficiaries under the revised scheme

Let t_i^0 be the amount of the tax that an individual i would pay on an e-bike they obtained under the current scheme. Let t_i^1 be the amount of the tax that i would pay on an e-bike they obtained under the revised scheme. Let both terms be a monotonic function of v_i , $t_i^0 = g(v_i^0)$ and $t_i^1 = g(v_i^1)$ respectively, where $g(v_i = 0) = 0$ and $t_i^0 \ge t_i^1$ if $v_i^0 \ge v_i^1$.

With the assumption $p_i^1 - \beta^1 \le 0$ as in the above proof, $v_i^0 \ge v_i^1$ and, therefore, $t_i^0 \ge t_i^1$. The difference in the tax collected between the current scheme and the revised scheme is $t_i^0 - t_i^1$ per i.

Now, let us assume that an increase in the maximum non-taxable value would create a new cohort of individuals, indexed by j ($i \neq j$), who would then obtain e-bikes (i.e., those who would not obtain e-bikes if the maximum non-taxable value did not increase). The difference in the tax collected for these individuals is $t_j^0 - t_j^1$. Because these individuals would not obtain e-bikes under the current scheme, $p_j^0 = 0$, $v_j^0 = 0$, and therefore $t_j^0 = 0$, for all j. Given $p_j^1 \geq p_j^0$ and $p_j^1 - \beta^1 \leq 0$, $g(v_j^0) - g(v_j^1) = t_j^0 - t_j^1 = 0$ for all j. It follows that $\sum_i (t_i^0 - t_i^1) + \sum_j (t_j^0 - t_j^1) = \sum_i (t_i^0 - t_i^1)$.