GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN IN PUBLICA COMMODA
SEIT 1737

DOCTORAL THESIS

Local Dimensions of Development

A thesis submitted in fulfillment of the requirements for the degree of doctor of economic sciences (Dr. rer. pol.) from the

Faculty of Business and Economics
Georg-August-Universität Göttingen

Submitted by
Lennart Reiners
born in Köln, Germany

May 25th, 2023

First Examiner
Prof. Dr. Krisztina Kis-Katos
Professor of Economics
Chair of International Economic Policy
Department of Economics
Georg-August-Universität Göttingen

Second Examiner
Prof. Dr. Andreas Fuchs
Professor of Economics
Chair of Development Economics
Department of Economics
Georg-August-Universität Göttingen

Third Examiner
Prof. Dr. Jochen Kluve
Professor of Economics
Chair of Empirical Labor Economics
School of Business and Economics
Humboldt-Universität zu Berlin

Abstract

Poverty remains one of the most pressing global challenges, affecting people's livelihood in far more than the mere income dimension: The poor suffer from limited access to education, health care and nutrition, face discrimination, and lack societal participation. Tackling poverty in all its dimensions is therefore at the core of national governments' and international institutions' policy agenda, contributing towards Sustainable Development Goal (SDG) 1. While the underlying causes of poverty are as complex and multidimensional as their implications, it is generally agreed that the role of the government and its policies is decisive: It establishes a conducive environment for economic development and provides social protection systems for people in need. The right policies are therefore at the heart of successful poverty alleviation, yet often constrained by limited financial resources and technical capacity.

Assessing the effectiveness of such policies is an essential prerequisite for their careful design. In the context of development cooperation, a longstanding scholarly debate has questioned its effectiveness. This comprises not only whether intended goals were accomplished, but also the less prominent debate on unintended—both positive and negative—consequences of development interventions. Only more recently, a growing body of research on the unintended and, in particular, adverse impacts of international donor-supported development interventions has emerged. This literature has been fueled by the growing use of rigorous evaluations, better data coverage and the realization that more scrutiny on unintended consequences is needed. Researchers can thus increasingly uncover policies' side effects, crucially complementing assessments of what development interventions bring about.

In this thesis titled *Local Dimensions of Development*, I aim to make a contribution towards understanding the unintended consequences of interventions aimed at promoting development. I discuss three policies in the context of three distinctive phenomena that developing countries often experience: (i) Governance reforms and environmental degradation; (ii) migration restrictions and natural disasters; and (iii) social protection systems and violence. As introductory analysis, I examine the success determinants of policy interventions supported by an international development institution.

Zusammenfassung

Armut ist eine der drängendsten globalen Herausforderungen und beeinträchtigt die Lebensgrundlagen von Menschen in mehreren Dimension: Die Armen haben eingeschränkten Zugang zu Bildung, Gesundheitsversorgung und Ernährung, und können sich weniger an der Gesellschaft beteiligen. Die Bekämpfung von Armut in all ihren Dimensionen steht daher im Mittelpunkt der Agenda von nationalen Regierungen und internationalen Institutionen, um die nachhaltigen Entwicklungsziele (SDGs) zu erreichen. Obwohl die Ursachen von Armut komplex sind, besteht allgemein Einigkeit darüber, dass die Rolle des Staates entscheidend ist: Er schafft ein passendes Umfeld für wirtschaftliche Entwicklung und stellt soziale Absicherung bereit. Die richtigen Politikmaßnahmen sind daher das Kernstück erfolgreicher Armutsbekämpfung, werden jedoch häufig durch begrenzte finanzielle und technische Mittel eingeschränkt.

Die Bewertung der Wirkungen solcher Maßnahmen ist eine wesentliche Voraussetzung für deren sorgfältige Gestaltung. Im Kontext von Entwicklungszusammenarbeit gibt es eine lange Debatte über deren Wirksamkeit. Dabei geht es nicht nur darum, ob angestrebte Ziele erreicht wurden, sondern auch um die weniger prominente Frage zu unbeabsichtigten Folgen von Entwicklungsmaßnahmen. Erst seit Kurzem gibt es eine wachsende Zahl von Studien über die unbeabsichtigten und insbesondere negativen Auswirkungen solcher Maßnahmen. Diese Literatur wird durch den Einsatz rigoroser Evaluierungsmethoden, größerer Datenverfügbarkeit und der Erkenntnis, dass unbeabsichtigte Folgen genauer untersucht werden müssen, vorangetrieben. Forschende können so zunehmend die Nebenwirkungen von Maßnahmen untersuchen, was die Bewertung von Entwicklungsmaßnahmen und deren Auswirkungen entscheidend ergänzt.

In dieser Arbeit mit dem Titel *Local Dimensions of Development* möchte ich einen Beitrag zum Verständnis der unbeabsichtigten Folgen von Entwicklungsprojekten leisten. Dabei erörtere ich drei Maßnahmen im Kontext von drei charakteristischen Phänomenen, die in Entwicklungsländern häufig auftreten: (i) Staatsreformen und Umweltauswirkungen; (ii) Migrationsbeschränkungen und Naturkatastrophen; und (iii) Soziale Sicherungssysteme und Gewalt. Einleitend untersuche ich die Erfolgsfaktoren von Maßnahmen, die von einer internationalen Entwicklungsinstitution unterstützt werden.

Acknowledgements

Für die längste Zeit meines Lebens lag eine Promotion außerhalb dessen, was ich mir vorstellen konnte und zugetraut habe. Deswegen möchte ich allen Menschen danken, ohne die dieses Projekt nicht mglich gewesen wäre:

Zuallererst möchte ich Prof. Dr. Eva Terberger danken. Dein entgegengebrachtes Vertrauen im Rahmen der Trainee-Doktoranden Stelle haben für mich eine Promotion erst greifbar gemacht.

Mein größter Dank geht ebenfalls an meine Doktormutter, Prof. Dr. Krisztina Kis-Katos. Ich weiß es sehr zu schätzen, dass du dich auf das 'Experiment' externer Doktorand eingelassen hast. Vom ersten Tag an hast Du dir immer Zeit genommen und trotz der räumlichen Distanz habe ich mich zu jeder Zeit hervorragend betreut gefühlt. Dank deiner wissenschaftlichen Exzellenz und Begeisterung für Forschung und Austausch werde ich immer gerne auf diese sehr lehrreiche Zeit zurückblicken.

Im Rahmen der Promotion hatte ich die Gelegenheit, mit vielen ausgezeichneten Wissenschaftlerinnen und Wissenschaftlern als Co-Autoren, Fakultäts-Kolleginnen und Konferenzteilnehmern in Kontakt zu kommen. Dieser Austausch hat meine Promotion sehr bereichert. Insbesondere geht mein Dank an Prof. Dr. Andreas Fuchs für das hilfreiche Feedback im Rahmen der Zweitbetreuung und an Dr. Elías Cisneros für die großartige Zusammenarbeit.

All diese Unterstützung wäre jedoch unerheblich gewesen, wenn nicht die wichtigsten Grundlagen gelegt worden wären und das richtige private Umfeld vorhanden gewesen wäre. Rita und Onno, ihr habt mir immer eine Neugier und Offenheit vorgelebt, die diese Promotion erst möglich gemacht haben. Euer Glaube an mich und Eure Unterstützung sind von unschätzbarem Wert. Für die schönen Momente abseits der Arbeit geht mein Dank zu guter Letzt an meine Familie und Freunde, die immer ein offenes Ohr für mich hatten.

Contents

List of Figures **viii**

List of Tables **x**

1 Introduction **1**

1.1 Motivation 1
1.2 Research agenda 4

2 Volume, risk, complexity: What makes development finance projects succeed or fail? **9**

2.1 Introduction 10
2.2 Development finance and project evaluations at KfW Development Bank 13
2.2.1 Bilateral development finance 13
2.2.2 Project evaluations 14
2.3 Data and descriptive analysis 15
2.3.1 Meta sample construction and summary statistics . . . 15
2.3.2 Descriptive analysis 18
2.4 Methodology 21
2.4.1 Empirical specification 21
2.4.2 Independence of ratings 23
2.5 Empirical results 25
2.5.1 Cluster (1): Project financing 26
2.5.2 Cluster (2): Project structure 27
2.5.3 Cluster (3): Project complexity 29
2.5.4 Cluster (4): Project risks 31
2.5.5 Contextual variables 32
2.6 Heterogeneities and robustness 34
2.6.1 Empirical results by region 34
2.6.2 Empirical results by sector 35
2.6.3 Individual DAC-criteria 36

2.6.4 Robustness 37
2.7 Conclusion 38

3 Losing territory: The effect of administrative splits on land use in the tropics 40
3.1 Introduction 41
3.2 Background 43
3.2.1 Indonesia's decentralization reforms 43
3.2.2 Forestry and natural resource management 46
3.3 Theoretical framework 47
3.4 Data and methodology 51
3.4.1 Data 51
3.4.2 Econometric framework 52
3.5 Results 55
3.5.1 Main results 55
3.5.2 Mechanisms 58
3.6 Conclusion 63

4 Confined to Stay: Natural Disasters and Indonesia's Migration Ban 65
4.1 Introduction 66
4.2 The Indonesian context 70
4.2.1 Natural disasters 70
4.2.2 International migration 71
4.2.3 The moratorium: Indonesia's emigration ban 72
4.3 Data 73
4.3.1 Indonesian village census 73
4.3.2 Additional sources 74
4.3.3 Village panel 75
4.4 Empirical strategy 76
4.4.1 Natural experiment: The emigration ban 76
4.4.2 Identification: The triple difference 77
4.4.3 Causal interpretation 78
4.5 Results 79
4.5.1 Migration flows after the ban 79
4.5.2 Disasters and migration under the ban 80
4.5.3 Parallel trends 82
4.5.4 Robustness checks 83
4.5.5 Mechanisms and discussion 88
4.6 Conclusion 92

5 Cash Transfers and Violent Crime in Indonesian Communities 94
5.1 Introduction 95
5.2 Background 98
5.3 Data 100
5.3.1 Data sources and samples 100
5.3.2 Variable construction 104
5.4 The effects of the CCT on violent crime 105
5.4.1 Econometric framework 105
5.4.2 Main results 107
5.4.3 Robustness checks 109
5.4.4 Additional evidence from crime victim surveys 112
5.5 Potential mechanisms 115
5.5.1 Benefits of crime 116
5.5.2 Costs of crime 118
5.6 The plausibility of the idleness mechanism 121
5.6.1 PKH and young men's idleness: External validity . . . 121
5.6.2 Idleness and crime: On the timing of criminal activities 123
5.6.3 Idleness and crime: A vignette experiment 124
5.7 Conclusion 127

6 Concluding Remarks 130

A Volume, risk, complexity: What makes development finance projects succeed or fail? 133
A.1 Tables 134
A.2 Figures 144
A.3 Methodology 145
A.3.1 Calculation of macro variables 145
A.3.2 Extra-/Interpolation of macro variables 146
A.3.3 Within- vs. between-country analysis 147

B Losing territory: The effect of administrative splits on land-use in the tropics 148
B.1 Tables 149
B.2 Figures 152

C Confined to Stay: Natural Disasters and Indonesia's Migration Ban 156
C.1 Background 157
C.1.1 Questions of key variables included in PODES 157

C.1.2 Criteria for the eligibility of poverty letters (*SKTM*) . . 158
C.1.3 Determinants of access to poverty letters (SKTM) . . . 159
C.2 Tables 160
C.3 Figures 168

D Cash Transfers and Violent Crime in Indonesian Communities 171
D.1 Tables 172
D.2 Figures 184
D.3 PKH and its Impact Evaluation 185
D.3.1 Data 185
D.3.2 A note on the construction of key variables 186
D.3.3 RCT-related Tables 187
D.4 SUSENAS Description 190
D.4.1 Sample construction and variables 190
D.4.2 PKH and Crime: PSM 192
D.4.3 SUSENAS Tables 194
D.5 Vignette Experiment 196

Bibliography 200

Declaration of Authorship 231

Author contributions 232

List of Figures

2.1 Forest plot of success ratings by region 19
2.2 Forest plot of success ratings by sector 20

3.1 Administrative reorganization in an exemplary district split . 44
3.2 District splits and forest cover across Indonesia 45
3.3 Spatial RDD: Initial forest cover and forest loss around new district boundaries 52
3.4 Dynamic SRDD effects: Deforestation 56
3.5 Dynamic SRDD effects: Expansion of oil palm area 60

4.1 Number of new restrictive immigration policies implemented worldwide 67
4.2 Main destination country and natural disasters in East Javanese villages 75
4.3 The effect of the moratorium: Change in female migrant stocks in villages with Saudi Arabia as main destination 80
4.4 Triple difference: Parallel trends 82
4.5 Average effect of disasters on poverty by type of disaster . . . 87

5.1 Official crime statistics 99
5.2 Community-level crime incidence and PKH roll-out 104
5.3 Robustness: Pre-trends and treatment effects by year, normalized 110

A.1 Distribution of OECD DAC-ratings 144

B.1 Descriptives: Frequency of splits 152
B.2 Identification check: Density of the forcing variable 153
B.3 Robustness: Deforestation effects for varying bandwidths . . . 154
B.4 Robustness: Shifting boundaries in space 155

C.1 Disaster events in the period 2003–2005 168
C.2 Stock of emigrants by gender and destination in 2005 169

C.3 Annual flows of documented migrants per destination 169
C.4 Placebo: Average effects of disasters by villages' top destination countries 170
C.5 Geocoded weather stations 170

D.1 Overview on the national roll-out of PKH 184
D.2 NVMS coverage 184

List of Tables

2.1 Results: Evaluation-specific characteristics 24
2.2 Results: Project financing 27
2.3 Results: Project structure 29
2.4 Results: Project complexity 30
2.5 Results: Project risks 32
2.6 Results: Country context 33

3.1 SRDD effects: Deforestation in child vs. mother districts 57
3.2 SRDD effects: Heterogeneities by ethnic composition 61
3.3 SRDD effects: Heterogeneities by closeness to the new political center 62

4.1 Average effect of disasters on poverty 81
4.2 Average effect of disasters on poverty by the type of rice production 89
4.3 Average effect of disasters on poverty or internal migration by terciles of initial international emigration rate 92

5.1 The effects of the CCT program on violent crime 108
5.2 Alternative measure: The CCT's effects on the probability of being a victim of violent crime (2007–2011) 114
5.3 RCT: The short-run effects of the CCT program on assets, expenditures and behavior 117
5.4 RCT: The short-run effects of the CCT program on work, schooling and idleness by cohort 120
5.5 The country-wide effects of the CCT program on work, schooling and idleness by cohort 122
5.6 The effects of PKH by type and timing of violent crime 124
5.7 Vignette Experiment: Likelihood of suspect committing the crime 126

A.1 Representativeness of sample 134
A.2 Summary statistics 135

A.3 Codebook: Outcome and project variables 136
A.4 Codebook: Macro, control and analytical variables 137
A.5 Results: Regional split 138
A.6 Results: Sectoral split I 139
A.7 Results: Sectoral split II 140
A.8 Results: OECD DAC ratings 141
A.9 Results: LASSO estimates 142
A.10 Robustness: Alternative estimations 143

B.1 Descriptives: Summary statistics 149
B.2 Placebo checks: Continuity of topographic and socio-economic characteristics in 2000 150
B.3 Robustness: Dynamic SRDD effects on deforestation 151
B.4 Robustness: SRDD effects on deforestation using quadratic fit 151

C.1 Summary statistics 160
C.2 Average effect of disasters on poverty: With and without control variables 161
C.3 Average effect of number of disasters on poverty 162
C.4 Average effect of disasters on poverty using alternative poverty measurements 162
C.5 Average effect of disasters on poverty in t-1 163
C.6 Average effect of disasters on poverty: Additional robustness checks 163
C.7 Average effect of disasters on poverty controlling for financial transfers 164
C.8 Average effect of disasters on poverty in the presence of spillovers 165
C.9 Average effect of disasters on poverty: Conley standard errors 165
C.10 Average effect of disasters on population growth and poverty 166
C.11 Average effect of extreme rainfall events on poverty 166
C.12 Probability to receive poverty letters (SKTM) 167

D.1 NVMS data: Descriptive statistics on the types of violent crime 172
D.2 Sample selection due to NVMS coverage 173
D.3 Overview on variable construction 174
D.4 Roll-out determinants: Explaining year of PKH roll-out 175
D.5 Robustness: Assessing the presence of spillover effects 176
D.6 Robustness: Results from randomization inference 177

D.7 Robustness: PKH effects after adjusting standard errors 177
D.8 Robustness: PKH effects by sample construction 178
D.9 PKH effects on the intensive margin of violent crime 178
D.10 PKH effects on community-level socio-economic development 179
D.11 Robustness: PKH effects on conflict 180
D.12 Alternative measure: CCT access intensity and the probability of being a victim of violent crime (2007–2011) 180
D.13 PKH effects on the community-level presence of police stations 181
D.14 RCT: PKH effects on peer-group inequality 181
D.15 PKH effects on crime: The role of community-level targeting inequality 182
D.16 Robustness: Middle-run effects of PKH on work, schooling and idleness by PKH eligibility 183
D.17 PKH conditionality criteria 185
D.18 PKH benefit payments (annual; Indonesian Rupiah) 185
D.19 Covariate balance at baseline 187
D.20 RCT sample: Covariate balance at the individual level at the time of the baseline 188
D.21 RCT sample: Compliance at the community level 188
D.22 RCT household survey: Receiving PKH at endline 189
D.23 RCT household survey: Attrition 189
D.24 RCT household survey: Determinants of attrition 189
D.25 SUSENAS 2007-2019: Share (%) of households experiencing crime 194
D.26 Alternative measure: PKH effects on the probability of being a victim of violent crime (PSM) 194
D.27 The effects of PKH on the reporting of violent crime to the police 195
D.28 Summary statistics (vignette experiment sample) 198
D.29 Balance Table (vignette experiment sample) 199

Chapter 1

Introduction

1.1 Motivation

Poverty remains one of the most pressing global challenges, affecting people's livelihood in far more than the mere income dimension: The poor suffer from limited access to education, health care and nutrition, face discrimination, and lack societal participation (Alkire et al., 2014). Tackling poverty in all its dimensions is therefore at the core of national governments' and international institutions' policy agenda, contributing towards Sustainable Development Goal (SDG) 1 *to end poverty in all its forms everywhere* by 2030 (UN General Assembly, 2015). While considerable progress has been made, in 2019 an estimated 700 million people—or 9% of world population—still lived in extreme poverty (World Bank, 2023). Recently the Covid-19 pandemic has even exacerbated these figures, pushing millions of people into poverty within a brief period (World Bank, 2021). Future prospects are also pessimistic, as climate change has the potential to endanger future eradication efforts (Hallegatte et al., 2017): At an estimated worldwide poverty rate of 7.4% in 2030 (Lakner et al., 2022), projections highlight the need to continuously address the causes and consequences of poverty around the globe.

The underlying causes of poverty are as complex and multi-dimensional as their implications (Brady, 2019). It is generally agreed however that economic development is a crucial ingredient to long-term poverty reduction (Islam, 2004; Dollar et al., 2016). While this perspective puts emphasis on the private sector, the role of the government and its policies is equally decisive: It establishes a conducive environment for economic development (Acemoglu et al., 2005) and provides social protection systems for people in need (Banerjee et al., 2022). The right policies are therefore at the heart of successful poverty alleviation, yet often constrained by limited financial resources and technical capacity. SDG 1a consequently calls for *enhanced development cooperation* (UN General Assembly, 2015), reflecting that developing countries often rely on financial and technical resources provided by international donor organizations. These institutions will continue to play a key role in shaping interventions targeted at poverty eradication (Greenhill et al., 2015).

From a global perspective, developing countries have made progress in terms of economic development and poverty eradication to very different degrees. Within the World Bank's *income status* framework, this is reflected in categories classifying countries' developmental status and needs. Nations that graduate to higher income groups often experience comparable socio-economic, political and environmental developments: Structural economic transformation (Herrendorf et al., 2014), altered migration patterns (Cattaneo et al., 2019a), environmental degradation (Grossman et al., 1995) and (social) governance reforms are only few of many changes. To reach economic development and poverty reduction targets as outlined in the SDGs, public policy needs to account for these intertwined phenomena with targeted interventions adapted to countries' specific needs. In an increasingly complex environment due to, e.g., climate change, such considerations will become ever more important in the future (Beegle et al., 2019).

Assessing the effectiveness of such policies is an essential prerequisite for their careful design. In the context of development cooperation, a long-standing scholarly debate has questioned its effectiveness (Bourguignon et al., 2007; Easterly, 2007; Doucouliagos et al., 2008). This comprises not only whether intended goals were accomplished, but also the less prominent debate on unintended—both positive and negative—consequences of development interventions (Koch et al., 2018). Only more recently, a growing body of research on the unintended and, in particular, adverse impacts of international donor-supported development interventions has emerged (Collier et al., 2007; Crost et al., 2014; Nunn et al., 2014; Isaksson et al., 2018). This literature has been fueled by the growing use of rigorous evaluations, better data coverage and the realization that more scrutiny on unintended consequences is needed (Jabeen, 2016). Researchers can thus increasingly uncover policies' side effects, crucially complementing assessments of what development interventions bring about. This particularly holds because adverse side effects have the potential to confine otherwise desirable impacts. Against the backdrop of immense financial gaps in fostering development (OECD, 2022b), more evidence is essential for informed policy-making that translates into intended, beneficial impacts (Jabeen, 2016; Koch et al., 2018; Marschall, 2018).

In this thesis titled *Local Dimensions of Development*—based on four chapters written with my co-authors—I aim to make a contribution towards understanding the unintended consequences of interventions aimed at promoting development. I discuss three policies in the context of three distinctive phenomena that developing countries often experience: (i) Governance reforms and environmental degradation; (ii) migration restrictions and natural disasters; and (iii) social protection systems and violence. As introductory analysis, chapter 2 examines the success determinants of policy interventions supported by an international development institution. Results from this meta-study indicate that policy design adapted to the country context is crucial, motivating a more detailed analysis of individual policies. In chapter 3, I analyze a policy often prescribed by international donors to improve public service delivery—decentralization reforms—in the context of environmental protection. Findings implicate that re-shaped government responsibilities altered incentives for local government actors, resulting in temporary, but not sustained improvements of forestry protection as an unintended, yet positive side effect. As part of chapter 4, I examine the consequences of a policy that restricted emigration in a disaster-prone and migration-dependent context. Results imply that curtailing migration and thus associated remittances has detrimental impacts on communities' capability to cope with natural disaster-induced income shocks. Lastly, in chapter 5 I analyse how an important poverty alleviation policy—conditional cash transfers—relates to violence. As a considerable adverse side effect, results suggest that such payments can fuel incidents of violence.

Of the four thesis chapters, all but chapter 2 are set in Indonesia. The country's recent history makes it the ideal case study to analyze policies in the context of developmental patterns experienced in the Global South: First, the country is a major emerging market with the world's fourth largest population. Second, it is a resource- and biodiversity-rich archipelago, making it prone to natural disasters and resource exploitation. Third, it has embarked on the largest governance reforms witnessed in recent decades, piloting social protection policies that are now commonplace around the world (Cahyadi et al., 2020). Fourth, it has a religiously and ethnically very diverse population, resulting among others in heterogeneous migration dynamics. In short, along with accessible and superior data from official statistics, Indonesia represents the ideal setting to study local dimensions of development policies and their unintended consequences. My empirical research combines detailed information on the implementation of the policies studied with official Indonesian village census (PODES) and remotely sensed data. The consistent unit of observation across my studies—Indonesian villages, or *Desa*—allows for tracing heterogeneous development patterns at the lowest administrative level. In combination with topical causal inference methods, I can therefore derive precise and evidence-based relationships.

The thesis consists of four chapters corresponding to four stand-alone research articles, followed by concluding remarks. Individual (co-) author contributions are appended to the thesis. In the following, a short summary of each chapter is given.

1.2 Research agenda

Development policy success determinants

Global challenges related to climate change, poverty and conflict are particularly felt in developing countries, rendering international cooperation via financial transfers and technical training ever more important (Greenhill et al., 2015). Yet the effectiveness of such development cooperation—where policies are often implemented as individual projects—is debated (Easterly, 2007). While the recent wave of rigorous project evaluations can contribute towards understanding the intended impacts at the individual policy level, it cannot explain *under which circumstances* impacts materialize on a more aggregate level (Denizer et al., 2013). This understanding is crucial to designing the most impactful policy measures. Existing literature has highlighted the role of policy design features—e.g., its funding and implementation—vs. country characteristics such as the economic growth for the success of measures, mostly in the framework of policies supported by multilateral donors (Denizer et al., 2013; Feeny et al., 2017; Briggs, 2020). In this context, the contribution I make together with my co-authors in chapter 2 is twofold:[1] First,

[1] Yota Eilers, Jochen Kluve, Jörg Langbein and Lennart Reiners: "Volume, risk, complexity: What makes development finance projects succeed or fail?".

I significantly extend the depth of information on policy design features, allowing for more detailed insights into the success determinants of development policies using more than 30 characteristics. Second, the data covers a bilateral donor agency, German KfW Development Bank, complementing existing insights based mostly on multilateral organizations data with more than 5,000 individual success measurements.

In an empirical framework, I first document that in line with multilateral donors, KfW development projects' success varies more within than between countries. This motivates investigating individual policy's characteristics in more detail, which our comprehensive data allows for. Correlating these characteristics with measures of policy success derived from evaluation reports, two observations stand out: First, I find that ex-ante risks such as partner capacities and negative externalities are well identified, yet often cannot be mitigated sufficiently to make policies successful. Second, policy design choices correlate with success very heterogeneously across regions and sectors, underlining that *one size does not fit all*. Results highlight the importance of each policies' careful design to the local context—even within a country—in order to foster intended developmental impact.

Based on three key Indonesian policies from the 21st century, in the following chapters I illustrate how design choices can yield unintended consequences that can peril their intended impact.

Governance reforms and environmental degradation

International organizations such as the World Bank and IMF often mandate large-scale decentralization reforms in developing countries. While these reforms are ultimately aimed at improving public service delivery, research has found mixed evidence with regards to this objective (Gadenne et al., 2014). Because local governance is key to protect natural resources (Wehkamp et al., 2018) and such reforms fundamentally reshape governance at all levels, they can indirectly and unintentionally affect resource exploitation patterns. In chapter 3, together with my co-authors, I study how decentralization in Indonesia affected local governments' incentive to foster land-use changes on forested lands.[2] The country embarked on "big-bang" decentralization reforms in 1998 (Fitrani et al., 2005), capacitating local governance by both

[2] Elías Cisneros, Krisztina Kis-Katos and Lennart Reiners: "Losing territory: The effect of administrative splits on land use in the tropics".

vertical power devolution and the creation of new administrative units. Existing literature has found conflicting evidence on the impact of these reforms on deforestation (Burgess et al., 2012; Alesina et al., 2019), and disregarded important incentives by local actors (Grossman et al., 2014). By analyzing the reforms and associated creation of new administrative units in a spatial regression discontinuity design framework at the lowest administrative level, I address the identified gaps and provide causal effects of the reforms on forest use.

Results indicate that the decentralization reforms induced immediate behavioral response at the local level: Not only did deforestation rates decelerate in new administrative units after their creation, but already up to two years before such administrative splits were legislated. These effects suggest considerable anticipatory, strategic action by local policy makers which led to temporary improvements in forest protection. This materializes via altered land-use decisions by local policy makers as a response to the reforms: I document decreased expansion rates of oil palm plantations in areas that will form part of new administrative units before the split. These plantations often replace primary forests, and associated rents are an important income source for local governments. In the medium run, both deforestation and oil palm expansion accelerate in new administrative units, reflecting a self-interest in rents on part of new local governments. The results illustrate how policies can have far-reaching impacts beyond their original intent.

Migration restrictions and natural disasters

International migration and accruing remittances are a major income source for many developing countries around the world (Yang, 2011). Migration also serves as an important risk-coping mechanism against the backdrop of aggravating climate change (Hornbeck, 2012; Kleemans et al., 2018). However, this response is increasingly curtailed, as both sending and receiving countries legislate more restrictive policies (Haas et al., 2018). While this trend has the potential to deprive sending communities' ability to absorb natural disaster-induced income shocks, this relationship has not been empirically demonstrated in the literature. Together with my co-author, in chapter 4 I show how a policy aimed at protecting migrants had adverse side effects:[3] Indonesia introduced an emigration ban for all women wanting to migrate to

[3]Andrea Cinque and Lennart Reiners: "Confined to Stay: Natural Disasters and Indonesia's Migration Ban".

Saudi Arabia in 2011 in order to protect them from abuse abroad. Saudi Arabia was an important destination for Indonesian migrants, and as a consequence the stock of overseas workers significantly decreased. However, due to heterogeneous migration ties, the policy constrained communities' migratory options abroad to very different degrees. I compare how this affected communities' ability to cope with natural disasters—an ever more common phenomenon in Indonesia—in the aftermath of the ban.

In the analysis, I exploit the ban's sudden implementation as a natural experiment in a triple-difference estimation framework. Results show that the policy considerably reduced the number of overseas workers in villages with strong migration ties to Saudi Arabia. Once hit by natural disasters in the aftermath of the ban, these villages experience significant poverty increases. These results are among the first causal estimates in the literature and further illustrate the mechanisms underlying the observed effect. In an environment of curtailed migratory options, natural disaster-induced income shocks can no longer be compensated for via emigration and remittances. Against global trends of intensifying climate change and migration restrictions, this nexus will translate to communities in developing countries around the world. At the same time, the results exemplify how policies can have unintended consequences: Protecting citizens came at an adverse cost for many communities, highlighting the need for careful policy design that takes account of all potential benefits and downsides.

Social protection and violent crime

Social protection systems are important policy instruments to alleviate poverty, particularly in developing countries where the share of vulnerable population is significant (Banerjee et al., 2022). In these often institutionally-weak and violence-prone countries, large-scale payments schemes can however incentivize elite capture and other behavioral responses that trigger impacts beyond policies' original intent. A growing literature has examined the effect of welfare policies on conflict and crime (Crost et al., 2016; Bratsberg et al., 2019; Carr et al., 2019), but gaps in methodological and regional focus remain, particularly in the context of development policies. Together with with my co-authors, in chapter 5 I add evidence on the welfare-violence nexus by examining Indonesia's flagship conditional cash transfer program, *PKH*, and its effect on violent crime.[4] By studying an unintended outcome, my analysis adds to a more holistic assessment of the program which has been found to be successful in targeted dimensions like educational attainment (Cahyadi et al., 2020).

In the analysis we combine data from both an experimental pilot and national program roll-out with panel data on local-level violence incidents. Assessing the program's impact using two-way fixed effects estimation, I find robust evidence for considerable violence increases in areas with program access. The results contradict crime-reducing impacts found for similar Latin American programs (Camacho et al., 2013; Chioda et al., 2016), mirroring how one policy can have conflicting impacts in different environments. In the framework of a supply-side crime model, I show that the program induced idleness among beneficiary household members not directly targeted by the program, contributing to the observed violence hikes. Such intra-household spillover effects as a response to social protection policies have been documented before (Bratsberg et al., 2019), yet present a novel finding in relation to cash transfers and crime. The results add an additional piece of evidence on the potentially adverse side effects of policies targeting development and poverty reduction. Policy making will crucially depend on accounting for such unintended consequences as otherwise the intended impact of welfare policies can be significantly thwarted.

[4] Elías Cisneros, Krisztina Kis-Katos, Jan Priebe and Lennart Reiners: "Cash Transfers and Violent Crime in Indonesian Communities".

Chapter 2

Volume, risk, complexity: What makes development finance projects succeed or fail?

Yota Eilers, Jochen Kluve, Jörg Langbein and Lennart Reiners

Abstract

In 2021, governments around the world committed more than USD 170 bn. to official development assistance. Despite these high contributions, systematic assessments of the determinants of success—or failure—of development projects are still limited, particularly for bilateral development cooperation. In this paper, we provide such a systematic, quantitative analysis: We construct a unique database covering over 5,600 evaluation results—success ratings—for bilateral development cooperation projects financed through one of the biggest global donors, KfW Development Bank. Our detailed data on project characteristics provides insights into important factors along the entire project life-cyle: (i) In terms of *project financing*, we find a significant positive correlation between the financial budget volume of the project and its success ratings, ceteris paribus. Second, concerning the (ii) *project structure*, the type of project partner—government, private sector, multilateral organizations—shows no significant correlation with project success; (iii) *project complexity* as measured by longer preparation and implementation exerts negative influence. Regarding (iv) *project risks*, a highly relevant and significant predictor for less successful projects is the share of ex-ante identified risks that eventually materialized—suggesting that project designs often correctly identify the relevant risks but are not able to mitigate (all of) them. Finally, concerning (v) the *project context* there is some indication that higher rates of GDP p.c. growth are positively associated with project success.

2.1 Introduction

Today's world is shaped by multiple crises like climate change, rising inequality and an increasing number of conflicts. Their consequences are particularly felt in developing countries, where the capacity to financially cushion the impact of these crises is limited. Many see development finance as a panacea to these issues with an increasing volume of development cooperation being committed over the past decades. At the same time, the effectiveness of such commitments is ambiguous, with only scarce evidence available (Qian, 2015). A common evaluative approach to infer the impact of development projects are rigorous analyses using counterfactual methods. These project-level results are key to assess individual project effectiveness; however, they cannot inform a more aggregate perspective as to under which circumstances and conditions projects can succeed in delivering intended impacts. Only more recently, studies have tried to analyze the role of country- as well as project-characteristics on project outcomes in more detail, mostly using data from the World Bank and the Asian Development Bank (ADB) (e.g., Denizer et al., 2013; Feeny et al., 2017; Ashton et al., 2023).[1] In the context of this emerging strand of aid effectiveness literature, project outcomes are commonly measured using success ratings assigned by institutions' independent evaluation units (Honig et al., 2022).

The systematic assessment of the degree of success—or failure—of development cooperation projects is key for several reasons. First, it is important for the individual project itself, to serve transparency and accountability, and to learn whether the project, in fact, worked or not. Second, it informs the design of future projects. Third, if these assessments are conducted systematically across a (large) set of development projects, they can feed into structures of institutional learning. Fourth, such a systematic analysis of development cooperation project results is informative for both the institution implementing the projects as well as for other donors implementing similar projects.

This paper analyzes the success determinants of KfW Development Banks' projects. KfW is one of the largest bilateral donors worldwide and manages Germany's development finance commitments. With a portfolio spanning across most developing countries and sectors, KfW's engagement scope is comparable to that of large multilateral donors. We compile a database of KfW project evaluations accumulating to more than 5,600 individual ratings

[1] See Ashton et al. (2023) for a recent literature review.

that are representative of the institution's entire portfolio. The data contain extensive information on project structure, financing, complexity and risks, which we coded from evaluation reports as well as internal project documentation. After merging these data with contextual country-level characteristics, we can thus analyze a set of holistic factors that matter for the success of these projects.

Our results document the most in-depth bilateral donor contributions yet. We concur with past research results that project success varies more within than between countries. Project characteristics thus deem particular attention in comparison to the country context: While a favorable economic context supports the success of the projects, factors such as the project financing structure and complexity exert much bigger influence. Notably, we find that initial project design matters less than expected—neither institutional setup of the recipient country implementing agency nor whether a project was co-financed is significant. In contrast, the financial volume of a project as well as counterpart contributions favor successful outcomes. Unsurprisingly, projects' complexity measured along multiple dimensions displays strong and negative correlations with assigned ratings.

In the debate on development projects' effectiveness, our results are particularly relevant because they implicate that characteristics under the influence of donor agencies and partners matter most for projects delivering on their intended impact. To improve the success of projects, particularly project risk anticipation and management during implementation is key. While identification appears to work well, mitigation measures implemented do not appear to be sufficient to keep the projects on track. We also uncover important heterogeneities across regions and sectors which further underlines that "one size fits all" does not hold. Lastly, partner ownership and integration in the project does matter. Given that research has found project characteristics to correlate similarly with project success across different donors (Bulman et al., 2017; Briggs, 2020), our results are also relevant for the diverse panorama of donor institutions as well as recipient countries.

With this paper, our contribution to the literature is fourfold: First, we considerably expand the depth of data on project characteristics. Existing research has found that these aspects matter more than the country context (Denizer et al., 2013; Bulman et al., 2017; Feeny et al., 2017), highlighting the need for more detailed data. By coding more than 30—partially novel—variables across projects' entire life-cycle from previously unexploited sources, we provide the most detailed project data to our knowledge. This directly addresses the gap in existing literature regarding quality and depth of micro-level variables.[2] Second, most systematic analyses of project success determinants has focused on multilateral agencies (e.g., Mubila et al., 2000; Denizer et al., 2013; Feeny et al., 2017).[3] Because bi- and multilateral development cooperation commitments function differently (Biscaye et al., 2017; Findley et al., 2017; Dreher et al., 2022), the lack of detailed bilateral donor analyses presents a notable gap. With more than 5,600 individually assigned ratings from projects in 96 countries, we introduce the most extensive and up to date single-donor database. As third contribution, we provide a more nuanced perspective of what constitutes development project success: Existing research solely focuses on the overall project rating, neglecting the fact that projects' success is measured following distinct dimensions. By analysing a total of five success dimensions as measured by OECD-DAC criteria and splitting the sample by region and sector, we provide further insights into important heterogeneities of project success. Lastly, our approach presents more evidence on the quality and independence of evaluative ratings used in aid effectiveness analyses. Comparable research partially relies on self-assigned grades from project leaders heading the project (Denizer et al., 2013; Feeny et al., 2017; Rommel et al., 2020; Ashton et al., 2023). In comparison, all our data points are independently assigned by an evaluation department, which we can empirically discuss.

More generally speaking, the paper is also related to literature discussing the relationship between individual project or country characteristics and success in more detail (e.g., Chauvet et al., 2010; Kilby, 2015). For example, we can provide additional empirical evidence on literature discussing the role of implementing agencies (Shin et al., 2017; Winters, 2019; Marchesi et al., 2021). By adding evidence on a previously understudied donor, we also add to the

[2]For example, our variables respond to Bulman et al. (2017), who argue that "[t]his points to the importance of further work to understand the sources of this variation, for example, by systematically measuring the contribution to project success of project implementing agencies within recipient governments.", or to Ashton et al. (2023) who, based on a recent literature review, conclude that "(...) the quality and suitability of project design, have rarely been investigated (...)" and that existing literature has been "(...) concerned mainly with easily observable characteristics like size, duration, and sector".

[3]Wood et al. (2020) researched Australian bilateral aid, a comparatively small donor. Honig et al. (2022) also compile data for bilateral donors including KfW, but the scope of project characteristics is considerably more limited and relies on web-scraped data.

discussion on the comparability of success correlates across donors (Bulman et al., 2017; Briggs, 2020).

The next section 2.2 provides background information on KfW Development Bank. Section 2.3 describes the data in detail and gives insights into descriptive and graphical analyses. Section 2.4 delineates the estimation strategy, and sections 2.5 and 2.6 present and discuss the regression results and robustness tests. Lastly, section 2.7 concludes.

2.2 Development finance and project evaluations at KfW Development Bank

2.2.1 Bilateral development finance

The German Development Bank (KfW) handles the majority of Germany's official Financial Cooperation (FC).[4] In 2021, for instance, KfW committed EUR 8.6 bn. (USD 10.2 bn.) to developing countries around the world (KfW, 2022a), making it one of the largest bilateral donors worldwide. The funds mainly stem from the German Federal Ministry for Economic Cooperation and Development (BMZ) and finance projects in Africa, Asia, Latin America and the Caribbean, and South-Eastern Europe. Sector-wise, these engagements address all areas from agriculture to water supply. The institution's breadth of engagement is comparable to that of large multilateral development financiers, but its bilateral nature makes it a particular interesting case to study given that bi- and multilateral aid operate differently (Biscaye et al., 2017; Dreher et al., 2022; Findley et al., 2017; Rommel et al., 2020).

Funds committed by KfW are implemented via projects that comprise dedicated investments. They are designed with and implemented by local—mostly public—partner agencies such as line ministries, with whom financing agreements are concluded, at times together with international co-financing institutions. This process entails defining the Theory of Change (ToC), outlining the results framework including target indicators for the project's performance. Once projects are completed, a completion report is conducted, summarizing the project's results from the perspective of KfW project managers. For a subset of projects, this is followed by an independent ex-post

[4] In contrast, Germany's other major institution dedicated to implementing development cooperation, GIZ, is responsible for Technical Cooperation (TC).

evaluation of project success (or failure), taking place approximately three years after project completion.

2.2.2 Project evaluations

At KfW, the independent evaluation department FCE (Financial Cooperation Evaluation) is responsible for carrying out project evaluations. A random sample of 50% of the projects, stratified by nine sectors, is drawn from all completed projects for each year. The sample is, hence, representative for KfW's entire FC portfolio. All ex-post evaluations conducted at KfW adhere to the internationally established OECD-DAC criteria. That is, each evaluation systematically assesses the five criteria (i) relevance, (ii) effectiveness, (iii) efficiency, (iv) impact and (v) sustainability of the given project.[5] Each criterion is rated on a discrete scale from 6 (best) to 1 (worst), i.e. ranging from "very good" to "highly unsatisfactory". In KfW parlance, scores 6-4 imply that the project was "successful", while scores 3-1 imply that the project was "unsuccessful" (KfW, 2022b). Each evaluation also assigns an overall rating.[6]

From the annual sample of projects, one third is evaluated by FCE staff, one third by external consultants, and one third by seconded colleagues from KfW's operational departments. The governance of every single evaluation, however, lies with FCE. That is, (i) FCE supervises external consultants and seconded colleagues; (ii) all reports are peer-reviewed internally within FCE; and (iii) the absence of conflicts-of-interest is ensured: Specifically, any person involved in the evaluation process must not have worked on the project or within the responsible department during its implementation. Each evaluation follows a structured process entailing conceptual design, desk study, on-site visit and/or support from a local consultant, and report writing. Summary evaluation reports—with standardized table of contents—are published on KfW's website.

[5] These five criteria were defined by OECD-DAC in the 1990s. A sixth criterion, "coherence", was only added in 2020, and therefore most evaluation reports in our sample cover five criteria. Note that one evaluation can cover two or more projects in one report if they are directly related. In this case individual grades for each DAC-criterion are still assigned for each project.

[6] In general, the overall rating is calculated as the rounded, unweighted average of the five criteria ratings. There is one specific exception, however: If one or more of the three criteria sustainability, effectiveness or impact are rated as unsuccessful (1-3), then the overall project cannot be rated as successful (4-6), independent of the ratings assigned to the other criteria. This particular scenario applying the so-called *Knock Out-Criteria* concerns 37 of the 1,124 project evaluations, or 3.3% of our sample.

The evaluations thus constitute an expert assessment of projects, following an internationally established methodology. Another benefit from exclusively relying on DAC-criteria is that all project evaluations are guided by the same normative framework—independent of regional or thematic focus—addressing concerns that development objectives cannot be compared across sectors (Denizer et al., 2013; Feeny et al., 2017). For the purpose of our study, therefore, the use of DAC-criteria provides us with a large sample of individual success ratings that were systematically and consistently assigned over the entire sampling period.

In contrast to the ADB and World Bank, KfW's evaluation portfolio does not include self-assessments from operational staff, and it is selected on a strictly random basis. Thus, our data are not prone to selection biases (Kilby et al., 2019), or to overly favorable ratings assigned by project managers themselves (Bulman et al., 2017; Ashton et al., 2023). Still, even when conducted by a formally independent evaluation department, the autonomy of such bodies can be called into question given that they are based within the institution (Denizer et al., 2013). While the same critique could, in principle, be translated to KfW, several reasons speak against it: First, the department head is recruited externally from academia, and reports directly and only to the executive board; second, evaluation results are publicly shared; third, the broad set of evaluators (FCE staff, external consultants, seconded operational staff) guarantees the absence of conflicts-of-interest; fourth, FCE's methodology is reviewed by an external body, the German institute for Development Evaluation (DEval). Moreover, we can empirically test the independence of assigned ratings and evaluator characteristics and discuss this in a dedicated part of section 2.4.

2.3 Data and descriptive analysis

2.3.1 Meta sample construction and summary statistics

The sample is constructed from all $N = 1,124$ evaluations of development finance projects that FCE conducted between 2007 and 2021, yielding a sample of $N = 5,608$ observations on project success ratings (five ratings per evaluation). This is, to our knowledge, the most extensive and up to date database on bilateral financial cooperation evaluations worldwide from a single donor.

It covers projects implemented in a total of 96 low and middle income partner countries and is representative for KfW's FC portfolio (cf. Appendix Table A.1). As evaluations are conducted after project completion, different project durations imply that our sample effectively contains development projects that started as early as 1990 and as late as 2019. Each of the 1,124 development finance projects in our sample has a unique ID, allowing us to merge system variables from KfW databases with information coded from evaluation reports, covering rich information on project characteristics (the "micro" variables). In addition, we combine these data with external statistics on contextual factors in the countries during the time of project implementation (the "macro" variables). Conceptually, the resulting analytical sample combines three types of data sources—(i) key variables coded from the hardcopy evaluation reports, (ii) KfW operational project databases, (iii) external data on economic indicators—and therefore corresponds to data constructed for quantitative meta-analysis (e.g., Card et al., 2018).

Dependent variable: Standardized project success ratings

The main variable of interest is the individual rating assigned for each DAC-criterion on a 6-1 scale for a given project. This allows us to utilize the granularity of the full set of individually assigned ratings instead of only relying on the overall project rating. Appendix Figure A.1 displays the respective distribution of each DAC-rating in the data at the project evaluation level ($N = 1,124$ each). The majority of overall success ratings (top left) are either "4" or "5", with more than 400 cases each, while only few evaluations rate projects in the most successful category "6" (44 projects in total). The overall share of ratings with "2" and "3"—i.e. projects evaluated as unsuccessful—amounts to 19%, or 204 projects. The overall mean success rating is 4.21. In sum, this descriptive statistic indicates aggregate project success and failure rates of about 80% and 20%, respectively. Whereas the success rate can thus be considered as high, the overall mean success rating is just above the threshold required for a successful appraisal ("4"), suggesting hitherto unused potential for improvement.

The remaining panels for the five individual criteria indicate several patterns: First, the distributions for "effectiveness" and "impact" are rather similar to the overall rating, each with an average rating of 4.34. Second, "relevance" displays the highest share of successful ratings with "5" and "6", and thus the highest overall mean rating (4.85). Third, both “efficiency" and “sustainability" show slightly less successful average ratings, attaining 4.07 and 4.18, respectively.

The core of our analysis uses the pooled sample of these individual ratings. For robustness and comparability with the related literature, we also construct two additional outcome variables: (i) An alternative overall project rating based on an unrounded, unweighted average of all five individual DAC-Criteria (“arithmetic rating"); and (ii) a binary variable indicating whether a given project was successful overall, i.e. whether the overall rating was “4" or above.

Explanatory variables (i): Micro-level characteristics

At the micro level, we construct more than 30 variables capturing all dimensions of key project characteristics. Specifically, we can distinguish the four clusters (i) *financing* of the project, (ii) *structure* of the project, (iii) *complexity* of the project, and (iv) *risks* for implementation. Appendix Table A.2 presents summary statistics for the main variables within each dimension, along with several macro variables and the distribution by sector.[7] The table shows the full sample (column 1) and a stratification by major regions, i.e. Sub-Saharan Africa (SSA, column 2), Asia/Oceania (3), Europe/Caucasus (4), Latin America and the Caribbean (5), and Middle East and Northern Africa (MENA, column 6). Recall that this sample of project evaluations is representative for KfW’s development finance portfolio, indicating that the majority of projects are in SSA ($N = 428$ evaluations), followed by Asia/Oceania ($N = 281$).

The “average" development finance project (column 1) has a total volume of EUR 41.7 million, 16% of which are contributed by the country counterpart (top panel on “Financing"). The panel on “Structure" shows that co-financing occurs in 21% of projects on average, varying across regions from 11% (MENA, column 6) to 30% in SSA (column 2). The average number of institutions involved in a development finance project is 4. The variable “project manager turnover" relates the total number of project managers in a given project to the project duration in years—implying that at an average of 0.48, every second year the project manager changes.

Looking at project “complexity" (third panel), the average project duration amounts to 7 years, ranging by region from 5.7 (Europe/Caucasus) to almost

[7] For a detailed codebook of all variables used, see Appendix Table A.3 and A.4.

9 years (MENA). These averages relate to the fact that in MENA the execution of projects is delayed in 48% of cases, while this is the case for only 12% in Europe/Caucasus. The overall average share of delayed execution is 23% (column 1). The share of technically complex projects ranges widely from 15% in Latin America to 67% in Asia/Oceania (average: 48%).

Project appraisals identify and specify potential risks for project success ex-ante. As the fourth panel (“Risk") indicates, the average number of ex-ante identified risks is 4, with very little variation across regional sample splits. Our data also capture to what extent these risks actually occurred during implementation: 55% of ex-ante identified risks occurred in practice, an average that is somewhat higher in SSA (62%), and somewhat lower in Europe/Caucasus (49%).

Explanatory variables (ii): Macro-level contextual factors

Data on the country context that projects were implemented in are taken from official, publicly accessible databases and are merged to our micro-variables using country ISO-codes and information on the project life-cycle: Indicators are always measured for the specific country at the specific time the project was implemented. We incorporate four variables in our analyses: GDP p.c. growth, measures of democracy as well as fragility and total population.[8]

The bottom panel of Table A.2 displays the distribution of projects by sector. Some patterns by region are notable: In SSA (column 2), water supply (17%) and health (19%) are major sectors, the latter also being the case in Asia/Oceania (column 3). In Europe/Caucasus, water supply (28%) and finance (24%) are the main sectors, while in Latin America agriculture and environment are predominant (34%). In MENA, again water supply plays a major role (32%, column 6).

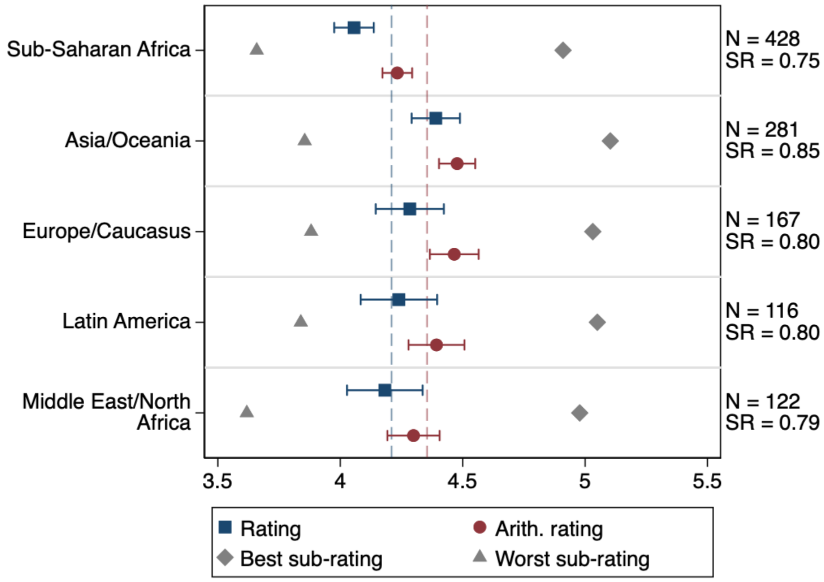
2.3.2 Descriptive analysis

Both the country and the sector where projects are allocated constitute two of the most distinctive project characteristics, as highlighted in the previous section. Indeed, typically donor institutions are institutionally organized along these dimension. This is also the case for KfW Development Bank and reflects how vital this distinction is for project implementation processes. The

[8]Sections A.3.1 and A.3.2 in the Appendix detail how averages over the time and missing observations are computed for these variables.

descriptive analysis therefore continues with a visual inspection of project success patterns by region and sector, respectively, using forest plots. This representation also reflects the meta-analysis nature of our data.

FIGURE 2.1: Forest plot of success ratings by region



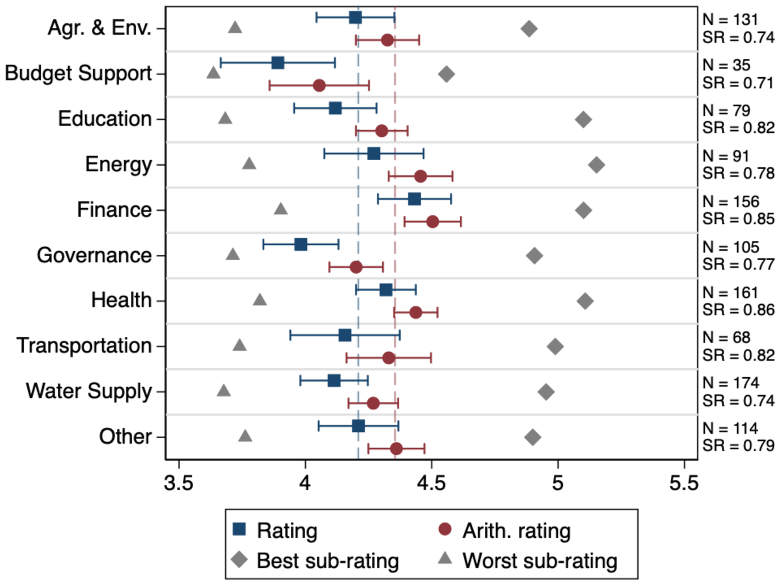
Note: The figure displays mean values of evaluation ratings by region: Blue squares denote average overall ratings (i.e. calculated from the rounded unweighted overall grade assigned to a project in the evaluation), red dots denote average arithmetic ratings (i.e. the unrounded arithmetic mean of the five DAC criteria ratings), diamonds denote means of the highest DAC-ratings per project and triangles denote means of the lowest DAC-ratings per project. 95% confidence intervals illustrated by whiskers. The blue and red dashed lines mark the sample mean of overall and arithmetic, respectively. "SR" denotes the success rate as the share of projects that received an overall rating of 4 or above. Observations are weighted by the inverse number of projects evaluated in the corresponding evaluation report. The y-axis on the right hand side gives the number of observations per category. 10 projects implemented in multiple regions excluded.

Overall success rating by region

Figure 2.1 shows a forest plot of overall success ratings by region. The blue square represents the average *overall grade*, i.e. the rounded (to a full grade) average of the five criteria grades that is reported in the evaluation report. The red dot represents the average *arithmetic grade*, i.e. the unrounded, unweighted average of the five criteria grades. The respective average of the lowest (triangle) and highest (diamond) DAC-criterion rating are also shown, as is the success rate (SR) for each region.

The figure indicates several patterns. First, the overall, rounded rating assigned to the project in the evaluation is always lower than the arithmetic mean of the individual criteria ratings. The difference, however, is not very large (4.21 overall vs. 4.36 arithmetic). Second, the respective regional averages are relatively close to the overall averages rather than widely dispersed. However, third, there are some visible regional differences. In SSA, both means are statistically significant below the overall means (4.06 overall, and 4.23 arithmetic). This is also reflected in the success rate, which is the lowest in this region at 75%. In Asia/Oceania, on the other hand, the mean ratings are statistically significant above the overall means (4.39 and 4.48, respectively). Finally, in SSA and MENA the worst sub-rating (triangle) is consistently lower than in the other regions.

FIGURE 2.2: Forest plot of success ratings by sector



Note: The figure displays mean values of evaluation ratings by sector. See notes for Figure 2.1.

Overall success rating by sector

Figure 2.2 displays the corresponding forest plot of average success ratings by main economic sector. The y-axis on the right indicates the sectoral distribution of project evaluations and reflects the summary statistic shown in Table A.2: Inter alia, the three largest sectors are finance, health, and water supply, with a share of 15% each ($N = 156$, $N = 161$, and $N = 174$ evaluations, respectively).

The figure illustrates notable variation in project success ratings across sectors: Looking at overall average grades (i.e. blue squares), finance (4.44) and health (4.32) display the most successful grades, the former statistically significant above the overall average. Projects in the energy sector are also comparatively successful and slightly above average (4.27), with relatively wide confidence bands. On the other hand, budget support (3.89) and governance sector (3.98) lie statistically significant below the overall average.

2.4 Methodology

2.4.1 Empirical specification

As delineated in the previous section, our meta sample combines rich information on several dimensions of project characteristics with contextual information. Given this structure of our data, we fit the following regression to explain variation in project success:

$$\begin{aligned} Rating_{irtc} = \alpha\, Fin_{ir} + \beta\, Struct_{ir} + \gamma\, Complex_{ir} + \eta\, Risk_{ir} + \lambda\, Eval_{r} \\ + \theta\, Macro_{cp} + \delta Z'_{ir} + \epsilon_{ct}, \end{aligned} \tag{2.1}$$

where $Rating_{irtc}$ denotes the respective DAC-rating dimension of project i, located in country c and evaluated as part of evaluation-report r, written in year t. Fin_{ir}, $Struct_{ir}$, $Complex_{ir}$, $Risk_{ir}$ and $Eval_{r}$ are vectors of relevant project-specific variables capturing the clusters financing, structure, complexity, risks, and evaluation, respectively, while vector $Macro_{cp}$ captures country-specific characteristics at the time of project implementation p. Specific variables within each dimension are discussed further in the results section.

Lastly, $\boldsymbol{Z}_{ir}$ controls for a comprehensive set of additional project-specific variables comprising fixed effects for sector, region, period of implementation as well as evaluation (5-year intervals). Robust standard errors are clustered at the country-evaluation-year level. We estimate equation (2.1) using weighted least squares (WLS), where the weights are given by the inverse number of projects evaluated in the corresponding evaluation report. This approach appropriately reflects the research question and data structure (Denizer et al., 2013; Card et al., 2018).

The main analysis focuses on the pooled sample, using the full set of projects' individual DAC-criteria ratings as outcome variable. It is organized along the key thematic dimensions of interest: That is, we first investigate evaluation features and the independence of assigned ratings (in section 2.4.2), and then in the results section (section 2.5) we stepwise introduce and present results for the four project characteristics clusters, as well as for the country context.

Adding to the full sample results, we stratify the sample by region and sector, respectively, to investigate and highlight heterogeneities in project success along these dimensions. From a methodological perspective, several additional analyses and robustness checks are added subsequently: First, we fit equation (2.1) for each DAC criterion separately to see whether micro and macro variables correlate across these dimensions differently. Second, sensitivity of the outcome variable is verified using (ordered) probit models. Finally, as a robustness check for the selection of the variables and to reduce the potential of overfitting, we also estimate the model using an adaptive Least Absolute Shrinkage and Selection Operator (LASSO) technique (Zou, 2006). Such an approach reduces the model to the key variables in a first step before the normal WLS model is estimated on the reduced set of variables.

It has to be noted that the obtained coefficients are prone to endogeneity similar to other related research (e.g., Denizer et al., 2013; Ashton et al., 2023). Development cooperation responds to macro-economic deterioration and political incentives, which will simultaneously affect the observed outcomes. Despite a rather detailed set of project characteristics, we cannot measure all project design features, which in the given context may also respond to unobservable conditions on the ground. Furthermore, finding valid instrumental variables in such settings has proven to be challenging, impeding the identification of causal effects (Bulman et al., 2017; Feeny et al., 2017). When discussing our findings, we therefore point to immediate as well as alternative interpretations that potentially underlie observed estimates. Given that the empirical analysis relies on an extensive set of fixed effects and control variables, we are confident that our results account for unobserved factors to the extent possible, thereby providing interpretable, relevant, and informative results on the success and failure of development finance projects, in particular in combination with past studies on the topic (Denizer et al., 2013; Feeny et al., 2017; Ashton et al., 2023).

2.4.2 Independence of ratings

Our analysis benefits from the fact that all success ratings are based on a coherent evaluation methodology. In fact, it is precisely due to the systematic rating framework provided by the DAC-criteria—and applied to more than 1,000 evaluations over 1.5 decades—that it is possible to construct these data. This coherent, systematic foundation of the data generation process notwithstanding, there is a possibility that other evaluation-specific characteristics may be significantly related to the assigned outcomes due to potential biases arising in the evaluation process. We test for these concerns in turn, and report the corresponding results in Table 2.1.

First, the sample time lag between the project completion report (i.e., the formal end of the project) and the evaluation report is 3.27 years. This duration might be structurally related to the success rating, since information for projects assessed later might not be as readily available. Also, certain evaluations may only be conducted with delay due to ongoing conflicts in a given country. Such instances might simultaneously affect the outcome. The first row of Table 2.1 reports some evidence for such a relationship: In column 1, the coefficient is negative but not significant; when including the entire set of controls, however, the estimate turns significant, indicating that projects assessed at a later stage receive lower ratings, on average (column 5). We therefore control for the time lag between project completion and evaluation in all models.

A second potential bias concerns the type of evaluator. Whereas all evaluations, ultimately, are conducted under the governance and quality assurance mechanisms of FCE, in practice there are four evaluator categories (cf. section 2.2): FCE staff, seconded colleagues from KfW operational units (“internal"), external evaluation consultants, and internal plus external combined. Ex-ante, it is a theoretical possibility that certain types of evaluators systematically assign, on average, too positive or too negative grades (e.g., one could plausibly speculate that internal evaluators might be tempted to rate too successfully given their expectation that at some time in the future their own project will also be evaluated). It is one strength of our data that they contain evaluator type information, allowing us to empirically investigate this potential bias. Rows 2–4 in Table 2.1 report the results. Both the reduced (columns 2 and 4) and full specifications (column 5) indicate that the magnitude of the point estimates is small and there is no statistically significant correlation between evaluator type and assigned rating. This is a reassuring finding: The unbiasedness of success ratings is not only plausible given the structural independence of the evaluation unit and its evaluators, but is in fact an empirical reality.

TABLE 2.1: Results: Evaluation-specific characteristics

Dep. variable: Rating (pooled)	(1)	(2)	(3)	(4)	(5)
Time between final review and EPE	-0.013			-0.016	-0.034**
	(0.012)			(0.012)	(0.015)
Evaluation type (base: FC E):					
-External		-0.046		-0.088	-0.088
		(0.222)		(0.221)	(0.200)
-Internal + external		-0.009		-0.008	-0.019
		(0.073)		(0.074)	(0.070)
-Internal		0.085		0.078	0.078
		(0.075)		(0.074)	(0.070)
Evaluation month (base: December):					
-January			0.184	0.201	0.206*
			(0.138)	(0.140)	(0.112)
-February			0.044	0.049	0.113
			(0.123)	(0.122)	(0.108)
-March			0.080	0.095	0.080
			(0.098)	(0.099)	(0.103)
-April			0.202**	0.211**	0.271***
			(0.094)	(0.098)	(0.099)
-May			0.014	0.023	-0.008
			(0.110)	(0.109)	(0.103)
-June			-0.019	-0.003	0.046
			(0.108)	(0.109)	(0.114)
-July			-0.012	0.008	0.039
			(0.112)	(0.113)	(0.105)
-August			0.070	0.087	0.003
			(0.090)	(0.089)	(0.092)
-September			-0.099	-0.080	-0.085
			(0.118)	(0.116)	(0.100)
-October			-0.024	-0.023	-0.001
			(0.091)	(0.090)	(0.092)
-November			0.016	0.021	0.025
			(0.082)	(0.081)	(0.078)
Full specification					Yes
Sector and region indicators	Yes	Yes	Yes	Yes	Yes
Sub-rating indicators	Yes	Yes	Yes	Yes	Yes
Other control variables	Yes	Yes	Yes	Yes	Yes
Observations	5,608	5,608	5,608	5,608	5,458
Adjusted R^2	0.14	0.14	0.14	0.14	0.23

Note: Table entries are coefficients from WLS regressions with the pooled rating as dependent variable. Weights are given by the inverse of the number of projects evaluated in the corresponding evaluation report. Other controls include the year of project start as well as evaluation year (both 5-year intervals). In the full specification, all other variables from Tables 2.3-2.6 are included. Standard errors in parentheses clustered at the country-evaluation-year level. */**/*** denote significance levels at 10/5/1% respectively.

A third internal process that could potentially influence ratings is the timing of evaluations during the calendar year. Due to the annual sampling process, the evaluation unit has to achieve a certain number of evaluations each fiscal year, and the annual count to achieve that number stops on December 31st. This leads to a clustering of evaluation reports at the end of the fiscal year: 30% of reports in our sample were finalized in December, and 15% in November. The remaining 55% are relatively equally distributed across the other ten months. Whereas the pure number of reports per month is no cause for concern, one might conjecture that last-minute reports might potentially be associated with either more positive (in order to finish the report on time) or more negative (the reason why the evaluation took so long) success ratings. Rows 5–15 in Table 2.1 report the corresponding estimation results and indicate that there is no such pattern recognizable in the data, in particular not concerning any end-of-the-fiscal-year pattern. Only April and January are (marginally) significantly different from the other months, but these are the two months with the lowest number of evaluation reports and simultaneously have slightly more positive mean overall grades (4.45 and 4.47, respectively, the remaining ten months all in-between 4.08 and 4.44), such that we interpret this as a deviation at random. Nonetheless, we include evaluation month as a control variable in all subsequent specifications.

A final issue in which institutional evaluation processes might be correlated with success ratings is trends: Over the years general trends toward better projects—and/or even more ambiguously—better ratings could potentially bias our results. In fact, we observe a slight trend towards better ratings over the sample period, however, mean ratings using five-year evaluation completion brackets from 1990 onward are not significantly different from one another (not shown in the table for brevity). Nonetheless, all regressions control for year of evaluation by means of these five-year period indicators.

2.5 Empirical results

The existing literature has highlighted the importance of project-level factors to explain the success or failure of development projects (Denizer et al., 2013; Bulman et al., 2017; Feeny et al., 2017). Our results provide support to the importance of project-level factors to explain the determinants of success or failure of development aid. To further motivate our analysis, we calculate the between-country variation in project success and regress country fixed effects on a binary success outcome variable for each year the projects in our sample were active.[9] From the resulting averaged R^2, we derive the share

[9] See section A.3.3 in the Appendix for a detailed description of the methodology.

of variation that can be explained by country factors, i.e. the environment in which the projects are implemented. Our result is comparable to that for World Bank projects (Denizer et al., 2013) and indicates that 34% (20% for the pooled sample) of project variation stems from between-country variation. In the following empirical analysis, we thus examine an extensive set of project-level micro variables to contribute towards better understanding the determinants of success and failure.

2.5.1 Cluster (1): Project financing

Whereas previous research had to revert primarily to financial volume as the only proxy for complexity, we are able to address project complexity and project financing separately. Project financing is the first cluster of project-level variables we analyze, providing a nuanced perspective by including information on seven financial variables such as cooperation type or share of counterpart contributions. The regression results are presented in Table 2.2.

There is some indication (row 1) that financially larger projects are systematically correlated with more successful ratings.[10] The point estimate is positive and statistically significant in the reduced specification (column 1), but becomes insignificant in the full specification (column 9). Financially larger projects may comprise straightforward infrastructure investments or politically prominent showcases receiving more attention, thus making implementation easier.

From a donor perspective, beyond total investment volume, more leverage potentially lies with the budget funds that are committed—i.e. broadly loan vs. grants—as well as the share of counterpart financing contributed by the partner government. KfW staff might, for example, be able to exert higher pressure on contractual and regulatory procedures like due diligence when funds are committed as a loan, possibly resulting in better outcomes. Looking at the results in Table 2.2, when compared to grants—which represent 90% of development finance projects in the sample—loans do not perform significantly better (row 2). At the same time, the correlation between the share of counterpart contributions and project success is (marginally) statistically significant and positive (row 3, columns 3 and 8, though not in the full specification in column 9). Intuitively, greater commitment by local partner governments could be expected to be associated with better ratings, such

[10]Recall that the total volume refers to the costs of the entire project, i.e. including commitments by the government itself and/or other donors.

that this finding underlines the importance of ownership as a key principle of development cooperation. The overall tendency of a positive association between variables related to project financing and project success that can be taken from Table 2.2 is further highlighted by a significant positive correlation between budget funds and project ratings in the reduced specification (row 4, column 4).

TABLE 2.2: Results: Project financing

Dep. variable: Rating (pooled)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Total volume (log)	0.049**							0.029	0.037
	(0.021)							(0.028)	(0.029)
Aid type (Base: Loan):									
-Grant		0.052						0.093	0.105
		(0.088)						(0.098)	(0.087)
% counterpart contributions			0.241**					0.196*	0.145
			(0.112)					(0.117)	(0.118)
Budget funds (log)				0.070**				0.057	0.095**
				(0.029)				(0.037)	(0.042)
% budget funds of ODA					0.000			-0.000	-0.000
					(0.000)			(0.000)	(0.000)
% project funds of GDP						0.000		-0.000	-0.000
						(0.000)		(0.000)	(0.000)
Disbursement vs. commitment							0.169	0.178	0.137
							(0.165)	(0.163)	(0.156)
Full specification									Yes
Sector and region indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sub-rating indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,608	5,608	5,608	5,608	5,608	5,608	5,608	5,608	5,458
Adjusted R^2	0.15	0.14	0.15	0.15	0.14	0.14	0.14	0.15	0.23

Note: Table entries are coefficients from WLS regressions with the pooled rating as dependent variable. % budget funds of ODA and % projects fund of GDP are re-scaled by 1 million. Weights are given by the inverse of the number of projects evaluated in the corresponding evaluation report. Other controls include the year of project start as well as evaluation year (both 5-year intervals) and evaluation month. In the full specification, all other variables from Tables 2.1-2.6 are included. Standard errors in parentheses clustered at the country-evaluation-year level. */**/*** denote significance levels at 10/5/1% respectively.

2.5.2 Cluster (2): Project structure

Details of the project structure are decided at project appraisal and at the discretion of KfW. Structural design features are, in theory, highly relevant from a policy perspective and could help improve development project effectiveness. There is some evidence that a tailored project design is a determinant for development outcomes on both the individual project (e.g., Khwaja, 2009) and aggregate level (e.g., Wane, 2004). This entails, e.g., deciding whether to implement a project along international partners in a co-financing arrangement, which is the case for 21% of projects in our sample. To increase cooperation is a common pledge among donors, largely due to the supposed positive effects attributed to it: More streamlined efforts toward developmental impacts and increased efficiency with regards to disbursement conditions have been affirmed in both the Paris Declaration and the Accra Agenda (OECD, 2022c). Our results provide only limited support for this hypothesis, as the coefficient on co-financing arrangements in Table 2.3 is positive and at the margin of significance (row 1, column 1, t-value 1.65).

Development finance projects are often implemented along with technical assistance to support local partner agencies (27% of projects in the sample). A plausible prior belief is that these measures are associated with improved project outcomes. However, the direction is not straight-forward, as it could be the particularly weak partners who receive such support in the first place. Such negative selection bias has been argued, e.g., to influence the relationship between more diligent project preparation time and unfavorable ratings (Denizer et al., 2013). The estimation results for accompanying measures in Table 2.3 are not statistically different from zero (row 2, columns 2, 8, and 9), a result that does not allow to disentangle the role these measures play or not. In fact, the insignificant point estimate could indicate that, on average, successful accompanying measures mitigate the negative selection effect.

Several more structural design features are worth considering: For instance, donors work with a multitude of local implementing partners, yet existing research cannot provide detailed insights regarding these agencies' capacities. Increasingly, projects are implemented with non-state actors, responding to the recognition that governmental partners' capacity is limited (Feeny et al., 2012), and potentially allowing for more participatory development partnerships with the civil society. In fact, such projects have been shown to perform better in some instances (Shin et al., 2017). While certain sectors such as micro-finance are already dominated by private agencies, in our sample most implementing partners—around 68%—are governmental institutions. Distinguishing different agency types, rows 3–6 in Table 2.3 find no significant relationship between any of these types and corresponding project success. This is an informative empirical finding for future project design: Agency type is not a key factor for project success, and neither is whether previous cooperation existed (row 7) nor the number of institutions involved (row 8).

In row 9 of Table 2.3 we estimate the role of project manager turnover. Project managers are in charge of the team at KfW and are the focal point for all interactions with the partner country and project implementing unit. Their importance for the success of a project is key, and therefore greater turnover could lead to knowledge loss and thus lower ratings (Ashton et al., 2023).[11]

[11] Since projects have different durations, we normalize the number of project managers by the number of operational years. A value of one therefore indicates that a project was managed by a new manager during each year when it was operational.

Our results indicate an—at first sight—counter-intuitive, positive relationship between the number of project managers per year and evaluation outcomes (reduced specification, column 6). Once we control for all other factors in the full specification (column 9), however, this association is no longer statistically significant. Finally, as a last hypothesis concerning this cluster of project variables, we explore whether a local KfW office in the project implementing country supports the success of a project. The assumption behind this is that such office presence might translate into higher engagement and knowledge in the partner country, resulting in more successful projects (Honig, 2020). We do not find any support for this hypothesis.

TABLE 2.3: Results: Project structure

Dep. variable: Rating (pooled)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Co-financing	0.091							0.075	0.002
	(0.055)							(0.056)	(0.064)
Accompanying measure		-0.066						-0.048	-0.015
		(0.058)						(0.058)	(0.056)
Agency type (Base: NGO):									
–Mixed			-0.034					-0.032	-0.099
			(0.126)					(0.128)	(0.130)
–Multilateral			0.113					0.080	-0.009
			(0.127)					(0.127)	(0.131)
–Private sector			0.031					-0.008	0.006
			(0.138)					(0.139)	(0.139)
–Government			-0.052					-0.056	-0.101
			(0.105)					(0.107)	(0.107)
Previous cooperation				0.079				0.074	0.066
				(0.053)				(0.053)	(0.051)
Number of institutions					0.008			0.005	0.005
					(0.009)			(0.010)	(0.009)
Project manager turnover						0.402*		0.348*	0.328
						(0.216)		(0.209)	(0.248)
Country office							-0.018	-0.008	-0.043
							(0.053)	(0.053)	(0.056)
Full specification									Yes
Sector and region indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sub-rating indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,608	5,608	5,608	5,608	5,608	5,608	5,608	5,608	5,458
Adjusted R^2	0.15	0.15	0.15	0.15	0.14	0.15	0.14	0.15	0.23

Note: Table entries are coefficients from WLS regressions with the pooled rating as dependent variable. Observations are weighted by the inverse of the number of projects evaluated in the corresponding evaluation report. Other controls include the year of project start as well as evaluation year (both 5-year intervals) and evaluation month. In the full specification, all other variables from Tables 2.1-2.6 are included. Standard errors in parentheses clustered at the country-evaluation-year level. */**/*** denote significance levels at 10/5/1% respectively.

2.5.3 Cluster (3): Project complexity

The design and, in particular, the implementation of development finance projects is often complex and challenging. Our meta sample allows us to investigate in more detail five features of this complexity. The overall finding from the corresponding results presented in Table 2.4 is that more complex projects have a lower likelihood of success.

In particular, the first dimension of project complexity captures the duration of the project. As row 1 of the table shows for all specifications (columns

1, 6, and 7, respectively), a longer project duration is strongly and significantly correlated with worse success ratings. The eventual duration of a project has both an implementational component, e.g., delays in contracting or executing, and a structural component, as it also depends on the sector or region where it is placed, which in turn also influence outcomes as described in 2.3.2. Row 2 of the table specifically investigates the role of delays, and shows that these are not a significant explanation of lower project ratings.

Concerning another, related factor of complexity, the length of time between the official commitment of governmental funds and their translation into actual projects as part of a contract, is theoretically ambiguous. While a longer duration could be an early flag for eventually hard-to-manage projects, they could also fare better due to thorough preparation (Deininger et al., 1998; Bulman et al., 2017; Kilby, 2015). Row 4 of Table 2.4 depicts some evidence for the former hypothesis, as the coefficient for the full specification (column 7) indicates a negative, marginally significant correlation between the length of time from mandate to contract and the success rating.

TABLE 2.4: Results: Project complexity

Dep. variable: Rating (pooled)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Project duration (log)	-0.234***					-0.215***	-0.149**
	(0.069)					(0.073)	(0.075)
Delay		-0.032				0.009	0.009
		(0.069)				(0.070)	(0.069)
Revised ToC			-0.071			-0.060	-0.048
			(0.049)			(0.049)	(0.047)
Years mandate to contract				-0.030		-0.026	-0.048*
				(0.024)		(0.023)	(0.027)
Technical complexity					-0.117**	-0.083	-0.130**
					(0.055)	(0.056)	(0.055)
Full specification							Yes
Sector and region indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sub-rating indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,608	5,608	5,608	5,598	5,608	5,598	5,458
Adjusted R^2	0.15	0.14	0.15	0.14	0.15	0.16	0.23

Note: Table entries are coefficients from WLS regressions with the pooled rating as dependent variable. Weights are given by the inverse of the number of projects evaluated in the corresponding evaluation report. Other controls include the year of project start as well as evaluation year (both 5-year intervals) and evaluation month. In the full specification, all other variables from Tables 2.1-2.6 are included. Standard errors in parentheses clustered at the country-evaluation-year level. */**/*** denote significance levels at 10/5/1% respectively.

Additionally, we consider whether the Theory of Change (ToC) outlined at the time of project appraisal was adjusted as part of the evaluation. A change could indicate that the project framework was not adequate in the first place or had to be updated to reflect operational adjustments, hinting towards increased complexity and thus potentially lower ratings (Blanc et al., 2016). However, we find this measure to be irrelevant for the rating obtained. As the last factor in this cluster, we analyze whether technically complex projects are correlated with better or worse evaluation ratings. This “Technical complexity" is a binary indicator variable taking on the value of one if the project required the support of a specific technical advisor, e.g., engineers for infrastructure projects. Row 5 of Table 2.4 shows that indeed technically complex projects—even when controlling for sector fixed effects—are significantly correlated with less successful project ratings.

2.5.4 Cluster (4): Project risks

A particularly interesting cluster of micro variables contained in the data is KfW’s internal risk assessment information. Specifically, the data contain information on (i) the number of risks that were identified ex-ante (i.e. before project start); (ii) the percentage of these that actually materialized during project implementation; (iii) the severity of the *overall* risk to project success ex-ante (low/medium/high); and (iv) the expected level of controllability of that overall risk (low/medium/high).

Row 1 of Table 2.5 presents estimation results for the number of risks identified ex ante. In theory, a larger number implies a more complex project, yet could also mean that the design is more deliberately thought through to cope with uncertainties during implementation. The results indicate no correlation between the pure number of identified risks and average project success. They key factor that matters for project success, however, is whether and at what rate these pre-identified risks actually materialize: Row 2 consistently shows a strong and statistically significant negative correlation between the share of risks that occurred and the success rating (columns 3, 5, and 6). In fact, the point estimate for the full specification (column 6) implies that projects for which all risks materialize are rated 0.5 points lower. This is a considerable effect size.

Furthermore, rows 3–5 of the table show that high-risk and medium-risk projects are statistically significantly associated with a lower success rating, relative to low-risk projects. Again, the effect is sizable (full specification, column 6) at -0.35 grades on average for high-risk projects, and -0.2 grades for medium-risk projects. Whether any of these risks was deemed controllable or not does not affect success ratings (rows 6–8 of the table).

TABLE 2.5: Results: Project risks

Dep. variable: Rating (pooled)	(1)	(2)	(3)	(4)	(5)	(6)
Number ex-ante identified risks	-0.004				0.002	0.001
	(0.013)				(0.013)	(0.013)
% ex-ante identified risks occurred		-0.504***			-0.464***	-0.486***
		(0.068)			(0.067)	(0.067)
Overall risk (base: low)						
-Medium			-0.251***		-0.185**	-0.203**
			(0.080)		(0.078)	(0.082)
-(Very) high			-0.460***		-0.326***	-0.352***
			(0.086)		(0.084)	(0.088)
-Not assigned			-0.285***		-0.159	-0.219*
			(0.109)		(0.116)	(0.116)
Overall risk control (base: low)						
-Medium				0.075	0.070	0.084
				(0.058)	(0.055)	(0.058)
-High				0.001	-0.065	-0.061
				(0.253)	(0.193)	(0.169)
-Not assigned				0.090	-	-
				(0.096)	(.)	(.)
Full specification						Yes
Sector and region indicators	Yes	Yes	Yes	Yes	Yes	Yes
Sub-rating indicators	Yes	Yes	Yes	Yes	Yes	Yes
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,608	5,608	5,608	5,608	5,608	5,458
Adjusted R^2	0.14	0.18	0.16	0.15	0.19	0.23

Note: Table entries are coefficients from WLS regressions with the pooled rating as dependent variable. Weights are given by the inverse of the number of projects evaluated in the corresponding evaluation report. Other controls include the year of project start as well as evaluation year (both 5-year intervals) and evaluation month. In the full specification, all other variables from Tables 2.1-2.6 are included. The risk control category "not assigned" is omitted due to collinearity in column 5–6. Standard errors in parentheses clustered at the country-evaluation-year level. */**/*** denote significance levels at 10/5/1% respectively.

2.5.5 Contextual variables

Historically, macroeconomic outcomes such as GDP have shaped the discussion around the success of development aid (Isham et al., 1999; Qian, 2015). However, development projects ultimately are not only supposed to fuel development, but they are simultaneously affected by the economic environment in which they operate. This holds particularly for GDP growth, the most immediate variable measuring the general economic environment and shocks. Related literature has shown that an environment conducive to growth is a significant predictor for project success—and this is an empirical relationship we also observe in most of our specifications as shown in row 1, columns 1 and 5 of Table 2.6.

The role of civil liberties and citizen freedom is theoretically more ambiguous: Policies in democracies could be more aligned with citizens' needs than in autocracies, yet the latter might provide a more stable institutional environment. Indeed, existing literature has found conflicting relationships for World Bank- and ADB-financed projects (Isham et al., 1997; Feeny et al., 2017). We correlate Freedom House Democracy scores with success ratings, however cannot confirm previous results in either direction (row 2). In light of donor-targeting decisions partially based on governance criteria (Feeny et al., 2017), this is highly relevant. This particularly holds for German bilateral cooperation, which has recently put more emphasis on good governance criteria in commitment decisions as part of its reform partnerships (BMZ, 2022).

TABLE 2.6: Results: Country context

Dep. variable: Rating (pooled)	(1)	(2)	(3)	(4)	(5)	(6)
GDP p.c. growth (annual)	0.017**				0.016*	0.011
	(0.008)				(0.009)	(0.008)
Freedom House Democr. score		0.000			-0.006	-0.018
		(0.018)			(0.021)	(0.021)
State Fragility Index			-0.008		-0.008	-0.006
			(0.006)		(0.007)	(0.008)
Population (log)				-0.002	0.002	-0.029
				(0.017)	(0.017)	(0.022)
Full specification						Yes
Sector and region indicators	Yes	Yes	Yes	Yes	Yes	Yes
Sub-rating indicators	Yes	Yes	Yes	Yes	Yes	Yes
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,608	5,468	5,468	5,608	5,468	5,458
Adjusted R^2	0.15	0.14	0.15	0.14	0.15	0.23

Note: Table entries are coefficients from WLS regressions with the pooled rating as dependent variable. Weights are given by the inverse of the number of projects evaluated in the corresponding evaluation report. Other controls include the year of project start as well as evaluation year (both 5-year intervals) and evaluation month. In the full specification, all other variables from Tables 2.1-2.4 are included. Standard errors in parentheses clustered at the country-evaluation-year level. */**/*** denote significance levels at 10/5/1% respectively.

The institutional environment plays a crucial role for the success of aid interventions, particularly because most projects are implemented jointly with governmental partners. A reasonable expectation is that in conflict-prone countries, i.e. where state fragility is more pronounced and institutional quality lower, it is more difficult for projects to deliver on their objectives (Caselli et al., 2021). For example, World Bank projects have been shown to be more fruitful in post-conflict settings with sustained peace (Chauvet et al., 2010). Using the State Fragility Index—incorporating measures of governance effectiveness and legitimacy—we however find no statistically significant relationship (row 3).

The size of a country in terms of population is potentially negatively related to the likelihood of success, as the complexity of governing more people increases (Feeny et al., 2017). We find no evidence for this in our meta sample either (row 4). Lastly, contextual factors beyond the country level that are not specific to projects and vary over the period of implementation likely also matter. An example would be institutional arrangements among donors that increase delivery on projected outcomes. While we cannot account for those directly, we include indicators for five-year brackets of the year of project appraisal, capturing changes in institutional arrangements over time.

2.6 Heterogeneities and robustness

Disaggregating the results potentially yields further insights and can unmask different trends within the sample. In addition to the findings for the pooled sample, we stratify the analysis by region and sector, and fit separate regressions for the individual DAC criteria.

2.6.1 Empirical results by region

Development institutions regularly identify striking differences in projects' success depending on the region where the project was implemented in, with Sub-Saharan Africa often providing the most challenging environment. Appendix Table A.5 disaggregates empirical results by region and shows that indeed the correlation between the various determining factors and the project success rating is heterogeneous. In particular, there are only few variables that play the same significant role throughout all regions.

In the SSA-sample shown in column 2, two variables stand out: Projects financed via grants fare considerably better than loans, potentially explained by the fact that these instruments are used particularly for fundamental public services such as water supply, where ownership could thus be greater.[12] At the same time, projects led by governmental agencies are significantly rated worse as compared to NGO-led ones. As the flip side of the ownership argument, it could hint towards public institutions' limited capacity when it comes to providing basic infrastructure. Turning to Asia in column 3, the country-context appears to matter more than elsewhere. While a positive relationship with GDP p.c. is intuitive, it runs counter comparable findings for state fragility (e.g., Chauvet et al., 2010). In contrast, larger population size is significantly associated with lower outcomes.

Micro-variables are more often significantly related with outcomes in Europe (column 4). Particularly project structure- as well as complexity-related variables correlate negatively with outcomes, implying that KfW would have a

[12]This instrument however only makes up 2% in the SSA-sample.

higher leverage to address underlying obstacles ex ante and during implementation. While for example the number of institutions involved appears to make projects over-complex, at the same time the projects' outcomes are less affected by ex-ante identified risks that eventually materialized. Notably, greater democracy is negatively related to project success in our data, which adds to the already ambiguous results found in the related literature (e.g., Isham et al., 1997; Kosack, 2003).

2.6.2 Empirical results by sector

In a next step, we conduct the sub-sample analysis on the projects' main sectoral focus as shown in Appendix Tables A.6 and A.7. While many of the patterns identified for the pooled sample and the regional stratification are visible in the sectoral results, too, several additional results warrant attention.[13]

First, in the energy sector (Table A.6, column 4), projects with more institutions involved are associated with weaker outcomes, possibly due to undue complexity in a sphere dominated by large-scale utility companies on the partner-side. Longer preparation times before contract closings however relate positively with ratings, potentially hinting towards the role of due diligence in these mostly large-scale infrastructure investments. Two variables are noteworthy for governance interventions (Table A.7, column 2): Against the pattern in most sectors, more institutions involved in the implementation appear to yield better outcomes. Due to the complexity of these projects, a holistic approach might therefore be beneficial for this sector. Similarly surprising, these projects fare better in more fragile contexts, where governance might have already been weak in the first place. On the contrary, fragility is negatively related with agriculture-related projects (Table A.6, column 2). In this sector, larger projects with greater counterpart contribution shares—thus potentially inducing more ownership on part of the partners—are also rated better on average.

For transport-themed projects (Table A.6, column 3), the number of ex-ante identified risks stands out. It's negative relationship with evaluation ratings raises the question how well institutions can mitigate these risks that were already apparent before project inception. The identification is a key component of any due diligence, yet the ex-post perspective suggests that investing

[13] For this thematic split, some variables had to be excluded due to the limited sample sizes for individual sectors.

in projects with large uncertainties should potentially be questioned more thoroughly in the first place. Lastly, water-sector projects (Table A.7, column 4) are the only ones significantly related to an ex-post revised ToC.

2.6.3 Individual DAC-criteria

In the next analytical step, we estimate our main specification for five DAC-criteria and the overall rating separately and present results in Appendix Table A.8. Each criterion addresses a unique dimension of project success and thus provides an additional, detailed perspective on relevant success determinants. While *relevance*'s focus is on the project layout at the time of inception, when adjustments are still viable, *efficiency*, *effectiveness* and *impact* evaluate actual outcomes during implementation. Lastly, *sustainability* concerns outcomes observed at the time of the evaluation, taking into account potential future scenarios of project outcomes. Across the criteria, a first glance reveals that existing heterogeneities in previous sub-samples do not necessarily translate to this level. Nevertheless, only one variable is consistently significant—the share of eventuated risks—and correlates vary considerably.

Given that *relevance* concerns project design and its ability to address developmental challenges, project *structure* variables are of particular interest (column 2). However we find that none of our micro variables are significantly related to the rating. This includes *financing* variables which are still partially—as in the case of the budget funds committed—at the scrutiny of KfW. In terms of *complexity*, the share of risks that materialized is negatively related with *relevance*. Due to the structural break in time—ex-ante relevance and operative risks—this relationship is not entirely concise and suggests a level of risk tolerance: KfW correctly anticipates operational risks at the time of appraisal, but from an evaluative point of view, they may have already been rooted in the project design itself. Furthermore, the negative relationship with an ex-post adjusted indicator framework and therefore potentially inappropriate ToC corroborates this deliberation. With regards to *efficiency* (column 3), projects with greater budget volumes appear to fare better. This potentially stems from large-scale infrastructure investments that undergo more extensive cost-benefit analyses on part of KfW than projects with regionally spread, small-scale investments. The *efficiency*-criterion is closely related to the rate of economic return, which has been found to be positively related with the country environment, particularly GDP (Isham et al., 1999).

When assessing the *effectiveness* of interventions (column 4), beyond risks that eventually occurred, the relationship with delays before actual implementation—i.e. the time between intergovernmental agreements and the actual project financing contract—is negative. Potentially, this could already constitute a red flag for later implementation challenges due to, e.g., partner capacity constraints. However, either relationship does not materialize further down the project logic as displayed in column 5 for the *impact* criterion. Projects with more budget funds are associated with greater developmental results. Lastly, the only positive significant correlates of project *sustainability* (column 6) are larger investments and previous cooperation with project implementing agencies. In contrast, technical complexity hampers the durability of achieved impacts. Because sustainability is evaluated several years after project completion, these findings could motivate more peculiar attention on part of project managers towards sustainability concerns during the operational phase. At the country level, projects in more fragile environments receive more pessimistic ratings, a result that is in line with the existing literature (Chauvet et al., 2010).

2.6.4 Robustness

Our selection of the explanatory variables relies on existing theories and past studies. In an robustness check, we automize the variable selection and apply an adaptive LASSO technique. This strips the model to its most predictive variables in a first step, before re-estimating equation (2.1) with the reduced set of variables. Results are presented in Table A.9. The approach drops several control and main analytical variables from the clusters of interest in our main specification: (i) The share of counterpart contributions, (ii) indicators for co-financing and accompanying measures, (iii) sub-categories of the agency type and risk control variables, (iv) the delay variable and (v) the number of ex-ante identified risks. Re-estimating the model with the reduced number of variables in a next step mainly confirms the previous results, as the coefficient size for most variables remains similar. However, two additional variables turn statistically significant: The total project volume and the number of project managers in a given project, counter-intuitively implying that a greater project manager turnover increases the rating. The reduction of variables goes along with a reduced adjusted R^2 of 0.19, compared to 0.23 in the WLS regression.

In our main specification, the outcome variable represents individual DAC-ratings that are assigned on an ordinal scale. While estimating the models using WLS allows for straight-forward interpretation of the coefficients, we also estimate equation (2.1) in other specifications to assess the robustness of the coefficients: Appendix Table A.10 displays results from OLS and ordered probit models for the pooled ratings, the overall project rating and a binary success measure (cf. section 2.3) as outcome variables. By and large, we find that the coefficients (significance levels) are comparable across these specifications.

2.7 Conclusion

This paper presents a systematic, quantitative analysis of the success—and failures—of three decades of German bilateral financial cooperation. We construct a unique meta database covering 5,600 evaluation ratings, the most comprehensive and up-to-date database on bilateral financial cooperation results worldwide from a single donor. Together with extensive and novel data on project characteristics, our analysis yields new insights on the question of what works in development finance. Those results are transferable to some degree given the scope of the the dataset and the comparability with previous research that has shown similar results for bi- and multilateral donors (Bulman et al., 2017; Briggs, 2020).

Four *general* findings emerge from our analysis: First, we find that project characteristics can explain variation in project success better than contextual country-factors. This does not only motivate further scrutiny in project implementation for practitioners, but also underlines the importance of more detailed variables on project characteristics, also for bilateral cooperation. Second, these characteristics show that variables related to (i) factors ex ante and, hence under the influence of project managers, do not matter significantly for eventual outcomes and, (ii) complexity consistently exerts negative influence on the success of interventions. Third, we find that different dimensions of project success as measured by DAC-criteria relate very heterogeneously to our variables of interest, indicating that considering the mere overall success rating masks important relationships otherwise not visible. And finally, our results show that when it comes to project design and implementation, *one size does not fit all*. Disaggregating the sample by region and theme suggests that all our variables influence success to varying signs and significance levels.

From the host of variables included in our four clusters of project characteristics plus contextual factors, the following *specific* results stand out: First, larger shares of partner counterpart financing are partially associated with greater likelihood of project success, favoring the view that ownership matters for development to work. Overall, there is also the indication that the total financial volume is positively correlated with project success. Second, we find that the type of implementing agency—private, public or non-governmental—does not matter for the eventual result of the project. Against the backdrop of increasing cooperation with non-state actors in fragile contexts or private agencies in middle-income countries, this is notable. Third, the share of ex-ante identified risks that eventually materialized is a significant predictor of success, suggesting that project designs often correctly identify the relevant risks but may not be able to mitigate (all of) them during implementation. Finally, neither democracy or fragility are significantly related to project success, further substantiating the ongoing debate on the inclusion of governance criteria in aid allocation.

Our research significantly extends the literature on aid effectiveness with extensive data on a bilateral donor and project characteristics. Nevertheless, there are several gaps in the analysis of the determinants of project success that need to be addressed by future research. Given the surprising finding on the irrelevance of structural project characteristics, more detailed information on project design could help to better understand how to favourably influence outcomes. Moreover, the depth of project characteristics presented here provides novel insights that would benefit from verification by research drawing on other donor data, such as that from the World Bank. Finally, our approach shares the fate that causal analysis based on, e.g., appropriate instruments on such a broad set of variables is difficult to implement. While we can control for a wide range of observable characteristics, causally identified impacts on topical issues in development finance, such as co-financing, ownership and governance, will be key to making projects more effective in the future.

Chapter 3

Losing territory: The effect of administrative splits on land use in the tropics

Elías Cisneros, Krisztina Kis-Katos and Lennart Reiners

Abstract

State decentralization is often promoted as a way to improve public service delivery. However, its effects on forest are ambiguous. Decentralization might not only improve local forest governance, but also change the incentives to promote agricultural expansion into forests. This study focuses on the power devolution stemming from the proliferation of new administrative units in Indonesia during the last two decades. The discontinuous changes in government responsibilities at new administrative borders provide exogenous spatial variation to study forest outcomes. Using a spatial boundary discontinuity design with 14,000 Indonesian villages, we analyze the effects of 115 district splits between 2002 and 2014. Results show a 35% deforestation decline within new (child) districts relative to the existing (mother) districts both immediately before and after the splitting. In pre-split years, these changes can be explained by agricultural divestment on part of the mother districts on territories that are soon to be lost. In post-split years, the short-term forest conservation benefits are neither rooted in an increased social cohesion nor stronger development. Instead, newly formed districts seem to be temporarily suffering from administrative incapacity to attract large-scale agricultural investments. In the long run, no lasting local forest conservation benefits persist as deforestation equalizes between child and mother districts few years later.

3.1 Introduction

Tropical forests are under strong pressure from the demand for land conversion for alternative use. Their existence is essential for both climate and biodiversity protection, making conservation efforts a key policy goal worldwide. To be successful, interventions crucially rely on local governance and institutions (Burgess et al., 2012; Wehkamp et al., 2018). In recent decades, sub-national administrations have gained substantial influence on conservation outcomes due to broad decentralization reforms that sought to improve public service delivery (Besley et al., 2003; Faguet, 2004). While the empirical evidence on the effects of decentralization is extensive, its results are at times mixed (Gadenne et al., 2014). Conceptually, decentralization policies often combine both a transfer of administrative responsibilities and an increase in the number of sub-national jurisdictions, also referred to as government fragmentation (Grossman et al., 2014; Pierskalla, 2016). In a decentralized state, these (new) administrative entities become influential actors, yet understanding how their proliferation affects developmental outcomes remains understudied (Pierskalla, 2016; Grossman et al., 2017).

We focus on Indonesia, which provides the ideal environment to study the relationship between sub-national government fragmentation and deforestation: After the fall of the Suharto-regime in 1998, the country embarked on far-reaching decentralization reforms labelled as a “big bang" (Fitrani et al., 2005). The new legislation paved the way for jurisdictional adjustments, allowing for the formation of new districts, which received considerable power as part of the reforms (Ostwald et al., 2016). Consequently more than 150 new administrative units across the entire Indonesian archipelago came into existence within 14 years, which were carved out of existing ones and had to establish new capitals and corresponding institutions from scratch. At the same time, Indonesia—home to one of the world’s most pristine tropical rainforests—has experienced rampant deforestation and land-use change (Austin et al., 2019).

Our analysis exploits the fact that administrative boundaries between the newly split entities were idiosyncratic to local conditions in both topographic and socioeconomic terms. In the framework of a spatial regression discontinuity design, the new administrative boundaries represent sharp cutoffs between otherwise comparable villages.[1] Existing literature in the Indonesian context has focused on the impact of splits at the district and provincial level, highlighting the role of inter-administrative competition and ethnic homogeneity. In contrast, our analysis (i) is conducted at the highly localized village level and by that deals with a series of important heterogeneities (Grossman et al., 2014); and (ii) studies a new mechanism through which the creation of new jurisdictions affects deforestation: Anticipatory strategic land-use decisions with regard to oil palm expansion by local administrations.

From a theoretical perspective, ex-ante it is ambiguous how villages' land-use trajectories in a close neighborhood of the new boundary might develop once a split is expected and implemented. In our cross-sectional analysis of 115 district splits realized between 2002–2014, we show that deforestation in villages located in the newly formed child districts decreases compared to the ones located in the existing mother districts. This effect materializes from up to two years before to three years after the split came into effect. At around 35%, the reduction in deforestation is considerable and is supported by a host of robustness and placebo tests.[2]

Guided by existing research in the context of decentralization and land-use decisions, we identify mechanisms related to altered cost-benefit considerations that both existing and new governments face with regards to promoting or preventing land-use change. Taking into account both anticipatory short-run and strategic medium-run effects before and after the splits, we discuss and empirically verify five potential mechanisms: The role of (i) immediate land-use rents from deforestation; (ii) medium-term land-use rents by strategic investments into oil palm plantations; (iii) changing constituency preferences through decreasing ethnic fractionalization; (iv) temporarily diminished administrative capacity in new districts; and (v) the creation of new political centers and the subsequent expansion of human settlements in the neighborhood of new capitals. From these proposed mechanisms, we find empirical evidence for strategic divestment from land-use conversion by the existing district government. Because medium-term rents from investments

[1] Note that we use the terms districts and jurisdictions as well as new district boundary and split boundary interchangeably.

[2] Annual deforestation in our sample is around 1.5% of 2000 forest cover. Our results thus imply that annual deforestation rates are temporarily reduced to 1.0% in villages at the boundaries of a child district.

into oil palm expansion on contested land will go towards the new government, deforestation pressures are temporarily reduced already before the district split takes place. While this mechanism has not been documented before, it reconciles well with the fact that in Indonesia, deforestation responds strongly to political-economic incentives (Burgess et al., 2012), fostering especially land-use change towards oil palm cultivation (Angelsen, 2007; Austin et al., 2019; Cisneros et al., 2021). A few years after the split, both oil palm expansion and deforestation in the child districts accelerate once again, yielding no sustained protection of natural resources at the boundaries of newly formed districts in the longer run.

Our paper is related to several strands of literature. First, by focusing on the role of district splits for deforestation, it contributes to the literature on the determinants of deforestation in the tropics, and especially on the political economy of deforestation (Burgess et al., 2012; Austin et al., 2019; Cisneros et al., 2021). Second, by showing temporary localized effects of government fragmentation, our paper also relates to the ongoing debate on decentralized natural resource management (Blackman et al., 2021). Third, the paper adds to the growing literature on the unintended outcomes of decentralization (Pierskalla, 2016), by showing that decentralization reshapes land-use incentives: While most studies find negative side-effects (Grossman et al., 2017), our result implies a positive temporary impact in terms of forest protection. Lastly, we add to the growing literature that uses administrative borders as spatial discontinuities in economics more broadly (Michalopoulos et al., 2013; Pinkovskiy, 2017) and in environmental economics in particular (Bonilla-Mejía et al., 2019; Burgess et al., 2019; Cuaresma et al., 2019).

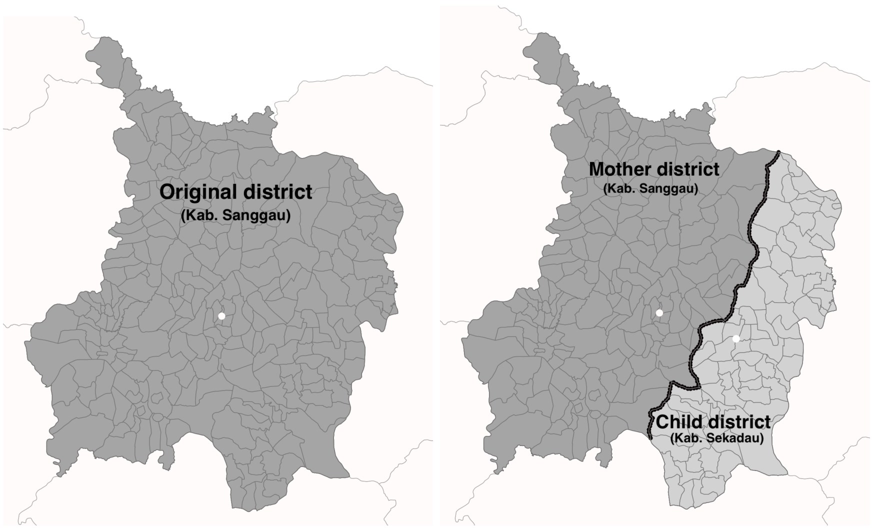
The remainder of the paper is organized as follows: Section 3.2 outlines our study's context, followed by a discussion of the theoretical framework in section 3.3. Section 3.4 presents an overview of data and the empirical methodology. Section 3.5 discusses the results and section 3.6 concludes.

3.2 Background

3.2.1 Indonesia's decentralization reforms

After the fall of the Suharto-regime in 1998, a period of rapid reforms triggered massive decentralization (Fitrani et al., 2005). It involved two related, however conceptually notably different components: On the one hand, classical decentralization resulted in vertical power devolution to lower tiers of government in administrative, fiscal, and political terms. While the administrative hierarchy remained unchanged, the second tier administrative districts (*Kabupaten*, or so-called regencies and *Kota*, or cities) received substantial new administrative and fiscal powers.[3] Increased fiscal transfers along some competencies to levy taxes were accompanied by the responsibility to deliver a large part of local public services (Ostwald et al., 2016). On the other hand, these reforms paved the way for the creation of new districts, additionally leading to horizontal power devolution by increasing the number of administrative units. Known as *pemekaran* (or the "blossoming" of districts), from 2001 onward, more than 150 new districts were created in a process of government fragmentation. This sequence of vertical, followed by horizontal power devolution is typical for developing countries' decentralization reforms worldwide (Grossman et al., 2014), yet it is considered particularly pronounced in Indonesia.[4]

FIGURE 3.1: Administrative reorganization in an exemplary district split



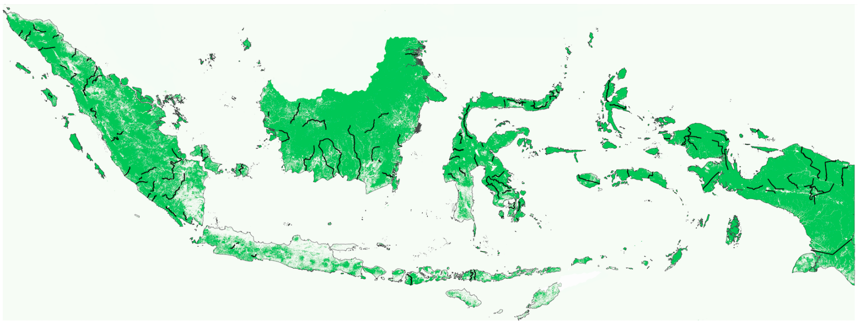
Note: The original mother district *Kabupaten Sanggau (left)* split into two units in 2003 *(right)*, establishing the new administrative child district, *Kabupaten Sekadau*. The dotted line depicts the new boundary between the mother and child district and the grey lines correspond to village boundaries. The white dots show the locations of the respective two capitals.

[3]Indonesia's administration is organized along provinces (*Propinsi*), districts (*Kabupaten/Kota*), sub-districts (*Kecamatan*) and villages or urban precincts (*Desa/Kelurahan*).

[4]For a discussion of the different dimensions of Indonesia's decentralization reform see for example Sjahrir et al. (2014), Ostwald et al. (2016), and Kis-Katos et al. (2017). The determinants of district splits are discussed in Fitrani et al. (2005) and Pierskalla (2016).

New districts were formed through administrative splits of existing ones, where the original district—referred to as the *mother*—retained its administrative capital and institutions, while the new district—referred to as the child—had to establish these institutions from scratch in a newly designated capital. Figure 3.1 illustrates this process for the district of *Sanggau*, from which the new district of *Sekadau* seceded in 2003. Between them, a new jurisdictional border was formed, which due to the preceding decentralization reforms, now divides the sphere of control between two local and influential decision-making units. Legislation foresaw that splits may only be facilitated within provincial boundaries, hence new boundaries do not overlap with existing provincial boundaries.[5] Given that mountain ranges and large rivers mostly coincide with upper-tier provincial boundaries, the newly established district boundaries are also largely independent of important geographical features (Burgess et al., 2012).

FIGURE 3.2: District splits and forest cover across Indonesia



Note: Black lines denote the new boundaries between mother and child district of the 115 splits included in our sample, described in section 3.4. Green shading indicates the extent of forest cover in 2000 from Global Forest Change (GFC) data based on 30×30m grid cells (Hansen et al., 2013), grey lines outline the extent of Indonesian land territory.

The legislation guiding the district splitting process was complex and required the fulfilment of numerous criteria (Alesina et al., 2019). This resulted in an average gap of one to three years between the first proposal and an official decree legislating the split (Burgess et al., 2012). Because early lobbying for splits was commonplace even before the proposal (Pierskalla, 2016), the actual waiting time from first plans to the final realization of the administrative split was at times even longer. In practice, more than 150 splits fulfilled the criteria, leading to an increase in the number of districts from 341 in year 2000 to 511 districts in year 2014. In terms of regional coverage, splits were

[5]Splits usually followed sub-district lines, which do not play a relevant role as polities within Indonesia and were themselves also subject to splits in the same period (Pierskalla, 2016).

dispersed across the entire Indonesian archipelago, covering all major islands as shown in Figure 3.2.[6]

3.2.2 Forestry and natural resource management

Districts also became in charge of forestry management, which underwent the most drastic decentralization reforms (Barr et al., 2006). Instead of reporting to the Ministry of Forestry, the newly created district forest departments became responsible for monitoring and levying taxes (Thung, 2019). In the early stages of decentralization, they were also granted the right to issue logging licenses, but continued to do so even in later years (Alesina et al., 2019). At the same time, legislation foresaw that districts receive 80% of forestry sector revenues and royalties from other natural resource extraction, e.g., from oil and mining, that originated on their own land.[7] While these revenues—generated from, e.g., concessionaire dues—are collected by the central government, the original fiscal distribution scheme remained in place despite later recentralization tendencies (Ostwald et al., 2016). As a result, resource rents have quickly become an essential source of funding for district governments and local elites (Thung, 2019).

In contrast to forestry revenues, fiscal decentralization did not mandate direct revenue sharing between central and local governments with respect to rents from oil palm, which became the dominant agricultural crop in Indonesia since the reforms. As the world's largest producer, the Indonesian oil palm sector employs more than 20 million people directly or indirectly (Nurfatriani et al., 2022) and is a crucial revenue source. Instead, the central government collects revenues related to palm oil as a commodity via, e.g., export levies, while district governments receive legal revenues from taxing land and income (Nurfatriani et al., 2022). However they have also been illegally selling land concessions (Smith et al., 2003; Barr et al., 2006), ignoring illegal deforestation (Amacher et al., 2012), and accepting electoral campaign contributions from the oil palm sector (Mongaby, 2018; Cisneros et al., 2021). District governments thus have an incentive to attract oil palm plantations, often by facilitating forest conversion (Irawan et al., 2013; UNEP, 2016), and consequently became important players in terms of their leverage to issue licenses (Sahide et al., 2015).

[6]Figure 3.2 displays only those splits that we use in our analysis (cf. section 3.4).

[7]Decentralization law UU 25/1999 Article 6.2 stipulates that such revenues go towards the originating district government. See Thung (2019) for a detailed discussion of Indonesian forestry sector decentralization.

Alongside the Amazon and Congo basins, Indonesia is home to the largest tropical forests worldwide. With its abundant wildlife and as a natural carbon sink, its protection plays a key role for reaching international climate-change and biodiversity targets. Over the past decades, this rich natural habitat has been under heavy deforestation pressures due to both human settlement and agricultural land expansion. In the first decade of the 21st century alone, Indonesian forests have been cut at an average rate of 47,600 hectares per year, reducing the extent of primary forest by 6% over 12 years (Margono et al., 2014). These trends have also persisted in the following decade. Recent estimates show that deforestation is primarily driven by large-scale oil palm and timber plantations (40%), followed by grassland conversion and small-scale agricultural activities (20%) (Austin et al., 2019). As a consequence of the decentralization reforms, district governments have become key actors for forest protection not only directly (Burgess et al., 2012), but also indirectly by controlling one of the major drivers of deforestation in Indonesia, oil palm expansion (Austin et al., 2019; Cisneros et al., 2021).

3.3 Theoretical framework

Land-use change creates large economic benefits for local administrations via revenues from land rents or illegal collusion (e.g., Alesina et al., 2019; Thung, 2019). 80% of revenues from the forestry sector are transferred to the originating district (cf. section 3.2). Beyond taxes and payments from the central government, such revenues have become an important income source for district governments and local elites (Thung, 2019). In the aftermath of the decentralization reforms, district governments had considerable influence on land-use change decisions. This included the expansion of agribusinesses, most notably oil palm, which became an important resource base for district governments and local elites, creating incentives for rent-seeking (Cisneros et al., 2021). In fact, opportunities to generate greater income from natural resources are seen as a key motivation behind district splits (Fitrani et al., 2005; Pierskalla, 2016). As a consequence, together with rising global demand for palm oil, the oil palm plantation area has significantly expanded since the early 2000s. At the same time, land-use change also bears political costs when associated for example with land grabbing, labor market marginalization, the loss of environmental services, or environmental damages (Krishna et al., 2017; Brito et al., 2019; Xu et al., 2022). Local administrations therefore face a benefit-cost calculation when deciding to support the conversion of natural forests into agricultural use. District splits change this benefit-cost calculation of both the existing (mother) districts and the newly formed (child) districts.

Immediate land-use rents District splits fundamentally alter the local governments' prospects to access land rents in the future, starting from the moment that a split becomes foreseeable and likely. Mother governments will have an increased incentive to extract immediate rents that are to be generated through legal (or illegal) deforestation before the split is legislated, and the jurisprudence of the territory is passed to the new child district. Once a split is formalized, the relevant district area is transferred to the sphere of influence of the new child government. Immediate rents from forest conversion now yield income opportunities for the new government. Together with the fact that new districts need to build their own institutions and resource base (Grossman et al., 2014), this might exert an upward pressure on deforestation after districts splits.

Medium-term land-use rents A mother district aims to maximise the economic benefits from land use and will therefore reassess its investment strategy in anticipation of future district splits. Once a district split is expected to take place, mother governments will face a much lower incentive to foster the establishment of new oil palm plantations on the area of the prospective child district as their future revenues will not go towards the mother government. This is especially true as there is a time-lag of about three years between the seeding of trees and the first harvest (Ismail et al., 2002). In the short-run, the anticipation of a split thus potentially de-incentivizes land conversion and hence deforestation. Once the split is effective, future rents associated with investments in oil palm in the new area will go towards the child government. Because palm oil is a key industry in rural areas—where most of the split (boundaries) are located—and payments contribute towards district governments' revenues, child districts will also face an incentive to expand oil palm plantations. In the medium run after the splits, deforestation thus potentially increases due to land-conversion pressures.

Constituencies' preferences Splits might also substantially reduce deforestation rates by moving government policies closer to the preferences of district constituencies. New districts have tended to become ethnically more homogeneous (Pierskalla, 2016; Alesina et al., 2019; Bazzi et al., 2021b),[8] which is associated with improved public service delivery—both in general (Alesina et al., 1999) and in Indonesia in particular (Bandiera et al., 2011).[9] This has been shown to improve forest protection, where homogeneous populations can control elected leaders more closely (Alesina et al., 2019). Local administrations therefore consider the political costs of land-use change and contrast them to the potential rents they generate. If a district split results in greater ethnic homogeneity in the new child district and forest conservation and the protection of small farmers are valued by the local population, we should observe a sustained longer-term slowdown of deforestation rates, but only in the aftermath and not before of district splits.

Administrative incapacity As long as the new district governments are in the process of formation, their capacity to monitor illegal deforestation might not yet be fully-fledged, which could increase deforestation rates immediately after the split. Moreover, as new district governments still need to set-up licensing processes and attract industries to exploit land-use associated rents, larger investments into new oil palm (or other) plantations may only materialize in the medium-run after the split. This could cause a decline in deforestation rates in the short-run but at the same time increase deforestation and oil palm expansion in the medium-run.

The creation of new political centers Existing literature suggests that when cities become administrative capitals, significant economic growth and an expansion of urban settlements is induced (Bluhm et al., 2021). While the mother district's capital retains its role in the process of jurisdictional splits, a new capital with all its relevant institutions has to be formed in the child district. This way, previously administrative-subordinate cities suddenly become central hubs for the new jurisdiction, inducing local construction booms and thereby stimulating the economy (Fitrani et al., 2005; Grossman et al., 2014; Thung, 2019). At the same time, urbanization is known to be a small, yet significant driver of deforestation in Indonesia (Austin et al., 2019). Spillovers from urban expansion in new centers thus potentially increase deforestation rates once a split has taken place, but not before. The distance to

[8]This has also been documented for other countries, where underrepresented areas tend to split off more frequently (Grossman et al., 2014).

[9]Greater homogeneity in new districts has also been shown to reduce conflicts (Bandiera et al., 2011; Bazzi et al., 2021b). Government fragmentation improves the fiscal resource base of new administrative entities by triggering yardstick competition between them (Grossman et al., 2017), which potentially improves service delivery for residents.

the new capital however potentially has heterogeneous effects as land rents decrease with the distance to cities (Angelsen, 2007). Once a district splits, the spatial relationship between child villages in close proximity to the new boundary and their expanding new capital changes profoundly. While on average most villages move closer to the capital, some are located in new peripheries and thus have less access to resources accumulated in the center (Grossman et al., 2017). In short, the literature suggests that in the aftermath of district splits, child villages that come closer to the newly formed capitals will face larger deforestation pressures.

In summary, we expect decentralization and the creation of new districts to affect deforestation patterns both before and after the district splits take place. The interplay of opposing incentives makes ex-ante predictions about the direction of the effect ambiguous: In anticipation of a split, the mother governments' land-use decisions depend on a trade-off between short-term gains from deforestation and the expected revenue losses in the medium-run if oil palm plantations locate in the soon-to-be-lost areas. If short-term incentives dominate, deforestation might increase before the split. By contrast, if anticipatory considerations—particularly with regard to oil palm plantations—play a dominant role, deforestation might decrease before the split.

After a split has taken place, deforestation trends in neighboring villages will depend on differences in the decisions by the mother and the newly formed child government. Children continue to face the incentive to extract short-term rents from forestry, and deforestation might also be exacerbated due to a temporarily more limited monitoring capacity. At the same time, increases in the ethnic homogeneity of the population might alleviate deforestation pressures. Changes in the spatial relationships with respect to the new capital will have heterogeneous effects: Spillovers from the development of new capitals could increase deforestation in more central locations, but decrease it in new district peripheries. Finally, in the medium run, the incentives to raise revenues from oil palm plantations are likely to foster land-use change and increase deforestation, whereas in the immediate aftermath of the split these dynamics might be still mitigated by limited administrative capacity.

3.4 Data and methodology

3.4.1 Data

We identify 115 newly created relevant boundaries between mother and child districts by relying on official district boundaries from 2014 from Statistics Indonesia (BPS), tracing back administrative entities to their historical boundaries for each year between 2000 and 2014.[10] These district splits reshape the administrative environment of 33,787 villages within the boundaries of mother and child districts, which we track from 2000 to 2018 using administrative and remotely sensed land-use data. Villages can appear multiple times in our analytical sample if they are located close to several new boundaries: For instance, because a child district was subject to a further split in a later year, or because they are located close to several newly formed child districts. Our sample therefore consists of repeated cross-sections of villages recorded at different periods in time, re-centered relative to the year of the district split. The empirical strategy will account for potential issues raised by duplicate villages as treated or controls.

Our main variables of interest are based on remotely sensed high-resolution data measuring different land-use dynamics: a) Forest losses between 2001 and 2018 (Global Forest Change data, Hansen et al., 2013); b) oil palm expansion between 2001 and 2018 (Gaveau et al., 2022); and c) settlement expansion between 2001 and 2015 (Global Settlement Footprint data, Marconcini et al., 2021). For each source, we construct measures of the initial area extent in the year 2000, as well as annual expansion measures. Socioeconomic data are taken from Indonesia's village census PODES. Further district-level characteristics, such as ethnic composition, are obtained from the 2010 Indonesian national census. To proxy for the discontinuous treatment of villages at the newly established boundaries we calculate the bee-lines distance from village centroids to the border.[11] Additionally, we also calculate distances to the respective district capitals before and after the split. Summary statistics and a list of sources are outlined in Appendix Table B.1.

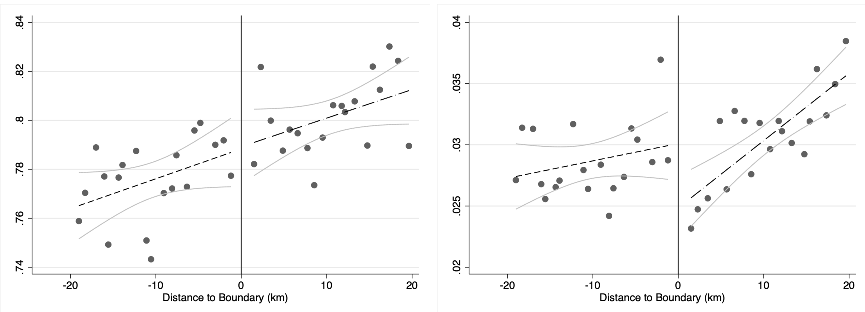
[10]We exclude splits where children do not share a physical boundary with the mother, including island splits (separated by water) and splits that involved several children at the same time and partially resulted in new boundaries only among the children. We further exclude areas where forest cover was relatively small to begin with by dropping splits of large urban centers into smaller administrative units, as well as splits that had less than 50% forest cover in 2000.

[11]On average, a mother district includes 172 villages and a child district includes 121 villages. Out of these, 64 and 60 villages are located "close" to—within 20km of—the newly established borders, respectively.

3.4.2 Econometric framework

To analyze whether the process of district splits changed deforestation dynamics, we employ a spatial regression discontinuity design (SRDD) strategy. This strategy relies on the main assumption that land-use change dynamics develop continuously in space and no systematic discontinuities arise across neighboring villages as long as they are located within the same district. If this identifying assumption holds, we can interpret all discontinuous jumps in deforestation on the two sides of a newly established (or soon-to-be established) district boundary as a causal effect of the district splitting process. As district administrations can adjust their decisions already in anticipation of an upcoming split, conceptually we expect changes in deforestation dynamics on the two sides of the boundary occurring after and also before a split has taken place; however only when the local government and economic actors could foresee along which lines the district will be splitting in the near future.

FIGURE 3.3: Spatial RDD: Initial forest cover and forest loss around new district boundaries



(A) Forest cover in 2000 (% of village area)

(B) Short-term forest loss post-split (% of 2000 cover)

Note: Dots represent 20 binned means at each side of the cutoff (the new district boundary) for our sample of 115 splits. The left side of each panel displays villages located in mother districts, whereas the right side shows villages in the newly formed child districts. Short-term forest loss in panel B captures cumulative deforestation from the year of the split to three years after the split. Dashed lines represent linear fits of the data with 90% confidence intervals.

In our SRDD strategy the newly established boundaries represent a sharp cutoff, and the running variables are defined by the villages' distance to the new boundary on each side of the border. Figure 3.3 visualizes our strategy by plotting initial forest cover against the distance to the new boundary in panel A, and total forest loss from the year of the split up to three years after the split against the same distance in panel B. Villages to the right of the cutoff are part of the new child district and thus are our treated units. As we move from the left to the right, towards the newly created districts, the distance to the original district capital increases monotonously and places become relatively more "remote" from the perspective of the original district administration. This leads to a monotonous increase in the initial forest cover, which simply reflects that more remote areas were generally more forested to begin with. More importantly, the forest cover is continuous across future boundaries in panel A and shows that future district splits were not linked to past discontinuities in forest cover (measured in 2000). This gives a first indication that the identifying assumption of variable continuity at the cut-off might hold, which we will support with further balance tests on a large number of topographical and socio-economic characteristics (reported below).

Panel B in Figure 3.3 shows that in the first three years after the district split, the extent of deforestation was generally increasing with remoteness, yielding a positively sloped linear fit at both sides of the boundary. Places that started with a larger forest cover also experienced on average more deforestation. However, in contrast to panel A, a sharp decrease in deforestation can be observed in the first three years after the split in the border area of the newly formed district. After the original district split up and a new child district was created, deforestation is substantially lower in villages that became part of the new child district than in their direct neighbors that remained part of the mother district. This can be taken as a first indicative evidence for a relative reduction in deforestation in the border regions of newly formed child districts.

We test this more formally by relying on an SRDD regression framework to assess whether deforestation dynamics developed smoothly in space across neighboring villages before and in the aftermath of the district split:

$$Deforest_{vs} = \eta\, Child_{vs} + f(Distance_{vs}, Child_{vs}) + \beta\, \mathbf{X}_v + \theta_s + \epsilon_{vs}, \quad (3.1)$$

where $Deforest_{vs}$ measures forest loss in village v before or after the district split s occurred in a given period, $Child_{vs}$ indicates a village's location in the new district and $Distance_{vs}$—measuring the distance between a village's centroid and the respective split boundary—is the continuous forcing variable.[12] The function $f(Distance_{vs}, Child_{vs})$ includes either two linear or quadratic

[12]We exclude villages with a centroid in close proximity to the boundary (<1km) as they represent random shapefile artefacts.

polynomials of distance, separately estimated on the two sides of the border. $\mathbf{X}_v$ is a vector of time-invariant village-specific controls, including village altitude and the initial share of the respective land-use type in 2000 that is being analysed in the regression (forest cover, oil palm, or human footprint area). Split fixed effects θ_s ensure that we only compare villages with their corresponding neighbors in our sample of pooled splits. Our main model is a pooled cross-section of 115 splits that took place at varying points in time between 2002 and 2014 (cf. Appendix Figure B.1). The split-level fixed effects θ_s further account for differential deforestation trends across the years, which we contrast with other, less strict, specifications.

We test our results based on fixed and optimal bandwidths. Our preferred specification uses a 20km window which eases the comparison across different estimations that rely on different outcomes. Results are robust to using robust-bias corrected (RBC) methods.[13] To account for potential serial correlation due to some villages being included more than once (either as treatment or control units), we cluster standard errors at the split level (e.g., Dube et al., 2010; Cantoni, 2020).[14] In our preferred specification, we fit our underlying outcome variable linearly on both sides of the cutoff, unless indicated differently. This helps to avoid overfitting and is supported by the visual examination of our data (cf. Figure 3.3) and by estimated information criteria (AIC/BIC).

Causal identification in the SRDD framework relies on two assumptions: a) Boundaries represent arbitrary thresholds across which all potential outcomes move continuously in the absence of treatment; and b) the absence of endogenous sorting, that is, villages cannot influence whether they end up as parts of the mother or the child district. While village boundaries are stable across time and space, new district boundaries are not randomly drawn in space but usually follow pre-existing sub-district borders. Our identifying assumptions require that the number of villages as well as topographical and economic characteristics are continuous across sub-district boundaries. If these assumptions are fulfilled, any differences in economic characteristics around the new borders must arise as a result of the decentralization process.

The assumption of no endogenous sorting can be assessed by a test of continuous density, for instance by estimating a local polynomial density function

[13] We use Calonico et al.'s (2014) dedicated STATA package *rdrobust*.

[14] Within the bandwidth 20km at each side of the boundary, a total of 1,325 observations (<10% of our sample) are villages that are included more than once.

as proposed by Cattaneo et al. (2020). Figure B.2 in the Appendix shows visually that the density plot is fairly continuous around the cutoff. The formal test of a discontinuity can be rejected, but only with the relatively low p-value of 0.102. However, a battery of balance checks in Appendix Table B.2—applying our SRDD design from eq. (3.1) to village-level socio-economic and topographical variables—does not show any significant discontinuities around the future district boundaries.[15] All 22 reported variables develop smoothly across future district boundaries, reducing concerns about endogenous border location.

Finally, we acknowledge that the timing of and reasons for each district split are not exogenous. While this might bias a cross-sectional analysis, we believe this is not an issue in our setting: First, in our main specification we only compare neighboring villages that appear on two sides of the same border before and after a split. Second, previous literature has shown that district-level correlates of deforestation such as forest cover in 2000, GDP and ethnic conflicts are not significantly related to the exact timing of the split, alleviating concerns regarding structural differences across time (Burgess et al., 2012; Alesina et al., 2019). And lastly, endogenous differences across the two districts resulting from a split, for example in their ethnic composition (Fitrani et al., 2005; Pierskalla, 2016; Bazzi et al., 2021b), are less of a concern. Results from balance checks discussed above lead us to assume continuity along unobserved dimensions (like village-level ethnic composition) as well.

3.5 Results

3.5.1 Main results

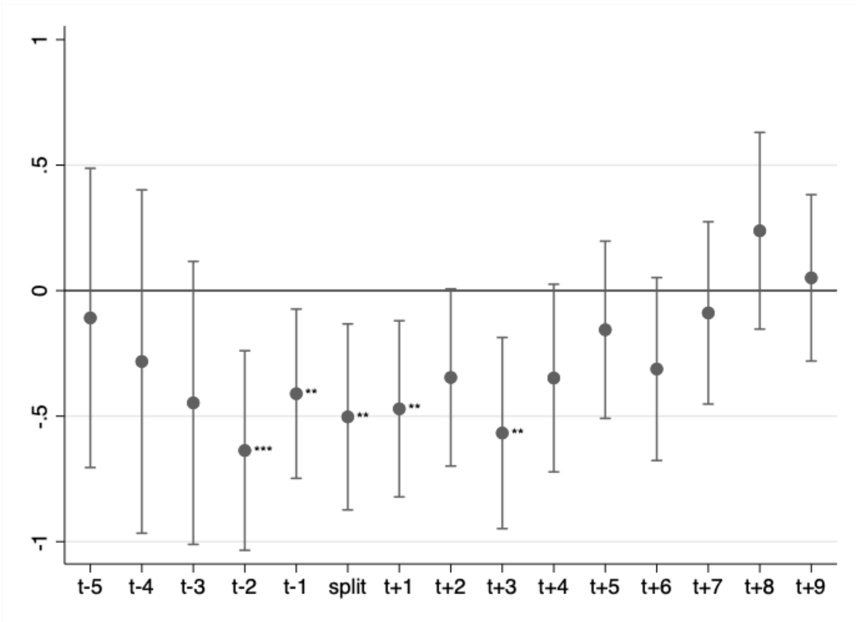
To investigate the dynamic effects of the district splitting process, we rely on yearly deforestation rates before and after each split as dependent variables in equation (3.1).[16] Figure 3.4 plots the estimates from 15 individual regressions, assessing deforestation starting five years before administrative splits occurred to up to nine years after. The results are based on our preferred specification relying on a linear fit, split-ID fixed effects, and controlling for

[15] We test the continuity of land-use characteristics in 2000 (forest cover, oil palm area, built-up settlement extent), geographic factors (altitude, coastal indicator, distance to nearest city by type), rural location, and initial conditions in 2000, including population size, socio-economic characteristics, and access to public services.

[16] Early and late splits lack information for pre- and post-split years, respectively, reducing the sample size at lags or leads of higher order (cf. Appendix Figure B.1).

initial ecological conditions. The results show that deforestation starts to significantly decrease in future child districts already up to two years before the split was actually implemented. Decreases in deforestation persist until up to three years after the split, but estimates get closer to zero over time, showing no statistical difference between mothers and children four years after the split. Thus, the pace of deforestation picks up in child districts in the long-run and catches up with that of mother districts over time.[17]

FIGURE 3.4: Dynamic SRDD effects: Deforestation



Note: The figure displays treatment coefficients η from separate regressions (eqn. (3.1)) that pool village observations based on their temporal distance to the district split year (denoted by "split"). The dependent variable measures deforestation in that given year, transformed by the inverse hyperbolic sine. The sample consists of villages whose centroids lie within a fixed bandwidth of 20km around the 115 district split boundaries. The SRDD relies on a linear fit. The graph displays 90% confidence intervals with standard errors clustered at the split level. */**/*** denote significance levels at 10/5/1% respectively.

Table 3.1 collects these results by focusing on the years around the official district split—from three years before up to three years after the split. It shows SRDD results that estimate the difference between average deforestation rates among neighboring villages located in a child and a mother district before the split (in panel A) and after the split (in panel B). The results again rely on a linear fit but introduce fixed effects and controls step-wise.

[17]Table B.3 in the Appendix aggregates deforestation into three-year intervals and shows no significant differences between neighboring villages from the fourth year after the split, nor in forest cover at the end of our sample period in 2018.

While column 1 reports the basic SRDD without any further controls, column 2 absorbs all macro-region-level shocks over time by introducing island-year fixed effects. Column 3 relies instead on split ID fixed effects, which restricts the comparison to villages that are located in the neighborhood of each split, controlling away all spatio-temporal variation at a district scale, whereas columns 4 and 5 also control for initial forest cover and altitude.

TABLE 3.1: SRDD effects: Deforestation in child vs. mother districts

	(1)	(2)	(3)	(4)	(5)
	Panel A: Dep.: asinh *Pre-split mean deforestation*				
Child	-0.816***	-0.549***	-0.498***	-0.483***	-0.652***
	(0.271)	(0.199)	(0.187)	(0.166)	(0.175)
Bandwidth	20	20	20	20	15 (42)
Observations	14,320	14,320	14,320	14,319	10,617
Adj. R^2	0.004	0.165	0.297	0.396	
	Panel B: Dep.: asinh *Post-split mean deforestation*				
Child	-0.566**	-0.405*	-0.404**	-0.390**	-0.568***
	(0.237)	(0.211)	(0.200)	(0.151)	(0.172)
Bandwidth	20	20	20	20	13 (35)
Observations	14,320	14,320	14,320	14,319	9,670
Adj. R^2	0.004	0.215	0.355	0.472	
Island-year FE	No	Yes	No	No	No
Split-ID FE	No	No	Yes	Yes	Yes
Controls	No	No	No	Yes	Yes

Note: The dependent variable is the average deforestation within three years before (Panel A) and from to three years after (Panel B) the split, transformed by the inverse hyperbolic sine. Child is a binary indicator for villages located in the new child district. The sample consists of villages whose centroids lie within a fixed bandwidth of 20km around the 115 district split boundaries. The bandwidth in column 5 is determined using an RBC estimator (Calonico et al., 2014). The SRDD relies on a linear fit. Controls include village altitude and forest cover in 2000. Standard errors are clustered at the district split ID. */**/*** denote significance levels at 10/5/1% respectively.

Across all specifications, child villages consistently experience statistically significantly lower deforestation rates than mother villages before as well as after the split. This difference is also considerable in economic terms: In our preferred specification in column 4, villages in child districts deforest 32–38% less than neighboring villages in mother districts.[18] Compared to the mean annual deforestation rate of 1.5% in our sample, it implies that the deforestation rate is around 0.5 percentage points lower in child districts. Results in this specification are based on a fixed bandwidth of 20km. Alternative

[18]Percentage changes in the outcome variables after regressing on binary variables are equally interpreted as in Log-binary variable regressions: $e^{\beta} - 1$ (Halvorsen et al., 1980; Bellemare et al., 2020).

specifications that rely on RBC-based bandwidths (in column 5) yield larger estimates.

The results are robust to using different specifications and outcome definitions. Estimates remain significant with somewhat larger effect sizes when fitting the data with a local quadratic polynomial (cf. Table B.4 in the Appendix). Results are furthermore robust when choosing alternative fixed bandwidths (cf. Appendix Figure B.3), as estimates remain significant at the 10% level for distances between 5 to 30km. Lastly, we run placebo regressions, artificially shifting borders up to 40km away from the actual boundaries. If the new administrations influence deforestation discontinuously only at the realized border, choosing other cutoffs in close neighborhood should lead to zero effects. Figure B.4 in the Appendix confirms this by showing insignificant and close to zero estimates for all placebo cutoffs.

3.5.2 Mechanisms

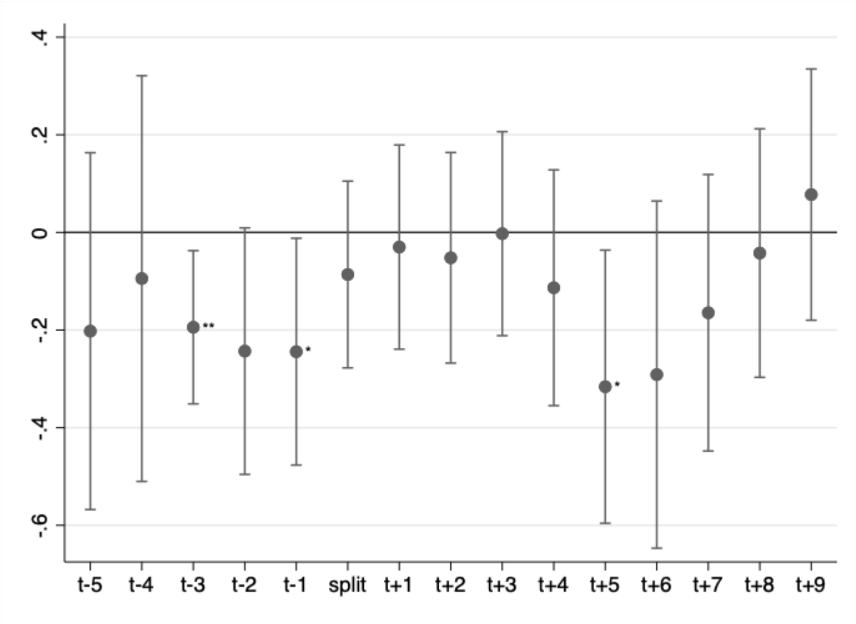
Our results document a temporary deceleration of deforestation in child districts as compared to mother districts, identified by a discontinuity at the newly established boundary. After splits, neighboring villages fall under the sphere of influence of new district administrations, so that these differences might reflect changing incentives to protect the remaining forest. However, our results show very similar decreases in the deforestation rate already in anticipation of district splits, which cannot yet be attributed to decisions made by the new district administrations. In this section, we analyze the interplay of different incentives induced by altered cost-benefit considerations before and after the split, focusing on how they affect the behaviour of both mother and child governments with regard to land-use decisions.

Immediate land-use rents Theory suggests that if forest conversion yields large immediate rents—e.g., through the sale of land-use licenses (Burgess et al., 2012) or wood products—mother governments have an incentive to try to extract as many resources as they can from the soon-to-be-lost areas. This would lead to a surge in deforestation rates on the area of the child district as soon as a district split is expected, which usually precedes the actual split by a few years. The results observed in Figure 3.4 speak against this hypothesis: Deforestation in areas that will belong to child governments after the split decelerates already before the jurisdictional change, showing no evidence for mother governments overusing the future child district's forestry resources in anticipation of a split. From the moment the split actually materializes, the rights to exploit forestry resources shift to the new child government for the same area. However, we also do not observe increased deforestation rates in the immediate aftermath of the split. Taken together, the deceleration of deforestation both before and after the split suggests that prospective benefits from short-term resource rents are over-compensated by other factors.

Medium-term land-use rents If deforestation is mainly driven by investments to expand agricultural production instead, then administrative decisions to support deforestation must follow a medium-term cost-benefit analysis. In consequence, the mother district's government will abstain from fostering land-use change in the soon-to-be lost areas and prefer to support agricultural development within its own remaining area. Starting when the wish for a new split is announced, medium-term rent considerations will create a gap in land-use dynamics between mothers and child districts.

Indonesia's decentralization reforms were accompanied by massive land-use change that shaped medium-term land-use rents: Triggered by a global palm oil boom, plantation area of oil palm increased from about 6% of village area in 2000 to 9.2% in 2018. This expansion was among the major drivers of deforestation in Indonesia. As oil palms take about three years to become productive after planting, remotely sensed oil palm expansion data offers us a useful opportunity to assess the role of medium-term agricultural rent considerations. To verify whether changes in oil palm expansion contribute to our findings, we rerun our main model in equation (3.1) with oil palm area expansion as the dependent variable. Figure 3.5 displays how oil palm area developed around the time of the splits. Estimates mirror the trends observed for deforestation closely around the time of the district split as oil palm expansion decelerates by around 20% in child villages already up to three years before the split took place. This suggests that the mother districts' unwillingness to promote agricultural development in the soon-to-be-lost areas contributes to forest protection in the short run, because costs associated with such investment fall short of obtainable rents. Once the split has taken place, the difference between villages at either side of the cutoff loses significance and the pace of land-use change in villages located in the child district catches up with that of the mother district. One explanation is that the new child governments now face the incentive to promote oil palm conversion for rent-extraction on their area as well.

FIGURE 3.5: Dynamic SRDD effects: Expansion of oil palm area



Note: The figure displays treatment coefficients η from separate regressions (eqn. (3.1)) that pool village observations based on their temporal distance to the district split year (denoted by "split"). The dependent variable measures new oil palm area in that given year, transformed by the inverse hyperbolic sine. The sample consists of villages whose centroids lie within a fixed bandwidth of 20km around the 115 district split boundaries. The SRDD relies on a linear fit. The graph displays 90% confidence intervals with standard errors clustered at the split level. */**/*** denote significance levels at 10/5/1% respectively.

Constituencies' preferences While the pre-split decline in land-use change points toward strategic divestment on the side of the mother district, the post-split decline could also result from socio-political considerations. For the period after the split, decreases in deforestation could have been especially pronounced in places where decentralization has led to a better matching of preferences between district administrations and their constituencies. In this scenario, excess deforestation would come at a political cost for elected leaders. To verify this hypothesis, we investigate the role of decreasing ethnic heterogeneity, which has been proposed as a main mechanism behind the improvements of public service delivery and deforestation reductions in Indonesia (Alesina et al., 2019). Using data from the 2010 national census, we construct ethnic fractionalization measures, as proposed by Alesina et al. (2003), both in the mother and child district. In our sample, average fractionalization is 0.57—a comparably large value, mirroring Indonesia's ethnically diverse population. This value decreases on average by about 1.1 points or 2% in the child districts after the splits. Table 3.2 augments our main model with a binary variable identifying splits that resulted in a more homogeneous population in the child districts.[19] If the theory holds, we would expect deforestation to decrease by more in the aftermath of a split if it resulted in a more homogeneous population. Although the interaction term is negative, we do not find statistically significant differences between child districts that became ethnically more homogeneous after the split. It therefore seems that, in contrast to Alesina et al. (2019), the decline in forest losses after a district split cannot be linked to the mechanism of constituencies' preferences. This is also in line with results presented in column 6, which do not show long-term improvements in forest conservation that would corroborate such a mechanism.

TABLE 3.2: SRDD effects: Heterogeneities by ethnic composition

Dependent:	*ln Mean deforestation*					*Forest cover*
Period	Pre 6-4	Pre 3-1	Post 0-3	Post 4-6	Post 7-9	in 2018
	(1)	(2)	(3)	(4)	(5)	(6)
Child	-0.148	-0.319	-0.337	0.112	0.101	-0.132
	(0.439)	(0.251)	(0.214)	(0.216)	(0.189)	(0.084)
Child $\times$ Decrease in	0.380	-0.226	-0.103	-0.429	-0.333	0.035
ethnic fractionalization	(0.511)	(0.326)	(0.299)	(0.317)	(0.325)	(0.147)
Split ID FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,695	12,822	12,822	12,822	12,822	12,822
Adjusted R^2	0.410	0.385	0.460	0.453	0.462	0.635

Note: The dependent variable is average deforestation in the years indicated, transformed by the inverse hyperbolic sine. Child is a binary indicator for villages located in the new child district. *Decrease in ethnic fractionalization* identifies villages in which the child district's ethnic fractionalization is smaller than the fractionalization of the original district. The sample consists of villages whose centroids lie within a fixed bandwidth of 20km around the 115 district split boundaries. The SRDD relies on a linear fit. Controls include village altitude and forest cover in 2000. Standard errors are clustered at the district split ID. */**/*** denote significance levels at 10/5/1% respectively.

Administrative incapacity A temporal administrative incapacity among new child districts could also influence land-use change dynamics after district splits. Monitoring and enforcement institutions might take some time to set up, which could increase illegal deforestation, especially in regions that are more remote and hence incur higher costs of monitoring and enforcement. However, deforestation could be also reduced if the new administrations are slow to start promoting regional development right after the split. Again, remoteness could play a moderating role in this process. To test this mechanism we create a binary variable that identifies splits in which the distance of child villages to their new capital is reduced by more than the sample

[19]This is the case for 51 out of 99 splits. We cannot compute changes in ethnic composition for 16 splits for reasons of data availability.

median.[20] Columns 1 and 2 in Table 3.3 display results from interacting the treatment variable in our main model with this measure. If monitoring and enforcement of forest conservation is the main driving force behind the differences in land-use change, we would expect an increase in deforestation in places that are relatively more remote from the perspective of the newly formed child districts as the costs of monitoring increase in distance. By contrast, we would expect relatively more favourable deforestation dynamics in areas that became less remote after the district split due to a larger ease of monitoring. There is no evidence for either of these hypotheses: (1) Deforestation does not increase but even significantly declines in the relatively more remote areas after the district split; and (2) the interaction effect is positive (and insignificant), which does not show more beneficial deforestation dynamics in places that become relatively less remote after the split.

TABLE 3.3: SRDD effects: Heterogeneities by closeness to the new political center

Dependent:	*ln Mean deforestation*		*ln Mean new settlement area*			
Period	Pre 3-1 (1)	Post 0-3 (2)	Pre 3-1 (3)	Post 0-3 (4)	Pre 3-1 (5)	Post 0-3 (6)
Child	-0.626***	-0.559***	-0.280	-0.243	-0.510*	-0.521**
	(0.211)	(0.198)	(0.216)	(0.164)	(0.288)	(0.200)
Child × Large decline in distance to capital	0.278	0.373			0.757*	0.893***
	(0.351)	(0.290)			(0.403)	(0.334)
Split ID FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	14,319	14,319	14,299	14,299	14,299	14,299
Adj. R^2	0.399	0.474	0.398	0.456	0.399	0.456

Note: The dependent variable in columns 1 and 2 (3 to 6) is average deforestation (expansion in settlement area) in the years indicated, transformed by the inverse hyperbolic sine. Child is a binary indicator for villages located in the new child district. *Large decline in distance to capital* is a split-level binary variable measuring whether the villages' average decline in distance to their capital cities lies above the median. The sample consists of villages whose centroids lie within a fixed bandwidth of 20km around the 115 district split boundaries. The SRDD is relies on a linear fit. Controls include village altitude and forest cover in 2000. Standard errors are clustered at the district split ID. */**/*** denote significance levels at 10/5/1% respectively.

The formation of new economic and political centers In addition to the new orientation to political centers, villages in some splits also find themselves close to quickly developing and increasingly urbanizing centers, while others move towards the new peripheries of the district, resulting in diverging deforestation pressures. Columns 3 to 6 in Table 3.3 investigate the relationship between district splits and urbanization using remotely sensed yearly human settlement expansion measures. On average, we do not observe significant discontinuities in settlement dynamics across villages at the

[20] On average, the distance to the new capital in the child district is 42km closer than that to the original mother district (cf. Appendix Table B.1).

new boundary (columns 3–4). However, interacting the treatment indicator with a binary variable that distinguishes between splits in which villages ended up closer than the median to their new capital than before, reveals divergent effects (columns 5–6). While these dynamics appear already in anticipation of the district split, the relationship is only marginally significant. After the split, villages in districts that are not experiencing a larger reduction in the distance to their administrative centers—and hence remain similarly peripheral as they were before—experience a substantially smaller relative decline in urbanization than their immediate neighbors. By contrast, urbanization increases in villages that move relatively closer to an administrative center—and hence become more central. These results lend further empirical support to the argument that new capitals trigger localized economic booms (Fitrani et al., 2005; Grossman et al., 2014; Thung, 2019; Bluhm et al., 2021). In summary, while new political centers accelerate urbanization in their close proximity, administrative incapacity could be delaying the same process in more remote areas, resulting in a reduction in deforestation. Close to cities, the economic effects of a new political center push deforestation pressures up, cancelling out the unintended forest conservation impacts of administrative incapacity.

3.6 Conclusion

In recent decades, Indonesia underwent wide-sweeping decentralization reforms that led to a considerable sub-national government fragmentation. Relying on a spatial regression discontinuity design, we show that the creation of over 100 new districts temporarily slowed down deforestation in the newly formed jurisdictions. An analysis of deforestation dynamics around the time of the splits suggests considerable anticipation effects that also translate into relatively lower deforestation rates in new districts up to three years before administrative splits. In the medium run, however, deforestation rates equalize at the boundary of mother and child districts, resulting in no differences in the remaining forest cover on both sides of the boundary in the long run.

The results point to a strategic investment behavior by existing governments that maximize medium-run revenues. Deforestation and oil palm area expansion both slow down in areas that will become part of the new jurisdiction even before splits officially take place, suggesting that local governments in the mother districts decelerate land-use change in these areas in expectation of losing the future economic rents from this process. However, deforestation rates at the boundaries of newly formed districts equalize over time once the new child districts build up enough capacities to foster agricultural expansion of their own. In addition, we do not find evidence for lower deforestation in more ethnically homogeneous child districts, and thus cannot confirm that the mechanism of better matching constituencies' preferences translated into sustained long-term forest protection. On the contrary, our results lend credibility to the interpretation that the slump in deforestation rates is likely related to a temporary administrative incapacity effect that slows down the profitable land-use investments in the short run.

Such anticipatory land-use decisions before jurisdictional adjustments have not yet been empirically documented. This mechanism thus provides another perspective on the process of government fragmentation at the subnational level, adding an unintended positive consequence for the protection of forests. Even a temporal decline in deforestation rates holds the potential to transform a local economy and make it more environmentally sustainable. Central governments, NGOs, and other policy makers might consider offering additional incentives for new district administrations to protect natural forests, before they build up a development strategy that relies on agricultural expansion.

Our results also pose questions that are beyond the scope of this paper: Given that political budget cycles play a major role in Indonesia (Sjahrir et al., 2013; Kis-Katos et al., 2017; Cisneros et al., 2021), an analysis of the interplay of the observed effects with local elections could provide additional insights. This is particularly relevant as public office is seen as a means to capitalize on successful but costly election campaigns (Pierskalla, 2016), whereby medium-run land development can help to generate the needed revenues. Finally, our results raise the question about anticipatory strategies and administrative incapacity effects that go beyond land-use decisions. District splits could also yield negative externalities in other policy areas. The quality of public services—impacting among others education, health, infrastructure, or social equity outcomes—could similarly worsen before and after the splits. Additional research in this area could help to better understand the potential dynamic effects of district splits. Such further analyses could especially highlight further the trade-offs of the district splitting process as its short-term and long-term effects may not be fully aligned.

Chapter 4

Confined to Stay: Natural Disasters and Indonesia's Migration Ban

Andrea Cinque and Lennart Reiners[1]

Abstract

This paper investigates the impacts of international migration restrictions on communities' ability to absorb income shocks in the aftermath of natural disasters. We exploit the implementation of an emigration ban on Indonesian women as a natural experiment. After a series of violent assaults against female servants in Saudi Arabia, the Indonesian government issued a moratorium in 2011, thus preventing millions of women from migrating there as domestic workers. Relying on the exogenous timing of the ban and that of natural disasters, we estimate the causal effect of the absence of international migration as an adaptive strategy. We use a panel of the universe of Indonesian villages in a triple difference strategy to compare poverty levels in the aftermath of natural disasters in villages whose main destination is Saudi Arabia as opposed to others, before and after the policy shock. We find that in villages with strong ex-ante propensity to migrate to Saudi Arabia, poverty increases by 13% after the ban in face of natural disasters, exacerbating the already severe consequences of these events.

[1]The study is currently under revision and has been published as CESifo Working Paper No. 9837, available *here.*

4.1 Introduction

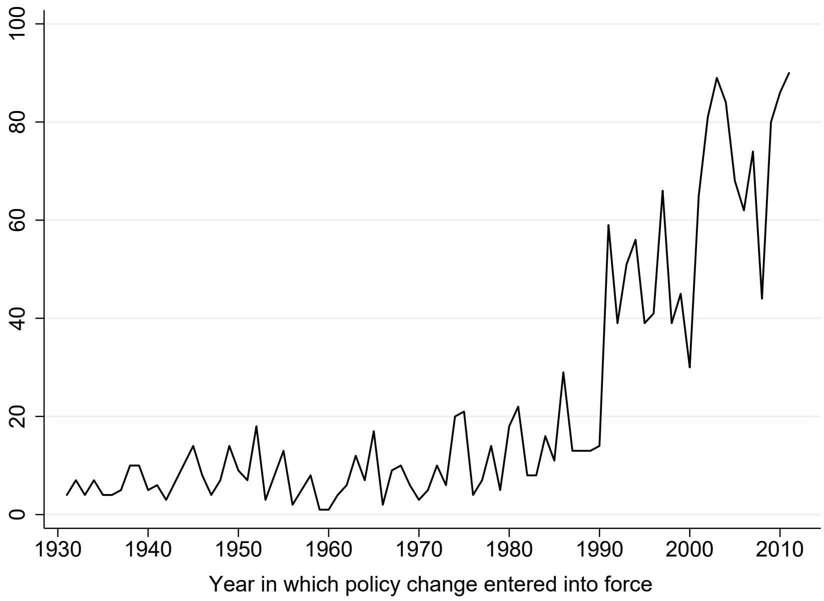
In the coming decades, climate change will exacerbate the frequency of extreme weather events such as floods, droughts and heat waves, affecting livelihoods in manifold ways (Jones et al., 2016; FAO, 2018; IPCC, 2021). The role of international migration as a coping strategy is becoming increasingly important: By 2050, the predicted number of *climate refugees* is estimated to reach hundreds of millions (Rigaud et al., 2018). These developments pose challenges for governments in both sending and receiving countries around the world. Historically, climate-induced migration has been little restricted because it mainly consisted of a within-border phenomenon (Cattaneo et al., 2019b) or legal hurdles to move across countries were loose (Spitzer et al., 2020). However, recently countries have been putting more emphasis on selective migration, resulting in more complex and restrictive regulations (Beine et al., 2016; Rayp et al., 2017; Haas et al., 2018). As shown in Figure 4.1, this trend suggests future scenarios where international migration will be further constrained, potentially undermining its role as a major coping strategy to climate change.

This paper examines how natural disasters affect poverty in a scenario where international emigration is heavily restricted. We exploit a unique natural experiment: The sudden implementation of an emigration ban in a country where 7% of the workforce was employed abroad (World Bank, 2017). After repeated cases of abuse and a death sentence for domestic workers in Saudi Arabia, the Indonesian government entirely banned the emigration of women who wanted to work there as domestic workers. This ban affected Indonesian villages to very different degrees due to heterogeneous destination-specific migration networks. We investigate whether restricting emigration deprived villages of their capacity to absorb income shocks induced by natural disasters, a widespread phenomenon in the country. The moratorium eliminated the possibility for Indonesians to emigrate to what was the second top destination country. In consequence, we show this inhibited an important adaptation strategy to natural disasters.

This paper is among the first to provide causal evidence of the effect of migration restrictions in the context of extreme climatic events.[2] We conduct our analysis in a highly localized setting for the universe of around 70,000

[2]Existing studies approach the relationship between tighter border restrictions and communities' climate-resilience either descriptively (McLeman, 2019) or theoretically (Benveniste et al., 2020; Burzyński et al., 2021).

FIGURE 4.1: Number of new restrictive immigration policies implemented worldwide



Note: Data taken from the DEMIG Policy Database (Haas et al., 2015).

Indonesian villages for the period 2005–2014. Our disaggregated data allow us to exploit: (i) Spatial variation in the main destination country of international emigrants across Indonesian villages; (ii) the exogeneity of natural disaster events, conditional on village fixed effects; and (iii) the implementation of an unexpected national emigration ban that affected emigration to one top destination country, but not others. Taken together, this allows us to causally estimate the effect of the moratorium in a triple difference (DDD) setting. We therefore compare the poverty levels of villages with migration links to Saudi Arabia—i.e. villages where the majority of emigrants went to in 2005—with the remaining villages, this depending on whether they were hit by natural disasters, before and after the moratorium was introduced in 2011.

This triple difference strategy overcomes several potential biases by fully saturating the estimation with all possible interaction terms. It allows controlling for time-varying changes in villages with migration links to Saudi Arabia, group-specific historical propensity to disasters and changes in resilience to natural disasters common to all villages. This way, we isolate the causal effect of natural disasters on villages more vulnerable to the introduction of the ban, i.e. those with migration links to Saudi Arabia. As the central prerequisite for causal identification, we follow Olden et al. (2022) in showing that the parallel trend assumption holds in the context of our triple difference model.

In our findings, we first demonstrate that villages with ex-ante migration ties to Saudi Arabia experienced a drastic reduction in international emigration. After the moratorium, the stock of overseas workers decreased by more than 30% compared to villages whose migratory networks were not affected, highlighting the importance of persistent migration networks (McKenzie et al., 2010). After the ban was implemented, these villages experienced a 13% increase in poverty once hit by natural disasters. Distinguishing between disaster types, we find that floods have the most devastating effect on poverty. Restrictions on emigration therefore further amplify the already significant implications for livelihoods induced by such catastrophic events. Estimates hold under manifold robustness checks, including alternative measurements of poverty and disasters, sub-sample adjustments, placebo regressions and falsification tests.

Internal migration is known to be an important response to climate-induced income shocks (Hornbeck, 2012; Marchiori et al., 2012; Kubik et al., 2016; Gröger et al., 2016; Dallmann et al., 2017; Kleemans et al., 2018). We find that substitution to internal migration only partially overcomes the effect of a lack of international migration: Poverty rises despite an increase in internal migration in affected villages after natural disasters. One potential explanation lies in villages' heterogeneous dependency on international emigration. We show that villages that rely ex-ante most heavily on workers abroad are those who suffer the most from the combined effect of the ban and natural disasters. Splitting the sample into terciles of ex-ante dependency on international migration, we find that communities least dependent are the most resilient while those in the middle manage to overcome these shocks by substituting for domestic migration.

We identify two mechanisms underlying our results. First, as the emigration ban caused an unexpected availability of female unskilled individuals, this potentially had repercussions on local labor markets. Research has shown that these extra workers were absorbed by the large demand of workforce in rice fields (Makovec et al., 2018). Exploring this channel, we show that poverty increased most drastically in villages with economies geared towards rainfed rice production: These villages lose the capacity to absorb workers once struck by natural disasters, in particular extreme floods. Second, we identify remittances to play a crucial role as an adjustment strategy.

Indonesian workers in Saudi Arabia tended to remit more than those living in other countries before the ban was introduced. A simple back-of-the envelope calculation suggests an increase of 8–10% of households in poverty due to the lack of remittances.

Our results are in line with studies showing that migration and remittances reduce disaster-induced income shocks. In a scenario where international migration is regulated but still viable, affected individuals can decide to move abroad to cope with natural disasters. Households can diversify their climate-induced income shock risks through ex-ante migration decisions (Stark et al., 1991; Kleemans, 2015) or ex-post by means of remittances (Yang et al., 2007; Blumenstock et al., 2016; Gröger et al., 2016; Giannelli et al., 2022). Related to our paper, Mbaye et al. (2017) show that migration and remittances reduce poverty rates particularly in disaster-affected countries. Our study contributes to this literature by analyzing the role of international migration in a setting where it is heavily restricted. We reach similar conclusions, but from the opposite and therefore novel perspective: The drastic reduction of migration and remittances makes communities more vulnerable to natural disasters.

This paper expands the literature on the nexus between climatic events and international migration. The existing evidence is mixed: Some studies show a positive link (Gray et al., 2012; Backhaus et al., 2015; Coniglio et al., 2015; Drabo et al., 2015; Mahajan et al., 2020; Giannelli et al., 2022), others find there is no association or heterogeneous mobility responses (Beine et al., 2015; Cai et al., 2016; Cattaneo et al., 2016; Bazzi, 2017; Gröschl et al., 2017; Martínez Flores et al., 2021; Bertoli et al., 2022). In some contexts, migration even diminishes because of climatic shocks at the origin (Halliday, 2006; Yang et al., 2007). One challenge in addressing this question is to empirically establish a causal link: Although extreme climatic events are exogenous, omitted biased responses could be correlated with both migration and natural disasters. We overcome these concerns by exploiting a national policy shock affecting only international migration, introduced with the purpose of protecting Indonesian domestic workers abroad and thus plausibly uncorrelated with local village characteristics. In addition, the ban unilaterally affects one important destination country but leaves others unaffected, thereby creating a natural control group.

Finally, this paper relates to the few studies on the effect of restrictive migration policies on development outcomes at origin.[3] Theoharides (2020) exploits an immigration ban from Japan on Filipino migrants, finding that the policy decreased income in sending communities. More closely to our paper, Makovec et al. (2018) study the effect of the Saudi Arabia ban on labor market outcomes in Indonesia. The authors find no effect on unemployment, but rather a shift towards the agricultural and informal sector. While we exploit the same policy shock, our study combines it with natural disasters—an additional natural experiment—it uses more granular data and focuses on a notably different outcome. We reconcile their findings as one of the mechanisms behind the observed poverty increases.

The paper is organized as follows: Section 4.2 outlines the context of our study, followed by a description of data and empirical strategy in sections 4.3 and 4.4. Results, mechanisms and extensive robustness checks are discussed in section 4.5. Section 4.6 concludes.

4.2 The Indonesian context

4.2.1 Natural disasters

Indonesia is extremely prone to climate change-induced disasters and nature-borne risks across its entire territory. It ranks 38th worldwide in terms of natural disasters susceptibility in the World Risk Report (Aleksandrova et al., 2021). According to global disaster database EM-DAT, the most common mass disasters in Indonesia since 1999 have been floods, earthquakes, landslides and volcanic activity (Guha-Sapir et al., 2021). Climate change-induced disaster such as prolonged periods of drought or rain-induced inundations have been on the rise since 2000 (BNBP, 2020). The 2014 Indonesian village-census PODES indicates that more than 40% of villages were affected by at least one disaster event over the previous three years, illustrated in the Appendix Figure C.1.

Among others, climate change-related disasters such as floods, droughts and heat waves can adversely affect crop yields. In a recent overview, the IPCC outlines that large swaths of crop lands may become barren over the next

[3] Existing studies have focused on the effects of selective migration restrictions on human capital formation at origin (Beine et al., 2008; Chand et al., 2011; Gibson et al., 2011; Batista et al., 2012; Shrestha, 2017).

decades (IPCC, 2019). Indonesia is no exception: The costs incurred by climate change already amounted to 1.4% of GDP in 2016, the majority of which resulted from agricultural productivity losses (Hecht, 2016). Indonesians from rural areas are therefore increasingly looking for other income opportunities to cope with these challenges. One common coping strategy involves migrating away from rural areas. Studies show that one of the major migratory push factors is climate change, either directly or indirectly through the loss of livelihood.[4] For example, it is estimated that the 2004 Indian ocean tsunami alone left 500 thousand Indonesians internally displaced (Gray et al., 2014).

4.2.2 International migration

Migration both within and outside Indonesia's borders has always played a vital role in shaping the country's development. According to estimates of the World Bank (2017), around nine million individuals, corresponding to nearly 7% of the country's labor force, were employed abroad in 2016. Most legal migrants leave through formal migration intermediaries and stay abroad for two to three years (Bazzi et al., 2021a). Historically, the main destination countries of Indonesian migrants have been Malaysia, Saudi Arabia, Singapore and Hong Kong as depicted in the Appendix Figure C.2.

A key characteristic in Indonesia's emigration patterns is the strong heterogeneity in villages' migration networks with certain countries. These ties are deeply rooted in villages' ethnic composition and hence tend to be sticky over time (Bazzi, 2012). For example, overseas workers from villages with a greater share of households of ethnic Arab origin have a greater propensity to emigrate to Arab countries as compared to destinations in South-East Asia. Migration agencies provide region-specific information, skill training and financing for migrants, further strengthening migration ties (Spaan et al., 2018). For an individual's choice among migratory destination countries, these networks are known to play a major role.

Indonesia is one of the few countries in the world that exhibits a higher international migration rate of women as compared to men as displayed in the Appendix Figure C.2. The share of documented female emigrants increased from 56% in 1996 to 78% in 2004, a phenomenon generally attributed to a rapid increase in the demand for foreign female unskilled workers in the

[4]See for example Flavell et al. (2020) for a recent literature review or Bohra-Mishra et al. (2014) and Thiede et al. (2017) for an analysis in the Indonesian context.

Middle East (IOM, 2010). In Saudi Arabia, for example, 84% of Indonesian emigrants were women in 2005. In these countries, immigrants are mainly employed as domestic workers and therefore educational requirements are low (World Bank, 2017).

Around 72% of Indonesian emigrants come from rural areas (World Bank, 2017). These areas are also more vulnerable to agriculture-related income shocks that affect migration decisions. Many low-skilled and informal workers see international migration as an essential element of their livelihood strategy and an entry point to formal work: Emigration increases their probability of having a formal work contract upon return (World Bank, 2017). In addition, Indonesian women working abroad earn five times more on average than those who stay (Bazzi et al., 2021a). Migration can also positively affect the income of household members at home through remittances. Cuecuecha et al. (2016) find that Indonesian households receiving remittances exhibit lower levels of poverty compared to those without. The volume of remittances sent differs significantly depending on the destination country: Migrants living in Saudi Arabia tend to remit more than migrants in other destination, despite earning on less on average (Bank Indonesia, 2009).

4.2.3 The moratorium: Indonesia's emigration ban

With increasing numbers of domestic workers in the Middle East, the number of reported abuses and harassment of Indonesian women rose too. In June 2011, Ruyati Binti Sapubi, an Indonesian maid in Saudi Arabia killed her employer's wife after suffering from repeated abuse. For this reason, she was sentenced to death by beheading (The Washington Post, 2011). The event caused a public outcry in Indonesia and provoked the government to step in and issue a moratorium. Enacted in August 2011 and still in place today, the moratorium bans all women from emigrating to Saudi Arabia as domestic workers.[5] Appendix Figure C.3 displays how the ban reflects in Indonesia's emigration flows, comparing Saudi Arabia to other countries. While Saudi Arabia was the most important destination in 2005 accounting for 43% of all emigrants, its share decreased to 11% after the ban in 2014.

[5]Similar restrictions have been gradually introduced to other countries: To the United Arab Emirates and Qatar in 2013; to 21 countries mainly across the Middle East and North Africa in 2015; and to Kuwait and Jordan already in 2009–10, however both countries only play a minor role in Indonesia's emigration as shown in Appendix Figure C.2.

Issued at the national level, the moratorium affected all Indonesian women wanting to emigrate to Saudi Arabia. However, given village-level heterogeneities in ethnic composition and migration networks, villages were affected to highly varying degrees. These structural relationships also implied that immediate substitution to other countries as a response was unlikely, especially given the fact that the ban was gradually extended to similar destination countries. No other destination experienced a large increase in Indonesian immigrants after the ban as depicted in the Appendix Figure C.3. Switching to illegal emigration to Saudi Arabia was not an option either. The ban was strictly enforced and the sheer geographical distance between Indonesia and the Saudi peninsula prevents the vast majority of workers to emigrate undocumented (World Bank, 2017; Friebel et al., 2018).

4.3 Data

4.3.1 Indonesian village census

We compile a highly granular dataset of Indonesian villages including all urban and rural precincts from four waves of the administrative census PODES (*Potensi Desa*). It is collected every three to four years and includes information on village characteristics of the entire country. We use the PODES waves of 2005, 2008, 2011 and 2014.[6] PODES contains, among others, detailed information on the stock of international out-migrants disaggregated by gender, natural disasters and aggregate socio-economic variables.[7] Across all waves, 2005 was the first census-year that collected information on the stock of international emigrants per village. Furthermore, it is the only wave to provide information on the main migratory destination country by village, which we use to identify villages with strong migration networks to Saudi Arabia.[8] We also extract information on the occurrence of natural disasters, categorized by disaster type and exact timing over the course of the three previous

[6]The year a given round is published includes data corresponding to the previous year. PODES 2005, 2008 and 2011 therefore constitute pre-ban periods.

[7]The primary sources for the data were key informants within the village administration, with additional information and validation provided by officials at the sub-district and district levels (Bazzi, 2017). Appendix C.1.1 lists definitions and the precise wording of key variables used.

[8]A limitation is that PODES does not provide a clear definition for "main" destination. Our empirical specification will target potential endogenous misreporting or measurement errors.

years.[9] Further variables taken from PODES include village population, the incidence of social conflict, rural status and agricultural activities.

Our main outcome of interest is poverty. The village census reports the number of issued poverty letters (*Surat Keterangan Tidak Mampu* (SKTM)) in the previous year, a measure used by literature in the Indonesian context (Morgans et al., 2018; Krishna et al., 2021). *SKTM* are letters issued at the village level, stating that the individual is poor and therefore eligible for social assistance including access to free medical treatment, preference in scholarship requests and basic food assistance, among others (Fiarni et al., 2013). These letters have a validity of 6 months, but can be renewed upon request. Eligibility criteria are based on the absolute poverty definition of individuals falling behind the poverty line as established by the Indonesian Statistical Office (BPS), outlined in Appendix C.1.2. Given that letters are issued by the village administrators, the criteria are potentially porous due to different interpretations (Fiarni et al., 2013). Besides addressing this potential issue in our identification strategy, we provide direct evidence that poverty letters are a suitable poverty measure. Using representative household-level data, in Appendix C.1.3 we show that poverty letters are well targeted at the poorest households. The probability of receiving poverty letters is greater among households at bottom quintiles of consumption expenditure as well as among those that are below the poverty line. In addition, we use two alternative PODES-based poverty measurements derived from the census rounds: The number of households living in slums as well people receiving assistance for public health services.

4.3.2 Additional sources

We use weather station data from the Indonesian Meteorological, Climatological and Geophysical agency (BMKG) as an alternative measure on villages' past disaster experience. It provides information on stations' precise coordinates and the date of extreme weather events in terms of temperatures, precipitation and wind speeds recorded.

Alternative sources to PODES for nationwide, time-variant village-level poverty data are scarce. One exception are poverty maps compiled by the

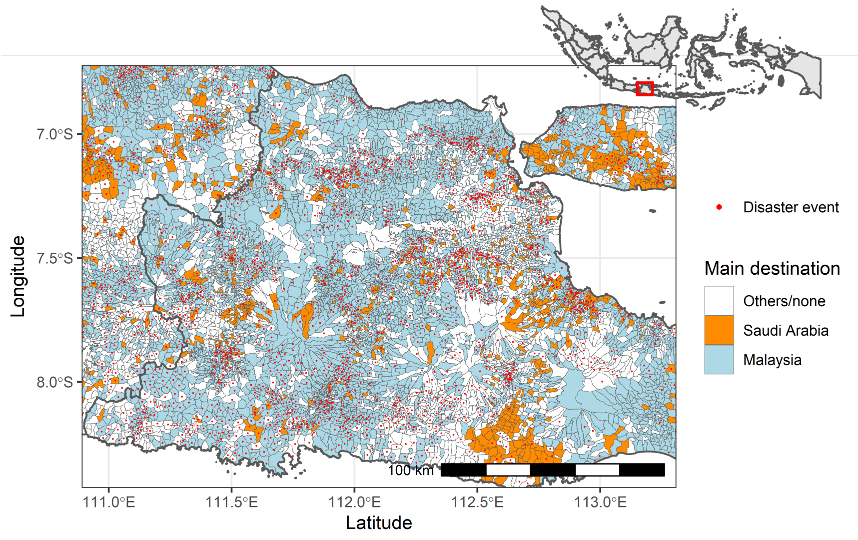
[9]Cameron et al. (2015) show that PODES correlates well with disaster records from other sources. We also validate our results using alternative measures of natural disasters in our robustness checks.

SMERU Research Institute, which are commonly used by research on Indonesia (Edwards et al., 2020). Their approach combines administrative statistics and household survey data from different sources to obtain poverty estimates at the village level for 2010 and 2015 (Suhayo et al., 2005). We can thereby verify our results with a further poverty measurement, the village-share of households below the poverty line.

4.3.3 Village panel

We combine all data sources at the village level, Indonesia's lowest administrative unit. To account for changing boundaries across time caused by administrative splits, we link different census years and aggregate all data to villages' 2005 boundaries, our unit of observation. The resulting dataset hence contains $N = 67,987$ villages over the years 2005, 2008, 2011 and 2014. Appendix Table C.1 provides basic summary statistics for all variables used in our analyses.

FIGURE 4.2: Main destination country and natural disasters in East Javanese villages



Note: Red dots represent centroids of villages that experienced at least one natural disaster between 2003 and 2005. Main destination refers to the country where most emigrants from a given village worked as of 2005. Data taken from PODES 2005.

Figure 4.2 maps villages in the province of West Java to illustrate the variation of our key variables. The majority of displayed villages had migration networks with Malaysia, followed by Saudi Arabia. At the same, many villages in this region experienced natural disasters in the displayed period. We explore the spatial variation of both natural disasters and migration links to different destination countries for our identification.

4.4 Empirical strategy

4.4.1 Natural experiment: The emigration ban

We investigate the effect of an international emigration restriction on communities' capacity to mitigate disaster-induced income shocks. To exploit the emigration ban as a natural experiment, we first analyse whether the ban significantly reduced the stock of international migrants in villages with migration ties to Saudi Arabia. Therefore we estimate an intention-to-treat (ITT) effect according to initial migration networks in 2005. The rationale is that villages with migration networks to Saudi Arabia in 2005 are more likely to experience a stronger reduction in the number of out-migrants after the ban was introduced in 2011. Therefore we estimate the following event-study model:

$$\begin{aligned} M_{vt} = \beta_1(T^{2005} \times SA_v) + \beta_2(T^{2008} \times SA_v) + \beta_3(T^{2014} \times SA_v) + \\ \lambda X_{vt} + \delta_t + \gamma_v + \eta_p t + \epsilon_{vt}, \end{aligned} \tag{4.1}$$

where M_{vt} measures the stock of emigrants from village v in year t. SA_v is a binary variable indicating whether Saudi Arabia is a village's main migratory destination country in 2005. T are year dummies. Coefficient β_1 to β_3 capture the yearly change of migration stocks in villages with Saudi Arabia as the main destination country in 2005 against all other villages. Compared to the excluded year 2011, we expect a decrease in migration stocks in 2014 for villages with migration ties to Saudi Arabia, i.e. a negative β_3.

X_{vt} controls for time-variant variables: Log of population and a binary variable for conflict in the previous year. Time fixed effects δ_t capture shocks common to the entire country. Village fixed effects (γ_v) control for time-invariant observable and unobservable characteristics such as soil suitability, propensity to be subject to natural disasters, cultural proximity with specific destination countries and established migration networks. Lastly, ($\eta_p t$) absorbs province-specific linear time trends. Standard errors are clustered at the village level.

4.4.2 Identification: The triple difference

An effective migration restriction policy implies that villages with strong Saudi Arabian migration networks experienced a larger reduction of out-migrants after 2011. These villages could therefore be more vulnerable to natural disaster-induced income shocks. To test this hypothesis, we run the following triple difference regression, our main specification of interest:

$$
\begin{aligned}
Poverty_{vt} = {} & \beta_1 D_{vt} + \beta_2 SA_v + \beta_3 Post2011_t + \\
& \beta_4 (D_{vt} \times SA_v) + \beta_5 (D_{vt} \times Post2011_t) + \beta_6 (SA_v \times Post2011_t) + \\
& \beta_7 (D_{vt} \times SA_v \times Post2011_t) + \lambda X_{vt} + \delta_t + \gamma_v + \eta_p t + \epsilon_{vt},
\end{aligned}
\tag{4.2}
$$

where $Poverty_{vt}$ stands for the number of new poverty letters issued in village v in year t = 2005, 2008, 2011 and 2014. D_{vt} is a binary variable for villages' disaster experience in the three years preceding t. $Post2011_t$ takes the value one if t = 2014, and zero otherwise. As time-variant controls, X_{vt} includes the inverse hyperbolic sine of the male emigrants stock, log of population and a binary variable for conflict events. Again, we add fixed effects for year (δ_t) and village (γ_v) on top of province linear-time trends ($\eta_p t$). Robust standard errors are clustered at the village level.

We include interactions of all three binary variables analogous to standard double difference models. $D_{vt} \times SA_v$ controls for time-invariant heterogeneous responses to disasters in villages with Saudi Arabia as the main migratory destination country. $D_{vt} \times Post2011_t$ captures natural disaster trends that could spuriously affect the dependent variable after the ban. The interaction $SA_v \times Post2011_t$ is essential to control for all observable and unobservable factors influenced by the moratorium that could affect poverty, other than being exposed to disasters. For example, it includes direct wealth shocks due to foregone remittances and expected income from migrating as well as common changes in population compositions due to altered migration patterns. The interaction also captures differential labor market responses as identified by Makovec et al. (2018): The increase in local labor supply by those no longer able to migrate could push wages down and potentially increase poverty. Finally, including the stock of international male migrants as control variable is crucial. Some villages could substitute the outflow of female domestic workers to Saudi Arabia with male emigration.

4.4.3 Causal interpretation

Our identification derives from the triple interaction $D_{vt} \times SA_v \times Post2011_t$. This term allows us to causally estimate the effect of natural disasters on poverty in villages that could no longer rely on international migration to Saudi Arabia as an adaptation strategy.

One potential threat to the identification stems from any potential anticipation effects of the ban. Would-be migrants could either anticipate the departure to Saudi Arabia or simply refrain from emigrating. At the same time, village heads could ex-ante issue more poverty letters to cope with the foregone income from diminishing remittances. These scenarios assume that village heads and individuals possessed prior information on the national government's move to implement a ban. Even if this was true, the interaction term $SA_v \times Post2011_t$ controls for this bias that would be common to all villages with ties to Saudi Arabia. The only residual variation in the dependent variable derives from natural disasters, quasi-random events once geographic factors are controlled for by village fixed effects.

Village authorities could still over-report disaster events and issue more poverty letters to receive greater government transfers. To upward bias our results, this would need to systematically happen in villages with links to Saudi Arabia hit by disasters after 2011. To rule out this hypothetical scenario, we adopt three strategies. First, we show that the main effect is robust to controlling for the inflow of different transfer types from local and central governments as well as foreign and private citizen aid. Secondly, we use two alternative PODES-based variables to proxy poverty: (i) The number of social health insurance cards (*Askeskin*) issued in year *t-1*, which Sparrow et al. (2013) find to be well targeted to the poorest and most vulnerable individuals; and (ii) the number of households living in slums.[10] Lastly, we use poverty data external to PODES from SMERU, measuring the village-share of individuals below the poverty-line. Self-reporting can analogously affect our measurement of natural disasters, hence we also use alternative data provided by BMKG. Results for all alternative data sources are discussed in section 4.5.4.

Although we argue that the moratorium date is unexpected, villages with migration links to Saudi Arabia could follow different pre-trends in poverty

[10] In 2016, 29 millions Indonesians lived in slums with limited access to basic services like sanitation and safe water (World Bank, 2016).

rates. While in theory this should not be an issue for villages hit by exogenous disasters, it could potentially violate the parallel trend assumptions for villages vulnerable to the ban. We provide full evidence on pre-treatment parallel trends in a triple difference context in section 4.5.3.

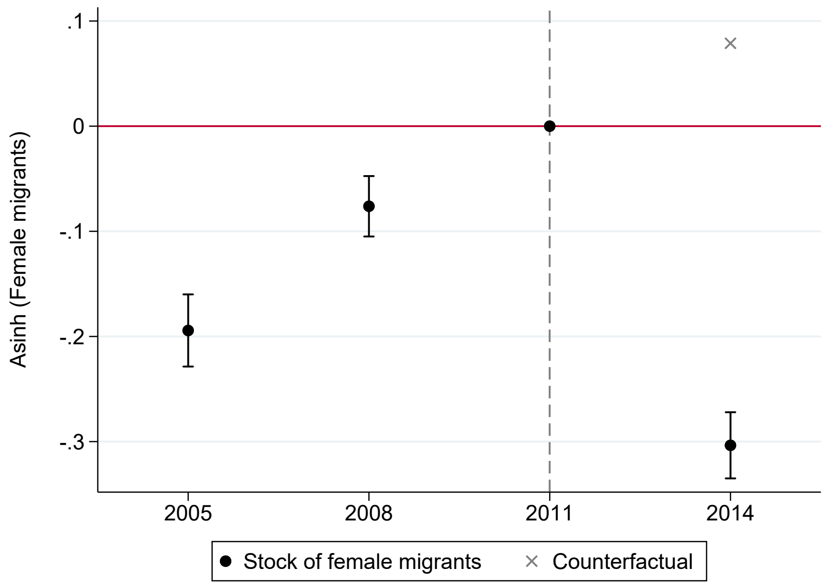
4.5 Results

4.5.1 Migration flows after the ban

We exploit the emigration ban on Saudi Arabia to causally estimate how villages are able to cope with disaster-induced income shocks in a scenario where international migration opportunities are curtailed. Our analysis relies on the assumption that the ban was effective, therefore we first quantify its impact on migrant stocks. Figure 4.3 plots the coefficients from the event-study in equation (4.1), indicating that the ban's impact on mobility is considerable. With respect to 2011, the emigrant stock drops by 30% in villages with Saudi Arabia as main migratory destination in 2005. Figure 4.3 also shows that the female migrant stock in these villages catches up over time until 2011. In a counterfactual scenario, the stock of female migrants in Saudi Arabia would have increased further had it followed the same linear growth rate of the period 2005–2011. This scenario suggests that the stock of female migrants would have been larger than that of the comparison group in 2014. However, estimates display a sharp u-turn: The drop in female migrant stocks even amounts to 38% with respect to this scenario indicated by the grey symbol x in 2014. The negative estimates for the years leading up to 2011 reflect the growing importance of Saudi Arabia as destination country, which, we argue, would have further intensified had the ban not been implemented. These results comfort our premise that using Saudi Arabian migration ties in 2005 successfully singles out villages most affected by the moratorium.

While the coefficients in Figure 4.3 do not follow parallel trends, there are two reasons why this does not threaten our identification. First, a bias would arise if the stock of female migrants was already decreasing before 2011, not increasing as in this context. And secondly, the non-parallel trend does not affect the causal interpretation of our results: The DDD estimator does not require two but only one parallel trend assumption. We will show that the only parallel trends assumption required holds in our baseline model further below.

FIGURE 4.3: The effect of the moratorium: Change in female migrant stocks in villages with Saudi Arabia as main destination



Note: Displayed coefficients capture the event-study in equation (4.1), i.e. the relative decrease in the inverse hyperbolic sine of female migrants' stocks for villages with Saudi Arabia as main destination country in 2005 vs. others, with 95% confidence intervals. The vertical dotted line indicates the implementation of the ban in 2011, which is also the baseline period. "x" indicates the value of female stocks in a counterfactual scenario where it follows the linear trend from 2005–2011. The sample is restricted to villages that indicate they have at least one Indonesian domestic worker abroad in 2005. Control variables include log(population), a conflict event binary indicator, village and year fixed effects and province-time trends. Standard errors are clustered at the village level.

4.5.2 Disasters and migration under the ban

A strong negative effect of the moratorium on migrant stocks could leave villages dependent on migration to Saudi Arabia more vulnerable to climatic shocks. This is where we introduce our main specification, with Table 4.1 presenting the baseline results. Columns 1–3 display the estimations of simple difference-in-difference (DD) models, where each double interaction is shown in a separate regression. Out of the three two-way interactions, only the double difference coefficient $SA \times Post2011$ in column 1 is statistically significant. This means that villages with migration ties to Saudi Arabia experience higher levels of poverty than all others after 2011. Potential explanations are a deterioration in labor markets or an overall decrease in remittances. Across all columns, the coefficient on disasters is statistically significant, implying that the number of poverty letters increased by around 9% in villages hit by natural disasters.

TABLE 4.1: Average effect of disasters on poverty

Dependent	Poverty cards			
	DD			DDD
	(1)	(2)	(3)	(4)
Disaster		0.085***	0.086***	0.093***
		(0.008)	(0.008)	(0.009)
SA × Post2011	0.062***			0.015
	(0.020)			(0.027)
Post2011 × Disaster		-0.011		-0.028*
		(0.014)		(0.015)
SA × Disaster			-0.025	-0.055**
			(0.020)	(0.023)
SA × Post2011 × Disaster				0.118***
				(0.039)
Village FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Province-time trend	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	268,194	268,194	268,194	268,194

Note: Poverty letters is transformed by the inverse asymptotic sine (asinh). Control variables include asinh(male migrants), log(population) and a conflict event binary indicator. Robust standard errors are clustered at the village level and reported in parentheses. */**/*** denote significance levels at 10/5/1% respectively.

Column 4 shows results of the DDD model. The interaction term $SA \times Disaster$ is significant and negative. It implies that villages with Saudi Arabian migration networks tend to cope better with natural disasters. Potentially, this is due to the fact that migrants in Saudi Arabia remit more on average than those working in other destination countries (Bank Indonesia, 2009). Estimates of the interaction $Post2011 \times Disaster$ suggest that all villages tend to cope better with disasters in 2014 than before. One conceivable explanation is that disaster prevention systems have improved over time. Lastly, the interaction $SA \times Post2011$ is no longer significant, hence the triple interaction almost entirely explains its coefficient from column 1. The triple interaction is positive and significant, with a coefficient of 0.118.[11] To interpret this effect, we compute marginal effects: Villages with migration links to Saudi Arabia hit by natural disasters experience a poverty increase by 13% after the ban was introduced in 2011. Results are also qualitatively similar when we consider the number of natural disasters experienced by villages as reported in the Appendix Table C.3.[12]

[11] This coefficient is robust to the choice of different control variables as displayed in the Appendix Table C.2.

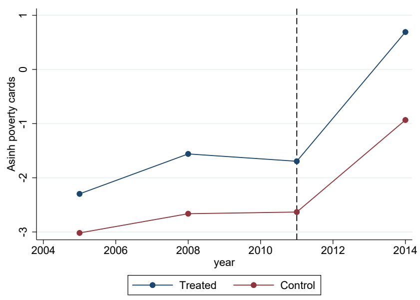
[12] This sample does not contain the census wave 2005 for lack of data on the number of disasters per village. Results for the extensive margin are robust to restricting the sample to 2008–2014 census waves (not shown).

4.5.3 Parallel trends

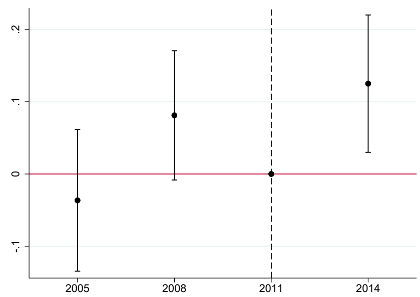
As a prerequisite for the causal interpretation of our findings, we show that the pre-ban parallel trend assumption holds. The DDD estimator requires only one parallel trend assumption instead of two, because common biases between the treatment and control group are partialled out by a first difference (Olden et al., 2022). In our context, the treatment group consists of villages with migration ties to Saudi Arabia and the control group of villages with migration ties to other countries and those without migrants in 2005.

To demonstrate that the pre-ban parallel trend assumption holds, we follow three steps: (i) Within the treatment group, we subtract poverty levels of villages struck by disasters from those without disasters, before 2011; (ii) we perform the same step for villages in the control group; (iii) we show that the two differences obtained follow the same trend before the ban. Panel (A) in Figure 4.4 shows that villages in the treatment and control group trended similarly in the time leading up to the ban in 2011. After 2011, both groups experienced an increase in the average number of newly issued poverty letters. However, this rise is larger for the treated group, in line with our baseline results. In panel (B), we present our baseline estimations in an event-study, highlighting the absence of pre-treatment trends before the ban was implemented in 2011.

FIGURE 4.4: Triple difference: Parallel trends



(A) Unconditional parallel trends



(B) Conditional parallel trends: Event study

Note: Panel (A) displays evidence on the unconditional parallel trend assumption. The treated group consists of villages with Saudi Arabia as main destination country, subtracted by the effect of being hit by natural disasters. The control group consists of villages that do not have Saudi Arabia as main destination, subtracted by the effect of being hit by natural disasters. Panel (B) shows an event study, plotting the triple difference coefficients from equation (4.2) with year 2011 as the baseline year and with 95% confidence intervals. Control variables include asinh(male migrants), log(population), a conflict event binary indicator, village and time fixed effects as well as province-time trends. Standard errors are clustered at the village level.

4.5.4 Robustness checks

Alternative measurements

Poverty estimates can be porous, particularly when relying on self-reported data. We demonstrate that our results are not measurement-specific and that the observed increases reflect in different poverty dimensions in the Appendix Table C.4. Columns 1–2 present results for our main model, using the number of issued social health cards and households living in slums as dependent variables. Albeit qualitatively different to SKTM letters, both measurements reflect dimensions of poverty experienced in villages. Coefficients of the triple interaction are positive and significant for both, indicating that our results are not measurement-specific.

This is further confirmed in column 3, where poverty is measured as the the share of individuals falling below the poverty line according to the international convention of USD 2 PPP.[13] Villages with Saudi Arabian migration links experience a 1.19 percentage point poverty increase after the migration ban once hit by natural disasters. Compared to pre-ban poverty levels, this amounts to a 6% increase. In terms of magnitude, the different estimate as compared to the main specification can be explained by measurement: Poverty letters measure absolute increments, whereas poverty shares reflect relative poverty rates.[14]

Sub-samples, placebos and falsification tests

Our identification relies on the exogenous timing of natural disasters and that of the migration ban. The former are quasi-random events, conditional on village fixed effects. We confirm this by performing a falsification test in the Appendix Table C.5, where we regress the lagged natural disaster binary indicator taken from the previous period on poverty in year t. Results show that natural disasters which occurred three to six years earlier do not have a significant effect on poverty in the following period.

To show that results are not driven by sample choice or outliers, we present our main estimates for different sub-samples in the Appendix Table C.6. In column 1, we exclude all villages on Java island as the most disaster-prone, populated and emigration-intensive region. In turn, column 2 contains only

[13]Given the data availability described in section 4.3, this measurement is only available for 2010 and 2015. We match these waves to PODES 2011 and 2014.

[14]A direct comparison is not viable given the lack of village poverty letter stock data in PODES.

villages on Java. For either analysis, the magnitude of the triple interaction is larger than our main effect and statistically significant. Next, we exclude villages without migrants in 2005 in column 3. This way, the binary variable SA compares villages with Saudi Arabia as main destination only to those with a positive number of emigrants in 2005. In column 4 we investigate the possibility that although the moratorium was only progressively extended to other important Middle Eastern destination countries, they were already indirectly affected right after 2011. This could be due to a more general negative sentiment towards Arab states provoked by the 2011 ban or the events leading up to it. For either specification, results are virtually unchanged as compared to our main estimates. In column 5 we show that the results are robust to excluding population outliers, trimming the sample at the 1^{st} and 99^{th} percentile of village population. Lastly, in column 6 we weight the estimation by population, showing that the coefficient of the triple interaction is slightly larger and still significant at the 1% level.

We further demonstrate that other time-variant, unobserved changes at the village-level do not alter the main results. In our main specification, SA_v is a binary variable taking the value one if Saudi Arabia is a village's main destination country in 2005, and zero otherwise. As a placebo test, we replace this binary indicator with a categorical variable for all twelve destination countries recorded in PODES. Appendix Figure C.4 displays the triple interaction coefficients for this specification where the base category is villages without migrants in 2005. The only destination country displaying a positive and significant coefficient is Saudi Arabia, reassuring that the ban constitutes the main treatment. Only villages with migration networks to this country experience greater poverty once hit by natural disasters after 2011.[15]

Different types of welfare transfers could be used as substitutes for poverty letters. If these payments ameliorated actual poverty experienced in a village, our estimates based on the issued letters would be upward biased. Therefore in the Appendix Table C.7 we include local and central government transfers as well as foreign and private citizen's aid as controls.[16] The triple difference coefficient remains unaltered across all columns, suggesting that the larger number of issued poverty letters for ban-affected villages hit by natural disasters after 2011 is not influenced by differential financial inflows.

[15]These coefficients are qualitatively comparable when excluding villages with no emigrants in 2005 from the sample or using any other destination country as base category (not shown).

[16]This sample does not contain census wave 2005 for lack of government transfers data.

Spillovers

The effects of natural disasters and the migration ban could spill over to neighboring villages, potentially questioning the stable unit treatment value assumption (SUTVA). For example, natural disasters could push individuals to seek jobs in villages nearby without strong Saudi Arabian migratory ties. If the emigration of these individuals had detrimental effects on the economy in these destinations, our estimates would be biased downward. On the contrary, upward biased estimations would arise if those emigrants fueled economic development in neighboring villages.

We address these potential biases directly by controlling for spillover effects and including spatial standard errors. For villages without Saudi Arabian migration networks, we calculate the distance to their closest neighbor with these ties. Based on that distance, we assign three binary variables for cutoffs of 0–10km, 10–20km and 20–30km. In our baseline regression, we then replace the indicator for Saudi Arabian migration ties with each binary variable in separate regressions. This allows us to analyze whether villages without migration ties to Saudi Arabia experience differential poverty rates depending on their distance to ban-affected villages. Appendix Table C.8 shows our main effect does not change with the inclusion of these variables. Interactions for villages distant up to 30km from our treatment villages (SA=1) are insignificant as well, pointing to the absence of spillovers within this radius. Furthermore, we account for Conley-type spatial correlations of the error term (Conley, 1999) in the Appendix Table C.9, where results remain statistically significant across different distance cutoffs.

Substitution to domestic migration or to other countries

The moratorium could push individuals to substitute Saudi Arabia with other countries or internal migration as alternative coping strategies. In practice, different educational requirements limit short-run migratory substitution options to other countries. Saudi Arabia demands foreign domestic workers to have primary education, whereas other important destinations such as Taiwan, Hong Kong and South Korea require that workers have completed at least secondary education. Furthermore, strong kinship migration networks impede easy substitution to other destinations. Descriptively, this can be seen in the Appendix Figure C.3, showing that emigration to other destination countries did not increase significantly after then ban.

Choosing to migrate internally is therefore a more viable option for individuals affected by the ban. This would pose a threat to our identification only in case the ban itself affected selection into domestic migration. If the majority of those able to afford internal migration moved in the aftermath of natural disasters, the composition of stayers would be skewed towards poorer individuals. This, in turn, could bias the coefficient upward because of potential general equilibrium effects towards a deteriorating economy in those villages. However, our measure of issued poverty letters captures the change in the absolute number of poverty letters emitted, or the "new poor" households, partially overcoming changes in composition. In the Appendix Table C.10 we still test for any potential bias from substitution to internal migration. We proxy internal migration as the change in population, given the lack of domestic migration data in PODES. Column 1 shows that that the overall population of villages affected by the ban drops by 1.3%. The coefficient thus suggests potential substitution from international to internal migration. However, once we further interact the DDD coefficient with the change in population, we do not find differential effects on poverty as shown in column 2. It implies that villages in the treated group with higher levels of domestic out-migration do not show differential poverty rates as compared to the control group.[17]

In absence of compositional changes, substitution to any alternative coping strategies would only lead to a downward bias of our main DDD coefficient. More specifically, our main results in the DDD model imply that natural disasters increase poverty by 13% in villages with strong migratory ties to Saudi Arabia. If those who would have emigrated there moved elsewhere, the effect of disasters on poverty would only be reduced. This implies that the 13% effect from the baseline estimation is a lower bound estimate.

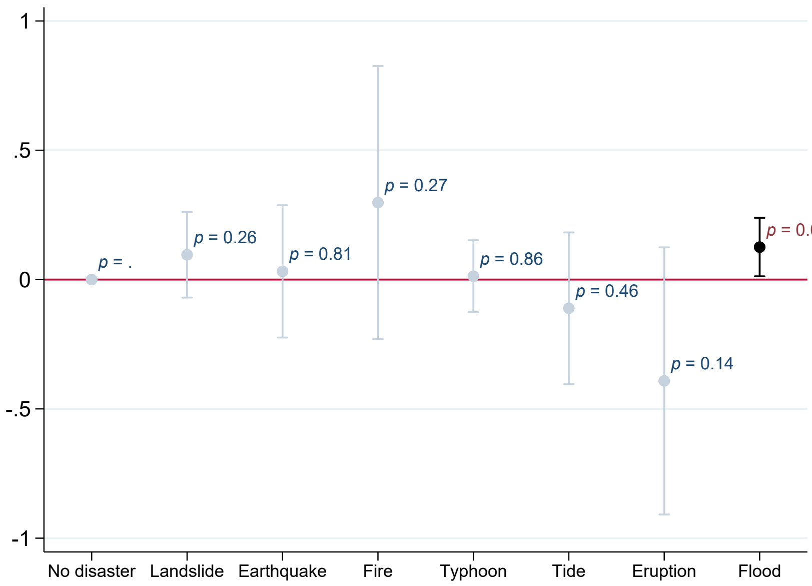
Rainfall, floods and Indonesian weather stations

We investigate which type of natural disaster drives our results. Figure 4.5 displays coefficients of the triple interaction where D now represents a categorical variable for different natural disaster types. Floods, in particular, lead to a significantly greater number of issued poverty letters as compared to the base category of not experiencing any natural disaster. This is in line with

[17]The DDD coefficients in column 2 is insignificant. However, the overall marginal effect of natural disasters after 2011 on villages with Saudi Arabia as main destination is 12.8% and significant (not shown).

research showing that floods are also one of the most devastating types of natural disasters in terms of losses and harvest failure (FAO, 2018).

FIGURE 4.5: Average effect of disasters on poverty by type of disaster



Note: The displayed coefficients capture the effect of natural disasters on the inverse hyperbolic sine (asinh) of emitted poverty letters after the migration ban in 2011 by type of disaster with 95% confidence intervals. The baseline type is "no disaster". We exclude the category "Tsunami" since we only have one observation in the control group after 2011. Control variables include asinh(male migrants), log(population) and a conflict event binary indicator. Village and year fixed effects as well as province-time trends are included. Standard errors are clustered at the village level. The sample includes census rounds 2008, 2011 and 2014.

Having established that our results are driven by heavy rain-caused events, we rely on alternative disaster definitions from Indonesian weather station data provided by BMKG to verify our results. This also allows us to address potential concerns related to reporting bias in PODES-recorded events. Rainfall data is collected from each of the 170 geocoded weather stations that operated uninterruptedly between 1990 and 2015. Extreme rainfall events are defined as the day each station recorded the largest precipitation over the ten previous years. The value of one is assigned if extreme rainfall events occurred either between 2003–2005, 2005–2006, 2009–2011 and/or 2012–2014, and zero otherwise.[18]

[18] For example, if a given weather station records the day with the largest rainfall between 1995 and 2005 between 2003 and 2005, then the binary variables takes the value one in t=2005. This is repeated for the period 1998–2008, where the variable is one again if the extreme rainfall was recorded for any day in 2005–2008.

Relying on precise coordinates of the stations, we use different bandwidths of either 10, 15, 20 or 30 km to assign villages to their corresponding precipitation records. Appendix Figure C.5 displays the location of the weather stations and respective buffer zones used in our analysis.[19] Across all specifications in the Appendix Table C.11, the effect of the triple interaction on poverty letters is significant and ranges between 25.3% and 62.9%, depending on the chosen radius. Compared to our main results, these larger effect sizes could be explained by the fact that with weather station data we identify particularly extreme events for a subset of villages in our sample.

4.5.5 Mechanisms and discussion

Labor market adjustments

Emigration diminishes the workforce that is available in sending communities, thereby driving wages up (Amuedo-Dorantes et al., 2006; Aydemir et al., 2007; Hanson, 2007; Mishra, 2007; Elsner, 2013). Conversely, if would-be migrants can no longer move abroad, this increases the local labor supply and could negatively affect wages, which could in turn increase poverty. Makovec et al. (2018) show that Indonesia's migration ban did not affect unemployment, but increased employment in the agricultural sector among those that could no longer migrate. More than half of all emigrants had already been working in this sector prior to moving abroad (Bank Indonesia, 2009). In Indonesia, agriculture is dominated by rice, the most cultivated and consumed staple, yet particularly vulnerable crop to weather shocks. A significant shift of workers into agriculture might therefore leave rice-producing villages even more prone to poverty once hit by disasters. When disasters such as floods damage crop production, they can limit the capacity of local labor market to absorb the excessive workforce of would-be migrants through jobs in the fields.

Rice production is located in rural areas of the country, where irrigation of fields is either rainfed or relies on man-made schemes (Khairulbahri, 2021). Indonesian rice farmers consider floods to be the greatest threat to their production (Rondhi et al., 2019), which particularly holds for more vulnerable rainfed irrigated areas (Panda et al., 2021). This is in line with our results that floods are the most impactful type of natural disaster on poverty. Therefore,

[19]For each radius, only villages within the respective buffer are selected, where the choice of the buffer size implies a trade-off between the precision of local weather measurement and the number of villages included in the analysis.

we explore the identified labor market adjustment mechanism by the type of rice-irrigation villages rely on. Based on PODES data, we estimate a quadruple difference model by interacting all binary variables in equation (4.2) with a variable taking the value one if a given village mainly cultivates rainfed paddy (15% of sample villages) and zero if it has an irrigation system (85%).

TABLE 4.2: Average effect of disasters on poverty by the type of rice production

Dependent	Poverty cards		
	All disasters	Floods	Other disasters
	(1)	(2)	(3)
Disaster	0.102***		
	(0.011)		
SA × Post2011 × Disaster	0.081		
	(0.061)		
SA × Post2011 × Disaster × Lowlands	0.260**		
	(0.125)		
Flood		0.142***	
		(0.017)	
SA × Post2011 × Flood		0.061	
		(0.078)	
SA × Post2011 × Flood × Lowlands		0.423***	
		(0.149)	
Other disaster			0.077***
			(0.014)
SA × Post2011 × Other disaster			0.099
			(0.081)
SA × Post2011 × Other disaster × Lowlands			0.056
			(0.170)
Village FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Province-time trend	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	183,337	145,256	147,415

Note: Poverty letters is transformed by the inverse asymptotic sine (asinh). Control variables include log(population), asinh(male migrants) and a conflict event binary indicator. All two-way and three-way interaction terms are included in the estimation but omitted here. "Flood" and "Other disasters" take the value one in case of a flood or any other disaster than flood occurred within the three previous years, and zero with no disasters. Robust standard errors are clustered at the village level and reported in parentheses. */**/*** denote significance levels at 10/5/1% respectively.

Table 4.2 displays the results of this regression for the sample of rural rice-cultivating villages (68% of sample villages). The binary variable on disasters in column 1 includes seven types of disasters, whereas columns 2–3 are coded to capture either floods or any other disaster. In all columns, the reference group consists of villages without natural disasters. Across all columns, the control group consists of villages that did not experience any disaster in year t. The impact of any natural disaster on the treated group after 2011 indicates 26% more emitted poverty letters in villages with rainfed lowlands as compared to those with irrigated areas (column 1). When restricting the analysis to flood-disasters in column 2, we find that villages relying on rain-fed irrigation receive 42.3% more poverty letters than villages with irrigated fields. On the contrary, column 3 shows that this heterogeneity does not prevail if these villages are hit by any another type of natural disaster. We can thereby reconcile and extend Makovec et al. (2018) results: Migrants confined to stay in agriculture can no longer adjust their labor market decisions towards agriculture if floods reduces crop yields.

Remittances

Sending money to support families at home is one of the key incentives to emigrate. An emigration ban reduces the remittances received, which in turn influences how receiving communities can smooth income shocks from natural disasters. There are two reasons why we believe this holds particularly for the setting of our study in Indonesia. First, annual remittance inflows amounted to USD nine billion in 2016, corresponding to 1% of national GDP (World Bank, 2017) and strongly affecting local development (Bal et al., 2020). With Saudi Arabia being one of the main destination countries, the observed 30% decrease in migrant stocks in villages affected by the ban strongly reduced the number of potential remitters. Secondly, Indonesian workers living in Saudi Arabia remitted more on average than migrants in any other main destination country before the ban (Bank Indonesia, 2009). This implies even stronger effects of the moratorium for villages relying on these inflows.

Data on remittances is scarce and, to the best of our knowledge, no nationally representative survey on migration and remittances was collected before and after the moratorium. Given this data constraint, we provide a simple back-of-the envelope calculation combining our results with additional sources. First, in Figure 4.3 we showed that the stock of female migrants in villages with links to Saudi Arabia drops by 30.4% in 2014 compared to 2011, or 37.9% in the counterfactual scenario without the ban. Secondly, 95% of Indonesians working abroad transfer money home at least once within the first year of departure (Bank Indonesia, 2009). Lastly, Indonesian households receiving remittances from abroad have a 27.8% lower probability of being poor than those not receiving any (Cuecuecha et al., 2016). Assuming that the share of emigrants remitting stays constant at 95%, it implies that after 2011, potentially around $0.95 \times 0.304 \times 0.278 = 8\%$ more households are in poverty because of the lack of remittances sent by emigrants abroad.

With respect to the counterfactual scenario, the estimate is slightly larger at $0.95 \times 0.379 \times 0.278 = 10\%$.[20]

Dependency on international and internal migration

Income shocks can lead to substitution from international and domestic migration. This effect is however heterogeneous and determined by the degree of communities' dependency on international migration (Gröger, 2021). We investigate these heterogeneous substitution dynamics by splitting the sample into terciles of initial international emigration rates. This way, we capture communities' historical propensity to rely on work overseas.[21] We first test if there is a heterogeneous effect of natural disasters on poverty after the ban by each tercile of initial international emigration rate. Secondly, we investigate the potential substitution to internal migration for each sub-sample. Coefficients in Table 4.3 point towards heterogeneous effects of the triple difference: Villages that have historically relied more on international migration, i.e. with a relatively higher pre-ban international emigration rate, are those most affected by natural disasters after 2011. The decrease in the population of stayers (i.e. an increase in out-migration to other villages) appears to be driven by villages in the second tercile of ex-ante emigration rate.

The sample in column 1–2 consists of villages that relied on international migration the least. Neither do these villages show significantly different levels of poverty (column 1), nor different levels of internal migration (column 2) once hit by natural disasters after 2011. In this sample, urban villages constitute an above-average share: 33% are urban, well above the sample mean of 18%. As urban precincts, these villages are potentially more resilient to climatic shocks and therefore less reliant on internal and international migration as a coping strategy. Villages in the middle tercile are those that experience the largest rise in internal out-migration and no significant changes in poverty (columns 3–4). The latter might be due to households not overshooting investment into international migration compared to households from the third tercile.

Our results indicate that communities more reliant on international migration might have over-invested in a riskier adaptation strategy with regards

[20]Because this only accounts for the drop of female migrant stocks, the estimates can be considered conservative.

[21]Initial international emigration rate is defined as the the stock of international emigrants divided by the population, averaged for the years leading up to the ban in 2011 (2005 and 2008).

to income shocks from natural disasters. Being over-dependent on international migration can make it more difficult to switch to alternatives such as moving elsewhere in Indonesia, potentially explaining the results in columns 5–6. Furthermore, nine out of ten villages in the sample are rural and therefore more dependent on agriculture, more likely to send international migrants (Bazzi, 2017) and more vulnerable to disaster-induced income shocks.

TABLE 4.3: Average effect of disasters on poverty or internal migration by terciles of initial international emigration rate

	Low initial ER		Middle initial ER		High initial ER	
Dependent:	(1) Poverty cards	(2) Internal migrants	(3) Poverty cards	(4) Internal migrants	(5) Poverty cards	(6) Internal migrants
SA $\times$ Post2011 $\times$ Disaster	0.091 (0.072)	0.000 (0.011)	0.106 (0.076)	-0.023** (0.009)	0.113* (0.066)	-0.003 (0.009)
Village FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Province-time trend	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	56,884	56,884	56,884	56,884	56,884	56,884
Share of rural villages	0.67	0.67	0.81	0.81	0.89	0.89

Note: Initial international emigration rate (ER) is defined as the the stock of international emigrants divided by population, averaged for 2005 and 2008. We exclude villages with zero stock of emigrants in years 2005 and 2008. The dependent variable is the inverse hyperbolic sine (asinh) in columns 1, 3 and 5; log(population-international stock of migrants) in columns 2, 4 and 6. Control variables include asinh(male migrants), log(population) and a conflict event binary indicator in columns 1, 3 and 5; and only conflict in columns 2, 4 and 6. All further interactions are included in the estimation but not displayed here. Robust standard errors are clustered at the village level and reported in parentheses. */**/*** denote significance levels at 10/5/1% respectively.

4.6 Conclusion

We investigate whether international migration restrictions affect the capacity of villages to absorb income shocks induced by natural disasters. Indonesia—a country with long emigration history and prone to weather shocks—abruptly implemented a ban preventing all women from emigrating to Saudi Arabia as domestic workers in 2011. Exploiting this large-scale natural experiment in a triple difference analysis, we show that villages whose migratory opportunities were curtailed experienced a 13% greater increase in poverty in the aftermath of disasters.

To the best of our knowledge, we are the first to causally quantify the unintended consequences of migratory restrictions in the context of natural disasters. Our results suggest that the aim of the Indonesian government to protect citizens overseas by inhibiting emigration came at a cost for Indonesians confined to stay. The burden of this policy was particularly high for areas relying on rainfed irrigation for rice production, a sector that absorbed many would-be international emigrants after the ban. We identify floods as the most consequential disaster type, particularly when hitting these agriculture-intensive villages. This finding points towards important heterogeneities in how villages come to adapt to the ban due to their economic structure and thereby shed light on an important mechanism other than remittances.

Our results are particularly relevant in light of two of the most pressing issues worldwide: The increasing frequency of climate-induced disasters and current political debates to create restrictive barriers to migration. In this respect, we extend findings highlighting the vast gains from reducing barriers to migration (Clemens, 2011; Bryan et al., 2019), yet from the opposite perspective. We show that suppressing international migration curbs one major adaptation strategy to natural disasters. With a rise in restrictive migration policies against the backdrop of sharpening climatic changes, this scenario has the potential to further put livelihoods in affected communities around the world under pressure.

While the setting of our study examines the less frequent case of restrictions implemented by the country of origin, we believe that the effects on households relying on migration would be similar for policies enacted by destination countries. More specifically, implications for policy makers in both countries of origin and destination can be derived from this setting: Migration restrictions can put further pressure on communities already affected by climate change, particularly when the opportunities to substitute are limited. In light of current projections on the number of climate-induced migrants going into hundred of millions (Cattaneo et al., 2019b), decision makers need to carefully take climatic changes into account when designing migration-related policies.

Chapter 5

Cash Transfers and Violent Crime in Indonesian Communities

Elías Cisneros, Krisztina Kis-Katos, Jan Priebe and Lennart Reiners[1]

Abstract

This study investigates the impact of Indonesia's flagship conditional cash transfer (CCT) program—PKH—on violent crime. Exploiting data from a randomized controlled trial and administrative data from the staggered nationwide program roll-out in combination with different causal identification strategies, we show that communities receiving access to the CCT experienced an increase in violent crime. Examining possible mechanisms, our analysis reveals that the program resulted in an increase in idleness among non-targeted young men within beneficiary households, which we believe contributed to the rise in violent crime. In contrast, we show that the surge in violent crime is neither related to PKH increasing the (monetary and non-monetary) rewards for committing crime nor to alternative reductions in the (material, psychic, punishment-related) costs of engaging in crimes.

[1]The study has been published as working paper on SSRN, available *here*.

5.1 Introduction

Public welfare programs are expensive. To justify their costs, research has increasingly begun to study whether these programs can create desirable externalities on other outcomes that are not directly targeted by the program. In this context, a growing body of research has examined whether welfare programs can reduce crime rates, with studies investigating the impact of housing programs (Ludwig et al., 2001; Jacob et al., 2014; Chin, 2018), youth employment schemes (Gelber et al., 2015), social assistance (Bratsberg et al., 2019; Deshpande et al., 2022), food programs (Foley, 2011; Carr et al., 2019), universal income schemes (Watson et al., 2020), and conditional cash transfer programs (Camacho et al., 2013; Chioda et al., 2016; Machado et al., 2018; Borraz et al., 2020; Attanasio et al., 2021). Overall, the empirical evidence on the impact of such welfare programs on crime is mixed.

Despite its policy relevance, there is a notable gap in understanding the mechanisms that connect welfare programs and crime. As most existing studies relate spatially aggregated data on the dynamics of crime and program beneficiaries, mechanisms are typically inferred by interpreting differences in the timing and types of crime. A lack of micro data on beneficiary households often limits the understanding of underlying behavioral channels outlined in prominent economic theories of crime (Becker, 1968; Sah, 1991; Freeman, 1999).

In this paper, we address this gap and provide detailed insights into the impact of a welfare program on local-level crime and potential underlying mechanisms. For this purpose, we study the roll-out of Indonesia's flagship anti-poverty program PKH (*Program Keluarga Harapan*). PKH is the world's second largest CCT (covering more than 10 million households) and focuses on poor families with young children (below the age of 16 years). Our empirical identification strategy leverages annual data on local incidences of violent crime[2] for the period 2005 to 2014 in combination with (i) household survey data from a randomized controlled trial (RCT) and (ii) administrative data on the nationwide community-level roll-out of PKH. We derive short- and medium-run causal effects of the CCT's impact on violent crime (two to four years after program implementation).

[2]Our principal data source on violent crime is obtained from Indonesia's National Violence Monitoring System (NVMS). NVMS data is based on local newspaper reports. For robustness checks we additionally utilize self-reported violent crime data from nationally representative crime victim surveys (SUSENAS).

We find that the CCT led to a substantial increase in the incidence of community-level violent crime of about 0.6 to 3.2 percentage points (between 10% to 33%). Examining possible mechanisms behind our results, we find that the CCT allowed young males (aged 18 to 25) from beneficiary households to stay idle (neither attending school, working, or performing household chores) and, therefore, potentially provided them with additional time to commit violent crimes. In contrast, we do not find empirical support for a number of alternative explanations that are highlighted in prominent economic theories of criminal behavior (Becker, 1968; Sah, 1991; Freeman, 1999). First, we do not observe that the CCT altered incentives related to the monetary and non-monetary rewards of engaging in crime. Neither did beneficiary households and local communities become wealthier in terms of assets and expenditures, nor did the CCT create relevant local-level peer group inequalities. Second, we also find no evidence for decreases in the costs of committing crimes, with no changes in material costs (such as social mobility- and information-related costs), psychic costs (captured by community engagement), or expected punishment costs (proxied by alcohol and drug use). Lastly, we provide suggestive evidence that the CCT recipient households are possibly not only the perpetrators of violent crime but that they are at the same time also more likely to become its victims.

In order to shed further light on the CCT-idleness-crime channel, we adopt two additional empirical strategies. First, we zoom closer into the temporal patterns of crime and show that violent crime only increases during normal workdays—when being idle can make a difference—but not on weekends or during public holidays. Second, in early 2023 we conducted an online factorial vignette experiment with about 1,800 Indonesians that was designed to causally link individuals' perception of guilt (has person X committed a crime) to person's X character traits (being poor, being idle). We find that Indonesians are more likely to consider a young male guilty of theft if he is described at the same time as idle and poor. Given robust evidence from the sociological crime literature (Dressel et al., 2018; Lin et al., 2020) showing that individuals tend to be somewhat good at correctly identifying actual crime offenders, we believe that the vignette experiment underscores that there is a real-world relationship between young males' idleness and the incidence of violent crime in Indonesia.

Our paper advances the relevant literature in three ways. First, we speak to the literature that examines the impact of welfare programs on juveniles' and young adults' criminal activities. Existing studies have predominantly looked at welfare programs and policies that targeted young people directly, providing them with better access to (or ensuring stronger enforcement of) schooling and work (Machin et al., 2011; Gelber et al., 2015; Hjalmarsson et al., 2015; Bratsberg et al., 2019; Bell et al., 2022). In contrast, we investigate the case in which juveniles and young adults benefit only indirectly from intra-household spillovers (when their parents receive CCT payments and their younger siblings become more likely to go to school). In this regard, our study is most closely related to the work of Ludwig et al. (2001), Jacob et al. (2014), and Chin (2018), who analyze the effects of households moving to better neighborhoods on youth criminal activities. Unlike these studies, however, we show that welfare programs can lead to an increase in crime, possibly due to their impact on youth idleness. Moreover, our findings also align with studies on the incapacitation effect of welfare programs (Bratsberg et al., 2019). While young children become substantially more likely to be enrolled at school due to PKH, non-targeted members of beneficiary households (juveniles and young adults) gain additional freedom to potentially engage in crime.

Second, we contribute to the external validity of empirical studies in the welfare programs vs. crime literature. To the best of our knowledge, existing studies have used a single data source on crime incidences (typically police reports) and a single causal identification strategy (mostly difference-in-differences estimates).[3] Since crime data are notoriously noisy and incomplete, and each econometric identification strategy hinges on a number of assumptions that might be violated, there is a need to replicate results with alternative data and identification strategies. Our paper addresses this concern of external validity by applying two distinct identification strategies—estimating local average treatment effects (LATE) based on an RCT using difference-in-differences estimates and based on a staggered country-wide roll-out using event study design estimates—and leveraging two independent data sources to gauge crime (newspaper articles and survey-based self-reports of being a victim of crime).

Third, we add to the scarce literature on intra-household spillover effects of CCTs. Though CCTs have been widely shown to have a positive impact on

[3]See Hindelang et al., 1981; Ludwig et al., 2001; Buil-Gil et al., 2020 for a discussion on biases in official crime data and in particular, crimes recorded by the police.

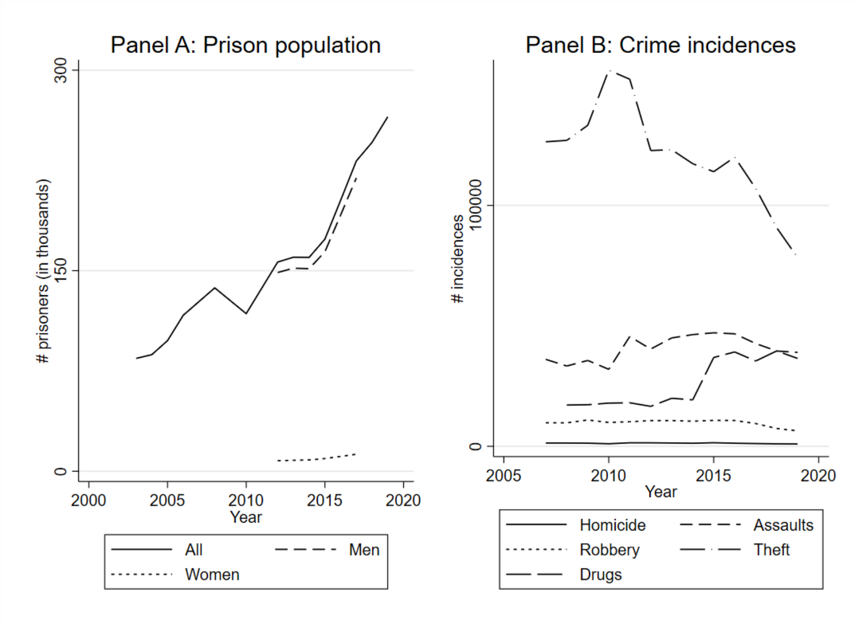
recipients' welfare (Millán et al., 2019), they can lead to distortions in the allocation of household resources (Kazianga et al., 2014; Bryan et al., 2021; Suarez et al., 2021) and unintended side effects in terms of negative impacts on education decisions and labor supply of non-targeted children (Barrera-Osorio et al., 2011; Ferreira et al., 2017; Hoop et al., 2019; Chuan et al., 2021). In this context, our study shows that CCT-induced intra-household spillover effects increased idleness among young men and, by that, possibly contributed to a rise in violent crime. Lastly, we expand the regional coverage of the CCT vs. crime literature from the Latin American context (Camacho et al., 2013; Chioda et al., 2016; Machado et al., 2018; Borraz et al., 2020; Attanasio et al., 2021) to Southeast Asia.

The remainder of the paper is structured as follows. Section 5.2 provides background information on PKH and violent crime in Indonesia. Section 5.3 describes our data. Section 5.4 presents the empirical strategy, main results, and a series of robustness checks. Section 5.5 examines different mechanisms to explain our main results, whereas section 5.6 assesses the plausibility of youth idleness as explanatory factor for the rise in crime. Section 5.7 concludes.

5.2 Background

Crime in Indonesia Indonesia has the eighth largest prison population in the world (Fair et al., 2021; UNODC, 2021), with about 271,000 incarcerated persons in 2020 (cf. Figure 5.1, panel A). Especially due to revisions in the penal code related to drug offenses, the number of prisoners has tripled between 2002 and 2020. According to administrative reports (Mutiarin et al., 2019), the majority of inmates is imprisoned because of drug-related crimes (about 70%), followed by robbery and theft (26%), and homicides (3%). Considering that not all crime leads to imprisonment, a more comprehensive picture is available from official police reports which regularly provide crime statistics since the mid-2000s. According to these reports, the vast majority of crimes comes from thefts and physical assaults (about 80% of all crimes; cf. Figure 5.1, panel B). Regarding the perpetrators of crime, the majority of suspects are male (about 95%), possess less than nine years of formal education (about 73%), and come from poorer socio-economic backgrounds, while about 50% of all crimes are committed by individuals younger than 25 years (BPS, 2013; POLRI, 2019).

FIGURE 5.1: Official crime statistics



Note: Data on prisoners (panel A) are taken from the United Nations Office on Drugs and Crime. Statistics on crime incidences (panel B) come from official police reports as tabulated in Statistics Indonesia's annual crime reports.

The conditional cash transfer program PKH is Indonesia's flagship anti-poverty program implemented by the Ministry of Social Affairs (MoSA). Eligible households have to be poor and possess a certain demographic structure (with at least one child below the age of 16 or at least one pregnant woman). PKH's conditionality criteria require participation in health screenings of pregnant mothers and young children below the age of seven and school attendance of children aged 7 to 15 (Cahyadi et al., 2020).[4] PKH provides substantial cash transfers to eligible households (Nazara et al., 2013). Depending on their demographic structure, households receive USD 83–290 per year (in 2012 prices). It is estimated that PKH transfers constitute on average about 15% of annual expenditures of beneficiary households (World Bank, 2012a). Empirical evidence shows that PKH has been successful in improving education and health outcomes among recipient households. While Alatas et al. (2011) document mixed results on schooling and child labor at program introduction, Cahyadi et al. (2020) demonstrate sustained improvements in a number of educational and health outcomes in the middle run.[5]

[4]See Tables D.17 and D.18 in Appendix section D.3 for more detailed information concerning PKH's conditionality criteria and payment structures.

[5]PKH has also served as the backdrop for an influential literature that studies the role of targeting and beneficiary selection mechanisms (Alatas et al. 2016a, Alatas et al. 2016b,Banerjee et al. 2020) and health-related spillover effects regarding suicides (Christian et al., 2019) and provider responses (Triyana, 2016).

At the time of its introduction in 2007, PKH covered about 500,000 households. Subsequently the program witnessed a gradual expansion, covering 3.2 million households by 2014, and 10 million households in 2020 (MoSA, 2020). At the beginning, it largely operated in the more urban parts of Java and Sumatra (cf. Figure D.1 in the Appendix). Throughout the following years, PKH expanded by covering more and more rural and remote areas of the country. By the year 2014, it managed to cover almost all Javanese subdistricts and significantly expanded its coverage on the islands of Sumatra, Sulawesi, Kalimantan, and several Eastern Indonesian locations. Expansion occurred at the sub-district and community level, whereby not necessarily all communities within a given sub-district were covered.[6] Once PKH began to operate in a given sub-district and community, it remained active throughout the entire time span analysed.

5.3 Data

5.3.1 Data sources and samples

NVMS crime data Our main measure of crime comes from the National Violence Monitoring System (NVMS) which covers 16 out of Indonesia's 33 provinces (at the time of recording). The geographic coverage of the NVMS was initially limited to ten conflict-prone provinces in 1998, to which six additional provinces were added in 2005 (cf. Figure D.2 in the Appendix).[7] While the selected 16 provinces are not formally representative of Indonesia, they span all major island groups and cover the majority of its population (53% in 2014). Based on systematic coding of print newspaper archives, the data captures daily local-level crime incidences at a high spatial resolution, which we aggregate to yearly frequency from 2000 to 2014. Between 2000 to 2004, the NVMS covered 75 newspapers and between 2005 to 2014, 123 newspapers. In total, the NVMS recorded and coded over 2 million articles. To code each recorded incidence, NVMS staff used standardized procedures based on the underlying motive. Broad groupings distinguish among others between violent crime and conflicts (cf. Barron et al., 2009; Barron et al., 2014;

[6]For convenience, we refer to both rural villages (*Desa*) and urban precincts (*Kelurahan*) as communities throughout this paper.

[7]The 16 provinces are: Aceh, Central Kalimantan, Central Sulawesi, DKI Jakarta, East Java, East Nusa Tanggara, Lampung, Maluku, North Maluku, North Sulawesi, North Sumatra, Papua, South Sulawesi, West Kalimantan, West Nusa Tenggara, and West Papua.

Bazzi et al., 2021b; Bazzi et al., 2022 for a discussion on the quality of the data).

Our study leverages NVMS's data on *violent crime*, which it defines as "an act of violence that occurs without any prior dispute between parties (due to monetary and/or non-monetary motives)". The violent crime indicators are split into 13 sub-categories, including assault, robbery, vandalism, and fights. In general, the vast majority of violent crime cases relate to physical assaults and robberies.[8] Since cases recorded in the NVMS have made it into local newspapers, the data does not typically cover petty crime. To the best of our knowledge, the NVMS is the only data source available for Indonesia that allows for the consistent construction of annual crime indicators for a large number of locations and a longer time horizon.[9]

Crime victimization data We complement our main results on crime by relying on self-reported crime victimization data from the national household survey (SUSENAS) for selected years. Collected by Statistics Indonesia (BPS), it consists of repeated cross-sections and interviews about 250,000 households yearly. The survey is Indonesia's principal data source on education, demographic, and labor market indicators. Since 2007, SUSENAS gathers information on the victims of crime by asking respondents about the type and number of crimes experienced themselves over the past 12 months. Due to changes in questionnaire design and data release policies, our analysis of crime victimization rates is restricted to the household survey rounds from 2007 to 2011 (cf. Appendix section D.4 for a detailed overview of the SUSENAS sample construction). While survey data for this period can be matched with the administrative PKH roll-out data at the sub-district level, in later years BPS removed sub-district-level identifiers from the data.[10]

Roll-out data and sample From MoSA we obtained administrative program data that documents the annual roll-out of PKH—indicating the first

[8]See Appendix Table D.1 for descriptive statistics on the various types of violent crime.

[9]Records from the police are only available at a spatially more aggregated provincial level. At this level (covering the period 2007–2014), we find that the number of violent crimes reported in NVMS is positively, albeit not perfectly correlated with administrative crime statistics from the police (with a correlation coefficient of 0.38). Crime data from other sources such as the community-level census PODES is self-reported, covers crime only for selected years and is subject to changing definitions over time.

[10]Further results in Appendix section D.4.2 cover the years 2013–2014, and 2017–2019. These survey years cannot be matched with the administrative roll-out data but include self-reported information on whether a household received PKH, which can be used to match recipient households to comparable poor households.

year in which implementation began—at the community level for the period 2007 to 2014. While the PKH roll-out occurred in most parts of the country, we restrict our main analytical sample to (i) those 16 provinces that are included in the NVMS, and (ii) those communities that received PKH at some point during the years 2007 to 2014. This latter restriction aims at reducing the threat of omitted variable bias, since communities that received access to the CCT program much later or not at all differ substantially from communities that were part of the roll-out during the 2007–2014 period. In our robustness checks we later relax restriction (i) to assess differences between crime rates reported in victim surveys in communities with and without NVMS coverage, and also relax restriction (ii) to provide estimates for all locations with NVMS coverage irrespective of whether the sub-districts were reached by the CCT program by 2014. The *Roll-out sample* covers the period 2005 to 2014 and all communities in which PKH has been introduced until the year 2014. By then Indonesia consisted of 82,190 communities, out of which 45,077 are covered by NVMS and out of which again 28,873 received PKH. Our main sample thus consists of 288,730 observations over 10 years.[11]

RCT data and sample For parts of our empirical identification strategy and for exploring possible mechanisms behind our main result, we leverage data from a randomized impact evaluation conducted by the World Bank. The RCT collected household survey data in 2007 (at baseline) and 2009 (at endline) to enable a rigorous impact evaluation of the pilot program on various child education and health outcomes (Alatas et al., 2011; Cahyadi et al., 2020). The RCT was fully integrated into the launch of PKH, covering 360 sub-districts across six provinces (DKI Jakarta, East Java, Gorontalo, North Sulawesi, Nusa Tenggara Timur, and West Java), with 180 sub-districts serving as control and 180 as treatment. The sampling frame of the RCT largely overlaps with the regions covered by the NVMS. In total, 250 sub-districts (123 control and 127 treatment sub-districts) of the RCT are also part of the NVMS. Based on this spatial overlap we construct what we refer to as the *RCT sample*. It covers 1,830 communities from 250 sub-districts and represents a sub-sample of the *Roll-out sample* in terms of geographic and temporal coverage. The *RCT sample* starts with the year 2005 (two years before the

[11]In the period from 2005 to 2014, Indonesia witnessed the creation of many new administrative units, including communities, sub-districts, and districts. Throughout the paper, our unit of analysis refers to the 2014 list of communities and community boundaries. We match datasets using administrative identifiers and a crosswalk based on RAND (2022).

program was introduced) and ends in 2010, after which a significant share of the RCT's control areas also received access to PKH.

Sub-sample selection and internal vs. external validity Restricting the *RCT sample* to areas covered by the NVMS data might threaten the internal validity of the original treatment assignment. Previous research showed that the randomization in PKH's impact evaluation resulted in covariate balance across a large number of socio-economic variables (Alatas et al., 2011; Cahyadi et al., 2020). Since the original randomization was stratified at the province level, restricting the sample to those provinces that are covered by the NVMS should in principle also result in covariate balance, albeit at the cost of lower power. Tables D.19 and D.20 in Appendix section D.3 provide results from balance tests for our *RCT sample*. By and large, we find no systematic differences across control and treatment areas.

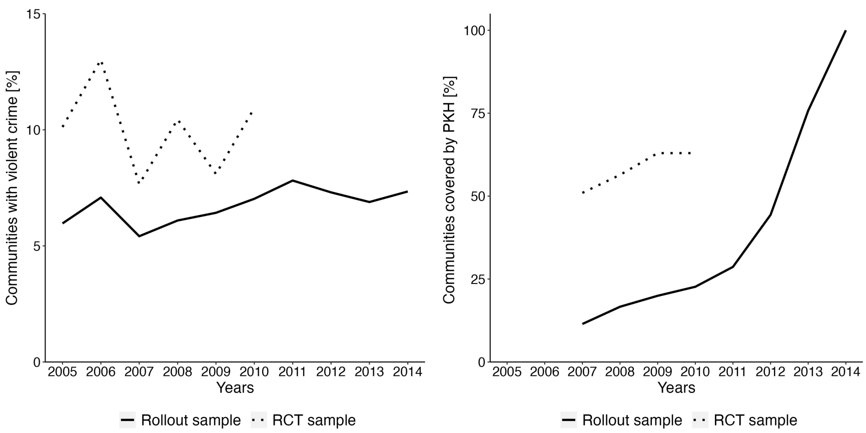
Restricting both the *Roll-out sample* and the *RCT sample* to areas covered by the NVMS data also raises questions about the external validity of our results. Table D.2 in the Appendix compares the NVMS and non-NVMS communities, restricted to communities that received PKH until 2014. Across different samples and consistent with NVMS's focus on more violence-prone regions, we find that communities covered by the NVMS tend to be less populous and somewhat more rural, whereas we do not observe clear-cut differences across the *RCT* and the *Roll-out samples* themselves. While this might raise the suspicion that our results are only relevant for the less developed regions of Indonesia, our subsequent analysis of crime victim surveys shows that self-reported victimization rates also increase in provinces without NVMS coverage.

Further data To complement our short-term analysis of potential mechanisms based on the RCT, we additionally rely on household data from SUSENAS to identify PKH eligible households, to assess the time use of household members and to build further controls. Finally, our empirical analysis also draws on information provided by PODES, the Indonesian village (or community) censuses that interview local government heads every three to four years and gather information on community population, economy, and infrastructure.

5.3.2 Variable construction

Violent crime Based on the NVMS data, we construct annual, community-level indicators of violent crime. Our main specification defines violent crime at the extensive margin and relies on a binary variable that takes the value of one if a community experienced at least one case of violent crime in a given year and zero otherwise. Panel A of Figure 5.2 depicts the crime trends in the *Roll-out sample* and the *RCT sample*. We observe a moderate increase in violent crime cases over time in the *Roll-out sample*: About 2,800 communities experienced violent crime incidences in 2005, which increased to about 3,500 communities in 2014 (averaging at about 6% of all observed communities in the sample period). In the *RCT sample* the share of communities experiencing violent crime is somewhat higher and fluctuates at around 9%.

FIGURE 5.2: Community-level crime incidence and PKH roll-out



(A) Crime incidence (B) PKH roll-out

Note: The *Roll-out sample* shows statistics on 28,873 communities that received PKH between 2007 and 2014. The *RCT sample* is restricted to 1,830 communities that were included at the RCT stage of the PKH program and ends in 2010.

Treatment indicators Across all Indonesia, PKH was scaled up rapidly from reaching 4,000 communities in 358 sub-districts in 2007 to 56,000 communities in about 4,800 sub-districts in 2014. Our principal treatment indicator measures the actual implementation of the CCT program at the community level at the extensive margin (based on MoSA's administrative data) and records whether PKH was actually operating in a community in a given year. Panel B of Figure 5.2 shows a marked expansion in the share of communities receiving transfers in our *RCT* and *Roll-out samples* over time.

To analyse treatment effects in the *RCT sample*, we also construct an additional indicator that refers to original treatment *assignment* of program implementation as part of the RCT, which took place at the sub-district level. The indicator takes the value of one if a community is located in an originally assigned treatment sub-district and zero otherwise. As discussed in Cahyadi et al. (2020) and also documented in Table D.21 in section D.3, the actual implementation of the treatment did not strictly follow its intended assignment. While we observe almost perfect compliance with treatment assignment in treatment group communities, control group compliance decreases over time with the progress of the national program roll-out.[12] We use our binary indicator of PKH treatment assignment to identify intention-to-treat (ITT) effects and as an instrument for actual program implementation in our IV framework, yielding LATE estimates.

Additional variables We rely on the baseline and endline surveys of the RCT to explore potential mechanisms and construct additional control variables. Outcome variables relate to household-level assets and expenditures, individual-level education and labor supply, as well as community-level peer-group inequality. For further analyses, we derive household-level measures of crime victimization and individual-level education and labor supply variables from SUSENAS, and a set of community characteristics from PODES.[13]

5.4 The effects of the CCT on violent crime

5.4.1 Econometric framework

We estimate three distinct econometric models to derive causal effects in the *RCT sample*. First, we employ a simple two-way-fixed effects estimator (TWFE) in which we regress community-level violent crime on the actual community-level implementation of the treatment:

$$Crime_{ckdt} \quad = \quad \eta \; PKH\text{-}Treat_{ckdt} + \lambda_c + \theta_{dt} + \epsilon_{ckdt}, \tag{5.1}$$

where $Crime_{ckdt}$ is a binary variable that indicates the occurrence of any violent crime in community c belonging to sub-district (*Kecamatan*) k of district

[12] By 2010, the CCT was implemented in about 33% of the RCT control communities.
[13] See Appendix Table D.3 for details on the construction of all variables.

(*Kabupaten* or *Kota*) d in year t. $PKH\text{-}Treat_{ckdt}$ is a binary indicator that denotes the actual implementation of the program. Community fixed effects, λ_c, control for time-invariant community-specific differences in the propensity of experiencing violent crime. District-year fixed effects, θ_{dt}, control for a wide-range of economic and political dynamics at the district level. To allow for potential serial correlation and to account for the group-wise treatment design (Angrist et al., 2008), we cluster standard errors at the level of randomization (sub-districts k).

Second, we provide ITT estimates by regressing crime on the original treatment assignment:

$$Crime_{ckdt} \quad = \quad \alpha\, PKH\text{-}Assign_{kd} \times Post_t + \lambda_c + \theta_{dt} + \epsilon_{ckdt}, \tag{5.2}$$

where $PKH\text{-}Assign_{kd}$ denotes the original treatment assignment and $Post_t$ denotes the post-treatment years, and α gives us the ITT effect.

Third, we address concerns about potential biases in our ITT estimates due to the fact that compliance with original treatment assignment was not perfect; in particular some control communities also received access to the program (cf. Table D.22 in Appendix section D.3). To correct for such biases, we estimate treatment effects by adopting an IV strategy that provides us with LATE estimates. We use the original treatment assignment, $PKH\text{-}Assign_{kd}$, interacted with a post-treatment indicator as an instrument for actual community-level program implementation. Results are obtained from estimating the following equation by two-stage least squares (2SLS):

$$\begin{aligned} PKH\text{-}Treat_{ckdt} &= \beta\, PKH\text{-}Assign_{kd} \times Post_t + \pi_c + \phi_{dt} + u_{ckdt} \\ Crime_{ckdt} &= \gamma\, PKH\text{-}Treat_{ckdt} + \lambda_c + \theta_{dt} + \epsilon_{ckdt} \end{aligned} \tag{5.3}$$

We contrast the estimates from the RCT sample with large-scale quasi-experimental estimates based on the national implementation of the program between 2005 and 2014. In the *Roll-out sample*, treatment effects are obtained from estimating equation (5.1) above, relying on standard TWFE specifications, but also comparing them with two difference-in-differences event study design estimators robust to heterogeneous treatment effects (Borusyak et al., 2021; Sun et al., 2021).

Since the program was rolled out in a (non-random) staggered fashion across the country, our national roll-out results rely on the assumption of parallel trends, conditional on district-year fixed effects and constant treatment effects (Goodman-Bacon, 2021). To study which communities received early access to the program, Table D.4 in the Appendix regresses the first year of program introduction on a set of initial local characteristics.[14] Results show that the nation-wide roll-out of the CCT targeted more populous, urban, and poorer communities first. The evidence is more mixed for supply-side preparedness indicators. Only the local presence of kindergartens is significantly associated with an earlier introduction of the program, whereas places with access to health facilities were actually significantly later covered by the program. Importantly, however, these correlates lose their significance entirely once district fixed effects are also accounted for (with the joint F-statistic declining from 28.6 to 1.4). Thus, while community characteristics explain the broad regional roll-out patterns, the within-district variation in roll-out timing appears to be substantially more idiosyncratic. Due to this, our preferred specifications always include district-year fixed effects. Additionally, we also present results that include all initial conditions reported in Table D.4 as further controls interacted with a full set of year fixed effects each.

5.4.2 Main results

Panel A of Table 5.1 presents our main findings using the *RCT sample*. Estimates from the TWFE specifications (columns 1–2) show a positive and statistically significant impact of the CCT program on community-level incidence of violent crime. Once controlling for district-level differences in crime dynamics over time (starting with column 2), coefficients are precisely estimated, with significance levels at five percent and below. Our ITT estimate (column 3) is qualitatively similar to the TWFE results, while the LATE estimate (column 4) corrects for non-compliance and yields even larger increases in violent crime as a result of randomized access to the program. The LATE estimate shows an increase in the likelihood of a community experiencing violent crime by 3.2 percentage points at the extensive margin. This translates to a rise of about 33% in violent crime compared to communities without access to the CCT program, which is a substantial increase.

[14]We measure initial characteristics in 2008, based on PODES village census data. This information is collected right after the first introduction of the program in 2007, and we do not expect to see changes in relevant community characteristics in response to the program.

TABLE 5.1: The effects of the CCT program on violent crime

	Panel A: RCT sample (2005–2010)			
Estimation	TWFE		TWFE (ITT)	IV (LATE)†
	(1)	(2)	(3)	(4)
PKH treatment	0.021*	0.026***	0.027***	0.032***
	(0.011)	(0.010)	(0.010)	(0.012)
Community FE, year FE	Yes	Yes	Yes	Yes
District-year FE		Yes	Yes	Yes
Observations	10,980	10,980	10,980	10,980
Sub-districts (clusters)	250	250	250	250
Adj. R-squared (F-stat^F)	0.307	0.321	0.447	1735.0^F
Mean (control)	0.101	0.101	0.101	0.101

	Panel B: Roll-out sample (2005–2014)				
Estimation	TWFE			BJS‡	
	(1)	(2)	(3)	(4)	(5)
PKH treatment	0.004*	0.007***	0.006**	0.015***	0.013***
	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)
Community FE, year FE	Yes	Yes	Yes	Yes	Yes
District-year FE		Yes	Yes	Yes	Yes
Controls			Yes		Yes
Observations	288,730	288,730	265,970	244,270	224,959
Sub-districts (clusters)	2,335	2,335	2,310	2,335	2,310
Adj. R-squared	0.206	0.221	0.220		
Mean (control)	0.060	0.060	0.061	0.060	0.064

Note: The dependent variable is a binary indicator that takes the value one if NVMS reported at least one violent crime incident in a community in a given year. The *RCT sample* (Panel A) is restricted to 1,830 communities. The *Roll-out sample* (Panel B) includes at most 28,873 communities. In the TWFE and IV models, PKH treatment measures the actual access to the CCT program by at least one household. In the ITT model, PKH treatment stands for the original treatment assignment status at the sub-district level. In the IV model, PKH access is instrumented by PKH assignment. † The first-stage coefficient of PKH treatment on PKH assignment is 0.844, with a standard error of 0.020; ^{F}Kleibergen-Paap F-statistic is reported instead of R-squared. ‡ Imputation-based estimates (Borusyak et al., 2021). Further controls include year fixed effects interacted with a set of initial conditions (as specified in Table D.4 in the Appendix). The displayed mean of dependent variable refers to the (assigned) control group only. Robust standard errors are clustered at the sub-district level and reported in parentheses. */**/*** denote significance levels at 10/5/1% respectively.

Panel B of Table 5.1 contrasts the above findings with the large-scale medium-run effects of the administrative roll-out of the program. In line with the RCT results, the TWFE specifications (columns 1–2) suggest that the CCT program led to an increase in violent crime, albeit the treatment effects of the nation-wide roll-out are somewhat lower (amounting to about 0.7 percentage points or an increase of about 12% compared to pre-program violent crime incidence). Adding a full set of time fixed effects interacted with a list of initial community characteristics in column 3 (as specified in Appendix Table D.4) changes the estimate only marginally. To address potential biases inherent to staggered treatment designs in the presence of time-varying treatment effects, columns 4 and 5 show results from an imputation-based estimator (BJS) relying on the method proposed by Borusyak et al. (2021).[15] The BJS estimator also yields a positive link between PKH coverage and violent crime, with a somewhat higher and more precisely estimated coefficient. Overall, these baseline results document that violent crime rates increased substantially in Indonesian communities in response to receiving access to the CCT program.

5.4.3 Robustness checks

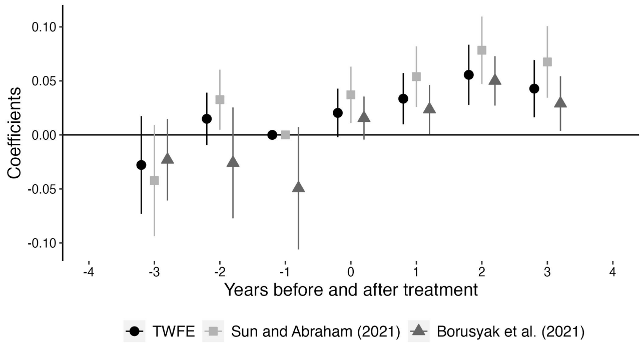
Applying an event study design provides empirical evidence on the parallel trend assumption and sheds light on dynamic treatment effects. To do so we replace the static treatment indicator in equation (5.1) with a set of dynamic treatment indicators *PKH-Treat*$_{ckd,t+\tau}$ with $\tau \in \{-4, ..., -2, 0, 1...4\}$, which take one in $t + \tau$ years before and after PKH was introduced in a community and zero otherwise. In both the *RCT* and the *Roll-out sample*, the baseline omitted category is $t - 1$.[16] Results are presented in Figure 5.3. Pre-treatment indicators are close to zero and insignificant for both samples, corroborating the parallel trends assumption in the difference-in-differences models. In post-treatment years, the impacts of the program increase gradually and are sustained for at least three years. Since in the course of the nation-wide program implementation the number of CCT recipient households was increasing over time also within treated communities, this could have also given rise to increasing spillover effects on local crime.

Given the staggered nature of our fixed effect difference-in-differences specifications, concerns might arise that our main treatment effect is driven by period- and group-specific differences in weights but not average treatment effects. Further estimation results alleviate this concern as the effects are generally less significant but show similar dynamics using Sun et al.'s (2021) interaction-weighted estimator. By contrast, the Borusyak et al. (2021) estimator results in somewhat lower increases in crime in the *RCT sample* and higher increases in the *Roll-out sample*, with effects rising two to three years after the introduction of the program.

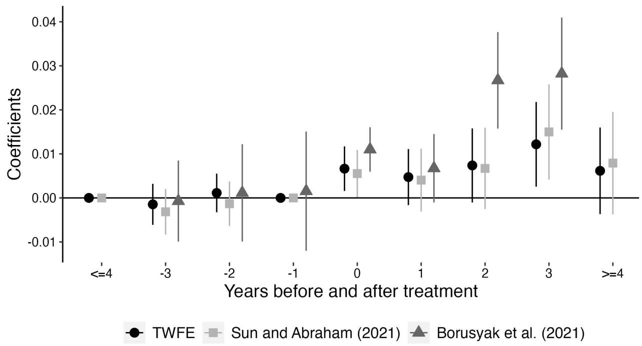
[15]The BJS estimator relies on first estimating fixed effects using "not-yet-treated" observations for all communities that still have such observations and then in a second step on projecting counterfactual outcomes on treated observations. Observations with no available imputed counterfactual are excluded from the analysis.

[16]For the *Roll-out sample*, as all communities are treated by the end of the time frame (2005–2014), it is good practice to conceptually include a second omitted category, a binned indicator for all observations before four years of treatment start (cf. Schmidheiny et al., 2019; Borusyak et al., 2021). For all years after four years, we also use a binned indicator assuming constant effects thereafter. For the *RCT sample* the last indicator in $t + 4$ is dropped as the shorter sample (2005–2010) can only capture effects up to three years after treatment.

FIGURE 5.3: Robustness: Pre-trends and treatment effects by year, normalized



(A) RCT sample



(B) Roll-out sample

Note: The figure provides dynamic PKH effect estimates based on the treatment indicator of *actual* PKH beginning (cf. equation (5.1)). The *RCT sample* is restricted to 1,830 communities that were included in the RCT of the PKH program. The *Roll-out sample* includes 28,873 communities that received PKH from 2007 to 2014. Both samples are restricted to communities with full NVMS data coverage. The dependent variable is a binary indicator that takes the value one if NVMS reported at least one instance of violent crime in a community in a given year. Regressions control for community, year, and district-year fixed effects. TWFE estimates and Sun et al.'s (2021) estimator are run using R's *fixest* package. Borusyak et al.'s (2021) estimator is run using STATA's *did_imputation* package. 95% confidence intervals are reported, based on robust standard errors, clustered at the sub-district level.

Causal identification in difference-in-differences specifications typically requires that the stable unit treatment value assumption (SUTVA) holds. Since our units of observation are spatial units (communities) across time, we believe that the main identification threat is related to spatial spillovers. In particular, our estimates will be biased upward in the case of negative spatial spillovers of crime. Crime will be displaced from control to treatment communities for instance if in response to program roll-out, criminals living in control communities decide to travel to communities receiving CCTs. We adopt two empirical strategies to assess the relevance of such spatial spillovers: First, we follow Crost et al. (2016) and examine heterogeneous treatment effects with respect to the Euclidean distance of treatment and control communities. Second, we follow Clarke (2017) and re-estimate our main results using the "spillover-robust double difference estimator" which, based on distributional assumptions on how spillover effects vary in distance between control and treatment communities, aims to take spatial spillovers into account explicitly. Findings shown in Appendix Table D.5 do not indicate any spillover effects to neighboring communities.

To provide further credibility of the causal nature of our estimates, we perform a series of additional sensitivity checks. In Table D.6 in the Appendix, we randomly assign treatment status in the *Roll-out sample* both across and within years (columns 1–2), which yields fairly precisely estimated zero effects. We also show that results in the *RCT sample* are not driven by the randomization itself (Athey et al., 2017). Standard errors adjusted for randomization inference indicate that our previous findings hold (column 3). Results using alternative types of standard errors adjustments are presented in Appendix Table D.7. We allow for potential spatial correlation in the error terms by clustering our regressions at the level of districts—the next administrative level—and by using Conley standard errors (Conley, 1999) within a 50 kilometer radius. The significance of the program effects in the *Roll-out sample* and the *RCT sample* remains stable at the 1% level. Finally, in Appendix Table D.8 we conduct a number of sensitivity checks to assess whether the way we construct our samples affects our findings. Column 1 presents results from an extended sample, in which we include all 3,323 communities within the 250 sub-districts that were part of the RCT and at the same time also covered by the NVMS.[17] Column 2 shows results that are estimated at the sub-district level (instead of community level), motivated by the circumstance that treatment assignment was at the sub-district level. Based on the *Roll-out sample*, column 3 uses a sample that considers a longer time period, starting with the year 2000 and therefore several years before PKH operations began anywhere in Indonesia. Column 4 contains estimates for an extended *Roll-out sample* that comprises all communities covered by NVMS irrespective of actual PKH roll-out until 2014 (47,680 communities instead of of 28,873 communities). Throughout the different sample definitions, we find that our main results hold qualitatively.

[17]This sample is larger than our *RCT sample* as it also includes communities in which no baseline and endline data was collected.

Lastly, we assess the sensitivity of our results to measurement issues along several dimensions: First, results at the intensive margin, defined by the hyperbolic sine of the number of violent crime incidences in a community in a given year, confirm increases in violent crime in communities receiving CCT access by 0.5 to 3 percentage points (Table D.9 in the Appendix). Second, we show that increases in violent crime in program communities are not driven by selective migration patterns (e.g., by young persons delaying their migration decisions). Exploiting data from PODES, we do not find that community-level population numbers changed as a result of the program (Appendix Table D.10).[18] Third, we test whether our violent crime variable might simply pick up local conflicts, in which case we might not measure “genuine" additional cases of violent crime but rather spillovers from conflict incidences.[19] Using NVMS incidents coded as “conflicts", we re-estimate the impact of the program on conflict incidences (Appendix Table D.11). We find no link between the CCT program and community conflict in Indonesia.

5.4.4 Additional evidence from crime victim surveys

Crime victim surveys (from SUSENAS) allow us to assess the sensitivity of our results to the quality and coverage of the NVMS crime data in order to address two main concerns. First, NVMS-based results could be driven by a non-random measurement error if newspapers systematically change their reporting practices and start reporting crime more frequently once the CCT program is (becoming) established in a community. The second concern relates to the external validity of our results: As NVMS covers the more violence-prone (and by that more remote and less wealthy) half of the country, our main findings might not carry over to the rest of Indonesia.

Crime victimization surveys address these concerns by offering an alternative measurement of local crime rates. Despite of being self-reported and hence subject to survey biases, individual survey response is unlikely to depend on what has been reported in provincial newspapers, mitigating the first concern. Moreover, as household surveys are also fielded in provinces

[18]The result is consistent with findings in Cahyadi et al. (2020) who use household survey data from the CCT’s impact evaluation to show that individual- and household-level attrition rates are unrelated to program implementation.

[19]In economics and political science there has been a long-standing debate on the impact of local-level income shocks on conflict (Berman et al., 2011; Crost et al., 2014; Nunn et al., 2014); including the role of CCTs (Crost et al., 2016). Given that Indonesia has a history of conflict events, particularly in the period between 1997 to 2003 (Barron et al., 2009; Pierskalla et al., 2017), the CCT program could have influenced the likelihood of local conflict events.

without NVMS coverage, we can assess the adjustments in self-reported crime throughout the country.

For this purpose, we link existing crime victimization data at the household level to the administrative program roll-out at the sub-district level. Our sample is restricted to the years 2007–2011 by the inclusion of detailed questions on crime in SUSENAS and the provision of sub-district identifiers. Since the yearly household surveys do not cover the same sub-districts in every round, we further restrict the sample to those sub-districts that were covered at least twice during the period 2007–2011. This allows us to estimate the following TWFE model at the household level:

$$\textit{Crime victim}_{jkdt} = \eta \, \textit{PKH-Treat}_{kdt} + X'_{jkdt}\gamma + \kappa_k + \theta_{dt} + \epsilon_{jkdt}, \qquad (5.4)$$

where $\textit{Crime victim}_{jkdt}$ indicates whether household j residing in sub-district k of district d reported having been a victim of a violent crime in year t. $\textit{PKH-Treat}_{kdt}$ captures either the share of communities within the sub-district k that were covered by the program or whether at least 100 households in the sub-district k had PKH access in year t. We control for a basic set of household characteristics, X'_{jkdt}, which includes indicators for urban status, the educational level, age and marital status of the household head, and household size quintiles. To mirror our previous difference-in-differences specifications, we control for sub-district fixed effects, κ_k, and district-year effects, θ_{dt} and hence focus on sub-district-level changes in crime victimization over time. As before, we restrict our sample to sub-districts that received access to CCTs by 2014. Table 5.2 reports the results. In line with the results documented in Table 5.1, we find that program roll-out also increased self-reported incidences of violent crime throughout the country. In column 1, when all communities within a sub-district receive access to the CCT program, the likelihood of having been the victim of a crime increases by 0.7 percentage points (or about 17%). Increases in crime victimization rates are comparable if at least 100 households get access to PKH.[20] The magnitude of the effect is comparable to our main results using community-level crime incidences of the NVMS data. This make it less likely that our main results merely reflect

[20] As documented in Table D.12 in the Appendix, self-reported crime increases significantly only when a substantial number of households receives PKH, but not yet if only a few households in the sub-district are targeted.

changes in newspaper reporting instead of actual crime rates.[21] Importantly, the almost identical coefficients to results presented in Table 5.1 also confirm that our findings are not limited by the partial spatial coverage of the NVMS data but carry over to the rest of the country as well, dealing with the second concern and increasing the external validity of our results.

TABLE 5.2: Alternative measure: The CCT's effects on the probability of being a victim of violent crime (2007–2011)

PKH roll-out:	Village share in sub-district			At least 100 households in sub-district		
Household sample:	All	Poor	Non-poor	All	Poor	Non-poor
	TWFE (1)	TWFE (2)	TWFE (3)	TWFE (4)	TWFE (5)	TWFE (6)
PKH roll-out	0.007***	0.018***	0.007**	0.008**	0.025***	0.006*
	(0.003)	(0.006)	(0.003)	(0.003)	(0.007)	(0.004)
Mean (control)	0.043	0.034	0.044	0.043	0.034	0.044
Sub-district FE	Yes	Yes	Yes	Yes	Yes	Yes
District-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,095,130	95,340	999,630	1,095,130	95,321	999,630

Note: All estimates are restricted to households living in sub-districts that received access to the program until 2014. The dependent variable is a binary indicator that takes the value one if a household reported being victim of at least one crime in a given year. The first treatment variable measures the share of CCT recipient communities within a sub-district; the second treatment variable turns to one if at least 100 households within the sub-districts received PKH. Controls include indicators for the completed education, age and marital status of the household head, the quintiles of household size, and urban status. Standard errors in parentheses are clustered at the level of sub-districts. */**/*** denote significance levels at 10/5/1% respectively.

Split sample estimates based on classifying households into poor and non-poor in columns 2/5 and 3/6, respectively, further reveal that crime victimization increases both among households living below and above the poverty line.[22] Coverage of the CCT program can be linked to significantly increased victimization among non-poor households that were not eligible to receive transfers themselves (by about 0.7 percentage points, or 16%), but the relative magnitude of the effects is even larger among poor households that were potentially among transfer recipients (by 1.8 and 2.5 percentage points, or 53% and 73% respectively).

A further set of results outlined in Appendix section D.4.2 analyses crime victimization in later years based on an alternative SUSENAS sample for the years 2013–2019 (cf. Table D.26). As the administrative program roll-out information cannot be matched to the more recent SUSENAS rounds, we rely

[21] An alternative test for the existence of reporting bias in the NVMS data is reported in Table D.13. Based on PODES data, it shows that the roll-out of PKH did not induce improvements in the local police infrastructure, which otherwise might have also spuriously resulted in a better documentation and reporting of crime.

[22] Wealth classifications of households are based on per capita expenditures benchmarked against Indonesia's official provincial rural and urban provincial poverty lines.

instead directly on households' reports on whether they receive PKH (available for selected years). Results from propensity score matching show that CCT-receiving households report having been a victim of a crime somewhat more often than comparable eligible poor households. One rival explanation for the increase in self-reported crime victimization among transfer recipients could also be due to a change their crime reporting behavior. We can test the relevance of this explanation in this latter sample explicitly, by regressing the share of crimes reported to the police from all experienced crimes on program beneficiary status. Table D.27 in the Appendix shows that CCT recipient households are not more likely to report crime incidents to the police than non-receiving households.

5.5 Potential mechanisms

Our main results show that the CCT program led to an increase in violent crime at the community level. To explore possible mechanisms behind our findings, we guide our analysis by the theoretical models and conceptual frameworks outlined in Becker (1968), Sah (1991), and Freeman (1999). In particular, we approach the analysis from a supply-side model of crime, in which increases in the benefits and/or decreases in the costs of pursuing a criminal activity determine whether individuals engage in violent crime.

The principal data in this section comes from the baseline and endline surveys of the World Bank's impact evaluation.[23] Treatment effects are estimated by 2SLS with the empirical specification following Cahyadi et al. (2020) closely:

$$\begin{aligned} \textit{PKH-Beneficiary}_{jk} &= \gamma \ \textit{PKH-Assign}_k + X'_{(i)j0}\phi + \theta Y_{(i)jk0} + \alpha_d + \epsilon_{(i)jk} \\ Y_{(i)jk} &= \beta \ \textit{PKH-Beneficiary}_{jk} + X'_{(i)j0}\lambda + \omega Y_{(i)jk0} + \pi_d + \mu_{(i)jk} \end{aligned} \tag{5.5}$$

where $Y_{(i)jk}$ denotes the outcome of interest for either individual i or household j residing in sub-district k at the time of the endline survey, with the value of the dependent variable at baseline labeled as $Y_{(i)jk0}$. *PKH-Beneficiary*$_{jk}$ refers to the treatment variable, which is a binary indicator of

[23] As discussed in Cahyadi et al. (2020), the impact evaluation showed balance in almost all covariates, a high level of compliance with treatment assignment, and very low levels of attrition. We replicate these findings using our own data in the Appendix section D.3 (Tables D.20 and D.23).

living in a transfer-recipient family. *PKH-Assign*$_k$ represents the instrument (treatment assignment at the sub-district level). $X'_{(i)j0}$ is a vector of individual and household-level controls from the baseline survey.[24] α_d and π_d refer to district fixed effects. Standard errors are clustered at the level of randomization (sub-districts).

5.5.1 Benefits of crime

Monetary rewards Although the CCT's benefits are meant to cover expenditures related to young children's education, healthcare, and food, households might use the cash payments to purchase valuable and easy to loot assets such as cellphones, TVs or motorbikes (Borraz et al., 2020). Purchases of valuable goods could increase the asset base in a CCT-receiving community, thereby increasing the monetary incentives to perpetrate theft and thus the incidence of violent crime. As shown in panels A, B, and C of Table 5.3 and consistent with Cahyadi et al. (2020), we do not find evidence for increased asset possession among program beneficiaries.[25]

Alternatively, ownership of lootable assets in program recipient communities might have increased among non-beneficiary households. Through within-community spillovers (general equilibrium effects) increased spending on food, health, and education by CCT beneficiaries might have translated into higher wealth of non-recipient households residing in the same community (Angelucci et al., 2009; Cunha et al., 2019; Filmer et al., 2021; Egger et al., 2022; Muralidharan et al., 2023). Relying on data from PODES and estimating equation (5.1), Table D.10 in the Appendix shows that program implementation did not lead to an increase in various community wealth proxies.

Non-monetary rewards Benefits of crime are not limited to monetary rewards but can include psychological aspects related to deriving satisfaction from achieving fairness, retribution, and a sense of accomplishment. As discussed in Fajnzylber et al. (2002) and Cameron et al. (2014), such perceptions and reactions can be triggered if government programs lead to increases in

[24]Controls include the age, gender, and marital status of the respondent as well as fifteen household head characteristics (indicators on the level of education and sector of work) and general household characteristics (household size and indicators on various dwelling features, e.g., toilet, source of drinking water, roof type, wall type, and floor material).

[25]Instead of a true null effect, results could also reflect non-random measurement error in asset and wealth variables due to selective reporting practices as transfer recipient households might tend to under-report asset/wealth in order to maintain their eligibility status. As shown in Banerjee et al. (2020), selective under-reporting of assets and wealth does not seem to be an issue in the context of the Indonesian PKH.

local inequality, in particular related to the exclusion of eligible poor households. Given that mistargeting of transfers to non-poor households is common during the implementation of PKH (World Bank, 2012b; Alatas et al., 2019), violent crime might have increased in communities as a result of addressing perceived unfairness related to local-level program implementation.

TABLE 5.3: RCT: The short-run effects of the CCT program on assets, expenditures and behavior

	Panel A: Wealth and expenditures			
	Assets	Total Exp.	Food Exp.	Non-food exp.
PKH beneficiary	-0.002	-0.008	-0.003	-0.018
	(0.051)	(0.032)	(0.033)	(0.047)
	Panel B: Transportation assets and expenditures			
	Bicycle	Motorcycle	Car	Transport Exp.
PKH beneficiary	-0.027	0.011	0.007	-0.117
	(0.121)	(0.030)	(0.005)	(0.092)
	Panel C: Assets related to information acquisition			
	Radio	TV	Antenna	Cell
PKH beneficiary	0.033	-0.005	0.011	0.000
	(0.103)	(0.027)	(0.010)	(0.041)
	Panel D: Expenditures on alcohol and drugs			
	Alcohol Exp.	Drug Exp.	Alc.& Drug Exp.	
PKH beneficiary	-0.384	-0.054	-0.019	
	(0.239)	(0.203)	(0.202)	
	Panel E: Community involvement			
	Engage 1	Engage 2	Engage 3	Engage 4
PKH beneficiary	0.012	0.012	-0.057	-0.080
	(0.027)	(0.027)	(0.119)	(0.100)

Note: Results are reported for 12,929 households, based on 2SLS estimates (LATE) from the pilot's impact evaluation, using the baseline and endline surveys. All regressions use district fixed effects and a set of controls as described in (5.5). Abbreviation 'Exp.' refers to per capita expenditure values in logs. 'Engage 1' captures whether a household is a member of any type of community organization. 'Engage 2' refers to the number of community organizations a household has joined. 'Engage 3' refers to the number of household members that have joined community organizations. 'Engage 4' captures the number of times any household member has joined a community organization meeting over the 3 months preceding the survey. Robust standard errors in parentheses are clustered at the sub-district level. */**/*** denote significance levels at 10/5/1% respectively.

Examining the baseline and endline survey data from the RCT's impact evaluation, Table D.14 in the Appendix shows that the CCT program did not lead to an increase in local-level inequality among the poor in treated communities. Moreover, building on Cameron et al. (2014), we link program implementation features (undercoverage and leakage) to increases in violent crime.[26] Combining information from SUSENAS on PKH undercoverage and leakage with the *RCT* and *Roll-out samples*, Appendix Table D.15 shows that increases in violent crime were not larger in districts that suffered from more substantial implementation errors. If at all, we find some evidence that undercoverage might be negatively related to increases in violent crime, which might suggest that violent crime increases only if a sufficient number of eligible poor households have received program benefits.

5.5.2 Costs of crime

Bearing data limitations in mind, we aim to shed light on four cost categories that theoretical models of crime have focused on: Material costs, psychic costs, opportunity costs, and expected punishment costs.

Material costs Receiving a CCT relaxes households' budget constraints and therefore can contribute to reducing the material costs of committing crime by (i) enabling the purchase of weapons (Duggan, 2001; Koenig et al., 2021); (ii) increasing mobility and social interactions (Glaeser et al., 1996); and (iii) improving information acquisition about potential crime victims (Glaeser et al., 1996). While our data does not capture whether households possess weapons, we find that the program most likely did not alter households' mobility as proxied by the possession of transport-related assets and expenditures (panel B in Table 5.3) and did not affect households' possession of information and communication related assets (panel C in Table 5.3).

Psychic costs Psychic costs relate—among others—to feelings of guilt. As stressed by the experimental literature in economics and psychology, guilt (aversion) towards others is strongly influenced by interpersonal relationships (Charness et al., 2006; Battigalli et al., 2007) and the social and communal ties that people are embedded in (Baumeister et al., 1994; Tangney, 1995; Morell, 2019). Considering that the empirical evidence on the impact

[26] Data on undercoverage and leakage is derived from the SUSENAS 2014 round, which is the first SUSENAS round to ask households whether they are program beneficiaries and which is closest in time to the NVMS data. Undercoverage is the share of untreated eligible households at the district level and is defined as the number of eligible households that do not receive PKH divided by the total number of eligible households. Leakage is the share of treated ineligible households at the district level and is defined as the number of ineligible households receiving the CCT program divided by the total number of households receiving PKH.

of CCTs on social ties and community participation is rather mixed (Attanasio et al., 2009; Cameron et al., 2014; Attanasio et al., 2015), we investigate whether PKH has led to a decrease in community involvement, which in turn might have made program beneficiaries feel less guilty when committing violent crimes. We contrast four different measures of household-level engagement in community organizations and observe no decreases in any measure of community involvement by beneficiary households (panel E in Table 5.3).

Opportunity costs and income effects The role of opportunity costs and income effects as explanatory factors for criminal behavior has been widely studied in the economic crime literature, with a specific emphasis on young men (Jacob et al., 2003; Deming, 2011; Chin, 2018; Bell et al., 2022) and their labor market options (Phillips et al., 1972; Myers Jr., 1983; Gelber et al., 2015; Freedman et al., 2016; Bell et al., 2018). To what extent the CCT program affects young men's opportunity costs is theoretically ambiguous. On the one hand, by requiring school attendance of youth as a pre-condition of transfer receipt, the program might increase the opportunity costs to engage in criminal activities. On the other hand, the increase in household income might reduce the need for young men not bound by the program's conditionality criteria to take up jobs, thereby increasing the incidence and duration of unemployment, and thus freeing up time to commit violent crimes (Ludwig et al., 2001; Bratsberg et al., 2019).

In the following we examine the CCT program's impact on the time use of men (boys) and women (girls) during the last week before the interview, focusing on four indicators: (i) Market work, (ii) household chores, (iii) school attendance, and (iv) being idle (neither engaged in (i), (ii), nor (iii)). Table 5.4 shows the LATE estimates based on the RCT's impact evaluation (relying on equation (5.5)).[27] We split our sample by gender and age groups, distinguishing between (i) the targeted cohort (aged 7 to 15 years); (ii) young household members who were not directly targeted by the intervention (aged 18 to 25 years), but might have been subject to within-household spillover effects; and (iii) adult household members (aged 26 to 35 years).[28] We exclude children in the intermediate age range (16 to 17 years) as they might have been

[27] See section D.3.2 in the Appendix for more details on variable construction.

[28] Theoretically adult household members could also benefit from intra-household spillovers. However, as they are usually more attached to the labor market than younger household members, their labor supply and idleness might not adjust that strongly in response to additional transfer income.

still partially targeted by the program.[29] In line with Cahyadi et al. (2020), we find that the program was successful in increasing school enrollment of young children (aged 7 to 15 years) who were the target population of the CCT's conditionality criteria. Unsurprisingly, school attendance among the other non-targeted age groups in beneficiary households did not change. However, we also find marginally significant declines in the likelihood of market work among young men and women (aged 18 to 25) in response to cash transfer access in their households. Finally, CCT access resulted in a significant reduction in idleness in favour of schooling within the targeted cohort as well as in a 8.7 percentage point increase in idleness among young men (aged 18 to 25). The increase in youth idleness was an unintended side effect of the program and likely reflects intra-household spillovers towards non-targeted youth who live within beneficiary households.

TABLE 5.4: RCT: The short-run effects of the CCT program on work, schooling and idleness by cohort

	Men			Women		
Age group:	7–15 (1)	18–25 (2)	26–35 (3)	7–15 (4)	18–25 (5)	26–35 (6)
Working	-0.010	-0.077*	-0.029	-0.006	-0.094*	0.002
	(0.007)	(0.047)	(0.022)	(0.006)	(0.055)	(0.038)
Household chores	0.0001	0.003	0.003	-0.003	0.076	-0.006
	(0.002)	(0.007)	(0.005)	(0.003)	(0.047)	(0.037)
Attending school	0.084***	-0.015	0.002	0.045***	-0.013	-0.001
	(0.016)	(0.024)	(0.005)	(0.013)	(0.023)	(0.003)
Staying idle	-0.075***	0.087**	0.024	-0.034***	0.038	0.004
	(0.014)	(0.043)	(0.021)	(0.012)	(0.040)	(0.010)
Observations	8,760	2,977	3,804	8,087	2,651	5,182
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: The dependent variables refer to the last week before the survey and indicate whether an individual reported having engaged in any (i) work activities outside the household, (ii) household chores, (iii) school attendance, or (iv) stayed idle (not engaged in work, household chores, or school). The sample includes all households included in the *RCT sample*. Models instrument treatment at community level by treatment assignment at sub-district level relying on 2SLS. Age categories refer to the age of a person at the time of the endline survey. Controls include whether a household lives in an urban area, age, marriage status and educational degree indicators of the household head as well as quintiles of household size. Standard errors in parentheses are clustered at the sub-district level. */**/*** denote significance levels at 10/5/1% respectively.

Expected punishment costs Crime rates of young men have been shown to be sensitive to actual punishment costs (Levitt, 1998; Drago et al., 2012).

[29]Results for the 16 to 17 years old are insignificant throughout all specifications. Results are available from the authors upon request.

While the CCT program is unlikely to have altered the stringency of actual punishment or formal and informal sanctions and litigation costs, the program might have changed subjective perceptions of expected punishment costs (Sah, 1991). As discussed in Carpenter (2007), the consumption of drugs such as alcohol and narcotics can lead people to forget or underestimate punishment costs and consequently can increase crime. Bearing in mind that Indonesia is a predominantly Muslim country and that we do not have data on individual-level alcohol and drug consumption, we examine whether the program altered household expenditures on alcohol and legal drugs (Betel nuts which is a stimulant drug that is consumed by many Indonesians). As shown in panel D of Table 5.3, we find no changes in household spending on these consumption items.[30]

5.6 The plausibility of the idleness mechanism

Our empirical analysis demonstrated that the CCT led to an increase in community-level violent crime. Moreover, we showed that young men within beneficiary households became more likely to stay idle due to the program. At the same time, we were able to rule out a number of alternative potential mechanisms that may otherwise help explain our findings. Since official police reports consistently point to young men with lower education levels and poorer socio-economic backgrounds as a major demographic group who commits violent crimes in Indonesia, we speculate that PKH, despite its otherwise beneficial effects, allowed young men to have more time to engage in violent crimes. Next, we corroborate this hypothesis by investigating the mechanism from additional perspectives.

5.6.1 PKH and young men's idleness: External validity

Our analysis on potential mechanisms provided short-run estimates using household survey data from PKH's official impact evaluation (World Bank RCT). To understand whether adjustments in young men's idleness hold for a larger geographical setting, we link the roll-out of PKH to labor supply and idleness information from SUSENAS (for the years 2004 to 2011). We

[30]Results persist if we restrict the sample to non-Muslim areas only. Results are available from the authors upon request.

focus on the sample of individuals living in PKH-eligible households only—households with per capita household expenditures that are below the rural/urban provincial poverty line and with at least one household member in the age group of 0 to 15. We estimate TWFE models as specified in equation (5.4), but run them at the individual level by age group, with the dependent variables indicating that the surveyed individual was primarily engaged in (i) market work, (ii) household chores, (iii) schooling over the last week, or (iv) stayed idle (none of the above). The pooled models regress each dependent variable on the share of treated communities within a sub-district, conditional on sub-district and year fixed effects and further controls.

TABLE 5.5: The country-wide effects of the CCT program on work, schooling and idleness by cohort

	Men			Women		
Age group:	7–15 (1)	18–25 (2)	26–35 (3)	7–15 (4)	18–25 (5)	26–35 (6)
Working	-0.025*** (0.007)	-0.030* (0.017)	0.008 (0.009)	-0.021*** (0.006)	-0.049*** (0.018)	-0.024 (0.015)
Household chores	-0.021** (0.010)	-0.028** (0.013)	-0.018 (0.015)	-0.056*** (0.014)	-0.031* (0.019)	-0.005 (0.007)
Attending school	0.036*** (0.009)	-0.004 (0.009)	-0.001 (0.001)	0.038*** (0.010)	0.006 (0.011)	0.000 (0.001)
Staying idle	-0.016** (0.007)	0.028* (0.015)	-0.007 (0.008)	-0.014** (0.006)	0.018 (0.012)	0.003 (0.003)
Observations	82,730	41,357	38,087	75,546	33,493	59,525
Sub-district FE, year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: The dependent variables refer to the last week before the survey and indicate whether an individual reported having engaged in any (i) work activities outside the household, (ii) household chores, (iii) school attendance, or (iv) stayed idle (not engaged in work, household chores, or school). The treatment variable measures the share of PKH recipient communities within a sub-district and year. Results are based on pooled cross-sections of SUSENAS national household survey data (2004 to 2011), restricted to sub-districts that received the program by 2014. Only households below the provincial poverty line and with at least one child of eligible age are included. Controls include household head's age, gender, education, marital status and household size in quintiles. Standard errors in parentheses are clustered at the sub-district level. */**/*** denote significance levels at 10/5/1% respectively.

The nation-wide results in Table 5.5 re-confirm the RCT results (Table 5.4). Among children targeted by the program (aged 7 to 15 years), all forms of work and idleness decline in favour of school attendance. By contrast, within the cohort of non-targeted youth (aged 18 to 25), idleness increases with the program's presence within the sub-district, but only significantly among males. Shifting the CCT program's coverage from no to all communities within a sub-district is linked to an increase in idleness by 2.8 percentage points among male youth, while average idleness within this group lies at 21.8%. Importantly, placebo checks in Table D.16 in the Appendix show no significant increases in idleness of young males who live in non-eligible poor or in non-poor households. This ensures that we are not merely picking up regional differences in the dynamics of youth idleness that are spuriously correlated with program expansion. These placebo results also make it more likely that we observe adjustments of labor supply among CCT recipient households and not general job destruction among poor youth.

5.6.2 Idleness and crime: On the timing of criminal activities

Table 5.6 offers additional empirical support to the notion that increases in violent crime might be indeed linked to the idleness channel. Based on the newspaper reports, NVMS provides rich contextual information for each recorded instance of crime. This allows us to differentiate between criminal activities along two additional dimensions and re-estimate our main models—the LATE model from equation (5.3) for the RCT and the TWFE model from equation (5.1) for the national roll-out—by splitting criminal activities by their perpetrators and timing.

Columns 1–2 first contrast criminal activities that according to the NVMS have been committed by either individuals or larger groups of people. While both types of criminal activities increase in the *RCT sample*, in the nationwide *Roll-out sample* increases in crime tend to be dominated by individual activities.[31]

A second notable distinction refers to the timing of criminal activities. Incapacitation effects (through work or schooling) should be more relevant during days of the workweek, but their relative importance vanishes during weekends and public holidays, when a large part of the population is expected to stay at home. Our results confirm this expectation as we find that most of the crime effects arise during the workweek, whereas estimated coefficients turn small and insignificant during weekends. We consider crime committed during Indonesia's public holidays (when schools but also most workplaces are shut) separately in column 5. Public holidays follow a rolling window across the years (especially due to the shifting timing of the Muslim religious holidays like the Indonesian *Idul Fitri*). As expected, we also see no link between PKH and crimes committed during weekends or public holidays.

[31] Gang violence that is based on pre-existing disputes is classified as conflict in NVMS and does not respond to CCT coverage (cf. Table D.11 in the Appendix).

We believe that these latter findings are interesting and important for two reasons: First and in contrast to much of the existing crime CCT literature (Ludwig et al., 2001; Chioda et al., 2016; Machado et al., 2018), they seem to rule out that school-aged children are the main perpetrators of violent crime. Second, the evidence is in line with the view that violent crime is linked to an increase in the supply of potential perpetrators in CCT receiving communities (idle young men).[32]

TABLE 5.6: The effects of PKH by type and timing of violent crime

Dependent:	Crime committed by		Crime committed on		
	individ.	group	work*days*	week*ends*	public holidays
	(1)	(2)	(3)	(4)	(5)
			Panel A: RCT sample (2005–2010) LATE		
PKH	0.028**	0.031**	0.023**	-0.002	-0.001
	(0.012)	(0.013)	(0.011)	(0.009)	(0.005)
Mean (control)	0.093	0.124	0.078	0.036	0.013
			Panel B: Roll-out sample (2005–2015 TWFE		
PKH	0.008***	0.002	0.006**	0.002	0.001
	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)
Mean (control)	0.060	0.071	0.044	0.019	0.006
Community FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
District-year FE	Yes	Yes	Yes	Yes	Yes

Note: The *Roll-out sample* includes 28,873 communities and results are based on TWFE. The *RCT sample* comprises 1,830 communities and results are based on 2SLS (LATE). The dependent variable is a binary indicator that takes the value one if NVMS reported at least one incident of violent crime of a given type in a community in a given year. Columns 1–2 distinguish between violence committed by individual actors and those with group affiliations. Columns 3–5 distinguish between crimes that occurred on workdays, during the weekend (excluding public holidays), or on public holidays. Standard errors in parentheses are clustered at the level of sub-districts. */**/*** denote significance levels at 10/5/1% respectively.

5.6.3 Idleness and crime: A vignette experiment

Above we established that the CCT program increases idleness among young men and provided suggestive evidence that time allocation (on workdays vs. weekends or public holidays) might matter for committing crimes. To

[32] An alternative interpretation of our main results would have been that PKH leads school children to become more effective in committing crimes by attending schools with schools being an information and coordination hub to plan criminal activities (Jacob et al., 2003). If true, increases in school attendance would lead to an increase in organized and group-level criminal activities. Nonetheless, as shown in column 2 and 3 of Table 5.6, we do not find evidence for the role of schools in fostering criminal activities. In fact, violent crime incidences seem to be driven by single individuals and occur during workdays, which is compatible with our hypothesis of idleness among young men.

provide further quantitative evidence on the plausibility of the idleness-crime channel, we conducted an online vignette factorial survey experiment in Indonesia. The vignette experiment was designed to explicitly elicit whether a young person's idleness status affects Indonesians' perceptions of them being a likely perpetrator of crime.

We believe that the vignette experiment sheds light on two aspects: First, it provides country-specific evidence on whether Indonesians believe that the idleness of young men in their country is a crime-enabling factor. It therefore provides a local perspective on our narrative; a narrative that is derived from the economic literature covering other regions of the world.[33] Second, vignettes can provide insights beyond perceptions into real-world actions and behaviors. Widely used in political science (e.g., Hainmueller et al., 2015) and sociology (e.g., Auspurg et al., 2017), scholarly work in criminal sociology has shown—using vignette experiments—that laypersons are actually good in predicting the perpetrators of real-world crimes (Dressel et al., 2018; Lin et al., 2020). Bearing in mind that research has demonstrated that people's perceptions of and judgements on crime are subject to a number of biases at the individual level—for instance, due to stereotyping related to taste-based and statistical discrimination—there is nonetheless a robust and strong positive correlation between people's perception of crime suspects (whether a person committed a crime or not) and actual criminals. Therefore, we conjecture that if people relate youth idleness to a higher propensity to engage in criminal activities, this is because there is real-world manifestation of this relationship in Indonesia.

The vignette experiment was fielded as an online survey in Indonesia in January 2023 and was completed by 1,763 participants aged 18 to 50 years. The average person who completed the survey was about 32 years old, and about 48 percent of respondents were female.[34] The vignette consisted of a 2x2x3 factorial design, resulting in 12 unique texts. We used a between-subject design such that each survey participant randomly received one of the twelve texts to read.

Each vignette described a specific young man, Budi (aged 20), who lives with his parents and younger siblings in Central Java and was recently arrested by

[33]See, e.g., the previously highlighted economic literature on incapacitation effects of school children in Latin America (Chioda et al., 2016; Machado et al., 2018), the U.S. (Jacob et al., 2003; Foley, 2011), and Europe (Bratsberg et al., 2019).

[34]See Appendix D.5 for details on the survey, the vignette, descriptives on survey participants, balance checks, and empirical models.

the police as a suspect for having committed a theft. The vignette experiment randomized three factors, namely the suspect's idleness status (working full-time in a small kiosk vs. not working or attending school), the value of the loot (about USD 100 vs. 2,000), and the suspect's wealth (rich vs. very poor vs. very poor and member of a PKH household). Our main interest is in the first factor, whereas the third factor sheds light on whether the idleness-crime relationship differs across socio-economic sub-groups. The second factor is explicitly taken from the literature in criminal sociology that stresses that the value of the loot can matter in terms of which socio-economic profiles are linked to criminal activities.[35]

TABLE 5.7: Vignette Experiment: Likelihood of suspect committing the crime

Coefficient	Full sample		Sub-samples		
			Poor=1	PKH=1	Rich=1
	(1)	(2)	(3)	(4)	(5)
Idle	0.297***	0.041	0.414**	0.469**	0.056
	(0.102)	(0.170)	(0.176)	(0.184)	(0.170)
Poor	0.156	-0.029			
	(0.123)	(0.180)			
PKH	-0.015	-0.225			
	(0.125)	(0.181)			
Loot	-0.048	-0.047	-0.070	-0.196	0.176
	(0.102)	(0.102)	(0.178)	(0.184)	(0.170)
Idle x Poor		0.361			
		(0.245)			
Idle x PKH		0.413*			
		(0.250)			
Mean dependent	6.62	6.62	6.72	6.56	6.58
Observations	1,763	1,763	588	579	596

Note: The dependent variable measures the likelihood that respondents believe the described person committed the crime (theft) under investigation. It is coded on a 10-item Likert scale with "1" meaning "extremely unlikely to have committed the crime" and "10" meaning "extremely likely to have committed the crime". "Idle" is a binary indicator indicating whether person was described as idle (neither work nor school) or as working (working in a kiosk). "Poor" is a binary indicator indicating whether suspect is poor (without mentioning PKH). "PKH" is a binary indicator indicating whether suspect is poor and coming from a PKH-recipient family. "Loot" is a binary indicator indicating whether the stolen things were of high value. All specifications include two control variables (a respondent's age and gender) and were estimated by OLS using robust standard errors. */**/*** denote significance levels at 10/5/1% respectively.

Directly after the respondent had read the text, he/she was asked to rate on a 10-item Likert scale whether he/she believed that Budi has committed the

[35]We do not present results that include sub-treatment effects by the value of the loot since we do not find any economically sizeable and statistically significant relationships. The respective sub-group/sub-treatment effect results are available from the authors upon request. The experiment was pre-registered under the Open Science Framework (https://osf.io/y9exw).

crime. As shown in Appendix section D.5 Table D.28, respondents' replies range from 1 (extremely unlikely that Budi committed the crime) to 10 (extremely likely that Budi committed the crime), with people on average rating the suspect with 6.6 (median of 7).

We examine the role of idleness (Budi is neither working nor at school) on respondents' perceptions of whether the subject had committed the crime. Table 5.7 depicts our findings based on OLS regressions. In the full-sample (column 1) we find that describing the suspect as idle increases the respondent's perceptions of the suspect's guiltiness by about 0.3 points (statistically significant at the 1% level). By contrast, crime perceptions do not change with any of the other two factors. In column 2, we additionally interact idleness with the poverty status of the suspect. Results show that while idleness is not worsening crime perceptions about a wealthy subject, the effects of idleness increase substantially (and marginally significantly) among PKH recipients. The differential effect of the suspect being simply poor and idle is imprecisely estimated but does not differ from that of an idle poor person. Sub-sample results in columns 3–5 confirm these findings: Idleness increases the perceptions of the suspect being guilty if he is also poor, or poor and PKH-recipient, but not if he is rich.

Overall, the vignette experiment lends further support to our proposed mechanism. Our findings of the CCT program's impact on crime do not only coincide with an increase in young men's idleness, but these two dynamics are also likely to be causally related.

5.7 Conclusion

This paper shows that welfare programs can lead to increases in violent crime at the local level. Studying the massive roll-out of a conditional cash transfer (CCT) program—Program Keluarga Harapan (PKH)—that reached about 3 million Indonesian households by 2014, we find that the program resulted in an increase in the community-level incidence of violent crime by about 0.6 to 3.2 percentage points. These results are robust to a variety of econometric identification strategies and measurement issues.

Furthermore, our study sheds light on the mechanisms behind these findings. In particular, we show that the CCT increased idleness among non-targeted young male household members in the age range of 18 to 25, which could have provided them with an increased opportunity to engage in criminal activities. Importantly, our study can rule out that increases in violent crime are driven by the program increasing the expected benefits for committing crimes. Neither did it lead to increases in households' asset possession, nor did it alter local-level peer group inequalities. Likewise, we believe that our study can rule out a number of alternative channels related to the costs of engaging in crime and local-level general equilibrium effects.

Our study has identified relevant negative spillover effects of an otherwise beneficial policy. What can policy makers do in order to mitigate such negative effects? The idleness of young men is likely to play a role in explaining the CCT's impact on local-level violent crime in Indonesia. While we are less aware of policy interventions aimed at breaking the idleness-crime relationship, we focus our discussion on interventions that aim at reducing the incidence of idleness among the target population. Related policy interventions broadly fall into two categories that can be viewed as "carrot" or "stick" approaches. The former approach recognizes that the poor face persistent constraints across multiple dimensions (e.g., information, aspirations, networks, preferences) and acknowledges that a CCT's financial benefits might not be enough to improve the employment outcomes of young men. In this context, linking beneficiary families to active labor market policies such as workfare programs or skill training could be a viable intervention to reduce young men's idleness. In contrast, the latter approach is motivated by an assessment of the youth idleness-crime nexus as arising due to psychological constraints (e.g., self-control problems) or expectations of positive reciprocity. In this regard, interventions that limit individuals' choices or increase punishments for committing crimes might reduce the incidence of both idleness and crime. Such policies could for instance expand the conditionality criteria for schooling (or training) to include older children and adults (beyond the age of 15), add further conditionality dimensions such as the labor market participation of older household members, or lower cash transfers if household members engage in criminal activities. While possibly appearing too strict, the latter strand of policies can be important to generate continued support for CCTs among the general population (Gelbach et al., 2001; Baute et al., 2021).

Finally, we would like to point to some limitations of our study: First, the generalizability of our findings to other welfare programs and country contexts remains unclear. Second, our results provide evidence on short- to medium-term impacts only. Given that the program led to increases in the educational attainment of young children, we expect that its impact on violent crime fades out over time as labor market opportunities for youth improve over the medium to long term. Third, we lack micro data that contains a direct proof that young adults in CCT households increase their criminal behaviour. Consequently, the presented evidence is circumstantial. Lastly, we would like to point out that our key outcome variable, violent crime, cannot be readily compared to other studies. While we believe that crime rates measured by newspaper reports or victimization surveys hold several advantages over data from police reports, we would have ideally liked to compare our results with administrative data collected by the police. Unfortunately, in Indonesia this data is not yet available to researchers in a spatially disaggregated manner.

Chapter 6

Concluding Remarks

Promoting development and fighting poverty remain at the heart of the international community's policy agenda. Over the past decades, considerable achievements have been made, but concurrent challenges will make delivering on SDG 1 more complex in the future. Climate change, state fragility and conflicts are rampant and exemplary for phenomena closely intertwined with poverty, both as cause and consequence (Beegle et al., 2019). Against the backdrop of limited financial resources to address global developmental challenges, the need for an effective and targeted policy mix in the framework of international development cooperation is more urgent than ever. Research can play a decisive role in making such interventions more evidence-based and thus effective, as numerous recent advancements allow for unprecedented opportunities in policy assessments: Causal inference methods make precise estimates of policy impacts more credible; the growing scope of accessible data facilitates new analytical perspectives; and strengthened consensus among international donors for evidence-use all imply that demand for such assessments will rise.

A targeted policy mix should however not only account for available evidence on the initially specified—i.e. the desired—impact of interventions. A growing body of literature suggests that development policies often have unintended outcomes, inducing positive or negative side effects beyond their initial intention (Jabeen, 2016). While notable examples have put particular focus on negative consequences such as conflicts (e.g., Collier et al., 2007; Crost et al., 2014; Nunn et al., 2014), the scope of potential side effects is much more comprehensive (Koch et al., 2018). If such effects are not accounted for, policies' intended impacts can be significantly diminished or even reversed. A credibly evidence-based policy agenda therefore needs to be grounded on a more holistic assessment of interventions' outcomes, appealing action on the part of two groups: On the one hand, policy makers and practitioners have to strengthen demand for such analyses. On the other hand, researchers and evaluators will continuously need to fill the knowledge gap on the inadvertent outcomes of development policies.

In this thesis, I add evidence on the unintended consequences of three distinct Indonesian policy interventions along common phenomena experienced in developing countries: Chapter 3 shows that decentralization reforms altered land-use change incentives among local governments, resulting in temporarily decelerated deforestation rates. The identified mechanism behind this—from an environmental perspective—positive side effect is novel in the existing literature. Chapter 4 documents that a policy aimed at protecting migrant workers altered sending communities' capacity to cope with natural disasters, inducing significant poverty increases. The results are among the first empirical estimates on the effect of migration restrictions in the context of disaster-induced income shocks. Lastly, chapter 5 demonstrates how a large-scale social protection measure triggered idleness within recipient households, contributing towards increasing incidents of violent crime. By extending the analysis to Indonesia and identifying the underlying mechanism behind the crime hikes, the study fills several gaps in existing research on the welfare-crime nexus.

Considered independently, each chapter allows for conclusions that can inform policy design of comparable interventions: Decentralization reforms should consider the incentives of local governmental actors that are empowered as part of the policies. Precisely setting the legislative framework for local-level decisions can help for these reforms to deliver on the promise of improved public service delivery. Migration restrictions—particularly when issued by sending countries—should carefully weight costs and benefits of such measures against each other. Social assistance for communities hit heaviest by curtailed migratory options can help to absorb forfeited income from remittances. And conditional cash transfer schemes should recognize that payments can not only affect targeted household members, but also trigger behavioral adjustments via intra-household spillovers. Expanding conditionality criteria or improving labor market opportunities could prevent criminal activity as a result of increased idleness.

Considered together, the three chapters expand the knowledge on unintended consequences in the context of development policies. The evident observation is that interventions—even if well-established in one context—might have inadvertent impacts once implemented in another environment, making an approach tailored to local conditions essential. This perspective corresponds to results presented as part of introductory chapter 2, highlighting the role of individual implementation choices for the success of development policies supported by international donors. Results from this aggregate analysis and the three case studies thus complement each other and forge one key conclusion. In order to tackle international development- and poverty-related objectives, successful policy design has to consider induced impacts more holistically. This entails two potentially overlapping dimensions: A positive, yet unintended side effect as described in chapter 3 can help to inform which measures are particularly effective and cost-efficient. Adverse side effects as outlined in chapters 4 and 5 however can advise on whether interventions are potentially attenuated beyond its original benefit. Because such policies crucially depend on international cooperation, this conclusion underlines the role of decision makers in both international organizations and national governments. All relevant actors need to institutionalise a more evidence-based approach that takes into account potential unintended consequences.

The conclusion of my thesis explicitly draws attention to the necessity of policy design that is adapted to the local context. This is linked to the debate on the external validity of individual policy assessments, whether of intended or unintended impacts. In itself, each article of my thesis contributes towards the knowledge base in its respective field of research: While chapter 3 presents evidence on a previously unstudied relationship, chapter 4 empirically verifies existing theories, and chapter 5 extends existing literature—with conflicting results—to a new region. Even though Indonesia exhibits many characteristics and experiences phenomena that are considered typical of developing economies, these results cannot be readily translated to other countries. Future research needs to fill the knowledge gaps on unintended impacts on a much broader scale, which can then be synthesized to inform better policy making. This can be a key contribution to achieving developmental objectives that the international community has agreed upon in the framework of the SDGs.

Appendix A

Volume, risk, complexity: What makes development finance projects succeed or fail?

A.1 Tables

TABLE A.1: Representativeness of sample

	Non-sample	Sample	Absolute difference
Budget funds (in mil. EUR)	10.15	9.68	0.47
	(10.6)	(11.6)	(1.06)
Share of grants	0.89	0.90	-0.01
	(0.32)	(0.30)	(-0.77)
Disbursement vs. commitment	1.00	0.98	0.02
	(0.01)	(0.02)	(-0.50)
Time mandate to contract	0.45	0.38	0.07
	(0.86)	(0.83)	(1.79)
Project duration	6.51	7.04	-0.53
	(4.14)	(3.89)	(0.58)
Delay	0.19	0.19	0.00
	(0.40)	(0.39)	(0.15)
Observations	1,048	1,124	2,172

Note: The first two columns show the mean values of project characteristics of $N = 1,048$ out-of-sample and $N = 1,124$ in-sample observations. The former represent the sample of projects that were not randomly selected for evaluation in our sample period (cf. section 2.2). Means presented here are unweighted. Standard deviations shown in parentheses. Column 3 displays the absolute difference of the mean values in columns one and two. The t-statistic for testing equality of both means is displayed in parentheses below. */**/*** denote significance levels at 10/5/1% respectively.

TABLE A.2: Summary statistics

	Full Sample (1)	SSA (2)	Asia/Oceania (3)	Europe/Caucasus (4)	Lat. America (5)	MENA (6)
Financing						
Total volume (in million)	41.66	43.14	58.88	23.57	28.66	29.74
	(125.56)	(146.38)	(159.22)	(51.48)	(59.79)	(44.14)
% counterpart contributions	0.16	0.11	0.18	0.17	0.25	0.15
	(0.21)	(0.17)	(0.23)	(0.22)	(0.21)	(0.18)
Budget Funds (in million)	9.68	7.42	12.96	8.34	7.51	13.91
	(12.10)	(6.30)	(18.53)	(10.08)	(5.88)	(13.12)
% budget funds of ODA (x 1000)	19.79	15.41	21.01	23.81	22.50	24.40
	(29.65)	(18.87)	(38.60)	(33.46)	(27.90)	(31.23)
% project funds of GDP (x 1000)	122.00	157.70	121.09	121.73	51.13	87.22
	(266.94)	(279.71)	(370.38)	(169.39)	(68.22)	(119.43)
Disbursement vs. commitment	0.98	0.99	0.97	0.98	1.00	0.95
	(0.11)	(0.08)	(0.15)	(0.12)	(0.00)	(0.15)
Structure						
Co-financing	0.21	0.30	0.15	0.16	0.20	0.11
	(0.41)	(0.46)	(0.36)	(0.37)	(0.40)	(0.32)
Accompanying measure	0.27	0.20	0.30	0.50	0.17	0.28
	(0.44)	(0.40)	(0.46)	(0.50)	(0.38)	(0.45)
Previous cooperation	0.23	0.29	0.23	0.05	0.17	0.31
	(0.42)	(0.45)	(0.42)	(0.21)	(0.38)	(0.47)
Number of institutions	4.00	4.26	3.67	3.69	4.46	3.79
	(2.52)	(2.66)	(2.02)	(2.37)	(3.01)	(2.52)
Project manager turnover	0.48	0.46	0.49	0.52	0.47	0.51
	(0.38)	(0.38)	(0.30)	(0.47)	(0.35)	(0.33)
Country office	0.46	0.36	0.59	0.47	0.38	0.57
	(0.50)	(0.48)	(0.49)	(0.50)	(0.49)	(0.50)
Complexity						
Project duration	7.04	6.82	7.09	5.69	7.55	8.97
	(3.71)	(3.30)	(3.75)	(3.11)	(4.11)	(4.24)
Delay indicator	0.23	0.19	0.27	0.12	0.21	0.48
	(0.42)	(0.39)	(0.44)	(0.33)	(0.41)	(0.50)
Revised ToC	0.37	0.35	0.37	0.33	0.38	0.44
	(0.48)	(0.48)	(0.48)	(0.47)	(0.49)	(0.50)
Technical complexity	0.48	0.35	0.67	0.63	0.15	0.65
	(0.50)	(0.48)	(0.47)	(0.48)	(0.36)	(0.48)
Risks						
Number ex-ante identified risks	3.99	3.93	4.25	3.67	4.00	4.05
	(2.02)	(1.94)	(2.24)	(1.88)	(2.04)	(1.84)
% ex-ante identified risks occurred	0.55	0.62	0.52	0.49	0.53	0.53
	(0.35)	(0.34)	(0.35)	(0.36)	(0.34)	(0.34)
Macro						
GDP p.c. growth (annual)	3.29	2.47	4.98	4.54	2.37	1.67
	(2.97)	(2.30)	(2.50)	(4.18)	(1.24)	(3.19)
Freedom House Democracy score	4.00	4.05	3.43	4.46	5.30	2.88
	(1.45)	(1.34)	(1.65)	(1.01)	(0.67)	(0.85)
State Fragility Index	12.19	15.00	12.91	7.21	8.46	11.02
	(4.71)	(3.85)	(3.71)	(2.96)	(3.38)	(4.58)
Sectors						
Agr. & Env.	0.13	0.13	0.12	0.07	0.34	0.03
Budget Support	0.02	0.04	0.00	0.00	0.02	0.00
Education	0.07	0.07	0.09	0.02	0.04	0.13
Energy	0.09	0.04	0.14	0.17	0.07	0.08
Finance	0.14	0.07	0.17	0.24	0.15	0.18
Governance	0.08	0.10	0.03	0.09	0.09	0.08
Health	0.13	0.19	0.19	0.03	0.03	0.05
Transportation	0.06	0.08	0.10	0.02	0.02	0.01
Water Supply	0.18	0.17	0.08	0.28	0.14	0.32
Other	0.09	0.11	0.07	0.07	0.10	0.12
Observations	1,124	428	281	167	116	122

Note: Table shows mean and standard deviation in parentheses of covariates, excluding categorial variables and population of the country. % budget funds of ODA and % projects fund of GDP are re-scaled by one thousand. Sector figures correspond to fraction of the respective sector in each sub-sample. Observations are weighted by the inverse of the number of projects evaluated in the corresponding evaluation report. The mean value and standard deviation of the Freedom House Democracy Score and the State Fragility Index is calculated from only 1,096 observations in the full sample due to unavailable data. Cross-regional projects ($N = 10$) are not shown as individual sub-sample but are included in the full sample figures (column 1). Abbreviations: SSA = Sub-Saharan Africa; MENA = Middle East and North Africa.

TABLE A.3: Codebook: Outcome and project variables

Label	Description	Source
Outcome variables		
Rating (pooled)	Discrete variable on a scale from 1 to 6. Each of the five DAC ratings of each project is considered as an individual observation	Evaluation reports.
Rating	Overall rating. Differs from mean because: (i) KO-criteria, for a project to be successful, sub-ratings "sust.", "impact" and "effect." must be successful. (ii) Differing weights for certain sub-ratings for most projects	Evaluation reports.
Rating (mean)	Continuous variable ranging 1 through 6. Mean of reported DAC ratings.	Evaluation reports.
Relevance	Discrete variable on a scale from 6 to 1.	Evaluation reports.
Effectiveness	Discrete variable on a scale from 6 to 1.	Evaluation reports.
Efficiency	Discrete variable on a scale from 6 to 1.	Evaluation reports.
Impact	Discrete variable on a scale from 6 to 1.	Evaluation reports.
Sustainability	Discrete variable on a scale from 6 to 1.	Evaluation reports.
Micro variables		
Aid type	Categorical variable. Aid types: (1) Loan, (2) Grant	Evaluation reports.
Total volume (log)	Discrete variable. Sum of project volume and counterpart contribution. In current EUR million, logarithmized.	Evaluation reports.
Budget funds (log)	Discrete variable. Budget funds in current EUR million, logged.	KfW data.
Share of counterpart contribution	Discrete variable. Share of counterpart contribution relative to investment volume.	KfW data.
% budget Funds of ODA	Continuous variable. Share of budget funds relative to (current) EUR value of ODA commitments (grants) in year of project start (x 1 million).	KfW data and OECD Statistics (OECD, 2022a).
% project funds of GDP	Continuous variable. Share of project volume relative to the country's GDP in the year of project start (multiplied by 1 billion). If the project operates in more than one country, the sum of GDP across countries is taken as reference value.	KfW data and WB/OECD Nat. Accounts (World Bank, 2023).
Disbursement vs. commitment	Continuous variable. Share of disbursements relative to commitments.	KfW data.
Co-financing	Binary variable. Indicates co-financing: 0 = no. 1 = yes.	Evaluation reports.
Accompanying measure	Binary variable. Indicates accompanying measure. 0 = no. 1 = yes.	Evaluation reports.
Agency type	Categorical variable. Type of programme executing agency.	Evaluation reports.
Number of institutions	Discrete (integer) variable. Number of institutions involved.	Evaluation reports.
Previous cooperation	Binary variable. Previous cooperation with implementing agency.	KfW data.
Project manager turnover	Continuous variable. Number of project managers assigned to project divided by project duration.	KfW data.
Country office	Binary variable for presence of KfW office in project country during entire project implementation	KfW data.
Number of ex ante identified risks	Discrete variable. Count of ex-ante identified risks.	Evaluation reports.
% ex ante identified risks that occurred	Continuous variable. Share of ex-ante identified risks that materialised.	Evaluation reports.
Overall risk	Categorical variable. Ex-ante overall project risk assessment by project manager	KfW data.
Overall risk control	Categorical variable. Ex-ante overall project risk assessment of controllability by project manager	KfW data.
Project duration (log)	Discrete variable. Year of first contract signed until year of final review, logarithmized.	KfW data.
Delay Indicator	Binary variable. "Delay" is identified as follows: (i) Identification of 80th percentile of project duration for each sector in 10-year intervals. (ii) Projects are classified "delayed" if project duration was longer than the 80th percentile of project duration in the sector. (iii) Sectoral comparison group needs to meet the following criteria: (a) Projects with project start year >= 1990 and (b) finished projects (existing final review except for financial sector).	Evaluation reports.
Revised ToC	Binary variable. This variable indicates whether the Theory of Change was revised, characterized by either or both of the following: (i) Impact objective of the project has changed. (ii) Outcome objective of the project has changed.	Evaluation reports.
Years mandate to contract	Discrete variable. Time between mandate and first contract in years.	KfW data.
Technical complexity	Binary variable. Indicates involvement of technical expert: 0 = no. 1 = yes.	Evaluation reports.

TABLE A.4: Codebook: Macro, control and analytical variables

Label	Description	Source
Macro variables		
Population	Logarithm of population.	World Development Indicators (World Bank, 2023).
GDP p.c. growth (annual)	GDP per capita growth (annual %).	WB/OECD Nat. Accounts (World Bank, 2023).
Democracy	Mean of score for: Political rights and Civil liberties Interpretation: Smaller values imply greater freedom	Freedom House (Freedom House, 2022)
Fragility Index	State Fragility Index as sum of Effectiveness Score + Legitimacy Score Interpretation: Higher value corresponds to higher fragility	Center for Systemic Peace (Integrated Network for Societal Conflict Research (INSCR), 2018).
Net ODA (commitments: Grants)	Net Official Development Aid (ODA) commitments (grants only) in the year of project start. If a project operated in more than one country, the value of this variable corresponds to the sum of ODA in respective countries.	OECD Statistics (OECD, 2022a).
Evaluation and Control variables		
Sector	Projects are either assigned one of nine sectors or classified as "other" if none of the sectors is applicable. List of sectors: Budget Support, education, energy, finance, health & population, governance, agriculture & environment, transportation, Water Supply.	Evaluation reports.
Region	Region of the project. List of regions: SSA, Asia/Oceania, Europe/Caucasus, Latin America, MENA. Projects operated in multiple regions are labeled as "Cross-regional".	Evaluation reports.
Time between final review and EPE	Discrete variable. Time between final review and ex post evaluation (financial sector: year of last disbursement). In case the final review is dated after the evaluation year (33 instances), the value is set to zero. Unit: years	Evaluation reports.
Year of project start	Categorical variable. 5-year intervals of year of project start. First interval: 1990–1994. Last interval: 2015–2019.	KfW data.
Year of evaluation	Categorical variable. 5-year intervals of year of ex post evaluation. First interval: 2005–2009. Last interval: 2020–2024.	Evaluation reports.
Analytical variables		
Report weight	Inverse of number of projects evaluated in the same report.	Evaluation reports.

TABLE A.5: Results: Regional split

Dep. variable: Rating (pooled)	(1) All	(2) SSA	(3) Asia/Oceania	(4) Europe/Caucasus	(5) Lat. America	(6) MENA
Financing						
Total volume (log)	0.037	0.030	-0.003	0.107	0.074	0.160*
	(0.029)	(0.041)	(0.054)	(0.070)	(0.054)	(0.082)
Aid type (Base: Loan):						
-Grant	0.105	1.151***	-0.157	-0.092	0.359**	-0.536**
	(0.087)	(0.271)	(0.136)	(0.153)	(0.147)	(0.215)
% counterpart contributions	0.145	0.005	0.232	0.322	0.457	1.665***
	(0.118)	(0.191)	(0.238)	(0.345)	(0.430)	(0.256)
Budget funds (log)	0.095**	0.047	0.140**	0.179**	-0.025	0.210*
	(0.042)	(0.063)	(0.069)	(0.086)	(0.085)	(0.122)
% budget funds of ODA	-0.000	-0.000	-0.000	-0.000***	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
% project funds of GDP	-0.000	0.000	-0.000	-0.000*	-0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Disbursement vs. commitment	0.137	-0.398	0.457	0.048	-	-
	(0.156)	(0.338)	(0.309)	(0.394)	(-)	(-)
Structure						
Co-financing	0.002	-0.006	-0.104	0.298**	-0.183	-0.365***
	(0.064)	(0.082)	(0.137)	(0.133)	(0.200)	(0.098)
Accompanying measure	-0.015	0.105	-0.123	-0.139*	-0.016	-0.502***
	(0.056)	(0.090)	(0.107)	(0.071)	(0.165)	(0.182)
Agency type (Base: NGO):						
-Mixed	-0.099	-0.324*	0.074	0.325	-0.073	-0.043
	(0.130)	(0.173)	(0.253)	(0.281)	(0.199)	(0.593)
-Multilateral	-0.009	-0.184	0.776	0.196	0.133	0.362
	(0.131)	(0.200)	(0.520)	(0.360)	(0.295)	(0.255)
-Private sector	0.006	-0.282	0.467	-0.247	0.139	2.156***
	(0.139)	(0.213)	(0.329)	(0.286)	(0.344)	(0.416)
-Government	-0.101	-0.313**	0.277	-0.226	-0.162	-0.209
	(0.107)	(0.142)	(0.219)	(0.259)	(0.182)	(0.464)
Previous cooperation	0.066	-0.011	0.050	0.399**	0.107	0.279**
	(0.051)	(0.071)	(0.108)	(0.195)	(0.125)	(0.116)
Number of institutions	0.005	0.011	-0.011	-0.050**	-0.031*	0.019
	(0.009)	(0.015)	(0.025)	(0.023)	(0.017)	(0.031)
Project manager turnover	0.328	0.702	0.694	-0.590***	0.511	-0.770***
	(0.248)	(0.533)	(0.449)	(0.181)	(0.809)	(0.216)
Country office	-0.043	-0.070	-0.047	0.003	0.234*	0.490**
	(0.056)	(0.111)	(0.122)	(0.118)	(0.119)	(0.194)
Complexity						
Project duration (log)	-0.149**	0.124	-0.037	-0.372**	-0.564***	-0.143
	(0.075)	(0.150)	(0.135)	(0.148)	(0.149)	(0.119)
Delay indicator	0.009	0.001	-0.149	0.167	0.466**	-0.069
	(0.069)	(0.104)	(0.125)	(0.216)	(0.228)	(0.118)
Revised ToC	-0.048	-0.034	-0.035	-0.265***	-0.003	0.447***
	(0.047)	(0.079)	(0.095)	(0.086)	(0.113)	(0.131)
Years mandate to contract	-0.048*	-0.065	0.027	-0.109*	-0.038	0.048
	(0.027)	(0.070)	(0.037)	(0.060)	(0.034)	(0.097)
Technical complexity	-0.130**	-0.227***	0.042	0.090	-0.267*	0.131
	(0.055)	(0.076)	(0.121)	(0.161)	(0.151)	(0.177)
Risks						
Number ex-ante identified risks	0.001	-0.025	-0.013	-0.024	0.040	-0.040
	(0.013)	(0.021)	(0.025)	(0.029)	(0.026)	(0.026)
% ex-ante identified risks occurred	-0.486***	-0.525***	-0.478***	-0.264**	-0.839***	-0.412**
	(0.067)	(0.119)	(0.135)	(0.122)	(0.182)	(0.159)
Overall risk (base: low)						
-Medium	-0.203**	-0.523***	0.122	-0.534**	0.480	-0.202
	(0.082)	(0.145)	(0.172)	(0.212)	(0.297)	(0.396)
-(Very) high	-0.352***	-0.655***	-0.131	-0.572**	0.436	-0.197
	(0.088)	(0.149)	(0.199)	(0.220)	(0.305)	(0.410)
-Not assigned	-0.219*	-0.305*	-0.173	-0.245	0.120	-0.479
	(0.116)	(0.174)	(0.208)	(0.257)	(0.341)	(0.557)
Overall risk control (base: low)						
-Medium	0.084	0.125	0.212*	0.167	-0.129	0.048
	(0.058)	(0.095)	(0.112)	(0.144)	(0.139)	(0.247)
-High	-0.061	0.058	-0.657***	0.760***	-	-0.189
	(0.169)	(0.230)	(0.228)	(0.274)	(-)	(0.744)
Macro variables						
GDP p.c. growth (annual)	0.011	0.016	0.042*	0.010	0.146***	0.025
	(0.008)	(0.015)	(0.022)	(0.012)	(0.038)	(0.022)
Freedom House Democracy score	-0.018	0.001	-0.019	-0.208***	-0.490***	-0.036
	(0.021)	(0.049)	(0.037)	(0.064)	(0.120)	(0.066)
State Fragility Index	-0.006	-0.017	0.048***	0.029	-0.072**	0.002
	(0.008)	(0.018)	(0.018)	(0.029)	(0.032)	(0.023)
Population log	-0.029	0.050	-0.074**	-0.043	-0.002	-0.281**
	(0.022)	(0.059)	(0.036)	(0.062)	(0.052)	(0.134)
Sector indicators	Yes	Yes	Yes	Yes	Yes	Yes
Sub-rating indicators	Yes	Yes	Yes	Yes	Yes	Yes
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,458	2,136	1,401	804	580	487
Adjusted R^2	0.23	0.29	0.27	0.51	0.54	0.57

Note: Table entries are coefficients from WLS regressions with the pooled rating as dependent variable. Weights are given by the inverse of the number of projects evaluated in the corresponding evaluation report. Observations are weighted by the inverse of the number of projects evaluated in the corresponding evaluation report. Other control variables include: Number of years between final project inspection and evaluation; the year of project start as well as evaluation year (both 5-year intervals); evaluation month. Standard errors in parentheses clustered at the country-evaluation-year level. */**/*** denote significance levels at 10/5/1% respectively.

TABLE A.6: Results: Sectoral split I

Dep. variable: Rating (pooled)	(1) Full Sample	(2) Agr. & Env.	(3) Education	(4) Energy	(5) Finance	(6) Health
Financing						
Total volume (log)	0.016	0.084	-0.035	0.220***	-0.060	-0.020
	(0.029)	(0.092)	(0.061)	(0.046)	(0.068)	(0.054)
% counterpart contributions	0.191	1.175***	0.177	0.500**	-0.099	-0.087
	(0.117)	(0.433)	(0.400)	(0.234)	(0.402)	(0.210)
Budget funds (log)	0.115***	0.045	0.066	0.287***	0.064	-0.035
	(0.041)	(0.114)	(0.158)	(0.079)	(0.089)	(0.118)
% budget funds of ODA	-0.000	-0.000**	-0.000	-0.000***	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
% project funds of GDP	-0.000	0.000	0.000***	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000))
Structure						
Co-financing	0.024	-0.136	-0.285*	0.155	0.005	-0.031
	(0.062)	(0.178)	(0.158)	(0.116)	(0.213)	(0.130)
Accompanying measure	-0.028	-0.001	0.315	0.063	0.099	-0.211
	(0.056)	(0.142)	(0.195)	(0.100)	(0.144)	(0.182)
Number of institutions	0.002	-0.051*	-0.009	-0.119***	-0.029	-0.007
	(0.009)	(0.026)	(0.022)	(0.028)	(0.033)	(0.022)
Project manager turnover	0.368	2.247*	-2.775***	0.195	-0.137	0.152
	(0.264)	(1.165)	(0.757)	(0.445)	(0.352)	(0.790)
Country office	-0.054	0.025	-0.197	-0.490***	-0.036	-0.081
	(0.056)	(0.168)	(0.260)	(0.156)	(0.159)	(0.112)
Complexity						
Project duration (log)	-0.138*	0.241	-0.100	-0.344*	-0.463***	0.362**
	(0.076)	(0.307)	(0.249)	(0.191)	(0.150)	(0.155)
Delay indicator	-0.190***	-0.131	-0.618***	0.160	-0.287	-0.285*
	(0.065)	(0.219)	(0.224)	(0.148)	(0.226)	(0.167)
Revised ToC	-0.066	0.095	-0.229*	-0.026	-0.043	-0.068
	(0.047)	(0.134)	(0.128)	(0.122)	(0.192)	(0.087)
Years mandate to contract	-0.050*	0.064	0.477***	0.122***	-0.065	-0.063
	(0.026)	(0.078)	(0.148)	(0.041)	(0.067)	(0.080)
Technical complexity	-0.123**	-0.361	0.046	-0.226	0.323	-0.112
	(0.054)	(0.223)	(0.201)	(0.155)	(0.284)	(0.132)
Risks						
Number ex-ante identified risks	-0.002	0.030	-0.127***	-0.037**	-0.003	-0.027
	(0.013)	(0.041)	(0.043)	(0.019)	(0.041)	(0.030)
% ex-ante identified risks occurred	-0.503***	-0.669***	-0.841***	0.337*	-0.327	-0.549***
	(0.070)	(0.202)	(0.173)	(0.186)	(0.205)	(0.157)
Macro variables						
GDP p.c. growth (annual)	0.012	0.023	0.010	-0.033**	0.005	0.009
	(0.008)	(0.032)	(0.021)	(0.016)	(0.014)	(0.027)
Freedom House Democracy score	-0.020	0.006	-0.054	-0.067*	-0.042	-0.052
	(0.022)	(0.059)	(0.087)	(0.038)	(0.058)	(0.042)
State Fragility Index	-0.008	-0.065**	-0.078**	-0.013	0.027	-0.029*
	(0.008)	(0.025)	(0.034)	(0.020)	(0.026)	(0.015)
Population (log)	-0.029	0.061	0.125**	-0.208***	-0.039	0.057
	(0.022)	(0.064)	(0.054)	(0.040)	(0.062)	(0.070)
Region indicators	Yes	Yes	Yes	Yes	Yes	Yes
Sub-rating indicators	Yes	Yes	Yes	Yes	Yes	Yes
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,283	654	370	455	815	803
Adjusted R^2	0.22	0.38	0.46	0.50	0.30	0.39

Note: Table entries are coefficients from WLS regressions with the pooled rating as dependent variable. Weights are given by the inverse of the number of projects evaluated in the corresponding evaluation report. Reduced set of covariates (exclusion of categorical variables, "Disbursement vs. commitment" and "Previous cooperation") due to lack of variation in small sub-samples. Sector "Budget support" is not displayed as sub-sample for the same reason. Other control variables include: Number of years between final project inspection and evaluation; the year of project start as well as evaluation year (both 5-year intervals); evaluation month. Standard errors in parentheses are clustered at the country-evaluation-year level. */**/*** denote significance levels at 10/5/1% respectively.

TABLE A.7: Results: Sectoral split II

Dep. variable: Rating (pooled)	(1) Full Sample	(2) Governance	(3) Transportation	(4) Water Supply	(5) Other
Financing					
Total volume (log)	0.016	-0.150***	-0.022	0.056	0.149**
	(0.029)	(0.051)	(0.072)	(0.089)	(0.063)
% counterpart contributions	0.191	0.445*	0.493	0.596*	-0.814***
	(0.117)	(0.264)	(0.341)	(0.346)	(0.255)
Budget funds (log)	0.115***	0.228**	-0.562***	0.159	-0.079
	(0.041)	(0.087)	(0.147)	(0.140)	(0.081)
% budget funds of ODA	-0.000	-0.000	0.000**	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
% project funds of GDP	-0.000	0.000	0.000***	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Structure					
Co-financing	0.024	-0.030	-0.236	0.161	0.054
	(0.062)	(0.230)	(0.166)	(0.208)	(0.142)
Accompanying measure	-0.028	-0.408***	-0.063	-0.030	-0.800***
	(0.056)	(0.140)	(0.260)	(0.112)	(0.143)
Number of institutions	0.002	0.009	0.251***	0.014	-0.007
	(0.009)	(0.017)	(0.045)	(0.027)	(0.021)
Project manager turnover	0.368	-0.499	0.810	-0.244	-0.107
	(0.264)	(0.528)	(1.361)	(0.496)	(0.608)
Country office	-0.054	-0.025	0.839***	-0.144	-0.152
	(0.056)	(0.126)	(0.255)	(0.150)	(0.096)
Complexity					
Project duration (log)	-0.138*	0.061	-0.007	-0.466**	-0.049
	(0.076)	(0.141)	(0.277)	(0.233)	(0.137)
Delay indicator	-0.190***	0.042	-0.492	0.158	-0.339**
	(0.065)	(0.122)	(0.306)	(0.172)	(0.146)
Revised ToC	-0.066	0.042	-0.252	-0.235**	-0.197*
	(0.047)	(0.144)	(0.170)	(0.106)	(0.107)
Years mandate to contract	-0.050*	-0.336***	0.080	-0.168**	0.153**
	(0.026)	(0.069)	(0.125)	(0.069)	(0.075)
Technical complexity	-0.123**	-0.221*	-0.021	-0.145	-0.480***
	(0.054)	(0.129)	(0.229)	(0.185)	(0.094)
Risks					
Number ex-ante identified risks	-0.002	0.014	-0.031	0.004	0.065**
	(0.013)	(0.030)	(0.046)	(0.028)	(0.031)
% ex-ante identified risks occurred	-0.503***	-0.473**	-0.670***	-0.735***	-0.362**
	(0.070)	(0.214)	(0.150)	(0.149)	(0.168)
Macro variables					
GDP p.c. growth (annual)	0.012	0.012	0.001	0.013	-0.034
	(0.008)	(0.014)	(0.024)	(0.028)	(0.031)
Freedom House Democracy score	-0.020	0.091	-0.245**	-0.056	0.020
	(0.022)	(0.092)	(0.092)	(0.065)	(0.044)
State Fragility Index	-0.008	0.016	-0.043	-0.000	-0.017
	(0.008)	(0.022)	(0.039)	(0.018)	(0.019)
Population (log)	-0.029	-0.077	0.264***	-0.021	-0.043
	(0.022)	(0.062)	(0.082)	(0.062)	(0.044)
Region indicators	Yes	Yes	Yes	Yes	Yes
Sub-rating indicators	Yes	Yes	Yes	Yes	Yes
Other control variables	Yes	Yes	Yes	Yes	Yes
Observations	5,283	503	340	824	519
Adjusted R^2	0.22	0.54	0.59	0.37	0.53

Note: Table entries are coefficients from WLS regressions with the pooled rating as dependent variable. Weights are given by the inverse of the number of projects evaluated in the corresponding evaluation report. Reduced set of covariates (exclusion of categorical variables, 'Disbursement vs. commitment' and 'Previous cooperation') due to lack of variation in small sub-samples. Sector 'Budget support' is not displayed as sub-sample for the same reason. Other control variables include: Number of years between final project inspection and evaluation; the year of project start as well as evaluation year (both 5-year intervals); evaluation month. Standard errors in parentheses are clustered at the country-evaluation-year level. */**/*** denote significance levels at 10/5/1% respectively.

TABLE A.8: Results: OECD DAC ratings

Dep. variable:	(1) Overall	(2) Relevance	(3) Efficiency	(4) Effectiveness	(5) Impact	(6) Sustainability
Financing						
Total volume (log)	0.037	0.015	0.047	0.053	0.014	0.055*
	(0.029)	(0.034)	(0.041)	(0.043)	(0.036)	(0.032)
Aid type (Base: Loan):						
-Grant	0.105	0.034	0.147	0.155	0.146	0.039
	(0.087)	(0.102)	(0.126)	(0.126)	(0.115)	(0.106)
% counterpart contributions	0.145	0.156	-0.019	0.283*	0.213	0.100
	(0.118)	(0.152)	(0.166)	(0.163)	(0.171)	(0.148)
Budget funds (log)	0.095**	0.054	0.092*	0.133**	0.127***	0.067
	(0.042)	(0.052)	(0.054)	(0.059)	(0.048)	(0.048)
% budget funds of ODA	-0.000	0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
% project funds of GDP	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Disbursement vs. commitment	0.137	0.019	0.212	0.207	0.184	0.048
	(0.156)	(0.178)	(0.250)	(0.267)	(0.228)	(0.261)
Structure						
Co-financing	0.002	-0.064	-0.009	0.028	0.076	-0.013
	(0.064)	(0.081)	(0.082)	(0.088)	(0.084)	(0.072)
Accompanying measure	-0.015	0.032	0.001	-0.062	-0.007	-0.040
	(0.056)	(0.062)	(0.078)	(0.079)	(0.079)	(0.068)
Agency type (Base: NGO):						
-Mixed	-0.099	-0.038	-0.168	-0.103	-0.078	-0.106
	(0.130)	(0.142)	(0.191)	(0.181)	(0.169)	(0.138)
-Multilateral	-0.009	0.123	-0.074	-0.191	0.040	0.051
	(0.131)	(0.156)	(0.206)	(0.175)	(0.177)	(0.153)
-Private sector	0.006	-0.116	0.023	0.048	0.062	0.010
	(0.139)	(0.159)	(0.202)	(0.185)	(0.186)	(0.147)
-Government	-0.101	-0.079	-0.128	-0.055	-0.097	-0.144
	(0.107)	(0.119)	(0.160)	(0.150)	(0.146)	(0.106)
Previous cooperation	0.066	-0.025	0.131*	0.027	0.086	0.115**
	(0.051)	(0.059)	(0.074)	(0.074)	(0.069)	(0.058)
Number of institutions	0.005	0.017	0.003	-0.003	0.019	-0.012
	(0.009)	(0.011)	(0.013)	(0.011)	(0.012)	(0.010)
Project manager turnover	0.328	0.237	0.258	0.640	0.631*	-0.129
	(0.248)	(0.277)	(0.278)	(0.395)	(0.322)	(0.252)
Country office	-0.043	-0.000	0.024	-0.128*	-0.048	-0.059
	(0.056)	(0.067)	(0.076)	(0.076)	(0.076)	(0.065)
Complexity						
Project duration (log)	-0.149**	-0.077	-0.150	-0.178*	-0.032	-0.314***
	(0.075)	(0.084)	(0.107)	(0.101)	(0.097)	(0.082)
Delay indicator	0.009	0.121	-0.124	-0.086	0.121	0.018
	(0.069)	(0.082)	(0.093)	(0.092)	(0.089)	(0.081)
Revised ToC	-0.048	-0.106*	-0.018	-0.028	-0.110*	0.026
	(0.047)	(0.058)	(0.066)	(0.064)	(0.066)	(0.054)
Years mandate to contract	-0.048*	-0.038	-0.034	-0.082***	-0.058	-0.028
	(0.027)	(0.031)	(0.038)	(0.031)	(0.038)	(0.033)
Technical complexity	-0.130**	-0.013	-0.215***	-0.109	-0.215***	-0.099
	(0.055)	(0.064)	(0.081)	(0.078)	(0.078)	(0.063)
Risks						
Number ex-ante identified risks	0.001	0.001	-0.002	-0.002	0.014	-0.004
	(0.013)	(0.014)	(0.018)	(0.018)	(0.016)	(0.014)
% ex-ante identified risks occurred	-0.486***	-0.264***	-0.593***	-0.558***	-0.480***	-0.537***
	(0.067)	(0.081)	(0.092)	(0.097)	(0.094)	(0.075)
Overall risk (base:low)						
-Medium	-0.203**	-0.011	-0.326**	-0.131	-0.260**	-0.284**
	(0.082)	(0.098)	(0.139)	(0.111)	(0.121)	(0.111)
-(Very) high	-0.352***	-0.080	-0.494***	-0.344***	-0.435***	-0.407***
	(0.088)	(0.112)	(0.146)	(0.120)	(0.133)	(0.117)
-Not assigned	-0.219*	0.017	-0.349*	-0.124	-0.259	-0.384**
	(0.116)	(0.149)	(0.178)	(0.162)	(0.174)	(0.151)
Overall risk control (base: low)						
-Medium	0.084	0.121*	0.050	0.103	0.160*	-0.022
	(0.058)	(0.070)	(0.082)	(0.081)	(0.083)	(0.073)
-High	-0.061	0.140	-0.300	-0.085	0.001	-0.060
	(0.169)	(0.219)	(0.239)	(0.169)	(0.226)	(0.211)
Macro variables						
GDP p.c. growth (annual)	0.011	0.003	0.020*	0.002	0.014	0.018**
	(0.008)	(0.008)	(0.011)	(0.011)	(0.011)	(0.009)
Freedom House Democracy score	-0.018	0.029	-0.034	-0.039	-0.030	-0.016
	(0.021)	(0.024)	(0.031)	(0.028)	(0.031)	(0.025)
State Fragility Index	-0.006	0.016	-0.007	-0.006	-0.009	-0.020**
	(0.008)	(0.010)	(0.011)	(0.011)	(0.011)	(0.009)
Population (log)	-0.029	-0.029	-0.047	-0.030	-0.037	-0.003
	(0.022)	(0.026)	(0.029)	(0.028)	(0.031)	(0.025)
Sub-rating indicators	Yes					
Sector and region indicators	Yes	Yes	Yes	Yes	Yes	Yes
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,458	1,092	1,094	1,093	1,093	1,086
Adjusted R^2	0.23	0.07	0.15	0.16	0.16	0.21

Note: Table entries are coefficients from WLS regressions with individual DAC-criteria as dependent variable. Weights are given by the inverse of the number of projects evaluated in the corresponding evaluation report. Other control variables include: Number of years between final project inspection and evaluation; the year of project start as well as evaluation year (both 5-year intervals); evaluation month. Standard errors in parentheses are clustered at the country-evaluation-year level. */**/*** denote significance levels at 10/5/1% respectively.

Table A.9: Results: LASSO estimates

Dep. variable: Rating (pooled)	(1) Overall	(2) LASSO estimation results	(3) Reduced-form estimation results
Financing			
Total volume (log)	0.037	0.042	0.047*
	(0.029)		(0.025)
Aid type (base: Loan):			
-Loan		-0.115	-0.127
			(0.084)
-Grant	0.105		
	(0.087)		
% counterpart contributions	0.145		
	(0.118)		
Budget funds (log)	0.095**	0.075	0.083**
	(0.042)		(0.039)
% budget funds of ODA	-0.000	-0.000	-0.000
	(0.000)		(0.000)
% project funds of GDP	-0.000	-0.000	-0.000
	(0.000)		(0.000)
Disbursement vs. commitment	0.137	0.062	0.131
	(0.156)		(0.156)
Structure			
Co-financing	0.002		
	(0.064)		
Accompanying measure	-0.015		
	(0.056)		
Agency type (base: NGO):			
-NGO		0.073	0.095
			(0.106)
-Mixed	-0.099	-0.001	-0.013
	(0.130)		(0.081)
-Multilateral	-0.009		
	(0.131)		
-Private sector	0.006	0.084	0.100
	(0.139)		(0.095)
-Government	-0.101		
	(0.107)		
Previous cooperation	0.066	0.058	0.064
	(0.051)		(0.051)
Number of institutions	0.005	0.002	0.005
	(0.009)		(0.009)
Project manager turnover	0.328	0.161	0.163**
	(0.248)		(0.066)
Country office	-0.043	-0.031	-0.050
	(0.056)		(0.054)
Complexity			
Project duration (log)	-0.149**	-0.154	-0.160**
	(0.075)		(0.073)
Delay indicator	0.009		
	(0.069)		
Revised ToC	-0.048	-0.036	-0.043
	(0.047)		(0.047)
Years mandate to contract	-0.048*	-0.043	-0.046*
	(0.027)		(0.027)
Technical complexity	-0.130**	-0.135	-0.139**
	(0.055)		(0.055)
Risks			
Number ex-ante identified risks	0.001		
	(0.013)		
% ex-ante identified risks occurred	-0.486***	-0.493	-0.492***
	(0.067)		(0.068)
Overall risk (base: low):			
-Medium	-0.203**	-0.172	-0.204**
	(0.082)		(0.079)
-(Very) high	-0.352***	-0.312	-0.344***
	(0.088)		(0.084)
-Not assigned	-0.219*	-0.161	-0.206*
	(0.116)		(0.114)
Overall risk control (base: low):			
-Medium	0.084	0.091	0.089
	(0.058)		(0.056)
-High	-0.061		
	(0.169)		
Macro variables			
GDP p.c. growth (annual)	0.011	0.011	0.011
	(0.008)		(0.008)
Freedom House Democracy score	-0.018	-0.013	-0.021
	(0.021)		(0.020)
State Fragility Index	-0.006	-0.005	-0.007
	(0.008)		(0.008)
Population (lo)g	-0.029	-0.019	-0.022
	(0.022)		(0.021)
Sector indicators	Yes	Yes	Yes
Sub-rating indicators	Yes	Yes	Yes
Other control variables	Yes	Yes	Yes
Observations	5,458	5,458	5,458
Adjusted R^2	0.23		0.19

Note: Table entries in column 1 are coefficients from WLS regressions with the pooled rating as dependent variable. LASSO (column 2) presents results from an adaptive LASSO regression. Reduced form estimates (column 3) runs the WLS regression on all variables with coefficients that are different from zero in the LASSO regression. Weights are given by the inverse of the number of projects evaluated in the corresponding evaluation report. Other control variables include: Number of years between final project inspection and evaluation; the year of project start as well as evaluation year (both 5-year intervals); evaluation month. Standard errors in parentheses clustered at the country-evaluation-year level. */**/*** denote significance levels at 10/5/1% respectively.

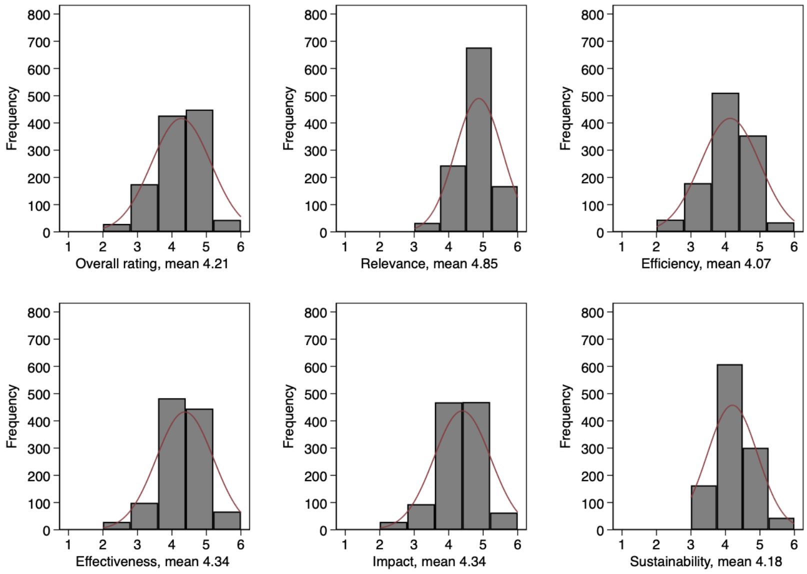
TABLE A.10: Robustness: Alternative estimations

	Rating (pooled)			Overall Rating		Arithmetic Rating		Binary
	(1) OLS	(2) O. Probit	(3) Vol. weighted	(4) OLS	(5) O. Probit	(6) OLS	(7) O. Probit	(8) O. Probit
Financing								
Total volume (log)	0.037	0.020	0.012	0.050	0.028	0.037	0.012	0.061
	(0.029)	(0.034)	(0.022)	(0.041)	(0.048)	(0.030)	(0.042)	(0.080)
Aid type (Base: Loan):								
-Grant	0.105	0.073	0.042	0.129	0.095	0.104	0.080	0.055
	(0.087)	(0.132)	(0.087)	(0.121)	(0.165)	(0.089)	(0.163)	(0.239)
% counterpart contributions	0.145	0.050	0.026	0.176	0.073	0.149	0.047	0.530
	(0.118)	(0.170)	(0.109)	(0.170)	(0.235)	(0.122)	(0.215)	(0.355)
Budget funds (log)	0.095**	0.128**	0.085**	0.123**	0.176**	0.096**	0.157**	0.355***
	(0.042)	(0.057)	(0.037)	(0.059)	(0.074)	(0.043)	(0.071)	(0.107)
% budget funds of ODA	-0.000	-0.000	-0.000	-0.000*	-0.000*	-0.000	-0.000	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
% project funds of GDP	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Disbursement vs. commitment	0.137	0.282	0.228	0.286	0.386	0.128	0.365	1.112**
	(0.156)	(0.302)	(0.212)	(0.234)	(0.366)	(0.159)	(0.377)	(0.501)
Structure								
Co-financing	0.002	0.044	0.028	0.083	0.221**	-0.001	0.059	0.274
	(0.064)	(0.081)	(0.053)	(0.088)	(0.112)	(0.066)	(0.101)	(0.171)
Accompanying measure	-0.015	-0.048	-0.037	-0.078	-0.109	-0.013	-0.053	-0.228
	(0.056)	(0.074)	(0.050)	(0.077)	(0.104)	(0.057)	(0.090)	(0.151)
Agency type (Base: NGO):								
-Mixed	-0.099	-0.070	-0.046	-0.092	-0.089	-0.103	-0.065	-0.384
	(0.130)	(0.177)	(0.118)	(0.180)	(0.241)	(0.134)	(0.223)	(0.353)
-Multilateral	-0.009	0.249	0.137	-0.045	0.326	-0.011	0.297	0.586
	(0.131)	(0.220)	(0.136)	(0.186)	(0.300)	(0.134)	(0.276)	(0.588)
-Private sector	0.006	-0.067	-0.049	0.049	-0.014	0.004	-0.032	-0.371
	(0.139)	(0.219)	(0.148)	(0.185)	(0.279)	(0.142)	(0.264)	(0.377)
-Government	-0.101	-0.072	-0.048	-0.094	-0.075	-0.101	-0.059	-0.389
	(0.107)	(0.149)	(0.099)	(0.145)	(0.199)	(0.110)	(0.185)	(0.298)
Previous cooperation	0.066	0.066	0.050	0.112	0.125	0.065	0.074	0.163
	(0.051)	(0.068)	(0.045)	(0.070)	(0.092)	(0.052)	(0.084)	(0.143)
Number of institutions	0.005	0.021	0.013	0.006	0.020	0.005	0.029*	0.013
	(0.009)	(0.013)	(0.009)	(0.011)	(0.016)	(0.009)	(0.016)	(0.026)
Project manager turnover	0.328	0.426	0.291	0.449*	0.548*	0.321	0.397	2.279**
	(0.248)	(0.328)	(0.215)	(0.267)	(0.319)	(0.256)	(0.382)	(0.899)
Country office	-0.043	-0.087	-0.058	-0.093	-0.180*	-0.044	-0.115	-0.400***
	(0.056)	(0.080)	(0.053)	(0.076)	(0.104)	(0.057)	(0.100)	(0.149)
Complexity								
Project duration (log)	-0.149**	-0.230**	-0.148**	-0.202**	-0.310**	-0.146*	-0.323***	-0.387*
	(0.075)	(0.099)	(0.063)	(0.099)	(0.130)	(0.078)	(0.123)	(0.200)
Delay indicator	0.009	-0.013	-0.022	-0.010	-0.036	0.005	-0.005	0.012
	(0.069)	(0.104)	(0.069)	(0.094)	(0.129)	(0.071)	(0.125)	(0.174)
Revised ToC	-0.048	-0.081	-0.050	-0.078	-0.095	-0.049	-0.101	-0.066
	(0.047)	(0.070)	(0.047)	(0.064)	(0.094)	(0.048)	(0.087)	(0.141)
Years mandate to contract	-0.048*	-0.048	-0.031	-0.046	-0.049	-0.049*	-0.064	-0.057
	(0.027)	(0.037)	(0.025)	(0.037)	(0.047)	(0.028)	(0.046)	(0.066)
Technical complexity	-0.130**	-0.214***	-0.140***	-0.121	-0.246**	-0.130**	-0.240**	-0.454***
	(0.055)	(0.078)	(0.052)	(0.076)	(0.103)	(0.057)	(0.094)	(0.157)
Risks								
Number ex-ante identified risks	0.001	0.006	0.003	0.000	0.002	0.001	0.008	-0.034
	(0.013)	(0.019)	(0.012)	(0.017)	(0.024)	(0.013)	(0.024)	(0.032)
% ex-ante identified risks occurred	-0.486***	-0.760***	-0.504***	-0.650***	-1.008***	-0.486***	-0.940***	-1.548***
	(0.067)	(0.103)	(0.069)	(0.092)	(0.136)	(0.069)	(0.128)	(0.209)
Overall risk (base: low)								
-Medium	-0.203**	-0.292**	-0.187**	-0.292**	-0.457**	-0.203**	-0.354**	-1.237***
	(0.082)	(0.128)	(0.080)	(0.124)	(0.201)	(0.084)	(0.166)	(0.440)
-(Very) high	-0.352***	-0.519***	-0.333***	-0.496***	-0.767***	-0.349***	-0.629***	-1.459***
	(0.088)	(0.134)	(0.084)	(0.131)	(0.211)	(0.091)	(0.174)	(0.465)
- Not assigned	-0.219*	-0.424***	-0.276***	-0.349**	-0.630**	-0.217*	-0.550***	-1.312**
	(0.116)	(0.163)	(0.103)	(0.175)	(0.249)	(0.119)	(0.211)	(0.511)
Overall risk control (base: low)								
-Medium	0.084	0.125	0.077	0.140*	0.182	0.086	0.143	0.101
	(0.058)	(0.085)	(0.055)	(0.076)	(0.112)	(0.060)	(0.107)	(0.155)
-High	-0.061	0.074	0.034	-0.106	-0.005	-0.061	0.074	-0.070
	(0.169)	(0.284)	(0.175)	(0.221)	(0.355)	(0.173)	(0.367)	(0.435)
Macro variables								
GDP p.c. growth (annual)	0.011	0.012	0.009	0.009	0.012	0.011	0.012	0.037
	(0.008)	(0.012)	(0.008)	(0.011)	(0.016)	(0.008)	(0.015)	(0.023)
Freedom House Democracy score	-0.018	-0.034	-0.022	-0.031	-0.068	-0.019	-0.052	-0.056
	(0.021)	(0.036)	(0.024)	(0.029)	(0.047)	(0.022)	(0.044)	(0.059)
State Fragility Index	-0.006	-0.006	-0.005	-0.009	-0.014	-0.006	-0.011	-0.028
	(0.008)	(0.013)	(0.009)	(0.010)	(0.017)	(0.008)	(0.016)	(0.022)
Population (log)	-0.029	-0.036	-0.024	-0.031	-0.042	-0.029	-0.036	-0.085
	(0.022)	(0.039)	(0.025)	(0.028)	(0.048)	(0.022)	(0.048)	(0.058)
Sector and region indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sub-rating indicators	Yes	Yes	Yes					
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,458	5,458	5,458	1,094	1,094	1,094	1,094	1,053
Adjusted R^2	0.23		0.24	0.19		0.19		

Note: Table entries are coefficients from WLS regressions with individual DAC criteria as dependent variable. Weights are given by the inverse of the number of projects evaluated in the corresponding evaluation report. Other control variables include: number of years between final project inspection and evaluation; the year of project start as well as evaluation year (both 5-year intervals); evaluation month. Standard errors in parentheses are clustered at the country-evaluation-year level. */**/*** denote significance levels at 10/5/1% respectively.

A.2 Figures

FIGURE A.1: Distribution of OECD DAC-ratings



Note: Distribution of the overall and individual DAC-ratings for $N = 1,124$ evaluations in our sample. Frequency refers to the number of projects with respective rating. The red line depicts the normal density.

A.3 Methodology

A.3.1 Calculation of macro variables

All macro variables we employ in our model are computed as averages using the respective variable's average value (i) during its project duration and (ii) across countries in which the project was implemented:

$$macro_i = \frac{1}{N}\sum_{k=1}^{N}[\frac{1}{T-t}\sum_{s=t}^{T} macro_{kt}], \tag{A.1}$$

where i denotes the respective project, N is the number of countries in which the project was implemented in, T is the year in which a project was completed and t is the year a project started.

A.3.2 Extra-/Interpolation of macro variables

The macro variables used are partially not available for the relevant time period. We impute these missing values in two steps:

Annual growth rate:

$$agr_t = var_t / var_{t-1}, \tag{A.2}$$

where var_t is the variable of interest in year t and agr is the annual growth rate.

Geometric mean:

$$geomgr_t = (agr_t \cdot agr_{t-1} \cdot agr_{t-2} \cdot agr_{t-3} \cdot agr_{t-4})^{1/5}, \tag{A.3}$$

where $geomgr_t$ is the geometric mean of the five last annual growth rates (last five years) in year t.

(i) If the latest data points are missing (forward extrapolation) or if data is missing in between available data points (interpolation):

$variable_t = variable_{t^*} \cdot geomgr_{t^*}$

t^* corresponds to the latest year for which the variable is available (and its five-year geometric growth rate). $t^* < t$.

(ii) If the earliest data points are missing (backward extrapolation):

$variable_t = variable_{t^*} \cdot geomgr_{t^*}$

t^* corresponds to the earliest year for which the variable is available (and its five-year geometric growth rate). $t^* > t$.

In case macro data observations are missing for certain years, we use the five-year geometric mean of annual growth rates to extrapolate (interpolate) missing values.

A.3.3 Within- vs. between-country analysis

We follow Denizer et al. (2013) and regress binary project success on a categorical indicator for the the country a project was implemented in to determine whether success varies more within-country or between-country.[1] This regression is run separately in sub-samples for each year projects in our sample were active in, i.e. 1990–2020.[2] We define active in year t as project start prior to or in t and project end before or in t. The following equation describes the model:

$$y_{ti} = \alpha_t + \beta_t \, country_i, \tag{A.4}$$

where y_{ti} is a binary measure of success of project i in sub-sample t and $country_i$ is a categorical variable, corresponding to the country project i was implemented in.

To obtain a single numeric value across sub-samples, we take the weighted average of regression coefficients. The weight corresponds to the number of observations in each sub-sample:

$$\hat{\beta} = \frac{\sum_{t=1}^{T} \hat{\beta}_t \cdot N_t}{\sum_{t=1}^{T} N_t}, \tag{A.5}$$

where T is the number of years a project was active in and $\hat{\beta}_t$ is the parameter for projects active in year t and N_t is the number of observations (projects) active in year t.

The explanatory power of *between*-country variation in project success is evaluated as follows: The estimated parameter $\hat{\beta}$ is interpreted as the explanatory power of between-country variation. The unexplained variation, $1 - \hat{\beta}$, is interpreted as within-country variation, i.e. project-specific characteristics.

Note that some projects ($N = 87$) were implemented in multiple countries. For these countries, a country group was created and entered as country. A possible caveat of this procedure is that the explanatory power of between-country variation may be inflated: If in a specific set of countries, only one project of our sample was active, the country or country group explains 100% of the variation in outcomes (cf. Bulman et al., 2017). In our dataset, this only applies to two observations.

[1] Or countries, depending on the number of countries a project was implemented in.
[2] Because only one project was active was active in 2021, we exclude this year.

Appendix B

Losing territory: The effect of administrative splits on land-use in the tropics

B.1 Tables

TABLE B.1: Descriptives: Summary statistics

Samples:	Entire sample		Bandwidth 20km	
	Mother (1)	Child (2)	Mother (3)	Child (4)
Split characteristics				
Number of villages	19,867	13,920	7,369	6,951
Distance to split (km)	39.7	29.1	10.3	9.9
	(37.8)	(31.1)	(5.4)	(5.3)
Distance to capital	39.7	34.0	28.7	26.6
	(38.9)	(31.5)	(26.6)	(23.1)
Distance to capital change (km)	-	42.0	-	22.7
	(-)	(45.8)	(-)	(30.0)
Length of split (km)	108.4	108.4	108.4	108.4
	(72.4)	(72.4)	(72.4)	(72.4)
Land use metrics				
Village size (sqkm)	40.7	45.1	26.6	27.9
	(109.0)	(128.3)	(76.3)	(75.9)
Forest cover 2000 (%)	79.2	80.1	77.6	80.1
	(23.0)	(23.4)	(23.0)	(23.1)
Forest cover 2018 (%)	66.6	69.0	66.1	68.1
	(24.3)	(25.6)	(23.5)	(24.7)
Oil Palm area 2000 (%)	5.4	7.0	5.8	5.8
	(15.7)	(18.6)	(17.0)	(16.7)
Human footprint area 2000 (%)	3.5	2.3	4.8	3.0
	(8.9)	(6.1)	(10.5)	(7.4)
Village topography				
Altitude (in meters)	396.3	454.7	449.5	537.2
	(598.2)	(670.3)	(592.2)	(720.6)
Located on shore (%)	17.8	18.3	12.4	13.6
	(38.2)	(38.6)	(32.2)	(34.2)
Distance to sub-district capital	20.0	23.6	16.9	18.4
in 2000 (km)	(32.5)	(50.1)	(31.6)	(30.6)
Distance to district capital	169.6	182.3	133.6	150.4
in 2000 (km)	(191.4)	(198.8)	(147.8)	(165.1)
Socio-economic composition (in 2000)				
Population	1,650	1,529	1,763	1,670
	(1,921)	(1,813)	(2,054)	(2,010)
Rural (%)	94.0	96.8	92.7	96.4
	(23.6)	(17.4)	(25.8)	(18.5)
Main income agricultural (%)	96.1	97.7	95.7	97.5
	(19.2)	(14.9)	(20.2)	(15.3)
Ethnic fractionalization	0.511	0.477	0.511	0.477
(at district-level)	(0.19)	(0.20)	(0.19)	(0.20)

Note: Distance to capital change is not available for mother villages because they retain their original capital as part of district splits. Forest cover, oil palm area and human footprint area relate the respective extent to village area. Standard deviations reported in parentheses.

TABLE B.2: Placebo checks: Continuity of topographic and socio-economic characteristics in 2000

	Panel A: Land-use characteristics in 2000						
	Forest cover (1)	Oil palm area (2)	Settlement area (3)				
Child	-0.003	-0.004	0.007				
	(0.019)	(0.005)	(0.009)				
Obs.	14,320	14,300	14,320				
Adjusted R^2	0.340	0.267	0.483				
	Panel B: Socio-geographic characteristics (in 2000)						
	ln Pop. (1)	% Rural (2)	% Agricult. Income (3)	Subdist. city distance (4)	District city distance (5)	% Coastal location (6)	Altitude (7)
Child	0.038	0.024	0.006	-2.360	7.702	0.025	1.274
	(0.046)	(0.015)	(0.007)	(1.828)	(6.728)	(0.016)	(24.745)
Obs.	13,568	14,227	13,568	13,568	13,568	13,568	14,319
Adjusted R^2	0.503	0.075	0.070	0.166	0.670	0.260	0.787
	Panel C: Socio-economic characteristics in 2000 (1)						
	No. Poverty card (1)	No. health card (2)	% Phone (3)	% Radio (4)	% Hospital (5)	% Sub-hospital (6)	% Kinder-garten (7)
Child	5.840	8.568	0.002	0.004	-0.003	0.0179	0.006
	(4.916)	(7.207)	(0.005)	(0.021)	(0.004)	(0.005)	(0.002)
Obs.	13,569	13,569	13,569	13,569	13,569	13,569	13,569
Adjusted R^2	0.141	0.271	0.052	0.143	0.007	0.116	0.242
	Panel D: Socio-economic characteristics in 2000 (2)						
	% Primary school (1)	% Bank index 1 (2)	% Bank index 2 (3)	% Market index 1 (4)	% Market index 2 (5)	# State electr. access (6)	# Private electr. access (7)
Child	-0.005	-0.001	0.003	0.011	0.016	3.979	3.116
	(0.004)	(0.017)	(0.017)	(0.019)	(0.010)	(19.160)	(4.062)
Obs.	13,569	13,569	13,569	13,569	13,569	13,569	13,569
Adjusted R^2	0.251	0.047	0.073	0.065	0.064	0.400	0.149
Split ID FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: % rural, agricult. income, coastal location, phone, radio, (sub-) hospital, kindergarten, primary school, bank, and market capture binary village access variables. Poverty and health cards, state and private electr. access capture the number of inhabitants with access. See section 3.4 for the source of the respective outcome variable used. Child is a binary indicator for villages located in the new child district. The sample consists of villages whose centroids lie within a fixed bandwidth of 20km around the 115 district split boundaries. The SRDD relies on a linear fit. Standard errors are clustered at the district split ID. */**/*** denote significance levels at 10/5/1% respectively.

TABLE B.3: Robustness: Dynamic SRDD effects on deforestation

Dependent:	ln Mean deforestation					Forest cover
Period:	Pre 6-4	Pre 3-1	Post 0-3	Post 4-6	Post 7-9	in 2018
	(1)	(2)	(3)	(4)	(5)	(6)
Child	-0.208	-0.483***	-0.390**	-0.190	-0.0625	-0.109
	(0.296)	(0.166)	(0.151)	(0.160)	(0.170)	(0.069)
Split ID FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,148	14,319	14,319	14,319	12,958	14,319
Adjusted R^2	0.435	0.396	0.472	0.457	0.462	0.628

Note: The dependent variable is average deforestation in the years indicated, transformed by the inverse hyperbolic sine. Child is a binary indicator for villages located in the new child district. The sample consists of villages whose centroids lie within a fixed bandwidth of 20km around the 115 district split boundaries. The SRDD relies on a linear fit. Controls include village altitude and forest cover in 2000. Standard errors are clustered at the district split ID. */**/*** denote significance levels at 10/5/1% respectively.

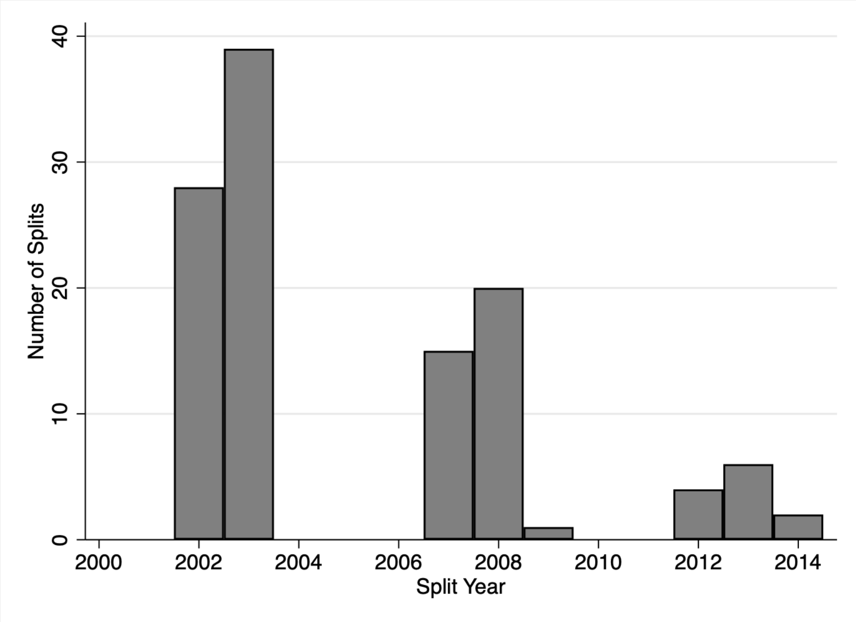
TABLE B.4: Robustness: SRDD effects on deforestation using quadratic fit

	(1)	(2)	(3)	(4)	(5)
	Panel A: Dep.: ln *Pre-split mean deforestation*				
Child	-1.065***	-0.880***	-0.704**	-0.627***	-0.670***
	(0.384)	(0.292)	(0.275)	(0.234)	(0.221)
Bandwidth	20	20	20	20	30 (66)
Observations	14,320	14,320	14,320	14,319	19,848
Adj. R^2	0.004	0.165	0.297	0.396	
	Panel B: Dep.: ln *Post-split mean deforestation*				
Child	-0.743*	-0.704**	-0.610**	-0.530**	-0.558***
	(0.381)	(0.288)	(0.268)	(0.223)	(0.214)
Bandwidth	20	20	20	20	26 (55)
Observations	14,320	14,320	14,320	14,319	17,746
Adj. R^2	0.004	0.215	0.355	0.472	
Island-year FE	No	Yes	No	No	No
Split-ID FE	No	No	Yes	Yes	Yes
Controls	No	No	No	Yes	Yes

Note: The dependent variable is average deforestation within three years before (Panel A) and from to three years after (Panel B) the split, transformed by the inverse hyperbolic sine. Child is a binary indicator for villages located in the new child district. The sample consists of villages whose centroids lie within a fixed bandwidth of 20km around the 115 district split boundaries. The bandwidth in column 5 is determined using an RBC estimator (Calonico et al., 2014). The SRDD is fitted relying on a quadratic trend. Controls include village altitude and forest cover in 2000. Standard errors are clustered at the district split ID. */**/*** denote significance levels at 10/5/1% respectively.

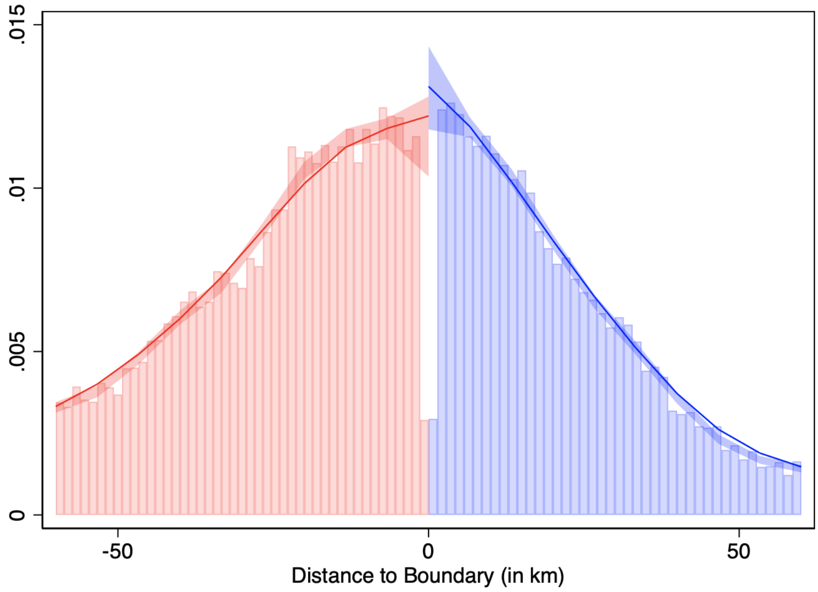
B.2 Figures

FIGURE B.1: Descriptives: Frequency of splits



Note: The figure displays the number of district splits in our sample by year they were legislated in.

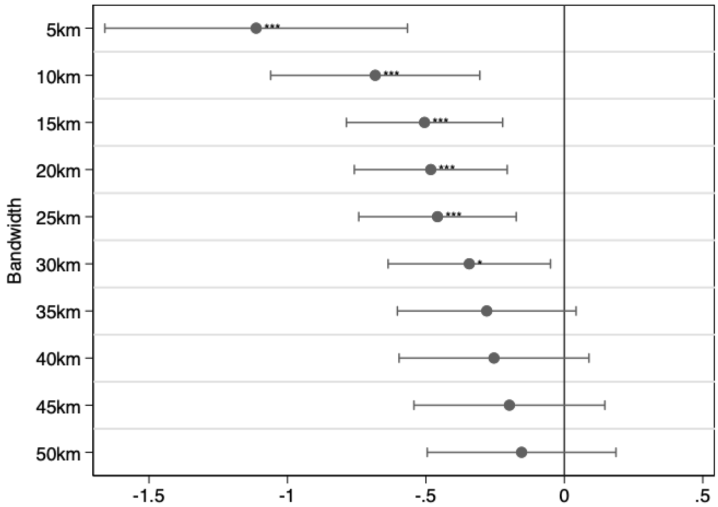
FIGURE B.2: Identification check: Density of the forcing variable



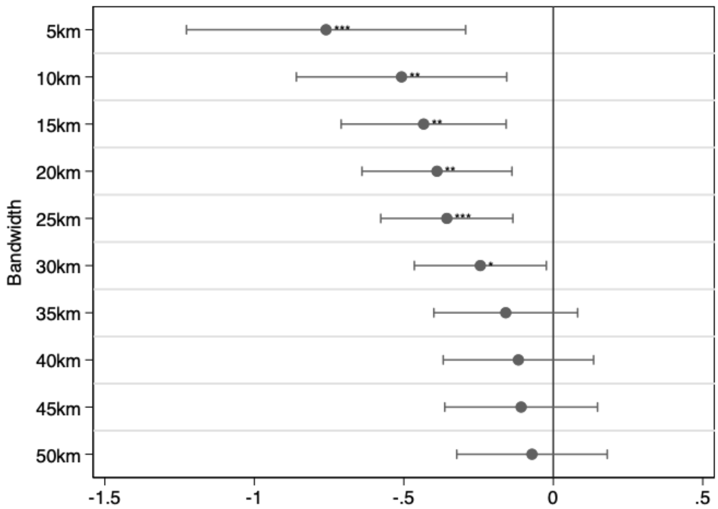
Note: Density of villages around the new district boundaries, measured in km. Figure constructed using *rdrobust* STATA package by Cattaneo et al., 2020. The corresponding local polynomial density estimator with quadratic fit is based on a 20km bandwidth, yielding a p-value of 0.102. The sample consists of villages whose centroids lie within the indicated bandwidth around the 115 district split boundaries.

FIGURE B.3: Robustness: Deforestation effects for varying bandwidths

(A) Pre-split

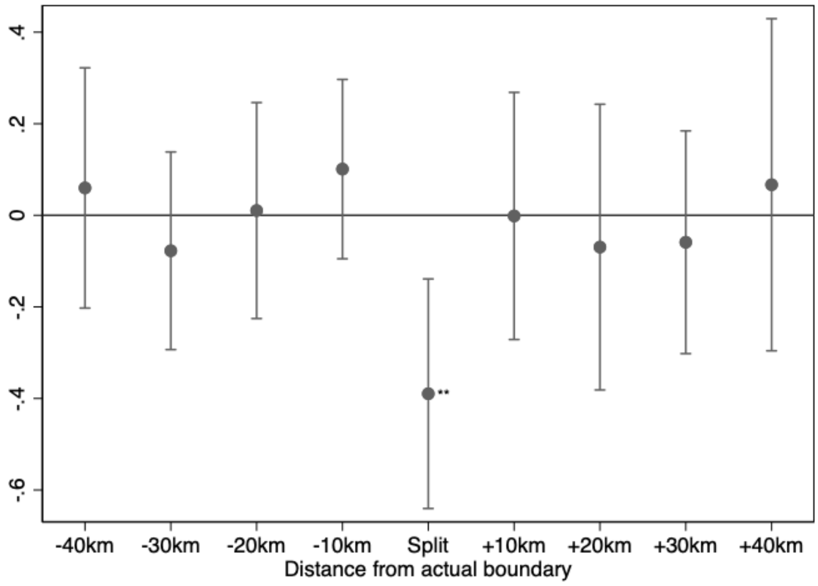


(B) Post-split



Note: The figure displays coefficients from individual estimates of the binary child indicator for villages located in the new child district (eqn. (3.1)), with the dependent variable measuring average deforestation within three years before (Panel A) and from to three years after (Panel B) the split, transformed by the inverse hyperbolic sine. The sample consists of villages whose centroids lie within a fixed bandwidth indicated on the y-axis around the 115 district split boundaries. The SRDD relies on a linear fit. Controls include village altitude and forest cover in 2000. The graph displays 90% confidence intervals with standard errors clustered at the split level. */**/*** denote significance levels at 10/5/1% respectively.

FIGURE B.4: Robustness: Shifting boundaries in space



Note: The figure displays coefficients from individual estimates of the binary child indicator for villages located in the new child district (eqn. (3.1)), with the dependent variable measuring average deforestation from to three years after the split, transformed by the inverse hyperbolic sine. The sample consists of villages whose centroids lie within a fixed bandwidth of 20km around the 115 district split boundaries for the coefficient labeled as “Split". All other coefficients are based on samples that artificially moved the boundary up to 40km away from the actual split boundary. The SRDD relies on a linear fit. Controls include village altitude and forest cover in 2000. The graph displays 90% confidence intervals with standard errors clustered at the split level. */**/*** denote significance levels at 10/5/1% respectively.

Appendix C

Confined to Stay: Natural Disasters and Indonesia's Migration Ban

C.1 Background

C.1.1 Questions of key variables included in PODES

Stock of migrants Are there any village residents who work abroad as "overseas workers" (*TKI*)? If yes, number of males/females currently working abroad.

Main destination country What is the destination country for the majority of overseas workers from this village?

Natural disasters Has there been natural disaster in the last three years, that caused damages and losses?

Poverty letters (*SKTM*) Number of poverty letters issued in the previous year.

Social health cards (*Askeskin*) Number of households who received "Health card/member card of the health aid program for poor people" during the previous year.

Population Residents and families - total male/female population (note that for PODES 2014, we use data on household electricity access to extrapolate population figures: The number of households with and without access yields the total number of households in a village, which we multiply by the average Indonesian household size (3.6 in 2014)).

Conflict Have there been any mass fights in the past year?

Source: List based on PODES 2005. See Appendix Table C.1 for summary statistics.

C.1.2 Criteria for the eligibility of poverty letters (*SKTM*)

1. The floor area of the building in which the household resides is less than eight square meters per member.
2. The floor of the household's residence is made of earth/cheap cement.
3. The walls of the household's residence are made of low quality wood or are damaged.
4. Household does not have their own sanitation facility.
5. Household's lighting sources do not use electricity or share it with other families.
6. Household's access to drinking and cooking water comes from wells.
7. Household's fuel for daily cooking is firewood or subsidized gas.
8. Household consumes meat/dairy/chicken less than once a week.
9. Household can purchase a maximum of one set of new clothes per member per year.
10. Household's frequency of eating for each member is maximum twice per day.
11. Household is not able to pay for a treatment at the public health centre ("puskesmas"/polyclinics).
12. The income of the head of the household is less than IDR 500,000 per month.
13. The educational attainment of the head of the household is less than primary schooling.
14. Household members do not have savings/assets with the minumum value of IDR 500,000.

Source: List based on Fiarni et al. (2013).

C.1.3 Determinants of access to poverty letters (SKTM)

To show that poverty letters are precisely targeted at the poorest households, we follow Priebe et al. (2014) in analysing the determinants of individual-level poverty letter access. For this exercise we use the fourth wave of the Indonesian Family and Life Survey (IFLS) panel data (Strauss et al., 2009). This wave was collected in 2007 and covers 13 provinces with a total of 13,535 households comprised of 50,580 individuals. We use the information about households' poverty letters uptake to estimate its determinants with the following linear probability model:[1]

$$SKTM_{ip} = \beta_0 + \beta_1\, Poverty_{ip} + \lambda_i + \mu_p + \varepsilon_{ip}, \qquad \text{(C.1)}$$

where $SKTM$ is a binary variable taking the value 1 if the individual i from province p lives in a household with poverty letter access (SKTM). We measure *Poverty* in two ways: First, we divide the sample into quintiles of yearly consumption expenditures. Second, we create a binary variable for individuals falling below the poverty line, i.e. those whose daily consumption expenditure is below USD 2 (2007 PPP). We introduce a vector of individual- and household-level controls λ_{ip} including rural/urban status, age and age squared, years of education, marital status, religion, number of household members and number of household members squared, number of children below 5, and number of elderly people above the age of 60. We further include province fixed effects μ_p and we cluster the standard errors at the household level.

Results are presented in Appendix Table C.12. It shows that individuals with lower consumption expenditures are more likely to reside in households with poverty letter access. More specifically, in column 1, individuals in the bottom expenditure quintile have a 11.3 percentage points greater probability to hold poverty letters than those the top quintile. This difference monotonically decreases, but remains significant at the 1% level for other quintiles. Furthermore, individuals whose daily consumption expenditure is below USD 2 (2007 PPP) have a 6.4 percentage points higher probability to hold a poverty letter. The results are similar if we restrict the sample to individuals in rural (columns 3–4) and urban areas (columns 5–6).

[1] The exact question in the survey is "Does this household have a "letter of poor" (Surat Keterangan Tidak Mampu, SKTM)?".

C.2 Tables

TABLE C.1: Summary statistics

	Mean	SD	Min	Max	Obs
Podes (2005, 2008, 2011 and 2014)					
Saudi Arabia as main destination	0.12	0.33	0	1	268,194
Stock of emigrants	18.39	67.40	0	5,912	268,194
Stock of female emigrants	10.78	39.76	0	3,022	268,194
Stock of male emigrants	7.61	38.54	0	4,670	268,194
Disaster in the last threee years	0.40	0.49	0	1	268,194
Number of disasters in the last three years	1.36	2.71	0	69	200,206
Poverty cards	66.58	210.62	0	41,448	268,194
Social health cards	431.93	939.95	0	55,307	268,194
Households living in slums	7.93	95.07	0	22,358	268,194
Population	3,346.63	4,731.86	4	199,996	268,194
Conflict in village	0.03	0.17	0	1	268,194
Rural village	0.82	0.38	0	1	267,724
Lowlands	0.19	0.39	0	1	183,337
Flood in the last three years	0.25	0.43	0	1	214,588
Landslide in the last three years	0.14	0.34	0	1	186,833
Forest fire in the last three years	0.05	0.22	0	1	169,542
Earthquake in the last three years	0.10	0.29	0	1	178,128
Tsunami in the last three years	0.01	0.08	0	1	162,051
Typhoon in the last three years	0.14	0.34	0	1	138,613
Tide in the last three years	0.03	0.18	0	1	123,888
Other disasters in the last three years	0.07	0.26	0	1	44,454
Smeru (2010 and 2015)					
Poverty rate (below 2$ PPP)	19.05	22.13	0	99.50	131,915

Note: Information on type of natural disaster and number of natural disasters is restricted to the years 2008, 2011 and 2014.

TABLE C.2: Average effect of disasters on poverty: With and without control variables

Dependent	Poverty cards							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SA × Post2011 × Disaster	0.110***	0.115***	0.113***	0.110***	0.118***	0.115***	0.113***	0.118***
	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)
Log(population)		0.391***			0.385***	0.389***		0.383***
		(0.014)			(0.014)	(0.014)		(0.014)
Asinh(male migrants)			0.060***		0.056***		0.060***	0.056***
			(0.004)		(0.004)		(0.004)	(0.004)
Conflict dummy				0.122***		0.110***	0.120***	0.108***
				(0.019)		(0.019)	(0.019)	(0.019)
Village FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province-time trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	268,194	268,194	268,194	268,194	268,194	268,194	268,194	268,194

Note: Poverty letters is transformed by the inverse asymptotic sine (asinh). All further interactions are included in the estimation but not displayed here. Robust standard errors are clustered at the village level and reported in parentheses. */**/*** denote significance levels at 10/5/1% respectively.

TABLE C.3: Average effect of number of disasters on poverty

Dependent	Poverty cards			
	DD			DDD
	(1)	(2)	(3)	(4)
Number of disasters		0.026***	0.019***	0.027***
		(0.002)	(0.002)	(0.002)
SA × Post2011	0.039*			0.015
	(0.021)			(0.024)
Post2011 × Number of disasters		-0.010***		-0.013***
		(0.003)		(0.003)
SA × Number of disasters			0.001	-0.011
			(0.005)	(0.007)
SA × Post2011 × Number of disasters				0.018**
				(0.008)
Village FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Province-time trend	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	200,173	200,173	200,173	200,173

Note: The sample is restricted to census years 2008, 2011 and 2014 for lack of intensive margin disaster data in PODES 2005. Poverty letters is transformed by the inverse asymptotic sine (asinh). Control variables include asinh(male migrants), log(population) and a conflict event binary indicator. Binary variables *SA* and *Post*2011 are omitted because they are absorbed by the fixed effects. Robust standard errors are clustered at the village level and reported in parentheses. */**/*** denote significance levels at 10/5/1% respectively.

TABLE C.4: Average effect of disasters on poverty using alternative poverty measurements

Dependent	Health cards	Households living in slums	Poverty rate 2$ PPP
	(1)	(2)	(3)
SA × Post2011 × Disaster	0.154*	0.279***	1.186***
	(0.081)	(0.042)	(0.382)
Village FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Province-time trend	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	268,194	268,194	128,772

Note: The dependent variables in columns 1–2 are transformed by the inverse asymptotic sine (asinh). The dependent variable in column 3 is the number of poor people below the poverty line of USD 2 PPP divided by total population. Control variables include asinh(male migrants), log(population) and a conflict event binary indicator. All further interactions are included in the estimation but not displayed here. The sample consists of census years 2005, 2008, 2011 and 2014 in columns 1–2, and 2011 and 2014 in column 3. Robust standard errors are clustered at the village level and reported in parentheses. */**/*** denote significance levels at 10/5/1% respectively.

TABLE C.5: Average effect of disasters on poverty in t-1

Dependent	Poverty cards		
	DD		DDD
	(1)	(2)	(3)
$\text{Disaster}_{\text{t-1}}$	-0.018*	-0.010	-0.020*
	(0.010)	(0.009)	(0.010)
Post2011 × $\text{Disaster}_{\text{t-1}}$	0.034**		0.032**
	(0.015)		(0.016)
SA × $\text{Disaster}_{\text{t-1}}$		0.026	0.020
		(0.025)	(0.029)
SA × Post2011			0.031
			(0.029)
SA × Post2011 × $\text{Disaster}_{\text{t-1}}$			0.013
			(0.043)
Village FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Province-time trend	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	199,899	199,899	199,899

Note: t-1 corresponds to the period of three to six years before *t*. Poverty letters is transformed by the inverse asymptotic sine (asinh). Control variables include asinh(male migrants), log(population) and a conflict event binary indicator. Robust standard errors are clustered at the village level and reported in parentheses. */**/*** denote significance levels at 10/5/1% respectively.

TABLE C.6: Average effect of disasters on poverty: Additional robustness checks

Dependent	Poverty cards					
	Java excluded	Java only	Only villages with migrants	Exclude Middle East	Trim population	Weighted by population
	(1)	(2)	(3)	(4)	(5)	(6)
SA × Post2011 × Disaster	0.177**	0.140***	0.118***	0.118***	0.085**	0.128***
	(0.076)	(0.048)	(0.042)	(0.042)	(0.041)	(0.040)
Village FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Prov-time trend	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	166,496	101,698	140,399	141,111	239,872	268,194

Note: Column 1 excludes villages on the island of Java from the sample, column 2 is restricted to villages on Java island. Column 3 restricts the sample to villages that had at least one Indonesian worker overseas in 2005. Column 4 excludes UAE, Jordan and Qatar as main destinations from the sample. Column 5 restricts the sample between the 1st and 99th population percentile. Column 6 weights the regression by population. Poverty letters is transformed by the inverse asymptotic sine (asinh). Control variables include asinh(male migrants) and a conflict event binary indicator. Log(population) is also included a control in columns 1–5. All further interactions are included in the estimation but not displayed here. Robust standard errors are clustered at the village level and reported in parentheses. */**/*** denote significance levels at 10/5/1% respectively.

TABLE C.7: Average effect of disasters on poverty controlling for financial transfers

Dependent	Poverty cards					
	(1)	(2)	(3)	(4)	(5)	(6)
SA × Post2011 × Disaster	0.097**	0.096**	0.096**	0.096**	0.096**	0.096**
	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)
Asinh(District transfers)		0.005**	0.004**	0.004*	0.004*	0.004*
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Asinh(Province transfers)			0.007***	0.007***	0.007***	0.007***
			(0.002)	(0.002)	(0.002)	(0.002)
Asinh(Central govn't transfers)				-0.002	-0.002	-0.002
				(0.002)	(0.002)	(0.002)
Asinh(Foreign aid)					0.021***	0.021***
					(0.006)	(0.006)
Asinh(Private aid)						-0.002
						(0.005)
Village FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Province-time trend	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	200,173	200,173	200,173	200,173	200,173	200,173

Note: Poverty letters is transformed by the inverse asymptotic sine (asinh). District transfer, province transfer and foreign and private aid are in million IDR. Control variables include asinh(male migrants), log(population) and a conflict event binary indicator. All further interactions are included in the estimation but not displayed here. The sample consists of census years 2008, 2011, 2014. Robust standard errors are clustered at the village level and reported in parentheses. */**/*** denote significance levels at 10/5/1% respectively.

TABLE C.8: Average effect of disasters on poverty in the presence of spillovers

Dependent	Poverty cards			
	DDD			
	(1)	(2)	(3)	(4)
SA × Post2011 × Disaster	0.119***	0.113***	0.0130***	0.131***
	(0.039)	(0.041)	(0.043)	(0.045)
Neighboring village with SA=1 in distance of:				
0-10km × Post2011 × Disaster		-0.009	0.007	0.008
		(0.029)	(0.031)	(0.034)
10-20km × Post2011 × Disaster			0.067	0.067
			(0.044)	(0.046)
20-30km × Post2011 × Disaster				0.005
				(0.055)
Village FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Province-time trend	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	268,194	268,194	268,194	268,194

Note: Poverty letters is transformed by the inverse asymptotic sine (asinh). Control variables include asinh(male migrants), log(population) and a conflict event binary indicator. The additional distance controls indicate whether a village is within xx km distance of a village with Saudi Arabia as main migratory destination (centroid based), and set to zero in case the village itself has Saudi Arabia as main destination. All further interactions are included in the estimation but not displayed here. Robust standard errors are clustered at the village level and reported in parentheses. */**/*** denote significance levels at 10/5/1% respectively.

TABLE C.9: Average effect of disasters on poverty: Conley standard errors

Dependent	Poverty cards			
Distance cut-off	5 km	10 km	20 km	30 km
	(1)	(2)	(3)	(4)
SA × Post2011 × Disaster	0.119**	0.119**	0.119**	0.119**
	(0.054)	(0.061)	(0.057)	(0.049)
Village FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Province-time trend	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	265,804	265,804	265,804	265,804

Note: Poverty letters is transformed by the inverse asymptotic sine (asinh). Control variables include asinh(male migrants), log(population) and a conflict event binary indicator. All further interactions are included in the estimation but not displayed here. 604 villages are excluded from the sample due to the absence of coordinates. */**/*** denote significance levels at 10/5/1% respectively.

TABLE C.10: Average effect of disasters on population growth and poverty

Dependent	(1) Population	(2) Poverty cards
SA × Post2011 × Disaster	-0.013**	-0.524
	(0.005)	(0.416)
SA × Post2011 × Disaster × log(population)		0.082
		(0.050)
Village FE	Yes	Yes
Time FE	Yes	Yes
Province-time trend	Yes	Yes
Controls	Yes	Yes
Observations	268,194	268,194

Note: The dependent variable is log(population) in column 1 and asinh(poverty letters) in column 2. Control variables include a conflict event binary indicator. All further interactions are included in the estimation but not displayed here. Robust standard errors are clustered at the village level and reported in parentheses. */**/*** denote significance levels at 10/5/1% respectively.

TABLE C.11: Average effect of extreme rainfall events on poverty

Dependent	Poverty cards			
Buffer	10 km	15 km	20 km	30 km
	(1)	(2)	(3)	(4)
SA × Post2011 × Extreme rainfall	0.576***	0.629***	0.562***	0.253***
	(0.196)	(0.144)	(0.119)	(0.091)
Village FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Province-time trend	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	15,453	23,735	32,360	51,657
Villages	5,112	7,781	10,524	16,656

Note: Poverty letters is transformed by the inverse asymptotic sine (asinh). Extreme rainfall events are defined as days of year t with the largest rainfall recorded in the 10 previous years (cf. 4.5. Control variables include asinh(male migrants), log(population) and a conflict event binary indicator. All further interactions are included in the estimation but not displayed here. Robust standard errors are clustered at the village level and reported in parentheses. */**/*** denote significance levels at 10/5/1% respectively.

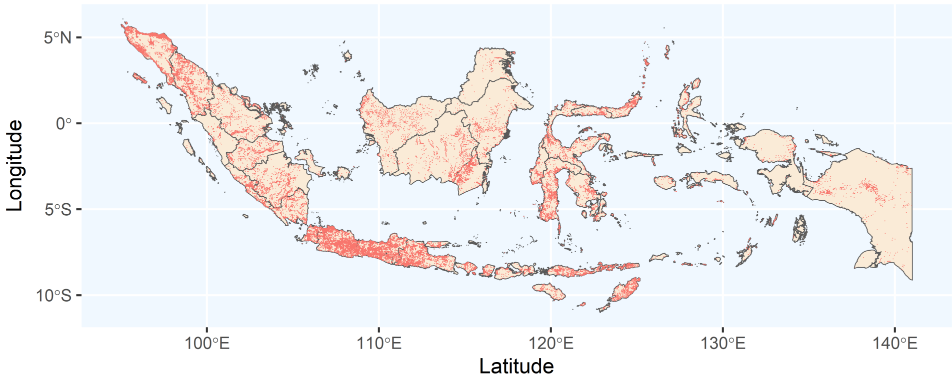
TABLE C.12: Probability to receive poverty letters (SKTM)

Dependent	Holding SKTM					
	All		Rural		Urban	
	(1)	(2)	(3)	(4)	(5)	(6)
Expenditure quintile 1	0.113		0.078		0.149	
	(0.012)		(0.016)		(0.019)	
Expenditure quintile 2	0.098		0.063		0.133	
	(0.011)		(0.015)		(0.017)	
Expenditure quintile 3	0.060		0.035		0.077	
	(0.009)		(0.013)		(0.012)	
Expenditure quintile 4	0.021		0.013		0.022	
	(0.007)		(0.012)		(0.009)	
Poor (below 2$ PPP)		0.064		0.043		0.098
		(0.010)		(0.013)		(0.018)
Avg. SKTM uptake in 5th qntl	0.050		0.059		0.046	
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	45,953	45,953	20,944	20,944	25,009	25,009

Note: The dependent variable captures whether a household holds a poverty letter (SKTM). Columns 2–3 restrict the sample to individuals living in rural, columns 5–6 to those in urban areas. The omitted group in columns 1, 3 and 5 is the expenditure quintile 5. Control variables include rural/urban residence (columns 1–2), age and age square, years of education, marriage status, religion, number of household members and number of household members squared, number of children below 5, and number of elderly people above the age of 60. Standard errors are clustered at the household level. Survey weights are applied. Data source: IFLS 4 (Strauss et al., 2009).

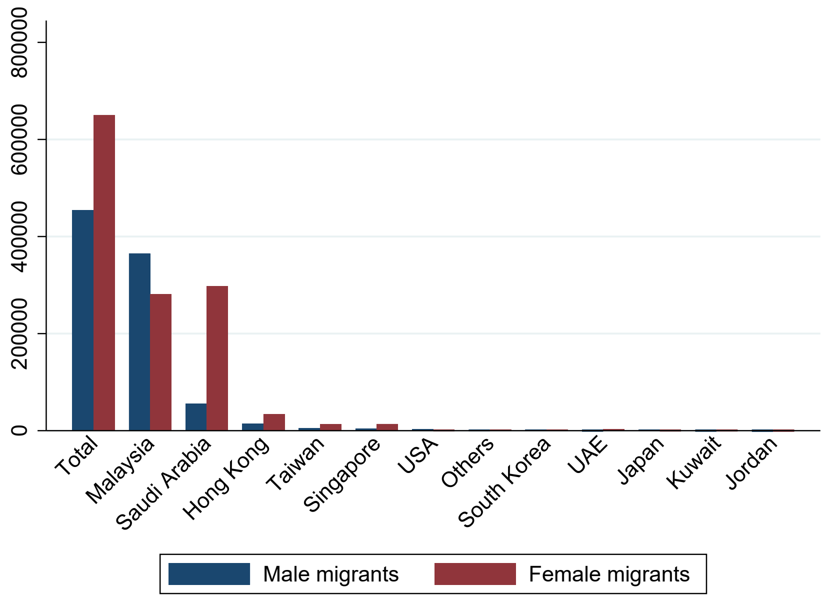
C.3 Figures

FIGURE C.1: Disaster events in the period 2003–2005



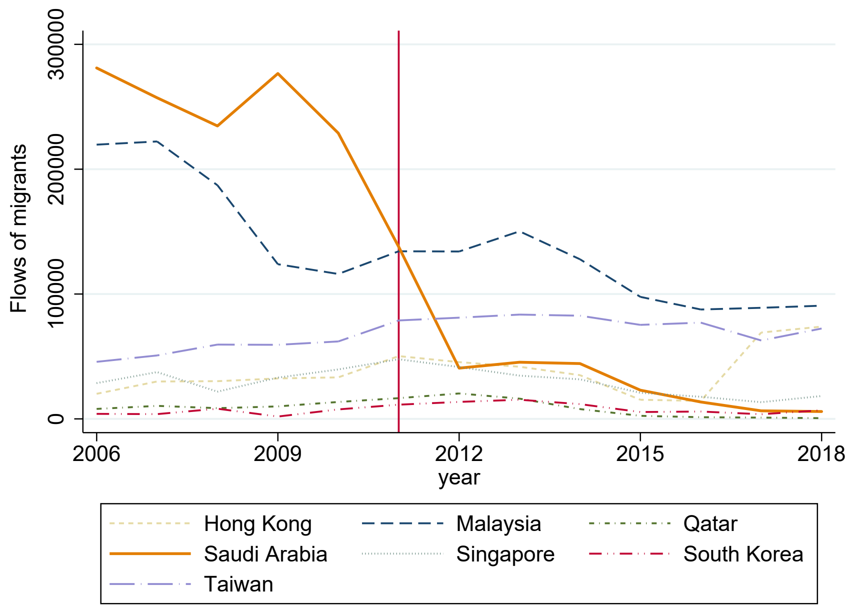
Note: Red dots represent centroids of villages that experienced at least one natural disaster between 2003 and 2005. Data taken from PODES 2005.

FIGURE C.2: Stock of emigrants by gender and destination in 2005



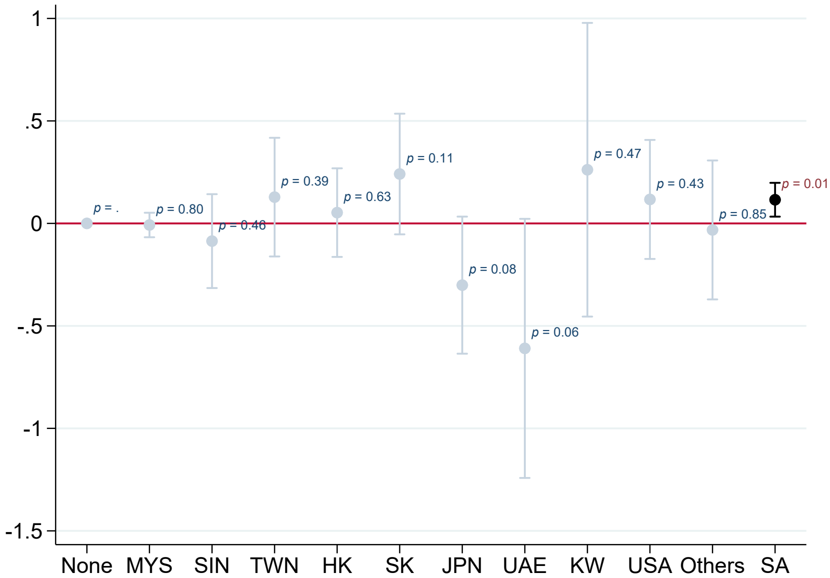
Note: Data taken from PODES 2005.

FIGURE C.3: Annual flows of documented migrants per destination



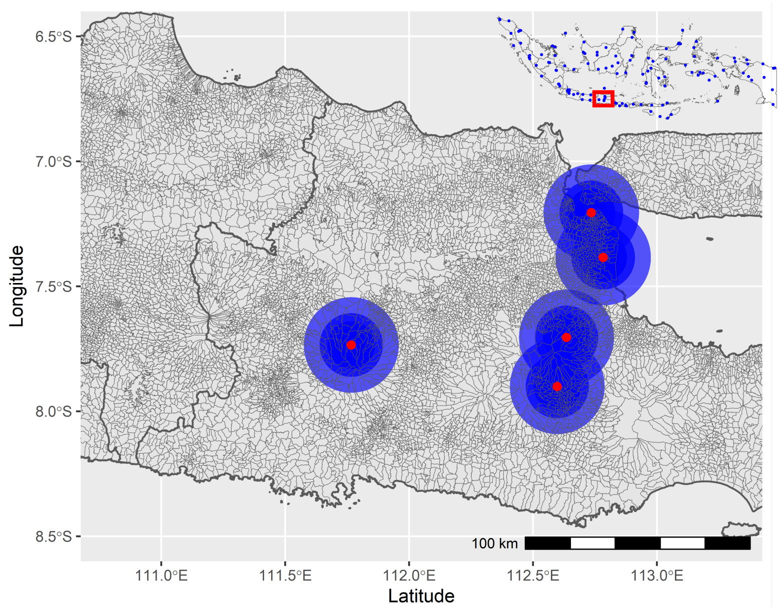
Note: The vertical line indicates the implementation of the ban in 2011. Data taken from the national placement agency for Indonesian workers abroad (BNP2TKI).

FIGURE C.4: Placebo: Average effects of disasters by villages' top destination countries



Note: The displayed coefficients capture the effect of natural disasters on poverty letters transformed by the inverse hyperbolic sine (asinh) with 95% confidence intervals. The baseline category is villages with no migrants ("None"). Countries from the left to the right are: Malaysia, Singapore, Taiwan, South Korea, Japan, United Arab Emirates, Kuwait, United States, other countries, Saudi Arabia. Control variables include asinh(male migrants), log(population) and a conflict event binary indicator. Village and year fixed effects as well as province-time trends are included. Standard errors are clustered at the village level. Number of observations: 268,194.

FIGURE C.5: Geocoded weather stations



Note: The larger map plots the zoomed in area delimited by the red box in the top right corner. Red dots represent the coordinates of each weather station. Different shades of blue indicate buffer zones of 10, 15, 20 and 30 km respectively. Data taken from BMKG and PODES.

Appendix D

Cash Transfers and Violent Crime in Indonesian Communities

D.1 Tables

TABLE D.1: NVMS data: Descriptive statistics on the types of violent crime

	RCT Sample Total cases	*Roll-out sample* Total cases
All	2,487	31,189
Group clash	35	762
Fight	173	1,888
Lynching	214	3,540
Vandalism	162	2,338
Assault	2,254	23,848
Sweeping	1	12
Kidnappping	7	237
Robbery	361	7,289
Others	7	345

Note: Each incident included in the totals ("All") can be categorized as pertaining to up to two types listed above. The *RCT sample* comprises 1,830 communities and the *Roll-out sample* 28,873 communities. Figures shown refer to the period 2005–2010 (*RCT sample*) and 2005–2014 (*Roll-out sample*) and to provinces covered by the NVMS. The table includes information on violent crime as defined by the NVMS (definition: "Violent crime comprises acts of violence that occur without any prior dispute between parties. The motivation behind a criminal act can be monetary, for example, robbery or abduction; or personal pleasure, for example, rape or serial killings. In contrast, violence in the context of conflict occurs due to pre-existing disputes between those involved such as dispute over land, election, religion or other such matters. As such, in the NVMS system, an act of killing can be coded as "Conflict" if there is a dispute behind it, e.g., in a killing of a certain group figure by other groups, or can be coded as "Crime" if there is no pre-existing dispute between parties, for example, serial killings.", (NVMS, 2014)). The "Others" category comprises cases of demonstrations, blockades, riots and terror attacks.

TABLE D.2: Sample selection due to NVMS coverage

Samples:	*Roll-out sample*			*RCT sample*			
Crime database coverage:	None (1)	NVMS (2)	Diff. (2–1)	None (3)	NVMS (4)	Diff. (4–3)	Diff. (4–2)
Socio-economic variables in 2008:							
Population [in 1,000]	4.159	2.977	-1.182***	6.671	3.215	-3.456***	0.278***
	(0.027)	(0.026)	(0.038)	(0.115)	(0.052)	(0.116)	(0.081)
Urban	0.136	0.102	-0.034***	0.076	0.068	-0.008	-0.038***
	(0.002)	(0.001)	(0.002)	(0.009)	(0.004)	(0.009)	(0.005)
% Househ. w. electricity	0.832	0.750	-0.082***	0.939	0.851	-0.088***	0.115***
	(0.001)	(0.001)	(0.002)	(0.004)	(0.004)	(0.009)	(0.005)
Community market available	0.197	0.178	-0.019***	0.112	0.174	0.062***	-0.003
	(0.002)	(0.002)	(0.003)	(0.010)	(0.006)	(0.014)	(0.007)
Community hospital available	0.023	0.020	-0.003**	0.019	0.015	-0.004	-0.005**
	(0.001)	(0.001)	(0.001)	(0.004)	(0.002)	(0.004)	(0.002)
Number communities (max.)	26,657	28,654		872	3,317		

Note: The *Roll-out sample* includes communities in which PKH was introduced until 2014. It is split into locations with NVMS coverage (our analytical sample; 16 provinces), and into non-NVMS communities that are located in the remaining 17 provinces. Likewise, the *RCT sample* is divided into communities located in the 16 NVMS provinces and those located in the other provinces. For the difference in column 4-2, the *Roll-out sample* does not contain communities that are part of the *RCT sample*. Statistical significance of differences is based on t-tests. All variables presented are taken from PODES 2008. */**/*** denote significance levels at 10/5/1% respectively.

TABLE D.3: Overview on variable construction

Data	Variable	Description
NVMS	Violent crime (extensive margin)	Binary; takes the value one if NVMS reported at least one incident of violent crime in a community in a given year
NVMS	Violent crime (intensive margin)	Numerical; # of NVMS reported cases of violent crime in a community in a given year
PODES	Population	Numerical; # of people reported to be living in a community as reported in the respective PODES round
PODES	# police stations	Numerical; # of existing police stations in a community as reported in the respective PODES round
SUSENAS	PKH	Binary; self-reported PKH membership by household
SUSENAS	Undercoverage	Share (0-1); # hh eligible for PKH but not receiving PKH divided by # PKH eligible households
SUSENAS	Leakage	Share (0-1); # non-eligible households receiving PKH divided by all households receiving PKH in a village
SUSENAS	Violent crime	Binary; takes the value 1 if hh stated to be victim of any of the following crimes: theft, robbery
SUSENAS	Reported crime	Share (0-1); ratio between # violent crime reported to police and # experienced violent crime (last 12 months)
PKH-RCT	Total assets	Asset index (PCA) considering asset possession of the following items: radio, tv, antenna, bike, motor bike, car, cellphone
PKH-RCT	Non-idleness	Binary; takes the value 1 if person is either in school or worked (at least 1 day) in the last 1 month
PKH-RCT	Working	Binary; takes the value 1 if person worked (at least 1 day) in the last 1 month; informal and formal work is included
PKH-RCT	Attending school	Binary; takes the value 1 if person is enrolled and attending school typically
PKH-RCT	Expenditures	Numerical; obtained from expenditure and consumption module, different recall periods are converted to monthly figures
PKH-RCT	Inequality	Gini, Theil, and CV are derived from monthly household per capita expenditures
PKH-RCT	Engage 1	Binary; Takes the value 1 if household has joined at least 1 community organization
PKH-RCT	Engage 2	Numerical; # community organizations a household has joined
PKH-RCT	Engage 3	Numerical; # hh members who are engaged in community organizations
PKH-RCT	Engage 4	Numerical; # meetings hh members have joined in the last 3 months

TABLE D.4: Roll-out determinants: Explaining year of PKH roll-out

Sample	*Roll-out (2007–2014)* (1)	(2)
Village characteristics		
Population (ln)	-0.448***	-0.042
	(0.049)	(0.034)
Rural	-0.660***	0.094
	(0.133)	(0.102)
Electricity access (hh %)	-0.892***	0.077
	(0.135)	(0.092)
Poverty cards (ln)	-0.070**	-0.024
	(0.029)	(0.019)
Roll-out criteria: Health facilities		
Hospital	0.120	-0.047
	(0.126)	(0.070)
Sub-hospital	0.503***	-0.011
	(0.044)	(0.019)
Puskesmas (health station)	0.523***	-0.006
	(0.049)	(0.022)
Roll-out criteria: Education facilities		
Kindergarten	-0.460***	-0.057
	(0.081)	(0.052)
Primary school	0.022	0.040
	(0.090)	(0.070)
High-school	-0.054	0.011
	(0.081)	(0.269)
District FE	No	Yes
Observations	26,597	26,597
Adj. R^2	0.078	0.594
F-statistic	28.60	1.35

Note: The dependent variable is the year of roll-out; the cross-sectional sample includes all communities in our national roll-out sample with full data in the village census PODES 2008. Robust standard errors are clustered at the sub-district level and reported in parentheses. */**/*** denote significance levels at 10/5/1% respectively.

TABLE D.5: Robustness: Assessing the presence of spillover effects

Sample	*Roll-out*					
Estimator	TWFE			Spillover-robust double diff. estimator (Clarke, 2017)		
	(1)	(2)	(3)	(4)	(5)	(6)
PKH treatment	0.006*	0.007*	0.008*	0.010***	0.007*	0.012**
	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.005)
Close to community with PKH treatment						
$0-10km$	0.000			0.004	0.001	0.006
	(0.000)			(0.003)	(0.004)	(0.005)
$0-10km \times$ PKH	0.000					
	(0.000)					
$10-20km$		0.000			-0.003	0.001
		(0.000)			(0.003)	(0.004)
$10-20km \times$ PKH		-0.000				
		(0.000)				
$20-30km$			0.000			0.007
			(0.000)			(0.004)
$20-30km \times$ PKH			-0.000			
			(0.000)			
Community FE, year FE	Yes	Yes	Yes	Yes	Yes	Yes
District-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	288,730	288,730	288,730	288,340	288,340	288,340

Note: The *Roll-out sample* includes 28,873 communities that ever received PKH during the period 2007–2014 and is restricted to communities with full NVMS coverage and population information. Additional controls indicate the number of communities with PKH access within *k* kilometers of the community (centroid based). In columns 4–6 neighboring indicators are set to zero in case the community was treated itself. Robust standard errors clustered at the sub-district level are reported in parentheses. */**/*** denote significance levels at 10/5/1% respectively.

TABLE D.6: Robustness: Results from randomization inference

Sample	*Roll-out*		*RCT*
Randomization	Across year	Within years	
	(1)	(2)	(3)
Mean PKH effect	0.000	0.000	
Mean standard error	(0.002)	(0.002)	
Median standard error	(0.002)	(0.002)	
Share of significant estimates ($p \leq 0.1$)	10.2%	9.5%	
PKH assignment			0.027***
			(0.010)
Rand. Inference p-value			0.00
Community FE, year FE	Yes	Yes	Yes
District-year FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
No. repetitions	10,000	10,000	1,000
Mean Adj. R^2	0.221	0.221	

Note: The *Roll-out sample* includes 28,873 communities, while the *RCT sample* comprises 1,830 communities. The dependent variable is a binary indicator that takes the value one if NVMS reported at least one violent crime incident in a community in a given year. The table shows statistics on TWFE estimates after randomly distributed PKH treatments for 10,000 repetitions. Randomization across years assigns the PKH start randomly across all years after 2006, whereas within-year randomization assigns the yearly PKH beginning randomly within the same year. For randomization inference p-values reported in column 3, the STATA package *ritest* (Heß, 2017) was used with $N = 1,000$ resampling iterations. Standard errors in parentheses are clustered at the level of sub-districts. */**/*** denote significance levels at 10/5/1% respectively.

TABLE D.7: Robustness: PKH effects after adjusting standard errors

Sample	*RCT* (LATE)			*Roll-out*		
Standard Errors	Basic	District	Conley	Basic	District	Conley
	(1)	(2)	(3)	(4)	(5)	(6)
PKH treatment	0.032***	0.032***	0.032***	0.007**	0.007**	0.007***
	(0.012)	(0.011)	(0.010)	(0.003)	(0.003)	(0.002)
Community FE, year FE	Yes	Yes	Yes	Yes	Yes	Yes
District-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10,980	10,980	10,956	288,730	288,730	287,940

Note: The *Roll-out sample* includes 28,873 communities, while the *RCT sample* comprises 1,830 communities. For Conley SEs, several communities are excluded due to missing GPS coordinates. The dependent variable is a binary indicator that takes the value one if NVMS reported at least one incident of violent crime in a community in a given year. Robust standard errors reported in parentheses are clustered at the sub-district level in column 1 and 4, clustered at the district-level in columns 2 and 5 and (Conley-) clustered within a 50km threshold around respective community centroids in column 3 and 6. Alternative thresholds at 100km or 200km do not alter the estimated SEs significantly. Results are available from the authors upon request. */**/*** denote significance levels at 10/5/1% respectively.

TABLE D.8: Robustness: PKH effects by sample construction

Sample	*RCT* (LATE)		*Roll-out*	
	Extended 2005-2010 (1)	Sub-district 2005-2010 (2)	Ever-teated 2000-2014 (3)	All 2005-2014 (4)
PKH	0.018**	0.023**	0.007**	0.004**
	(0.010)	(0.012)	(0.003)	(0.002)
Community FE, year FE	Yes	Yes	Yes	Yes
District-year FE	Yes	Yes	Yes	Yes
Observations	19,950	1,500	206,985	476,800

Note: The sample in column 1 is restricted to 3,323 communities that were included in the extended RCT of the PKH program. Column 2 displays sub-district level estimates for the *RCT sample* consisting of 250 sub-districts with data coverage. Column 3 includes 13,779 communities that ever received PKH in 2007–2014 and had NVMS coverage already in 2000. Column 4 covers all Indonesian communities with NVMS coverage. The dependent variable is a binary indicator that takes the value one if NVMS reported at least one incident of violent crime within the community in a given year in columns 1, 3 and 4, and it is the share of communities within the sub-district that experience at least one conflict in column 2. PKH treatment indicates whether any household within a given community received transfers from the CCT program in a given year. Robust standard errors are clustered at the sub-district level and reported in parentheses. */**/*** denote significance levels at 10/5/1% respectively.

TABLE D.9: PKH effects on the intensive margin of violent crime

Sample	*RCT*			*Roll-out*	
	TWFE (1)	TWFE (ITT) (2)	IV (LATE) (3)	TWFE (4)	BJS (5)
PKH	0.027**	0.027**	0.032**	0.009***	0.015***
	(0.013)	(0.013)	(0.016)	(0.003)	(0.004)
Community FE, year FE	Yes	Yes	Yes	Yes	Yes
District-year FE	Yes	Yes	Yes	Yes	Yes
Observations	10,980	10,980	10,980	288,730	244,270

Note: The *Roll-out sample* includes 28,873 communities, while the *RCT sample* comprises 1,830 communities. The dependent variable is the inverse hyperbolic sine of the number of NVMS reported violent crime incidents within a community in a given year. Robust standard errors are clustered at the sub-district level and reported in parentheses. */**/*** denote significance levels at 10/5/1% respectively.

TABLE D.10: PKH effects on community-level socio-economic development

Variable	HH share electricity	Bank access Bank 1	Bank 2	Bar	Availability of Soccer	Clinic	Doctor	Phone	Access to Internet	Market 1	Market 2	*ln* Population
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(8)	(9)	(10)	(11)
					Panel A: RCT sample							
PKH	-0.025*	0.030	0.015	-0.016	-0.060	-0.012	-0.014	0.000	0.016	0.023	-0.006	0.006
	(0.013)	(0.061)	(0.034)	(0.011)	(0.037)	(0.014)	(0.038)	(0.010)	(0.010)	(0.020)	(0.017)	(0.016)
Observations	3,650	1,826	1,826	1,826	1,826	3,650	3,650	1,824	1,824	3,650	3,650	3,650
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
					Panel B: Roll-out sample							
PKH	0.007	-0.003	-0.013	0.002	0.005	-0.009	0.014	0.003	0.019**	-0.023	-0.031**	-0.005
	(0.005)	(0.009)	(0.008)	(0.002)	(0.010)	(0.006)	(0.022)	(0.005)	(0.008)	(0.016)	(0.012)	(0.005)
Observations	85,782	57,610	57,610	57,610	57,610	85,782	85,782	56,338	56,338	85,782	85,782	85,782
Community FE, year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Bank 1 refers to the number of commercial banks in a community, while Bank 2 refers to the number of public credit banks. The remaining dependent variables are binary and refer to the availability of entertainment places (bars/karaoke), soccer fields, a health clinic, a medical doctor, a landline phone connection, internet cafes, markets with (market 1) and without (market 2) a permanent building in a given community. All variables are based on PODES. The *Roll-out sample* includes up to 28,873 communities for the years 2007, 2010 and 2013. The *RCT sample* comprises up to 1,830 communities and includes observations for the years 2007 and 2010. It does not include a community fixed effect because RCT assignment began only in 2007. For both samples, some variables are available for 2010 and 2013 only. Results are estimated using equation (5.1). PKH treatment captures actual treatment by indicating whether any household in a given community received PKH in a given year. Robust standard errors are clustered at the sub-district level and reported in parentheses. */**/*** denote significance levels at 10/5/1% respectively.

TABLE D.11: Robustness: PKH effects on conflict

Sample	*RCT*			*Roll-out*	
	TWFE (1)	TWFE (ITT) (2)	IV (LATE) (3)	TWFE (4)	BJS (5)
PKH	-0.003	0.004	0.005	-0.002	0.005
	(0.008)	(0.009)	(0.010)	(0.002)	(0.003)
Community FE, year FE	Yes	Yes	Yes	Yes	Yes
District-year FE	Yes	Yes	Yes	Yes	Yes
Observations	10,980	10,980	10,980	288,730	244,270

Note: The *Roll-out sample* includes 28,873 communities, while the *RCT sample* comprises 1,830 communities. The dependent variable is a binary indicator that takes the value one if NVMS reported at least one incident of conflict (i.e. cases that are not categorized as "violent crime") in a community in a given year. Robust standard errors are clustered at the sub-district level and reported in parentheses. */**/*** denote significance levels at 10/5/1% respectively.

TABLE D.12: Alternative measure: CCT access intensity and the probability of being a victim of violent crime (2007–2011)

Intensity threshold	< 50 or ≥ 50	< 100 or ≥ 100	< 200 or ≥ 200	< 300 or ≥ 300
	TWFE (1)	TWFE (2)	TWFE (3)	TWFE (4)
PKH low intensity	0.001	0.001	0.002	0.005
	(0.005)	(0.004)	(0.004)	(0.003)
PKH high intensity	0.008**	0.008**	0.008**	0.007*
	(0.004)	(0.004)	(0.004)	(0.004)
Mean (control)	0.043	0.043	0.043	0.043
Sub-district FE	Yes	Yes	Yes	Yes
District-year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	1,095,130	1,095,130	1,095,130	1,095,130

Note: All estimates are restricted to households living in sub-districts that received access to the program by 2014. The dependent variable is a binary indicator that takes the value one if a household reported being victim of at least one crime in a given year. The treatment variable turns to one if the indicated number of households within the sub-districts received PKH, where *low* intensity refers to the lower, and *high* intensity to the upper indicated threshold. Controls include indicators for the completed education, age and marital status of the household head, the quintiles of household size, and urban status. Standard errors in parentheses are clustered at the level of sub-districts. */**/*** denote significance levels at 10/5/1% respectively.

TABLE D.13: PKH effects on the community-level presence of police stations

Dependent:	Village Police Station (0/1)			
Sample:	*RCT* (LATE)		*Roll-out*	
	(1)	(2)	(3)	(4)
PKH treatment	0.015	0.014	0.003	-0.001
	(0.015)	(0.011)	(0.003)	(0.003)
Community FE, year FE	No/Yes	No/Yes	Yes	Yes
District-year FE		Yes		Yes
Observations	3,650	3,650	85,782	85,782

Note: The *Roll-out sample* includes 28,873 communities, while the *RCT sample* comprises 1,830 communities. Columns 1–2 only include observations for the years 2007 and 2010 and do not include community fixed effect because RCT assignment began only in 2007. Columns 3–4 only include 2007, 2010 and 2013 given data availability reasons. PKH treatment captures actual treatment by indicating whether any household within a given community received PKH in a given year. Robust standard errors are clustered at the sub-district level and reported in parentheses. */**/*** denote significance levels at 10/5/1% respectively.

TABLE D.14: RCT: PKH effects on peer-group inequality

	ITT			LATE		
Inequality index:	Gini	Theil	CV	Gini	Theil	CV
	(1)	(2)	(3)	(4)	(5)	(6)
PKH Assignment	-0.003	-0.002	-0.008			
	(0.003)	(0.002)	(0.007)			
PKH Treatment				-0.004	-0.003	-0.011
				(0.004)	(0.003)	(0.010)
Observations	2,212	2,212	2,212	2,212	2,212	2,212

Note: Inequality measures are calculated based on household-level expenditures per capita, taking into account all households included in the *RCT sample*. In the analysis only communities with at least 5 household observations per survey round were used. "CV" refers to "Coefficient of Variation". Additional controls: Baseline inequality measure. LATE estimates are obtained from 2SLS. Standard errors in parentheses are clustered at the sub-district level. */**/*** denote significance levels at 10/5/1% respectively.

TABLE D.15: PKH effects on crime: The role of community-level targeting inequality

Sample	*RCT* (LATE)		*Roll-out*	
	(1)	(2)	(3)	(4)
PKH	0.079*	-0.059	0.015***	-0.010
	(0.040)	(0.063)	(0.006)	(0.010)
PKH $\times$ Undercoverage	-0.222		-0.035*	
	(0.159)		(0.019)	
PKH $\times$ Leakage		0.132		0.025
		(0.098)		(0.016)
Community FE, year FE	Yes	Yes	Yes	Yes
District-year FE	Yes	Yes	Yes	Yes
Observations	10,980	10,980	267,680	267,680

Note: The *Roll-out sample* includes 28,873 communities, while the *RCT sample* comprises 1,830 communities. The dependent variable is a binary indicator that takes the value one if NVMS reported at least one incident of violent crime in a community in a given year. Undercoverage/leakage measure the share of PKH non-receiving eligible/receiving ineligible households in surveyed households on the district-level (based on SUSENAS 2014). Standard errors in parentheses are clustered at the level of sub-districts. */**/*** denote significance levels at 10/5/1% respectively.

TABLE D.16: Robustness: Middle-run effects of PKH on work, schooling and idleness by PKH eligibility

Sample:	Young men (aged 18–25)			
PKH eligibility criteria:				
Poverty status:	Poor		Non-poor	
Children (aged < 16) in hh.:	Yes	No	Yes	No
	(1)	(2)	(3)	(4)
Working	-0.030*	-0.011	-0.005	0.017
	(0.017)	(0.007)	(0.006)	(0.020)
Household chores	-0.028**	-0.005	-0.016*	-0.025
	(0.013)	(0.008)	(0.008)	(0.019)
Attending school	-0.004	0.008	-0.010**	0.001
	(0.009)	(0.006)	(0.005)	(0.008)
Staying idle	0.028*	-0.007	-0.005	-0.017
	(0.015)	(0.007)	(0.005)	(0.018)
Observations	41,357	23,915	156,236	219,933
Sub-district FE, year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Note: The dependent variables refer to the last week before the survey and indicate whether an individual reported having engaged in any (i) work activities outside the household, (ii) household chores, (iii) school attendance, or (iv) stayed idle (not engaged in work, household chores, or school). The treatment variable measures the share of PKH recipient communities within a sub-district and year. Results are based on pooled cross-sections of SUSENAS national household survey data (2004 to 2011), restricted to sub-districts that received PKH by 2014. Results distinguish between young men living in PKH eligible and non-eligible households (defined by per capita household expenditures w.r.t. provincial poverty lines and the presence of children of PKH eligible age). Controls include household head's age, gender, education, marital status and household size in quintiles. Standard errors in parentheses are clustered at the sub-district level. */**/*** denote significance levels at 10/5/1% respectively.

D.2 Figures

FIGURE D.1: Overview on the national roll-out of PKH

Notes: PKH access on the sub-district level. Roll-out refers to the first year PKH is operating in a given community. Dashed areas indicate areas in which PKH had not been rolled out by 2014. Data taken from MoSA.

FIGURE D.2: NVMS coverage

Notes: Provinces shaded in grey are covered by the NVMS (16 in total), dashed provinces have only partial coverage if at all. Data taken from NVMS, 2014.

D.3 PKH and its Impact Evaluation

D.3.1 Data

The World Bank collected baseline and endline data as part of an impact evaluation that accompanied the initial roll-out of PKH. The baseline survey was conducted between June to August 2007 (about a month before the first PKH transfers occured), while the endline survey was fielded between October to December 2009.

TABLE D.17: PKH conditionality criteria

Household category	Conditions
Household with pregnant or lactating women	Complete 4 pre-natal visits and take iron tablets during pregnancy; Give birth assisted by a trained professional; Complete two post-natal visits (lactating mothers)
Households with children aged 0–6 years	Ensure children receive all immunizations and take Vitamin A twice a year; Take children for growth monitoring check-ups (monthly for infants up to 11 months old, quarterly for children 1–6 years old
Households with children aged 6-15 years	Enroll children in primary/secondary school and ensure minimum 85% attendance
Households with children aged 16-18 who have not yet completed nine years of schooling	Enroll children in education program to complete nine years of schooling

Note: Table is adapted from Cahyadi et al., 2020.

TABLE D.18: PKH benefit payments (annual; Indonesian Rupiah)

Criteria	Year 2012	Years 2013/14
Fixed base transfer	200,000	300,000
Each child below age of five	800,000	1,000,000
Each woman pregnant/lactating	800,000	1,000,000
Each child attending primary school	400,000	300,000
Each child attending junior secondary school	800,000	1,000,000
Maximum benefit amount	2,200,000	2,800,000

Note: Table is adapted from Nazara et al., 2013.

D.3.2 A note on the construction of key variables

Variables are constructed based on PKH's impact evaluations surveys which interviews one main adult per household (typically the household head). The surveys start with enumerators collecting a household roster that among other information (age, gender, education) collects labor market information on each household member aged 5 and older. More specifically, the respondent is asked the following question for each household member separately: What was the main activity of ... in the last week? The question has six response options: (i) Working, (ii) attending school, (iii) not working, (iv) being retired, (v) unemployed, or (vi) doing household chores.

Based on this information we construct a binary idleness indicator that takes the value of one if the person is neither working, nor doing household chores, nor attending school in the previous week. The other three labor market indicators that we use in the analysis are constructed following the same logic. "Doing market work" refers to having indicated to work in the last week; "household chores" refers to having indicated to have done household chores in the last week; and "being currently enrolled at school" refers to the response "attending school".

D.3.3 RCT-related Tables

TABLE D.19: Covariate balance at baseline

	Control	Treatment	Difference	
	(1)	(2)	(3)	(4)
Panel A: Community level				
Population in 2007 [in 1,000]	3.448	3.472	-0.024	(0.068)
Violent crime in 2005	0.100	0.101	0.001	(0.014)
Violent crime in 2006	0.141	0.119	-0.022	(0.015)
Panel B: Household level				
Rural location	0.714	0.721	0.008	(0.033)
Age (resp.)	24.973	25.020	0.047	(0.379)
Female (resp.)	0.502	0.497	-0.005	(0.008)
Married (resp.)	0.492	0.499	0.007	(0.011)
HH size	5.155	5.115	-0.040	(0.079)
HH wealth	-0.011	-0.030	-0.019	(0.051)
HH has a radio	0.386	0.376	-0.010	(0.017)
HH has a TV	0.527	0.514	-0.013	(0.028)
HH has an antenna	0.010	0.009	-0.001	(0.002)
HH has a fridge	0.026	0.026	0.000	(0.004)
HH has a bike	0.480	0.470	-0.010	(0.036)
HH has a motor bike	0.160	0.159	-0.001	(0.014)
HH has a car	0.004	0.003	-0.001	(0.001)
HH has a cellphone	0.094	0.096	0.002	(0.009)
Total expenditures (monthly)	928.613	905.294	-23.319	(21.728)
Total food expenditures (monthly)	649.693	635.273	-14.420	(15.330)
Total non-food expenditures (monthly)	278.920	270.021	-8.898	(8.961)
PCA index on community engagement	-0.003	-0.059	-0.056	(0.081)
HH engages in community organization	0.780	0.773	-0.007	(0.018)
# community organization HH is involved	0.780	0.774	-0.006	(0.018)
# hh members involved with comm. organizations	2.193	2.090	-0.103	(0.109)
# meetings attended with comm. organ. (last 3 months)	1.659	1.573	-0.086	(0.099)
Works in agriculture (head)	0.647	0.663	0.016	(0.027)
Works in services (head)	0.135	0.124	-0.011	(0.014)
Highest degree is primary (head)	0.101	0.106	0.005	(0.006)
Highest degree is junior secondary (head)	0.029	0.025	-0.004	(0.003)
Highest degree is senior secondary or more (head)	0.014	0.013	-0.001	(0.002)

Note: Panel A shows statistics on 932 treatment and 897 control group communities. Panel B shows statistics on 7,184 treatment and 7,118 control group households. Mean values are reported in columns 1 and 2, respectively. In column 3 the simple differences between control and treatment group communities are displayed. Standard errors are clustered at the sub-district level and reported in parentheses in column 4. */**/*** denote significance levels at 10/5/1% respectively.

TABLE D.20: RCT sample: Covariate balance at the individual level at the time of the baseline

	Group	N	Control	Treatment	Difference	
	(1)	(2)	(3)	(4)	(5)	(6)
Working	Male (age 13–17)	3,552	0.096	0.083	-0.013	(0.013)
Working	Male (age 18–25)	3,821	0.303	0.293	-0.010	(0.021)
Working	Male (age 26–49)	12,540	0.346	0.345	-0.001	(0.020)
Working	Female (age 13–17)	4,068	0.140	0.137	-0.003	(0.013)
Working	Female (age 18–25)	3,760	0.628	0.646	0.018	(0.021)
Working	Female (age 26–49)	12,200	0.947	0.939	-0.008	(0.006)
Household chores	Male (age 13–17)	3,552	0.034	0.036	0.002	(0.007)
Household chores	Male (age 18–25)	3,821	0.465	0.467	0.002	(0.025)
Household chores	Male (age 26–49)	12,540	0.614	0.614	0.001	(0.021)
Household chores	Female (age 13–17)	4,068	0.003	0.004	0.001	(0.002)
Household chores	Female (age 18–25)	3,760	0.007	0.004	-0.002	(0.003)
Household chores	Female (age 26–49)	12,200	0.003	0.003	0.001	(0.001)
Attending school	Male (age 13–17)	3,552	0.636	0.633	-0.003	(0.023)
Attending school	Male (age 18–25)	3,821	0.053	0.045	-0.009	(0.008)
Attending school	Male (age 26–49)	12,540	0.003	0.002	-0.000	(0.001)
Attending school	Female (age 13–17)	4,068	0.613	0.598	-0.016	(0.024)
Attending school	Female (age 18–25)	3,760	0.068	0.056	-0.012	(0.009)
Attending school	Female (age 26–49)	12,200	0.003	0.002	-0.001	(0.001)
Staying idle	Male (age 13–17)	3,552	0.240	0.255	0.016	(0.021)
Staying idle	Male (age 18–25)	3,821	0.180	0.197	0.017	(0.016)
Staying idle	Male (age 26–49)	12,540	0.038	0.039	0.001	(0.004)
Staying idle	Female (age 13–17)	4,068	0.251	0.272	0.020	(0.020)
Staying idle	Female (age 18–25)	3,760	0.299	0.296	-0.003	(0.021)
Staying idle	Female (age 26–49)	12,200	0.048	0.056	0.008	(0.006)

Note: Standard errors are clustered at the sub-district level and reported in column 6. */**/*** denote significance levels at 10/5/1% respectively.

TABLE D.21: RCT sample: Compliance at the community level

	Control	Treatment
	(1)	(2)
Community-level compliance in..		
2007	0.995	0.998
2008	0.890	0.998
2009	0.736	0.998
2010	0.735	0.999
2011	0.572	1
Sub-district-level compliance in..		
2007	0.955	0.985
2008	0.829	0.985
2009	0.676	0.985
2010	0.667	0.993
2011	0.559	1

Note: Table show yearly compliance shares of the 897 (111) control and 933 (127) treated communities (sub-districts).

TABLE D.22: RCT household survey: Receiving PKH at endline

	Control	Treatment
	(1)	(2)
% Receiving PKH	0.095	0.484

Note: Information on PKH recipients are derived from households' self-reports at the endline survey in 2009. The baseline survey did not include information on whom was considered eligible as this was only determined later by the MoSA. Therefore, the compliance rates represent a lower bound of actual compliance.

TABLE D.23: RCT household survey: Attrition

	Control		Treatment		All	
	# obs.	%	# obs.	%	# obs.	%
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline (2006)	7,131	100	7,196	100	14,327	100
Endline (2009)	6,946	0.974	7,024	0.976	13,970	0.975

Note: Table shows attrition rates between PKH's RCT baseline and endline survey. In both, control and treatment group, about 97.5% of all baseline households could be re-surveyed.

TABLE D.24: RCT household survey: Determinants of attrition

	Specification 1	Specification 2
	(1)	(2)
Treatment assignment	-0.002	-0.002
	(0.003)	(0.003)
Observations	14,327	14,327

Note: The table examines the extent to which PKH is related to attrition. It shows coefficients after regressing (OLS) a binary indicator of "being re-surveyed at the endline" on the original treatment assignment indicator. Specification 1 includes district fixed effects and basic controls. Specification 2 in addition includes a set of extended controls. Standard errors are depicted in parentheses and are clustered at the sub-district level. */**/*** denote significance levels at 10/5/1% respectively.

D.4 SUSENAS Description

D.4.1 Sample construction and variables

SUSENAS is a cross-sectional household survey collected annually by Statistics Indonesia (BPS). The data constitutes the empirical foundation for a number of the country's official statistics—in particular poverty, expenditure, demographic, and education statistics. Its sampling frame is determined by BPS and is aimed to be representative at the district level.

Crime data Since 2007, SUSENAS gathers disaggregated information on crime. Crime information relates to self-reported information on whether a household experienced certain types of crime. Over time the SUSENAS crime module became more comprehensive. While in earlier years crime information was collected at the household level, later rounds collected individual-level crime information. Crime types collected in SUSENAS relate to: Theft, robberies, homicide, rape, and other types of crime. For this study we use a binary indicator that is one if any of the above-mentioned types were experienced by the household. Given that SUSENAS's data changed over time we construct two separate samples:

1. **Sample 2007–2011:**
   Our main sample reported in the paper is compiled by pooling SUSENAS data for the years 2007 until 2011. The sample matches approximately the time period of our main analysis that uses the NVMS data. While crime data disaggregated by type is not available from SUSENAS for the pre-2007 period, BPS decided to strip the data from sub-district identifiers from 2011 onwards. Hence, we cannot link SUSENAS with the administrative PKH roll-out data at the sub-district level in later SUSENAS rounds. Data is available both in provinces with and without NVMS coverage.

2. **Sample 2013–2019:**
   The second sample is compiled by pooling SUSENAS data for the years 2013, 2014, 2017, 2018, and 2019. This sample is used in Appendix section D.4.2. The rounds were selected since they include both, information on crime (self-reported) and a household's PKH membership (self-reported). Survey items related to PKH were introduced in 2013, though excluded in 2015 and 2016, until reintroduced again in 2017. In addition, SUSENAS also asks if households reported crime incidents to

the police (by type). The sample is available for both provinces with and without NVMS coverage. Table D.25 presents descriptive crime statistics based on SUSENAS for the years 2007 to 2019.

Time use data For analysing mechanisms, we pool the yearly SUSENAS modules on labor market participation and time use. We rely on the years 2004 to 2011 that contain sub-district identifiers and hence can be connected over time as well as with our PKH access data.

SUSENAS conducts interviews with one main person per household, typically the household head or her/his spouse. During the interviews (in the labor market section of the questionnaire) the respondent is asked to answer the following four questions regarding each single household member aged 10 or older:

1. Did he/she work last week? (yes/no)
2. Did he/she attend school last week? (yes/no)
3. Did he/she do household chores last week? (yes/no)
4. Was he/she engaged in other activities (personal, sports, mosque, etc.) last week? (yes/no)

Based on this information we construct four binary indicator variables. *Market work* takes the value of one if a person was engaged in any type of work except for household chores (answered yes to question 1). *Household chores* takes the value of one if a person has reported doing household chores (question 3). *In school* takes the value of one if the person has reported having attended school during the last week (question 2). Finally, the indicator *Idle* takes the value of one if a person did not pursue any of the three listed activities in the previous week. Thus, while market work, household chores, and school attendance are not mutually exclusive categories but may overlap, idleness is defined by the absence of any other reported activity.

Regressions based on SUSENAS utilize a range of further variables:

- Poverty status of households (defining potential eligibility as CCT recipient households) is measured by comparing monthly household expenditures per capita to the value of the provincial poverty line (in per capita terms) in a given year.
- Household head characteristics record the years of completed education, age and marital status.
- Further household characteristics include household size (in quintiles), and urban status (defined by BPS).
- Individual characteristics capture the completed degree of education (none/ primary/lower secondary/upper secondary/tertiary), age and marital status.

D.4.2 PKH and Crime: PSM

To derive the average treatment effect (ATE) from propensity score matching (PSM) we rely on STATA's *teffects* package. Given that in observational studies, propensity scores are not known, we do not match on the true propensity scores, $p(X)$, but on an estimate of it. More specifically and as common in economics, we consider a generalized linear specification for the propensity score, $p(X) = F(X'\theta)$ and use a Logit as link function.

The link function is specified as follows:

$$logit\ \theta_{icdt} = \ln \frac{\theta_{icdt}}{1-\theta_{icdt}} \quad = \quad Z'_{icdt}\beta + \lambda_d + \theta_t + \epsilon_{icdt}, \tag{D.1}$$

where θ_{icdt} is the probability to participate in PKH for household i in community c in district d in year t. Z'_{icdt} represents an array of household characteristics to predicting PKH participation. Controls at the household level include household size in quintiles, the number of children below the age of 12 and between 13 and 18, rural status, and a set of indicator variables that classify households into expenditure per capita deciles. Further controls relate to the age, gender, marital status, and education of the household head. λ_d and θ_t represent district and time fixed effects. Standard errors are clustered at the community level taking into account that propensity scores are estimated and treatment is set at the community level (Abadie et al., 2016).

Regressions reported in Table D.26 regress an indicator of self-reported crime victimization status on a household's PKH beneficiary status and further controls:

$$Crime\ victim_{jdt} = \psi\, PKH\text{-}Beneficiary_{jdt} + X'_{jdt}\gamma + \theta_{dt} + \epsilon_{jdt}, \qquad \text{(D.2)}$$

where dependent and controls are defined as before. The elements of the control vector X_{jdt} are outlined above.

D.4.3 SUSENAS Tables

TABLE D.25: SUSENAS 2007-2019: Share (%) of households experiencing crime

Year	Theft	Robbery	Violent Crime	Police	Obs.
	(1)	(2)	(3)	(4)	(5)
2007	0.0275	0.0245	0.0520	.	285,186
2008	0.0252	0.0222	0.0474	.	282,387
2009	0.0265	0.0049	0.0314	.	291,753
2010	0.0273	0.0031	0.0304	.	293,716
2011	0.0237	0.0032	0.0269	.	285,186
2012	0.0185	0.0023	0.0208	.	286,113
2013	0.0202	0.0021	0.0223	0.1909	284,063
2014	0.0231	0.0019	0.0250	0.1919	285,400
2015	0.0284	0.0007	0.0291	.	285,908
2016	0.0275	0.0008	0.0283	.	291,414
2017	0.0350	0.0008	0.0358	0.1623	297,276
2018	0.0341	0.0011	0.0352	0.3634	295,155
2019	0.0315	0.0010	0.0325	0.3684	315,672

Note: Statistics are derived by the authors from SUSENAS data and refer to the 16 NVMS provinces. We define violent crime as the sum of theft and robbery incidences. "Police" captures the share of violent crimes that according to survey respondents was reported to the police.

TABLE D.26: Alternative measure: PKH effects on the probability of being a victim of violent crime (PSM)

Region:	NVMS	Non-NVMS
Household sample:	Poor	
	PSM	PSM
	(1)	(2)
PKH beneficiary	0.004*	0.004***
	(0.002)	(0.001)
District-year FE	Yes	Yes
Controls	Yes	Yes
Observations	722,084	722,084

Note: The estimation sample is based on pooled SUSENAS rounds from the years 2013, 2014, 2017, 2018, 2019 and restricted to sub-districts that received the PKH program by 2014. Columns 1 and 2 distinguish between households residing in provinces with and without NVMS coverage. The dependent variable is a binary indicator that takes the value of one if households reported being victim of violent crime (robbery and/or theft) in a given year. Results rely on PSM as described above. Standard errors in parentheses are clustered at the level of sub-districts. */**/*** denote significance levels at 10/5/1% respectively.

TABLE D.27: The effects of PKH on the reporting of violent crime to the police

Parameter	Share (%) reported	
	(1)	(2)
PKH	-0.008	0.004
	(0.018)	(0.023)
Obs.	722,084	722,084
District-year FE	Yes	Yes
Controls	Yes	Yes
Expenditure Controls	No	Yes

Note: The estimation sample is based on pooled SUSENAS rounds from the years 2013, 2014, 2017, 2018, 2019 for NVMS provinces. The dependent variable is derived as the ratio of violent crime reported to the police divided by the total number of violent crime experienced by a household. Robust standard errors reported in parentheses. */**/*** denote significance levels at 10/5/1% respectively.

D.5 Vignette Experiment

The vignette experiment was conducted as an online survey between January 15th and February 1st 2023 in Indonesia. The survey for the vignette experiment was programmed using the platform UNIPARK. Respondents were recruited via one of Asia's largest online panel providers ("dataSpring"). To become eligible for the survey, respondents had to be between 18 and 50 years of age. Survey implementation involved quotas regarding age, education, and region (defined as major islands) in order to obtain a sample that is likely to be more representative of Indonesia's general population (in that age group).

In total, 1,763 persons completed the survey successfully. As shown in Table D.28, the average person who completed the survey was about 32 years old. About 48 percent of respondents were female.

The vignettes in the survey follow a factorial design and portray the same person as the perpetrator: A male person named Budi who is 20 years old and who lives in a city on Indonesia's most populous island (Java).

The three factors were implemented as follows:

- Factor 1: Idleness vs. work (2 attributes)

  Idle: "Is not working or attending school these days"

  Work: "Working 40h per week in a small kiosk"

- Factor 2: Value of the loot (2 attributes)

  Low: "1 Million Rupiah" (approx. USD 100)

  High: "20 Million Rupiah" (approx. USD 2,000)

- Factor 3: Socio-economic background (3 attributes)

  Poor: "Is very poor"

  Poor and PKH: "Is very poor and receives social assistance "Program Keluarga Harapan" by the government"

  Rich: "Is rather rich"

The implemented vignettes read as follows:

"Budi, 20 years old, lives in the city of Purwokerto in Central Java. He still lives with parents and has two younger siblings (aged 5 and 12). His father works as a cab driver while his mother handles the household. Budi is **working 40h per week in a small kiosk. [is not working or attending school these days.]** This morning Budi got arrested by the police and is suspected to have stolen items from a store in a local shopping mall worth IDR **1 Million [20 Million]**. If found guilty a fitting punishment must be imposed. Budi comes from a family that is **very poor [is very poor and receives social assistance "Program Keluarga Harapan" by the government] [is rather rich.]**."

Directly after the respondent had read the text, he/she was asked to rate on a 10-item Likert scale whether he/she believed that Budi has committed the crime. As shown in Table D.28 respondents replies range from 1 (extremely unlikely that Budi committed the crime) to 10 (extremely likely that Budi committed the crime) with people on average rating the suspect with 6.6 (median of 7).

We evaluate the experiment by estimating the following regression model by OLS:

$$Y_i = \alpha + \beta T_i + \mathbf{X}_i^{'}\gamma + \mathbf{Z}_i^{'}\theta + \epsilon_i, \tag{D.3}$$

where Y_i refers to the outcome variable for individual i, α indicates the intercept, $\mathbf{X}^{'}$ refers to individual-level control variables and $\mathbf{Z}^{'}$ includes the other two vignette factors (value of stolen items, socio-economic background of the suspect). Lastly, T_i is a binary indicator variable that takes the value one if the suspect was described as idle. Our main coefficient of interest is β.

Furthermore, we interact the idleness factor with the characteristics describing the subject as poor, or poor from a PKH-recipient household, using the following model:

$$Y_i = \alpha + \beta T_i + \delta T_i \times P_i + \mathbf{X}_i^{'}\gamma + \mathbf{Z}_i^{'}\theta + \epsilon_i, \tag{D.4}$$

where P_i denotes the categories of a poor and poor and PKH-recipient.

TABLE D.28: Summary statistics (vignette experiment sample)

Variable	Mean	Median	SD	Min.	Max.	Obs.
	(1)	(2)	(3)	(4)	(5)	(6)
Covariates						
Age (years)	32.02	32.00	8.74	18.00	55.00	1,763
Female	0.48	0.00	0.50	0.00	1.00	1,763
Dep. variable						
Likelihood perpetrator	6.62	7.00	2.15	1.00	10.00	1,763

Note: The dependent variable reflects the likelihood that respondents believe the described person committed the crime (theft in our case) under investigation. It is coded on a 10-item Likert scale with "1" meaning "extremely unlikely to have committed the crime" and "10" meaning "extremely likely to have committed the crime".

TABLE D.29: Balance Table (vignette experiment sample)

Variable	Mean values for each sub-sample					Test for differences between sub-samples			
	Poor	PKH	Rich	Works	Idle	Poor vs. PKH	Poor vs. Rich	PKH vs. Rich	Works vs. Idle
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Loot	0.495	0.482	0.502	0.480	0.506	0.013	-0.007	-0.020	0.027
						(0.029)	(0.029)	(0.029)	(0.024)
Idle	0.488	0.489	0.522	0.000	1.000	-0.001	-0.034	-0.033	1.000
						(0.029)	(0.029)	(0.029)	(0.000)
Age	32.218	31.889	31.950	31.793	32.246	0.328	0.268	-0.060	0.454
						(0.513)	(0.507)	(0.512)	(0.416)
Female	0.468	0.489	0.492	0.486	0.479	-0.021	-0.024	-0.003	-0.007
						(0.029)	(0.029)	(0.029)	(0.024)
Poor	1.000	0.000	0.000	0.341	0.326	1.000	1.000	0.000	-0.016
						(0.000)	(0.000)	(0.000)	(0.022)
PKH	0.000	1.000	0.000	0.336	0.321	-1.000	0.000	1.000	-0.014
						(0.000)	(0.000)	(0.000)	(0.022)
Rich	0.000	0.000	1.000	0.323	0.353	0.000	-1.000	-1.000	0.030
						(0.000)	(0.000)	(0.000)	(0.023)
Guilt verdict	6.721	6.556	6.579	6.475	6.763	0.165	0.142	-0.023	0.288
						(0.128)	(0.123)	(0.125)	(0.102)
Obs.	588	579	596	882	881	1,167	1,184	1,175	1,763

Note: The balance table divides the sample into three groups according to the description of the perpetrator's socio-economic background. The background categories are "Poor" (poor background without explicit mentioning of PKH), "PKH" (poor background in which parents receive PKH), and "Rich" (wealthy background). Other variables are defined as follows: "Guilt verdict" refers to our outcome variable to what extent the suspect committed the crime. "Loot" refers to the share of the sample that received vignettes in which the crime suspect is stealing is something of high value. "Idle" refers to the share of the sample in which the crime suspect is described as being idle (neither work nor schooling). Standard errors are shown in parentheses and were estimated via OLS using heteroskedastic robust standard error adjustments.

Bibliography

Abadie, A. and G. Imbens (2016). "Matching on the Estimated Propensity Score". In: *Econometrica* 84.2, pp. 781–807.

Acemoglu, D., S. Johnson, and J. A. Robinson (2005). "Institutions as a fundamental cause of long-run growth". In: *Handbook of economic growth* 1, pp. 385–472.

Alatas, V., A. Banerjee, R. Hanna, B. Olken, R. Purnamasari, and M. Wai-Poi (2016b). "Self-targeting: Evidence from a field experiment in Indonesia". In: *Journal of Political Economy* 4.2, pp. 371–427.

Alatas, V., A. Banerjee, R. Hanna, B. Olken, R. Purnamasari, and M. Wai-Poi (2019). "Does elite capture matter? Local elites and targeted welfare programs in Indonesia". In: *American Economic Association: Papers and Proceedings* 109, pp. 334–339.

Alatas, V., N. Cahyadi, E. Ekasar, S. Harmoun, B. Hidayat, E. Janz, and J. Jel-Lema (2011). "Program Keluarga Harapan: Main Findings from the Impact Evaluation of Indonesia's Pilot Household Conditional Cash Transfer Program". Working Paper 72506. Washington, D.C.: World Bank.

Alatas, V., A. Banerjee, A. G. Chandrasekhar, R. Hanna, and B. A. Olken (2016a). "Network Structure and the Aggregation of Information: Theory and Evidence from Indonesia". In: *American Economic Review* 106.7, pp. 1663–1704.

Aleksandrova, M., S. Balasko, M. Kaltenborn, D. Malerba, P. Muche, O. Neuschäfer, K. Radtke, et al. (2021). "World Risk Report 2021". Berlin: Bündnis Entwicklung Hilft, Ruhr University Bochum – Institute for International Law of Peace, and Armed Conflict (IFHV).

Alesina, A., R. Baqir, and W. Easterly (1999). “Public goods and ethnic divisions”. In: *The Quarterly Journal of Economics* 114.4, pp. 1243–1284.

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003). “Fractionalization”. In: *Journal of Economic Growth* 8.2, pp. 155–194.

Alesina, A., C. Gennaioli, and S. Lovo (2019). “Public Goods and Ethnic Diversity: Evidence from Deforestation in Indonesia”. In: *Economica* 86.341, pp. 32–66.

Alkire, S. and M. E. Santos (2014). “Measuring acute poverty in the developing world: Robustness and scope of the multidimensional poverty index”. In: *World Development* 59, pp. 251–274.

Amacher, G. S., M. Ollikainen, and E. Koskela (2012). “Corruption and forest concessions”. In: *Journal of Environmental Economics and Management* 63.1, pp. 92 –104.

Amuedo-Dorantes, C. and S. Pozo (2006). “Migration, Remittances, and Male and Female Employment Patterns”. In: *American Economic Review* 96.2, pp. 222–226.

Angelsen, A. (2007). “Forest cover change in space and time: Combining the von Thunen and forest transition theories”. World Bank Policy Research Working Paper Series 4117. Washington, D.C.: The World Bank.

Angelucci, M. and G. De Giorgi (2009). “Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles’ Consumption”. In: *American Economic Review* 99.1, pp. 486–508.

Angrist, J. and J.-S. Pischke (2008). “Mostly Harmless Econometrics: An Empiricist’s Companion”. Princeton University Press.

Ashton, L., J. Friedman, D. Goldemberg, M. Z. Hussain, T. Kenyon, A. Khan, and M. Zhou (2023). “A Puzzle with Missing Pieces: Explaining the Effectiveness of World Bank Development Projects”. In: *The World Bank Research Observer* 38.1, pp. 115–146.

Athey, S. and G. W. Imbens (2017). "The econometrics of randomized experiments". In: *Handbook of economic field experiments*. Vol. 1. Elsevier, pp. 73–140.

Attanasio, O., L. Pellerano, and S. Polanía-Reyes (2009). "Building Trust? Conditional Cash Transfer Programmes and Social Capital". In: *Fiscal Studies* 30.2, pp. 139–177.

Attanasio, O., S. Polania-Reyes, and L. Pellerano (2015). "Building social capital: Conditional cash transfers and cooperation". In: *Journal of Economic Behavior & Organization* 118, pp. 22–39.

Attanasio, O., L. C. Sosa, C. Medina, C. Meghir, and C. M. Posso-Suárez (2021). "Long Term Effects of Cash Transfer Programs in Colombia". Working Paper 29056. National Bureau of Economic Research.

Auspurg, K., T. Hinz, and C. Sauer (2017). "Why Should Women Get Less? Evidence on the Gender Pay Gap from Multifactorial Survey Experiments". In: *American Sociological Review* 82.1, pp. 179–210.

Austin, K. G., A. Schwantes, Y. Gu, and P. S. Kasibhatla (2019). "What causes deforestation in Indonesia?" In: *Environmental Research Letters* 14.2, p. 024007.

Aydemir, A. and G. J. Borjas (2007). "Cross-Country Variation in the Impact of International Migration: Canada, Mexico, and the United States". In: *Journal of the European Economic Association* 5.4, pp. 663–708.

Backhaus, A., I. Martinez-Zarzoso, and C. Muris (2015). "Do Climate Variations Explain Bilateral Migration? A Gravity Model Analysis". In: *IZA Journal of Migration* 4.1, pp. 1–15.

Bal, C. S. and W. Palmer (2020). "Indonesia and Circular Labor Migration: Governance, Remittances and Multi-Directional Flows". In: *Asian and Pacific Migration Journal* 29.1, pp. 3–11.

Bandiera, O. and G. Levy (2011). "Diversity and the power of the elites in democratic societies: Evidence from Indonesia". In: *Journal of Public Economics* 95.11, pp. 1322–1330.

Banerjee, A., R. Hanna, B. Olken, and S. Sumarto (2020). "The (lack of) distortionary effects of proxy-means tests: Results from a nationwide experiment in Indonesia". In: *Journal of Public Economics Plus* 1.

Banerjee, A., R. Hanna, B. A. Olken, and D. Sverdlin-Lisker (2022). "Social Protection in the Developing World". Mimeo.

Bank Indonesia (2009). "Laporan Survei Nasional Pola Remitansi TKI". Jakarta: Direktorat Statistik Ekonomi dan Moneter Bank Indonesia.

Barr, C. M., I. A. P. Resosudarmo, A. Dermawan, J. McCarthy, M. Moeliono, and B. Setiono (2006). "Decentralization of forest administration in Indonesia: Implications for forest sustainability, economic development, and community livelihoods". CIFOR Report. Center for International Forestry Research (CIFOR).

Barrera-Osorio, F., M. Bertrand, L. Linden, and F. Perez-Calle (2011). "Improving the Design of Conditional Transfer Programs: Evidence from a Randomized Education Experiment in Colombia". In: *American Economic Journal: Applied Economics* 3.2, pp. 167–195.

Barron, P., S. Jaffrey, and A. Varshney (2014). "How large conflicts subside: Evidence from Indonesia". Indonesian Social Development Paper 18. Jakarta: World Bank.

Barron, P., K. Kaiser, and M. Pradhan (2009). "Understanding Variations in Local Conflict: Evidence and Implications from Indonesia". In: *World Development* 37.3, pp. 698–713.

Batista, C., A. Lacuesta, and P. C. Vicente (2012). "Testing the 'Brain Gain' Hypothesis: Micro Evidence from Cape Verde". In: *Journal of Development Economics* 97.1, pp. 32–45.

Battigalli, P. and M. Dufwenberg (2007). "Guilt in Games". In: *American Economic Review* 97.2, pp. 170–176.

Baumeister, R., A. Stillwell, and T. Heatherton (1994). "Guilt: An Interpersonal Approach". In: *Psychological Bulletin* 115.2, pp. 243–267.

Baute, S., F. Nicoli, and F. Vandenbroucke (2021). "Conditional generosity and deservingness in public support for European unemployment risk sharing". In: *Journal of Common Market Studies* 60.3, pp. 721–740.

Bazzi, S. (2012). "International Migration from Indonesia: Stylized Facts". Mimeo.

Bazzi, S. (2017). "Wealth Heterogeneity and the Income Elasticity of Migration". In: *American Economic Journal: Applied Economics* 9.2, pp. 219–255.

Bazzi, S., R. A. Blair, C. Blattman, O. Dube, M. Gudgeon, and R. Peck (2022). "The Promise and Pitfalls of Conflict Prediction: Evidence from Colombia and Indonesia". In: *The Review of Economics and Statistics* 104.4, pp. 764–779.

Bazzi, S., L. Cameron, S. G. Schaner, and F. Witoelar (2021a). "Information, Intermediaries, and International Migration". NBER Working Paper 29588. National Bureau of Economic Research.

Bazzi, S. and M. Gudgeon (2021b). "The Political Boundaries of Ethnic Divisions". In: *American Economic Journal: Applied Economics* 13.1, pp. 235–66.

Becker, G. (1968). "Crime and Punishment: An Economic Approach". In: *Journal of Political Economy* 76.2, pp. 169–217.

Beegle, K. and L. Christiaensen (2019). "Accelerating poverty reduction in Africa". Washington, D.C.: World Bank Publications.

Beine, M., A. Boucher, B. Burgoon, M. Crock, J. Gest, M. Hiscox, P. McGovern, H. Rapoport, J. Schaper, and E. Thielemann (2016). "Comparing Immigration Policies: An Overview from the IMPALA Database". In: *International Migration Review* 50.4, pp. 827–863.

Beine, M., F. Docquier, and H. Rapoport (2008). "Brain Drain and Human Capital Formation in Developing Countries: Winners and Losers". In: *Economic Journal* 118.528, pp. 631–652.

Beine, M. and C. Parsons (2015). "Climatic Factors as Determinants of International Migration". In: *The Scandinavian Journal of Economics* 117.2, pp. 723–767.

Bell, B., A. Bindler, and S. Machin (2018). "Crime Scars: Recessions and the Making of Career Criminals". In: *The Review of Economics and Statistics* 100.3, pp. 392–404.

Bell, B., R. Costa, and S. Machin (2022). "Why does education reduce crime?" In: *Journal of Political Economy* 130.3, pp. 732–765.

Bellemare, M. F. and C. J. Wichman (2020). "Elasticities and the Inverse Hyperbolic Sine Transformation". In: *Oxford Bulletin of Economics and Statistics* 82.1, pp. 50–61.

Benveniste, H., M. Oppenheimer, and M. Fleurbaey (2020). "Effect of Border Policy on Exposure and Vulnerability to Climate Change". In: *Proceedings of the National Academy of Sciences of the United States of America* 117.43, pp. 26692–26702.

Berman, E., J. Shapiro, and J. Felter (2011). "Can Hearts and Minds Be Bought? The Economics of Counterinsurgency in Iraq". In: *Journal of Political Economy* 119.4, pp. 766–819.

Bertoli, S., F. Docquier, H. Rapoport, and I. Ruyssen (2022). "Weather Shocks and Migration Intentions in Western Africa: Insights from a Multilevel Analysis". In: *Journal of Economic Geography* 22.2, pp. 289–323.

Besley, T. and S. Coate (2003). "Centralized versus decentralized provision of local public goods: a political economy approach". In: *Journal of Public Economics* 87.12, pp. 2611–2637.

Biscaye, P. E., T. W. Reynolds, and C. L. Anderson (2017). “Relative effectiveness of bilateral and multilateral aid on development outcomes”. In: *Review of Development Economics* 21.4, pp. 1425–1447.

Blackman, A. and R. Bluffstone (2021). “Decentralized forest management: Experimental and quasi-experimental evidence”. In: *World Development* 145, p. 105509.

Blanc, M., T. Esmail, C. Mascarell, and J. R. Rodriguez (2016). “Predicting project outcomes: A simple methodology for predictions based on project ratings”. World Bank Policy Research Working Paper 7800.

Bluhm, R., C. Lessmann, and P. Schaudt (2021). “The political Geography of cities”. CESifo Working Paper 9376. Munich: CESifo.

Blumenstock, J. E., N. Eagle, and M. Fafchamps (2016). “Airtime Transfers and Mobile Communications: Evidence in the Aftermath of Natural Disasters”. In: *Journal of Development Economics* 120, pp. 157–181.

BMZ (2022). “BMZ reform partnerships”. BMZ Policy Paper.Bonn: BMZ.

BNBP (2020). “Data Informasi Bencana Indonesia”. `https://dibi.bnpb.go.id.`. Accessed February 12th, 2021.

Bohra-Mishra, P., M. Oppenheimer, and S. M. Hsiang (2014). “Nonlinear Permanent Migration Response to Climatic Variations but Minimal Response to Disasters”. In: *Proceedings of the National Academy of Sciences* 111.27, pp. 9780–9785.

Bonilla-Mejía, L. and I. Higuera-Mendieta (2019). “Protected areas under weak institutions: Evidence from Colombia”. In: *World Development* 122, pp. 585–596.

Borraz, F. and I. Munyo (2020). “Conditional Cash Transfers and Crime: Higher Income but also Better Loot”. In: *Economics Bulletin* 40.2, pp. 1804–1813.

Borusyak, K., X. Jaravel, and J. Spiess (2021). "Revisiting Event Study Designs: Robust and Efficient Estimation". Working Paper.

Bourguignon, F. and M. Sundberg (2007). "Aid Effectiveness – Opening the Black Box". In: *American Economic Review* 97.2, pp. 316–321.

BPS (2013). "Statistik Kriminal 2012". BPS Statistics. Jakarta: Badan Pusat Statistik.

Brady, D. (2019). "Theories of the Causes of Poverty". In: *Annual Review of Sociology* 45, pp. 155–175.

Bratsberg, B., Ø. Hernes, S. Markussen, O. Raaum, and K. Røed (2019). "Welfare Activation and Youth Crime". In: *The Review of Economics and Statistics* 101.4, pp. 561–574.

Briggs, R. C. (2020). "Results from single-donor analyses of project aid success seem to generalize pretty well across donors". In: *The Review of International Organizations* 15.4, pp. 947–963.

Brito, B., P. Barreto, A. Brandão, S. Baima, and P. H. Gomes (2019). "Stimulus for land grabbing and deforestation in the Brazilian Amazon". In: *Environmental Research Letters* 14.6, p. 064018.

Bryan, G., S. Chowdhury, A. M. Mobarak, M. Morten, and J. Smits (2021). "Encouragement and distortionary effects of conditional cash transfers". IZA Discussion Paper 14326. Bonn: IZA Institute of Labor Economics.

Bryan, G. and M. Morten (2019). "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia". In: *Journal of Political Economy* 127.5, pp. 2229–2268.

Buil-Gil, D., A. Moretti, and S. Langton (2020). "The integrity of crime statistics: Assessing the impact of police data bias on crime mapping". Center for Open Science.

Bulman, D., W. Kolkma, and A. Kraay (2017). "Good countries or good projects? Comparing macro and micro correlates of World Bank and Asian

Development Bank project performance". In: *The Review of International Organizations* 12, pp. 335–363.

Burgess, R., F. Costa, and B. A. Olken (2019). "The Brazilian Amazon's double reversal of fortune". SocArXiv Working Paper.

Burgess, R., M. Hansen, B. A. Olken, P. Potapov, and S. Sieber (2012). "The political economy of deforestation in the tropics". In: *The Quarterly Journal of Economics* 127.4, pp. 1707–1754.

Burzyński, M., C. Deuster, F. Docquier, and J. de Melo (2021). "Climate Change, Inequality, and Human Migration". In: *Journal of the European Economic Association* 20.3, pp. 1145–1197.

Cahyadi, N., R. Hanna, B. Olken, R. Prima, E. Satriawan, and E. Syamsulhakim (2020). "Cumulative Impacts of Conditional Cash Transfer Programs: Experimental Evidence from Indonesia". In: *American Economic Journal: Economic Policy* 12.4, pp. 88–110.

Cai, R., S. Feng, M. Oppenheimer, and M. Pytlikova (2016). "Climate Variability and International Migration: The Importance of the Agricultural Linkage". In: *Journal of Environmental Economics and Management* 79, pp. 135–151.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs". In: *Econometrica* 82.6, pp. 2295–2326.

Camacho, A. and D. Mejia (2013). "The externalities of Conditional Cash Transfer programs on crime: The case of Bogota's "Familias en Acción" program". Working Paper. Universidad de los Andes.

Cameron, L. and M. Shah (2014). "Can Mistargeting Destroy Social Capital and Stimulate Crime? Evidence from a Cash Transfer Program in Indonesia". In: *Economic Development and Cultural Change* 62.2, pp. 381–415.

Cameron, L. and M. Shah (2015). "Risk-taking Behavior in the Wake of Natural Disasters". In: *Journal of Human Resources* 50.2, pp. 484–515.

Cantoni, E. (2020). "A Precinct Too Far: Turnout and Voting Costs". In: *American Economic Journal: Applied Economics* 12.1, pp. 61–85.

Card, D., J. Kluve, and A. Weber (2018). "What works? A meta analysis of recent active labor market program evaluations". In: *Journal of the European Economic Association* 16.3, pp. 894–931.

Carpenter, C. (2007). "Heavy Alcohol Use and Crime: Evidence from Underage Drunk-Driving Laws". In: *The Journal of Law & Economics* 50.3, pp. 539–557.

Carr, J. and A. Packham (2019). "SNAP Benefits and Crime: Evidence from Changing Disbursement Schedules". In: *The Review of Economics and Statistics* 101.2, pp. 310–325.

Caselli, F. G., A. F. Presbitero, R Chami, R Espinoza, and P. Montiel (2021). "Aid effectiveness in fragile states". In: *Macroeconomic Policy in Fragile States*. Ed. by R. Chami, R. Espinoza, and P. J. Montiel. Oxford University Press, pp. 493–520.

Cattaneo, A. and S. Robinson (2019a). "Economic development and the evolution of internal migration: moving in steps, returnees, and gender differences". Working Paper 2019/03. FAO.

Cattaneo, C., M. Beine, C. J. Fröhlich, D. Kniveton, I. Martinez-Zarzoso, M. Mastrorillo, K. Millock, E. Piguet, and B. Schraven (2019b). "Human Migration in the Era of Climate Change". In: *Review of Environmental Economics and Policy* 13.2, pp. 189–206.

Cattaneo, C. and G. Peri (2016). "The Migration Response to Increasing Temperatures". In: *Journal of Development Economics* 122, pp. 127–146.

Cattaneo, M. D., M. Jansson, and X. Ma (2020). "Simple local polynomial density estimators". In: *Journal of the American Statistical Association* 115.531, pp. 1449–1455.

Chand, S. and M. A. Clemens (2011). "Skilled Emigration and Skill Creation: A Quasi-Experiment". In: *SSRN Electronic Journal*.

Charness, G. and M. Dufwenberg (2006). “Promises and Partnership”. In: *Econometrica* 74.6, pp. 1579–1601.

Chauvet, L., P. Collier, and M. Duponchel (2010). “What explains aid project success in post-conflict situations?” World Bank Policy Research Working Paper Series 5418.

Chin, E. (2018). “Moved to Opportunity: The Long-Run Effects of Public Housing Demolition on Children”. In: *American Economic Review* 108.10, pp. 3028–3056.

Chioda, L., J. De Mello, and R. Soares (2016). “Spillovers from conditional cash transfer programs: Bolsa Família and crime in urban Brazil”. In: *Economics of Education Review* 54, pp. 306–320.

Christian, C., L. Hensel, and C. Roth (2019). “Income Shocks and Suicides: Causal Evidence From Indonesia”. In: *The Review of Economics and Statistics* 101.5, pp. 905–920.

Chuan, A., J. List, and A. Samek (2021). “Do financial incentives aimed at decreasing interhousehold inequality increase intrahousehold inequality?” In: *Journal of Public Economics* 196, p. 104382.

Cisneros, E., K. Kis-Katos, and N. Nuryartono (2021). “Palm oil and the politics of deforestation in Indonesia”. In: *Journal of Environmental Economics and Management* 108, p. 102453.

Clarke, D. (2017). “Estimating Difference-in-Differences in the Presence of Spillovers”. MPRA Paper 81604. University Library of Munich, Germany.

Clemens, M. A. (2011). “Economics and Emigration: Trillion-Dollar Bills on the Sidewalk?” In: *Journal of Economic Perspectives* 25.3, pp. 83–106.

Collier, P. and A. Hoeffler (2007). “Unintended Consequences: Does Aid Promote Arms Races?” In: *Oxford Bulletin of Economics and Statistics* 69.1, pp. 1–27.

Coniglio, N. D. and G. Pesce (2015). "Climate Variability and International Migration: an Empirical Analysis". In: *Environment and Development Economics* 20.4, pp. 434–468.

Conley, T. (1999). "GMM Estimation with Cross Sectional Dependence". In: *Journal of Econometrics* 92.1, pp. 1–45.

Crost, B., J. Felter, and P. Johnston (2014). "Aid under Fire: Development Projects and Civil Conflict". In: *American Economic Review* 104.6, pp. 1833–1856.

Crost, B., J. H. Felter, and P. Johnston (2016). "Conditional cash transfers, civil conflict and insurgent influence: Experimental evidence from the Philippines". In: *Journal of Development Economics* 118, pp. 171–182.

Cuaresma, J. C. and M. Heger (2019). "Deforestation and economic development: Evidence from national borders". In: *Land Use Policy* 84, e347–e353.

Cuecuecha, A. and R. H. Adams (2016). "Remittances, Household Investment and Poverty in Indonesia". In: *Journal of Finance and Economics* 4.3, pp. 12–31.

Cunha, J., G. De Giorgi, and S. Jayachandran (2019). "The Price Effects of Cash Versus In-Kind Transfers". In: *Review of Economic Studies* 86.1, pp. 240–281.

Dallmann, I. and K. Millock (2017). "Climate Variability and Inter-State Migration in India". In: *CESifo Economic Studies* 63.4, pp. 560–594.

Deininger, K., L. Squire, and S. Basu (1998). "Does economic analysis improve the quality of foreign assistance?" In: *The World Bank Economic Review* 12.3, pp. 385–418.

Deming, D. (2011). "Better Schools, Less Crime?" In: *The Quarterly Journal of Economics* 126.4, pp. 2063–2115.

Denizer, C., D. Kaufmann, and A. Kraay (2013). “Good countries or good projects? Macro and micro correlates of World Bank project performance”. In: *Journal of Development Economics* 105, pp. 288–302.

Deshpande, M. and M. Mueller-Smith (2022). “Does Welfare Prevent Crime? the Criminal Justice Outcomes of Youth Removed from Ssi”. In: *The Quarterly Journal of Economics* 137.4, pp. 2263–2307.

Dollar, D., T. Kleineberg, and A. Kraay (2016). “Growth still is good for the poor”. In: *European Economic Review* 81, pp. 68–85.

Doucouliagos, H. and M. Paldam (2008). “Aid effectiveness on growth: A meta study”. In: *European Journal of Political Economy* 24.1, pp. 1–24.

Drabo, A. and L. M. Mbaye (2015). “Natural Disasters, Migration and Education: an Empirical Analysis in Developing Countries”. In: *Environment and Development Economics* 20.6, pp. 767–796.

Drago, F. and R. Galbiati (2012). “Indirect Effects of a Policy Altering Criminal Behavior: Evidence from the Italian Prison Experiment”. In: *American Economic Journal: Applied Economics* 4.2, pp. 199–218.

Dreher, A., V. Lang, B. P. Rosendorff, and J. R. Vreeland (2022). “Bilateral or Multilateral? International Financial Flows and the Dirty-Work Hypothesis”. In: *The Journal of Politics* 84.4, pp. 1932–1946.

Dressel, J. and H. Farid (2018). “The accuracy, fairness, and limits of predicting recidivism”. In: *Science Advances* 4.1, eaao5580.

Dube, A., T. W. Lester, and M. Reich (2010). “Minimum wage effects across state borders: Estimates using contiguous counties”. In: *The Review of Economics and Statistics* 92.4, pp. 945–964.

Duggan, M. (2001). “More Guns, More Crime”. In: *Journal of Political Economy* 109.5, pp. 1086–1114.

Easterly, W. (2007). “Was development assistance a mistake?” In: *American Economic Review* 97.2, pp. 328–332.

Edwards, R. B., R. L. Naylor, M. M. Higgins, and W. P. Falcon (2020). "Causes of Indonesia's forest fires". In: *World Development* 127, p. 104717.

Egger, D., J. Haushofer, E. Miguel, P. Niehaus, and M. Walker (2022). "General equilibrium effects of cash transfers: Experimental evidence from Kenya". In: *Econometrica* 90.6, pp. 2603–2643.

Elsner, B. (2013). "Does Emigration Benefit the Stayers? Evidence from EU Enlargement". In: *Journal of Population Economics* 26.2, pp. 531–553.

Faguet, J.-P. (2004). "Does decentralization increase government responsiveness to local needs? Evidence from Bolivia". In: *Journal of Public Economics* 88.3-4, pp. 867–893.

Fair, H. and R. Walmsley (2021). "World prison population". Institute for Crime & Justice Policy Research.

Fajnzylber, P., D. Lederman, and N. Loayza (2002). "Inequality and Violent Crime". In: *The Journal of Law and Economics* 45.1, pp. 1–39.

FAO (2018). "The Impact of Disasters and Crises on Agriculture and Food Security". Rome: Food and Agriculture Organization of the United Nations.

Feeny, S. and A. de Silva (2012). "Measuring absorptive capacity constraints to foreign aid". In: *Economic Modelling* 29.3, pp. 725–733.

Feeny, S. and V. Vuong (2017). "Explaining aid project and program success: Findings from Asian Development Bank Interventions". In: *World Development* 90, pp. 329–343.

Ferreira, F., D. Filmer, and N. Schady (2017). "Own and Sibling Effects of Conditional Cash Transfer Programs: Theory and Evidence from Cambodia". In: *Research on Economic Inequality*. Ed. by S. Bandyopadhyay. Vol. 25. Research on Economic Inequality. Emerald Publishing Ltd, pp. 259–298.

Fiarni, C., A. Gunawan, and A. Lestari (2013). "A Fuzzy AHP Decision Support System for SKTM Recipient Selection". In: *Open Access Journal* *of Information Systems* 2013.Information Systems International Conference (ISICO).

Filmer, D., J. Friedman, E. Kandpal, and J. Onishi (2021). “Cash transfers, food prices, and nutrition impacts on ineligible children”. In: *The Review of Economics and Statistics*, pp. 1–45.

Findley, M. G., H. V. Milner, and D. L. Nielson (2017). “The choice among aid donors: The effects of multilateral vs. bilateral aid on recipient behavioral support”. In: *The Review of International Organizations* 12, pp. 307–334.

Fitrani, F., B. Hofman, and K. Kaiser (2005). “Unity in diversity? The creation of new local governments in a decentralising Indonesia”. In: *Bulletin of Indonesian Economic Studies* 41.1, pp. 57–79.

Flavell, A., A. Milan, and S. Melde (2020). “Migration, Environment and Climate Change: Literature Review”. Dessau: German Environment Agency (UBA).

Foley, C. (2011). “Welfare Payments and Crime”. In: *The Review of Economics and Statistics* 93.1, pp. 97–112.

Freedman, M. and E. Owens (2016). “Your Friends and Neighbors: Localized Economic Development and Criminal Activity”. In: *The Review of Economics and Statistics* 98.2, pp. 233–253.

Freedom House (2022). “Freedom in the World Data”. `https://freedomhouse.org/report/freedom-world#Data` (accessed March 8th, 2022).

Freeman, R. (1999). “The economics of crime”. In: *Handbook of Labor Economics*. Ed. by O. Ashenfelter and D. Card. 1st ed. Vol. 3, Part C. Elsevier. Chap. 52, pp. 3529–3571.

Friebel, G., M. Manchin, M. Mendola, and G. Prarolo (2018). “International Migration Intentions and Illegal Costs: Evidence Using Africa-to-Europe Smuggling Routes”. In: *CEPR Discussion Paper* 11978.

Gadenne, L. and M. Singhal (2014). "Decentralization in developing economies". In: *Annual Review of Economics* 6.1, pp. 581–604.

Gaveau, D., M. A. Salim, Husnayaen, and T. Manurung (2022). "Industrial and Smallholder Oil Palm Plantation Expansion in Indonesia from 2001 to 2019 [Data set]". Zenodo. `https://doi.org/10.5281/zenodo.6069212`.

Gelbach, J. and L. Pritchett (2001). "Indicator targeting in a political economy: Leakier can be better". In: *Journal of Economic Policy Reform* 4.2, pp. 113–145.

Gelber, A., A. Isen, and J. Kessler (2015). "The Effects of Youth Employment: Evidence from New York City Lotteries". In: *The Quarterly Journal of Economics* 131.1, pp. 423–460.

Giannelli, G. C. and E. Canessa (2022). "After the Flood: Migration and Remittances as Coping Strategies of Rural Bangladeshi Households". In: *Economic Development and Cultural Change* 70.3, pp. 1159–1195.

Gibson, J. and D. McKenzie (2011). "The Microeconomic Determinants of Emigration and Return Migration of the Best and Brightest: Evidence from the Pacific". In: *Journal of Development Economics* 95.1, pp. 18–29.

Glaeser, E., B. Sacerdote, and J. Scheinkman (1996). "Crime and Social Interactions". In: *The Quarterly Journal of Economics* 111.2, pp. 507–548.

Goodman-Bacon, A. (2021). "Difference-in-differences with variation in treatment timing". In: *Journal of Econometrics* 225.2. Themed Issue: Treatment Effect 1, pp. 254–277.

Gray, C., E. Frankenberg, T. Gillespie, C. Sumantri, and D. Thomas (2014). "Studying Displacement After a Disaster Using Large-Scale Survey Methods: Sumatra After the 2004 Tsunami". In: *Annals of the Association of American Geographers* 104.3, pp. 594–612.

Gray, C. L. and V. Mueller (2012). "Natural Disasters and Population Mobility in Bangladesh". In: *Proceedings of the National Academy of Sciences of the United States of America* 109.16, pp. 6000–6005.

Greenhill, R., P. Carter, C. Hoy, and M. Manuel (2015). "Financing the future: How international public finance should fund a global social compact to eradicate poverty". London: Overseas Development Institute.

Gröger, A. (2021). "Easy Come, Easy Go? Economic Shocks, Labor Migration and the Family Left Behind". In: *Journal of International Economics* 128, p. 103409.

Gröger, A. and Y. Zylberberg (2016). "Internal Labor Migration as a Shock Coping Strategy: Evidence from a Typhoon". In: *American Economic Journal: Applied Economics* 8.2, pp. 123–53.

Gröschl, J. and T. Steinwachs (2017). "Do Natural Hazards Cause International Migration?" In: *CESifo Economic Studies* 63.4, pp. 445–480.

Grossman, G. M. and A. B. Krueger (1995). "Economic Growth and the Environment". In: *The Quarterly Journal of Economics* 110.2, pp. 353–377.

Grossman, G. and J. I. Lewis (2014). "Administrative Unit Proliferation". In: *American Political Science Review* 108.1, pp. 196–217.

Grossman, G., J. H. Pierskalla, and E. Boswell Dean (2017). "Government fragmentation and public goods provision". In: *The Journal of Politics* 79.3, pp. 823–840.

Guha-Sapir, D, R. Below, and P. Hoyois (2021). "EM-DAT: The CRED/OFDA International Disaster Database". `https://www.emdat.be` (accessed February 12, 2021).

Haas, H. de, K. Natter, and S. Vezzoli (2015). "Conceptualizing and Measuring Migration Policy Change". In: *Comparative Migration Studies* 3.1, pp. 1–21.

Haas, H. de, K. Natter, and S. Vezzoli (2018). "Growing Restrictiveness or Changing Selection? The Nature and Evolution of Migration Policies". In: *International Migration Review* 52.2, pp. 324–367.

Hainmueller, J. and D. Hopkins (2015). "'The hidden American immigration consensus: A conjoint analysis of attitudes toward immigrants'". In: *American Journal of Political Science* 59.3, pp. 529–548.

Hallegatte, S. and J. Rozenberg (2017). "Climate change through a poverty lens". In: *Nature Climate Change* 7.4, pp. 250–256.

Halliday, T. (2006). "Migration, Risk, and Liquidity Constraints in El Salvador". In: *Economic Development and Cultural Change* 54.4, pp. 893–925.

Halvorsen, R. and R. Palmquist (1980). "The Interpretation of Dummy Variables in Semilogarithmic Equations". In: *American Economic Review* 70.3, pp. 474–75.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, et al. (2013). "High-resolution global maps of 21st-century forest cover change". In: *Science* 342.6160, pp. 850–853.

Hanson, G. H. (2007). "Emigration, Labor Supply, and Earnings in Mexico". In: *Mexican Immigration to the United States*. University of Chicago Press, pp. 289–328.

Hecht, J. E. (2016). "Indonesia: Cost of Climate Change 2050". USAID Policy Brief. Washington, D.C.

Herrendorf, B., R. Rogerson, and Valentinyi (2014). "Growth and Structural Transformation". In: *Handbook of Economic Growth*. Ed. by P. Aghion and S. N. Durlauf. Vol. 2. Handbook of Economic Growth. Elsevier, pp. 855–941.

Heß, S. (2017). "Randomization inference with Stata: A guide and software". In: *The Stata Journal* 17.3, pp. 630–651.

Hindelang, M., T. Hirschi, and J. Weis (1981). "Measuring delinquency". Maurice Taylor Collection. Sage Publication.

Hjalmarsson, R., H. Holmlund, and M. Lindquist (2015). "The Effect of Education on Criminal Convictions and Incarceration: Causal Evidence from Micro-data". In: *The Economic Journal* 125.587, pp. 1290–1326.

Honig, D. (2020). "Information, power, and location: World Bank staff decentralization and aid project success". In: *Governance* 33.4, pp. 749–769.

Honig, D., R. Lall, and B. C. Parks (2022). "When Does Transparency Improve Institutional Performance? Evidence from 20,000 Projects in 183 Countries". In: *American Journal of Political Science*.

Hoop, J., J. Friedman, E. Kandpal, and F. Rosati (2019). "Child Schooling and Child Work in the Presence of a Partial Education Subsidy". In: *Journal of Human Resources* 54.2, pp. 503–531.

Hornbeck, R. (2012). "The Enduring Impact of the American Dust Bowl: Short- and Long-Run Adjustments to Environmental Catastrophe". In: *American Economic Review* 102.4, pp. 1477–1507.

Integrated Network for Societal Conflict Research (INSCR) (2018). "State Fragility Index and Matrix, Time-Series Data, 1995-2018". `https://www.systemicpeace.org/inscrdata.html` (accessed March 8th 2022.

IOM (2010). "International Migration and Migrant Workers' Remittances in Indonesia". Manila: International Organisation fot Migration.

IPCC (2019). "Climate Change and Land: an IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems". P.R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.-O. Pörtner, D. C. Roberts, P. Zhai, et al., (eds.).

IPCC (2021). "Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change". Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, et al. (eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. In press.

Irawan, S., L. Tacconi, and I. Ring (2013). "Stakeholders' incentives for land-use change and REDD+: The case of Indonesia". In: *Ecological Economics* 87, pp. 75–83.

Isaksson, A.-S. and A. Kotsadam (2018). "Chinese aid and local corruption". In: *Journal of Public Economics* 159, pp. 146–159.

Isham, J. and D. Kaufmann (1999). "The forgotten rationale for policy reform: The productivity of investment projects". In: *The Quarterly Journal of Economics* 114.1, pp. 149–184.

Isham, J., D. Kaufmann, and L. H. Pritchett (1997). "Civil Liberties, Democracy, and the Performance of Government Projects". In: *The World Bank Economic Review* 11.2, pp. 219–242.

Islam, R. (2004). "The nexus of economic growth, employment and poverty reduction: An empirical analysis". Vol. 14. Recovery and Reconstruction Department, International Labour Office Geneva.

Ismail, A. and M. N. Mamat (2002). "The Optimal Age of Oil Palm Replanting". In: *Oil Palm Industry Economic Journal* 2.1, pp. 11–18.

Jabeen, S. (2016). "Do we really care about unintended outcomes? An analysis of evaluation theory and practice". In: *Evaluation and Program Planning* 55, pp. 144–154.

Jacob, B., M. Kapustin, and J. Ludwig (2014). "The Impact of Housing Assistance on Child Outcomes: Evidence from a Randomized Housing Lottery". In: *The Quarterly Journal of Economics* 130.1, pp. 465–506.

Jacob, B. and L. Lefgren (2003). "Are Idle Hands the Devil's Workshop? Incapacitation, Concentration, and Juvenile Crime". In: *American Economic Review* 93.5, pp. 1560–1577.

Jones, B. and B. C. O'Neill (2016). "Spatially Explicit Global Population Scenarios Consistent with the Shared Socioeconomic Pathways". In: *Environmental Research Letters* 11.8, p. 084003.

Kazianga, H., D. de Walque, and H. Alderman (2014). “School feeding programs, intrahousehold allocation and the nutrition of siblings: Evidence from a randomized trial in rural Burkina Faso”. In: *Journal of Development Economics* 106, pp. 15–34.

KfW (2022a). “KfW Annual Review 2021”. `https://www.kfw.de/About-KfW/Newsroom/Latest-News/Pressemitteilungen-Details_703296.html` (accessed October 30th, 2022).

KfW (2022b). “KfW Development Bank evaluation criteria”. `https://www.kfw-entwicklungsbank.de/International-financing/KfW-Development-Bank/Evaluations/Evaluation-criteria/` (accessed October 9th, 2022).

Khairulbahri, M. (2021). “Analyzing the Impacts of Climate Change on Rice Supply in West Nusa Tenggara, Indonesia”. In: *Heliyon* 7.12, e08515.

Khwaja, A. I. (2009). “Can good projects succeed in bad communities?” In: *Journal of Public Economics* 93.7-8, pp. 899–916.

Kilby, C. (2015). “Assessing the impact of World Bank preparation on project outcomes”. In: *Journal of Development Economics* 115, pp. 111–123.

Kilby, C. and K. Michaelowa (2019). “What influences World Bank project evaluations?” In: *Lessons on foreign aid and economic development: Micro and macro perspectives*. Ed. by N. Dutta and C. R. Williamson. Springer, pp. 109–150.

Kis-Katos, K. and B. S. Sjahrir (2017). “The impact of fiscal and political decentralization on local public investment in Indonesia”. In: *Journal of Comparative Economics* 45.2, pp. 344–365.

Kleemans, M. (2015). “Migration Choice under Risk and Liquidity Constraints”. Paper presented at the 2015 AAEA & WAEA Joint Annual Meeting, July 26-28, San Francisco, California.

Kleemans, M. and J. Magruder (2018). “Labour Market Responses To Immigration: Evidence From Internal Migration Driven By Weather Shocks”. In: *Economic Journal* 128.613, pp. 2032–2065.

Koch, D.-J. and L. Schulpen (2018). "Introduction to the special issue 'unintended effects of international cooperation'". In: *Evaluation and Program Planning* 68, pp. 202–209.

Koenig, C. and D. Schindler (2021). "Impulse Purchases, Gun Ownership, and Homicides: Evidence from a Firearm Demand Shock". In: *The Review of Economics and Statistics*, pp. 1–45.

Kosack, S. (2003). "Effective aid: How democracy allows development aid to improve the quality of life". In: *World Development* 31.1, pp. 1–22.

Krishna, V., M. Euler, H. Siregar, and M. Qaim (2017). "Differential livelihood impacts of oil palm expansion in Indonesia". In: *Agricultural Economics* 48.5, pp. 639–653.

Krishna, V. V. and C. Kubitza (2021). "Impact of Oil Palm Expansion on the Provision of Private and Community Goods in Rural Indonesia". In: *Ecological Economics* 179, p. 106829.

Kubik, Z. and M. Maurel (2016). "Weather Shocks, Agricultural Production and Migration: Evidence from Tanzania". In: *The Journal of Development Studies* 52.5, pp. 665–680.

Lakner, C., D. G. Mahler, M. Negre, and E. B. Prydz (2022). "How much does reducing inequality matter for global poverty?" In: *The Journal of Economic Inequality* 20.3, pp. 559–585.

Levitt, S. (1998). "Juvenile Crime and Punishment". In: *Journal of Political Economy* 106.6, pp. 1156–1185.

Lin, Z., J. Jung, S. Goel, and J. Skeem (2020). "The limits of human predictions of recidivism". In: *Science Advances* 6.7, eaaz0652.

Ludwig, J., G. Duncan, and P. Hirschfield (2001). "Urban Poverty and Juvenile Crime: Evidence from a Randomized Housing-Mobility Experiment". In: *The Quarterly Journal of Economics* 116.2, pp. 655–679.

Machado, D. B., L. C. Rodrigues, D. Rasella, M. Lima Barreto, and R. Araya (2018). "Conditional cash transfer programme: Impact on homicide rates and hospitalisations from violence in Brazil". In: *PloS One* 13.12, e0208925.

Machin, S., O. Marie, and S. Vujić (2011). "The Crime Reducing Effect of Education". In: *Economic Journal* 121.552, pp. 463–484.

Mahajan, P. and D. Yang (2020). "Taken by Storm: Hurricanes, Migrant Networks, and US Immigration". In: *American Economic Journal: Applied Economics* 12.2, pp. 250–77.

Makovec, M., R. S. Purnamasari, M. Sandi, and A. R. Savitri (2018). "Intended versus Unintended Consequences of Migration Restriction Policies: Evidence from a Natural Experiment in Indonesia". In: *Journal of Economic Geography* 18.4, pp. 915–950.

Marchesi, S. and T. Masi (2021). "Delegation of implementation in project aid". In: *The Review of International Organizations* 16, pp. 655–687.

Marchiori, L., J.-F. Maystadt, and I. Schumacher (2012). "The Impact of Weather Anomalies on Migration in Sub-Saharan Africa". In: *Journal of Environmental Economics and Management* 63.3, pp. 355–374.

Marconcini, M., A. Metz-Marconcini, T. Esch, and N. Gorelick (2021). "Understanding current trends in global urbanisation - the world settlement footprint suite". In: *GI_Forum* 9.1, pp. 33–38.

Margono, B. A., P. V. Potapov, S. Turubanova, F. Stolle, and M. C. Hansen (2014). "Primary forest cover loss in Indonesia over 2000–2012". In: *Nature Climate Change* 4.8, pp. 730–735.

Marschall, P. (2018). "Evidence-oriented approaches in development cooperation: experiences, potential and key issues". Working Paper 08/2018. German Development Institute.

Martínez Flores, F., S. Milusheva, and A. R. Reichert (2021). "Climate Anomalies and International Migration". World Bank Policy Research Working Papers 9664.

Mbaye, L. M. and A. Drabo (2017). “Natural Disasters and Poverty Reduction: Do Remittances Matter?” In: *CESifo Economic Studies* 63.4, pp. 481–499.

McKenzie, D. and H. Rapoport (2010). “Self-Selection Patterns in Mexico-US Migration: the Role of Migration Networks”. In: *The Review of Economics and Statistics* 92.4, pp. 811–821.

McLeman, R. (2019). “International Migration and Climate Adaptation in an Era of Hardening Borders”. In: *Nature Climate Change* 9.12, pp. 911–918.

Michalopoulos, S. and E. Papaioannou (2013). “National Institutions and Subnational Development in Africa”. In: *The Quarterly Journal of Economics* 129.1, pp. 151–213.

Millán, T., T. Barham, K. Macours, J. Maluccio, and M. Stampini (2019). “Long-Term Impacts of Conditional Cash Transfers: Review of the Evidence”. In: *The World Bank Research Observer* 34.1, pp. 119–159.

Mishra, P. (2007). “Emigration and Wages in Source Countries: Evidence from Mexico”. In: *Journal of Development Economics* 82.1, pp. 180–199.

Mongaby (2018). “How corrupt elections fuel the sell-off of Indonesia’s natural resources”. https://news.mongabay.com/2018/06/how-corrupt-elections-fuel-the-sell-off-of-indonesias-natural-resources/ (accessed December 11th, 2020).

Morell, A. (2019). “The short arm of guilt – An experiment on group identity and guilt aversion”. In: *Journal of Economic Behavior & Organization* 166, pp. 332–345.

Morgans, C. L., E. Meijaard, T. Santika, E. Law, S. Budiharta, M. Ancrenaz, and K. A. Wilson (2018). “Evaluating the Effectiveness of Palm Oil Certification in Delivering Multiple Sustainability Objectives”. In: *Environmental Research Letters* 13.6, p. 064032.

MoSA (2020). “What is program keluarga harapan?” MoSA Policy Paper. Jakarta: Ministry of Social Affairs, Indonesia.

Mubila, M. M., C. Lufumpa, and S. Kayizzi-Mugerwa (2000). "A Statistical Analysis of Determinants of Project Success: Examples from the African Development Bank". In: *African Development Bank Economic Research Paper* 56.

Muralidharan, K., P. Niehaus, and S. Sukhtankar (2023). "General equilibrium effects of improving public employment programs: Experimental evidence from India". In: *Econometrica*. forthcoming.

Mutiarin, D., Q. P. Tomaro, and D. Almarez (2019). "The war on drugs of Philippines and Indonesia: A literature review". In: *Journal of Public Administration and Governance* 9.1, pp. 41–59.

Myers Jr., S. (1983). "Estimating the Economic Model of Crime: Employment Versus Punishment Effects". In: *The Quarterly Journal of Economics* 98.1, pp. 157–166.

Nazara, S. and S. Rahayu (2013). "Program Keluarga Harapan (PKH): Indonesian Conditional Cash Transfer Programme". Policy Research Brief 42. International Policy Center for Inclusive Growth.

Nunn, N. and N. Qian (2014). "US Food Aid and Civil Conflict". In: *American Economic Review* 104.6, pp. 1630–1666.

Nurfatriani, F., Ramawati, G. K. Sari, W. Saputra, and H. Komarudin (2022). "Oil Palm Economic Benefit Distribution to Regions for Environmental Sustainability: Indonesia's Revenue-Sharing Scheme". In: *Land* 11.9, pp. 1–24.

NVMS (2014). "National Violence Monitoring System". Database. Jakarta: Government of Indonesia / World Bank.

OECD (2022a). "Aid (ODA) commitments to countries and regions [DAC3a]". `https://stats.oecd.org/` (accessed March 24th, 2022).

OECD (2022b). "Global Outlook on Financing for Sustainable Development 2023". Paris: OECD.

OECD (2022c). “The High Level Fora on Aid Effectiveness: A history”. `https://www.oecd.org/dac/effectiveness/thehighlevelforaonaideffectivenessahistory.html` (accessed October 24th, 2022).

Olden, A. and J. Møen (2022). “The Triple Difference Estimator”. In: *The Econometrics Journal* 25.3, pp. 531–553.

Ostwald, K., Y. Tajima, and K. Samphantharak (2016). “Indonesia’s decentralization experiment: Motivations, successes, and unintended consequences”. In: *Journal of Southeast Asian Economies (JSEAE)* 33.2, pp. 139–156.

Panda, D. and J. Barik (2021). “Flooding Tolerance in Rice: Focus on Mechanisms and Approaches”. In: *Rice Science* 28.1, pp. 43–57.

Phillips, L., H. Votey, and D. Maxwell (1972). “Crime, Youth, and the Labor Market”. In: *Journal of Political Economy* 80.3, Part 1, pp. 491–504.

Pierskalla, J. H. (2016). “Splitting the Difference? The Politics of District Creation in Indonesia”. In: *Comparative Politics* 48.2, pp. 249–268.

Pierskalla, J. and A. Sacks (2017). “Unpacking the Effect of Decentralized Governance on Routine Violence: Lessons from Indonesia”. In: *World Development* 90, pp. 213–228.

Pinkovskiy, M. L. (2017). “Growth discontinuities at borders”. In: *Journal of Economic Growth* 22, pp. 145–192.

POLRI (2019). “Jurnal Kriminalitas dan lalu lintas: Dalam angka tahun 2018 dan semester 1 2019”. Pusiknas Bareskim Polri, Pusat Informasi Kriminal Nasional, National Criminal Information Center. Jakarta.

Priebe, J., F. Howell, and P. Pankowska (2014). “Determinants of Access to Social Assistance Programmes in Indonesia: Empirical Evidence from the Indonesian Family Life Survey East 2012”. TNP2K Working Paper 11b-2014.

Qian, N. (2015). “Making Progress on Foreign Aid”. In: *Annual Review of Economics* 7.1, pp. 277–308.

RAND (2022). "Indonesia Family Life Survey 5 (IFLS5) crosswalk". `https://www.rand.org/well-being/social-and-behavioral-policy/data/FLS/IFLS/datanotes.html` (accessed March 20th, 2022.

Rayp, G., I. Ruyssen, and S. Standaert (2017). "Measuring and explaining cross-country immigration policies". In: *World Development* 95, pp. 141–163.

Rigaud, K., A. De Sherbinin, B. Jones, J. Bergmann, V. Clement, K. Ober, J. Schewe, et al. (2018). "Groundswell: Preparing for Internal Climate Migration." Washington, D.C.: The World Bank.

Rommel, T. and P. Schaudt (2020). "First impressions: How leader changes affect bilateral aid". In: *Journal of Public Economics* 185, p. 104107.

Rondhi, M., A. Fatikhul Khasan, Y. Mori, and T. Kondo (2019). "Assessing the Role of the Perceived Impact of Climate Change on National Adaptation Policy: The Case of Rice Farming in Indonesia". In: *Land* 8.5.

Sah, R. (1991). "Social Osmosis and Patterns of Crime". In: *Journal of Political Economy* 99.6, pp. 1272–1295.

Sahide, M. A. K. and L. Giessen (2015). "The fragmented land use administration in Indonesia - Analysing bureaucratic responsibilities influencing tropical rainforest transformation systems". In: *Land Use Policy* 43, pp. 96–110.

Schmidheiny, K. and S. Siegloch (2019). "On Event Study Designs and Distributed-Lag Models: Equivalence, Generalization and Practical Implications". Cesifo Working Paper 7481. Munich: CESifo.

Shin, W., Y. Kim, and H.-S. Sohn (2017). "Do different implementing partnerships lead to different project outcomes? Evidence from the World Bank project-level evaluation data". In: *World Development* 95, pp. 268–284.

Shrestha, S. A. (2017). "No Man Left Behind: Effects of Emigration Prospects on Educational and Labour Outcomes of Non-migrants". In: *Economic Journal* 127.600, pp. 495–521.

Sjahrir, B. S., K. Kis-Katos, and G. G. Schulze (2013). "Political budget cycles in Indonesia at the district level". In: *Economics Letters* 120.2, pp. 342–345.

Sjahrir, B. S., K. Kis-Katos, and G. G. Schulze (2014). "Administrative Overspending in Indonesian Districts: The Role of Local Politics". In: *World Development* 59, pp. 166–183.

Smith, J., K. Obidzinski, S. Subarudi, and I. Suramenggala (2003). "Illegal logging, collusive corruption and fragmented governments in Kalimantan, Indonesia". In: *International Forestry Review* 5.3, pp. 293–302.

Spaan, E. and T. van Naerssen (2018). "Migration Decision-Making and Migration Industry in the Indonesia–Malaysia Corridor". In: *Journal of Ethnic and Migration Studies* 44.4, pp. 680–695.

Sparrow, R., A. Suryahadi, and W. Widyanti (2013). "Social Health Insurance for the Poor: Targeting and Impact of Indonesia's Askeskin Programme". In: *Social Science & Medicine* 96, pp. 264–271.

Spitzer, Y., G. Tortorici, and A. Zimran (2020). "International Migration Responses to Natural Disasters: Evidence from Modern Europe's Deadliest Earthquake". In: *CEPR Discussion Paper, DP15008*.

Stark, O. et al. (1991). "Migration in LDCs: risk, remittances, and the family". In: *Finance and Development* 28.4, pp. 39–41.

Strauss, J., F. Witoelar, B. Sikoki, and A. Wattie (2009). "The Fourth Wave of the Indonesian Family Life Survey (IFLS4): Overview and Field Report".

Suarez, D. and P. Maitra (2021). "Health spillover effects of a conditional cash transfer program". In: *Journal of Population Economics* 34.3, pp. 893–928.

Suhayo, W., A. Akhmadi, Hastuti, R. Filaili, S. Budiati, and W. Munawar (2005). "Developing a Poverty Map for Indonesia: A Tool for Better Targeting in Poverty Reduction and Social Protection Programs Book 1: Technical Report." Jakarta: SMERU Research Institute.

Sun, L. and S. Abraham (2021). "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects". In: *Journal of Econometrics* 225.2. Themed Issue: Treatment Effect 1, pp. 175–199.

Tangney, J. (1995). "Recent Advances in the Empirical Study of Shame and Guilt". In: *American Behavioral Scientist* 38.8, pp. 1132–1145.

The Washington Post (2011). "Saudi Beheading Fuels Backlash in Indonesia". Newspaper article (accessed April 5th, 2022). `https://www.washingtonpost.com/world/asia-pacific/saudi-beheading-fuels-backlas`.

Theoharides, C. (2020). "The Unintended Consequences of Migration Policy on Origin-Country Labor Market Decisions". In: *Journal of Development Economics* 142, p. 102271.

Thiede, B. C. and C. L. Gray (2017). "Heterogeneous Climate Effects on Human Migration in Indonesia". In: *Population and Environment* 39.2, pp. 147–172.

Thung, P. H. (2019). "Decentralization of government and forestry in Indonesia". CIFOR Working Paper 249. Center for International Forestry Research (CIFOR).

Triyana, M. (2016). "Do Health Care Providers Respond to Demand-Side Incentives? Evidence from Indonesia". In: *American Economic Journal: Economic Policy* 8.4, pp. 255–288.

UN General Assembly (2015). "Transforming our world: the 2030 Agenda for Sustainable Development". UN Resolution A/RES/70/1. United Nations.

UNEP (2016). "Fiscal Incentives for Indonesian Palm Oil Production: Pathways for alignment with green growth". United Nations Environment Program Technical Report.

UNODC (2021). "Country profile Indonesia". `https://dataunodc.un.org/content/Country-profile?country=Indonesia` (accessed December 12th, 2021).

Wane, W. (2004). “The quality of foreign aid: Country selectivity or donors incentives?” World Bank Policy Research Working Paper 3325.

Watson, B., M. Guettabi, and M. Reimer (2020). “Universal Cash and Crime”. In: *The Review of Economics and Statistics* 102.4, pp. 678–689.

Wehkamp, J., N. Koch, S. Lübbers, and S. Fuss (2018). “Governance and deforestation — a meta-analysis in economics”. In: *Ecological Economics* 144, pp. 214–227.

Winters, M. S. (2019). “Too many cooks in the kitchen? The division of financing in World Bank projects and project performance”. In: *Politics and Governance* 7.2, pp. 117–126.

Wood, T., S. Otor, and M. Dornan (2020). “Australian aid projects: What works, where projects work and how Australia compares”. In: *Asia & the Pacific Policy Studies* 7, pp. 171–186.

World Bank (2012a). “PKH conditional cash transfer: Social assistance program and public expenditure review 6”. Jakarta: World Bank.

World Bank (2012b). “Protecting poor and vulnerable households in Indonesia”. Jakarta: The World Bank.

World Bank (2016). “Indonesia: Improving Infrastructure for Millions of Urban Poor”. Press release (accessed April 17th, 2022). `https://www.worldbank.org/en/news/press-release/2016/07/12/indonesia-improving-infrastructure-for-millions-of-urban-poor`.

World Bank (2017). “Indonesia’s Global Workers: Juggling Opportunities and Risks”. Jakarta: World Bank.

World Bank (2021). “Poverty and Shared Prosperity 2022: Correcting Course”. Washington, D.C.: World Bank.

World Bank (2023). “World Development Indicators Online Database”. `https://databank.worldbank.org/source/world-development-indicators` (accessed February 10th, 2023).

Xu, S., H. A. Klaiber, and D. A. Miteva (2022). "Impacts of forest conservation on local agricultural labor supply: Evidence from the Indonesian forest moratorium". In: *American Journal of Agricultural Economics*.

Yang, D. (2011). "Migrant Remittances". In: *Journal of Economic Perspectives* 25.3, pp. 129–52.

Yang, D. and H. Choi (2007). "Are Remittances Insurance? Evidence from Rainfall Shocks in the Philippines". In: *The World Bank Economic Review* 21.2, pp. 219–248.

Zou, H. (2006). "The Adaptive Lasso and Its Oracle Properties". In: *Journal of the American Statistical Association* 101, pp. 1418–1429.

Declaration of Authorship

I, Lennart Reiners, confirm

- that the dissertation "Local Dimensions of Development" that I submitted was produced independently without assistance from external parties, and not contrary to high scientific standards and integrity,
- that I have adhered to the examination regulations, including upholding a high degree of scientific integrity, which includes the strict and proper use of citations so that the inclusion of other ideas in the dissertation are clearly distinguished,
- that in the process of completing this doctoral thesis, no intermediaries were compensated to assist me neither with the admissions or preparation processes, and in this process,
  - no remuneration or equivalent compensation were provided,
  - no services were engaged that may contradict the purpose of producing a doctoral thesis,
- and that I have not submitted this dissertation or parts of this dissertation elsewhere for the purpose of obtaining a doctoral degree.

I am aware that false claims (and the discovery of those false claims now, and in the future) with regards to the declaration for admission to the doctoral examination can lead to the invalidation or revoking of the doctoral degree.

Signed:

Date:

Author contributions

The main part of the thesis builds on four research papers. The contributions to each have been divided among the respective co-authors as follows:

1. **Volume, risk, complexity: What makes development finance projects succeed or fail?**
   The paper is co-authored with Yota Eilers, Jochen Kluve and Jörg Langbein. Conceptualization and writing of the draft manuscript are my own work. Yota Eilers conducted the data preparation. The research design is joint work with Jochen Kluve and Jörg Langbein, who also revised the manuscript. The empirical analysis is joint work of all co-authors.

2. **Losing territory: The effect of administrative splits on land use in the tropics**
   The paper is co-authored with Elías Cisneros and Krisztina Kis-Katos. Conceptualization and research design are the work of Krisztina Kis-Katos and myself. The writing of the draft manuscript, data preparation and analysis are my own work. Elías Cisneros and Krisztina Kis-Katos revised the manuscript.

3. **Confined to Stay: Natural Disasters and Indonesia's Migration Ban**
   The paper is co-authored with Andrea Cinque. Conceptualization and data analysis were conducted by Andrea Cinque. We equally contributed towards data preparation, research design and writing of the manuscript.

4. **Cash Transfers and Violent Crime in Indonesian Communities**
   The paper is co-authored with Elías Cisneros, Krisztina Kis-Katos and Jan Priebe. All authors have equally contributed to the conceptualization, research design, data preparation as well as analysis and writing of the manuscript.