

Impact assessment for market-building interventions in capital markets: Review, critiques and improvements

The ODI research series for financial development in Africa

Issouf Soumaré, Désiré Kanga and Judith Tyson

May 2021

Abstract

To develop African capital markets and increase financial inclusion, specific interventions have been introduced to better design policies and programmes for capital market development. One important question is how to assess the impacts of these interventions in capital markets, since most of them are implemented at small scale and it is usually very difficult to find an appropriate control group in order to perform a rigorous impact assessment. This paper reviews existing techniques and methods used to evaluate the impact of interventions in general and discusses their application in the financial sector. Most interventions in financial development conducted by development finance institutions were evaluated through the Development Assistance Committee's five criteria: relevance, effectiveness, efficiency, impact, and sustainability. Recent developments in micro-econometrics allow the use of a variety of impact evaluation methods such as randomised control trials, instrumental variables, propensity score matching, regression discontinuity and double difference. These techniques, however, are more suitable for interventions at micro unit level and for evaluating the direct impact of the interventions. To account for spillover and macro-economic effects of interventions implemented at macro level, general equilibrium, synthetic controls, and vector autoregression approaches can be used. These approaches are also reviewed in the current paper.

Key words: Capital market, impact assessment, impact evaluation, capital market intervention, Africa.

About this series



The ODI research series for financial development in Africa funded by FSD Africa

FSD Africa is a specialist development agency working to build and strengthen financial markets across sub-Saharan Africa. Its mission is to reduce poverty through a ‘market systems development’ approach addressing the structural, underlying causes of poverty by improving how financial market systems function.

As part of its work, FSD Africa invests in breakthrough research, analysis and intelligence-gathering projects to provide policy makers, regulators, development actors, investors and financial service providers with vital market insights and information and that contributes to long-term change.

As part of this research, FSD Africa has partnered with ODI, the independent global development think-tank, to address policy issues in financial sector development in Africa. This includes a series of working papers by and for leading academics and an accompanying series of briefing papers focused on the policy implications of research targeted at policy-makers, regulators and the broader stakeholder community.

The research programme is being accompanied by joint FSD Africa and ODI dissemination events in the region and internationally to present and debate the policy implications of the research findings.

Acknowledgements

Readers are encouraged to reproduce material for their own publications, as long as they are not being sold commercially. ODI requests due acknowledgement and a copy of the publication. For online use, we ask readers to link to the original resource on the ODI website. The views presented in this paper are those of the author(s) and do not necessarily represent the views of ODI or our partners.

This work is licensed under CC BY-NC-ND 4.0.

How to cite: Soumaré, S., Kanga, D. and Tyson, J. (2021) Impact assessment for market-building interventions in capital markets: Review, critiques and improvements. Joint FSD Africa–ODI research programme for financial sector development in Africa. Working Paper 3. London: ODI.

Acknowledgements

We thank participants at the ODI–FSD Africa–Université Laval Virtual Workshop, held on 10 December 2020, for their very helpful comments and suggestions, which have greatly improved the quality of the paper, especially, Ahmad H Ahmad, Abbi M. Kedir, Mark Napier, Evans Osano, Kevin Munjal, Philip Buyskes, Paddy Carter, Jacqueline Theresa Irving, Rebecca Waghorn, Joost Zuidberg, and Sherillyn Raga. All errors and omissions are the authors' sole responsibility.

About the authors

Professor Issouf Soumaré is Professor of Finance and the Director of the Laboratory for Financial Engineering of Université Laval in Canada. He was the Director of International Relations of the Faculty of Business Administration at Université Laval from 2012 to 2020. His research and teaching interests include financial sector development, international finance, risk management and financial engineering. His theoretical and applied financial economic works have been published in leading international economics and finance journals. He worked for the African Development Bank at the former Risk Management Unit from 1996 to 1998. He is the President of the Scientific Council of the *Ecole nationale supérieure de statistique et d'économie appliquée d'abidjan* (Côte d'Ivoire). He is a Resource Person for the African Economic Research Consortium, member of the Scientific Council of the Moroccan Financial Market Authority, member of the Programme Steering Committee of Financial Sector Development programme of the West African Economic and Monetary Union and member of the Scientific Committee of the *Revue économique et monétaire* of the *Banque Centrale des États de l'Afrique de l'Ouest*. He is also part of the consortium of researchers on the “FCDO-ESRC Growth Research Programme Call 3: ESRC Ref. ES/N013344/2; Research Project: Delivering Inclusive Financial Development & Growth in Africa, 2016-2021.” Prof. Soumaré holds a PhD in Business Administration in Finance from the University of British Columbia, Canada and an MSc in Financial Engineering from Université Laval.

Dr Desire Kouamé KANGA received his doctoral degree in Economics from the University of Orleans, France. His research interests are in econometrics, monetary policy, impact evaluation and access to finance. He is an economist at the *Banque Centrale des États de l'Afrique de l'Ouest*. Before this, he worked as a Lecturer at the National Applied Economics and Statistics School of Côte d'Ivoire and as a Postdoctoral Research Fellow at SOAS University of London. His work has been published in *Journal of Banking and Finance*, *Revue économique*, *Revue d'économie politique*, *Economics and Statistics*, and *Revue française d'économie*.

Dr Judith Tyson is a Research Fellow at the ODI and a specialist in finance for development including financial market development, private investment and financial macroeconomics. She is a ‘subject matter expert’ for the UK’s Independent Commission on Aid Impact and a number of Foreign, Commonwealth and Development Office (FCDO) panels and is co-investigator for a £2 million research program for African financial development led by the School of Africa and Oriental Studies (SOAS) at the University of London. Clients include the UN Economic Commission for Africa, CDC Group, FSD Africa, FCDO and national governments including in Kenya, Liberia, Nigeria, Rwanda and Tanzania. She also regularly advises private clients. She holds a Master and PhD in financial sector development in Africa from the School of Africa and Oriental Studies at the University of London.

Contents

About this series	2
Acknowledgements.....	3
Acronyms	7
1 Introduction	8
2 Drivers of capital market development	10
2.1 Basic functions of capital market for economic development	10
2.2 Macroeconomic and institutional factors for capital market development.....	11
2.2.1 Institutional quality	11
2.2.2 Domestic resource mobilisation.....	12
2.2.3 International capital flows.....	13
2.2.4 Fiscal policies and crowding-out effect	13
2.2.5 Infrastructure development and transactions costs	13
2.2.6 Market demand size	14
3 Current interventions: Case studies	15
3.1 Impact assessment of the Capital Market Development Program in Bangladesh (Asian Development Bank, 2005).....	15
3.2 CDC's mobilisation of private investment (Spratt et al., 2019).....	17
3.3 The experience from the World Bank Group (World Bank, 2016).....	18
4 New approaches in impact assessment	21
4.1 Impact assessment: Definition and challenges.....	21
4.2 Econometric techniques for micro assessment.....	24
4.2.1 Randomised control trial	24
4.2.2 Instrumental Variables	27
4.2.3 Regression discontinuity design	29
4.2.4 Difference-in-differences (double-difference)	31
4.2.5 Propensity score matching	32
4.3 Econometric techniques for macro assessment.....	34
4.3.1 Computable General Equilibrium models	35
4.3.2 Dynamic Stochastic General Equilibrium models.....	35
4.3.3 Agent-based models.....	37
4.3.4 Heterogeneous agent new Keynesian models	38
4.3.5 Synthetic controls.....	39
4.3.6 Vector autoregression model.....	40
4.4 Summary table of the econometric techniques for micro- and macro-assessment.....	41
5 Conclusion.....	47
References	48
Appendix 1 Application to selected FSD projects.....	53

Acronyms

ABM	agents-based models
ACE	allowance for corporate equity
CGE	computable general equilibrium
CMDP	Capital Market Development Program
DD	difference-in-differences
DSGE	dynamic stochastic general equilibrium
EME	emerging market economies
GDP	gross domestic product
HANK	heterogeneous agent new Keynesian
IFC	International Finance Corporation
ITT	intention-to-treat
IV	instrumental variables
KID	key information document
LATE	local average treatment effect
MSME	micro, small and medium enterprise
PSM	propensity score matching
RCT	randomised control trial
RD	regression discontinuity
SEC	Securities and Exchange Commission
SME	small- and medium-sized enterprise
TOT	treatment-on-the-treated
VAR	vector autoregression

1 Introduction

Financial institutions play a critical intermediary role between providers of funds and those in need of funds (households, firms and governments). For instance, by providing effective risk management vehicles, financial institutions and capital markets contribute to risk diversification, thereby channelling investments toward high risk-adjusted return projects, typical in Africa. A large body of empirical research indicates that improvements in financial systems foster economic development, e.g., Allen et al. (2014), Atindehou et al. (2005), Beck et al. (2009), Bekaert et al. (2005), Coulibaly et al. (2018), Klein and Olivei (2008), Levine et al. (2000), Otchere et al. (2016), Soumaré (2015), among many others. Moreover, capital markets can contribute to technological innovation and diffusion as well as lower transaction costs, and hence facilitate trade and specialisation. However, if capital markets are underdeveloped – small and less liquid – and perform these functions poorly, they can hinder economic growth.

Financial systems in many African countries remain underdeveloped (e.g. Allen et al., 2011 and Beck and Cull, 2013), which results in credit constraint for households and firms, especially small and medium-sized enterprises (SMEs), and low investment rates. The African financial sector is dominated by commercial banks, with very few investment banks. National development banks and specialised banks, except for the African Development Bank and the Development Bank of South Africa, are very limited in their capacity to raise enough external finance to fill the financing needs of businesses. Consequently, finance has been identified as the most severe obstacle to doing business in Africa and financial constraint is a major hindrance to business start-ups and innovation by firms (Ayyagari et al., 2011; Gorodnichenko and Schnitzer, 2013). All these cause very low investment rates in Africa (24%) compared to other emerging world economies, such as China (40%), South Asia (28%) and East Asia and Pacific countries (32%).¹

To ease access and foster the development of African capital markets, several initiatives have to be put in place. One example is the expansion of capital market access to micro, small and medium enterprises (MSMEs) and to financially excluded households. For the success of these initiatives, policies must be designed to meet the needs and aspirations of local, national, and regional stakeholders (public and private) with the support of international development partners. As an international specialist development agency, FSD Africa is working to build and strengthen financial markets across sub-Saharan Africa. To achieve its goals, FSD Africa partners with all the critical branches of Africa's financial markets – banking, capital markets, fintech, insurance, market infrastructure, credit, and informal economy – to

¹ Investment (gross capital formation) is measured by the total value of the gross fixed capital formation and changes in inventories and acquisitions less disposals of valuables for a unit or sector. It is expressed as a percentage of GDP. The percentages are computed using data from the World Bank's World Development Indicators database 2019.

develop effective financial sector development programmes to address the most pressing challenges facing Africa's financial sector, from long-term finance to climate change to affordable housing.

One pressing question is how to assess the effectiveness of FSD Africa's interventions in African capital markets. Impact assessment for interventions in capital markets are sometimes difficult given the intrinsic nature of these interventions. Indeed, most of these interventions are small in scale and sometimes it is very difficult to find an appropriate control group to perform a rigorous impact evaluation. Even when this last condition is met, it is often difficult to dissociate the direct and indirect impacts of the policies and programmes being evaluated. Therefore market-building interventions in capital markets seem to lend themselves less to formal assessments. Hopefully, some of the new techniques we reviewed below allow the evaluation of such interventions either at micro or macro level by considering spillover and macro-economic effects.

To shed light on some of the issues raised above, this paper is organised as follows. Section 2 provides a synthetic literature review of causes and effects of capital market development. In section 3, we review some past interventions in capital markets by development finance institutions. Section 4 presents existing and new approaches in micro and macro impact assessments and discusses their applications to capital market interventions. We conclude in section 5.

2 Drivers of capital market development

Traditionally, ‘capital markets’ refers to venues or places for trading long-term debt instruments (those that mature in more than one year), including stock and bond markets. More recently, the term has been used in a more general sense to refer to markets for equities, bonds, derivatives and other investments.

Theory is equivocal on the link between capital markets development and economic growth. On the one hand, capital markets can have a positive effect on long-run economic growth (Dabla-Norris et al., 2020; Laeven et al., 2015); on the other, they can jeopardise growth (Kaminsky and Reinhart, 1999; Calamanti, 1983). Other authors identified a two-way causality relationship between capital market development and economic growth (e.g., Patrick, 1966). The remaining of this section provides a literature review on the role of capital market for economic development and factors of capital market development.

2.1 Basic functions of capital market for economic development

A literature survey conducted by Laeven (2014) highlights the key role played by (local) capital markets, especially bond markets and the rationale for their development. According to this work, the main contributions of capital market development to the economy can be summarised as follows. First, local bond markets allow governments to finance large fiscal deficits without having to resort to financial repression or foreign borrowing. Second, the development of money and bond markets supports the conduct of monetary policy. Third, the development of local capital markets can improve the availability of long-term financing, allowing households and firms to better manage interest rate and maturity risks associated with long-term investments. Indeed, the development of capital markets through the mobilisation of savings can: (i) reduce transaction costs associated with collecting savings from different individuals; and (ii) mitigate informational asymmetries associated with making savers feel comfortable in relinquishing control of their savings. Fourth, the development of local capital markets can improve access to local currency financing. Fifth, local capital markets allow for financial deepening alongside the development of banking markets, improving the efficiency of capital allocation in the economy. Sixth, local capital markets, when opened to foreign investments, increase financial integration by attracting foreign capital, which can lower the cost of capital for local firms and household and improve risk sharing across countries. Finally, the development of local capital markets can enhance financial stability by strengthening the ability of financial institutions to manage risk.

It is well known that financial development is closely associated with economic growth through the role of financial intermediaries or markets. Financial intermediaries reduce the costs of acquiring and processing information and improve

resource allocation within the economy (Boyd and Prescott, 1986), which leads to sustained economic growth. Most academic publications found a positive relationship between financial development and economic growth (e.g., Calderon and Liu, 2003; Ibrahim and Alagidede, 2018; Levine et al., 2000; Levine and Zervos, 1996; Masten et al., 2008; Rioja and Valev, 2004). The main question is how does it work? One channel is innovation which is central to economic development. Finance is essential to enabling the significant investment required for innovation (Laeven et al., 2015). Because innovation is characterised by long lead times, uncertainty, cumulativeness and collectiveness, it requires special financial vehicles (e.g., Mazzucato and Semieniuk, 2017). For example, uncertainty means that finance must be willing to bear high risks while the long-run nature of innovation and its cumulativeness imply that the kind of finance must be patient. A developed financial market should channel funds to highly productive sectors. This is probably the reason why Atje and Jovanovic (1993) found that stock market development could have a substantial effect on economic growth, while they failed to find a similar effect of bank lending on economic growth.

Nevertheless, well-developed financial sector can contribute to dampen (magnify) the impact of real (monetary) shocks on business cycle and long-run volatility components (Ibrahim and Alagidede, 2017). Therefore, development of the financial sector does not necessarily translate into higher growth and may even distort sustained path towards development (see Kaminsky and Reinhart, 1999; Duffie, 2019; Stiglitz, 2000). Ocampo and Stiglitz (2008) have shown that there are inherent risks in the liberalisation of capital markets. For example, pro-cyclical capital flows have been associated with crisis, particularly in developing countries since the 1980s. Although they are not necessarily the cause, they still help to propagate the crisis. In addition, the study by Calamanti (1983) on selected African securities markets concluded that securities markets cannot significantly contribute to growth in these economies because, even in the few country cases where securities markets were linked with economic growth, the markets seemed to be the result rather than the cause of such growth. She argued that capital markets may actually jeopardise growth in developing countries with underdeveloped markets – by producing economic instability and adversely affecting savings allocation and the reallocation of existing real wealth.

2.2 Macroeconomic and institutional factors for capital market development

The development of local capital markets has been a long-standing policy question. Sound macroeconomic policies, strong institutional and legal setting, and a well-functioning financial infrastructure are the main preconditions for proper functioning of local capital markets (Coulibaly et al., 2019; Laeven, 2014; Otchere et al., 2016; Soumaré and Tchana Tchana, 2015). Policies and laws matter for the development of bond markets (Burger and Warnock, 2006).

2.2.1 Institutional quality

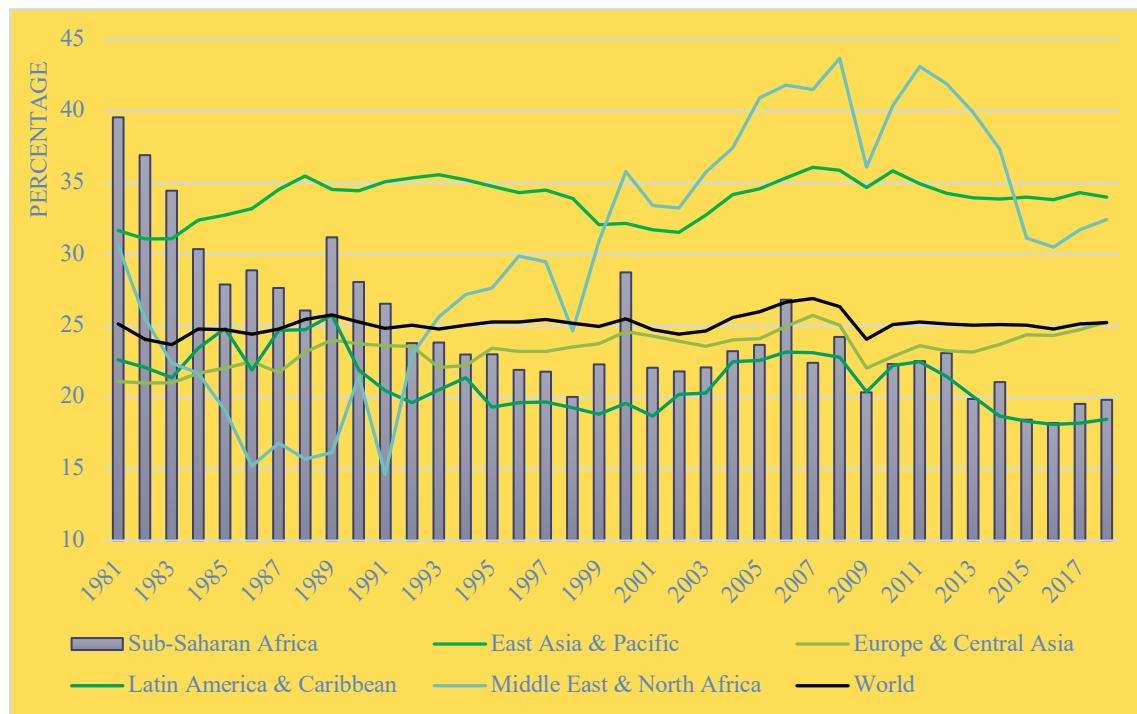
A sound macroeconomic framework and stable macroeconomic policies will attract foreign capital and ensure that monetary policy actions can be taken without causing excessive interest rate volatility. Burger and Warnock (2006) found that countries with stable inflation rates have more developed local bond markets and rely less on foreign-currency-denominated bonds. This is also true for countries with strong

creditor rights (Burger and Warnock, 2006, 2012). In fact, stronger rule of law is associated with deeper local bond markets, while countries with better creditor rights can issue a higher share of bonds in their local currency. Strong institutions and a well-functioning legal system provide the basis for the protection of investor rights, including minority interests. This is consistent with the literature stating that economies with investor-friendly laws tend to have deeper capital markets (La Porta et al., 1997, 1998; Claessens et al., 2007; Eichengreen and Luengnaruemitchai, 2006; among many others).

2.2.2 Domestic resource mobilisation

Another factor that can limit the development of local capital markets is the level of domestic savings. Figure 1 shows that domestic savings in sub-Saharan African countries have been among the lowest in the world since 2000. In fact, it is the second lowest after Latin America and Caribbean region. The average gross domestic savings in the sub-Saharan Africa region was estimated at 19.8% of GDP compared to the world average of 25.2%.

Figure 1: Evolution of gross domestic savings (% of GDP)



Source: Word Development Indicators, 2020.

In theory, the low level of domestic savings can be compensated by international capital inflows (from the rest of the world) depending on the degree of openness of the country. Integration of global financial markets and integration of global goods markets are needed to achieve net transfers of capital (Ford and Horioka, 2017), which are essential for the development of international capital markets. However, capital is not freely moving from one country to another, not even among developed countries with good quality of institutions.² According to scholars like Eaton et al.

² See the puzzle highlighted by Feldstein and Horioka (1980) in Organisation for Economic Co-operation and Development countries.

(2016), trade barriers on goods and services is a major obstacle to international capital mobility, and this can hamper economic growth (e.g., Henry, 2003).

2.2.3 International capital flows

The potential reversal of portfolio flows in emerging markets, as experienced during the recent spread of coronavirus (Covid-19), highlights the challenges of managing volatile portfolio flows and risks they may pose to financial stability if domestic capital markets in emerging markets rely on foreign capital (IMF, 2020). Indeed, foreign portfolio flows may be an important source of funding for emerging markets governments and corporations when underpinned by sound macroeconomic fundamentals since they can help expand and diversify the investors base in these markets and lower the cost of funding (e.g., Desai et al., 2006; Henry, 2000, 2003). However, reliance on foreign financing can also entail risks. Heightened uncertainty in the global economy can lead to a significant tightening of global financial conditions and increased portfolio flow volatility, with reversal of capital flows (e.g., Calvo, 1998; Reinhart and Reinhart, 2008).

2.2.4 Fiscal policies and crowding-out effect

Local bond markets allow governments to finance fiscal deficits. However, this is not sustainable if fiscal policy is not well managed and can lead to crowding out of private-sector borrowers. In practice, the government can resort to the central bank, the domestic financial institutions, the international private financial institutions, and international organisations including donors to finance its deficit. If the local market is developed the government can finance its deficit by issuing local debt instruments instead of international debt instruments. Local financing puts less pressure on the government as it does not bear exchange rate risk compared to international borrowing.

Nevertheless, two potential risks can negatively affect the whole economy if the government relies heavily on domestic market. Firstly, local borrowing from a less developed capital market will generate a crowding-out effect and can probably shrink the production due to the limited access to finance of non-financial private firms which in turn will create inflation and likely external imbalances. Secondly, if the government debt becomes unsustainable after relying on the domestic market, it will negatively affect the balance sheets of financial institutions, which will lead to credit-crunch in the economy and generate low growth and high inflation. Adequate and responsible fiscal policy is needed for the growth of the domestic financial market. A limited fiscal deficit assures investors.

2.2.5 Infrastructure development and transactions costs

The development of domestic financial infrastructure is necessary to facilitate trading and the exchange of information (Laeven, 2014). The efficiency and security with which securities can be listed and traded on the exchange, together with the quality and flow of information to value those securities, will to a large extent determine the market's success. In addition, the quality of information disclosed to investors needs to be of better quality.

2.2.6 Market demand size

For capital markets to flourish, they need to meet minimum size. In this regard, investor base and corporate demand are two key constituents to consider. A large and diversified investor base ensures a strong and stable demand for a growing bond market (Asifma, 2019). More generally, the investor base includes domestic and foreign participants from a variety of institutions such as commercial and investment banks, insurance companies, pension funds, hedge funds, mutual funds, and individual investors. The lack of critical mass of investors needed to provide for market depth and liquidity is a challenge for many emerging markets. One key constraint commonly impeding local capital market development in developing countries is (particularly local bond markets) is that the role of banks as investors is greater than that of local institutional investors with longer-term liabilities such as local pension funds and life insurers.

To address such an issue, governments can open the domestic markets to foreign investors in order to diversify investors. The diversity of investors will create demand for a wide range of time horizons, risk preferences and trading motives, which are vital to boost market liquidity. These developments will enable government entities, corporates and financials to execute funding strategies under a wide range of market conditions. Another complementary option is to create a larger relative role for or building the capacity of local institutional investors with longer-term investment horizons in order to develop the buy-side of many of the national and regional capital markets. Taking steps to reduce the relative role of banks as investors in local capital markets could also help advance financial inclusion – where more bank finance is made available at more affordable rates to borrowers further down the credit spectrum, such as MSMEs and households.

3 Current intervention case studies

One way to design sound and effective policies to support the development of capital markets is to be able to assess upfront the possible effects of the policies to be implemented at a large scale. Before implementing a given policy on a large scale, for instance at the national or regional level, it may be useful to run an experiment or intervention at a small scale. However, interventions in capital markets can be harder to evaluate formally for several reasons. For example, changes in laws or regulations may occur at an economy-wide level, or a large loan may only be given to a small number (one or two) of banks or firms (McKenzie, 2010).

Nevertheless, many impact evaluations have been undertaken at sectoral level, including microfinance, micro-enterprises, insurance, etc. However, the approaches used in these impact assessments are not feasible for interventions such as reforms in the regulatory environment. In this case, other impact assessment methods have been developed to assess the impact of the interventions in capital markets. These methods are discussed in more details in section 4. For example, the financial sector modernisation support programme financed by the African Development Bank in Tunisia in 2016 elaborated a logical framework containing objectives, measures, targeted output and outcome indicators, and data sources within the institutions in charge. This framework serves as the basis for the programme's monitoring and evaluation. In this case, the Development Assistance Committee (DAC) criteria – relevance, effectiveness, efficiency, impact and sustainability – are widely used.

In the remainder of this section, we present two case studies of interventions in capital markets by two development finance institutions, the Asian Development Bank and the CDC, a public limited company whose only shareholder is the UK's Foreign, Commonwealth & Development Office. We discuss how these institutions have handled the impact assessment. These two case studies are supplemented by the World Bank's experience.

3.1 Impact assessment of the Capital Market Development Program in Bangladesh (Asian Development Bank, 2005)

After the crash of the Bangladesh stock market in 1996, the Asian Development Bank approved the Capital Market Development Program (CMDP) to the People's Republic of Bangladesh for \$80 million, divided into two tranches. The CMDP aimed to broaden market capacity and develop a fair, transparent, and efficient domestic capital market to attract larger amounts of investment capital to augment the capital resources provided through the banking system. Ultimately, the programme aimed to restore investor confidence which was significantly damaged during the financial crash because of excessive speculations, allegedly aggravated by widespread irregular activities.

The CMDP was to achieve its objective by (i) strengthening market regulation and supervision, (ii) developing the capital market infrastructure, (iii) modernising capital market support facilities, (iv) increasing the limited supply of securities in the market, (v) developing institutional sources of demand for securities in the market, and (vi) improving policy coordination.

The DAC criteria (relevance, effectiveness, efficiency, sustainability and impact) have been used to make an overall assessment of the program:

1. Firstly, on the *relevance* criteria, the CMDP was assessed as being *partly relevant*. The relevance is the extent to which the programme is suited to the priorities and policies of the target group, recipient and donor. Although the programme objective was consistent with the Government's strategy over the years, and its comprehensiveness in scope can be justified, the implementing agencies relevant to subsectors – such as insurance and pension reforms and privatisation – did not actively support the programme objective.
2. Secondly, regarding the *effectiveness*, which is a measure of the extent to which a programme activity attains its objectives, the CMDP was assessed *partly effective*. The main reason was that the weak prioritisation as well as the lack of timelines for the second tranche release reflected deficiencies in programme design.
3. Thirdly, the *efficiency* of a programme measures the outputs – qualitative and quantitative – in relation to the inputs. Efficiency is an economic term which signifies that the programme uses the least costly resources possible to achieve the desired results. The CMDP was assessed as *less efficient* because: (i) the activities of some implementing agencies were not coordinated with the Securities and Exchange Commission (SEC) and the stock exchanges, hence complementarity between programme components was less than expected; and (ii) the performance measures indicate that the capital market has not developed as expected.
4. Fourthly, the programme was assessed as *likely sustainable* in terms of whether the benefits of an activity are likely to continue after donor funding has been withdrawn. Indeed, most of the achievements of the programme – such as the introduction of SEC Automation System and the introduction of automated trading system in the stock exchanges, can be considered generally irreversible. However, stakeholders generally feel that the SEC does not have enough expertise to exercise broad rule-making power to create a market-friendly regulatory environment.
5. Finally, the *impact* of the programme was assessed. The *impact* refers to the positive and negative changes produced by an intervention, directly or indirectly, intended or unintended. This involves the main impacts and effects resulting from the activities on the local, social, economic, environmental and other development indicators. The examination should be concerned with both intended and unintended results and must also include the positive and negative impact of external factors, such as changes in terms of trade and financial conditions. The CMDP has had changes at organisational level through its contribution to build the capacity of the SEC especially in the areas of market surveillance, and audit and inspection of stock exchanges and their member brokers. On the socio-economic impact, the evaluators concluded that the CMDP's contribution to job creation in

capital market-related businesses was marginal due to the low performance of the stock markets. It is worth noting that socio-economic impact evaluation was difficult to conduct because of the insufficient participation of stakeholders in programme formulation.

Although useful, DAC impact evaluation criteria are subject to criticisms. The first criticism is that investigating and understanding all five areas is not plausible, feasible or even appropriate for capturing all these dimensions in one evaluation (Pasanen, 2018). Secondly, some of the criteria may not be the most relevant for the project to evaluate. For example, gender, equity, financial inclusion of the most vulnerable groups may be of interest for a project but are not covered by DAC criteria. A stakeholder consultation (OECD DAC, 2018) concluded that the following themes should be added to the criteria:

- Greater and more explicit harmonisation with Sustainable Development Agenda narrative
- Recognise complexity or systems models
- Specify and emphasise interconnectedness, i.e. the criteria stand “in relation” not “in isolation”
- Ensure applicability to policy, programme, systems, institutional and strategic evaluations
- Ensure strong presence of gender, equity, human rights throughout.

3.2 CDC’s mobilisation of private investment (Spratt et al., 2019)

The objective of the study by Spratt et al. (2019) was to understand the determinants of CDC’s mobilisation in developing contexts, and to provide evidence and learning that can help in realising the full potential of the mobilisation approach. CDC is one of the largest and more influential bilateral development finance institutions with £5.1 billion net assets, and £3.9 billion portfolio size in 2017. CDC was specialised in indirect investments in funds managed by external partners, but its proportion of direct investments, as well as the provision of debt financing has grown substantially. It has multiple mandates: investment return, job creation, building better capital markets, payment of local taxes by investee companies, to name a few. Mobilisation of private capital features alongside these mandates.

The two research questions of the study can be summarised as follows:

- a) To what extent have CDC investments been effective at directly mobilising private sector investment? Mobilisation is additional private sector investment that is attributable, to some degree, to current or historical CDC activities.
- b) What, if any, have been the systemic impacts of CDC on the private sector investment market, including their influence (if any) on investor sentiment and behaviour, as well as broader indicators of activity such as macroeconomic data?

The research tries to establish causal effects of CDC investments on capital mobilisation and macroeconomic variables. The team adopts a two-stage approach. At the first level, a survey is conducted with CDC staff and investors. The survey with CDC staff focuses on projects that may have generated demonstration effects. Demonstration effects are where a private investment takes place following a specific and observable CDC activity in the past, such as where returns obtained from a particular investment were higher than assumed, leading to increased private investment in similar investments in the future. This helps to identify project characteristics that seem to be predictors of demonstration effects. The second survey with investors aims to understand the internal and external drivers of decision-making for CDC-type deals in key markets and sectors, and to identify examples of previous CDC projects that have positively influenced their investment decisions.

Based on these two surveys, at the second stage, a logic model is developed for demonstration effects for each type of financial instrument and validated with workshops with CDC and other development finance institutions. Logic models capture the set of factors that influence an outcome or set of outcomes. This is similar to a *theory of change* and may describe the causal links between activities and intended outcomes (ex-ante) or capture how particular outcomes were caused by a set of preceding activities (ex post).

In addition to the construction of the logic model, the research also develops Investor Sentiment Indices for each of CDC's major target markets from a combination of risk appetite and investor behaviour data, which are used to identify significant spikes in investor sentiment. Regular media monitoring of key markets and sectors and meetings with an Investor Focus Group are complementary tools used to contextualise the indices. This is a blended approach based on qualitative and quantitative datasets.

Finally, to assess the macroeconomic effects of the interventions, (i) an instrumental variable approach in a cross-country framework, and (ii) quasi-experimental estimations at country levels, including time-series analyses (notably structural vector autoregression – VAR) were favoured by the research team. These methodologies depend on available datasets, and extensive descriptive analysis was conducted prior to econometric estimations.

Although the methodology is suitable for the analysis of the macroeconomic effects of the interventions, VAR models rely on (i) long time series which are not always available in developing countries, and (ii) a limited number of variables. In addition, VAR models are somewhat atheoretical, even if structural VAR models aim to deal with identification purposes. These characteristics may limit the model's ability to measure the spillover effects of interventions, and in particular the heterogeneity of the effects among different groups of firms and types of households.

3.3 The experience from the World Bank Group (World Bank, 2016)

The World Bank Group has worked in different countries over the years to support the development of capital markets. It is worth noting that capital market development has received special attention only since March 2007. Before this date there have been few formal articulations of financial sector strategy at the World Bank.

The World Bank Group's interventions cover many segments including public sector and private corporate issuers. For instance, it has undertaken local currency bond issues through its Treasury departments, and support for the development of markets in asset-backed securities has been undertaken mainly through parts of its housing finance portfolio. In the same vein, the International Finance Corporation (IFC) has numerous private equity investments, which support the financing of small firms. The World Bank Group has a substantial portfolio in insurance and pensions. The IFC directly supported insurance companies, while the World Bank focused on new product development and risk management. Moreover, the World Bank and the IFC have made use of capital market instruments to finance their own real sector investments, through bond issues and guarantees.

In 2016, the World Bank Group undertook an evaluation of its interventions related to capital market development. The core purpose of the evaluation was to assess how well the World Bank Group supported its client countries in the development of their capital markets, across the full spectrum of activities that contribute to this. The overarching evaluation question was: *"Has the Bank Group been relevant, effective, and efficient in supporting the development of its client countries' domestic capital markets to deepen their financial systems, realise real sector development, and support the achievement of the Bank Group's twin goals of poverty alleviation and shared prosperity?"*

The evaluation examines the *relevance* of objectives and design, the *effectiveness* of outcomes and impact, and programme *efficiency*. *Relevance of objectives* refers to the extent to which the Bank Group's capital market interventions reflected prevailing financial sector knowledge and diagnostics. *Relevance of design* looks at the extent to which intervention focused on the right issues in the country and sector contexts. The *effectiveness* is evaluated in terms of the extent to which the Bank Group's interventions achieved their objectives, primary or secondary, relevant to capital market development, in terms of both immediate outputs and outcomes, for domestic capital market development or real sector support – and whether these results were sustained over time. Finally, *efficiency* refers to programme funding and sustainability, programme monitoring, tracking, and results measurement, internal and external coordination, and quality control.

The underlying *theory of change* is that all interrelated areas of capital markets and their surrounding environment together achieve the final output of market strengthening, more robust financial systems, supporting growth, and the reduction of poverty and inequality.

Evaluation questions were answered through a combination of methodologies including desk reviews of policy and strategy documents, theme-focused portfolio reviews based on customised questionnaires, and field visits.

Before the review of the Bank Group's capital markets portfolio, the joint IMF/World Bank Financial Sector Assessment Program is used to examine the extent to which there is an adequate and in-depth diagnostic basis for such interventions. Overall, the review of 39 of these assessments in 20 countries concludes that coverage of most areas relevant to capital market development received reasonable even if diminishing coverage.

Among the challenges faced was that about half to three-quarters of interventions were advisory services and for many interventions, capital market development was of secondary or indirect relevance. Even on the investments and lending side, there was limited evaluative material. An additional challenge was that activities such as insurance, pensions or housing provided indirect, or secondary, support to capital market development.

4 New approaches in impact assessment

This section is substantially based on Gertler et al. (2016) and Khandker et al. (2010)

Many policy questions involve cause-and-effect relationships. For example, does the development of capital markets increase wages and create more jobs in the domestic economy? Impact assessments seek to answer such cause-and-effect questions precisely.

Although cause and effect relationships are common in daily life, establishing that the observed effects are caused by the intervention alone can be challenging. We know that there is a close connection between financial markets systems and the real economy,³ simply observing an increase in access to finance, job creation and wages after interventions in capital market development is not enough to establish causality. Access to credit might have increased even if the government did not directly intervene in the financial markets – because of actions to sustain government budget, debt and deficit, hence creating trust in the economy, or because structural reforms to remove barriers that might affect access to credit such as credit bureaus, land reforms, etc. Impact assessments help to overcome the challenge of establishing causality by empirically proving to what extent a programme – and the programme alone – contributed to the change in an outcome.

In practice, different methods have been developed to rule out the possibility that any factors other than the programme of interest explain the observed impact. This section presents the econometric techniques used in the literature for micro assessment, on the one hand, and for macro impact evaluation, on the other hand. Before presenting the econometric techniques in the second and third subsections, the first subsection is devoted to main challenges and types of the impact assessment. Subsection 4.4 presents a summary table of the micro- and macro-impact evaluation techniques. Finally, subsection 4.5 provides an overview of the possible application of the studied techniques to some of FSD Africa's projects.

4.1 Impact assessment: Definition and challenges

The impact or causal effect of a programme (P) on an outcome of interest (Y) is given by the following formula:

$$\Delta = (Y|P = 1) - (Y|P = 0) \quad (1)$$

³ Recent studies show that financial uncertainty and real uncertainty move together (Jurado et al., 2015; Bloom et al., 2018). See also section 2.1 above where key references linking financial development and economic development have been highlighted.

This formula states that the impact (Δ) of a programme (P) on an outcome (Y) is the difference between the outcome (Y) *with* the programme (i.e. $P = 1$) and the same outcome (Y) *without* the programme (i.e. $P = 0$).

To evaluate this formula, one must measure the outcome at the same point in time for the same unit of observation, but in two different states of the world. It is impossible in the real world to measure the same unit in two different states (with and without the programme) at the same time, and that because at a given moment in time, a unit either participates in the programme or does not participate. The question is how to measure what would have happened if the other circumstance had prevailed? This is the so-called *counterfactual problem*. The basic idea is to estimate the unobserved term ($Y|P = 0$) in the formula (1) in order to evaluate the impact of the intervention. Solving the counterfactual problem would be possible if the evaluator could find a “perfect clone” for a programme participant. But it is known that perfect clone does not exist. Although no perfect clone exists for a single unit, we can rely on statistical properties to generate two groups – *treatment* ($P = 1$) and *control* ($P = 0$) groups – of units that, if their numbers are large enough, are statistically indistinguishable from each other at the group level. So, in practice, the challenge of an impact evaluation is to identify a treatment group and a comparison group that are statistically identical, on average, in the absence of the programme. Without a comparison group that yields an accurate estimate of the counterfactual, the true impact of a programme cannot be established.

Therefore, the main challenge for identifying impacts is to find a *valid comparison group* (or control group) that has the same characteristics as the treatment group in the absence of a programme. The comparison group must be similar in at least three ways to the treatment group to be valid:

- i. First, the average characteristics of the treatment group and the comparison group must be identical in the absence of the programme.
- ii. Second, the treatment should not affect the comparison group either directly or indirectly.
- iii. Third, the outcomes in the control group should change the same way as outcomes in the treatment group, if both groups were given the programme.

When these three conditions are met, then only the existence of the programme of interest will explain any differences in the outcome (Y) between the two groups. In the next subsection, various methods that can be used to construct valid comparison groups will be presented.

Let us present the basic evaluation problem comparing outcomes Y across treated and non-treated individuals i . For that, consider Equation (2):

$$Y_i = \alpha X_i + \beta P_i + \varepsilon_i \quad (2)$$

where P is a dummy equal to 1 for those who participate and 0 for those who do not participate. X is a set of other observed characteristics of the individual and perhaps of his or her household and local environment. Finally, ε is an error term reflecting unobserved characteristics that also affect Y . Equation (2) reflects an approach commonly used in impact evaluations, which is to measure the direct effect of the

programme P on outcome Y . The impact is measured by the parameter β . Indirect effects of the programme (that is, those not directly related to participation) may also be of interest. The subsection 4.2 covers different methods to measure direct programme effects. Indirect programme effects measurements are discussed in subsection 4.3.

Estimating the impact of the programme based on Equation (2) might be biased because treatment assignment is not often random due to (a) purposive programme placement and (b) self-selection into the programme. In other words, programmes are placed according to the need of the communities and individuals, who in turn self-select given programme design and placement. Self-selection could be based on observed characteristics, unobserved factors, or both. In the case of unobserved factors, the error term in the estimating equation will be correlated with the treatment dummy P (i.e. $\text{cov}(P, \varepsilon) \neq 0$). This violates one of the main assumptions for applying the ordinary least square to estimate Equation (2). This correlation biases the estimates of the equation, including the programme effect β .

From Equation (2), we can rewrite Equation (1) as follows:

$$\begin{aligned}\Delta &= E(Y_i(1)|P_i = 1) - E(Y_i(0)|P_i = 0) \\ &= E(Y_i(1)|P_i = 1) - E(Y_i(0)|P_i = 0) + [E(Y_i(0)|P_i = 1) - E(Y_i(0)|P_i = 1)]\end{aligned}$$

Which yields:

$$\Delta = \underbrace{E(Y_i(1)|P_i = 1) - E(Y_i(0)|P_i = 1)}_{\text{Average Treatment Effect (ATE)}} + \underbrace{[E(Y_i(0)|P_i = 1) - E(Y_i(0)|P_i = 0)]}_B \quad (3)$$

ATE stands for *Average Treatment Effect*, that is the average gain in outcomes of participants relative to nonparticipants, as if nonparticipating households were also treated. It corresponds to a situation in which a randomly chosen household from the population is assigned to participate in the programme, so participating and nonparticipating households have an equal probability of receiving the treatment P .

The term B is the selection bias, the extent of selection bias that crops up in using Δ as an estimate of the ATE. Because one does not know $E(Y_i(0)|P_i = 1)$, one cannot calculate the magnitude of selection bias. As a result, if one does not know the extent to which selection bias makes up Δ , one may never know the exact difference in outcomes between the treated and the control groups.

The basic objective of a sound impact assessment is to find ways to get rid of selection bias ($B = 0$) or to find ways to account for it. One approach to deal with it is to randomly assign the programme (see section 4.2.1). Another approach is grounded on the basis that the selection bias would disappear if one could assume that households or individuals receive treatment conditional on a set of covariates, X , were independent of the outcomes that they have. This assumption is called the *assumption of unconfoundedness*, also referred to as the *conditional independence assumption*:

$$(Y_i(1), Y_i(0)) \perp P_i | X_i \quad (4)$$

One can also make a weaker assumption of *conditional exogeneity of programme placement*. These different approaches and assumptions will be discussed in the following sub-sections. The soundness of the impact estimates depends on how

justifiable the assumptions are on the comparability of participant and comparison groups, as well as the exogeneity of programme targeting across treated and nontreated areas. However, without any approaches or assumptions, one will not be able to assess the extent of bias B .

4.2 Econometric techniques for micro assessment

This section reviews the different econometric techniques available for micro assessment: randomised control trial (RCT), instrumental variables (IV), Regression Discontinuity Design, Difference-in-Differences (Double Difference), Propensity Score Matching (PSM).

4.2.1 Randomised control trial

Randomised assignment of treatment or randomised control trial is considered the gold standard of impact evaluation. It uses a random process to decide who is granted access to the programme and who is not. Under randomised assignment, every eligible unit has the same probability of being selected for treatment by a programme.

When units are randomly assigned to treatment and comparison groups, this randomised assignment process in itself will produce two groups that have a high probability of being statistically identical. More precisely, with a large enough number of units, the randomised assignment process will produce groups that have statistically equivalent averages for all their characteristics, ensuring equivalence between the treatment and comparison groups in both observed and unobserved characteristics (*balancing hypothesis*).

Before the programme starts, and based on the baseline data from the evaluation sample, this balancing assumption can be empirically tested to verify that there are no systematic differences in observed characteristics between the treatment and comparison groups. Therefore, after launching the programme, if differences are observed in outcomes between the treatment and comparison groups, those differences can be explained only by the introduction of the programme, since by construction the two groups were identical at the baseline, before the programme started, and are exposed to the same external environmental factors over time. In this sense, the comparison group controls for all factors that might also explain the outcome of interest.

To estimate the impact of a programme under randomised assignment, we simply take the difference between the outcome under treatment (the mean outcome of the randomly assigned treatment group) and our estimate of the counterfactual (the mean outcome of the randomly assigned comparison group). We can be confident that our estimated impact constitutes the true impact of the programme since we have eliminated all observed and unobserved factors that might otherwise plausibly explain the difference in outcomes. Therefore, the selection bias has been eliminated, i.e. $B = 0$.

Moreover, randomised assignment of treatment ensures both the internal and the external validity of the impact estimates. *Internal validity* means that the estimated impact of the programme is net of all other potential confounding factors – that is the comparison group provides an accurate estimate of the counterfactual, so that we are estimating the true impact of the programme. An evaluation is *externally valid* if the

evaluation sample accurately represents the population of eligible units. The results of the evaluation can then be generalised to the population of eligible units. The randomisation process implies random selection of a sample for *external validity* and is used as randomised assignment of treatment as an impact evaluation method for *internal validity*.

Nevertheless, there is a lack of internal validity in many RCTs as shown in the recent experiment by Bulte et al. (2020). This issue derives from the fact that RCTs in economics are not double-blind, unlike RCTs in medicine. Double-blinding is the preferred approach to eliminate placebo effects, or other effects that would bias the average treatment effect. The challenge in economics is that if subjects know their treatment status, they may adjust their behaviour in order to fully benefit from the treatment. While the analyst randomises subjects into treatment, she therefore cannot assume that all relevant covariates are held constant across treatment arms. Disentangling the overall treatment effect that is measured into its behavioural component and the direct effect of the intervention is difficult.

In their experiment, Bulte et al. (2020) consider the provision of improved seed varieties to smallholders in Tanzania. They organised a study where half of the farmers participated in a standard RCT and the other half participated in a double-blind RCT. They find that while standard RCTs found large treatment effects, double-blind RCTs revealed that a large share of this impact is due to farmers allocating extra effort and their best plots to the cultivation of new seeds.

Randomised assignment can be used as a programme allocation rule when the eligible population (demand for the program) is greater than the number of programme spaces available (supply side). Randomisation can also be used when a programme needs to be gradually phased in until it covers the entire eligible population. When a programme is phased in, randomisation of the order in which participants receive the programme gives each eligible unit the same chance of receiving treatment in the first phase or in a later phase of the programme. As long as the last group has not yet been phased into the programme, it serves as a valid comparison group from which the counterfactual for the groups that have already been phased in can be estimated. This setup can also allow for the evaluation to pick up the effects of differential exposure to treatment: that is, the effect of receiving a programme for more or less time.

RCTs have been extensively used in evaluating projects' impacts in financial sector development, including the microfinance and insurance sectors. For example, Belissa et al. (2019) carry out an experiment in Ethiopia to examine whether uptake of index-based insurance is enhanced if we allow farmers to pay after harvest (addressing a liquidity constraint). They test to what extent uptake can be enhanced by promoting insurance via informal risk-sharing institutions (*Iddirs*). They find that delaying the payment of insurance premium increases uptake substantially when compared to standard insurance. Moreover, promoting this new product via *Iddirs* results in even greater uptake. In addition, they find suggestive evidence that the delayed premium product is indeed better at targeting the liquidity constrained. However, default rates associated with delayed payments are relatively high and concentrated in a small number of *Iddirs* – potentially compromising the economic viability of the new insurance product.

RCTs have also been used by the Dutch Authority for the Financial Markets (see AFM, 2019) to assess the extent to which retail investors are able to identify and choose the best investment using different types of mandatory disclosure. In fact, when securities are offered to the public or admitted to trading on a regulated market, a prospectus often needs to be published. For retail investors, this disclosure also must be summarised in a summary prospectus and in a key information document (KID). The objective of the study was to assess the effectiveness of the summary prospectus, the KID, and proposed combination of these two documents. The key finding is that Dutch retail investors objectively allocate assets better using the KID than with the summary prospectus.

A randomisation design is subject to several concerns, including ethical issues, political sensitivity, external validity, partial or lack of compliance, selective attrition, and spillovers.

- **Ethical issues:** Withholding a treatment from a random group of people and providing access to another random group of people may be simply unethical under some circumstances.
- **Political sensitivity:** Carrying out randomised design is often politically unfeasible because justifying such a design to people who might benefit from it is hard.
- **External validity:** A project of small-scale may not achieve aggregate macroeconomic effects, whereas a large-scale project might. That is, impact measured by the pilot project may not be an accurate guide to the project's impact on a national scale. The problem is how to generalise and replicate the results obtained through randomised evaluations.
- **Compliance:** This arises when a fraction of individuals who are offered the treatment do not take it, and/or, some members of the comparison group may receive the treatment. This situation is referred to as partial (or imperfect) compliance and may also be a problem with randomisation. To be valid and to prevent selection bias, the analysis needs to focus on groups created by the initial randomisation. The analysis cannot exclude subjects or cut the sample according to behaviour that may have been affected by the random assignment. More generally, interest often lies in the effect of a given treatment, but the randomisation affects only the probability that the individual is exposed to the treatment, rather than the treatment itself.

In the absence of full compliance in the treatment group, the estimated impact Δ is called the *intention-to-treat* (ITT) when comparing groups to which the programme has randomly been *offered* (in the treatment group) or not (in the comparison group) regardless of whether or not those in the treatment group actually enrol in the programme. The ITT is a weighted average of the outcomes of participants and nonparticipants in the treatment group compared with the average outcome of the comparison group. The ITT is important for those cases in which we are trying to determine the average impact of offering a programme, and enrolment in the treatment group is voluntary. By contrast, we might also be interested in knowing the impact of a programme for the group of individuals who are offered the programme and actually participate. This estimated impact is called the *treatment-on-the-treated* (TOT). The ITT and TOT will be the same when there is full compliance.

- **Spillover effects:** Potential spillover effects arise when treatment helps the control group as well as the sample participants, thereby confounding the estimates of the programme impact. This can happen for example, if people outside the sample move into a village where health clinics have been randomly established, thus contaminating programme effects.

As stated above, RCT is the gold standard of impact evaluation. It appears to be the perfect impact assessment method in medical sciences but not necessarily in social sciences. Nevertheless, it remains an excellent methodological development as its prominence culminated with the award of the 2019 Nobel Prize in economic sciences to Abhijit Banerjee, Esther Duflo and Michael Kremer. However, the identification issue in social sciences research is not really settled even though reasoned intuition and parametric techniques (e.g. Instrumental Variables) are still valid (Basu, 2014). While RCT can establish “*circumstantial causality*”, it is not able to demonstrate *generalised causality*” (*ibid*).

4.2.2 Instrumental Variables

In the discussion of randomised assignment, it was assumed that the programme administrator has the power to assign units to treatment and comparison groups, and units that are assigned to each group comply with their assignment. However, in real-world, it might be unrealistic to think that the programme administrator will be able to ensure full compliance with the group assignment. In practice, many programmes allow potential participants to choose to enrol and thus are not able to exclude potential participants who want to enrol. In addition, some programmes have a budget that is big enough to supply the programme to the entire eligible population immediately, so that randomly assigning people to treatment and comparison groups and excluding potential participants for the sake of an evaluation would not be ethical. In such circumstances, we need alternative ways to evaluate the impact of these kinds of programmes.

Instrumental variables (IV) can help to evaluate programmes with imperfect compliance, voluntary enrolment, or universal coverage.⁴ Generally, to estimate impact, the IV method relies on some external source of variation to determine treatment status. Intuitively, we can think of an IV as something outside the control of the individual that influences her likelihood of participating in a programme but is otherwise not associated with her characteristics. In other words, the instrumental variable method relies on some external source of variation to determine treatment status. Therefore, an instrumental variable influences the likelihood of participating in a programme but is outside of the participant’s control and is unrelated to the participant’s characteristics. Formally, an instrument Z must be correlated with P (i.e. $\text{cov}(Z, P) \neq 0$) but uncorrelated with ε (i.e. $\text{cov}(Z, \varepsilon) = 0$).

IV approach requires two-stage approach to estimate Equation (2). At the first stage, known as *first-stage regression*, the treatment variable is regressed on the instrument Z and the other covariates X as follows:

$$P_i = \gamma Z_i + \phi X_i + u_i \quad (5)$$

The predicted treatment from this regression, \hat{P} ($\equiv \hat{\gamma} Z_i + \hat{\phi} X_i$), therefore reflects the part of the treatment affected only by Z and thus embodies only exogenous variation

⁴ IV method has a wide range of applications beyond impact evaluation.

in the treatment. \hat{P} is then substituted for treatment in Equation (2) to create the following reduced-form outcome regression (*second-stage regression*):

$$Y_i = \alpha X_i + \beta \hat{P}_i + \varepsilon_i \quad (6)$$

The IV estimate of the programme (*two-stage least squares*) is $\hat{\beta}_{IV}$ and, the programme impact can be written as follows:

$$\beta_{IV} = \frac{\text{cov}(Y, Z)}{\text{cov}(P, Z)} \quad (7)$$

Equation (7) shows that the estimate depends on the instrument. To highlight the importance of the instrument, we can show that:

$$\beta_{IV} = \beta + \frac{\text{cov}(Z, \varepsilon)}{\text{cov}(P, Z)} \quad (8)$$

Three main conclusions can be drawn from Equation (8). First, if the instrument is correlated with the unobserved factors affecting the outcome ($\text{cov}(Z, \varepsilon) \neq 0$), the estimates of the programme effect will be biased. Therefore, the exogeneity of the instrument is key. However, in practice, it is difficult to find a good instrument and the exogeneity of an instrument cannot be tested in a homogenous effects model when the model is *just identified* (i.e. the number of instruments is equal to the number of endogenous variables). Even the Durbin-Hausman-Wu test which checks whether the OLS and IV estimates are the same does not tell us if the instrument is invalid. Recent research has developed test procedures for instrument validity in the heterogenous treatment effect to ensure non-bias estimates of programme impact (see Huber and Mellace, 2015; Kitagawa, 2015; Mourifié and Wan, 2017). *Over-identification* is commonly tested by using Sargan and Hansen statistics.

Second, if the instrument is only weakly correlates with the treatment variable P , the standard error of the IV estimate is likely to increase because the predicted impact on the outcome will be measured less precisely. Consistency of the IV estimate (that is, asymptotic bias) is also likely to be large when Z and P are weakly correlated ($\text{cov}(P, Z) \approx 0$), even if the correlation between Z and ε is low. Weak instruments can be tested by using a range of statistics (see Stock and Yogo, 2005 and Kleibergen, 2007 for presentation of the main statistics).

Third, the IV estimate of the impact of an intervention is a *local average treatment effect (LATE)*. That is the treatment effect only for those who decide to participate because of a change in the instrument Z (Imbens and Angrist, 1994). The LATE avoids the problem of unobserved forecasting of programme gains by limiting the analysis to individuals whose behaviour is changed by local changes in Z in a way unrelated to potential outcomes. However, LATE needs to be interpreted carefully, as it represents programme effects for only a specific subgroup of the population.

When conditional exogeneity does not hold, one can estimate the marginal treatment effect, which is a limited form of the LATE (see Carneiro et al., 2010; Heckman and Vytlacil, 2005; Moffitt, 2008; Todd, 2007).

On the application of the technique, it is worth noting that instrumental variables have been used extensively to deal with endogeneity in empirical research. For example, in a recent paper, Adler et al. (2019) rely on the instrumental-variables panel

approach to study the effect of foreign exchange intervention on the exchange rate. They find that a purchase of foreign currency of 1 percentage point of GDP causes a depreciation of the nominal and real exchange rates in the ranges of [1.7-2.0] percent and [1.4-1.7] percent respectively. The effects are found to be quite persistent and there are no asymmetric effects, meaning that positive and negative interventions are equally effective. This last result indicates that interventions in the foreign exchange market is a valid policy instrument both when facing appreciation and depreciation pressures. Similarly, ElBannan (2017) examines the extent to which capital structure decisions are affected by capital markets, in particular security liquidity. She uses the two-stage least squares (2SLS) regressions approach and lagged explanatory variables as instruments to address the endogeneity problem between liquidity and leverage. The study concludes that liquid equity has no effect on firm leverage.

4.2.3 Regression discontinuity design

Social programmes often use an index to decide who is eligible to enrol in the programme and who is not. An eligibility threshold or cut-off is set so that, for example, people below the threshold are eligible for the programme and those above are not. In a nonexperimental setting, such programme eligibility rules can sometimes be used as instruments for exogenously identifying participants and nonparticipants. To establish comparability, one can use participants and nonparticipants within a certain neighbourhood of the eligibility threshold as the relevant sample for estimating the treatment impact. Known as *regression discontinuity* (RD), this method allows observed as well as unobserved heterogeneity to be accounted for. Although the cut-off or eligibility threshold can be defined nonparametrically, the cut-off has in practice traditionally been defined through an instrument.

To model the effect of a particular programme on individual outcomes y_i through an RD approach, one needs a variable S_i that determines programme eligibility (such as age, asset holdings, or the like) with an eligibility cut-off s^* . The estimating equation is

$$y_i = \beta S_i + \varepsilon_i \quad (9)$$

where individuals with $S_i \leq s^*$, for example, receive the programme, and individuals with $S_i > s^*$ are not eligible to participate. Individuals in a narrow band above and below s^* need to be “comparable” in that they would be expected to achieve similar outcomes prior to programme intervention.

If one assumes that limits exist on either side of the threshold s^* , the impact estimator for an arbitrarily small $\epsilon > 0$ around the threshold would be the following:

$$E[y_i|s^* - \epsilon] - E[y_i|s^* + \epsilon] = E[\beta S_i|s^* - \epsilon] - E[\beta S_i|s^* + \epsilon] \quad (10)$$

Taking the limit of both sides of Equation (10) as $\epsilon \rightarrow 0$ would identify β as follows:

$$\beta = \frac{y^- - y^+}{S^- - S^+} \quad (11)$$

where $y^- = \lim_{\epsilon \rightarrow 0} E[y_i|s^* - \epsilon]$, $y^+ = \lim_{\epsilon \rightarrow 0} E[y_i|s^* + \epsilon]$, $S^- = \lim_{\epsilon \rightarrow 0} E[S_i|s^* - \epsilon]$ and $S^+ = \lim_{\epsilon \rightarrow 0} E[S_i|s^* + \epsilon]$.

In practice, the determination or enforcement of eligibility may not be “sharp”. In this case, S can be replaced with a probability of participating $\mathbb{P}(S) = E(P|S)$, where $P = 1$ if treatment is received and $P = 0$ otherwise. This is the case of *fuzzy* or *stochastic* discontinuity.

RD design can be used for programmes that have a continuous eligibility index with a clearly defined eligibility threshold (cut-off score) to determine who is eligible and who is not. To apply a RD design, the following main conditions must be met:

- i. The index must rank people or units in a continuous or “smooth” way.
Indexes like poverty scores or test scores have many values that can be ordered from small to large, and therefore they can be considered smooth. By contrast, variables that have discrete or “bucket” categories that have only a few possible values or cannot be ranked are not considered smooth.
- ii. The index must have a clearly defined cut-off score: that is, a point on the index above or below which the population is classified as eligible for the programme.
- iii. The cut-off must be unique to the programme of interest; that is, there should be no other programmes, apart from the programme to be evaluated, that uses the same cut-off score.
- iv. The score of a particular individual or unit cannot be manipulated by enumerators, potential beneficiaries, programme administrators, or politicians.

The advantages of the RD method are that: (a) it yields an unbiased estimate of treatment effect at the discontinuity; (b) it can, many times, take advantage of a known rule for assigning the benefit that is common in the design of social policy; and finally, (c) a group of eligible households or individuals need not be excluded from treatment.

The concerns with RD are that: (a) it produces local average treatment effects that are not always generalisable; (b) the effect is estimated at the discontinuity, so, generally, fewer observations exist than in a randomised experiment with the same sample size; and (c) the specification can be sensitive to functional form, including nonlinear relationships and interactions. In addition, programme officials may not always know precisely the eligibility criteria; hence, behavioural responses to the programme intervention may be confused with actual targeting rules (Ravallion, 2007). Moreover, if the eligibility rules are not adhered to or change over time, the validity of the discontinuity approach also needs to be examined more carefully.

Using the regression discontinuity approach, Kuersteiner et al. (2018) investigates the effectiveness of sterilised foreign exchange interventions in Colombia. The authors exploit a discontinuous policy rule used by the Central Bank of Colombia to identify the surprise component of rule-based interventions and use this variation to measure how they affect exchange rates and capital flows. The results indicate that interventions had significant effects on the exchange rate, albeit short-lived (2–3 weeks). In addition, capital controls amplify the effect of interventions. A methodological contribution of the paper is the extension of regression discontinuity designs to a time-series environment and to show how these techniques can be used to identify and estimate local non-linear impulse response functions.

On the capital market side, Chang et al. (2015) use the regression discontinuity to assess the impact of indexing on price. They focus on the Russell 1000 list which contains the list of the top 1000 most well capitalised companies. They find that stocks from the Russell 1000 that just landed in the Russell 2000 – that is the top 2000 most well capitalised companies – have discontinuously higher returns by about 5 percentage points compared with stocks that just missed making it into the Russell 2000. In addition, stocks whose market capitalisation placed them above the 1000 cut-off (and hence moved into the Russell 1000) have lower returns by about 5.4 percentage points than stocks just below the 1000 cut-off (that stayed in the Russell 2000).

4.2.4 Difference-in-differences (double-difference)

The *difference-in-differences (DD)* method compares the changes in outcomes over time between a population that is enrolled in a programme (the treatment group) and a population that is not (the comparison group). That is, given a two-period setting where $t = 0$ before the programme and $t = 1$ after programme implementation, letting Y_t^T and Y_t^C be the respective outcomes for programme beneficiary and nontreated units in time t , the DD method will estimate the average programme impact as follows:

$$DD = E(Y_1^T - Y_0^T | P_1 = 1) - E(Y_1^C - Y_0^C | P_1 = 0) \quad (12)$$

where $P_1 = 1$ denotes treatment or the presence of the programme at $t = 1$, whereas $P_1 = 0$ denotes untreated areas.

The DD estimator allows for *unobserved heterogeneity* (the unobserved difference in mean counterfactual outcomes between treated and untreated units) that may lead to selection bias. More specifically, DD assumes this unobserved heterogeneity is time-invariant, so the bias cancels out through differencing.

When baseline data are available, one can thus estimate impacts by assuming that unobserved heterogeneity is time invariant and uncorrelated with the treatment over time. This assumption renders the outcome changes for a comparable group of nonparticipants (that is, $E(Y_1^C - Y_0^C | P_1 = 0)$) as the appropriate counterfactual, namely, equal to $E(Y_1^T - Y_0^T | P_1 = 1)$.

In a regression framework, the DD estimator is given by the parameter β – the coefficient on the interaction between the post-programme treatment variable (P_{i1}) and time ($t = 1, \dots, T$) – as given in the following regression equation

$$Y_{it} = \alpha + \beta P_{i1} t + \rho P_{i1} + \gamma t + \varepsilon_{it} \quad (13)$$

Basic assumptions such as (a) the model is correctly specified, (b) $cov(\varepsilon_{it}, P_{i1}) = 0$, (c) $cov(\varepsilon_{it}, t) = 0$ and (c) $cov(\varepsilon_{it}, P_{i1}t) = 0$ must hold for the DD estimator to be valid. Assumption (c), known as *the parallel-trend assumption*, is the most critical. It means that unobserved characteristics affecting programme participation do not vary over time with treatment status.

An example of the application of this method is done by Belissa et al. (2020) who study whether adopting index insurance improves access to financial markets and reduces credit rationing in Ethiopia. They differentiate between credit rationing from supply side and that from demand side. From the supply-side, the credit rationing

implies that potential borrowers who need credit are involuntarily excluded from the credit market, while the demand-side credit rationing refers to borrowers who self-select and voluntarily withdraw from the credit market to avoid transaction costs and threats to their collateral. The authors find that 38% of sample households are credit constrained and that index insurance significantly reduces supply-side credit rationing.

Equation (13) can be generalised with multiple time periods, which may be called the *panel fixed-effects model*. This is particularly important for a model that controls not only for the unobserved time-invariant heterogeneity but also for heterogeneity in observed characteristics over a multiple-period setting. More specifically, Y_{it} can be regressed on P_{it} , a range of time-varying covariates X_{it} , and unobserved time-invariant individual heterogeneity η_i that may be correlated with both the treatment and other unobserved characteristics ε_{it} . One can consider the following revision of Equation (13):

$$Y_{it} = \phi P_{it} + \delta X_{it} + \eta_i + \varepsilon_{it} \quad (14)$$

Equation (14) may be estimated by using *first difference, within or dummy variables* transformations. Although the first difference is consistent, one needs to account for serial correlation (see Bertrand et al., 2004).

Berck and Villas-Boas (2016) argue that when the outcome variable is determined by policy, time, place and another variable, a triple difference strategy is better than the double difference strategy to estimate the effect of the policy change. This technique has been used by Hamermesh and Trejo (2000), Kellogg and Wolf (2008) and McIntosh (2008), among many others.

The advantage of DD is that it relaxes the assumption of conditional exogeneity or selection only on observed characteristics. It also provides a tractable way to account for selection on unobserved characteristics. The main drawback is the assumption of time-invariant selection bias which is implausible for many targeted programmes in developing countries.

4.2.5 Propensity score matching

Propensity score matching (PSM) constructs a statistical comparison group that is based on a model of the probability of participating in the treatment, using observed characteristics X , that is $\mathbb{P}(X) = \mathbb{P}(P = 1|X)$. Participants are then matched based on this probability, or propensity score, to nonparticipants. Rosenbaum and Rubin (1983) show that, under certain assumptions, matching on $\mathbb{P}(X)$ is as good as matching on X . The necessary assumptions for identification of the programme effect are: (a) conditional independence and (b) presence of a common support.

Conditional independence states that given a set of observable covariates X that are not affected by treatment, potential outcomes Y are independent of treatment assignment P . This is given by Equation (4) above. This assumption, known as *unconfoundedness*, implies that uptake of the programme is based entirely on observed characteristics. Conditional independence is a strong assumption and is not a directly testable criterion; it depends on specific features of the programme itself. If unobserved characteristics determine programme participation, conditional independence will be violated, and PSM is not an appropriate method.

A second assumption is the *common support or overlap condition*: $0 < \mathbb{P}(P_i = 1|X_i) < 1$. This condition ensures that treatment observations have comparison observations “nearby” in the propensity score distribution. Specifically, the effectiveness of PSM also depends on having a large and roughly equal number of participant and nonparticipant observations so that a substantial region of common support can be found.

If conditional independence holds, and if there is a sizable overlap in $\mathbb{P}(X)$ across participants and nonparticipants, the PSM estimator for the TOT can be specified as the mean difference in Y over the common support, weighting the comparison units by the propensity score distribution of participants. A typical cross-section estimator can be specified as follows:

$$TOT = E_{\mathbb{P}(X)|P=1}\{E(Y^T|P = 1, \mathbb{P}(X)) - E(Y^C|P = 0, \mathbb{P}(X))\} \quad (15)$$

To calculate the programme treatment effect, one must first calculate the propensity score $\mathbb{P}(X)$ on the basis all observed covariates X that jointly affect participation and the outcome of interest. This is done through an estimation of logit or probit model. The aim of matching is to find the closest comparison group from a sample of nonparticipants to the sample of programme participants. “Closest” is measured in terms of observable characteristics not affected by programme participation. Different algorithms can be used for the matching: *nearest-neighbour* matching, *calliper* or *radius* matching, *stratification* or *interval* matching, *kernel* or *local* matching, *difference-in-difference* matching (only when data on participant and control observations is available before and after programme intervention). Before the matching, the region of common support should be defined, and balancing tests should be conducted to ensure the quality of the matching.

PSM has been used to evaluate projects in different sectors, including education, labour markets, the financial sector, among many others. In the financial sector, Fatum and Hutchison (2010) use this technique to assess the effect of official foreign exchange market intervention on the exchange rate movement. A direct measurement will be biased since that intervention entails a “self-selection” choice made by the authorities. An estimate of the “counterfactual” exchange rate movement in the absence of intervention is provided by the method of propensity score matching based on the central bank intervention reaction functions. The authors provide the ATE of the intervention and conclude that only sporadic and relatively infrequent intervention is effective.

Similarly, Oh et al. (2009) use the propensity score matching to estimate the effect of credit guarantee policy on growth rates of different performance indicators including productivity, sales, employment, investment, R&D, wage level, and the survival of firms post-crisis. PSM is used to avoid selection problems. Oh et al. find that credit guarantees significantly influenced firms’ ability to maintain their size, and increase their survival rate, but not to increase their R&D and investment and hence, their growth in productivity.

While PSM is a popular technique to process data for causal inference in order to reduce imbalance in pre-treatment covariates between the treated and control groups, thereby reducing the degree of model dependence and bias, King and Nielsen (2019) argue that PSM often accomplishes the opposite of what is intended, i.e., increasing imbalance, inefficiency, model dependence, research discretion and

bias. Hence, results based on this method should be taken with some caution. In addition, as stated above, PSM relies on observable characteristics, and does not account for unobserved factors. This is another limitation of the technique.

4.3 Econometric techniques for macro assessment

Economic models can help in understanding the potential interactions and interdependence of a programme with other existing policies and/or individual behaviour. Unlike reduced-form estimations, which focus on a one-way, direct relationship between a programme intervention and ultimate outcomes for the targeted population, structural estimation approaches explicitly specify interrelationships between endogenous variables (such as household outcomes) and exogenous variables or factors. *Structural approaches* can help create a schematic for interpreting policy effects from regressions, particularly when multiple factors are at work. These models specify interrelationships among endogenous variables (such as outcomes Y) and exogenous variables or factors. One simple structural model is a system of two equations as follows:

$$\begin{cases} Y_{1i} = \alpha_1 + \beta_1 Y_{2i} + \rho_1 Z_{1i} + \varepsilon_{1i} \\ Y_{2i} = \alpha_2 + \beta_2 Y_{1i} + \rho_2 Z_{2i} + \varepsilon_{2i} \end{cases} \quad (16)$$

where Y_1 and Y_2 are endogenous variables (outcomes) and Z s are exogenous variables. Since the treatment is assumed to be exogenous, it must be included in one of the Z s.

A system of equations can be used to estimate the effects of policies. The system can even be extended to account for dynamics or persistence. However, the evaluations of policies are very much context specific. Therefore, it is impossible to come out with a single approach for modelling policy effects in an economic framework. In practice, the evaluation of economic policies calls on different tools for questions of allocation (of resources to different possible uses), stabilisation (reduction of deviations from equilibrium) and redistribution (changes in income distribution).

Modelling the effects of macroeconomic policies such as taxes, trade liberalisation, or financial regulation can be very complex, because these policies are likely to be concurrent and to have dynamic effects on household and firm behaviour. Economic shocks such as commodity price increases or liquidity constraints stemming from the recent global financial crisis also jointly affect the implementation of these policies and household outcomes; the distributional impacts of these shocks also depend on the extent to which heterogeneity among economic agents is modelled (Essama-Nssah, 2005).

Studies have constructed general equilibrium models – e.g., *computable general equilibrium* (CGE), *dynamic stochastic general equilibrium* (DSGE), *agents-based models* (ABM), *heterogeneous agent new Keynesian* (HANK) – to examine the effects of macroeconomic policy changes and shocks on the behaviour of economic agents (households and firms) across economic sectors.⁵ In some cases, authors conduct counterfactual analysis to isolate the effects of specific shocks. We must

⁵ Pereira da Silva et al. (2008) and Essama-Nssah (2005) provide a discussion of different macroeconomic models. ABM models have been developed after the 2007 financial crisis (see Geanakoplos et al., 2012) similar to HANK (Guerrieri and Lorenzoni, 2017; Oh and Reis, 2012; McKay and Reis, 2016).

keep in mind that the type of model to be constructed and used and its adequacy depends entirely on the tasks envisaged for the model. In economic modelling, one size does not fit all: a model can be good at addressing some issues but inappropriate for addressing others. In the next subsection, we describe these macro-assessment models in more details. In addition to these general equilibrium models, we also present the *synthetic controls*, a new evaluation method, as well as the *vector autoregression* (VAR) model.

4.3.1 Computable general equilibrium models

Typical CGE models describe the motivations and behaviour of all consumers and producers in an economy and their linkages. It highlights the effects on the level and composition of industry activity and employment, trading performance, household incomes and consumption, macro-economic outcomes and the government's budgetary position of: (i) policies and changes occurring at sectoral level (industrial sector or commodity sector) or at the household level; (ii) policies and changes which occur at the macro-economic level such as changes in the aggregate level of government spending, a change in the exchange rate, or a change in the economy-wide real wage; and (iii) changes external to the domestic economy but to which it must adjust. The model comprises of two distinct set of variables (exogenous and endogenous variables) as well as market clearing constraints. The exogenous variables are provided by the user and the values of endogenous variables are determined as solutions to the equations of the model. The solutions of the equations at equilibrium are the set of prices which make the quantities of supply and demand equal in every market.

For many years, CGE models have been extensively used to assess the impact of trade policies, of providing anti-retroviral treatment for HIV/AIDS patients, of labour market policies on employment, among many other issues (see Decaluwé and Martens, 1988 for a review). CGE models have been extended to include the financial sector and to analyse issues such as financial fragility. In their recent publication, Beyers et al. (2020) use CGE model to examine the impact of banking regulations on banks and their consumers and apply it as a risk assessment tool in South Africa. They find that the effect of a rise in the default penalty was higher than that of the capital requirements infringement penalty. Moreover, banks react differently depending on whether the Reserve Bank sets its base money or the interbank rate as its monetary policy instrument. A similar exercise has been performed by Haqiqi and Mirian (2015) to investigate the linkage between natural resources and the financial sector in resource-abundant countries such as Iran. A positive resource shock affects the level of government intervention in the financial sector: the government offers low interest loans; it purchases equities and bonds and invests in production sectors.

4.3.2 Dynamic stochastic general equilibrium models

One advantage of CGE models is the level of disaggregation which depends, nevertheless, on the social accounting matrix containing the input data. Newer ideas and techniques make CGE models challenging. For instance, the combination of dynamic programming with explicit representation of uncertainty is computationally challenging. These two ideas are united in the area of DSGE (Horridge et al., 2013). DSGE models are built based on the inter-temporal optimisation of the utility of

economic agents. Thus, they are commonly used for the evaluation of allocation and stabilisation policies.

Basic DSGE models capture elements of the New Keynesian paradigm, of the New Classical school and of the Real Business Cycle approach, with several features of apparently irreconcilable traditions of macroeconomic thought. It is a fully dynamic coherent micro-founded environment that helps to better understand the transmission mechanisms of policy interventions and of shocks. In this environment, it should be possible to escape the Lucas critique, in contrast to the traditional macro-econometric models in which the estimated parameters are not invariant to policy shifts or to expected policy changes. DSGE models are flexible and shed new light on the linkages among monetary, financial and fiscal policy, inflation and the business cycle.

Following the 2007 financial crisis,⁶ DSGE models have been extended to consider the financial sector, and more precisely financial frictions and the balance sheet of the financial intermediaries (see Gertler and Kiyotaki, 2010; Meh and Moran, 2010; Gerlali et al., 2010; Gertler and Karadi, 2011; Curdia and Woodford, 2011; Verona et al., 2013; among many others). For example, Verona et al. (2013) simulated anticipated and unanticipated monetary policies in a model with bond financing via a shadow banking system. They show that the US boom-bust was caused by the combination of (i) too low for too long interest rates, (ii) excessive optimism, and (iii) a failure of agents to anticipate the extent of the abnormally favourable conditions. Similarly, Miao et al. (2015) estimate a DSGE model to study the link between stock market bubbles and business cycles. In the model, bubbles emerge through a positive feedback loop mechanism supported by self-fulfilling beliefs. They identify a sentiment shock that drives the movements of bubbles. This shock explains most of the stock market fluctuations and is transmitted to the real economy through endogenous credit. Because of that channel of transmission, bubbles are responsible of sizable fractions of the variations in real quantities.

Related to financial reforms, Funke and Paetz (2012) use a DSGE model to evaluate various financial system reform initiatives – phasing out benchmark interest rates, deepening the direct finance market, reducing government's quantity-based intervention on financial institutions – on the design of monetary policy in China. They conclude that the reforms will be beneficial only if Chinese monetary policy continues to rely on quantity-based interventions on financial institutions or tightens the interest rate rule.

In the same vein, Cantú and Chui (2020) show that “original sin redux” leads to a welfare improvement over “original sin” when there are global shocks by using a two-country DSGE model. The “original sin” refers to a situation in which a country cannot borrow abroad using its own currency. This creates a currency mismatch on emerging market economies’ (EME) balance sheets, triggers financial constraints and negatively affecting output after an increase in global interest rates. The development of local currency government bond helps mitigate the impact on EME output and investment of a global shock, as the currency mismatches are on foreign balance sheets. But as EME borrowers still rely on foreign banks, they are not fully

⁶ DSGE models were subject to very violent criticisms following the financial crisis of 2008 (see Galí, 2018 on this issue): the inability of DSGE models to predict the crisis; a weakness often attributed to the absence of the financial sector in the model which would have made it possible to explain the key factors of the crisis; the assumptions of rational expectations; the perfect information framework and the infinite lifespan of the agents (households) of the model.

shielded from the tightening in global financial conditions: this is termed “original sin redux”.

Other extensions of DSGE models include: (i) the relaxation of the hypothesis of rational expectations (Woodford, 2010); (ii) labour market friction (Trigari, 2009; Blanchard and Galí, 2010); (iii) the relaxation of the hypothesis of representative agents to reflect the heterogeneity among economic agents, known as HANK; (iv) information asymmetry (Boissay et al., 2016); (v) bubbles (Galí, 2020) and (vi) non-linearity (Basu and Bundick, 2017).

4.3.3 Agent-based models

Agent-based models (ABM) use computer simulation to explore emerging dynamical patterns, free from any top-down assumptions.⁷ In contrast to conventional models, ABM make no assumptions about the existence of efficient policies or general equilibrium. These may or may not emerge due to the dynamical rules. Big fluctuations and even crashes are often inherent features, not shocks coming from the outside to upset the system’s normal state of equilibrium as in conventional DSGE models.

Agent-based models use a dynamic system of interacting, autonomous agents to allow macroscopic behaviour to emerge from microscopic rules. The models specify rules that dictate how agents will act based on various inputs. Each agent individually assesses its situation and makes decisions based on its rules. The agents are heterogeneous and can act with some degree of independence. At the start of each time period, each agent observes its environment and acts accordingly. The agent’s environment is only a local view of the overall system. The agents’ actions change the environment. In the next period, each agent sees its new environment, altered based on the actions of the previous period, and acts again accordingly. Thus, there is an interaction between agents and the environment, and between one agent and another.

Agent-based models allow for more descriptive richness, as they describe ecologies of agents, locally interacting through non-obvious network structures, learning using incomplete information, and competing within imperfect markets (Fagiolo et al., 2019). Since agents are heterogeneous, they may widely differ in their utility functions, goals, views, resources, etc. This makes them especially useful to study scenarios or cases where different groups of stakeholders are involved. The system does not need to be populated with representative agents such as identical decision-makers, firms or governments whose individual behaviour mirrors the system. In models that produce equilibrium states, the dynamics come to an end. In social systems, on the other hand, there are a large number of different equilibrium states between which the system erratically jumps, a phenomenon known as punctuated equilibrium.

Agent-based models are used to study complex systems, including ecosystems, pandemics, markets, finance, energy generation and distribution, weather and climate, as well as societal phenomena such as urbanisation, traffic flows and migration. In finance,⁸ Lauretta (2018) uses an agent-based model to investigate the interaction between financial innovation and securitisation. Financial innovation,

⁷ See the New Approaches to Economic Challenges Initiative for a full description and application to the labour market (NAEC, 2017).

⁸ See Botta et al. (2019) and Mazzocchetti et al. (2020) for other applications of agent-based models in finance.

measured by the rate of financial innovation, captures the level of development of financial tools, processes and services, given the financial operators' business decisions on how to make use of them. The study postulates that financial innovation and securitisation have increased macro-finance systemic instability and can generate two business cycles. The author shows that high values of financial innovation harm the economic system. Therefore, the rate of financial innovation can trigger financial and economic instability and should be monitored closely by financial regulators.

4.3.4 Heterogeneous agent new Keynesian models⁹

HANK models address shortcomings of representative agent models highlighted during the financial crisis which originated from housing and credit markets. The collapse in house prices affected households differently, depending on the composition of their balance sheets. The extent to which this wealth destruction translated into a fall in expenditures was determined by marginal propensities to consume, which are also very heterogeneous and closely related to households' access to liquidity (Mian et al., 2013). This drop in aggregate consumer demand and the contemporaneous breakdown in bank lending to businesses resulted in a severe contraction of labour demand, which materialised unevenly across different occupations and skill levels. All these took place against the backdrop of a secular rise in income and wealth inequality. Thus, portfolio composition, credit, liquidity, marginal propensities to consume, unemployment risk, and inequality were all central to the unfolding of the Great Recession. A representative agent model framework was not suitable to discuss all these issues. In addition, there was a call from a number of high officials and governors of central banks to move beyond the representative agent fiction in business cycle analysis.

HANK models offer a much more accurate representation of household consumption behaviour and can generate realistic distributions of income, wealth, and, albeit to a lesser degree, household balance sheets. At the same time, they can accommodate many sources of macroeconomic fluctuations, including those driven by aggregate demand. In sum, they provide a rich theoretical framework for quantitative analysis of the interaction between cross-sectional distributions and aggregate dynamics. They differ from representative agent models along two dimensions: differences in average consumption at any point in time between constrained and unconstrained households, and consumption heterogeneity within the subset of unconstrained households (Debortoli and Galí, 2017). A typical unconstrained household has access to the financial market and can save and invest in financial assets as opposed to constrained household who consumes only the revenue from labour.

This class of models was primarily developed to analyse the transmission of monetary policy, and more specifically, to understand the transmission of monetary policy, including the relative contribution of direct and indirect effects (Kaplan et al., 2018) or its redistributive effects across income groups (Auclert, 2019) such as inequality and monetary policy. For example, by using a two-agent version of the standard New Keynesian model, Broer et al. (2020) find that under rigid prices, monetary policy affects the distribution of consumption, but it has no effect on output as workers choose not to change their hours worked in response to wage

⁹ This section is mainly based on Kaplan and Violante (2018).

movements, whereas in the corresponding representative-agent model hours rise after a monetary policy loosening due to a wealth effect on labour supply.

4.3.5 Synthetic controls

In addition to general equilibrium models, *synthetic controls* is a new evaluation method presented by Athey and Imbens (2017) as “the most important innovation in the policy evaluation literature in the last 15 years”. It was proposed by Abadie and Gardeazabal (2003) and Abadie et al. (2010) and aims at estimating the effects of aggregate interventions¹⁰ on some aggregate outcome of interest. Consider a setting where one aggregate unit ($j = 1$), such as a state, or a school district, is exposed to an intervention of interest. The synthetic control method is based on the idea that, when the units of observation are a small number of aggregate entities, a combination of unaffected units ($j = 2, \dots, J + 1$) often provides a more appropriate comparison than any single unaffected unit alone. The methodology seeks to formalise the selection of the comparison units using a data-driven procedure. It opens the door to a mode of quantitative inference for comparative case studies.

We assume that the data span T periods and that the first T_0 periods are before the intervention. For each unit, j , and time, t , we observe the outcome of interest, Y_{jt} . For each unit, j , we also observe a set of k predictors of the outcome, X_{1j}, \dots, X_{kj} , which may include pre-intervention values of Y_{jt} and which are themselves unaffected by the intervention. For the unit affected by the intervention, $j = 1$, and a post-intervention period, $t > T_0$, we will define Y_{1t}^I to be the potential response under the intervention. Then, the effect of the intervention of interest for the affected unit in period t (with $t > T_0$) is:

$$\tau_{1t} = Y_{1t}^I - Y_{1t}^N \quad (17)$$

The methodology provides an estimate of the counterfactual as a *synthetic control*, that is a weighted average of the units in the donor pool ($j = 2, \dots, J + 1$). Formally, a synthetic control can be represented by a $J \times 1$ vector of weights, $W = (w_2, \dots, w_{J+1})'$ and the synthetic control estimators of Y_{1t}^N is given by:

$$\hat{Y}_{1t}^N = \sum_{j=2}^{J+1} w_j Y_{jt} \quad (18)$$

The weights are restricted to be non-negative and to sum to one, so synthetic controls are weighted averages of the units in the donor pool. Abadie and Gardeazabal (2003) and Abadie et al. (2010) propose to choose the weights so that the resulting synthetic control best resembles the pre-intervention values for the treated unit of predictors of the outcome variable. That is, given a set of non-negative constants, v_1, \dots, v_k , Abadie and Gardeazabal (2003) and Abadie et al. (2010) propose to choose the synthetic control, $W^* = (w_2^*, \dots, w_{J+1}^*)'$, that minimises

¹⁰ Aggregate interventions refer to interventions that are implemented at an aggregate level affecting a small number of large units (such as a cities, regions, or countries). However, the technique has been applied to settings with a large number of units.

$$\|X_1 - X_0 W\| = \left(\sum_{h=1}^k v_h (X_{h1} - w_2 X_{h2} - \cdots - w_{J+1} X_{h,J+1})^2 \right)^{\frac{1}{2}} \quad (19)$$

subject to the restriction of that w_2, \dots, w_{J+1} are non-negative and sum to one. $X_0 = [X_2, \dots, X_{J+1}]$ is the $k \times J$ matrix of the values of the predictors for the donor pool and X_1 is the $k \times 1$ vector of the values of the predictors of the treated unit.

The synthetic control approach has been used by quite a number of authors in the literature (see Athey and Imbens, 2017 for a survey and Abadie, 2020 for recent developments). For example, Chamon et al. (2017) use this approach to determine whether the intervention of the Central Bank of Brazil in foreign exchange markets after the taper tantrum of May 2013 was successful. Recall that the taper tantrum generated sharp fall in risky assets prices, including the depreciation of several emerging market currencies. The Central Bank of Brazil designed two interventions aiming to fight excess volatility and exchange rate overshooting. They find that the first foreign exchange intervention programme mitigated the depreciation of the real against the dollar. However, the second announcement had a smaller effect and this effect was not significant. Moreover, both programmes did not have an impact on the volatility of the exchange rate. Chen and Nugent (2019) use a similar approach to evaluate the impact of capital controls on financial stability and cash flows management. They found that capital controls are not effective in reaching financial stability outcomes but are consistent in reaching cash flows management outcomes.

Hebous and Ruf (2017) use this technique to evaluate the effects of adopting the allowance for corporate equity (ACE) introduced by Belgium on debt financing, passive investment, and active investment of multinational firms. ACE was introduced to achieve tax neutrality or to remedy tax discrimination against equity. They conclude that ACE reduces the corporate debt ratio of multinational affiliates, increases intra-group lending and other forms of passive investment but has no effects on production investment of multinational affiliates.

Although synthetic control methods have been widely applied in empirical literature, Becker and Klößner (2018) highlight some implementation issues, as the current implementations fail to find the correct synthetic control unit and therefore may lead to biased results. For instance, Becker and Klößner (2017) found that the loss in GDP per capita due to the presence of the mafia in Italy has been slightly overestimated by Pinotti (2015).

In their recent publication, Becker and Klößner (2018) provided (i) results on the theory of the optimisation problems that must be solved when synthetic controls are calculated, (ii) algorithmic solutions leading to exact results for the synthetic control, and (iii) generalised the methods to more than one outcome. They also provide a software to implement these extensions.

4.3.6 Vector autoregression model

The VAR model is based on empirical regularities embedded in the data. The VAR model may be viewed as a system of reduced form equations in which each of the endogenous variables is regressed on its own lagged values and the lagged values of all other variables in the system. A p^{th} -order VAR is written as

$$y_t = c + A_1 y_{t-1} + \cdots + A_p y_{t-p} + e_t \quad (20)$$

where y is a set of k endogenous variables, e_t is a k -vector of error terms with mean 0 and covariance matrix Ω , and each A_i is a $k \times k$ matrix.

The VAR technique is able to characterise the dynamic structure of the model as well as its ability to avoid imposing excessive identifying restrictions associated with different economic theories. In other words, VAR does not require any explicit economic theory to estimate the model. But VAR models can be well connected to structural models. For instance, a first order approximation of a DSGE model can be written as a VAR(1) model. VAR model has been extensively used in the economic literature to generate empirical evidence to support or infirm economic theories (see Blanchard and Watson, 1986 and Bernanke, 1986, among others). Related to capital market development, Pradhan et al. (2015) use this approach to examine the long-run relationship between bond market development and economic growth in the G-20 countries. They found the presence of both unidirectional and bidirectional causality between bond market development and economic growth, implying that the economic policies should recognise the differences in the development of bond market and economic growth to maintain sustainable development in the G-20 countries.

The analysis of the effects of exogenous shocks on the endogenous variables uses e_t in equation (20). Since the matrix Ω is not necessarily a diagonal matrix, restrictions are used to identify the shocks to analyse. Different identification schemes exist in the literature such as the Cholesky factorisation, the sign restrictions, the short and long run restrictions, etc. More recently, VAR models have been extended to account for time-varying parameters or to be applied to a large number of variables (factor VAR and dynamic factor analysis; see Stock and Watson, 2005).

4.4 Summary table of the econometric techniques for micro- and macro-assessment

Table 1 provides a summary of the econometric techniques for micro- and macro-assessment reviewed above.

Table 1: Comparing impact evaluation methods

Methodology	Description	Who is in the comparison group?	Key assumption	Required data
<i>Micro-assessment</i>				
Randomised control trial	Eligible units are randomly assigned to a treatment or comparison group. Each eligible unit has an equal chance of being selected. Tends to generate internally valid impact estimates under the weakest assumptions.	Eligible units that are randomly assigned to the comparison group.	Randomisation effectively produces two groups that are statistically identical with respect to observed and unobserved characteristics (at baseline and through endline).	Follow-up outcome data for treatment and comparison groups. Baseline outcomes and other characteristics for treatment and comparison groups to check balance.
Instrumental variable	Uses the change in outcomes induced by the change in participation rates to estimate programme impacts.	“Complier” units whose participation in the programme is affected by the instrument (they would participate if exposed to the instrument but would not participate if not exposed to the instrument).	The instrument affects participation in the programme but does not directly affect outcomes (that is, the instrument affects outcomes only by changing the probability of participating in the program).	Follow-up outcome data for all units; data on effective participation in the program; data on baseline outcomes and other characteristics.
Regression discontinuity design	Units are ranked based on specific quantitative and continuous criteria, such as a poverty index.	Units that are close to the cut-off but are	To identify unbiased programme impacts for the population close to the cut-off, units that are	Follow-up outcome data; ranking index and eligibility cut-off; data on

	There is a cut-off that determines whether or not a unit is eligible to participate in a programme. Outcomes for participants on one side of the cut-off are compared with outcomes for nonparticipants on the other side of the cut-off.	ineligible to receive the programme.	immediately below and immediately above the cut-off are statistically identical. To identify unbiased programme impacts for the whole population, the population close to the cut-off needs to be representative of the whole population.	baseline outcomes and other characteristics.
Difference-in-differences	The change in outcome over time in a group of non-participants is used to estimate what would have been the change of outcomes for a group of participants in the absence of a programme.	Units that did not participate in the programme (for any reason), and for which data were collected before and after the programme.	If the programme did not exist, outcomes for the groups of participants and nonparticipants would have grown in parallel over time.	Baseline and follow-up data on outcomes and other characteristics for both participants and nonparticipants.
Propensity score matching	For each programme participant, the method looks for the “most similar” unit in the group of non-participants (the closest match based on observed characteristics).	For each participant, the non-participant unit that is predicted to have the same likelihood to have participated in the programme based on observed characteristics.	There is no characteristic that affects programme participation beyond the observed characteristics used for matching.	Follow-up outcome data for participants and non-participants; data on effective participation in the programme; baseline characteristics to perform matching.
Macro-assessment				

Computable general equilibrium models	A CGE model is a system of simultaneous non-linear equations. It describes the motivations and behaviour of consumers and producers in an economy and their linkages. It highlights the effects of: (i) policies and changes occurring at sectoral level or at the household level; (ii) policies and changes which occur at the macro-economic level; and (iii) changes in external economy.	The policy analysis compares a scenario (that assumes the economy will change under the influence of future changes due to the interventions) to a baseline scenario. The baseline scenario provides a reference point for evaluating the impact of policy changes or other events.	The parameters that govern the reactions of households and firms to shocks are either estimated or calibrated. Parameters are fixed or change over time (dynamic).	Disaggregated and up-to-date social accounting matrix.
Dynamic stochastic general equilibrium models	Full structural micro-founded model describing the behaviour of economic agents over time in a consistent manner.	The policy analysis compares the deviation from the equilibrium. In absence of shocks, the model does not deviate from the equilibrium	Modelling of the behaviour of economic agents, that is the choice of utility functions. The parameters that govern the reactions of households and firms are either estimated or calibrated.	Time series if one wants to estimate the model; or key macro-economic indicators for the calibration.
Agent-based models	ABM use computer simulation to explore emerging dynamical patterns. There are no assumptions about the	ABM can simulate the behaviour of agents with and without the intervention. The method exploits spatial	Research questions that require significant heterogeneity within and between agents and diverse spatial and	Qualitative or quantitative data for the calibration of the model.

	<p>existence of efficient policies or general equilibrium. It can accommodate high heterogeneity in agent characteristics and interactions among agents and environments, as well as features like dynamics, feedbacks, and adaptation.</p>	<p>and temporal differences in roll-out.</p>	<p>relational elements are well-suited to ABM. ABM require experts from different fields depending on the research question.</p>	
Heterogeneous agent New Keynesian models	<p>HANK models offer a much more accurate representation of household consumption behaviour. They can also accommodate many sources of macroeconomic fluctuations, including those driven by aggregate demand. They differ from representative agent models along two dimensions: differences in average consumption at any point in time between constrained and unconstrained households, and consumption</p>	<p>The policy analysis compares the deviation from the equilibrium. In absence of shocks, the model does not deviate from the equilibrium.</p>	<p>Modelling of the behaviour of economic agents, that is the choice of utility functions. The parameters that govern the reactions of households and firms are either estimated or calibrated.</p>	<p>Time series if one wants to estimate the model; or key macro-economic indicators for the calibration.</p>

	heterogeneity within the subset of unconstrained households.			
Synthetic controls	For the participant unit, the method constructs a synthetic comparison group based on a weighted average of the unaffected units.	The methodology seeks to formalise the selection of the comparison units using a data-driven procedure based on data of unaffected units.	The synthetic control is constructed from a pool of similar units and a good fit is achieved over a sufficient period pre-implementation. There are no shocks other than the intervention.	Preintervention data on both treated unit and units of comparison.
Vector autoregression models	The VAR model is a system of reduced form equations in which each of the endogenous variables is regressed on its own lagged values and the lagged values of all other variables in the system.	Model-based simulations assuming there is no shock.	Time series properties of the variables and the model need to be studied before any regression.	Time series over a long period to estimate parameters.

Source: Gertler et al. (2016) and authors' compilation

5 Conclusion

It is well documented that Africa has one of the lowest financial inclusion rates in the world. The financing of the economy is essentially bank-based, with underdeveloped capital markets across the continent, the exception being South Africa. Despite this underdeveloped stage, there is still a huge potential for capital market expansion given the continent's large financial development and infrastructure development gaps.

Capital market interventions can help test experimental policies and programmes to understand their impacts on the economy. If successful, these interventions can be scaled up to a much larger sample of economic agents. However, one main challenge with these interventions is how to evaluate their true impact. It is well known that impact assessments for interventions in capital markets are sometimes difficult to achieve given the intrinsic nature of these interventions and their non-negligible indirect effects. Most of these interventions are implemented locally and at a small scale, and often it is hard to construct counterfactual groups. Even when this last condition is met, it is often difficult to dissociate the direct and indirect impacts of the policies and programmes being evaluated. All these make it harder to rigorously evaluate the impact of market-building interventions in capital markets.

This paper reviews the most up-to-date impact assessment techniques with illustrative examples for their application to market-building interventions in capital markets. The techniques include, on the one hand, micro assessment approaches which are more appropriate to evaluate the direct impact of capital market interventions; and on the other, macro evaluation techniques based on general equilibrium, synthetic controls and VAR, which allow spillover and macro-economic effects to be accounted for.

References

- Abadie, A. (2020). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature (Forthcoming)*.
- Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, 93(1), 113-132.
- Abadie, A., Diamond, A., & Hainmueller, J. (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.
- Adler, G., Lisack, N., & Mano, R. C. (2019). Unveiling the effects of foreign exchange intervention: A panel approach. *Emerging Markets Review*, 40, 100620.
- AFM (2019). A randomized controlled trial on the effectiveness of mandatory investment information. Online article, February. Amsterdam: Dutch Authority for the Financial Markets (<https://www.afm.nl/en/professionals/onderwerpen/consumentengedrag-artikelen/trial-mandatory-investment>).
- Allen, F., Carletti, E., Cull, R., Qian, J., Senbet, L., & Valenzuela, P. (2014). The African financial development and financial inclusion gaps. *Journal of African Economies*, 23, 614-642.
- Allen, F., Otchere, I., & Senbet, L. (2011). African financial systems: A review. *Review of Development Finance*, 1, 79-113.
- Asian Development Bank (ADB, 2005). *Capital Market Development Program (Loan 1580-BAN[SF]) in Bangladesh*. Program Performance Audit Report.
- Asifma (2019). China's capital market: the pace of change accelerates. June 2019. Asia Securities Industry and Financial Markets Association.
- Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2), 3-32.
- Atindehou, R., Gueyie, J. P., & Amenounve, E. K. (2005). Financial intermediation and economic growth: Evidence from Western Africa. *Applied Financial Economics*, 15, 777-790.
- Atje, R., & Jovanovic, B. (1993). Stock markets and development. *European Economic Review*, 37(2-3), 632-640.
- Auclert, A. (2019). Monetary policy and the redistribution channel. *American Economic Review*, 109(6), 2333-67.
- Ayyagari, M., Demirguc-Kunt, A., & Maksimovic, V. (2011). Firm innovation in emerging markets: The role of finance, governance, and competition. *Journal of Financial and Quantitative Analysis*, 46(6), 1545-1580.
- Basu, K (2014) Randomisation, Causality and the Role of Reasoned Intuition, *Oxford Development Studies*, 42:4, 455-472
- Basu, S., & Bundick, B. (2017). Uncertainty shocks in a model of effective demand. *Econometrica*, 85(3), 937-958.
- Beck, T. & Cull, R. (2013). *Banking in Africa*. Policy Research Working Paper No. 6684, World Bank, Washington DC.
- Beck, T., Fuchs, M., & Uy, M. (2009). *Finance in Africa: Achievements and challenges*. Policy Research Working Paper No. 5020, World Bank, Washington DC.
- Becker, M., & Klößner, S. (2017). Estimating the economic costs of organized crime by synthetic control methods. *Journal of Applied Econometrics*, 32(7), 1367-1369.
- Becker, M., & Klößner, S. (2018). Fast and reliable computation of generalized synthetic controls. *Econometrics and Statistics*, 5, 1-19.
- Bekaert, G., Harvey, C. R., & Lundblad, C. (2005). Does financial liberalization spur growth? *Journal of Financial Economics*, 77, 3-56.
- Belissa, T., Bulte, E., Cecchi, F., Gangopadhyay, S., & Lensink, R. (2019). Liquidity constraints, informal institutions, and the adoption of weather insurance: A randomized controlled Trial in Ethiopia. *Journal of Development Economics*, 140, 269-278.

Working Paper 3: Impact assessment for market-building interventions in capital markets: Review, critiques and improvements

- Belissa, T., Lensink, R., & Winkel, A. (2020). Effects of Index Insurance on Demand and Supply of Credit: Evidence from Ethiopia. *American Journal of Agricultural Economics*, 102(5), 1511-1531.
- Berck, P., & Villas-Boas, S. B. (2016). A note on the triple difference in economic models. *Applied Economics Letters*, 23(4), 239-242.
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1), 249-275.
- Beyers, C. F., De Freitas, A., Essel-Mensah, K. A., Seymore, R., & Tsomocos, D. P. (2020). A computable general equilibrium model for banking sector risk assessment in South Africa. *Annals of Finance*, 1-24.
- Blanchard, O., & Galí, J. (2010). Labor markets and monetary policy: A New Keynesian model with unemployment. *American Economic Journal: Macroeconomics*, 2(2), 1-30.
- Blanchard, O. J., & Watson, M. W. (1986). Are business cycles all alike? In *The American business cycle: Continuity and change* (pp. 123-180). University of Chicago Press.
- Bloom, N., Floetotto, M., Jaimovich, N., Saporta-Eksten, I., & Terry, S. J. (2018). Really uncertain business cycles. *Econometrica*, 86(3), 1031-1065.
- Boissay, F., Collard, F., & Smets, F. (2016). Booms and banking crises. *Journal of Political Economy*, 124(2), 489-538.
- Botta, A., Caverzasi, E., Russo, A., Gallegati, M., & Stiglitz, J. E. (2019). Inequality and finance in a rent economy. *Journal of Economic Behavior & Organization*. Article in Press.
- Boyd, J. H. & Prescott, E. C. (1986). Financial intermediary-coalitions. *Journal of Economic Theory*, 38(2), 211-232.
- Broer, T., Harbo Hansen, N. J., Krusell, P., & Öberg, E. (2020). The New Keynesian transmission mechanism: A heterogeneous-agent perspective. *The Review of Economic Studies*, 87(1), 77-101.
- Bulte, E., Di Falco, S., & Lensink, R. (2020). Randomized interventions and “real” treatment effects: A cautionary tale and an example. *World Development*, 127, 104790.
- Burger, J. D., & Warnock, F. E. (2006). Local currency bond markets. IMF Staff papers, 53(1), 133-146.
- Burger, J. D., Warnock, F. E., & Warnock, V. C. (2012). Emerging local currency bond markets. *Financial Analysts Journal*, 68(4), 73-93.
- Calamanti, A. (1983). *The Securities Market and Underdevelopment: the security exchanges in Ivory Coast, Morocco and Tunisia*. Finafrica.
- Calderon, C., & Liu, L. (2003). The direction of causality between financial development and economic growth. *Journal of Development Economics*, 72(1), 321-334.
- Calvo, G. A. (1998). Capital flows and Capital markets crises: The Simple Economics of Sudden Stops. *Journal of Applied Economics*, 1(1), 35-54.
- Cantú, C., & Chui, M. (2020). *Financial market development and financial stability*. BIS Papers No. 113.
- Carneiro, P., Heckman, J. J., & Vytlacil, E. (2010). Evaluating marginal policy changes and the average effect of treatment for individuals at the margin. *Econometrica*, 78(1), 377-394.
- Chamon, M., Garcia, M., & Souza, L. (2017). FX interventions in Brazil: a synthetic control approach. *Journal of International Economics*, 108, 157-168.
- Chang, Y. C., Hong, H., & Liskovich, I. (2015). Regression discontinuity and the price effects of stock market indexing. *The Review of Financial Studies*, 28(1), 212-246.
- Chen, M. T., & Nugent, R. J. (2019). *On the effectiveness of capital controls: A synthetic control method approach*. Unpublished Working Paper.
- Claessens, S., Klingebiel, D., & Schmukler, S. L. (2007). Government bonds in domestic and foreign currency: the role of institutional and macroeconomic factors. *Review of International Economics*, 15(2), 370-413.
- Coulibaly, S. S., Gakpa, L. L. & Soumaré, I. (2018). The role of property rights in the relationship between capital flows and economic growth in SSA: Do natural resources endowment and country income level matter? *African Development Review*, 30(1), 112-130.
- Curdia, V., & Woodford, M. (2011). The central-bank balance sheet as an instrument of monetary policy. *Journal of Monetary Economics*, 58(1), 54-79.
- Dabla-Norris, E., Townsend, R. M., & Distinguishing constraints on financial inclusion and their impact on GDP, TFP, and the distribution of income, *Journal of Monetary Economics* (forthcoming).
- Debortoli, D., & Galí, J. (2017). *Monetary policy with heterogeneous agents: Insights from TANK models*. Manuscript, September.

Working Paper 3: Impact assessment for market-building interventions in capital markets: Review, critiques and improvements

- Decaluwé, B., & Martens, A. (1988). CGE modeling and developing economies: A concise empirical survey of 73 applications to 26 countries. *Journal of Policy Modeling*, 10(4), 529-568.
- Desai, M. A., Foley, C. F., & Hines Jr., J. R. (2006). Capital Controls, Liberalizations, and Foreign Direct Investment. *Review of Financial Studies* 19, 1434-1464.
- Duffie, D. (2019). Prone to Fail: The Pre-Crisis Financial System. *The Journal of Economic Perspectives*, 33(1), 81-106.
- Eaton, J., Kortum, S., & Neiman, B. (2016). Obstfeld and Rogoff's international macro puzzles: a quantitative assessment. *Journal of Economic Dynamics and Control*, 72, 5-23.
- Eichengreen, B. & Luengnaruemitchai, P. (2006). *Why doesn't Asia have bigger bond markets?* BIS Papers chapters, in: Bank for International Settlements (ed.), Asian bond markets: issues and prospects, volume 30, pages 40-77 Bank for International Settlements.
- EIBannan, M. A. (2017). Stock market liquidity, family ownership, and capital structure choices in an emerging country. *Emerging Markets Review*, 33, 201-231.
- Essama-Nssah, B. (2005). *Simulating the poverty impact of macroeconomic shocks and policies*. The World Bank.
- Fagiolo, G., Guerini, M., Lamperti, F., Moneta, A., & Roventini, A. (2019). Validation of agent-based models in economics and finance. In *Computer Simulation Validation* (pp. 763-787). Springer, Cham.
- Fatum, R., & Hutchison, M. M. (2010). Evaluating foreign exchange market intervention: self-selection, counterfactuals and average treatment effects. *Journal of International Money and Finance*, 29(3), 570-584.
- Feldstein, M., & Horioka, C. (1980). Domestic savings and international capital flows. *The Economic Journal*, 90(358), 314-329.
- Ford, N., & Horioka, C. Y. (2017). The 'real' explanation of the Feldstein–Horioka puzzle. *Applied Economics Letters*, 24(2), 95-97.
- Funke, M., & Paetz, M. (2012). *Financial system reforms and China's monetary policy framework: A DSGE-based assessment of initiatives and proposals*. BOFIT Discussion Paper No. 30/2012.
- Gali, J. (2018). The state of New Keynesian economics: a partial assessment. *Journal of Economic Perspectives*, 32(3), 87-112.
- Gali, J. (2020). Monetary policy and bubbles in a new Keynesian model with overlapping generations. *American Economic Journal: Macroeconomics* (Forthcoming).
- Geanakoplos, J., Axtell, R., Farmer, J. D., Howitt, P., Conlee, B., Goldstein, J., Hendrey, M., Palmer, N. M. & Yang, C. Y. (2012). Getting at systemic risk via an agent-based model of the housing market. *American Economic Review*, 102(3), 53-58.
- Gerali, A., Neri, S., Sessa, L., & Signoretti, F. M. (2010). Credit and Banking in a DSGE Model of the Euro Area. *Journal of Money, Credit and Banking*, 42, 107-141.
- Gertler, M., & Karadi, P. (2011). A model of unconventional monetary policy. *Journal of monetary Economics*, 58(1), 17-34.
- Gertler, M., & Kiyotaki, N. (2010). Financial intermediation and credit policy in business cycle analysis. In *Handbook of Monetary Economics* (Vol. 3, pp. 547-599). Elsevier.
- Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2016). *Impact evaluation in practice*. Washington, DC: Inter-American Development Bank and World Bank.
- Gorodnichenko, Y., & Schnitzer, M. (2013). Financial constraints and innovation: Why poor countries don't catch up. *Journal of the European Economic association*, 11(5), 1115-1152.
- Guerrieri, V., & Lorenzoni, G. (2017). Credit crises, precautionary savings, and the liquidity trap. *The Quarterly Journal of Economics*, 132(3), 1427-1467.
- Hamermesh, D. S., & Trejo, S. J. (2000). The demand for hours of labor: Direct evidence from California. *Review of Economics and Statistics*, 82(1), 38-47.
- Haqiqi, I., & Mirian, N. (2015). A Financial General Equilibrium Model for Assessment of Financial Sector Policies in Developing Countries. MPRA Paper No. 96105.
- Hebous, S., & Ruf, M. (2017). Evaluating the effects of ACE systems on multinational debt financing and investment. *Journal of Public Economics*, 156, 131-149.
- Heckman, J. J., & Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, 73(3), 669-738.
- Henry, P. B. (2003). Capital-Account Liberalization, the Cost of Capital, and Economic growth. *American Economic Review*, 93, 91-96.
- Henry, P.B. (2000). Do Stock Market Liberalizations Cause Investment Booms? *Journal of Financial Economics* 58, 301-334

Working Paper 3: Impact assessment for market-building interventions in capital markets: Review, critiques and improvements

- Horridge, M., Meeraus, A., Pearson, K., & Rutherford, T. F. (2013). Solution software for computable general equilibrium modeling. In *Handbook of computable general equilibrium modeling* (Vol. 1, pp. 1331-1381). Elsevier.
- Huber, M., & Mellace, G. (2015). Testing instrument validity for LATE identification based on inequality moment constraints. *Review of Economics and Statistics*, 97(2), 398-411.
- Ibrahim, M., & Alagidede, P. (2017). Financial sector development, economic volatility and shocks in sub-Saharan Africa. *Physica A: Statistical Mechanics and its Applications*, 484, 66-81.
- Ibrahim, M., & Alagidede, P. (2018). Effect of financial development on economic growth in sub-Saharan Africa. *Journal of Policy Modeling*, 40(6), 1104-1125.
- Imbens, G. W. & Angrist, J. D. (1994) Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2), 467-475.
- IMF (2020). *Chapter 3: Emerging and Frontier markets: Managing volatile portfolios flows*. In *Global Financial Stability Report: Markets in the Time of COVID-19* (April 2020). International Monetary Fund, Washington, DC.
- Jurado, K., Ludvigson, S. C., & Ng, S. (2015). Measuring uncertainty. *American Economic Review*, 105(3), 1177-1216.
- Kaminsky, G. L., & Reinhart, C. M. (1999). The twin crises: the causes of banking and balance-of-payments problems. *American Economic Review*, 89(3), 473-500.
- Kaplan, G., & Violante, G. L. (2018). Microeconomic heterogeneity and macroeconomic shocks. *Journal of Economic Perspectives*, 32(3), 167-94.
- Kaplan, G., Moll, B., & Violante, G. L. (2018). Monetary policy according to HANK. *American Economic Review*, 108(3), 697-743.
- Kellogg, R., & Wolff, H. (2008). Daylight time and energy: Evidence from an Australian experiment. *Journal of Environmental Economics and Management*, 56(3), 207-220.
- Khandker, S., B. Koolwal, G., & Samad, H. (2010). *Handbook on impact evaluation: quantitative methods and practices*. The World Bank.
- King, G., & Nielsen, R. A. (2019). Why propensity scores should not be used for matching. *Political Analysis*, 27(4), 435-454.
- Kitagawa, T. (2015). A test for instrument validity. *Econometrica*, 83(5), 2043-2063.
- Kleibergen, F. (2007). Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics. *Journal of Econometrics*, 139(1), 181-216.
- Klein, M. & Olivei, G. (2008). Capital account liberalization, financial deepness and economic growth. *Journal of International Money and Finance*, 27, 861-875.
- Kuersteiner, G. M., Phillips, D. C., & Villamizar-Villegas, M. (2018). Effective sterilized foreign exchange intervention? Evidence from a rule-based policy. *Journal of International Economics*, 113, 118-138.
- La Porta, R. L., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. W. (1998). Law and finance. *Journal of Political Economy*, 106(6), 1113-1155.
- La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. W. (1997). Legal determinants of external finance. *The Journal of Finance*, 52(3), 1131-1150.
- Laeven, L., Levine, R., & Michalopoulos, S. (2015). Financial innovation and endogenous growth. *Journal of Financial Intermediation*, 24(1), 1-24.
- Laeven, M. L. (2014). *The development of local capital markets: rationale and challenges*. Working Paper No. 14-234. International Monetary Fund.
- Lauretta, E. (2018). The hidden soul of financial innovation: An agent-based modelling of home mortgage securitization and the finance-growth nexus. *Economic Modelling*, 68, 51-73.
- Levine, R., & Zervos, S. (1996). Stock market development and long-run growth. *World Bank Economic Review*, 10(2), 323-339.
- Levine, R., Loayza, N., & Beck, T. (2000). Financial intermediation and growth: Causality and causes. *Journal of Monetary Economics*, 46, 31-77.
- Masten, A. B., Coricelli, F., & Masten, I. (2008). Non-linear growth effects of financial development: Does financial integration matter? *Journal of International Money and Finance*, 27(2), 295-313.
- Mazzocchetti, A., Lauretta, E., Roberto, M., Teglio, A., & Cincotti, S. (2020). Systemic financial risk indicators and securitised assets: an agent-based framework. *Journal of Economic Interaction and Coordination*, 15(1), 9-47.
- Mazzucato, M., and Semeniuk, G. (2017). Public financing of innovation: new questions. *Oxford Review of Economic Policy*, 33(1), 24-48.
- McIntosh, C. (2008). Estimating treatment effects from spatial policy experiments: an application to Ugandan microfinance. *The Review of Economics and Statistics*, 90(1), 15-28.

Working Paper 3: Impact assessment for market-building interventions in capital markets: Review, critiques and improvements

- McKay, A., & Reis, R. (2016). The role of automatic stabilizers in the US business cycle. *Econometrica*, 84(1), 141-194.
- McKenzie, D. (2010). Impact Assessments in Finance and Private Sector Development: What have we learned and what should we learn? *The World Bank Research Observer*, 25(2), 209-233.
- Meh, C. A., & Moran, K. (2010). The role of bank capital in the propagation of shocks. *Journal of Economic Dynamics and Control*, 34(3), 555-576.
- Mian, A., Rao, K., & Sufi, A. (2013). Household balance sheets, consumption, and the economic slump. *The Quarterly Journal of Economics*, 128(4), 1687-1726.
- Miao, J., Wang, P., & Xu, Z. (2015). A Bayesian dynamic stochastic general equilibrium model of stock market bubbles and business cycles. *Quantitative Economics*, 6(3), 599-635.
- Moffitt, R. (2008). Estimating marginal treatment effects in heterogeneous populations. *Annales d'Economie et de Statistique*, 239-261.
- Mourifié, I., & Wan, Y. (2017). Testing local average treatment effect assumptions. *Review of Economics and Statistics*, 99(2), 305-313.
- NAEC – New Approaches to Economic Challenges (2017) NAEC Innovation Lab: Workshop on agent based modelling: applications in labour markets (www.oecd.org/naec/Agent-based_models_background.pdf).
- Ocampo, A. J., & Stiglitz, J.E. (2008). Capital Market Liberalization and Development. Oxford: Oxford University Press.
- OECD DAC (2018). OECD DAC Evaluation Criteria: Summary of consultation responses. November (https://ieg.worldbankgroup.org/sites/default/files/Data/DAC-Criteria/ConsultationReport_EvaluationCriteria.pdf).
- Oh, H., & Reis, R. (2012). Targeted transfers and the fiscal response to the great recession. *Journal of Monetary Economics*, 59, S50-S64.
- Oh, I., Lee, J. D., Heshmati, A., & Choi, G. G. (2009). Evaluation of credit guarantee policy using propensity score matching. *Small Business Economics*, 33(3), 335-351.
- Otchere, I., Soumaré, I., & Yourougou, P. (2016). FDI and Financial Market Development in Africa. *The World Economy*, 39(5), 651-678.
- Pasanen, T. (2018). 2018: time to update the DAC evaluation criteria? ODI blog, 17 January (<https://odi.org/en/insights/2018-time-to-update-the-dac-evaluation-criteria>).
- Patrick, H. T. (1966). Financial development and economic growth in underdeveloped countries. *Economic development and Cultural change*, 14(2), 174-189.
- Pereira da Silva, L. A., Bourguignon, F., & Bussolo, M. (Eds.). (2008). *The impact of macroeconomic policies on poverty and income distribution: macro-micro evaluation techniques and tools*. The World Bank.
- Pinotti, P. (2015). The economic costs of organised crime: Evidence from Southern Italy. *The Economic Journal*, 125(586), F203-F232.
- Pradhan, R. P., Zaki, D. B., Maradana, R. P., Dash, S., Jayakumar, M., & Chatterjee, D. (2015). Bond market development and economic growth: The G-20 experience. *Tékhne*, 13(1), 51-65.
- Ravallion, M. (2007). Evaluating anti-poverty programs. *Handbook of Development Economics*, 4, 3787-3846.
- Reinhart, C. M., & Reinhart, V. R. (2008). Capital Flow Bonanzas: An Encompassing View of the Past and Present. *NBER International Seminar on Macroeconomics*, 5(1), 9-62.
- Rioja, F., & Valev, N. (2004). Does one size fit all? A reexamination of the finance and growth relationship. *Journal of Development Economics*, 74(2), 429-447.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.
- Soumaré, I. (2015). Does foreign direct investment improve economic development in North African countries? *Applied Economics*, 47(51), 5510-5533.
- Soumaré, I., & Tchana Tchana, F. (2015). Causality between FDI and Financial Market Development: Evidence from Emerging Markets. *World Bank Economic Review (WBER)*, 29 (Supplement), S205-S216.
- Spratt, S., Barnett, C., O'Flynn, P., Gadhavi, V., Tranchant, J. P., Saha, A., Punton, M. & Spafford, K. (2019). *Study of CDC's Mobilisation of Private Investment*. Inception report submitted by Itad in association with Institute of Development Studies, Open Capital Advisors and Ipsos MORI.
- Stiglitz, Joseph E (2000) Capital Market Liberalization, Economic Growth, and Instability, World Development, Volume 28, Issue 6, Pages 1075-1086.
- Stock, J. H., & Watson, M. W. (2005). *Implications of dynamic factor models for VAR analysis*. Working Paper No. w11467. National Bureau of Economic Research.

- Stock, J., & Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In D. Andrews & J. Stock (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg* (pp. 80-108). Cambridge: Cambridge University Press.
- Todd, P. E. (2007). Evaluating social programs with endogenous program placement and selection of the treated. *Handbook of Development Economics*, 4, 3847-3894.
- Trigari, A. (2009). Equilibrium unemployment, job flows, and inflation dynamics. *Journal of Money, Credit and Banking*, 41(1), 1-33.
- Verona, F., Martins, M. M., & Drumond, I. (2013). (Un) anticipated monetary policy in a DSGE model with a shadow banking system. *International Journal of Central Banking*, 9(3), 73-117.
- Woodford, M. (2010). Robustly optimal monetary policy with near-rational expectations. *American Economic Review*, 100(1), 274-303.
- World Bank (2016). The World Bank Group's Support to Capital Market Development: Main Report. The World Bank Group.

Appendix 1 Application to selected FSD projects

Table 2 below provides for each of the selected FSD projects the potential impact evaluation technique that can be used. Of course, for all of these interventions DAC criteria may be applied as well. But the emphasis is on the econometric techniques described above.

The final choice of the evaluation methodology will depend on five considerations:

- (i) the outcome the funder or project promoter is seeking to achieve;
- (ii) the context, such as the availability of data or the presence of a comparison group;
- (iii) the timeline of the contract and the time allocated to data collection;
- (iv) the evaluation budget: collecting baseline data, following large samples, and organising RCTs can be expensive and time-consuming; and
- (v) the political sensitivities around the intervention.

Table 2: Application of the impact evaluation techniques to selected FSD projects

Project name	Start date	End date	Anticipated project impact	Potential evaluation method
Capital Markets Authority of Uganda Development of Capital Markets Master Plan Institutional Capacity Assessment	01 Jan 2016	31 Dec 2019	Development of a capital markets master plan and support for its implementation At least 100 capital markets professionals certified under the Chartered Institute for Securities & Investment (CISI) IISI qualification	Double difference may be used to evaluate the capacity building component if all practitioners do not attend classes in order to create a valid control group. The same methodology may be used to assess the impact of the SME Advisory Centre. SMEs contemplating a stock market listing can be grouped into two categories (treated or access to the Centre, and control groups). If baseline data is not available, we can rely on propensity score matching.
CMA Kenya Project Change Management and HR Specialist Management Office for Islamic Finance	01 Nov 2015	31 Dec 2017	At least 500 capital markets professionals certified under CISI's IISI qualification Set up of a suitable framework for a National Shariah Supervisory Board (NSSB) to operate within Kenya	Double difference for the capacity building component if baseline data is available. Otherwise, we can rely on propensity score matching.
Chartered Institute for Securities & Investments (CISI) Africa	01 Feb 2018	31 Oct 2020	Capacity of over 3,000 practitioners enhanced Capacity of over 30 trainers enhanced 20 countries (including on-going project in WAMU) officially recognising CISI qualifications	Double difference or propensity score matching

Long Term Finance	01 Apr 2017	31 Mar 2021	<p>Policy and regulatory reforms that will increase the volume and accessibility of LTF. Closing information gaps and drawing investors' attention to individual markets and unexploited opportunities.</p> <p>Raising awareness to the current use of less well-known sources of long-term financing</p>	Structural simulation model to assess the potential impact of LTF, including firms, government, etc.
Frontclear Technical Assistance	15 Feb 2016	31 Mar 2022	<p>Greater liquidity in the interbank market, resulting in enhanced access to foreign exchange and longer-term finance for Tier 2/3 banks</p> <p>Increased availability of capital for local banks to on-lend to their clients, especially in sectors that need longer term finance or forex (e.g. SME, agro-business)</p>	Double difference, propensity score matching, IVs, macro-model (interbank and the macro-economy)

African Local Currency Bond Fund (ALCBF)	01 Apr 2017	31 Dec 2030	<p>Enables the first-time issuance of new bonds from corporate issuers across Sub-Saharan Africa</p> <p>Alternative and less expensive non-bank finance reduces service providers' maturity, currency and interest rate risk</p> <p>Low-income households and MSMEs benefit from access to basic services (finance, housing, agriculture, energy) with more attractive terms</p> <p>New best practices enhance skills and standards among intermediaries that structure and market issuances</p> <p>Increased awareness around opportunity to finance through bonds stimulates new issuances</p> <p>Enhanced transparency and improved efficiency and depth of market transactions attracts local institutional investors channelling capital to the real economy</p> <p>Increased knowledge around the functioning of bond markets supports enabling regulation</p>	All techniques can be applied. However, RCT needs to be designed before the beginning of the programme to sample the treated and control groups. In addition, macro-simulation techniques can be designed to assess the overall macroeconomic impact of the intervention.
Private Equity and Debt Programme	10 Sep 2020	31 Mar 2025	<p>i. Facilitative regulatory environment for the development of private markets.</p> <p>ii. Improved capacity and knowledge on private markets (private equity and private debt) by</p>	All techniques can be applied. However, RCT needs to be designed before the beginning of the programme to sample the treated and control groups. In addition, macro-simulation techniques can

			<p>pension funds, regulators and other relevant stakeholders.</p> <p>iii. Adequate structures and/or demonstration leading to increased investment in private markets by pension funds.</p> <p>iv. Achievement of social impacts such as: (a) creating and sustaining jobs, (b) increasing individuals' access to basic services (for instance housing and healthcare), (c) helping individuals, households and communities in target countries cope with climate change.</p>	be designed to assess the overall macroeconomic impact of the intervention.
Blue Orchard Liquidity Debt Fund	10 Sep 2020	2027	<p>Stability of MFIs (averting failure of institutions serving key market segments not reached by Banks)</p> <p>Survival of MSMEs</p> <p>Jobs protected</p> <p>Stability of income flows for low-income and poor households</p>	All techniques can be applied. However, RCT needs to be designed before the beginning of the programme to sample the treated and control groups.