Evaluating the impact of BII's investment in broadband fibre infrastructure in DRC

Insights from Liquid Intelligent Technologies

Authors: René Kim and Vivek Sakhrani

Date: December 2025

Acknowledgements

Itad and Steward Redqueen would like to extend our sincere gratitude to the staff of British International Investment (BII) for generously providing their time and information in support of this research. We would also like to thank Liquid Intelligent Technologies for their participation in the research process and for supplying the data that underpins this report.

Disclaimer

This report has been prepared by Itad for the Foreign, Commonwealth and Development Office (FCDO) for services specified in the Terms of Reference and contract of engagement.

For reports that are formally put into the public domain, any use of the information in this report should include a citation that acknowledges Itad as the author of the report.

This material has been funded by UK aid from the UK government; however, the views expressed in this report are those of the authors. They do not represent those of British International Investment (BII), the FCDO or any of the individuals and organisations referred to in the report.

Suggested citation

Itad (2025) Evaluating the impact of BII's investment in broadband fibre infrastructure in DRC: Insights from Liquid Intelligent Technologies. Brighton: Itad.

Copyright

Contents

List of figures

List of tables

List of acronyms

AI	Artificial Intelligence
ADM	Add/Drop Multiplexer
AWI	Asset Wealth Index
BII	British International Investment
CNN	Convolutional Neural Network
DFI	Development Finance Institution
DHS	Demographic and Health Surveys
DRC	Democratic Republic of the Congo
FCDO	Foreign, Commonwealth & Development Office
GDP	Gross Domestic Product
GSMA	GSM Association (Global System for Mobile communications)
HREA	High Resolution Electricity Access
ICT	Information and Communication Technology
ILA	Intermediate Line Amplifier
MNO	Mobile Network Operator
OECD	Organisation for Economic Co-operation and Development
OSM	OpenStreetMap
PCA	Principal Component Analysis
RCT	Randomised Control Trial
SCM	Synthetic Control Method
SNEL	Société Nationale d'Électricité
SSA	Sub-Saharan Africa
USAID	United States Agency for International Development

Executive Summary

Investment context and purpose

This study examined the impact of broadband internet backbone fibre on household asset wealth and spending in the Democratic Republic of the Congo (DRC), focusing on the impact of British International Investment's (BII) investment in Liquid Intelligent Technologies. It was undertaken as part of a broader evaluation of BII's infrastructure portfolio by Itad and Steward Redqueen, designed to strengthen BII's evidence base on the role of infrastructure investments in delivering economic and climate outcomes.

Liquid Intelligent Technologies is a pan-African technology group present in 20 countries, mainly in sub-Saharan Africa (SSA). BII first made an equity investment of US$180 million in Liquid to accelerate the expansion of its fibre cables along the Cape–Cairo route and further into Central and Western Africa, with the objective of improving access to affordable and quality connectivity in Africa. In 2020, it was followed by a US$40 million equity investment to enhance Liquid's data centre business.[1]

This study explored how Liquid's broadband internet fibre cable between Kinshasa and Lubumbashi – the SNEL line,[2] which became operational in March 2021 – has resulted in increased household wealth and spending. Internet penetration in the DRC is among the lowest in the world and the internet speed and quality are poor. The fibre line will allow mobile network operators (MNOs) to offer customers faster and more reliable internet access and speed in regions close to the fibre line.

The study had two purposes. First, it aimed to **develop an evidence-based understanding of the impact** generated as a result of faster and more reliable internet access resulting from communities connected to a Liquid Technologies fibre backbone. Second, it aimed to further **develop and test a cost-effective, replicable and scalable approach to evaluation** using satellite data and machine learning techniques, which BII can use in ongoing or future investments.

Method

The methodology underpinning this study aimed to quantify the impact of infrastructure investments in settlements that received fibre cable treatment against comparable ones that did not, using an innovative approach. Central to the analysis was the construction of robust datasets which combined high-resolution daytime and nighttime satellite imagery with survey data provided by AtlasAI. AtlasAI use advanced machine learning techniques to integrate these two types of data, allowing for a comprehensive and granular understanding of household asset wealth (as expressed by Atlas AI's proprietary Asset Wealth Index (AWI), and household spending at a 1km x 1km scale. This fusion of satellite and survey data enabled the detection of subtle changes in living standards that might otherwise have gone unnoticed.

---

[1] The impact of Liquid's data centres is the subject of another evaluation.

[2] The line is constructed along a power transmission line of Société Nationale d'Électricité.(SNEL)

Once the datasets were established, the study applied the synthetic control method (SCM), a rigorous statistical approach used to estimate the causal effects of interventions when randomised experiments are not feasible. By creating a synthetic control group – a weighted combination of settlements that did not receive access to the fibre line – the analysis provided a counterfactual against which to compare the settlements that received fibre cable treatment. This allowed for the quantification of differences in asset wealth and spending attributable to the arrival of faster, more reliable internet.

The results

Liquid's SNEL broadband fibre cable is providing an estimated 2.5 million internet users with access to faster and more reliable internet.

For reasons of comparability with settlements that do not benefit from fast internet, this study focused on nine urban settlements in which MNOs can connect their networks to Liquid's fibre cable. In these settlements, based on population statistics and World Bank and GSMA estimates of internet penetration among urban/rural and male/female populations in DRC, we estimated that 726,691 male and 465,200 female, i.e., 1,191,891 internet users, would potentially[3] have access to faster and more reliable internet. The total number of users that benefit from the fibre line is likely twice as large, i.e. approximately 2.5 million. Moreover, these numbers will increase as internet penetration rates continue to go up and as more settlements along the SNEL line (or further away from it) are connected through network investments by MNOs.

Although well below the US$4.20 lower-middle income threshold of the World Bank, the studied population live in urban areas and are wealthier than most of the DRC's population

In terms of average AWI, the households in the treatment areas are wealthier than 98% of the entire DRC population. Their average spending per capita in 2020 of US$3.38 (in 2021 PPP terms), although higher than 91% of the DRC's population, was well below the US$4.20 and US$8.30[4] per day per capita that the World Bank defines as the lower-middle income and upper-middle income thresholds, and close to the international poverty line of US$3.00.

There is emerging evidence that households in the nine urban settlements studied have benefitted because of the fibre line.

Almost four years after the arrival of faster internet, the households in these nine urban settlements experienced the following:

- An increase in the average household asset wealth (as measured by AWI) has moved the treated settlements from the 98th to the 99th percentile settlements. This compares to the synthetic control settlements moving from the 97th to the 98th percentile.
- US$1.25 PPP higher spending per capita in 2024, compared to an increase of US$0.30 for the synthetic control. In other words, the spending of households in settlements that were

[3] Whether or not existing internet users can benefit from faster internet does not only depend on their vicinity to Liquid's fibre line but also whether the networks of MNOs and the handheld devices of the users allow for faster data transmission.

[4] These thresholds are in terms of 2021 PPP spending. In terms of 2017 PPP spending, they correspond to US$3.65 and US$6.85.

connected to the fibre line increased by US$0.95 per capita per day more compared to households in settlements that were not. This has moved the treated settlements from the 91st to the 97th percentile for spending set against the distribution of all households in the region, compared to the synthetic control, which moved from the 89th to the 94th percentile.

These results, however encouraging, are not sufficiently robust yet to claim strong causal inference.

There is a 32% probability that the observed AWI results could have arisen by chance, which is well above the usual 5% considered as statistically significant. Although the spending results have greater statistical significance, at 7% in 2024, the higher confidence should be interpreted with caution. Since spending is partially derived from AWI, the additional assumptions and propagating errors within it make it a less stable dataset than AWI. Consequently, the true reliability of these findings may be overstated.

Care must be observed to infer causality.

The SNEL line covers a very large area and unobserved confounding factors in the pre-treatment period and idiosyncratic post-treatment shocks can introduce sources of bias. The fact that the first years after the fibre line became operational coincided with the Covid 19 pandemic and the recovery from it is an example of this. Any regional differences in the severity or recovery speed could have affected the findings.

Recommendations

1. **Based on the emerging evidence collected in this study, BII should consider other broadband investments in underserved regions.** Because of its size, inaccessibility and limited economic development, the DRC is among the most underserved countries in terms of backbone fibre. Although globally the largest gaps in backbone infrastructure are being filled, there are still regions where backbone internet investments are much needed. Within the DRC, the completion of the Kananga – Goma line is recommended. Some examples other than DRC are Ethiopia and South Sudan.
2. **We recommend that BII continues to invest in the 'last mile' through which people access the internet.** The presence of broadband fibre is necessary but not sufficient to increase internet quality and penetration. Last mile investments in internet connectivity involve MNOs through the (co)hosting of towers and higher bandwidth mobile antennas. They also involve financing internet-capable handsets, which the Global System for Mobile Communications Association (GSMA) now considers the single greatest barrier keeping people offline.
3. **If possible, pursue opportunities to ground-truth the findings for this investment and to understand the drivers of change at a household level**. BII can seek opportunities to ground-truth the findings of this study by comparing the results to evidence collected through other methods, including primary (survey) data. This could involve estimating (with MNOs that use the fibre line) how internet penetration and usage have developed in the regions along the line, as well as how business growth and productivity have changed in these areas. While broadband fibre forms the essential backbone for internet connectivity, it alone does not guarantee increased usage or penetration, and a ground-

truthing study should also examine other elements along the evidence chain to provide a more comprehensive understanding.

4. **Repeat the analysis in two years to increase the statistical robustness of the emerging impact and its durability and include more settlements that were connected after 2021**
   a. When the impact is durable over a longer post-treatment window, the cumulative gap between treated units and the synthetic control should become more pronounced. Statistical confidence also increases as random (placebo) effects dissipate over time.
   b. More settlements have obtained fibre access points after 2021, both along the SNEL line as well as the Kananga–Goma line, if fully completed (see recommendation 1). Inclusion of these settlements could also strengthen the robustness of the study. An important requirement, based on the results in this study, is that four years of post-treatment data should be available. Including settlements that were connected from 2022 onwards would also mitigate the possible impact of any regional heterogeneity resulting from the intensity of and the recovery from the Covid 19 pandemic.

1. Introduction

In 2019, the Foreign, Commonwealth & Development Office (FCDO) commissioned Itad and Steward Redqueen to independently evaluate British International Investment's (BII's) infrastructure portfolio. The purpose of this evaluation was to better understand the development outcomes and impacts associated with BII's investments in the infrastructure sector. The assignment consisted of two phases: an evidence and portfolio-level review (Phase 1), published in 2022, and a subsequent series of in-depth case studies (Phase 2).

This study had two purposes. First, it aimed to **develop an evidence-based understanding of the impact** generated because of new or improved access to broadband Internet through BII's investment in Liquid Technologies. Second, it aimed to further **develop and test a cost-effective, replicable and scalable approach to evaluation**, using satellite data and machine learning techniques, which BII can use in ongoing or future investments. Itad and Steward Redqueen have been working together with AtlasAI to develop this approach, which was used first in the impact assessment of Virunga Energies (Itad, 2024). The report is structured as follows:

Section 2: Study context. This provides the background to the study and the link to the previous phase of our evaluation work, in which we systematically identified areas of BII's infrastructure portfolio that could benefit from more in-depth evidence. This section also discusses what the study is trying to achieve, emphasising its strong learning focus, with the aims of (i) helping BII understand where it is generating impact in its infrastructure portfolio and (ii) demonstrating a new, cost-effective, flexible and replicable approach to impact evaluation. It provides an overview of the challenges associated with assessing the impact of infrastructure investments and the study design adopted for addressing these challenges.

In **Section 3: Evaluation approach and concepts**. This discusses how the study defines impact, focusing on household asset wealth and spending and how these measures have been adapted to assess the impact of infrastructure investments on underserved urban communities for the purpose of this study.

Section 4: Evaluation methodology. This provides a step-by-step description of how the study was implemented in practice. This includes a discussion of the data used, the identification and selection process for the settlements to include in the analysis, and how the geospatial impact evaluation approach, with the synthetic control method (SCM), has been defined and used to isolate impact.

In **Section 5: Key findings**. This discusses the asset wealth and spending results study, focusing on the development impact of the investment by applying the geospatial impact evaluation approach; how the robustness of the results was tested; how the results compare to each other; and what has been learned about applying the approach in comparison to alternatives.

Section 6: Conclusions and Recommendations. This discusses the key takeaways from the study and proposes a set of recommendations and next steps for BII and FCDO, based on the study's findings and the lessons learned from applying the new approach.

2. Study context

This section outlines the overall purpose of the study and how it fits with the wider evaluation of the BII infrastructure portfolio. It also includes a discussion of the rationale for the focus on broadband investments and their strategic importance to BII. This is followed by an overview of the investment studied – Liquid Intelligent Technologies – in which BII invested US$180 million in 2018 and another US$40 million in 2020. Liquid is the largest independent fibre and cloud services provider in Africa. In this report, the focus is on Liquid's backbone fibre investments in the Democratic Republic of the Congo (DRC).

2.1. Purpose

This study had the following two primary aims:

- First, it sought to develop an **evidence-based understanding of the impact generated as a result of BII's investment in Liquid, specifically its backbone fibre line in the DRC**. This is linked to the wider goals of Itad and Steward Redqueen's evaluation, which sought to **deepen BII's evidence base on the impact it is generating through infrastructure investments**.
- Second, it aimed to further develop a **low-cost, flexible and scalable approach to SCM, which can be replicated by BII to evaluate the impact of localised infrastructure investments,** using satellite data and machine learning techniques allied to recent thinking in the use of synthetic control analysis. The approach was used before to analyse the impact of electrification in the DRC.

The study focused on **three areas of BII's impact framework for the Information and Communication Technology (ICT) sector**: i) additional capacity (for data transport); ii) improved quality and reliability of Internet access; and iii) (ultimately) an improved standard of living. Phase 1 of the evaluation reviewed global evidence against the infrastructure sector's impact framework. In Figure 1, the evidence base presented in the Phase 1 Evaluation Report[5] is illustrated against the BII impact framework. Phase 1 found limited or weak evidence linking broadband investments to most impact pathways, outcomes and ultimate impacts, with two exceptions: greater productivity of companies and greater economic opportunity. The most tangible evidence, based on two studies,[6] is that a doubling of the average Internet speed is associated with a 0.3% increase in the Gross Domestic Product (GDP) in OECD countries.[7] A third study[8] concluded that fast Internet infrastructure may have the greatest employment-creating potential in Africa, through new firm entries and increasing productivity. There was limited evidence of the impact of broadband on standards of living and this study aimed to fill that gap.

This study, therefore, aimed to deepen the evidence base around the three links highlighted in solid blue in Figure 1, building **further evidence on the impact of improved Internet access on**

---

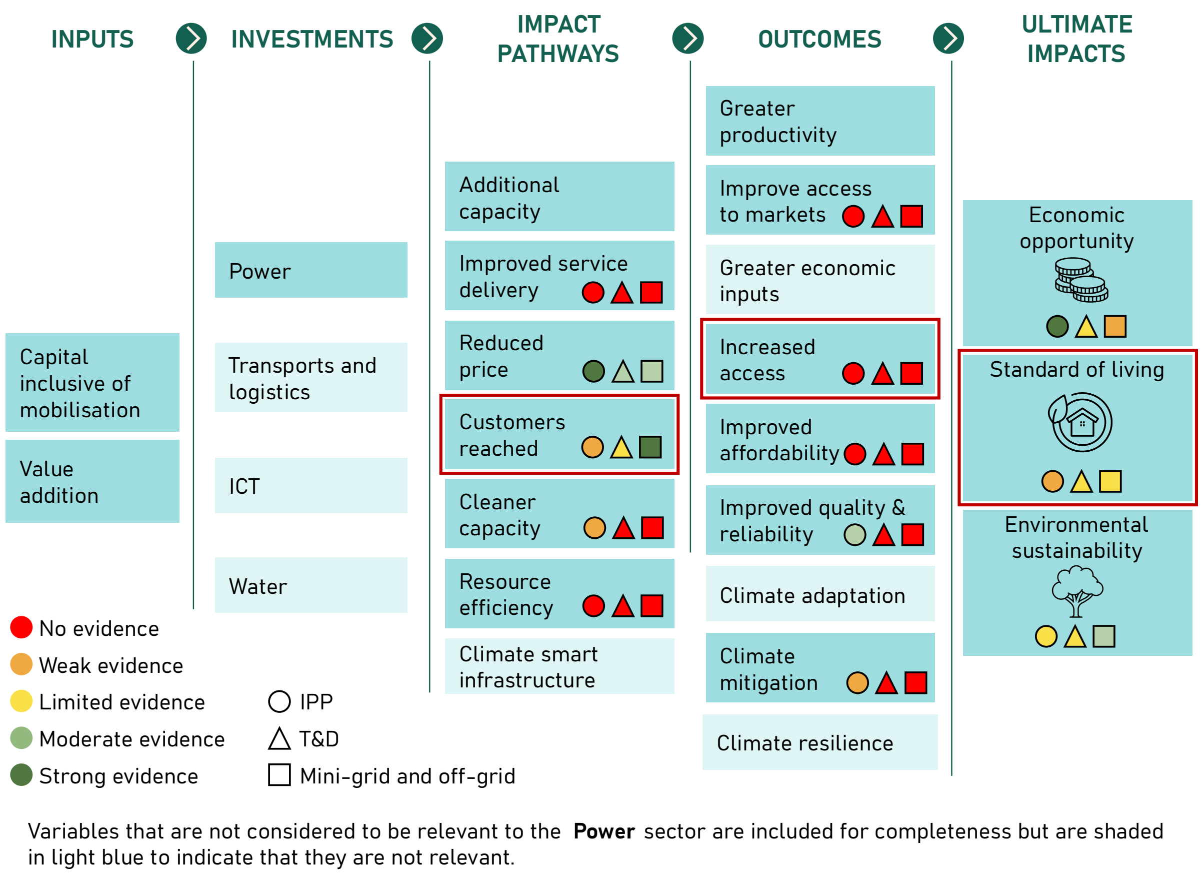
[6] Koutroumpis, P. (2019) The economic impact of broadband: Evidence from OECD countries. *Technological Forecasting and Social Change* 148(119719); Regeneris (2018) The Economic Impact of Full Fibre Infrastructure in 100 UK Towns and Cities.

[7] OECD: Organisation for Economic Co-operation and Development.

[8] Hjort, J. and Poulsen, J. (2019) The Arrival of Fast Internet and Employment in Africa. *American Economic Review* 109(3): 1032–1079.

standards of living. As explained in Section 3, asset wealth is used as a proxy for improved living standards, as captured in the Asset Wealth Index (AWI) developed by AtlasAI. By using some additional survey data, the AtlasAI-generated AWI maps can be translated in estimations of household spending per capita, in purchasing power parity terms.

Figure 1. BII impact framework for the infrastructure sector, with strength of evidence from Phase 1



The study also offered an opportunity to 'extend the toolkit' of impact assessment options available to BII, given that much of the current evidence of impact across the portfolio is reliant on modelling techniques (using methodologies such as the Joint Impact Model)[9] and on self-reporting by investees. This study went further by assessing observational data on impact but aimed to do so in a way that was appropriate and feasible for BII to replicate. To a large extent, the methodology deployed in this study replicated the impact study for Virunga Energies on the impact of mini-grid access to electricity in rural areas.[10]

Because of data limitations, this study was not able to cover the non-marked elements of the sector framework. For example, increased access and greater productivity could not be directly observed, because Liquid has no line of sight to the end customer. Improved affordability, which was emphasised in the Development Impact (DI) thesis for the investment, could not be observed directly either; however, lower prices directly due to a local backbone fibre line were unlikely

---

[9] https://www.jointimpactmodel.org/

[10] Itad, Evaluating the impact of a hydroelectric power investment in the Democratic Republic of the Congo, BII, 2024

because of the national pricing strategies of mobile network operators (MNO).[11] As shown in Table 1 (on page 6), the general affordability of mobile internet (usage and devices) in all of the DRC improved from 2020 to 2022 but decreased thereafter. It was impossible (with the chosen methodology for this study) to identify a causal link with the Société Nationale d'Électricité (SNEL) line. By looking at the AWI and average household spending, this study amalgamated several of the pathways and outcomes through which the presence of a fibre backbone impacts standards of living.

2.2. Strategic importance and relevance to BII

This study was undertaken as part of the wider evaluation of BII's infrastructure portfolio. It followed the Phase 1 evaluation, conducted by Itad and Steward Redqueen, which reviewed BII's infrastructure portfolio. The Phase 1 evaluation identified a long list of 13 evidence opportunities, where the existing evidence of impact in the infrastructure sector was weak and where it was feasible that BII could deepen its evidence base. One of these was "*the impact of broadband backbone investments on service delivery and onward linkages to customers reached and increased access*".

Extending the evidence base on the impact of broadband backbone is of strategic importance, given the urgent need for new investments to increase Internet penetration across Africa. According to the World Bank, in sub-Saharan Africa (SSA), only 36% of the population were connected to the Internet in 2024, compared to 52% in South Asia, the next least-connected region. Although only 13% (167 million people) of the people in Africa live in a zone where there is no mobile broadband Internet, 60% of the people who live in connected zones do not use mobile Internet services (a so-called usage gap of 770 million people).[12]

Extending the evidence base on impact for internet backbone investments is relevant to BII given that the data compiled for the 2012–2024 Portfolio Review indicated that ICT investments make up approximately 19% of BII's investment in infrastructure, 21% of which is in broadband backbone fibre and data centres. This is mostly through its investment in Liquid Technologies. Backbone infrastructure is necessary but not sufficient for increasing access to affordable, high-quality and reliable Internet access. MNOs are needed to provide the wired and wireless internet access that is carried by backbone providers such as Liquid

Background on Liquid Technologies

The investment selected for study was BII's investment in Liquid Technologies. In 2018, BII made a US$180 million equity investment in Liquid Technologies to accelerate its expansion along the Cape–Cairo route and further into Central and Western Africa, with the objective of improving access to affordable and quality connectivity in Africa. In 2020, BII made a follow-on US$40 million investment to support Liquid's data centre expansion.[13] The long-term aim of BII's investment was to improve access to affordable and quality internet.

[11] It may be that, in the much longer run, the fibre line attracts more price competition between and new entry of MNOs. According to 2024 GSMA data, in only three countries is internet less affordable than in DRC: South Sudan, Burundi and Zimbabwe.

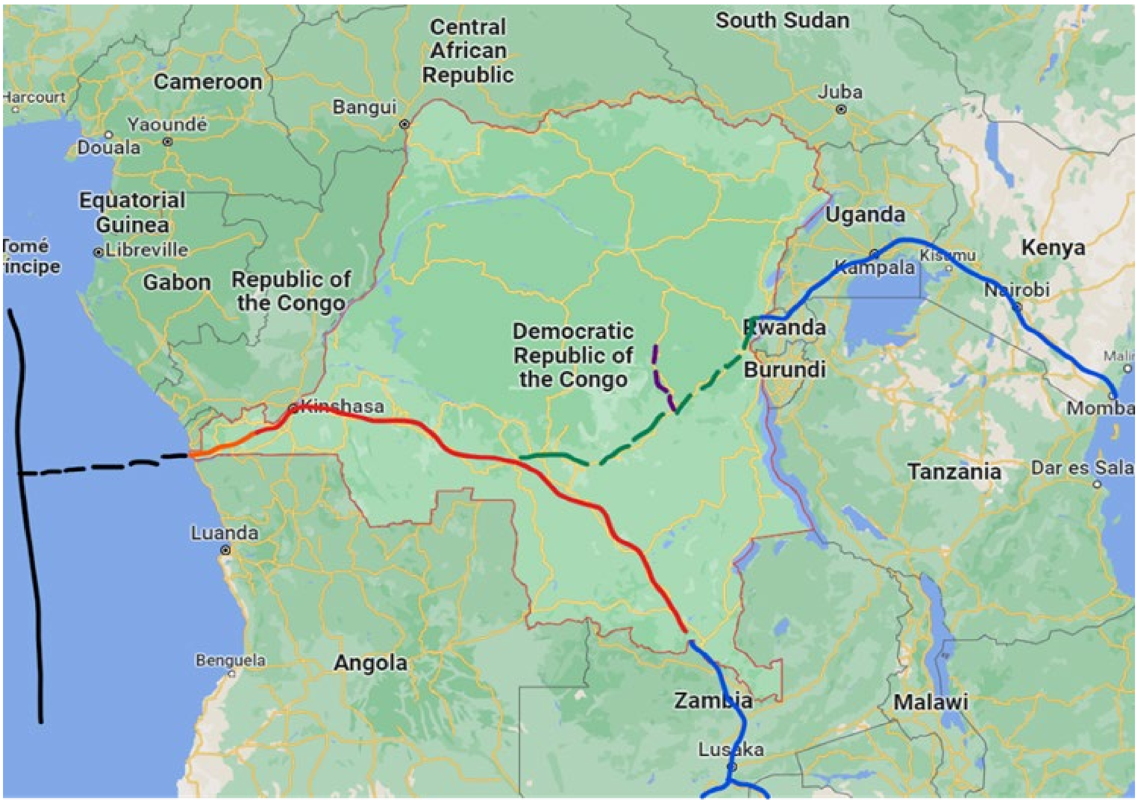
[12] GSMA (2023) The State of Mobile Internet Connectivity 2023.

[13] The impact of Liquid data centres and cloud services is covered in a separate BII report.

The focus of this report is on fibre backbone for broadband Internet in the DRC. The SNEL line (see Figure 2) is a 2,500 km fibre line connecting Kinshasa and Lubumbashi; it became operational in March 2021. The Kananga – Goma line (in green dashed) was completed in an initial phase but is not yet fully operational or secure. After landslide damage, repair works on a 150km stretch in South Kivu could not be completed due to heavy rains and civil unrest, and across 580km of the fibre line, ongoing military and rebel activities prevent maintenance and completion works. Although the Mbuji-Mayi node on the Kananga–Goma line went live in May 2021, the incomplete status of the line limits the extent to which large technology companies (Google, Meta, Microsoft) can provide their data-dependent digital platform services to the market. Apart from a reduced uptake of modern work practices such as Microsoft Teams and Google workspace, this also reduces data consumption and revenues for MNOs and through that their ability to invest in 'last-mile' connectivity.

The lines are open access, which increases competition between providers, which was expected to increase broadband penetration and better-quality access for end users. Broadband access is essentially internet access with much faster data transfer rates, allowing for streaming and modern work services for one or multiple users. Along the fibre lines, MNOs can add or drop traffic using so-called add/drop multiplexers[14] (ADMs). Liquid has no line of sight to the end customer, nor can it track the data volumes transported on its fibre lines.

Figure 2. Different fibre lines of Liquid Telecom in the DRC



***Key:** red — SNEL line; green — Kananga–Goma line; dotted line — under construction.*

[14] A multiplexer combines several lower bandwidth streams of data into a single beam of light. An ADM has the additional capability to add lower bandwith signals or to extract (drop) them and remove them from the stream.

The Liquid fibre lines connect the four largest cities in the DRC. The SNEL line covers 2,500 km and connects three of the four largest cities in the DRC (Kinshasa, Kananga and Lubumbashi). It runs parallel to the SNEL electricity transmission line. The partially completed line to Goma also connects the third-largest city, Mbuji-Mayi. The line goes through many urban and rural areas between these cities. The extent to which these areas gain access to better-quality Internet depends on the presence of ADMs. As will be shown later, most ADMs are in towns along the line.

Internet penetration in the DRC is low. According to the World Bank's 2023 data, Internet penetration stood at 31% of the population, up from 16% in 2019. With a population of 109.3 million people in 2024, the World Bank figures indicate that approximately 34 million people have Internet access.

Mobile connectivity in the DRC is still underdeveloped. Internet access in the DRC is almost entirely through mobile networks. According to the Global System for Mobile Communications Association (GSMA),[15] the DRC has one of the highest coverage gaps: In 2024, 32% of the population did not have access to mobile broadband networks (i.e., 3G or better). However, this was down from 46% in the years prior. As shown in Table 1, the Mobile Connectivity Index in the DRC is rising, largely because of improved affordability and better content and despite only a slightly improved score on infrastructure. Network performance (an amalgamation of download and upload speeds and latency) across the DRC hardly changed in the first four years since the SNEL line became operational in 2021, however, it improved substantially in 2024. The initial lack of improvement was likely due to other factors constraining the speed and reliability of the entire chain. For example, MNOs must upgrade their equipment (e.g., antennas and tower density) to accommodate higher speeds and there can be bottlenecks in how the DRC's backbone network connects to other countries or the fibre line around Africa (see, for example, the Google Equiano line, shown in black in Figure 2).

Table 1. GSMA indicators of the Mobile Connectivity Index and some of the underlying indicators[16]

	2020	2021	2022	2023	2024
Mobile Connectivity Index	21.6	22.5	24.5	25.2	28.2
Enablers					
Infrastructure index	48.1	41.0	40.3	42.2	50.4
Affordability index	10.4	13.6	19.6	17.5	17.0
Consumer readiness index	29.6	29.9	27.8	33.2	28.4
Content and services index	14.6	15.4	16.4	16.5	26.2
Coverage & Performance					
Network coverage of population Index	53.4	45.9	45.9	46.3	57.7

[15] GSMA (2024) The State of Mobile Internet Connectivity 2024; mobile connectivity data.

[16] https://www.mobileconnectivityindex.com/index.html. Index ranges from 0-100. The only countries with a lower overall score than DRC are South Sudan, Central African Republic, Burundi, Niger, Afghanistan and Chad (in increasing order).

Network performance index	30.1	27.3	25.6	30.0	35.6

This study aimed to answer the following research questions:

1. How has fibre broadband access developed geographically over time following BII's investment in Liquid?
2. What is the socioeconomic profile of the communities that have been reached by new connections?
3. How have communities that have been connected to new fibre broadband infrastructure developed over time?
4. Is there evidence that connected communities have demonstrated increased growth rates in their average household wealth and spending relative to non-connected communities?

This study was performed in 2025 using the most recent AWI data available (from 2024). This means that, at the time, the SNEL line had been operational for three years and nine months. Some of the of pathways in Figure 1 may take a long time to be fully activated. For example, MNOs need to connect their existing and new infrastructure to the ADMs; companies need to adapt to faster Internet and grow their business and/or increase their productivity in response; new firms need to enter; new employees must be hired and/or salaries need to adjust; and these changes, subsequently, need to feed through to changes in standards of living, which can be inferred from satellite observations. Based on Hjort and Poulson's study,[17] we estimate that it takes roughly two years after the arrival of fast Internet before changes in night light intensity can be observed, and it likely takes more than double that time before economies have fully adjusted. Based on this estimate, the three years and nine months since the SNEL line became operational should be long enough to discern impacts.

This report is similar in approach to the previously mentioned study on Virunga Energies, with one important difference: the identification of treatment areas. In the Virunga study, the treatment areas could be defined exactly because connections to the mini-grid were known at a household level. In this study, such a clear definition was not possible because customers' access to the faster internet enabled by Liquid's fibre line is wireless, through MNOs. Instead, we chose to focus on urban areas where the fibre cable and, thus, more reliable and faster internet have arrived. The identified control areas were comparable in every aspect except for the arrival of broadband fibre.

[17] Hjort, J. and Poulsen, J. (2019) The Arrival of Fast Internet and Employment in Africa. *American Economic Review* 109(3): 1032-1079.

3. Evaluation approach and concepts

This section highlights the research challenges typically associated with evaluating infrastructure projects (time, cost and methodology) and describes the research solution, which considered recent advances in **synthetic control impact evaluation design** and combined these with **artificial intelligence (AI)-derived geospatial data** to develop a **low-cost and scalable geospatial impact assessment approach,** which BII can use in future studies. This section also discusses how the study defined impact, focusing on household asset wealth and spending and how and why this measure was adopted for the purposes of this study, as well as the approach to utilising AtlasAI's AWI and the derived household spending as proxies for livelihood improvements. Finally, this section outlines some of the limitations of the study.

3.1. The research challenges

Evaluating infrastructure impacts is difficult and tends to be expensive and time-consuming. This study was designed to offer a lower-cost, more flexible approach to meeting these challenges. Challenges in evaluating infrastructure projects include the following: i) non-randomisation; ii) controlling differences between control and treatment groups; iii) incomplete datasets; and iv) the high costs associated with evaluating infrastructure impacts.

First, new infrastructure is not randomised in delivery, therefore, it is difficult to identify adequate counterfactuals. This can be resource-intensive and can be difficult to achieve when dealing with large and diverse treatment areas. Although the issue of non-randomisation cannot be eliminated entirely, in this study we exploited the fact that smaller towns along the SNEL line were connected accidentally rather than intentionally, because they happened to be very close to the fibre line and because of the low marginal cost of adding an ADM.[18] A further complication was that the accidentally connected towns would have had internet access before through mobile 2G or 3G networks; although this does not allow for meaningful or productive use, it makes the arrival of faster and more reliable internet less of a change.

Second, the recipients of the infrastructure may be different from the surrounding, untargeted populations (e.g., they may have higher underlying rates of economic growth), which complicated the use of more traditional evaluation designs, such as difference-in-difference.

Third, national surveys of living standards typically do not revisit the same households or locations across survey waves, or they are repeated infrequently, making it difficult to construct repeated local-level measurements using secondary datasets. This was true for the available DRC datasets. Fourth, traditional impact evaluation designs, which seek to close gaps in existing datasets through on-the-ground surveys, tend to be time-consuming and expensive to administer. This is particularly true for the DRC, where challenges in access further complicate and increase the cost of large-scale survey work.

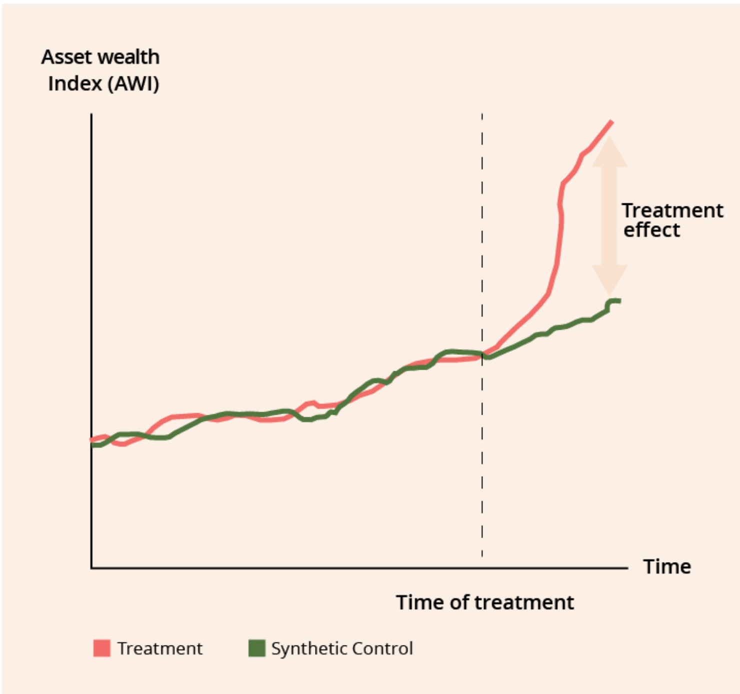
[18] Of course, this does not eliminate the fact that the location of the SNEL line is not entirely random. For example, it runs parallel to the SNEL electricity transmission line, which may be associated with non-random effects, such as better or more reliable electricity access.

3.2. The study approach

The approach developed through this study to address the challenge of evaluating telecommunications investments was to combine recent advancements in synthetic control impact evaluation design with the latest developments in AI-derived geospatial datasets. The SCM is a particularly useful approach for measuring impact where it is difficult to identify real-life, on-the-ground counterfactuals (or real-life control groups) that share sufficient similarities with the treatment group and are not impacted by the project over time (i.e., they become contaminated). Both challenges were present in this evaluation. The SCM differs from traditional (difference-in-difference) impact evaluation approaches in that it does not attempt to identify real-life control units on the ground or track their progress over time. Rather, it is based on a series of simulated control units, which are developed to best mimic the behaviour of the treated units in the years pre-treatment.[19]

The synthetic controls act as the (unobservable) counterfactual of what would have happened without the intervention. To date, they have typically been used in the evaluation of large policy decisions, which affect large treatment units.[20] They have not yet been used as frequently in other areas, such as the evaluation of infrastructure projects. In this study, the approach was tailored to be more relevant to the identification of impact in multiple treatment units, in a more localised investment. Because they were simulated, a further advantage of the method was that it could be tailored to the purposes of the study.[21]

Figure 3. Illustration of treatment effect (not based on actual data)



[19] These are developed as the weighted average of non-treated units across a series of predetermined metrics.

[20] A good example is the impact evaluation of the introduction of the California Tobacco Control Program. Abadie, A. *et al.* (2010) Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association* 105(490): 493–505.

[21] For example, confounding variables, which might influence the results of the study, can be stripped out during construction.

In this study, **synthetic control analysis was used to isolate the net impact** (as captured by settlements' changing asset wealth) of having access to the newly constructed backbone fibre line. The process of implementing the synthetic control approach is described in detail in Section 4. As illustrated in Figure 3, synthetic control units were developed to closely **match the behaviour of the settlements that have been** connected to the fibre line (our treatment settlements) in the period *before* they were connected (the time of treatment). We then used a statistical analysis to identify differences in the behaviour of the treated units and synthetic control units in the period *after* the time of treatment. This is the 'treatment effect' of the intervention. SCM was earlier applied to BII's Virunga Energies investment. The main difference with that study was that, here, we needed to identify treatment areas without data on who the customers were and where they were located. In other words, treatment areas in this study were more akin to catchment areas, based on proximity to ADMs that connect to the fibre backbone.

The data used in the geospatial impact evaluation design was fed through a machine learning model developed by AtlasAI to address any issues arising from missing or incomplete data. AtlasAI's large-scale proprietary datasets make use of daytime and nighttime satellite imagery, in combination with publicly available data and machine learning techniques, to develop comprehensive datasets covering key livelihood indicators. This technique closes gaps in the time series records of publicly available datasets and offers opportunities to customise indicators to better capture the impact of particular investments. This approach is, in part, inspired by recent work undertaken by Ratledge *et al.* (2022)[22] at the University of Stanford, which used similar datasets to estimate the impact of electricity grid access improvements on the rate of growth in village-level assets. This study built on learning from this work, which demonstrated how recent advancements in machine learning and satellite imagery can help ameliorate data gaps from traditional wealth indices, such as those from the Demographic and Health Surveys (DHS) Programme[23]. The process followed by AtlasAI to build its AWI and spending dataset from the available secondary data is outlined in Section 3.3 and Annex 4, including the steps taken to test and evaluate the accuracy of the model.

The use of geospatial data in the evaluation of investment projects is still in its infancy. This study, therefore, seeks to further 'expand the envelope' in terms of the methodological tools and data sources available to investors in this space. The approach holds a series of advantages for BII and other investors, including that the following: i) it relies primarily on remote sensing rather than on-the-ground data collection; ii) it does not place a significant burden on investees (in terms of either data or time to engage) – the primary data requested of Liquid is geotagged data on their installed infrastructure (e.g., the fibre line route and installed ADMs); and iii) it does not require the collection of baseline data but instead makes use of established geospatial records to enable researchers to 'go back in time' before the investment was made, which makes the approach much more flexible than traditional evaluation alternatives. However, the method currently has limitations, for example, with regard to the extent to which the impacts identified in

[22] Ratledge, N. *et al.* (2022) Using Machine Learning to Assess the Livelihood Impact of Electricity Access. *Nature* 611: 491–495. https://rdcu.be/cZOHV

[23] Funded by USAID, the future of the DHS Program is uncertain. International organizations, national governments and private organization may step in to fund data collection or provide an alternative program. Country surveys were conducted approximately once every five years

secondary datasets can be disaggregated to the household level or by gender. However, here, we can look to other studies to understand further how impact is generated.

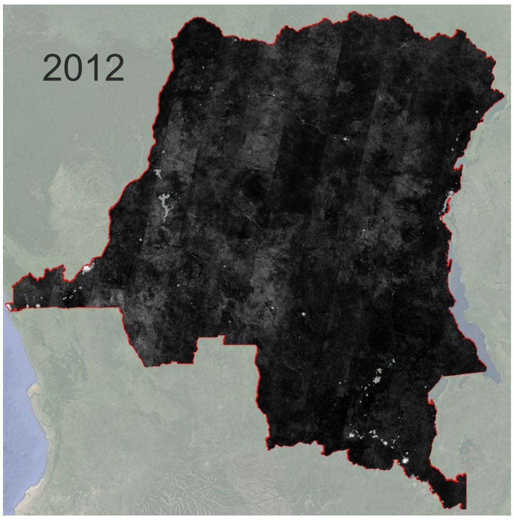
The solution developed in this study was designed to be replicable, scalable and cost-effective and, therefore, to offer an alternative to traditional impact assessment approaches, which are time-consuming and expensive to implement. The solution was intended to be particularly applicable to data-sparse and fragile contexts, where traditional approaches may not be feasible. This study has produced learnings for BII on how and in what circumstances it can replicate this study and apply similar methods to measure impact across its portfolio. As previously mentioned, the application here to ICT investments builds on an earlier application of the methodology for BII's investment in the Virunga mini-grid[24] which was peer-reviewed. An extension relative to that study is that the AWI results are translated into spending (in purchasing power parity) and, thereby, directly connect to BII's impact framework.

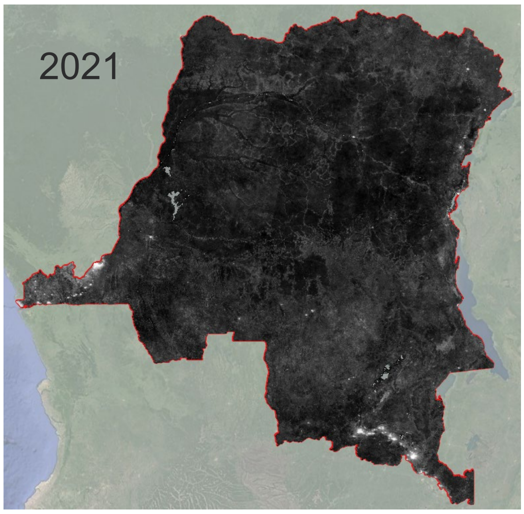
3.3. Our definition of impact – asset wealth

This study used AtlasAI's AWI to capture changes in standard of living as a result of living in an area with better and more reliable Internet access. Asset wealth was selected as a robust proxy for living standards in this study because it is based on multiple dimensions of wealth and is considered to be a more reliable measure of households' longer-term economic well-being than alternative monetary measures, such as spending. AtlasAI's AWI is based on data sourced from the many georeferenced, nationally representative surveys conducted in SSA, which collect data on asset wealth. A principal data source was the United States Agency for International Development's (USAID's) DHS. These data have been collected through representative household surveys for 30 years. Through these surveys, USAID calculates a household wealth index. A key advantage of this index is that it is less susceptible to errors in data collection than alternatives, given that many of the enumerated assets are directly observable to surveyors. The methodology AtlasAI used to construct the AWI is aligned to that used by USAID to construct the wealth index. It is calculated based on a household's ownership of selected assets, including televisions and bicycles, materials used for housing construction, and types of water access and sanitation facilities. Annex 4 provides additional detail on the assets included in the construction of the AWI. Figure 4 visualises AWI value across the DRC in two different years, with each pixel representing a 1 km x 1 km polygon. Kinshasa in the western part and the Lubumbashi (and nearby mining areas) in the southeast can be seen more clearly in 2021 than in 2012.

[24] British International Investment (2024) *How does access to green energy transform rural communities? Insights from Virunga Energies.* Available at: https://www.bii.co.uk/en/news-insight/insight/articles/how-does-access-to-green-energy-transform-rural-communities-insights-from-virunga-energies/?fl=true

Figure 4. Visualisation of AWI in 2012 and 2021. Lighter colours indicate higher AWI values.





In developing the AWI, AtlasAI used an AI model trained on satellite imagery to close data gaps in the secondary data record. Historically, a key challenge in using secondary datasets to measure the impact of investments is that they are updated infrequently and there are often gaps in the data record, especially in conflict-affected countries. The process developed by AtlasAI combines available secondary data on asset wealth with publicly available daytime and nighttime satellite imagery[25] to overcome this challenge. An AI model, based on a convolutional neural network (CNN), was trained to make predictions on asset wealth, drawing associations between satellite imagery and underlying secondary data on asset wealth.[26] The model's accuracy was tested against multiple datasets. In doing so, AtlasAI could use publicly available satellite data to close data gaps in the secondary data on asset wealth. The version of AtlasAI's AWI dataset[27] used in this study provided annual estimates of asset wealth for the period 2012–2024, at a resolution of 1 km x 1 km polygons. For more information on the AWI please refer to Annex A4.2.

3.4. Translating AWI to Spending

AtlasAI's AWI can be translated into household spending (in purchasing power parity, per capita). Spending provides an important bridge between a multidimensional measure of living standards and a more familiar, interpretable economic indicator. AWI captures a broad spectrum of household well-being, such as asset ownership, housing quality and access to essential services. Because AWI is grounded in more direct and observable data points and less on self-reported monetary measures, it is particularly well- suited for analysing differences between populations and over time. But its composite nature can make direct interpretation challenging, especially when comparing to standard benchmarks. By converting AWI values to an estimated household spending figure, the results become more readily comparable with widely used standards, such

[25] 'Earth Observation' datasets drawn from publicly available satellite image sources, with coverage over the last 25 years, including multispectral Landsat bands over multiple generations and satellite imagery of nighttime lights.

[26] The approach to training the AI model using publicly available satellite imagery is discussed by Yeh, C. *et al.* (2020) Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature Communications* 11(1): 2583.

[27] AWI is currently available for SSA and Southeast Asia.

as those employed within BII's impact framework, facilitating the easier communication of findings and alignment with established evaluation practices.

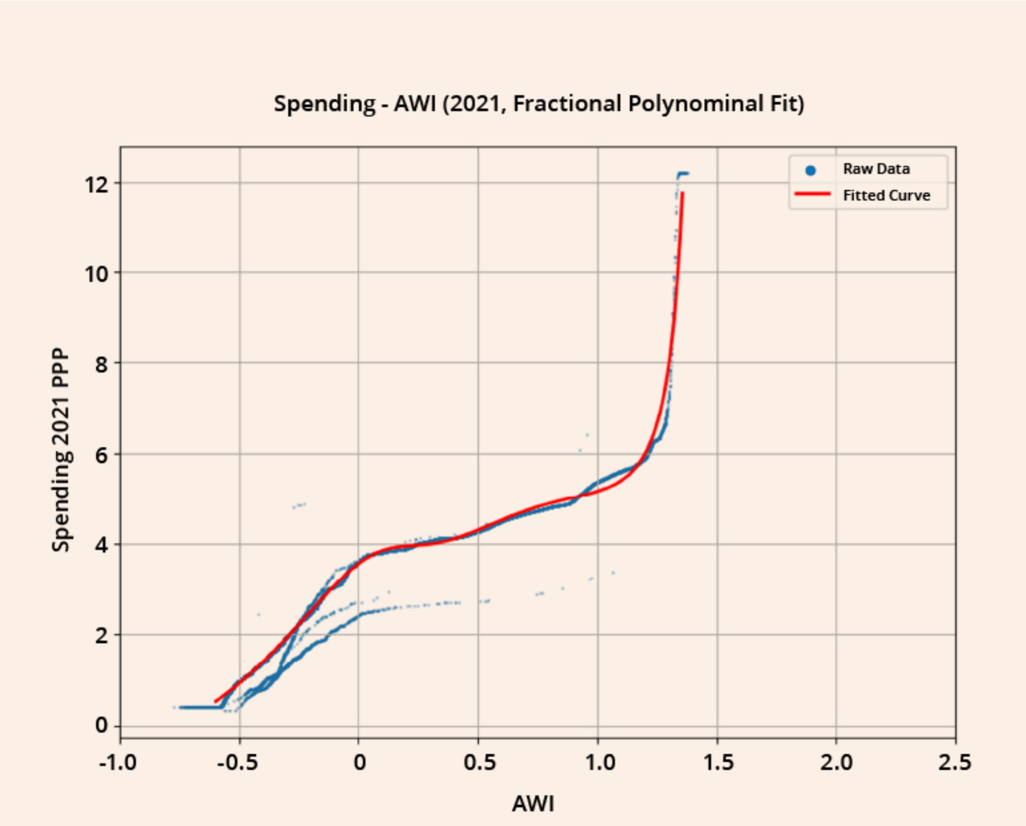
AtlasAI derived its spending layer by first using the same georeferenced household survey data and satellite imagery employed in the AWI to estimate asset wealth and then supplementing these sources with additional survey data that directly measure household expenditure. By establishing statistical relationships between asset ownership patterns and reported spending from these surveys, AtlasAI was able to translate AWI values into estimates of household spending in purchasing power parity terms.[28]

The relationship between AWI and household spending varies per country and per year, as is shown in Figure 5. While spending monotonically increases, it does so in a highly non-linear fashion. Below AWI values of -0.5 (spending US$1 PPP), changes in AWI cannot be meaningfully translated into changes in spending, whereas AWI values above 1.25 (spending US$6 PPP), small changes in AWI correlate with very large changes in spending. High AWI values are frequently observed in urban centres, where households have typically accumulated a significant array of assets and income increases do not necessarily translate into substantial rises in AWI, as households already possess most of the assets included in the index. As a result, additional assets acquired tend to yield diminishing marginal returns, meaning that the AWI becomes less sensitive to further improvements in household wealth at these higher levels.

The methodology developed in this study, thus, required that the AWI values of treatment and control sites were between -0.5 and 1.25, i.e., spending was between US$1 and US$6 PPP per capita per day. As will be shown below, this was indeed the case for all the DRC.

[28] For more information: https://docs.atlasai.co/economic%20well-being/spending/

Figure 5. Household spending (in 2021 US$ PPP per capita per day) vs. AWI for the DRC for 2017–2021. The fitted curve represents the best fit fractional polynomial for the year 2021.



3.5. Limitations

Although the approach used in the study has several notable benefits,[29] it does have some limitations.

- It is not currently possible to disaggregate impacts for different socioeconomic groups (including for men and women), given that the resolution of the AWI dataset is a maximum of 1 km x 1 km polygons. It is also not possible to 'look under the hood' to isolate the precise drivers of increased asset wealth due to access to broadband internet.
- The time between treatment (i.e. presence of fibre access in a settlement) and robust impact evidence in terms of AWI and spending is substantial – a minimum of four years as described in this report. Over such a long period, known and unknown confounding factors may become important.
- Above per US$ 6.00 (2021 PPP) per capita spending in DRC, the relationship with AWI breaks down, which potentially limits the approach in some areas. Given the spending distribution in DRC this does not seem to pose much of an issue in DRC. However, in other countries this may be a limitation, because the relationship between AWI and spending is country dependent.

[29] Including the completeness of the AI-derived AWI dataset, the ability to examine impacts retrospectively, and the fact that it does not require the collection of primary data, as discussed elsewhere in this report.

- The AWI dataset relies in part of DHS surveys, the funding for which has become uncertain due to recent budget cuts in USAID. New sources of funding or alternative data sources may or may not become available in the future, which poses some questions on the future continuity of the AWI dataset.

4. Evaluation methodology

This section provides a step-by-step explanation of how the approach was defined in practice and implemented. The methodology is broken down into three key steps:

- Step 1: Identification of geotagged data on new connections and new infrastructure to map where and when the project was rolled out over time. This enabled the treated location to be identified, selected, and included in the study.
- Step 2: Identification of non-treated locations in the same province that shared sufficient similarities with the treated locations to form the basis of synthetic controls.
- Step 3: Comparisons of asset wealth accumulation over time between the treated and synthetic control units are then used to identify the net effects of the intervention.

Each step is discussed in turn, explaining the design choices made and the sub steps and actions in each part of the process.

4.1. Step 1: Identifying and selecting treated locations

4.1.1. Data identification, entry and cleaning

Two principal datasets were identified and used to identify treatment settlements:

1. Time series satellite imagery and AtlasAI proprietary datasets.
2. Geotagged data on Liquid Technologies installed equipment along the fibre line.

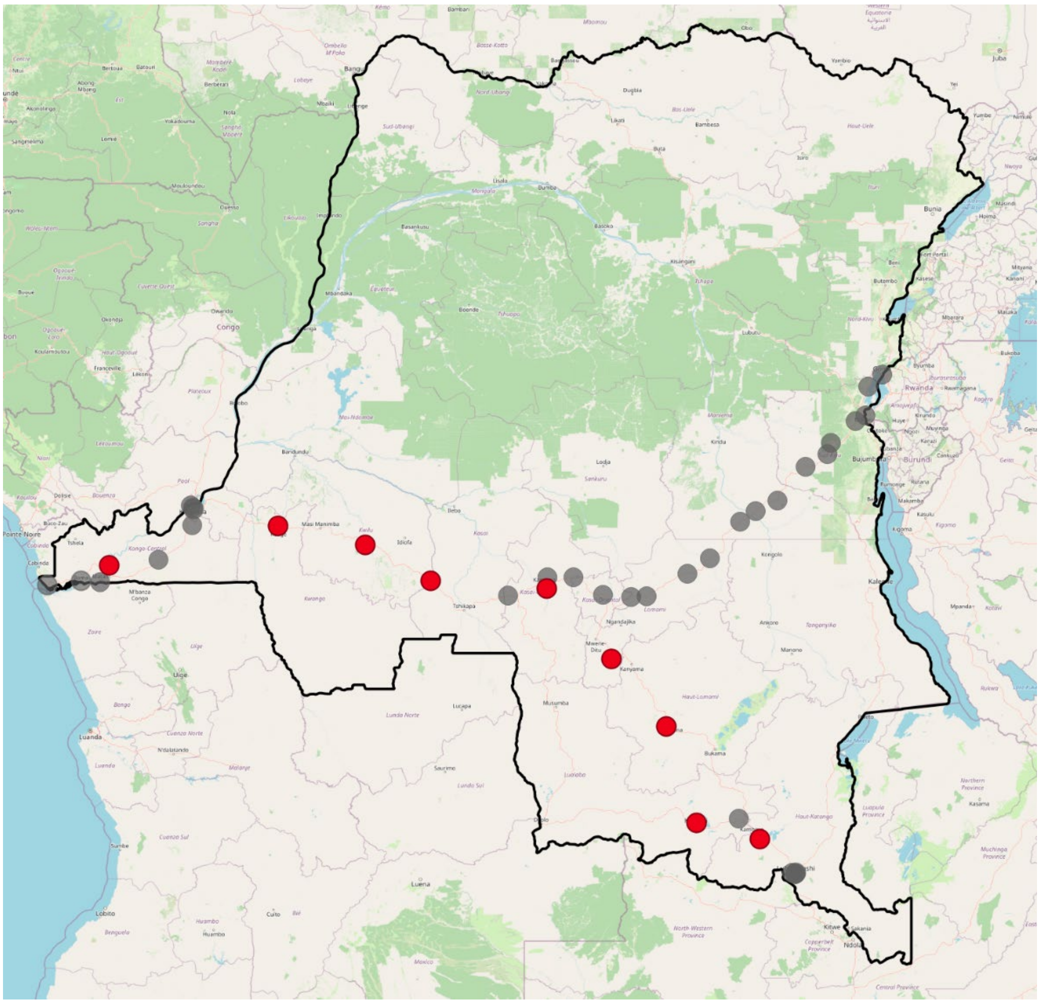
The AtlasAI proprietary dataset[30] used to identify treated and non-treated settlements in this study is the Human Settlement layer. The analysis period is 2012–2024, and the outcomes of interest (asset wealth and spending) vary annually. Further information on how these datasets were used is outlined below. AtlasAI's databases provide a high degree of customisability. The ability to adjust and fine-tune the data points used during the study offers a degree of flexibility which is usually not available in a traditional impact evaluation design once on-the-ground data collection has commenced.

Regarding the second dataset, geotagged data on infrastructure locations enables us to track where service has been provided throughout the DRC. We know that the SNEL line became available in March 2021, and the Kananga-Goma line became operational in May 2023. As per June 2025, a total of 37 nodes (both ADMs and intermediate line amplifier (ILA) locations) were present along both fibre lines, as shown in Figure 6. Of these, 14 were ADMs[31] installed during 2021. Exclusion of the ADMs in Kinshasa and Lubumbashi, which were intentionally connected and combination of ADMs that were within 5km of each other, resulted in 9 'treatment' towns. The arrival of fast internet through broadband fibre in these towns was an exogenous event (i.e. connection was accidental) as opposed to endogenous (connection was intentional).

[30] Available in 1 km x 1 km polygons.

[31] ILAs were excluded because they do not allow for adding or dropping internet traffic.

Figure 6. Location of add/drop locations considered in this study. Red dots indicate treatment towns and are dots are dropped either because they are not ADMs or because they were installed post 2021.



4.1.2. Defining the treatment areas

Unlike its MNO clients, Liquid has no line of sight to the customers who use the Internet. In addition, because most people access the Internet through mobile devices, often through prepaid cards, MNOs do not know where their customers live.[32] In order to define the areas that have access to faster Internet, one has to define catchment areas based on the location of ADMs. This situation is markedly different compared to when geotagged information is available for customers, in which case the treatment area or group can be defined exactly.[33] Below we discuss how treatment areas have been defined using data from AtlasAI's Human Settlement layer.

AtlasAI's Human Settlement layer fuses a range of input data sources at varying spatial resolutions to detect human settlements.[34] In this way, settlement areas have been identified around the Liquid ADM locations. The coverage ratio of a mobile telecommunications tower in

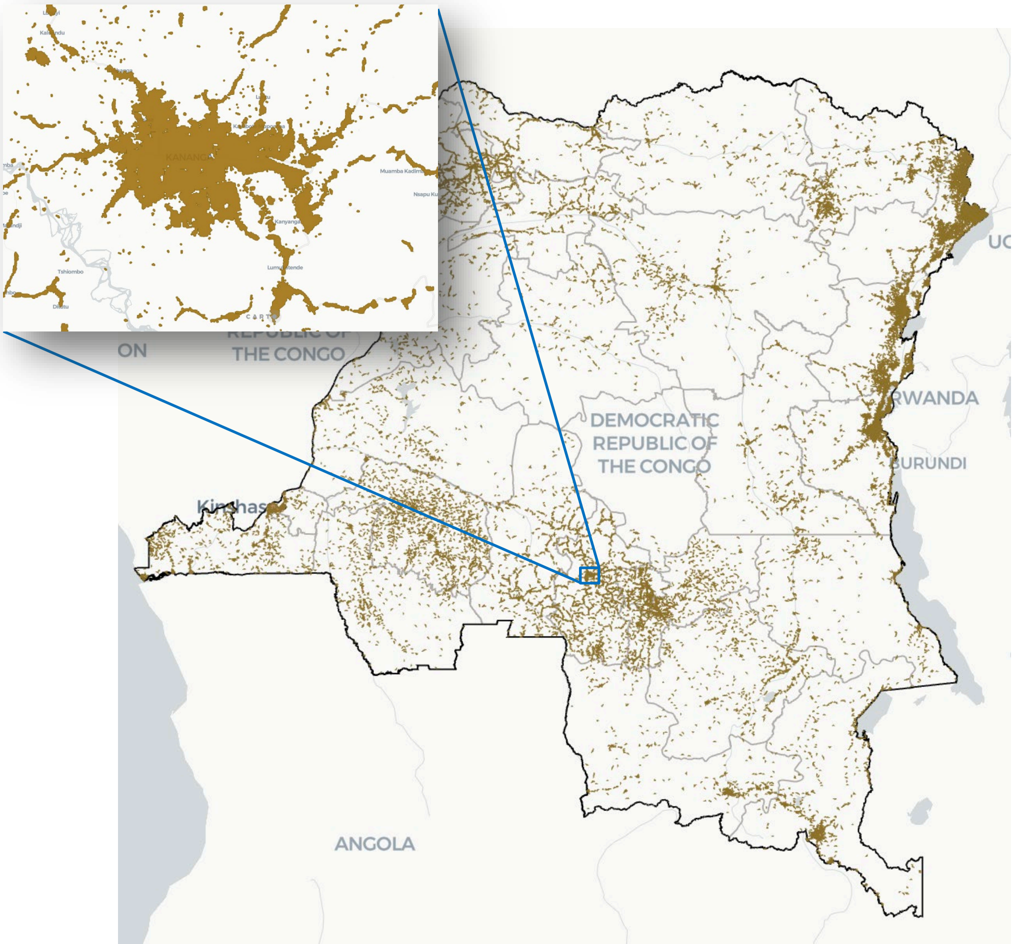
[32] Through triangulation between different antennas, MNOs can determine from where customers access the Internet.

[33] A good example is Virunga Energies ([How does access to green energy transform rural communities? Insights from Virunga Energies - British International Investment]), where access to electricity was known at the level of individual (geotagged) households.

[34] For more information on this, see [https://www.atlasai.co/blog/atlas-of-human-settlements]

the 4G band ranges from 3 km to 6.5 km,[35] depending on atmospheric conditions and the density of the built environment. The higher the bandwidth of the signal, the higher the signal's frequency and the lower the range of a mobile tower. Based on this, a range of 10 km around the ADM is a defensible catchment area. In such a 314 km^2 area[36] around an ADM, typically there are many settlements. By defining the actual treatment area as the most populous settlement within that radius[37], we assure that the characteristic treatment areas are most comparable. This is because these settlements are likely to be entirely urban, whereas the full circular catchment areas can have very different urban/rural distributions. This is illustrated in Figure 7.

Figure 7. All settlement areas in the DRC and an example of a settlement area around a Liquid add/drop location



4.1.3. People reached in the treatment areas

According to AtlasAI's Human Settlement layer, 3,131,622 people live in the nine treatment areas. This is 2.8% of the approximately 109.3 million population of the DRC. In this section we estimate the number of people that can be expected to benefit from faster and more reliable Internet.

---

[35] Source: https://dgtlinfra.com/cell-tower-range-how-far-reach/

[36] Area = 3.14 x 10^2 = 314 km^2

[37] When the most populous settlement extends outside of the radius, the settlement population outside the radius is included

Because the treatment areas are entirely urban, and because there are differences between males and females' use of the Internet, we first break down the entire population into four quadrants, based on the urban/rural split of 47.1%/52.9% and the male/female split of 49.6%/50.4%[38], as shown in Table 2.

Table 2. Breakdown of the DRC's population[39]

DRC Population	Urban	Rural	*Total*
Male	25,534,229	28,678,571	***54,212,800***
Female	25,946,071	29,141,129	***55,087,200***
Total	***51,480,300***	***57,819,700***	***109,300,000***

Per 2024 data, the World Bank estimates that Internet penetration in the DRC was 31%, or 33.8 million people. GSMA reports that in Sub Saharan Africa, people in rural areas are 54% less likely to use the (mobile) Internet than urban users. Using the population breakdown in Table 2, this means penetration rates of 38.1% in urban areas and 24.7% in rural areas. Because all treatment areas are in urban locations, we estimate that there are 1,191,891 urban internet users in the treatment settlements that potentially benefit from faster internet. But there are also internet users outside the treatment areas, and the number is therefore likely higher.

According to the GSMA, female users are 37% less likely than men to use the Internet. Combining this with the number of people who live in the urban treatment areas allows disaggregation of all people that benefit from faster internet into 726,691 men and 465,200 women. This is about 3.5% of all 33.8 million Internet users and 6.0% of all urban Internet users in the DRC.[40] These numbers will increase as internet penetration rates continue to go up in the treated settlements, and more fibre access points will be added. Most of the affected people and companies had rudimentary internet access before the SNEL line became operational, but the low speeds and high latency would not have allowed them the full range of services, such as streaming, working from home, and accessing government, education and other services. With the SNEL line operational, MNOs can invest in 4G and 5G infrastructure.[41]

In Africa, 4G antenna towers are typically connected through a dedicated backhaul fibre line to the backbone fibre line when the distance is less than 10km. For longer distances, microwave backhaul can be used up to distances of about 20km. By counting the urban and rural and population that live within these distances (i.e. in strips of 20km and 40km around the SNEL line) and by applying the above-mentioned urban and rural internet penetration rates, the total number of users that can potentially benefit from better quality internet are estimated in Table 3.

---

[38] According to the AtlasAI data, women make up 50.6% of the population in the treatment settlements. For sake of consistency and simplicity, we here use country average data on the urban/rural and male/female splits.

[39] Steward Redqueen analysis based on World Bank and Datareportal data. In this estimation there are four unknowns and three degrees of freedom in the data, meaning there are infinite solutions. To close the system, we impose that the distortion of proportionality on both dimensions is minimised.

[40] In all likelihood, these numbers will be higher – there are people outside of the treatment areas but inside the catchment radius of the add/drop points – but here we focus only on the largest settlements.

[41] Liquid offers shared connections to end customers as well, but that was outside the scope of this study.

Table 3. Urban and rural population and internet users that live within 10km and 20km from the SNEL line[42].

	Urban	Rural	*Total*
Population <10 km	4,063,519	3,096,016	**7,159,535**
Population <20 km	5,238,496	4,835,534	**10,074,030**
Internet users <10 km	1,546,569	382,222	**1,928,892**
Internet users <20 km	1,993,764	492,743	**2,486,507**

By assuming that the 10km and 20km distances constitute lower and upper estimates, the above-mentioned 1,191,891 internet users in the treatment settlements make up 60%-77% of the urban internet users and 48%-62% of all internet users along the SNEL line. Connecting all these people to high-quality internet will require more ADMs to be installed along the SNEL line, more towers to be installed by MNOs, and capable handsets to be bought by consumers. As both the total population and the internet penetration continue to increase, it is reasonable to assume that the total number of people that will benefit from the SNEL line will be at least double the 1,.2 million people in the treated settlements, say 2.5 million.

4.1.4. AWI and household spending characteristics of people in treatment areas

As shown in Figure 8, the treated (urban) settlements have considerably higher AWI value than the DRC's average. In fact, the AWI value of the treated settlements is higher than 98% of all settlements in the DRC. Translation of these AWI values into spending, using the relationship shown in Figure 5 results in a mean spending of US$3.38 PPP per capita per day. This is higher than 91% of the DRC's population, which spend on average US$2.25; As per Figure 9, the treatment sites are above the World Bank's international poverty line of US$3.00 but well below the US$4.20 poverty line for lower-middle income countries. We also note that substantially all settlements in the DRC are below the US$8.30 (or US$6.85 in 2011 PPP) World Bank upper-middle income threshold.

[42] The large cities like Kinshasa and Lubumbashi are not included because they were not 'accidentally' connected as explained before.

Figure 8. 2020 AWI for the treated donor pool settlements.

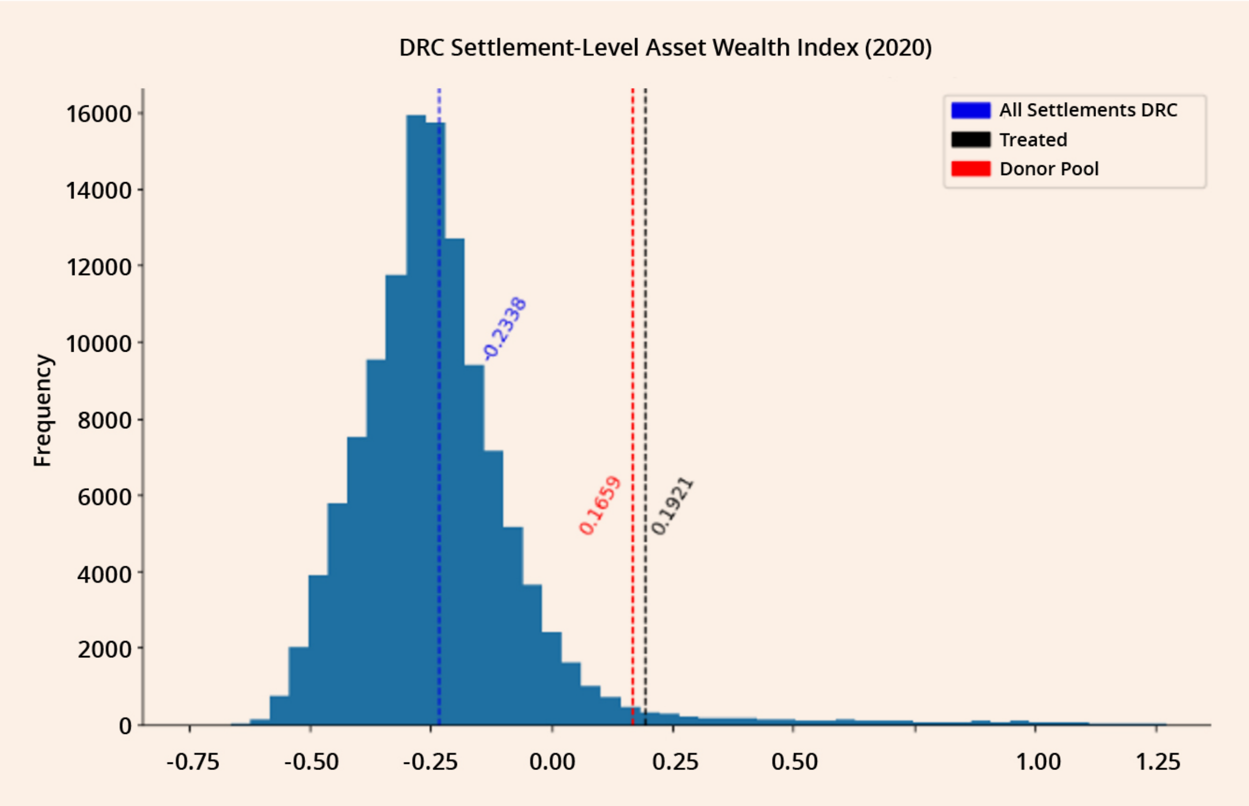
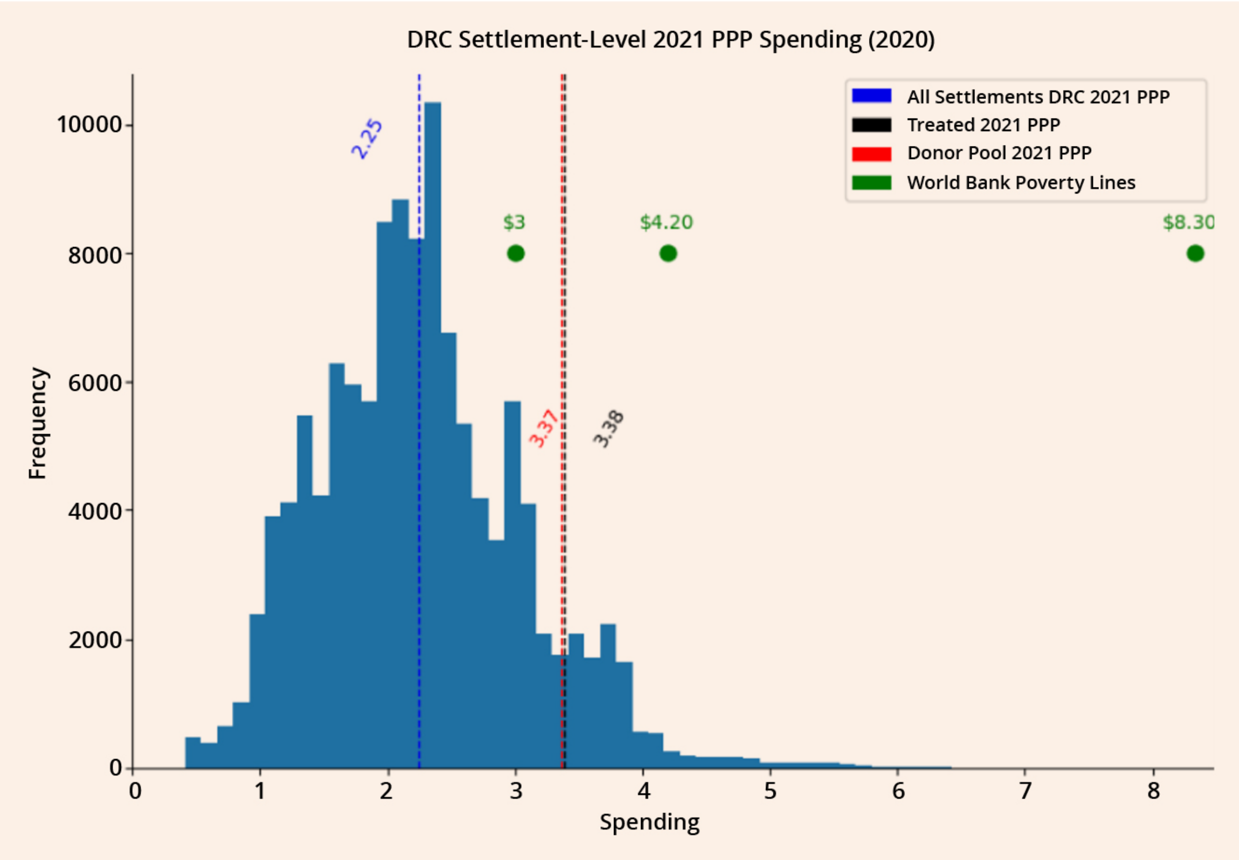


Figure 9. 2020 Household spending for the treated donor pool settlements in US$ 2021 PPP. Also indicated are the three World Bank poverty thresholds



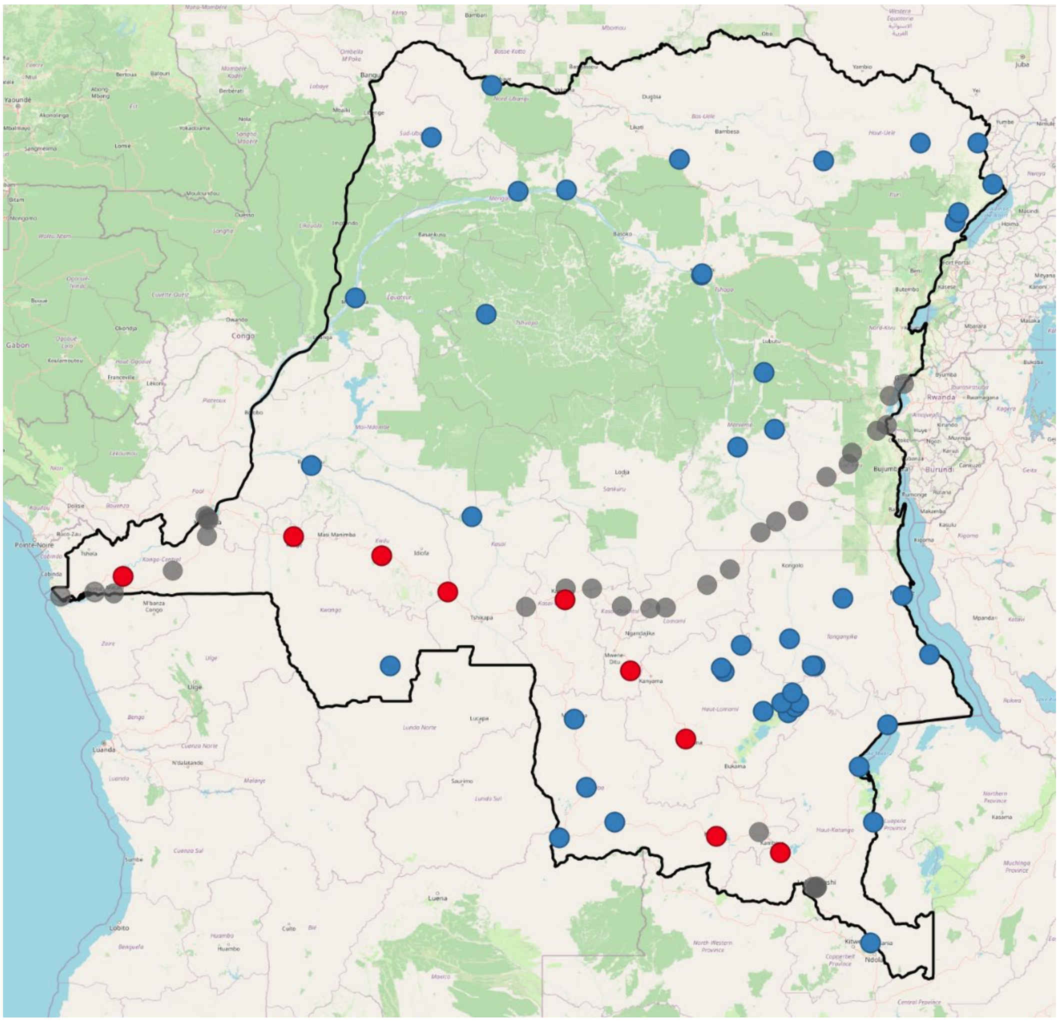
4.2. Step 2: Identification of a 'donor pool' of non-treated settlements

SCM requires a pool of donor units from which for each treatment a selection will be made which will serve as the unobserved counterfactual after treatment. Out of a candidate pool of over 120,000 urban settlements in the DRC, we selected 58 candidates in the donor pool of non-treated settlements across all the DRC, based on the following criteria:

- Only settlements with a population over 50,000 were considered
- Cosine similarity testing[43] on population, AWI (2012 and 2020), and spending (2012 and 2020) levels and trends

The 58 candidate settlements in the donor pool are shown in Figure 10 and the demographic and spending characteristics of the nine treatment and the 58 donor units are summarized in Table 4 and in Figure 8 and Figure 9.

Figure 10. Donor pool of 58 settlements (blue) superimposed on Figure 6



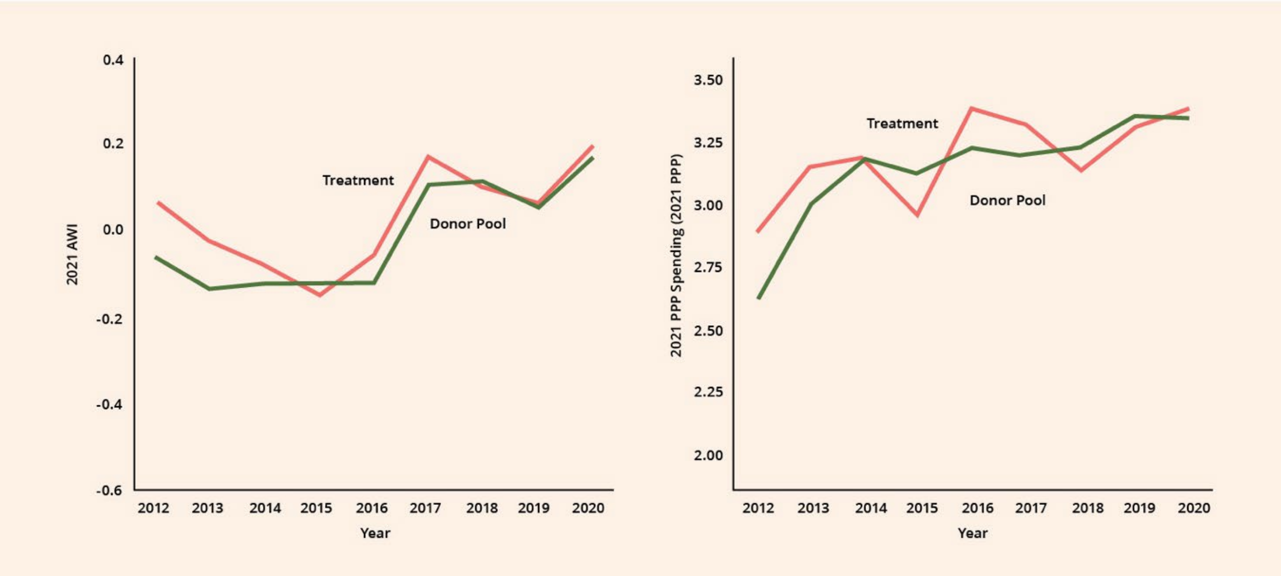
[43] Cosine similarity testing is a simple way to see how alike two sets of data are. In mathematical terms, it measures the cosine of the angle between two multi-dimensional vectors.

Table 4. Summary statistics of settlements in the treated settlements and all settlements in the DRC

	9 Treatment sites	58 Donor sites
Population (2020)		
Average	335,984	141,039
Median	210,013	79,911
Mean spending levels (2020)		
2021 US$ PPP per day per capita	3.38	3.37
2011 US$ PPP per day per capita	2.00	2.08

Figure 11 shows how both the average AWI and spending levels are very similar for the nine treatment sites and 58 donor pool, both in terms of values and evolution until the last full pre-treatment year (2020). This means that the cosine similarity search has yielded a suitable pool of donor units that will enable the SCM to construct a realistic synthetic control for each of the treatment sites.

Figure 11. Evolution of AWI and household spending for the candidate pool and the selected units from the donor pool

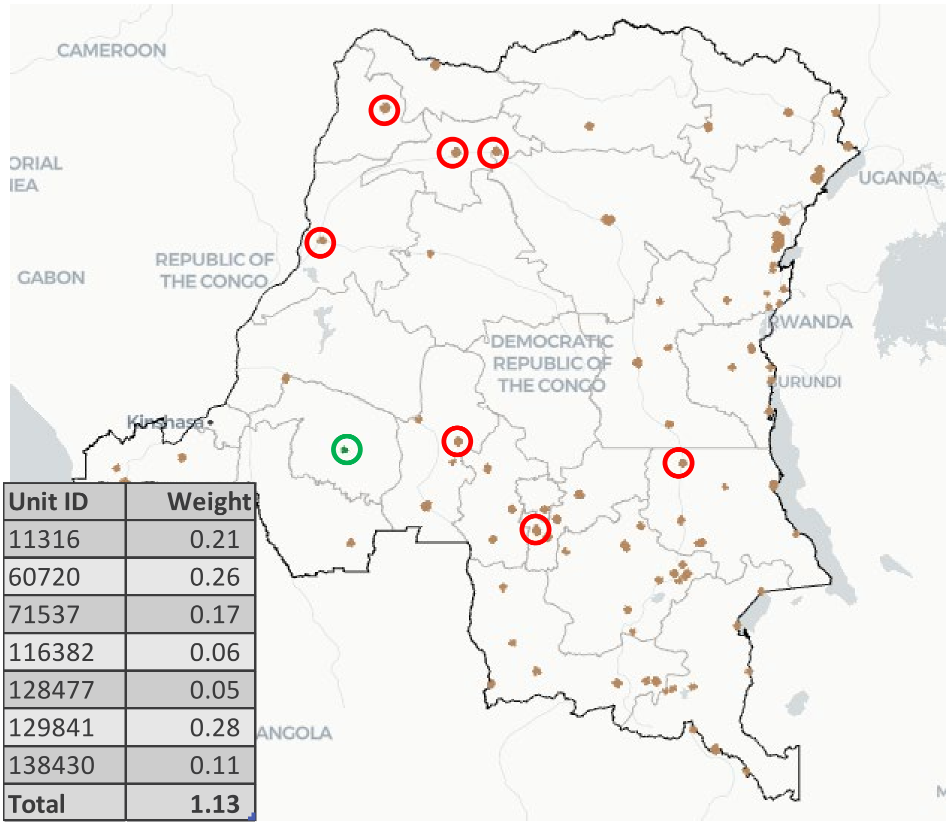


4.3. Step 3: Conducting synthetic control analysis

4.3.1. Illustration of SCM algorithm using a specific example

The synthetic control comprises a weighted average of some settlements in the donor pool. In the SCM with elastic net approach, we iteratively calculate the synthetic control for each treated unit. Figure 12 shows that the synthetic control for the treated settlement near the city Kikwit is composed of a weighted average of seven settlements in the donor pool. The selection of the units and the weights assigned to each of them were determined by the SCM algorithm. The weighted average AWI behaviour of the seven units most closely resembles that of the treated unit before Liquid's fibre backbone treatment in 2021 and serves as the unobserved counterfactual after treatment. The weights sum to a value slightly greater than 1. The reason for this is that while on average the treatment and donor settlements are very similar in terms of AWI and spending, this may not be the case for each individual treated settlement.

Figure 12. Treated settlement (green) and the non-treated settlements that constitute the synthetic control (red)



Unit ID	Weight
11316	0.21
60720	0.26
71537	0.17
116382	0.06
128477	0.05
129841	0.28
138430	0.11
Total	**1.13**

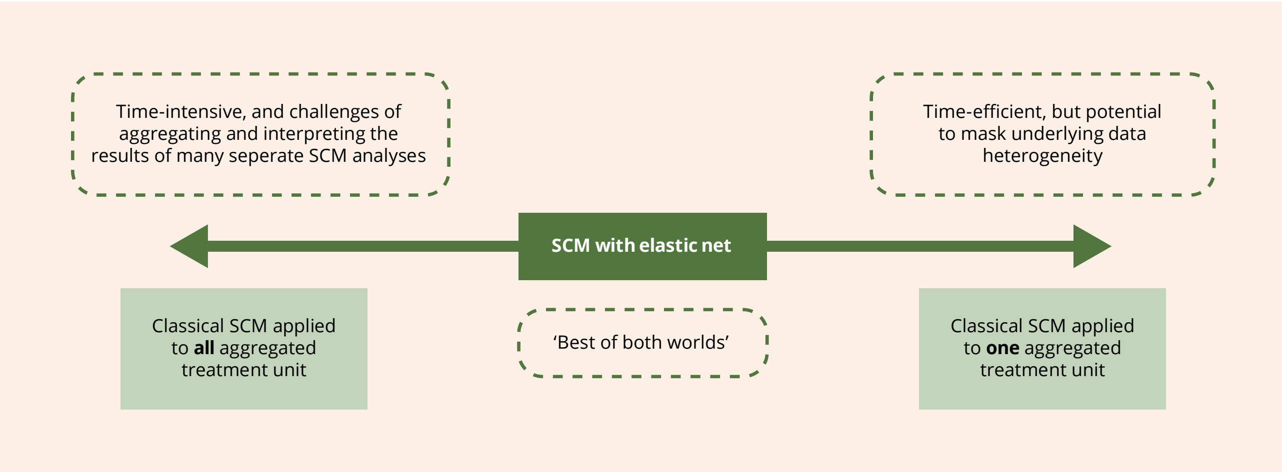
4.3.2. Synthetic control method with elastic net

In the description above we used the so-called elastic net SCM method. In this section we provide some more background on the different ways in which SCM can be performed.

The classical SCM approach was designed primarily to estimate the effects of large aggregate interventions focused on a small number of large treatment units (typically one) where data is available (for both dependent and independent variables) over a longer time horizon. This study applies this approach in a slightly different context in which there are multiple treated and non-treated units and where data is available over a relatively short time horizon (both pre- and post-treatment). In this context, the data could have been aggregated before the analysis stage, and the classical SCM approach could have been applied across all the treated and non-treated pool units as one, but this would have risked masking underlying data heterogeneity. Alternatively, classical SCM analyses could have been conducted for all individual treated and non-treated units separately, but this would have been highly resource-intensive and would have risked overcomplicating the analysis stage.

The elastic net SCM approach which was deployed in this study offers a more relaxed definition, with a regularisation function that helps the model to make generalisations when predicting the counterfactual for post-treatment years but avoids overfitting. This is helpful in situations (such as this one) where the data record is not extensive and there are residual differences between treatment units and the donor pool.[44] It is a variation which has been used frequently in the literature.[45] As illustrated in Figure 13, in this context SCM with elastic net offers the 'best of both worlds' in terms of striking a pragmatic balance between applying the classical SCM approach across all treated units separately and running a single classical SCM across all treated units as one.

Figure 13. Different versions of the SCM algorithm



Annex 3 provides a more detailed comparison between the classical SCM approach and SCM with elastic net and provides more detail on the specific technical definition of 'SCM with elastic net' used in this study, including some of the implications of relaxing the constraints of the classical SCM used.

In practice, the version of SCM with elastic net used in the analysis was defined as follows:

1. Weights of donor units are not restricted to sum to 1
2. Weights of donor units cannot be negative (which forces a smaller number of donor units to be selected)
3. No intercepts are allowed (an intercept allows for a level shift where the trend remains the same).

This is a variation on the standard SCM with elastic net. The standard SCM definition typically also relaxes the classical SCM approach to allow weights to be negative (which is a further advantage when dealing with data over shorter time horizons). We felt that this further relaxation of the model was not necessary in this case, given the selection procedure to identify a donor pool which is broadly similar to treated settlements pre-treatment (see Section 4.2)..

This model predicts the counterfactual by weighting each post-period control variable. These weights are determined through panel-like regression within the pre-treatment period, in which we are regressing a single treated unit on the full panel of control units. We did not include any

[44] Based on the use of an elastic net drawing on a combination of lasso and ridge penalties.

[45] Ratledge, N. *et al.* (2021) Using Satellite Imagery and Machine Learning to Estimate the Livelihood Impact of Electricity Access. https://doi.org/10.3386/w29237

covariates in the model, and we used an alpha of 0.5 (elastic net), which is a mix of lasso and ridge penalties. More information on penalty factor can be found in Annex 3.3.

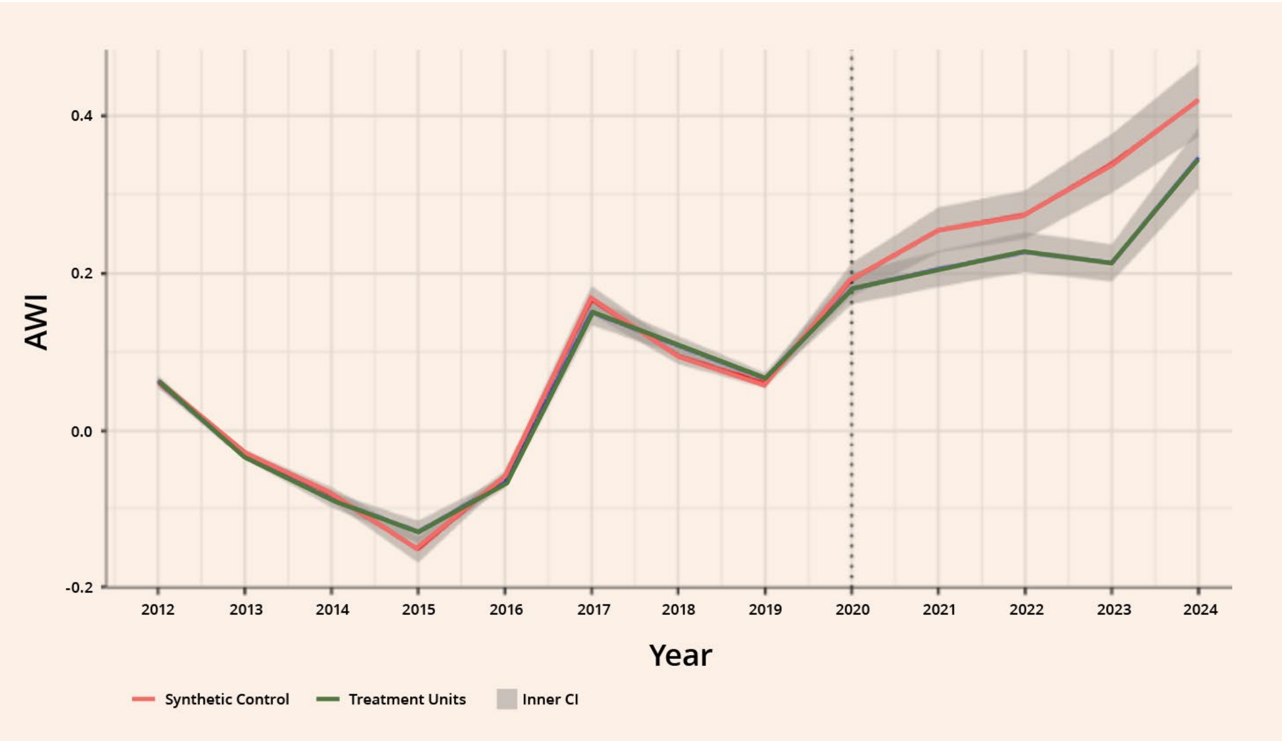
5. Key findings

The nine treatment sites (see Figure 6) received access to fast internet at different dates in 2021. Because the AWI and spending data are available on a yearly basis, we consider 1 January 2021 to be the start of the treatment period.[46]

5.1. Change in AWI for connected settlements

The rise in asset wealth observed in connected settlements suggests a positive impact attributable to the treatment. Figure 14 shows the results of the AWI evolution of the aggregated nine treated settlements. As intended, before treatment, the Synthetic Control behaves very similar[47] to treated settlements and demonstrates is validity as a counterfactual against which to assess post-treatment effects. In terms of the AWI histogram in Figure 8, the treated settlements have moved from the 98th percentile in 2020 to the 99th percentile in 2024, whereas the synthetic control has moved from the 97th to the 98th percentile.

Figure 14. AWI of the Synthetic Control and treatment units before and after the arrival of fast internet in 2021



From the start of the treatment period, the AWI of treatment units increases faster than that of the synthetic control. While this points to a causal effect due to the arrival of fast internet, one must be cautious for three reasons:

[46] In the graphs the start of the treatment period is shown at 2020 because that indicates 31 December 2020.

[47] The match is not identical meaning that the SCM does not overfit (or overly customizes). This is a good indication that the SCM distills genuine trends instead of capturing idiosyncratic patterns that do not persist after the intervention.

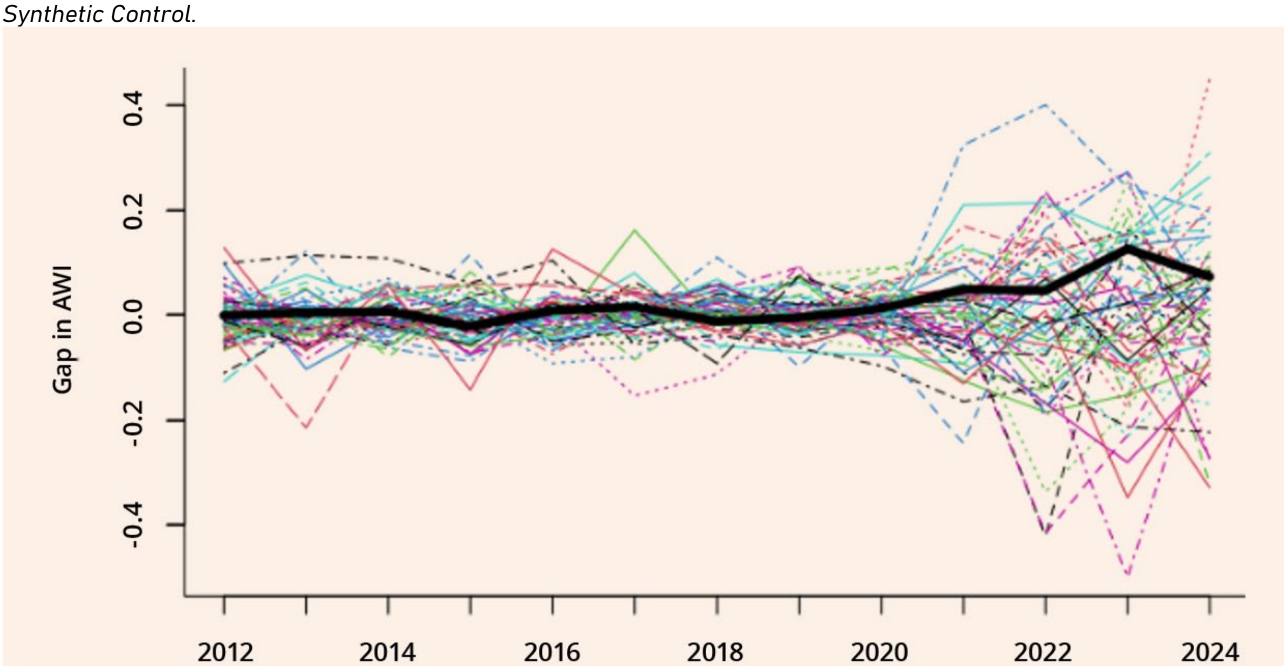
1. The possible presence of unobserved confounding factors that are not captured in the pre-treatment period. These confounders may influence the treated and the synthetic control units differently.
2. Idiosyncratic post-treatment shocks or events in the southern DRC region. For example, the first years after the arrival of fast internet in this region in 2021 coincide with the Covid 19 pandemic and the subsequent recovery. It is possible that regional differences in the pandemic's severity and the strength of the subsequent recovery can be a source of bias.
3. The fact that the confidence intervals of the treatment units and the SC overlap in 2024.

The last point can be analysed more precisely by determining the probability that the treatment result could have been obtained by chance. Every donor pool unit can be viewed as having undergone a placebo treatment[48]. By examining the gap between the AWI trajectories of each of the 58 placebos in the donor pool and the treatment units, one can count how many exhibit a treatment effect stronger than that observed in the treatment units. The number of placebos that show a larger effect than the treatment group divided by all 58 placebos is often referred to as a pseudo p-value. The results of placebo testing are depicted in Figure 15 and show that following the pre-treatment period[49], each donor unit's outcome is influenced by its unique set of post-treatment shocks and circumstances, leading to greater divergence in their trajectories compared to the pre-treatment period. Although the pseudo p-value post treatment is trending lower than the pre-treatment value of 0.48, it also means that the probability that observed treatment effect could have come about randomly is 0.32, which is much higher than 0.05 which is typically seen as statistically significant.

[48] Abadie, A., A. Diamond and J. Hainmueller, *Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco control program,* NBER Working Papers Series 12831, 2007.

[49] One expects before treatment the pseudo p-value to be around 0.5 on average, which is an indication that the average of the donor pool behaves very similar to the treatment group (i.e. an equal number of donor units above and below the treatment group)

Figure 15. AWI gap between of each of the donor units (dotted line), the treatment group (bold solid line) and the Synthetic Control.



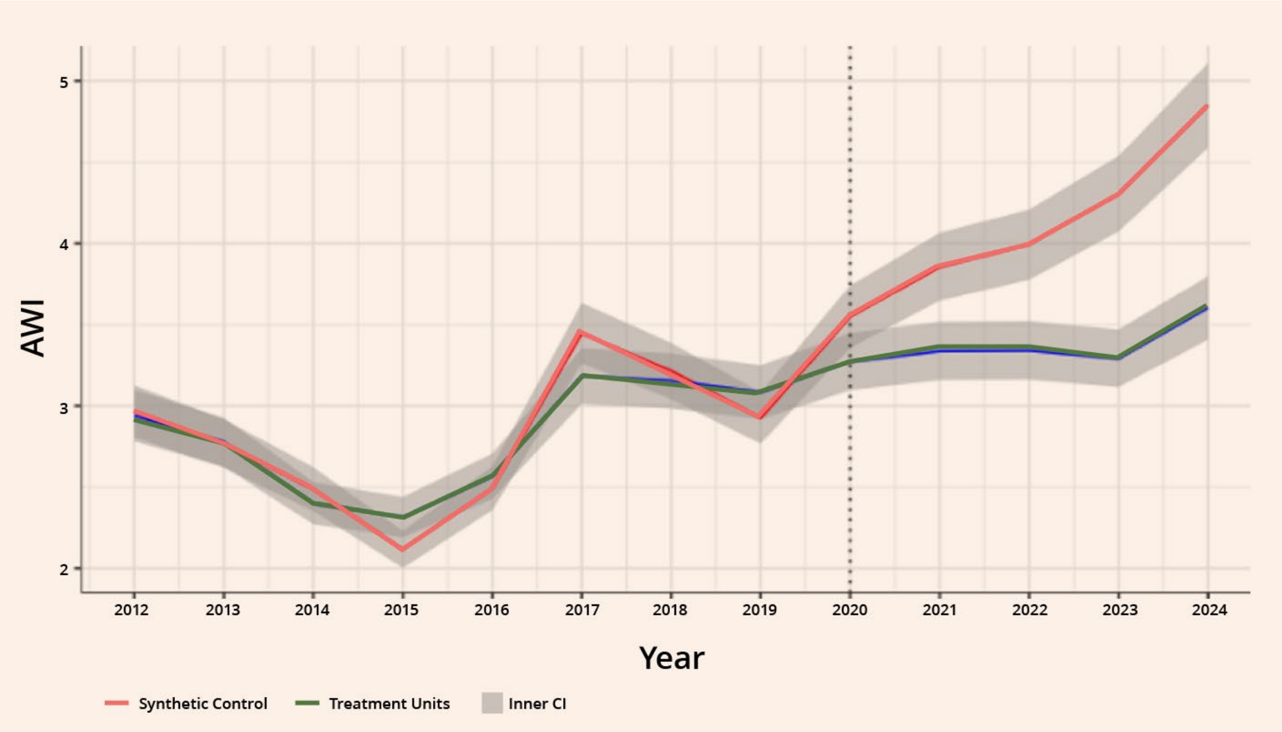
pseudo p-value	0.43	0.31	0.43	0.66	0.40	0.33	0.74	0.53	0.47	0.22	0.40	0.29	0.36

Greater robustness of the results primarily requires a longer post-treatment period. Many synthetic control studies[48] have shown that as time passes the cumulative gap between treated units and the synthetic control becomes more pronounced. Moreover, random placebo effects dissipate over time, which further decreases the pseudo p-value. Given that the SCM has been set up, repeating the analysis in the future can be done at minimal cost.

5.2. Change in spending for connected settlements

The increased spending observed also points to a positive impact attributable to the treatment. The AWI results can be translated into spending by using the relationship shown in Figure 5. This translation has been done individually for all 1x1km pixels that constitute the treatment and Synthetic Control units. The results shown in Figure 16. In terms of the spending histogram in Figure 9, the treated settlements have moved from the 91st percentile in 2020 to the 98th percentile in 2024, whereas the synthetic control has moved from the 89th to the 94th percentile.

Figure 16. Average household spending (in 2021 US$ PPP per capita per day) of the Synthetic Control and treatment units before and after the arrival of fast internet in 2021.

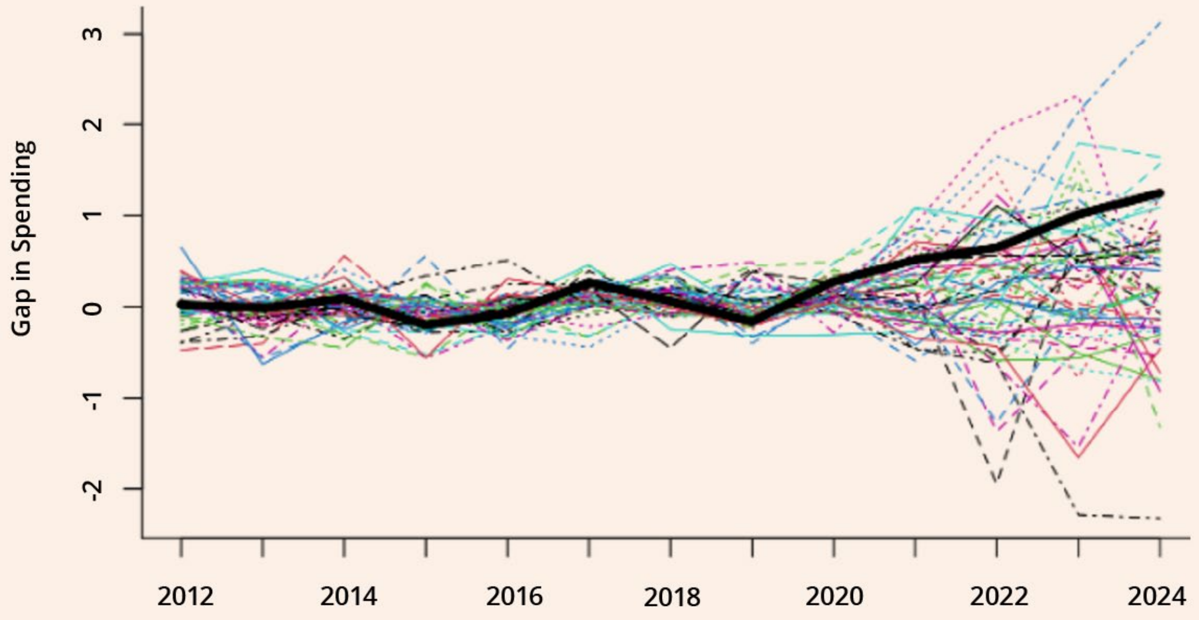


The spending results mirror the trends observed in the AWI results, but in a more pronounced way: in the treated settlements, spending went up from by US$ 1.25 to US$ 4.80 per capita day in 2021 PPP terms (+33%), compared with an increase of US$ 0.30 to US$ 3.60 (+9%) in the Synthetic Control. In other words, spending of households in the settlements that were connected to the fibre line increased by US$ 0.95 per capita per day more compared to households in settlements that were not. When taking the confidence intervals into account, the findings for spending appear to be more robust, lending further weight to the positive impact attributed to the arrival of fast internet.

The results of placebo robustness testing are shown in Figure 17. The number of placebo units with a larger impact than the treatment units is much smaller compared to AWI. Importantly, the pseudo p-value is trending down from 0.24 after treatment to 0.07 in 2024. It is important to note, however, that since spending is derived from AWI, the additional assumptions and propagating errors in the derivation make spending a less stable dataset than AWI. Consequently, the seemingly greater robustness of the spending results should be interpreted with caution, as it may overstate the true reliability of these findings.[50]

[50] The spending results are derived for all individual 1km x 1km pixels. Because of the nonlinear relationship between AWI and spending shown in Figure 5, inter-settlement heterogeneity can artificially amplify the strength of the impact signal.

Figure 17. Spending gap between of each of the donor units (dotted line), the treatment group (bold solid line) and the Synthetic Control.



pseudo p-value	0.62	0.55	0.36	0.74	0.66	0.12	0.50	0.83	0.16	0.24	0.24	0.16	0.07

6. Findings and recommendations

This section outlines the key conclusions and the recommended next steps for BII and the FCDO. The key conclusions are framed according to the original two key study purposes, focusing on i) the extent to which the study has been able to identify evidence of the impact of BII's investment in Liquid Technologies and its DRC broadband fibre backbone and, in doing so, fill a strategic evidence gap for BII in the infrastructure portfolio and ii) whether this study has been able to demonstrate proof of concept for an innovative, low-cost and rigorous approach to assessing impact in infrastructure projects.

6.1. Findings

Liquid's SNEL broadband fibre cable is providing an estimated 2.5 million internet users with access to faster and more reliable internet.

For reasons of comparability with settlements that do not benefit from fast internet, this study focused on nine urban settlements in which MNOs can connect their networks to Liquid's fibre cable. In these settlements, based on population statistics and World Bank and GSMA estimates of internet penetration among urban/rural and male/female populations in DRC, we estimated that 726,691 male and 465,200 female, i.e., 1,191,891 internet users, would potentially[51] have access to faster and more reliable internet. The total number of users that benefit from the fibre line is likely twice as large, i.e. approximately 2.5 million, because these numbers do not include people outside of the nine urban areas considered here. Moreover, these numbers will increase as internet penetration rates continue to go up and as more settlements along the SNEL line (or further away from it) are connected through network investments by MNOs.

Although well below US$4.20 lower-middle income threshold of the World Bank, the studied population live in urban areas and are wealthier than most of the DRC's population

In terms of average AWI, the households in the treatment areas are wealthier than 98% of the entire DRC population. Their average spending per capita in 2020 of US$3.38 (in 2021 PPP terms), although higher than 91% of the DRC's population, was well below the US$4.20 and US$8.30[52] per day per capita that the World Bank defines as the lower-middle income and upper-middle income thresholds, and close to the international poverty line of US$3.00.

There was emerging evidence that households in the nine urban settlements studied have benefitted because of the fibre line.

Almost four years after the arrival of faster internet, the households in these nine urban settlements experienced the following:

- An increase in the average household asset wealth (as measured by AWI) has moved the treated settlements from the 98th to the 99th percentile settlements. This compares to the synthetic control settlements moving from the 97th to the 98th percentile.

[51] Whether or not existing internet users can benefit from faster internet does not only depend on their vicinity to Liquid's fibre line but also whether the networks of MNOs and the handheld devices of the users allow for faster data transmission.

[52] These thresholds are in terms of 2021 PPP spending. In terms of 2017 PPP spending, they correspond to US$3.65 and US$6.85.

- US$1.25 PPP higher spending per capita in 2024, compared to an increase of US$0.30 for the synthetic control. In other words, the spending of households in the settlements that were connected to the fibre line increased by US$0.95 per capita per day more compared to households in settlements that were not. This has moved the treated settlements from the 91st to the 97th percentile for spending set against the distribution of all households in the region, compared to the synthetic control, which moved from the 89th to the 94th percentile.

These results, however encouraging, are not sufficiently robust yet to claim strong causal inference.

There is a 32% probability that the observed AWI results could have arisen by chance, which is well above the usual 5% considered as statistically significant. Although the spending results have greater statistical significance, at 7% in 2024, the higher confidence should be interpreted with caution. Since spending is partially derived from AWI, the additional assumptions and propagating errors within it make it a less stable dataset than AWI. Consequently, the true reliability of these findings may be overstated.

Care must be observed to infer causality.

The SNEL line covers a very large area and unobserved confounding factors in the pre-treatment period and idiosyncratic post-treatment shocks can introduce sources of bias. The fact that the first years after the fibre line became operational coincided with the Covid 19 pandemic and the recovery from it is an example of this. Any regional differences in the severity or recovery speed could have affected the findings.

Recommendations

1. **Based on the emerging evidence collected in this study, BII should consider other broadband investments in underserved regions.** Because of its size, inaccessibility and limited economic development, the DRC is among the most underserved countries in terms of backbone fibre. Although globally the largest gaps in backbone infrastructure are being filled, there are still regions where backbone internet investments are much needed. Within the DRC, the completion of the Kananga – Goma line is recommended. Some examples other than DRC are Ethiopia and South Sudan.
2. **We recommend that BII continues to invest in the 'last mile' through which people access the internet.** The presence of broadband fibre is necessary but not sufficient to increase internet quality and penetration. Last mile investments in internet connectivity involve MNOs through the (co)hosting of towers and higher bandwidth mobile antennas. They also involve financing internet-capable handsets, which the Global System for Mobile Communications Association (GSMA) now considers the single greatest barrier keeping people offline.
3. **If possible, pursue opportunities to ground-truth the findings for this investment and to understand the drivers of change at a household level**. BII can seek opportunities to ground-truth the findings of this study by comparing the results to evidence collected through other methods, including primary (survey) data. This could involve estimating (with MNOs that use the fibre line) how internet penetration and usage have developed in the regions along the line, as well as how business growth and productivity have changed in these areas. While broadband fibre forms the essential backbone for internet

connectivity, it alone does not guarantee increased usage or penetration, and a ground-truthing study should also examine other elements along the evidence chain to provide a more comprehensive understanding.

4. **Repeat the analysis in two years to increase the statistical robustness of the emerging impact and its durability and include more settlements that were connected after 2021**
    a. When the impact is durable over a longer post-treatment window, the cumulative gap between treated units and the synthetic control should become more pronounced. Statistical confidence also increases as random (placebo) effects dissipate over time.
    b. More settlements have obtained fibre access points after 2021, both along the SNEL line as well as the Kananga–Goma line, if fully completed (see recommendation 1). Inclusion of these settlements could also strengthen the robustness of the study. An important requirement, based on the results in this study, is that four years of post-treatment data should be available. Including settlements that were connected from 2022 onwards would also mitigate the possible impact of any regional heterogeneity resulting from the intensity of and the recovery from the Covid 19 pandemic.

Annexes

Annex 1: References

Abadie, A., Diamond, A. and Hainmueller, J. (2007) Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association* 105(490): 493–505.

GSMA (n.d.) 'GSMA Mobile Connectivity Index'. https://www.mobileconnectivityindex.com/index.html

GSMA (2023) The State of Mobile Internet Connectivity 2023.

GSMA (2024) The State of Mobile Internet Connectivity 2024.

Hjort, J. and Poulsen, J. (2017) The Arrival of Fast Internet and Employment in Africa. National Bureau of Economic Research, Working Paper 23582. http://www.nber.org/papers/w23582

Hjort, J. and Poulsen, J. (2019) The Arrival of Fast Internet and Employment in Africa. *American Economic Review* 109(3): 1032-1079.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B. and Ermon, S. (2016) Combining satellite imagery and machine learning to predict poverty. *Science* 353(6301): 790–794. https://doi.org/10.1126/science.aaf7894

Joint Impact Model (n.d.) 'The Joint Impact Model: The non-profit platform to calculate financed emissions, impacts, and climate risk'. https://www.jointimpactmodel.org/

Kemp, S. (2022) 'Digital 2022: The Democratic Republic of the Congo'. https://datareportal.com/reports/digital-2022-democratic-republic-of-the-congo#:~:text=Internet%20use%20in%20the%20Democratic,at%20the%20start%20of%202022

Kemp, S. (2023) 'Digital 2023: The Democratic Republic of the Congo'. https://datareportal.com/reports/digital-2023-democratic-republic-of-the-congo?rq=DRC

Kim, R., Sutherland, Z., Verhoeven, S., Binet, S., Düring, N., Barnett, C., Lemma, A. and Beckmann, L. (2022) Final Report: Evaluating the Impact of British International Investment's Infrastructure Portfolio. e-Pact consortium: Itad, Steward Redqueen, Overseas Development Institute. https://assets.publishing.service.gov.uk/media/623b3cb98fa8f540f6c2322c/BII_Infrastructure_-_Formal_Evaluation_Report_-_Final_Report_230322.pdf

Koutroumpis, P. (2019) The economic impact of broadband: Evidence from OECD countries. Technological Forecasting and Social Change 148(119719).

Ouzonis, G. K. (2024) 'Atlas of Human Settlements: A New Perspective on Human Presence Across the Globe'. https://www.atlasai.co/blog/atlas-of-human-settlements

Ratledge, N., Cadamuro, G., de la Cuesta, B., Stigler, M. and Burke, M. (2021) Using Satellite Imagery and Machine Learning to Estimate the Livelihood Impact of Electricity Access. https://doi.org/10.3386/w29237

Ratledge, N., Cadamuro, G., de la Cuesta, B., Stigler, M. and Burke, M. (2022) Using machine learning to assess the livelihood impact of electricity access. *Nature* 611: 491–495. https://doi.org/10.1038/s41586-022-05322-8

Regeneris (2018) The Economic Impact of Full Fibre Infrastructure in 100 UK Towns and Cities.

Simmons, A. (2024) 'Cell Tower Range: How Far Do They Reach?' https://dgtlinfra.com/cell-tower-range-how-far-reach/

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S. and Burke, M. (2020) Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature Communications* 11(1): 2583. https://doi.org/10.1038/s41467-020-16185-w

Annex 2: Data accessed

Source	Year(s)	Variable	Definition	Format	Use
Liquid Technologies	2024	ADMs	Location of access points on the fibre cable	Vector shapefile	Used to identify treatment communities.
AtlasAI datasets	2012–24	Per capita spending	Estimate of poverty (US$/person/day)	High-resolution rasters at 1 km x 1 km scale	Used to understand general characteristics of settlements in the treatment and donor pools but not used in cosine similarity analysis or in developing synthetic controls (AWI is preferred as a superior measure).
		Asset wealth	Relative wealth of community (index)	High-resolution rasters at 1 km x 1 km scale	The dependent variable in the causal inference analysis and used in the cosine similarity analysis.
		Population	Population count (number)	High-resolution rasters at 1 km x 1 km scale	Used as a comparison to the WorldPop population dataset.
		Electrification	Status of electrification (yes/no)		Used as a comparison to the High-Resolution Electricity Access (HREA) electrification dataset.
		Atlas of Human Settlements	All built-up areas (yes/no)	Vector shapefile	Used to delineate settlements and represent each unit in the treatment or donor pool.
Demographic and Health Surveys (DHS)	2013–14	Occupation	Percentage of population employed in various industries	Country-wide survey data	Used to understand general characteristics of settlements but not used in cosine similarity analysis or in developing synthetic controls (not deemed to add additional value beyond selected variables).
WorldPop	2012–20	Population density	Population density (people/km^2)	High-resolution rasters at 100 m x 100 m scale	Used to understand general characteristics of settlements but not used in cosine similarity analysis or in developing synthetic controls (not deemed to add additional value beyond selected variables).
		Population	Population count (number)	High-resolution rasters at 10 m x 10 m scale	Used in the cosine similarity analysis.
OpenStreetMap (OSM)	Current	Health facilities	Location of health facilities	Vector shapefile	Used to understand general characteristics of settlements but not used in cosine similarity analysis or in developing synthetic controls (not deemed to add additional value beyond selected variables).
		School facilities	Location of school facilities	Vector shapefile	Used to understand general characteristics of settlements but not used in cosine similarity analysis or in developing synthetic controls (not deemed to add additional value beyond selected variables).
		Roads	Location of roads	Vector shapefile	Used in the cosine similarity analysis.

Annex 3: SCM approach

A3.1. Different versions of the synthetic control approach

There are many versions of the SCM, which can be tailored to different implementation scenarios.

Version	Description	Useful in situations where…	Implications for this study
Classical SCM	Designed primarily to estimate the effects of large aggregate interventions focused on a small number of large treatment units (typically one). Can, in theory, be applied to a larger number of treatment units through separate classical SCM for all treated units, but this has implications.	…there is one treatment unit of interest, with many potential donor units, and where data (for both dependent and independent variables) is available over a longer time horizon, for example an assessment of policy changes at a national level.	High degree of rigidity: donor unit weights must be non-negative; donor unit weights must sum to 1; values of the predictors for the treated unit should be near or inside the convex hull of the values for the donor pool. Risk of masking underlying data heterogeneity if data is aggregated before analysis, or risk of overcomplicating analysis if individual classical SCM is performed for all treatment units. Resource-intensive if separate SCM applied to all treated units.
SCM with elastic net	Similar to classical SCM, but with a more relaxed definition with a regularisation function to reduce overfitting,[53] which has been used frequently in the literature.[54] Allows for aggregation at analysis stage rather than of the underlying data.	…there are multiple treatment units, and a balance is sought between applying a single classical SCM and running many individual classical SCMs in parallel for all treated units. Typical applications include assessments at regional or settlement level. More relevant in situations where data is available over shorter time horizons as a result of relaxations.[55]	More flexibility/less rigidity allows for more treatment units than years of treatment; donor weights can be negative; donor unit weights do not have to sum to 1. Less susceptible to overfitting in situations where the data record is not extensive. Offers a 'middle ground' between either applying a single classical SCM to all treatment units together, as the approach was initially intended (this leads to challenges in aggregating underlying data pre-analysis), or attempting to run separate classical SCMs for each treatment unit (this leads to challenges associated with the complexity of later analysis).

[53] Based on the use of an elastic net drawing on a combination of lasso and ridge penalties.

[54] Ratledge, N. *et al.* (2021) Using Satellite Imagery and Machine Learning to Estimate the Livelihood Impact of Electricity Access. https://doi.org/10.3386/w29237

[55] Nevertheless, a minimum of five years pre- and post-treatment is recommended.

A3.2. Implications of relaxing the constraints of the synthetic control method

The SCM constructs a synthetic control unit by combining weighted observed units in order to estimate the counterfactual outcome for a treated unit. One of the assumptions of the classical SCM is that the weights assigned to the observed units should sum to 1. We decided to relax this constraint in our analysis. We are aware of a few potential implications of doing this:

- **Bias in estimated treatment effect.** The constraint that weights sum to 1 ensures that the synthetic control is a convex combination of the observed units. This convexity property is important for maintaining a balanced and unbiased estimate of the treatment effect. If the constraint is relaxed, it is possible that the synthetic control will become skewed towards certain units, leading to biased treatment effect estimates.
- **Model overfitting.** Relaxing the weights sum constraint can lead to overfitting, whereby the synthetic control becomes too tailored to the pre-treatment outcomes of the treated unit.
- **Loss of interpretability.** The weights sum constraint enhances the interpretability of the synthetic control. When weights are required to sum to 1, each weight represents the proportion of the corresponding observed unit's characteristics in the synthetic control. Without this constraint, it is possible that the resulting weights will not have clear interpretive value.

Although recognising these potential implications, we opted to relax the classical SCM requirement that weights should sum to 1. This is a recognised tactic when using our chosen analytical approach (SCM with elastic net). It offers more flexibility/less rigidity and is especially appropriate in cases where there is a relatively large number of treatment units but relatively few treatment years. We combined this approach with a regularisation function to reduce susceptibility to overfitting.

A3.3. Exploring penalty factor to fine-tune the synthetic control model performance

In our regression approach for synthetic controls, we tested three different penalty factors to see how they would affect the results, particularly the match between the synthetic control and treatment units in our pre-treatment time period (2012–20). This penalty factor ranges from 0 to 1, representing ridge and lasso penalties respectively. Any value between 0 and 1 is considered elastic net. The goal of the penalty factor is to control the complexity of the model and improve generalisation performance.

Below is a description of how each penalty affects the synthetic control model and how we fine-tuned our model's performance:

- **Ridge penalty** (alpha = 0): The ridge penalty is like a safety net that helps our causal inference model maintain stability. It does not forcefully remove any donor units from our model, but it encourages them to be more consistent. When we use ridge in our causal inference analysis, it reduces the risk of our model being too sensitive to small changes in our data. This is important because it helps ensure that our results are reliable over time. In essence, ridge helps keep our model steady and dependable when matching synthetic control units with treatment units from the pre-treatment period (2012–20).

- **Lasso penalty** (alpha = 1): The lasso penalty is like a strict editor for our causal inference model. It emphasises picking only the most important pieces of information from our data. When we apply lasso in our analysis, it forces some donor units to be completely excluded, and the focus is on the most influential ones. This simplifies our model and makes it easier to understand. In the context of causal inference, lasso helps us pinpoint the key units that drive a strong match between the synthetic control and treatment units; it streamlines our model and boosts its performance.
- **Elastic net penalty** (alpha = 0.5): The elastic net penalty (with an alpha value of 0.5) is a compromise between ridge and lasso. It combines the stability of ridge with the donor unit selection of lasso. In our causal inference model, elastic net gives us a way to balance the complexity of the model, ensuring it is both robust and interpretable.

We experimented with all three penalty values in our model to determine which one yielded the optimal outcomes when aligning our synthetic control with the treatment sites. Our objective was to find the right equilibrium between achieving a strong fit between the synthetic control and treatment sites and avoiding overfitting to any noise in the data. Results for all three methods are shown, but for future analysis we recommend the elastic net penalty, because it strikes a balance between stability and complexity.

Annex 4: Key definitions and derived metrics

Asset Wealth Index

AtlasAI's Asset Wealth layer estimates household asset wealth based on asset ownership. The original AWI is a derived construct that considers assets such as electric appliances, nightlights, and land owned by a household.

Population and population density

Population and *population density* are measures of population count and population per km^2. This variable will be used to assess how population and population density have been changing over time. The layer has been attributed to the settlement areas with data for the years 2012–21.

Road access, length and density

Road access is a derived indicator using road locations from OSM and AtlasAI settlement areas. This metric measures the Euclidean distance between each catchment area and the nearest major road, defined as primary, secondary or tertiary. The distance to the nearest roads provides a way to determine the accessibility of each settlement. *Road length* and *density* are similar metrics and provide a way to further describe settlements that contain roads. *Road length* is the total length of major roads within a settlement area; *density* is the total length of major roads divided by the settlement area. Of the 148,157 catchment areas, only 31,328 contained a major road. If a settlement area contains a higher density of major roads compared to another settlement area, this could indicate that the settlement is more connected.

A4.1. Asset Wealth Index interpretation

The AWI is an economic construct that estimates the accumulated wealth and well-being of a household, derived from an inventory of the valuable items purchased and collected over time, for example appliances, livestock, property and vehicles. The AWI is a valuable metric when income statistics, tax records or other evidence of monetary wealth are not available.

Although the AWI can be calculated household by household, a more robust statistical estimate is obtained by cluster households within a community or across small proximate communities. The interpretation of AWI is therefore the average indexed wealth per household in a community of interest. Furthermore, by comparing the non-dimensional index across space and time (spatial time series), we can draw insights about the changes in well-being within and across communities on average at household level.

The AWI has a 2 km x 2 km resolution. This resolution refers to the level of detail and granularity present in each pixel (or polygon) of the raster image. Resolution is typically measured in terms of the size of the smallest discernible unit on the ground, often represented in metres, feet, or other units of distance. Therefore, a 2 km x 2 km pixel (or polygon) has one value for that entire area. We combined this raster with our settlement areas and calculated the average household AWI for each settlement area. Because we have an average of household-level AWI for each settlement, it would not be possible to pull out an individual household AWI; using this dataset, it is only possible to perform a community-level impact assessment.

A4.2. Asset Wealth Index production

The Asset Wealth layer is produced from a deep learning model that predicts survey-based estimates from satellite imagery. To facilitate comparison within and across countries, we transformed asset wealth into a normalised index. To generate this data, we collate locally representative survey data on household asset ownership to create an AWI, which is the first principal component of a principal component analysis (PCA) computed on those assets over those households. We then train a random forest model to predict village-aggregated values with satellite imagery, validating on data the model was not trained on.

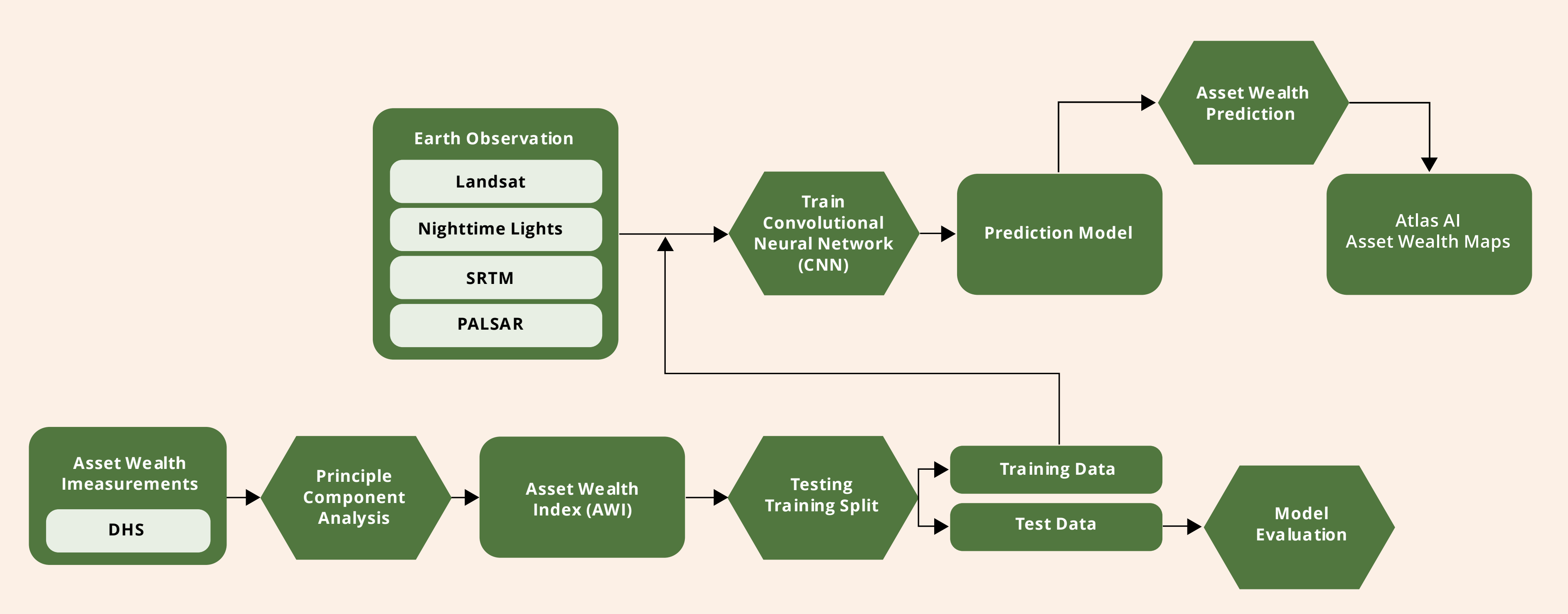


Figure source: https://docs.atlasai.co/economic%20well-being/asset_wealth/

Annex 5: Summary of comparisons to other approaches

Note: we include randomised control trials (RCTs) here for reference purposes as the 'gold standard'. It is rarely feasible to apply RCTs to investment projects, given the requirement to randomise treatment.

Approach	Cost	Time required	Data requirements	Engagement by investment owners	Flexibility	Evidence standard
RCT	High	High	High	High	Low	Gold standard (but rarely feasible)
	Costs will vary by scope.	Requires before and after data collection on the ground, with sufficient time intervals for impact to emerge.	Primary data typically required at household level before and after.	Requires adaptation of implementation models to enable randomised treatment.	Typically, not possible to adapt or scale up after baseline data collected.	Requires randomised assignment of treatment and control units.
Quasi-experimental designs (e.g. difference-in-difference)	Medium	High	High	Medium	Medium	Very strong (but limited flexibility)
	Costs will vary by scope.	Requires before and after data collection on the ground, with sufficient time intervals for impact to emerge.	Primary data typically required at household level before and after.	Requires support to identify and access treated and untreated locations on the ground.	Allows for some flexibility in application post-baseline.	Flexibility is limited once baseline data is collected; there is risk of contamination of control units over time. Difficult to apply to investments with rapidly expanding customer base. Can be difficult to identify plausible counterfactual on the ground.
Geospatial analysis with synthetic controls	Low	Low	Low	Low	High	Strong (and flexible)
	Marginal cost will fall in subsequent applications and as approach matures.	Can be done quickly and retrospectively.	Requires access to secondary geospatial datasets and geotagged data on clients, but no other primary data.	Limited engagement needed beyond providing geotagged data on clients.	Can be scaled and repeated quickly; does not require a baseline.	Flexible – can offer a robust counterfactual even where difficult to identify physical control groups.

We provide expert monitoring, evaluation, learning and strategy services to help build a more equitable and sustainable world for all.

itad.com @ItadLtd Itad mail@itad.com

Itad Ltd
International House
Queens Road
Brighton, BN1 3XE
United Kingdom

Tel: +44 (0)1273 765250

Itad Inc
c/o Open Gov Hub
1100 13th St NW, Suite 800
Washington, DC, 20005
United States