CESifo WORKING PAPERS

10969
2024

February 2024

Wedded to Prosperity? Informal Influence and Regional Favoritism

Pietro Bomprezzi, Axel Dreher, Andreas Fuchs, Teresa Hailer, Andreas Kammerlander, Lennart Kaplan, Silvia Marchesi, Tania Masi, Charlotte Robert, Kerstin Unfried

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10969*

Wedded to Prosperity? Informal Influence and Regional Favoritism

Abstract

We investigate the informal influence of political leaders' spouses on the subnational allocation of foreign aid. Building new worldwide datasets on personal characteristics of political leaders and their spouses as well as on geocoded development aid projects (including new data on 19 Western donors), we examine whether those regions within recipient countries that include the birthplace of leaders' spouses attract more aid during their partners' time in office. Our findings for the 1990–2020 period suggest that regions including the birthplaces of political leaders' spouses receive substantially more aid from European donors, the United States, and China. We find that more aid goes to spousal regions prior to elections and that developmental outcomes deteriorate rather than improve as a consequence. For Western aid but not for China, these results stand in some contrast to those for leader regions themselves. This suggests that aid from Western donors is directed from serving obvious political motives to promoting more hidden ones.

JEL-Codes: D720, F350, O190, O470, P330, R110.

Keywords: informal influence, ODA, favouritism, birth regions, development, political economy.

Pietro Bomprezzi
Kiel Institute for the World Economy
Kiel / Germany
pietro.bomprezzi@ifw-kiel.de

Axel Dreher
Heidelberg University / Germany
mail@axel-dreher.de

Andreas Fuchs
University of Göttingen / Germany
mail@andreas-fuchs.net

Teresa Hailer
Heidelberg University / Germany
teresa.hailer@awi.uni-heidelberg.de

Andreas Kammerlander
Albert-Ludwigs-University Freiburg
Freiburg / Germany
andreas.kammerlander@vwl.uni-freiburg.de

Lennart Kaplan
University of Göttingen / Germany
lennart.kaplan@wiwi.uni-goettingen.de

Silvia Marchesi
University of Milano Bicocca / Italy
silvia.marchesi@unimib.it

Tania Masi
University "G. d'Annunzio" Chieti Pescara / Chieti / Italy
tania.masi@unich.it

Charlotte Robert
Heidelberg / University / Germany
charlotte.robert@awi.uni-heidelberg.de

Kerstin Unfried
Bernhard-Nocht-Institute for Tropical Medicine (BNITM), Hamburg / Germany
kerstin.unfried@bnitm.de

February 19, 2024

We thank participants at the following conferences, seminars, and workshops for helpful comments: the Beyond Basic Questions Workshop (University of Erlangen-Nuremberg, August 2021), the European Public Choice Society (Hannover, March 2023), the Beyond Basic Questions Workshop (University of Stuttgart, June 2023), the International Conference on European Studies (ETH Zurich, June 2023), the Milan PhD Economics Workshop (June 2023), the Political Economy of Aid Conference (Dublin, October 2023), the Paris Sorbonne Sustainable Development Seminar (Paris, November 2023), and the CESifo Workshop on Political Economy (Dresden, November 2023). We are grateful to Christopher Kilby for sharing code that helped us downloading data on geocoded World Bank projects. We also thank Maria Perotta Berlin, Christian Bommer, Raj M. Desai, Richard Jong-A-Pin, Anders Olofsgård, Brad Parks, Marcello Pérez-Alvarez and Shu Yu for sharing their data. We are grateful for support from our excellent research assistants: Eliana Patricia Chavarría Pino, Regina Dworschak, Erin Flannagan, Pascal Haas, Alexander Hudlet, Vivien Hülsen, Timo Kretschmer, Matthew Rudh, Dustin Stanizek, Levani Veshapidze, and Levin Wiebelt from the University of Göttingen; Aurora Brahimllari, and Ramón Rey from the University of Freiburg; Giannis Hafizi and Franziska Hartung from the Helmut-Schmidt University Hamburg; and Tim Deisemann, Yannik Hausmann, Tobias Hellmundt, Felix Höfer, Daniel Overbeck, Paul Schlowack, Constantin Schmidt, Samuel Siewers, and Paul Vogel from Heidelberg University. We thank Tobias Korn for detailed comments on an earlier version of the paper.

1 Introduction

In 1979, when asked about the wealth accumulation among her family's acquaintances and allies during the tenure of her husband, Imelda Marcos, the wife of the then Philippine president, responded: "Well, some are smarter than others" (Branigin 1984). One individual who appears to have profited from his association with the president's wife is Herminio Disini, who at the time was married to one of Imelda Marcos's cousins. Disini obtained a virtual monopoly for his cigarette-filter business by taking advantage of a presidential decree that raised import tariffs for his foreign competitors to 100 percent (Branigin 1984). Ms. Marcos held no elected office—yet, it seems she was able to exercise substantial influence over the allocation of funds.

While anecdotes illustrate the informal influence wielded by powerful politicians' unelected spouses,[1] there is no evidence as to whether and to what degree this influence systematically benefits individuals, companies, or even entire subnational regions associated with powerful spouses. This paper addresses this research gap, analyzing the spouses of political leaders of a maximum of 163 countries over the 1990–2020 period. We focus on spousal birth region to capture areas where we expect them to take particular interest in. Specifically, we test whether subnational regions where the spouse of a leader was born attract more foreign aid at the time the spouse's partner is in office, compared to how much aid that same region receives at other times.

The importance of unelected spouses for policy outcomes and the lack of solid evidence on it has long been recognized. After he left the White House, U.S. President Harry S. Truman expressed the hope that "someday someone will take the time to evaluate the true role of the wife of a President" (Gonnella-Platts and Fritz 2017, 5). As Dahl (1961) points out, indirect influence [of leaders' spouses] is even more difficult to measure than those of elected officials. To ignore such indirect influence, however, is to exclude "what may well prove to be a highly significant" factor in answering who governs (Dahl 1961, cited in O'Connor et al. 1996, 836). According to O'Connor et al. (1996, 849), "[t]he failure of political scientists and historians to consider the political role of first ladies neglects the role of a key player in the president's inner circle"—a failure that we argue extends to economists as well.

[1] Such influence can be defined as "any attempts to affect public policy, executive decision-making, or the course of a political career" (O'Connor et al. 1996, 837). Examples include the wife of the Democratic Party's nominee for the 2022 South Korean presidential election, who had to apologize for using her husband's aides to run errands (see `https://www.thetimes.co.uk/article/leaders-wives-dragged-into-koreas-election-of-unlikeables-w0hw8vcsh`, last accessed July 14, 2023). Another example is first lady Hinda Déby of Chad, who "is also in the management of Chad's oil industry and has gained from lucrative contracts, acted as a mediator between private companies and foreign investors, and appointed relatives and members of her inner circle to strategic positions. Ahmat Khazali Acyl, for example, is Hinda Déby's elder brother and managing director of Société des Hydrocarbures du Tchad (SHT), the national Chadian oil company. Mahamat Kasser Younous, a member of her inner circle, was also a managing director in SHT, while Ibrahim Hissein Bourma, a brother-in-law, is SHT's marketing director" (Van Wyk et al. 2018, 5).

Spouses can exercise political influence via a number of channels. They are important advisors to their partners. U.S. President Truman, for example, consulted his wife on whether or not to use the atomic bomb, fight in Korea, or on the European Recovery Program after World War II (O'Connor et al. 1996). Observers often attribute more power to spouses than to elected vice presidents.[2] They derive status from their marriage, have influence over their partner, and often control access to them (Van Wyk et al. 2018). This might make it difficult for many to turn down requests.[3] Often, spouses develop their own political agendas and patronage networks, allowing them to deliver spoils in exchange for support.[4] It thus seems that spouses of country leaders have the means to favor their birth regions.[5]

We expect spouses of country leaders will leverage their influence to channel resources to their home regions for a number of reasons. First, spouses could cultivate political support for their partners during their time in office, especially during election periods.[6] To achieve the highest voter turnout in their strongholds, political leaders strategically allocate resources to these areas (Dreher et al. 2019). To the extent that regional populations affiliated with the leader's spouse extend their support to the leader, we expect the same to hold true for spouses' birth regions. Adida et al. (2016) argue that such support should be forthcoming because marriage is a credible signal of coalition building between (ethnic) regions. They identify expressive as well as instrumental reasons, highlighting how marriage serves as a signal regarding the likelihood of resource allocation to a particular region. Their findings show that people who share the spouse's ethnicity express substantially greater support for the leader. Second, spouses might use their role to prepare their own future political campaign (Gonnella-Platts and Fritz 2017). Famous examples include Cristina Fernández de Kirchner (Argentina), Sonia

[2]As Hay (1988) points out, while the "vice president [is] a heart beat away (from the president) [...] the first lady can hear it" (cited in O'Connor et al. 1996, 846).

[3]As Uruguayan first lady Maria Julia Pou put it, "[y]ou are not a simple citizen and nobody looks at you as a simple citizen. [W]hen a first lady picks up the phone and calls somebody, it's Somebody."

[4]As Van Wyk et al. (2018, 31) point out, most of the first ladies in their study, among them "[f]irst ladies Janet Museveni, Grace Mugabe and Hinda Déby, for example, have developed a public policy agenda independent of and/or parallel to that of their husbands' government. Their status as the "social-worker-in-chief" (Troy 2006, 142) should not be underestimated as it gives them access to people and money, for example, through the NGOs they are involved with.

[5]Anecdotes abound. Consider Uganda, where according to a local newspaper report first lady Janet Museveni thanked her husband in the name of the Ntungamo people during a 2020 campaign meeting: "thank you for granting Ntungamo a district status, they also commend you for the good roads of Ntungamo-Mirama hills, Kagamba-Ishaka and Ntungamo-Rukungiri that are all tarmacked." The report continues explaining that "Mrs Museveni added that the people of Ntungamo were also very happy for the pineapple factory, youth and women funds, education and health services among many development programs the NRM government has extended to them." Incidentally, Ntungamo District is Janet Museveni's birth region. See `https://www.independent.co.ug/janet-thanks-museveni-for-building-roads-in-ntungamo/,lastaccessedJuly13,2023`.

[6]Following Mrs. Museveni's praise for her husband cited in footnote 5, she assured Museveni of solid support from the region's voters in the upcoming elections.

Gandhi (India), and Hillary Clinton (United States).[7] Again, to the extent that political support from birth regions is particularly strong, focusing resources there in order to maximize turnout at election time can be a winning strategy. Finally, just like leaders themselves, spouses might be economically privileging their home regions in anticipation of returning to these places after their partner's political career comes to an end or might be motivated by parochial altruism (Dreher et al. 2019).

Taken together, spouses have motive and opportunity. We thus investigate whether subnational regions receive more foreign aid at the time a person who originates from such regions is married to the country's political leader compared to how they perform at other times. Western donors exercise substantial effort to avoid the misappropriation of their funds for reasons of recipient country politics. For example, previous work has shown that the birth regions of a country's leader receive more foreign aid from China, but not from the World Bank (Dreher et al. 2019, Dolan and McDade 2020).[8] While Western donors of foreign aid seem more interested in preventing the use of their aid for within-country political purposes than China, it seems unlikely that they can entirely prevent such allocation. To the extent they successfully prevent aid from being targeted to the leaders' birth regions, we expect part of the aid to go to the birth regions of their spouses instead. This difference in scrutiny could result from the donors' primary concern of avoiding any direct political influence or misuse of aid in the leaders' birthplaces. Because political leaders' spouses may not hold formal political office and are not officially involved in government affairs, Western donors might not focus their efforts as intensely on preventing aid from flowing into the regions tied to spouses, potentially allowing some aid to end up benefiting these regions.

Our panel data allow us to test whether the home regions of leaders' spouses receive more aid during their partners' tenures compared to other periods. We use a binary variable equal to one when the region is the birth region of the spouse of the effective political leader. The data allow us to use subnational region fixed effects and fixed effects for country-years. We are in particular interested in whether regions receive more aid in years the spouse of the country's leader was born in a region relative to the years just before. That way, we can disentangle increases in aid from trends in aid that are correlated with the probability that leaders' spouses are born in a specific region. This approach enables us to interpret the difference in aid during these periods as causal. Overall, controlling for the birth regions of the leaders themselves, the focus on spouses allows us to test the effect of informal influence on resource allocation and the contextual

[7] The list of examples is long. Spain's first lady Ana Botella later successfully campaigned for Mayor of Madrid (Gonnella-Platts and Fritz 2017). Other examples include Margarita Cedeño de Fernandez (Dominican Republic), Margarita Zavala de Calderon (Mexico), and Julia Pou (Uruguay) (Gonnella-Platts and Fritz 2017, 35). Likewise, Grace Mugabe (Zimbabwe) demonstrated a clear interest in higher office for herself (Van Wyk et al. 2018).

[8] Dreher et al. (2019) also report first evidence that Chinese aid in Africa flows more freely in regions affiliated with political leaders' spouses.

factors that facilitate such influence.

Our findings suggest that regions including the birthplaces of political leaders' spouses receive substantially more aid, in particular during election periods. We find that these increases in aid are driven by European donors, the United States, and China, but not by the World Bank. European donors and China also give more aid to birth regions of the countries' leaders themselves, compared to what these regions received in the year just before the leader was born there. For European donors, however, the increase in aid to leader birth regions is smaller by an order of magnitude compared to the effect on spousal birth regions. This suggests that Western donors try to avoid channeling their aid to regions where political motives are obvious, while China does not. While prior research implies that Western aid would be less influenced by political interests compared to aid from China (Dreher et al. 2019), we find that the aid is instead given to areas where the motive is more hidden, but political interests might be as stark. Consequently, the effectiveness of such aid may be constrained. Indeed, we find that spousal regions develop less—as measured by nighttime light emissions and infant mortality—and that the aid given while the leaders' spouse was born there is less effective compared to aid given to these regions at other times. These results highlight the informal influence wielded by unelected spouses and its impact on the allocation of resources and developmental outcomes.

This paper contributes to three literatures. The first investigates the influence of elected politicians' relatives. We are not aware of quantitative evidence on elected leaders' spouses on the allocation of resources or developmental outcomes.[9] While this question has received some attention, previous work is qualitative and selective. Examples include studies by Gonnella-Platts and Fritz (2017) on the role of 12 first ladies from five continents, O'Connor et al. (1996) on 38 wives of U.S. Presidents, and Van Wyk et al. (2018) on wives of the 10 longest-serving African presidents. These papers point to substantial influence of some of the first ladies they investigate but given their method of choice necessarily remain selective. This paper is thus the first to systematically evaluate the effects of politicians' unelected spouses on development aid and outcomes.

The second strand of literature to which we contribute investigates the effect of political decision makers on the prosperity and development of favored regions. Thanks to the growing availability of data at the subnational level, scholars have assessed the extent to which country leaders favor their birth region (Barkan and Chege 1989, Moser 2008, Mu and Zhang 2014, Burgess et al. 2015, Do et al. 2017) or their ethnic region (De Luca et al. 2018). Scholars have found substantial evidence of this regional preferential treatment. For instance, home regions experience higher economic growth (Hodler and Raschky

[9]There is, however, quantitative evidence on how children of political decision makers affect policies. Studying voting behavior of U.S. congressmen, Washington (2008) finds that congressmen with daughters are substantially more likely to vote in line with feminist views. McGuirk et al. (2023) show that congressmen with draft-age-sons are less likely to vote for conscription.

2014), receive more public goods (Kramon and Posner 2013, 2016, Burgess et al. 2015), and benefit from larger government transfers and biased taxation (Kasara 2007). Firms located in such regions receive easier access to credit and experience stronger growth in sales and employment after leaders from these regions assume office (Asatryan et al. 2022, Osei-Tutu and Weill 2023). A number of recent papers focus on foreign aid as one specific form of favoritism. They find that home regions of leading politicians receive a disproportionate amount of development aid (Briggs 2014, Dreher et al. 2019, Bommer et al. 2022). Contrary to this literature, we focus on the informal influence of unelected spouses rather than elected politicians. What is more, due to data availability, the previous literature has either focused on one particular recipient country or on aid from China and/or the World Bank. We are the first to investigate the subnational allocation of Western bilateral donors across a large number of recipient countries.

Finally, and less directly, we also relate to the literature on informal governance. This literature investigates the influence of unwritten rules that modify or substitute formal provisions. Such informal influence can override or complement legal procedures, in particular when important rules are unwritten (Stone 2011, 2013). This literature has mostly focused on international organizations and treaties. In this paper, we thus examine the role of informal governance in a new context, that is, we focus on the informal influence on governments that may be exerted by powerful spouses and the settings that facilitate such influence. Donors of development aid may overlook this source of regional favoritism as it is less visible to them or their domestic constituencies and/or may even be advantageous for them, to the extent that it allows to interfere with the politics of recipient countries unnoticed.

Beyond the substantive contributions to the literatures on formal and informal influence, this paper introduces two novel databases, both of which we expect to be useful for research over and beyond this one. First, we introduce what we call the *Political Leaders' Affiliations Database (PLAD)*. The database includes the precise geolocated birthplaces of leaders and their spouses all around the world, as well as information on their education, profession, ethnicity, and their number of children, covering a sample of 177 countries worldwide over the 1989–2020 period. We thus complement existing data on political leaders that provide information on leaders' tenure and details on turnover (Goemans et al. 2009, Licht 2022).

The second database we introduce in this paper—the *Geocoded Official Development Assistance Database (GODAD)*—is a geocoded, dyadic panel dataset of development aid projects assembled from various sources. We geocode data for 18 bilateral European donors and the United States from raw project data in the OECD DAC's Creditor Reporting System (CRS). In a nutshell, CRS project titles and descriptions provide text data from which geographic entities can be identified. We collect and process data from the CRS on our 19 donors for a total of around 1,600,000 unique projects. We then exploit project titles and descriptions to extract candidate geographical entities that can be matched to known cities, regions, or other administrative entities within the recipient country. To do so, we run a Named Entity Recognition (NER) algorithm through a (pre-trained) RoBERTa base transformer model for entity identification on the corresponding project titles or descriptions. This particular class of algorithms employs deep learning models to identify specific categories within a text, including geographic entities. The model finds at least one geopolitical entity for 220,369 projects. From these projects, we identify 285,558 unique project-location pairs, as many projects spread across multiple locations. The extracted entities undergo a series of data cleaning and cross-checking.[10] For this paper, we make use of Western bilateral aid data for these 220,369 projects, which amount to almost US$190 billion over the 1990–2020 period and include a wealth of details regarding lending arrangements and modalities.[11] We complement these data with geocoded information on Chinese aid commitments and World Bank projects, resulting in a first-of its kind dataset of geocoded aid projects that will hopefully serve as important input for future studies on the allocation and effects of development aid.[12]

We proceed as follows. Section 2 describes the construction of PLAD and GODAD and presents descriptive statistics. In Section 3, we explain how we identify the effects of informal influence. Section 4 shows our results, for total aid and separately by donors. This section also investigates heterogeneity, and the consequences of informal influence on development outcomes, and includes an analysis of dynamic treatment effects. We conclude in Section 5.

2 Data

2.1 The Political Leaders' Affiliation Database (PLAD)

To measure the informal influence of political leaders' spouses, we build a new database that contains systematic information on personal characteristics of political leaders and their spouses, including their respective regions of birth. To build the *Political Leaders'*

[10] Appendix B explains the evaluation of the NER model accuracy on our dataset, data cleaning, and how we deal with false positives and negatives.

[11] The CRS provides data from 1973 onwards. However, the share of missing information is substantially higher for earlier years. While we provide more details on the full sample in Appendix B, descriptive statistics in the main text refer to the 1990–2020 period. These are the years that we can use in the econometric analysis below, restricted by the availability of PLAD.

[12] Raw data on Chinese aid projects are available from Custer et al. (2021) and Dreher et al. (2022) for the 2000–2017 period. We clean these raw data and merge them to our administrative regions. AidData (2017) provides geocoded data for World Bank projects for the years 1995–2014. In order to make use of a larger sample we merge them with project-level data from the International Aid Transparency Initiative (IATI), available for the years 1998–2020. We thank Christopher Kilby for sharing code that helps performing this task. Data on World Bank disbursements are from Kersting and Kilby (2021).

Affiliation Database (PLAD), we collect information on the universe of effective political leaders of all countries worldwide, following the definitions in the Archigos Database on Political Leaders (Goemans et al. 2009).[13] Effective political leaders are typically presidents in presidential systems, prime ministers in parliamentary systems, and party chairpersons in communist states. We rely on Archigos data until 2015 and update leader information until the year 2020, relying on online libraries and databases such as the CIA World Factbook, Munzinger, Encyclopaedia Britannica, and Ethnicity of Celebs, as well as on reports from popular news services. We describe the data collection process in detail in Appendix A. Altogether, our database covers 1,078 political leaders with 1,229 terms in office from 176 countries over the 1989–2020 period.

Once we had identified the universe of effective political leaders, we used these same sources to collect data on names and personal characteristics of their spouses and information on their joint offspring. Among others, we have collected information on spouses' date and place of birth as well as their educational and professional background. We also documented how many children the couple had, the gender of these children, along with details on how the relationship with the leader ended (when applicable).[14] We gathered information on up to four spouses a leader has had throughout their lifetime.[15] In polygamous marriages, we identify what we refer to as the 'first spouse.' This designation is assigned to the individual who holds the official or informal title of 'first lady' or 'first gentleman.' In cases where such titles are not employed, our criterion for designation is based on the spouse with whom the leader is most frequently observed in public, as determined by their representative functions. Table A.1 in Appendix A.2 lists all variables with their definitions and sources.

Our estimation sample contains information on spouses for 93% of the leaders in our database, totaling 809 spouses. Among these individuals, 532 spouses were the spouse at some point during the respective leader's tenure. On average, a spouse is 'in power' for 5.6 years. The median spouse holds at least a Bachelor degree or equivalent and has three children with the respective leader.

We geocode birth regions using information from Geonames, relying on seven precision codes (Strandow et al. 2011), with the highest precision being the village or city level.[16]

[13]We make these data available to the public through our project website at `www.plad.me`. We currently include information on leaders only but will add spouse characteristics upon publication of this article. We plan to update these data on a yearly basis. A future version of PLAD will also include geocoded information on birth regions of cabinet members (Asatryan et al. 2023).

[14]We are currently in the process of expanding the PLAD with information on the political careers of spouses. More precisely, we plan to collect information on whether the spouses hold administrative or elected positions before, during, and after the respective leaders' tenure.

[15]Leaders in our dataset have on average 1.3 spouses. The only monogamous leaders in our dataset that had more than four spouses are Michael Manley in Jamaica and Gerhard Schröder in Germany, who both had five spouses. In these two cases, we collected information on all spouses except the most recent one.

[16]See `http://www.geonames.org`. As a secondary source, we rely on Google Maps.

We aggregate the birth region information for administrative regions at the first (ADM1) and second (ADM2) subnational levels.[17] Specifically, we geocoded the birth regions of 298 'spouses in office' (56%) to the respective ADM1 regions. Figure 1 shows a world map of the ADM1 birth regions of political leaders and their spouses, highlighting significant variation across space.[18] 236 spouses (44.4%) were geocodable to the more fine-grained ADM2 region level, leading to 1,210 ADM2 region-year observations.

We observe some overlap between the birth regions of leaders and their spouses. For 23.7% of the ADM2 region-year observations that contain the spouse birth regions, the leader also originates from that same region. The raw correlation between the two variables is 0.207. Out of the 1,210 spousal-year observations in our sample, 287 are leader-year observations as well. Of the 150 regions that change their status as birth region at least once, 12% of the new regions are the birth region of both the leader and their spouse, while their predecessors were born in two different regions. 4.7% correspond to a matching change in leader region—e.g., while both Kolinda and her husband Jakov Kitarović were born in Rijeka, their successors as president and first spouse of Croatia (Zoran Milanovic and Sanja Musić) both originated from Zagreb.

Table A.2 shows summary statistics for the whole PLAD sample. On average (in the 1989–2020 period), an ADM2 region is coded as spousal region in about 0.14%—or 0.044 years—across all regions. There is some variation across continents: an ADM2 region is coded as spousal region in about 0.25% (0.08%) of all years in Africa (the Americas). In our estimation sample, a number of the 150 ADM2 regions that change their status as spousal birth region do so more than once, resulting in 277 switches in total. This is the variation that we will use in most of the empirical analyses below. The number of switches is lowest in Europe, with only 22 changes, and highest in the Americas with 90 changes.

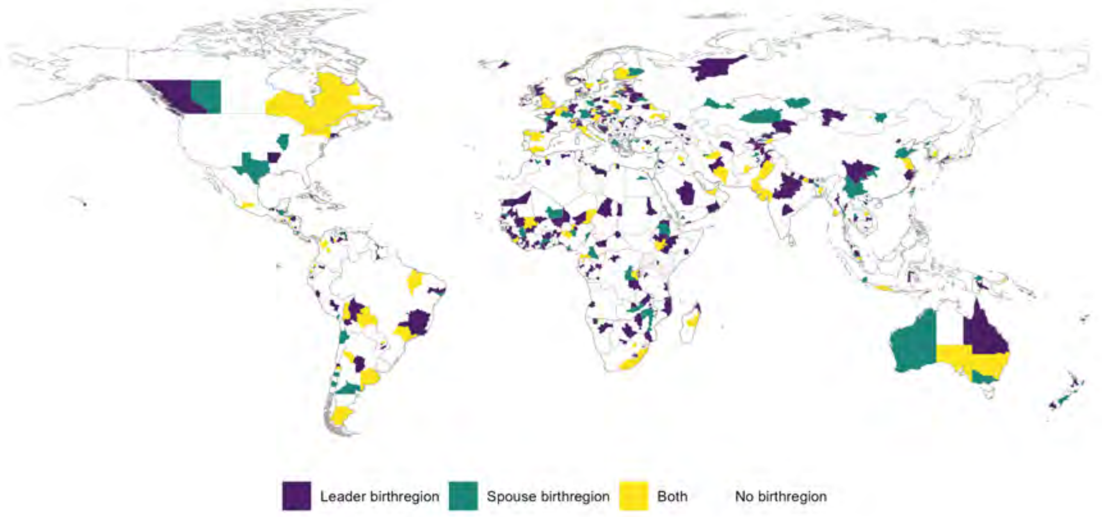
2.2 Geocoded Official Development Assistance Dataset (GODAD)

In order to analyze the subnational allocation of aid for a large set of donors, we build the *Geocoded Official Development Assistance Dataset (GODAD)*. It contains locational information on aid projects of 19 bilateral donors from the OECD's Creditor

[17] ADM1 regions are provinces, states, or governorates. ADM2 regions are smaller and typically correspond to counties or districts. The Database of Global Administrative Areas (GADM) provides shapefiles with information on subnational administrative regions and their boundaries (Hijmans et al. 2020).

[18] Figure C.6 in Appendix C replicates this map at the ADM2 level for three world regions. The figure shows that birthplaces are to some extent geographically concentrated. In Latin America, a comparably small number of regions are coded as spousal birth regions, as many spouses are born in the same regions as leaders. Africa shows a similar concentration of birth regions, to some extent because leaders (and thus their spouses) tend to stay in power for a prolonged period of time.

Figure 1 – Birth regions of political leaders and their spouses, ADM1, 1990–2020



Note: The world map indicates whether an ADM1 region has been a leader birth region (in purple), spouse birth region (in green), both (in yellow), or none (in white) over the 1990–2020 period.

Reporting System (CRS): the 18 European member countries of the OECD's Development Assistance Committee (DAC) and the United States.[19] We also include data on Chinese aid commitments available from AidData's Global Chinese Development Finance Dataset version 2.0 (Custer et al. 2021, Dreher et al. 2022) for the 2000–2017 period and World Bank projects for the years 1995–2020 from AidData and IATI.[20] The resulting database enables researchers to tackle a large set of research questions, against the backdrop of a literature that has previously been limited to investigate the allocation and effectiveness of subnational aid either only for individual recipient countries or the two donors (China and the World Bank) for which geocoded data had previously been available.[21]

To geocode projects for European donors and the United States, we utilize textual information associated with projects to identify geographic entities within the recipient country. The OECD's CRS data provide project-level information on OECD donors beginning in 1973. The GODAD uses all available years. However, given the time

[19]The European donor countries covered are Austria, Belgium, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom.

[20]We take data until 2014 from AidData (2017) and merge project-level data from the International Aid Transparency Initiative (IATI), available for the years 1997-2023.

[21]In a future version of GODAD, we aim to include these previous datasets. Specifically, we will include data from individual recipient countries' Aid Information Management System (AIMS), which are available for a small number of countries: Colombia, Burundi, DRC, Honduras, Iraq, Uganda, Nepal, Nigeria, Senegal, Sierra Leone, Somalia and Timor-Leste. For France, we plan to make use of geocoded data provided by the Agence Française de Developpement (AFD). Finally, we aim to include updated data for aid from India, which Asmus et al. (2023) provide for the years 2007–2014.

coverage of PLAD and other geocoded data, our empirical analysis uses aid data for the 1990–2020 period only. Project numbers and descriptive statistics that we report below all refer to this shorter sample period.

The raw data contain financial information on commitments and disbursements (in US$) as well as information on project characteristics, such as implementing agencies, scope, and project descriptions. We exploit text data on project titles and descriptions to identify and geocode projects at the ADM2 level, allowing us to study the subnational allocation of aid.

Project titles and descriptions provide text data from which geographical entities can be identified. We apply the following procedure for the extraction and identification of geographical entities:[22] First, we collect raw data from the OECD CRS for our 19 donors from 1973 to 2020. For our 1990–2020 sample period, the CRS reports a total of 1,605,748 unique, bilateral projects. We drop projects in sectors that are not allocated to specific regions, either because aid is spent in the donor country or it is not addressed to subnational regions (e.g., aid spent on refugees in donor countries, budget support, donors' administrative costs, or debt relief), reducing the number to 1,541,106 projects. We then exploit the text descriptions of projects to extract candidate geographical entities that can be matched to known cities, regions, or administrative entities within the recipient country. CRS data provide titles, short descriptions, as well as long descriptions of aid projects, which we all use as sources of information. For each project we run a Named Entity Recognition (NER) algorithm through a (pre-trained) RoBERTa base transformer model for entity identification. This particular class of algorithms uses deep learning models to identify specific categories within a text, including geographic entities. The NER model finds at least one geopolitical entity for 462,543 projects (roughly 30% of all reported CRS projects).[23] The extracted entities then undergo a series of data cleaning and cross-checking, resulting in 220,369 unique geocoded projects.[24] From these projects, we identify 285,558 project-location pairs, as some projects may be destined for more than one location.

We investigate whether or not these projects are a representative sample of all aid projects (see Appendix B.4 for details). Unsurprisingly, we find that more recently reported projects and projects in more location-dependent sectors, such as infrastructure, are more likely to be geocoded and are thus over-represented in GODAD. We also find differences with respect to donor countries and the type of financial flows. Our results reported below thus cannot necessarily be generalized to all aid from all donors but have to be interpreted with respect to the specific set of geocoded projects available to us.

[22] See Appendix B for details.

[23] Note that this does not mean that the model misses 70% of the geopolitical entities in the data. The bulk of projects do not contain entities that can be geocoded. Appendix B discusses this in detail.

[24] Appendix B provides details on the evaluation of the NER model accuracy, data cleaning, and how we deal with false positives and negatives.

We test the accuracy of our final data sample by making use of a random sample of 10,000 projects, for which we manually code geographical information. Overall, we find a low incidence of errors, with only 5% of projects in this random sample containing missing or erroneous geocoded information. Errors are classified as false negatives (missed information) or false positives (erroneous geocoding). Missed information is limited, as the entity extraction model responsible for identifying candidate geographical locations has an accuracy of almost 90%. Among wrongfully geocoded entities, most of these cases occur because of semantic reasons, when the model is 'tricked' by descriptions in the text, with references to projects or organizations in a grammatically similar manner to location descriptions.[25] In other cases, the location may be referring to a wider geographical area (Sahara desert), which provides not true subnational information on the aid project. Appendix B describes in detail all the steps taken to obtain the final sample, including the accuracy metrics for the NER model and the errors found in the final sample.

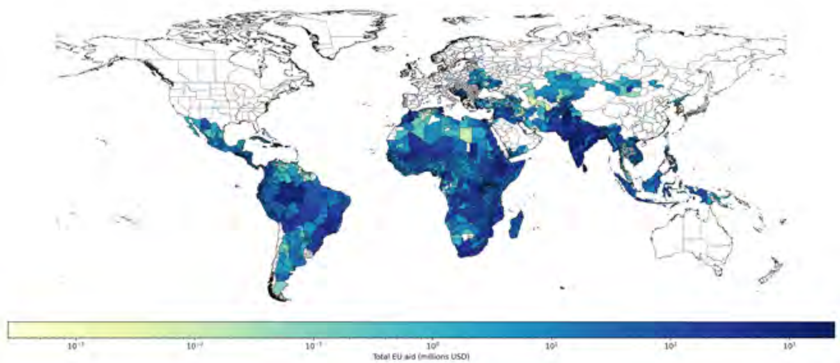
Table B.2 in Appendix B summarizes the number of projects available over the 1990–2020 period by showing the main descriptive statistics both before and after the geocoding. As can be seem, both number and size of aid projects considerably shrink after the geocoding. To a large extent, this reduction is explained by a lack of accuracy in the CRS, which does not include the necessary information we need for geocoding. Moreover, even among the sample of potentially geocodable project, the vast majority is not earmarked for specific, subnational purposes, preventing geocoding. In terms of project size, very large loans and grants are typically countrywide aid flows or finance large infrastructure projects, which do not generally indicate their specific region. Hence, these cases drive much of the changes in the distribution of project sizes before and after geocoding. As shown in Figure B.3, however, the distribution of commitments by sectors remains overall stable before and after the geocoding.

A skeptical reader might argue that the share of projects we managed to geocode is rather low. However, when we compare our data to one of the most widely used datasets on an individual country (Uganda), the number of reported projects in GODAD compares favorably to those reported in the AIMS Geocoded Research Release (AidData 2016). Specifically, AIMS covers 224 projects by those donors we cover here while GODAD includes 5,970 projects. Notably, AIMS reports no World Bank projects in Uganda, whereas we include 1,122 projects from the Bank. In sum, our dataset offers a considerably more comprehensive view compared to previous sources utilized in the literature on aid allocation and effectiveness, even in the few cases where such data for recipient countries are available at all (see Section B.5 for details).

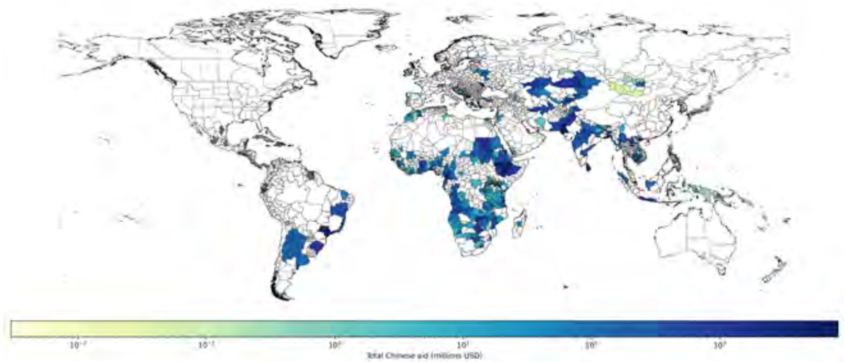
[25] The NGO "Moving to Freedom" will appear to the model as a locality named Freedom.

Figure 2 – Aid commitments in US dollars by ADM1 region, 1990–2020

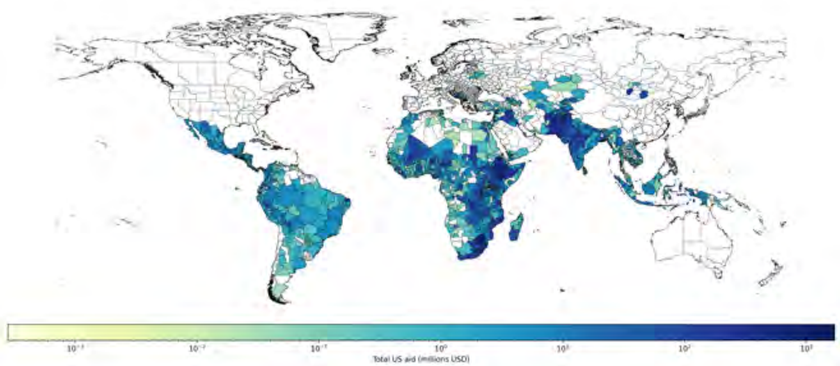
Panel A: European OECD-DAC bilateral donors



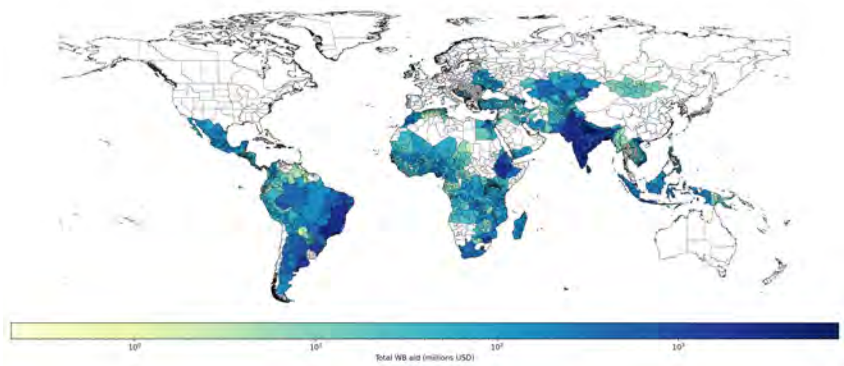
Panel C: China



Panel B: United States

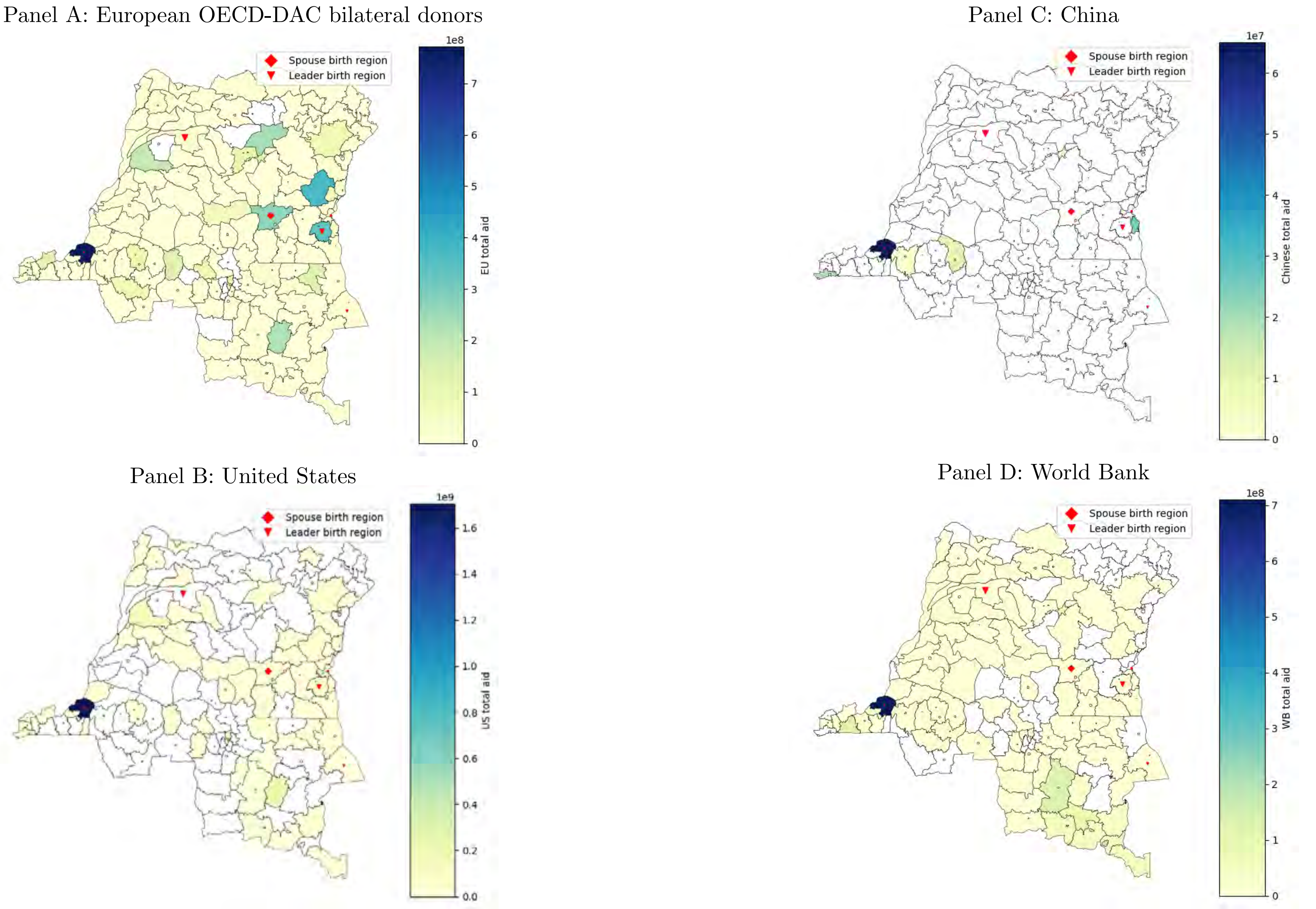


Panel D: World Bank



Note: The world map displays for each ADM1 region the (log) amount of aid in constant 2014 US dollars it has received over the 1990–2020 period on a color scale, with bluer colors indicating larger aid recipients. The panel header indicates the respective donor (group).

Figure 3 – Aid commitments in US dollars by ADM1 region, 1990–2020



Note: The map displays for each ADM1 region of Democratic Republic of Congo (DRC) the amount of aid in constant 2014 US dollars it has received over the 1990–2020 period on a color scale, with bluer colors indicating larger aid recipients. The red markers indicate the birth regions of the DRC's leaders and spouses, where the increasing size accounts for the various term lengths. The panel header indicates the respective donor (group).

As a final step, we merge our new data to the existing datasets described above. As can be seen from Table B.2, this results in a total number of 220,369 geocoded CRS projects (of which 181,775 are for the European bilateral donors combined and 38,594 are US projects), 5,645 projects for the World Bank, and 3,071 projects for China. The next step is to aggregate these projects to the level of subnational regions, resulting in a panel with 97,569 (1,446,339) region-year observations at the ADM1 (ADM2) level, respectively. Figure 2 provides a first glimpse at the data. The world maps show aid amounts in millions of US$ at the level of ADM1 regions over the 1990–2020 period, with darker colors representing larger amounts.

To illustrate the co-location of aid and birth places, we display four administrative maps of the Democratic Republic of Congo in Figure 3, where each panel represents our four donor (groups) as indicated in the column header. The darker colored a region is the more aid it has received by the respective donor. The red markers indicate the birth regions of the DRC's leaders and spouses, where the increasing size accounts for the various term lengths. The map shows the birth region and tenure of different notable leaders and their spouses in the Democratic Republic of Congo. On the far right, the red triangle corresponds to the birth region of Laurent-Désiré Kabila, leader of the country from 1997 until 2001. Shortly after, his son Joseph Kabila assumed power, staying in office until 2019. In 2006 Joseph Kabila married Olive Lembe di Sita, born in Kailo, a region marked on the center right of the map with a red diamond. As can been seen, the pattern of European aid is most in line with birth-region favoritism with respect to both leader and spouse, consistent with what one would expect given the example of the 13-year stay in the role of Olive Lembe di Sita. Of course, the case study presented here is purely descriptive and implies no causality. We next turn to the econometric analysis to analyze systematic patterns.

3 Empirical Strategy

We analyze potential effects of the informal influence leaders' spouses might hold exploiting within-region variation of aid over time, where we compare aid flows in a subnational region at times the region should benefit from informal influence compared to what this same region receives at other times. We run our baseline regressions at the more detailed level of ADM2 regions, but will show results at the ADM1 level for comparison. We estimate the following equation with ordinary least squares:

$$Aid_{c,i,t} = \beta_1 Spouseregion_{c,i,t-1} + \beta_2 Leaderregion_{c,i,t-1} + \beta_3 Pop_{c,i,t} + \gamma_i + \tau_{ct} + \varepsilon_{c,i,t} \quad (1)$$

Our main dependent variable is $Aid_{c,i,t}$, which captures the logarithm of aid in constant

2014 US$ given to recipient region i of country c in year t.[26] The main analysis focuses on aid aggregated over four (groups of) donors: gross aid disbursements from European bilateral OECD-DAC donors and the United States (both available for the years 1973–2020) and the World Bank (1995–2020), as well as Chinese aid commitments (available for the 2000–2017 period). Given that geocoded World Bank aid data are not available prior to 1995, our analysis of total aid covers the 1995–2020 period. Chinese aid was negligible in the earlier years of the sample, so we set it to zero there. We do the same for the years after 2018, but will include more recent data on Chinese aid as soon as they become available.[27]

We measure informal influence at the subnational level with a binary variable $Spouseregion_{c,i,t-1}$—$Spouseregion$ for short in all tables below—that is equal to 1 if the spouse of the leader of country c in year $t-1$ is born in region i.[28] Mirroring our definition of spouse birth regions, we code a binary variable $Leaderregion_{c,i,t-1}$ that takes a value of 1 if the effective political leader of country c in year $t-1$ is born in region i. Given that leaders and their spouses are often born in the same region, this variable allows us to control for leaders' formal influence. By doing so, we can rule out that the informal influence that we attribute to spouses actually measures the formal influence of the politician in power. Note that we lag these variables by one year to allow for some time for disbursements to adjust and to exclude the year in which a region's birth-region status changes.

Our regressions control for the logarithm of a region i's population size in year t (CIESIN 2018, $Pop_{c,i,t}$).[29] Most of our regressions include region fixed effects to control for unobserved variables that affect all regions equally at any point in time. An obvious example are regions that include the capital city. Such regions might differ from others for a number of reasons that are potentially correlated with development aid and the

[26] We add a value of 1 before taking logs. Some values in the DAC's gross disbursements are negative, arguably representing coding mistakes. We set them to 1. For projects with multiple locations that spread across regions, we allocate amounts equally across locations. For example, for a project with two locations in region A and one location in region B, we allocate two thirds of disbursement amounts to A and one third to B. Data for China report outlines of project infrastructure instead of project locations. We therefore split amounts across those regions that receive parts of the project.

[27] We provide separate analyses for (groups of) donors below, excluding the years no data for a donor are available. Note that data on Chinese aid disbursements are not available. However, commitment data exist for other donors, which we use in additional analyses below for a direct comparison. Also note that we focus on Official Development Assistance. World Bank aid thus covers disbursements from the International Development Association only; Chinese aid commitments do not include Other Official Flows.

[28] If there is no information on spousal birth regions, we code the spouse birth region variable as zero. This substantially increases the number of observations, which allows us to use information for a larger number of regions for which we have information for birth regions in a limited number of years only.

[29] We do not include additional covariates. Most of the variables included in country-level aid allocation studies that one might want to include here are arguably correlated with informal influence and would bias our coefficients of interest. We thus follow the tradition of subnational studies to estimate parsimonious regressions (e.g., Dreher et al. 2019). Hoewever, we test robustness to including a larger set of control variables below.

likelihood a spouse is born there. Our fixed effects capture such variables. Finally, we add country-year fixed effects to account for factors that affect aid given to the entire country in a given year. We cluster standard errors at the country level.[30]

While the fixed effects for regions allow us to rule out that better developed regions or those that receive more aid are correlated with unobserved variables that also make it more likely a region is the birth region of the leader's spouse, estimated coefficients do not necessarily represent causal effects. This is because the amount of aid given to a region might change over time due to factors that also affect the likelihood a leader's spouse was born there. In other words, the parallel trends assumption might be violated. We address this by including binary variables that take a value of 1 in the first or second year *before* a region becomes the birth region of a leader's spouse, respectively. We also include a binary indicator for the year a region's birth region status changes (i.e., the first year a region is a birth region, *b1*) and binary variables for the two years *after* a region was the birth region of a leader's spouse:

$$\begin{aligned} Aid_{c,i,t} = &\sum_{j=0}^{2} \beta_{1,j} Spouseregion_{c,i,b+j} + \beta_2 Spouseregion_{c,i,t-1} + \\ &\sum_{j=-3}^{-2} \beta_{3,j} Spouseregion_{c,i,b+j} + ... + \beta_4 Pop_{c,i,t} + \gamma_i + \tau_{ct} + \varepsilon_{c,i,t} \end{aligned} \tag{2}$$

We label these additional variables with the subscripts $b+j$ and $b-j$, respectively, as they are not simple leads and lags of the birth region indicators. $Spouseregion_{b+1}$, for example, indicates that a region will become the birth region of the leader's spouse next year. Likewise, $Spouseregion_{b-1}$ indicates that in the previous year—but not this year—the leader's spouse originated from a region. Significant effects in years directly *after* a region's status as birth region changes would not be surprising. Effects on aid disbursements are unlikely to evaporate immediately after a spouse's partner leaves office. Some informal influence might remain, aid committed earlier might be disbursed with lags, or donors might be slow to adjust their aid budgets. Significant coefficients in years before a leader's spouse originates from a region could however violate the parallel trends assumption. In addition to testing whether birth regions develop differently compared to how these same regions develop at other times, we thus also test whether the effect

[30] De Chaisemartin and d'Haultfoeuille (2020) show that two-way fixed effects estimators with staggered treatment provide a weighted estimate of average treatment effects over periods and units of observation, where weights can turn negative so that they bias the estimates (and coefficients can even switch sign). As we discuss in more detail below, just 32 of the 1056 Average Treatment Effects of the Treated receive negative weights in our baseline regression, and are thus extremely unlikely to affect our results. This is in line with results in Widmer and Zurlinden (2022), in a similar setting. Like they do, we thus mainly report results from traditional two-way fixed effects estimators. We however show below that our results are not affected by this choice. Also note that country-year fixed effects—while being included in most regressions—are not required for our exclusion restriction to hold, and all results below are similar if we exclude them.

of birth regions is stronger compared to the effects for the years just before the spouse's partner assumes office. In other words, even if there is a trend suggested by significant coefficients on $Spouseregion_{b+2}$ and $Spouseregion_{b+1}$, a significant *difference* in the size of the coefficients would still indicate an effect of informal influence. We thus explicitly test whether the effect of birth regions is larger (rather than different), reporting results of one-sided t-tests in the regression tables below. Recall that we lag the *Spouseregion* indicator by one year. The event-time specification thus estimates two separate coefficients for the years before a region is the birth region of a leader's spouse (b+2, b+1) and a separate coefficient for the first year (b1), which is the year a region's status changes (i.e., it typically is a *Spouseregion* in just parts of this year). The year in which a region's status as *Spouseregion* switches back to "normal status" (b-1) is covered by the lagged *Spouseregion* indicator, and we then estimate two additional coefficients for the years that follow (b-2, b-3). While we omit the respective variables from Equation (2) to ease exposition, the corresponding variables for *Leaderregion* are included in all regressions.

4 Results

4.1 Main Results

Column 1 of Table 1 shows the results for total aid corresponding to Equation (1), but excluding fixed effects for country-years;[31] column 2 shows the corresponding regression according to the event-time specification (Equation (2)). As can be seen, spousal regions see their total aid increase by 280–320 percent, compared to the amounts these same regions receive at other times. This effect is significant from zero at the one-percent level. More importantly, the coefficient reported in column 2 is also significantly larger than those corresponding to the year just before a region turns into being the birth region of a leader's spouse (see p-value reported at the bottom of Table 1). What is more, the coefficients indicating that a region will be a birth region in two years, one year, and the first year into being a birth region are small and estimated imprecisely. The results further show that aid disbursements decrease to their average levels one and two years after the spouse of a leader originated from a region, but the spouse of the current leader does not. The coefficients indicating that leaders themselves are born in a region are much smaller in comparison (indicating an increase in aid between 20–22 percent) and estimated imprecisely; its coefficient in the event-time specification is not significantly larger compared to the year before the leader originates from a region.

Columns 3 and 4 report results for our preferred specification (i.e., including fixed effects for country-years). Our results for *Spouseregion* hold, though the estimated

[31] Recall that we control for (log) population as well as fixed effects for ADM2 regions. We do not show their coefficients to reduce clutter.

increase in aid amounts to 130 percent and is thus substantially smaller. The most notable difference compared to the results estimated without the fixed effects for country-years regards the effect of leaders' birth regions. While the estimated effects remain smaller compared to those of spousal regions (between 29–52 percent), they are now estimated more precisely and indicate an increase in aid compared to the year before a region is the birth region of the country's leader. Overall these results are in line with previous work regarding the importance of leader birth regions for a region's ability to attract foreign aid. They also indicate the importance of less obvious influences connected to the leaders' spouses and are in line with the hypothesis that donors are more careful to avoid the placement of aid in regions where domestic political motives are more obvious (i.e., leader regions) compared to regions such influences are more hidden (i.e., birth regions of the leaders' spouses).

Columns 5–8 of Table 1 present variations in how we estimate our regressions, continuing to focus on total aid: Column 5 includes separate binary variables for the first six years into being a *Spouseregion* in addition to a binary variable that captures all additional years, in case the spouses' "tenure" is longer than that. As can be seen, aid significantly increases in the fourth year. Aid then remains high until the (new) leaders' spouse is no longer born in that region. Columns 6–8 focus on the event-time specification and changes the sample that we use to estimate this regression. First, we include only those ADM2 regions that register as *Spouseregion* in our sample period at least once (column 6). We do this because readers might be concerned with us including a large number of regions that never change their status in our control group. Column 7 is less conservative and includes all ADM2 regions that belong to the same ADM1 region, provided that at least one ADM2 region there registered as *Spouseregion* in the sample period. Column 8 focuses on this same sample, but in addition excludes ADM2 regions that did not receive any aid throughout the sample period. These modifications result in dramatic changes in the number of observations (now ranging between 4,602–88,192) but does not change our conclusion in qualitative and quantitative terms.[32]

Finally, in columns 9 and 10, we replicate our key specifications focusing on ADM1 regions, which are substantially larger than ADM2 regions (ADM1 regions cover 42,372 km^2, on average, compared to 2,793 km^2 for ADM2 regions). A priori, the expected effect on the size of the birth-region effect is unclear. On the one hand, the larger size of the regions might make it easier for domestic politicians to hide the political intentions of directing aid to specific areas, to the extent that donors and domestic audiences might be less likely to detect such influence. On the other hand, if aid is targeted to the smaller ADM2 regions instead of the larger ADM1 regions, effect sizes should be smaller and less

[32]Note that our results for the estimated effects of leaders' birth regions turn insignificant. Given that we select these samples based on the status of spousal regions rather than leader regions (and the number of included leader regions is thus comparably small) this is unsurprising.

precisely estimated when we focus on larger areas. According to the results, this latter effect dominates, with estimated increases in aid to *Spouseregions* ranging between 54–58 percent, while effects for leaders' birth regions are small and estimated imprecisely at the ADM1 level.[33]

While we opted for parsimonious specifications in our baseline regressions, we also replicated those regressions including a set of additional subnational determinants of aid, closely following the specification of Dreher et al. (2019). We control for development by including a region's road density and (log) nightlight.[34] We control for geographical size of a region in addition to just population, given as the area of the recipient region.[35] We also control for a series of geo-economic determinants, given that bilateral aid projects might be tied to the presence of key economic infrastructure. From the World Port Index, we construct a binary variable that indicates the presence of ports in a region; we also include binary indicators for the availability of oil or gas in a region (from Lujala et al. 2007) as well as for the presence of mines (based on the Mineral Resources Data System from the U.S Geological Survey). Finally, we include a dummy that captures if the region is also host to the country's capital. To minimize the risk of endogeneity, we include these variables with their values in the first year of our sample (or the first year these data are available) and consequently omit fixed effects for ADM2 regions. As Table C.1 in Appendix C shows, our results remain unchanged, the exception being that regions already register increases in aid starting in the first year a region becomes the birth place of a leader's spouse. While exact results for the additional variables to some extent vary across regressions, aid tends to increase with development, in line with subnational evidence reported elsewhere (e.g., Dreher et al. 2022). Capital cities, and regions with mines and ports receive more aid, while the size of a region's geographic area is estimated imprecisely in most regressions.[36]

In summary, birth regions of leaders' spouses register substantial increases in aid.

[33]As a potential threat to identification, *Spouseregions* might receive more scrutiny at times of political importance. This might result in a higher number of geocoded projects even if the number of actual projects remains unchanged. Arguably, smaller projects are more likely to be found as a consequence of additional scrutiny, while large ones will likely be included either way. This should result in a lower project size in geocoded *Spouseregions*. We find no evidence in line with that. We also do not find a larger share of geocoded projects in all projects in election years, which arguably is a time of greater political scrutiny at large. See Section B.4 for details.

[34]We use data on roads from CIESIN and ITOS (2013) and we measure road density on the ADM1 and ADM2 level as the total length of road in the region divided by the area of the region. We use nightlight data from Li et al. (2020), who combine Defense Meteorological Satellite Program (DMSP) satellite data, available from 1992 to 2013, and Visible Infrared Imaging Radiometer Suite (VIIRS) satellite data, available from 2012 to 2020. By leveraging the common years of 2012 and 2013, the authors provide a consistent time series by simulating DMSP-like nightlight observations using the VIIRS data for the years 2014 to 2020.

[35]This is computed as the log of square kilometers as given by the shapefile boundaries, which coincide with the boundaries provided by AidData and are in turn based on the GADM 3.6.

[36]Also note that the way we cluster standard errors does not affect results (see Table C.2 in Appendix C).

Table 1 – Birth Regions and Total Aid, 1995-2020

	(1) ADM2	(2) ADM2	(3) ADM2	(4) ADM2	(5) ADM2	(6) ADM2	(7) ADM2	(8) ADM2	(9) ADM1	(10) ADM1
Spouseregion (b+2)		0.52		-0.21	-0.20	0.46	-0.22	-0.26		0.21
		(1.37)		(0.66)	(0.63)	(1.05)	(0.68)	(0.59)		(0.71)
Spouseregion (b+1)		0.40		-0.33	-0.33	0.07	-0.65*	-0.98*		0.01
		(0.96)		(0.89)	(0.90)	(0.16)	(1.72)	(1.78)		(0.03)
Spouseregion (b1)		0.22		0.36	0.37	-0.22	0.42	0.38		0.23
		(0.42)		(0.83)	(0.84)	(0.44)	(0.88)	(0.68)		(0.75)
Spouseregion	1.34***	1.44***	0.84***	0.83**		0.78**	0.70**	0.71**	0.43**	0.46*
	(3.93)	(3.70)	(3.02)	(2.60)		(2.26)	(2.21)	(2.04)	(2.18)	(1.90)
Spouseregion (b2)					-0.03					
					(0.06)					
Spouseregion (b3)					0.47					
					(1.06)					
Spouseregion (b4)					1.18**					
					(2.52)					
Spouseregion (b5)					1.49**					
					(2.55)					
Spouseregion (b6)					1.38**					
					(2.28)					
Spouseregion (other year)					1.09**					
					(2.15)					
Spouseregion (b-1)					0.57					
					(1.51)					
Spouseregion (b-2)		0.57		0.02	0.03	0.01	0.26	0.14		-0.11
		(1.26)		(0.07)	(0.07)	(0.02)	(0.62)	(0.23)		(0.37)
Spouseregion (b-3)		0.05		0.06	0.07	-0.47	0.08	-0.10		0.00
		(0.14)		(0.15)	(0.20)	(1.16)	(0.20)	(0.19)		(0.01)
Leaderregion	0.20	0.18	0.25*	0.42**	0.43**	0.07	0.50	0.85	0.16	0.05
	(1.31)	(0.73)	(1.66)	(2.10)	(2.14)	(0.15)	(1.37)	(1.56)	(1.18)	(0.26)
Country-year FE	no	no	yes	yes	yes	yes	yes	yes	yes	yes
Number of countries	141	141	141	141	141	113	108	81	163	163
Number of regions	44506	44506	44506	44506	44506	241	5073	1410	3033	3033
Number of observations	854418	842702	854418	842702	849060	4602	88192	34313	67304	66676
Prob > F Spouse		0.01		0.00		0.06	0.00	0.00		0.04
Prob > F Leader		0.14		0.08		0.44	0.06	0.05		0.34
R squared (within)	0.0046	0.0047	0.0003	0.0003	0.0003	0.0095	0.0015	0.0016	0.0003	0.0004

Note: The dependent variable is *Aid*, i.e., the logarithm of aid (plus 1) given to region i of country c in year t: ODA disbursements of 18 European donors, IDA, and the U.S. and ODA commitments from China (which we set to zero for the years 1995-1999 and 2018-2020). *Spouseregion* and *Leaderregion* are lagged by one year. *Spouseregion (bx)* indicates the xth year into being a birth region. *Prob > F Spouse/Leader* tests whether the coefficients of *Spouseregion/Leaderregion* are larger than those of $Spouseregion_{b+1}/Leaderregion_{b+1}$. All regressions except columns 1, 3, and 9 control for $Leaderregion_{b+2}$, $Leaderregion_{b+1}$, $Leaderregion_{b1}$, $Leaderregion_{b-2}$, and $Leaderregion_{b-3}$. All regressions include the logarithm of a region's population size. Columns 1–8 (9–10) include ADM2 (ADM1) fixed effects. Columns 3–10 include country-year fixed effects (FE). Standard errors are clustered at the country level.
t-statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01.

This effect is largest for aid given as economic infrastructure, as we show in Table C.3 of Appendix C: Birth regions of spouses receive an increase of between 146–167 percent there. We estimate increases in social infrastructure aid of between 106–108 percent; the estimated effect in the production sector is an increase of comparably modest 57 percent (also note that while the increase is statistically significant from zero there, there is no significant increase in production aid compared to the year just before the leader's spouse is born in a region).[37] Social and economic infrastructure includes support for hospitals, schools, presidential palaces, and physical transport projects that might be more visible to potential voters than less tangible projects in the production sector.

4.2 Heterogeneous Effects

In Table 2, we further investigate potential heterogeneity of the average effect of aid. Column 1 tests electoral motives. To the extent that leaders' electoral support is particularly strong in their own birth regions as well as those of their spouses, leaders might want to increase turnout and thus steer aid to those regions particularly at election time (Dreher et al. 2019, Jablonski 2023). However, donors might be particularly weary at election time, making sure that aid is not used to promote the leaders' electoral goals. To test these hypotheses we interact the indicators for *Spouseregion* and *Leaderregion* with a variable that indicates whether or not an executive election will be held in one or two years. Column 2 instead uses electoral competitiveness as part of the interactions, while column 3 focuses on whether or not the leader is eligible for reelection.[38] We find evidence that the *Spouseregion* effect turns stronger prior to elections, in contrast to the effect of *Leaderregions*, which becomes weaker (and is estimated imprecisely).[39] This would be in line with recipient leaders attempting to promote reelection by steering more aid to the birth regions of their spouses, which donors are less likely to be aware of, while donors more effectively prevent them from steering aid to their own regions of birth. We find similar effects when electoral competitiveness is high. Though both coefficients are estimated less precisely, the positive coefficient of the interaction with *Spouseregion* is significantly different from the negative coefficient of the interaction with *Leaderregion* (p-value: 0.019). Finally, column 3 shows no differential effects of *Spouseregion* or

[37]We follow the OECD's DAC in coding these sectors: Social Infrastructure & Services includes Education, Health, Population Policies/Programs & Reproductive Health, Water Supply & Sanitation, Government & Civil Society, and Other Social Infrastructure & Services. Economic Infrastructure & Services includes Transport & Storage, Communications, Energy, Banking & Financial Services, and Business & Other Services. The Production Sector includes Agriculture, Forestry, Fishing, Industry, Mining, Construction, Trade Policies & Regulations, and Tourism.

[38]We take these data from Scartascini et al. (2021). Their index measuring the electoral competitiveness of executive elections is based on both electoral rules and electoral outcomes. It ranges between 0-7, with higher values indicating greater competitiveness. Note that the levels of these variables do not vary within countries and are thus captured by the country-year fixed effects.

[39]The coefficients of the two interactions are significantly different from each other (p-value: 0.012).

Leaderregion on aid at times leaders are eligible for reelection compared to times that they are not (p-value of t-test for equal coefficients of the interactions: 0.16).

Column 4 tests interactions between the birth regions of leaders and their spouses. To the extent that both the leaders and their spouse are born in the same region, aid might increase more. We find no evidence in line with this.

In columns 5–7 of Table 2 we focus on variables relating to aspects of the first couple's marital relations and their level of education. To this end, we interact the birth-region indicators with the number of spouses a leader has (in column 5) and the number of children they have (column 6). We expect the effect of a region's *Spouseregion* status to be weaker when it is shared with more regions, either because the number of spouses or the number of children is high. We also test whether leaders and spouses with higher levels of education differ in their use of aid for their own purpose (in column 7). The results however show that none of these interactions is significant at conventional levels, and neither are the differences between them.[40]

The final column 8 reports results from a placebo regression, where we include an indicator for regions where a future or past spouse of the country's leader is born, but the spouse of the current leader is born elsewhere. The coefficient is much smaller than those of *Spouseregion* and not significant at conventional levels.

Table C.4 in Appendix C tests additional heterogeneity. We find that the *Spouseregion*-effect does not differ significantly between world regions, with the exception of Oceania, where the effect is significantly smaller, and effect sizes turn the overall effects there negligible. Spouse regions benefit more in more corrupt countries and where law and order is less likely to be upheld (though the latter coefficient is estimated imprecisely), while a country's overall political risk does not mediate the effect of leader or spousal birth regions on aid. We further find that leader regions benefit less in more democratic countries, and those where media censorship is less prevalent, but more when the national party is more powerful.[41]

In summary, there is evidence that electoral motives and institutional quality matter for the size of the *Spouseregion* effect. To the contrary, results do not hinge on leaders and their spouses being born in the same region, nor the number of spouses, number of children, or the education of leaders and spouses.

[40] We have also tested interactions including the level of education of leaders themselves or the difference in education between leaders and spouses. None of the interactions are significant at conventional levels.

[41] We take these variables from the PRS Group's International Country Risk Guide and the Varieties of Democracy Project (Coppedge et al. 2023). The notes in Table C.4 provide details.

Table 2 – Birth Region Interactions and Total Aid, ADM2, 1995-2020

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Spouseregion	0.76**	-0.94	1.40***	0.84***	1.48***	0.34	1.21	0.85***
	(2.42)	(0.79)	(2.85)	(2.86)	(2.67)	(0.97)	(1.22)	(3.00)
Leaderregion	0.27	1.16*	0.07	0.25*	0.27	0.53	0.24	0.25*
	(1.62)	(1.66)	(0.22)	(1.67)	(1.13)	(1.56)	(0.25)	(1.67)
Spouseregion*election	0.59*							
	(1.92)							
Leaderregion*election	-0.27							
	(1.50)							
Spouseregion*competetiveness		0.28						
		(1.58)						
Leaderregion*competetiveness		-0.15						
		(1.42)						
Spouseregion*reelection			-0.67					
			(1.29)					
Leaderregion*reelection			0.20					
			(0.62)					
Spouseregion*leaderregion				-0.002				
				(0.00)				
Spouseregion*number spouses					-0.45			
					(1.50)			
Leaderregion*number spouses					-0.01			
					(0.08)			
Spouseregion*number children						0.07		
						(1.14)		
Leaderregion*number children						-0.07		
						(0.80)		
Spouseregion*education							-0.18	
							(1.03)	
Leaderregion*education							0.01	
							(0.05)	
Spouseregion past/future spouse								0.19
								(0.44)
Number of countries	136	140	135	141	140	130	88	141
Number of regions	43308	44345	43930	44506	44438	42084	29965	44506
Number of observations	797362	851981	830568	854418	837723	639077	329280	854418
R squared (within)	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0004	0.0003

Note: The dependent variable is *Aid*, i.e., the logarithm of aid (plus 1) given to region i of country c in year t: ODA disbursements of 18 European donors, IDA, and the U.S. and ODA commitments from China (which we set to zero for the years 1995-1999 and 2018-2020). *Spouseregion* and *Leaderregion* are lagged by one year. Column 1 interacts *Spouseregion* and *Leaderregion* with a variable that indicates whether or not an executive election will be held in one or two years. Column 2 instead uses electoral competitiveness as part of the interactions, column 3 whether or not the leader is eligible for reelection. Columns 5 interacts with the number of spouses a leader has, column 6 with the number of children they have, column 7 with an indicator for higher education. Column 8 is a placebo regression with an indicator for regions where a future or past spouse of the country's leader is born. All regressions include ADM2 fixed effects and country-year fixed effects. Standard errors are clustered at the country level. t-statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01.

Table 3 – Birth Regions and Development, ADM2

	(1) light	(2) light	(3) light	(4) light	(5) light	(6) light	(7) mortality	(8) mortality	(9) mortality	(10) mortality	(11) mortality	(12) mortality
Spouseregion (b+2)		0.041		0.125*		0.120*		4.149		3.487		3.333
		(0.74)		(1.92)		(1.83)		(0.59)		(0.47)		(0.44)
Spouseregion (b+1)		-0.025		0.060		0.057		-2.127		-2.507		-3.335
		(0.49)		(0.92)		(0.87)		(0.47)		(0.49)		(0.65)
Spouseregion (b1)		-0.070		0.012		0.006		-1.943		-0.635		-1.199
		(0.93)		(0.14)		(0.07)		(0.25)		(0.07)		(0.13)
Spouseregion	-0.115**	-0.135**	-0.101*	-0.103	0.122*	0.112	-0.349	1.336	-0.274	2.871	-31.080**	-28.256**
	(2.01)	(2.18)	(1.72)	(1.56)	(1.66)	(1.43)	(0.04)	(0.18)	(0.03)	(0.38)	(2.25)	(2.19)
Aid			0.0004	0.0005	0.0005	0.0005			0.047*	0.046*	0.045*	0.044*
			(0.75)	(0.77)	(0.80)	(0.82)			(1.76)	(1.80)	(1.69)	(1.73)
Spouseregion*Aid					-0.008**	-0.007**					0.847***	0.855***
					(2.57)	(2.49)					(3.21)	(3.13)
Spouseregion (b-2)		-0.134***		-0.114*		-0.117*		27.237		45.432		45.638
		(2.88)		(1.87)		(1.94)		(0.78)		(1.00)		(1.00)
Spouseregion (b-3)		-0.130**		-0.134*		-0.139*		-4.695		4.117		4.702
		(2.25)		(1.80)		(1.86)		(0.44)		(0.47)		(0.54)
Leaderregion	0.068*	0.109***	0.058	0.155***	0.056	0.155***	-7.973**	-2.786	-6.404	-3.261	-6.621	-3.367
	(1.96)	(2.87)	(1.36)	(3.17)	(1.29)	(3.15)	(2.11)	(0.50)	(1.42)	(0.55)	(1.46)	(0.57)
First year	1992	1992	1998	1998	1998	1998	1995	1995	1998	1998	1998	1998
Last year	2020	2020	2020	2020	2020	2020	2017	2017	2017	2017	2017	2017
Number of countries	142	142	132	132	132	132	62	62	61	61	61	61
Number of regions	44799	44799	39667	39667	39667	39667	6897	6895	6853	6851	6853	6851
Number of observations	962384	945698	709644	703883	709644	703883	106005	104845	87834	86902	87834	86902
Prob > F Spouse		0.06		0.02		0.26		0.34		0.27		0.04
Prob > F Leader		0.07		0.05		0.05		0.37		0.36		0.38
R squared (within)	0.0026	0.0027	0.0016	0.0016	0.0016	0.0017	0.0000	0.0001	0.0001	0.0002	0.0001	0.0002

Note: The dependent variable in columns 1–6 is $log(lights)$, defined as the log of mean nightlight emissions in region i of country c in year t (+0.01). The dependent variable in columns 1–6 is $(infantmortality)$—the rate of infants dying before reaching one year of age, per 1,000 live births. *Aid* is the logarithm of aid (plus 1): ODA disbursements of 18 European donors, the U.S., IDA, and ODA commitments from China. *Spouseregion* and *Leaderregion* are lagged by one year. *Spouseregion (b1)* indicates the 1st year into being a birth region. *Prob > F Spouse/Leader* tests whether the coefficients of *Spouseregion/Leaderregion* are larger than those of $Spouseregion_{b+1}/Leaderregion_{b+1}$. Even columns control for $Leaderregion_{b+2}$, $Leaderregion_{b+1}$, $Leaderregion_{b1}$, $Leaderregion_{b-2}$, and $Leaderregion_{b-3}$. All regressions include ADM2 fixed effects, country-year fixed effects, and the logarithm of a region's population size. Standard errors are clustered at the country level.
t-statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01.

4.3 Consequences of Informal Influence

We proceed with testing potential consequences of spouses' informal influence. In order to evaluate the overall consequences of increased aid in spousal regions we would need to compare the effects of aid in such regions to the opportunity costs of aid. Aid could be redirected from another country or region within the same country, where it might be more effective than it is when being rerouted for political reasons. Alternatively, the aid might be additional, so that the gains in $Spouseregions$ would have to be compared to potential uses of this money in the donor country. We instead focus on the consequences of aid on spouses' birth regions exclusively. Independent of the motive for giving the aid, donors might enforce that the aid is put to good use, ensuring its effectiveness. It is however also possible that the aid is wasted or even hurts development, for example because elites are strengthened at the expense of larger populations or leads to polarization or even conflict. Table 3 investigates effects of a region's status as a birth place of the leader's spouse on development. Our first measure for development is $Light_{c,i,t}$, which measures the logarithm of average nightlight emissions in an ADM2 region i of country c in year t.[42] It is the sum of emissions of all pixels in a region weighted by the fraction of each cell that falls within a specific polygon. In order to cover as many years as possible in our panel, we rely on Li et al. (2020), who combine satellite data from the Defense Meteorological Satellite Program (DMSP) and DMSP converted Visible Infrared Imaging Radiometer Suite (VIIRS). These data provide a useful proxy for regional GDP estimates in parts of the world where data coverage is spotty and cover almost all countries in our sample starting in 1992. Following Hodler and Raschky (2014), our measure then corresponds to the log of the average nighttime light intensity in region r at time t.[43]

Columns 1 and 2 show that birth regions of spouses develop less, as measured by (log) average nighttime light emissions in the ADM2 region. The coefficient implies that such regions' nightlight emissions decrease by 11–12.5 percent at times the leaders' spouse originates from there. The effect is statistically significant at the five-percent level and—according to the event-time specification reported in column 2—significantly larger in absolute terms compared to nightlights just before the leaders' spouse was born in a region. This effect stays negative after a region's "tenure," indicating persistence. To the contrary, the coefficient of leader birth region is positive and would imply an increase in lights by between 7–11.5 percent. Again, the coefficient is both statistically different from zero and from the level of development just before a region turns into the birth region of a leader. The size of the leader birth-region effect exceeds those reported in Hodler and Raschky (2014) (of around four percent, for a much smaller sample) by roughly 50

[42] We add 0.01 before taking logs in order to not lose zero observations.

[43] As Hodler and Raschky (2014) point out, we thus measure nightlight intensity per area. They prefer this measure over light per capita mainly due to concerns regarding the quality of subnational population data. Note that the inclusion of fixed effects for regions is likely to make this choice inconsequential.

percent.

In columns 3–6 we investigate the potential role of aid in how birth regions affect development. To this end, we sum the amount of aid received over three years and include it with a lag of one year. As can be seen in columns 3 and 4, results for *Spouseregion* and *Leaderregion* remain similar, while the effect of aid is small and measured imprecisely. In columns 5 and 6 we test whether the effect of aid on development depends on whether or not it is given to a *Spouseregion*. If aid given for political reasons is sufficiently less effective than aid given for other reasons, the total effect of aid received by a region at the time the leader's spouse was born there should be lower compared to the effect of aid a region receives at other times (Dreher et al. 2021). The negative and significant coefficient of the interaction between aid and the *Spouseregion* indicator is in line with this expectation.

The remaining columns of Table 3 replicate the analysis focussing on an alternative indicator of development—infant mortality, measured as the rate of infants dying before they reach one year of age, per 1,000 live births.[44] Overall, results are similar. Mortality is lower in *Leaderregions* but not *Spouseregions*; however, neither of these effects is significantly different compared to the years just before a region gains that status.[45] Aid increases mortality, in line with the subnational results reported for Chinese aid in Cruzatti et al. (2023).[46] Again, this average effect hides important heterogeneity, as indicated by columns 11 and 12. Regions that do not include the birth place of a leader's spouse see their aid to decrease mortality substantially and significantly. The more aid they receive however, the less effective the aid becomes in reducing mortality, in line with the expectation that at least parts of the additional aid are not spent on projects that effectively promote development. Overall, it seems the additional aid is less effective than what the aid would have achieved elsewhere.

4.4 Donor-specific Effects

Table 4 returns to *Aid* but looks at different (groups of) donors separately rather than aggregate amounts of aid. Columns 1 and 2 focus on aid disbursements from the group of 18 European donors, columns 3 and 4 on the United States, both for the 1990–2020 period. Results reported in columns 5 and 6 refer to IDA disbursements in the years 1995–2020. Columns 7 and 8 show results for Chinese ODA commitments, which are

[44] Data on infant mortality are from Cruzatti et al. (2023), who combine data from 92 Demographic and Health Surveys (DHS) across 53 countries and nearly 55,000 sub-national locations (enumeration areas) from 2002 to 2014. We summarize their data to the level of ADM2 regions.

[45] Our results for leaders' birth regions are comparable in magnitude to those reported in Widmer and Zurlinden (2022) who find that infant mortality in leader birth regions is eight deaths lower per 1,000 live births, in a sample of 36 African countries. Contrary to similar results for health ministers, the effect is however estimated imprecisely.

[46] Cruzatti et al. (2023) find opposing results for regional and country level outcomes and explain that difference with fungibility of aid.

Table 4 – Birth Regions and Aid by Donor Group, ADM2

	(1) Europe	(2) Europe	(3) USA	(4) USA	(5) WB	(6) WB	(7) China	(8) China
Spouseregion (b+2)		0.161		-0.427		0.126		0.343
		(0.50)		(1.61)		(0.50)		(1.45)
Spouseregion (b+1)		-0.152		0.044		0.254		0.316
		(0.46)		(0.16)		(0.99)		(1.53)
Spouseregion (b1)		0.668		0.068		-0.285		
		(1.62)		(0.21)		(0.99)		
Spouseregion	1.080***	1.165***	0.933***	0.961***	0.237	0.213	0.182*	0.232*
	(3.48)	(3.27)	(2.85)	(2.65)	(1.60)	(1.39)	(1.76)	(1.86)
Spouseregion (b-1)								-0.015
								(0.12)
Spouseregion (b-2)		0.682*		0.201		-0.354		-0.162
		(1.74)		(0.69)		(1.19)		(0.95)
Spouseregion (b-3)		0.680**		0.385		0.013		
		(2.07)		(1.28)		(0.04)		
Leaderregion	0.136	0.256	-0.095	-0.224	0.108	0.256	0.174	0.118
	(1.03)	(1.55)	(0.58)	(1.52)	(1.12)	(1.45)	(1.60)	(1.09)
First year	1990	1990	1990	1990	1995	1995	2000	2000
Last year	2020	2020	2020	2020	2020	2020	2017	2017
Number of countries	142	142	142	142	141	141	137	137
Number of regions	44799	44799	44799	44799	44506	44506	40921	40921
Number of observations	1028584	1009949	1028584	1009949	854418	842702	568950	568950
Prob > F Spouse		0.00		0.01		0.44		0.32
Prob > F Leader		0.02		0.21		0.17		0.00
R squared (within)	0.0009	0.0009	0.0009	0.0010	0.0000	0.0000	0.0003	0.0006

Note: The dependent variable is *Aid*, i.e., the logarithm of aid (plus 1) given to region i of country c in year t: ODA disbursements of 18 European donors (columns 1–2), the U.S. (columns 3–4), IDA (columns 5–6), and ODA commitments from China (columns 7–8). Except for columns 7–8, *Spouseregion* and *Leaderregion* are lagged by one year. *Spouseregion (b1)* indicates the 1st year into being a birth region. *Prob > F Spouse/Leader* tests whether the coefficients of *Spouseregion*/*Leaderregion* are larger than those of $Spouseregion_{b+1}/Leaderregion_{b+1}$. Columns 2, 4, 6, and 8 control for $Leaderregion_{b+2}$, $Leaderregion_{b+1}$, $Leaderregion_{b1}$, $Leaderregion_{b-2}$. Columns 2, 4, and 6 also control for $Leaderregion_{b-3}$. All regressions include the logarithm of a region's population size, ADM2 fixed effects and country-year fixed effects. Standard errors are clustered at the country level.
t-statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01.

available in the 2000–2017 period.[47]

Column 1 shows that aid from the 18 European donors combined increases substantially at times a region is the birth region of a leader's spouse. According to our estimates, European aid to those regions increases by around 200%, significant at the 1% percent level.[48] Note that there are again no pre-trends in aid prior to a region's change in its status as *Spouseregion*; what is more, aid flows more generously compared to the year just before a region achieves its *Spouseregion* status. Turning to the results for political leaders we find that while the coefficient for leader birth regions is also positive, it is much smaller in magnitude and imprecisely estimated. However, the event-time specification of column 2 shows that aid does increase compared to the year before a leader is born in a region. Overall, these weaker effects suggests that European donors prevent the direct use of their aid to the benefit of leaders to some extent, but not entirely, while much larger amounts of aid flow to the birth regions of leaders' spouses.

Columns 3 and 4 show very similar results for US aid. According to the estimates of column 3, US aid to spouse birth regions increases by 221%, significant at the 1% percent level, while birth regions of leaders themselves do not experience increases in aid (with coefficients not only being small, but negative. The event-time specification shows that aid to *Spouseregions* increases significantly over what these same regions received just before their status as birth region changes, but aid to *Leaderregions* does not (see column 4).

In line with previous results for Africa, and a shorter period of time (Dreher et al. 2019), World Bank aid is not different at times the country's leader or their spouse originate from a region, compared to what the same region receives at other times. Results do however show increases in Chinese aid, for both regions that include the birth place of a leader's spouse, and birth regions of leaders themselves. The coefficients for *Spouseregion* and *Leaderregion* in column 7 are similar in magnitude (though the latter is estimated less precisely), implying an increase in aid of around 20 percent. However, column 8 shows no significant increase in Chinese aid over the amount received two years before a region becomes the birth region of the leader's spouse. This is different when compared to *Leaderregions*, where aid increases compared to the year before the leader was born there, in line with results previously reported in Dreher et al. (2019) for Africa.[49]

[47]Recall that we do not lag commitments, expecting that they will change faster than disbursements, on average. This implies that the year of a region's change in status (b1) is covered by the birth region indicators, while the year in which regions loose their status (b-1) is not, and we therefore control for it. What is more, we continue to estimate two coefficients after we define regions as birth region and thus exclude the third lag (b+3). This does not affect our results.

[48]This effect is somewhat larger in magnitude when compared to those of leader birth regions on China's aid in Africa, reported in Dreher et al. (2019). Results are very similar when we do not lag aid by one year or lag by two years instead.

[49]When we focus on more commercially-oriented Other Official Flows, birth regions of leaders again receive larger amounts compared to what they get in the year before they become *Leaderregion* but this is not the case for birth regions of spouses.

Table C.5 in Appendix C shows results for the 18 European donors separately rather than jointly, focusing on the event-time specification. For many donors, results are less precisely estimated, as one would expect given the lower number of non-zero aid observations. The spouse birth region effect is significantly different from zero for Germany, Ireland, Spain, and Switzerland. Aid is significantly larger compared to the year before a leader's spouse originates from a region for seven of 18 donor countries: Germany, Greece, Iceland, Ireland, Portugal, Spain, and Switzerland.

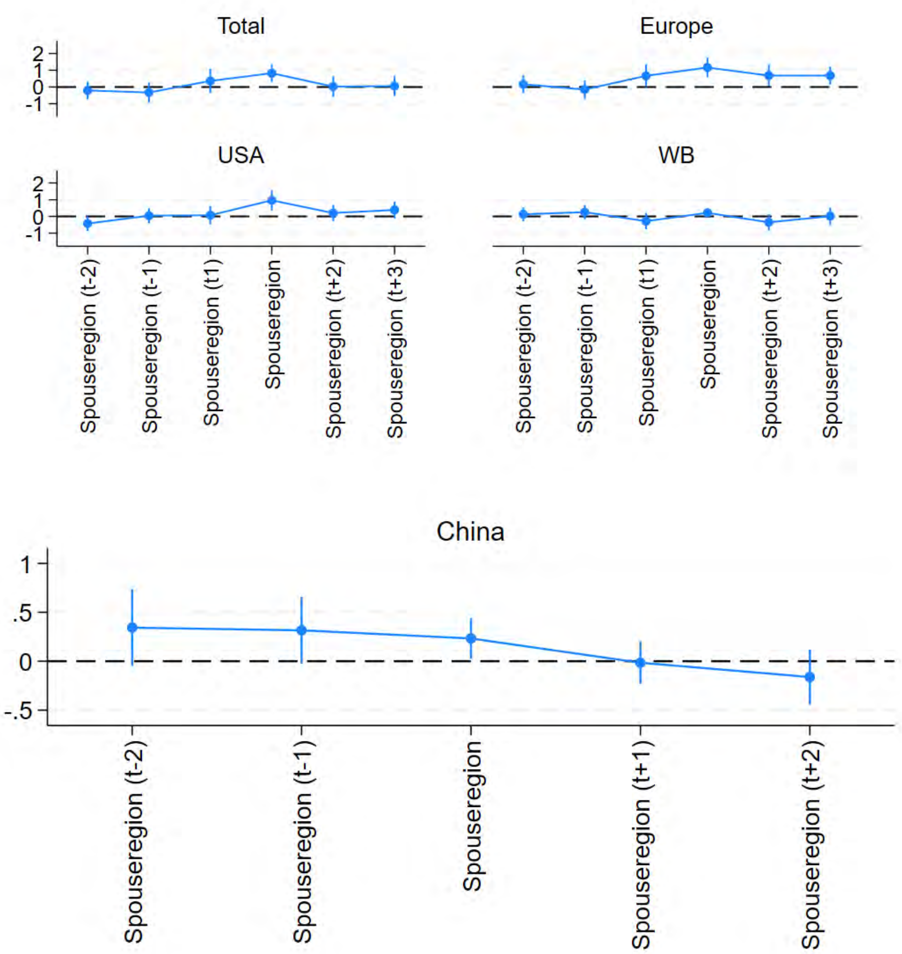
Figure 4 summarizes the effects of the event-time specification for total aid and the main (groups of) donors.[50] The figure illustrates that there are no significant pre-trends in aid by any donor before an ADM2 region becomes the birth place of a leaders' spouse, as well as the significant increases in total aid, driven by aid from Western bilateral donors. While aid disbursements from the group of European donors and the U.S. increase at times the leaders' spouse is born in a region, aid commitments from China do not.[51]

In summary, our results illustrate that both birth regions of spouses and of leaders themselves benefit from increased aid inflows. Results differ however across donors: Chinese aid flows more generously to the birth region of a country's leader. European donors and the United States give substantially more aid to the birth regions of leaders' spouses; aid from European donors also flows more generously to regions at times the leader is born there compared to the year before, but the increase is substantially less pronounced compared to *Spouseregions*. This is in line with the expectation that Western donors to some extent prevent aid from being directed for political reasons where such reasons seem obvious. Where reasons are more hidden, aid increases substantially instead, either because donors do not realize the political importance of spousal regions or they do not mind spending aid in such regions because audiences in the donor countries pay less attention there. This result puts those of previous work in perspective. To the extent that previous work found Chinese aid to flow abundantly to birth regions of political leaders, observers expected aid from Western donors to be more effective. Our results instead show that rather than being put to more productive uses, the aid is redirected for political purposes, where such motives are less obvious. In line with this interpretation, our above results have shown that the effect of *Spouseregions* on aid indeed become stronger prior to elections and that aid given to such regions is less effective compared to aid given at other times.

[50] See Figure C.1 for the analogous figure of the *Leaderregion* effects.

[51] We have also tested whether the use of disbursements vs. commitments drives these differences. While data on Chinese aid disbursements are unavailable, Table C.6 replicates the analysis for commitment amounts for all (groups of) donors as well as using a binary variable indicating that a project has been committed in any year. Again, European donors give more aid to *Spouseregions*, while China gives more generously to *Leaderregions* (note that according to the binary indicator World Bank aid to *Spouseregions* increases compared to the year before, but is below what these regions receive at other times).

Figure 4 – Effect of Spouse Birth Regions on Aid, ADM2



Note: The figure plots the coefficients of *Spouseregion* in concert with its 'leads' and 'lags' and 90% confidence intervals, corresponding to column 4 of Table 1 and columns 2, 4, 6, and 8 of Table 4.

4.5 Dynamic Treatment Effects

A skeptical reader might be concerned that our fixed effects estimations rely on post-treatment information, potentially resulting in a biased estimate of our coefficient of interest. Recall however that our identification strategy does not require the inclusion of fixed effects for country-years and we have shown that their exclusion does not affect our results. Moreover, our main strategy for identifying causal effects consists in comparing aid received while being the birth region of a spouse to the years just before, which should not be biased by the inclusion of regions treated at different points in time. Also note that our setting is unlike those typically discussed in the recent heterogeneous treatment effects literature.[52] Contrary to this literature, we do not estimate two-way fixed effects models with fixed effects for cross-sectional units and time, but rather include fixed effects for country-year in addition to those for ADM2 regions. Our treatment—birth region—in most cases reverses over time, and our prior is that the effect of birth regions on aid quickly fades out once a region loses its status as spousal birth region (an expectation that finds support in our event-time specifications shown above). A number of regions are treated in the first year, so we have no prior information to compare with. What is more, regions might or might not have been treated before they are included in our sample, so "never treated" or "not yet treated" regions do not comprise adequate control groups. Most of the recently developed estimators such as those of Callaway and Sant'Anna (2021) or Wooldridge (2021) thus cannot directly be applied to our setting. Most importantly, when we estimate the weights attached to the two-way fixed effects regressions studied in De Chaisemartin and d'Haultfoeuille (2020), we find that the share of negative weights is very low. Using De Chaisemartin et al.'s (2019) estimator—under the common trends assumption—we find that just 32 of the 1056 Average Treatment Effects of the Treated receive negative weights, and are thus extremely unlikely to affect our results.

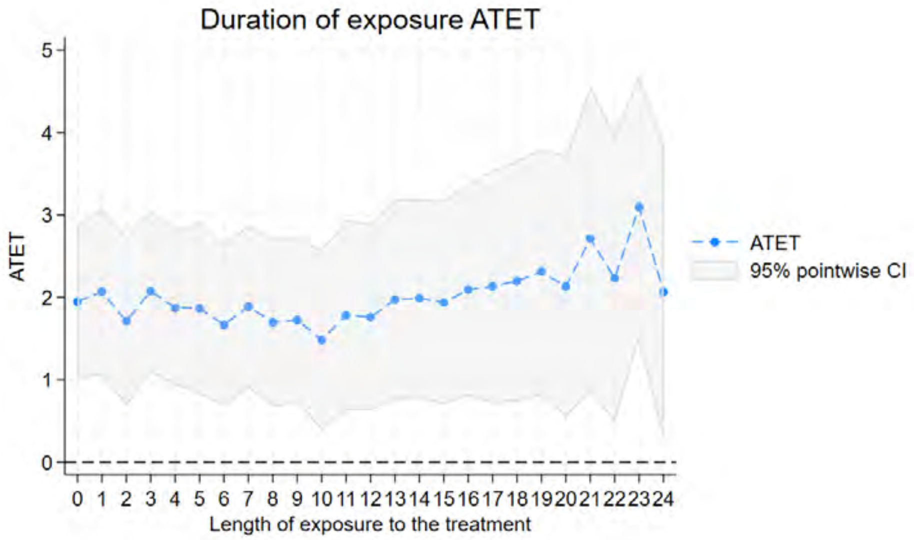
We nevertheless twist our model to better fit the assumptions of the recent two-way fixed effects literature. We drop those regions that are the birth region of a spouse in the first year of our estimation sample and estimate the modified model using Wooldridge's (2021) Extended Two-Way Fixed Effects Estimator, allowing for heterogeneous treatment effects in any year.[53] As before, we include the (lagged) binary leader birth region indicator and logged population as control variables for the outcome model but use fixed effects for years rather than country-years, as is usually assumed by this model. What is more, we do not allow for treatment reversal, implying that all regions change their status once a spouse first originates from the region.

Figure 5 shows the dynamic treatment effects of being a (lagged) spouse region on

[52] While the literature on heterogeneous treatment effects is voluminous and rapidly expanding, recent replication work has also shown that results of conventional two-way fixed effects estimators are rarely affected by accounting for potentially heterogeneous treatment (Chiu et al. 2023).

[53] The estimator includes interactions between treatment-year cohorts and years, allowing for heterogeneous treatment.

Figure 5 – Spouse Birth Regions and Total Aid, Extended Two-Way Fixed Effects



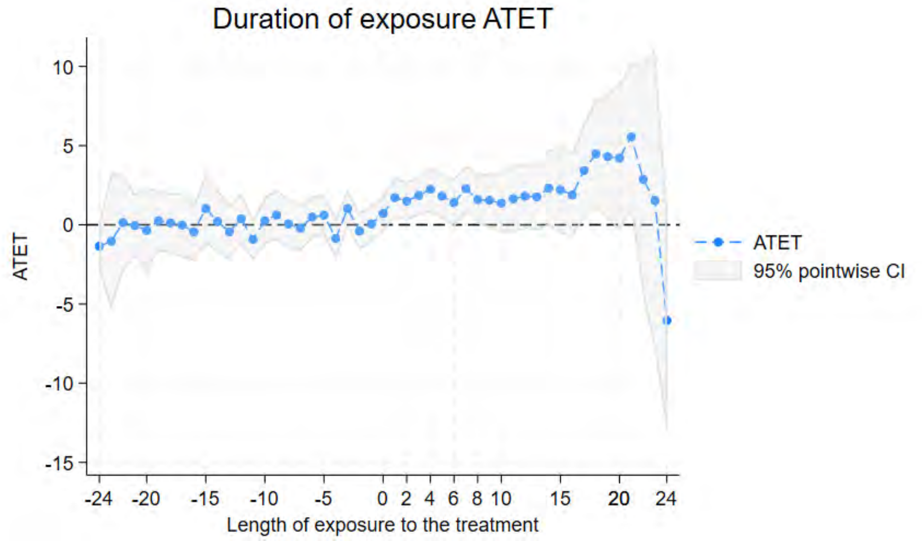
Note: The figure plots the Average Treatment Effect on the Treated (ATET) of (lagged) *Spouseregion* on total aid, using Wooldridge's (2021) Extended Two-Way Fixed Effects Estimator at the ADM2-level in the 1995–2020 period. We drop regions that are the birth region of a spouse in the year 1995 and estimate the modified model allowing for heterogeneous treatment effects in any year. The (lagged) binary leader birth region indicator and the logarithm of population size are included as control variables in concert with fixed effects for years and ADM2 regions. Not yet treated regions represent the control group.

logged total aid disbursements, compared to regions that have not yet been treated. All coefficients are positive and considerably larger than that corresponding to the base estimate. One reason for this is the exclusion of country-year fixed effects, which leads to higher effect sizes (see columns 1 and 2 of Table 1). In addition, it seems that the regions we exclude here to fit the model's requirements bias the coefficients upwards.

Figure 6 shows results using the doubly robust augmented inverse probability-weighted (AIPW) estimator. The AIPW estimator is robust to misspecification of the treatment or outcome models. While the ATET parameters of Figure 5 are shown for each cohort at the time of treatment exposure and the years thereafter, they are not calculated for the years prior to treatment.[54] To the contrary, Figure 6 shows treatment effects for the entire sample range, from 24 years before to 24 years after treatment. As can be seen, there are no significant effects prior to the treatment. Effects turn larger and significant at the one percent level in t=1, which is the second year after a leader's spouse is born in a particular region (recall that we lag the *Spouseregion* indicator by one year). Effects then stay significant at least at the five percent level until t=8, with effect sizes comparable to but somewhat larger than those reported in column 5 of Table 1. Effects are less precisely estimated for longer lags, with the exception of some lags exceeding t=16, which are estimated based on comparably few regions treated early in the sample

[54]This is because parameters are identified based on the parallel-trends assumption (see Wooldridge 2021).

Figure 6 – Spouse Birth Regions and Total Aid, Augmented Inverse Probability-weighted



Note: The figure plots the Average Treatment Effect on the Treated (ATET) of (lagged) *Spouseregion* on total aid, using the doubly robust augmented inverse probability-weighted (AIPW) estimator at the ADM2-level in the 1995–2020 period. We drop regions that are the birth region of a spouse in the year 1995 and estimate the modified model allowing for heterogeneous treatment effects in any year. The (lagged) binary leader birth region indicator and the logarithm of population size are included as control variables in the treatment and outcome models, in concert with fixed effects for years and ADM2 regions. Not yet treated regions represent the control group.

period.

5 Conclusions

This study shed light on the informal influence of political leaders' spouses in shaping regional outcomes. We introduce two novel datasets—the Political Leaders' Affiliation Database (PLAD) and the Geocoded Official Development Assistance Dataset (GODAD). PLAD provides new horizons for scholars to investigate the impact of personal characteristics on political decision-making processes. In parallel, GODAD empowers researchers to delve into a wide spectrum of critical research inquiries associated with the allocation and effects of aid below the country level.

Our analysis of these new datasets demonstrates that leaders' spouses significantly influence the allocation of resources, so that their birth regions experience substantial increases in foreign aid inflows. European donors and the United States appear to direct aid to these regions, possibly as a result of informal political networks and connections. Regions associated with the leaders themselves do not receive similar preferential treatment from Western donors: they do not receive additional European aid and much smaller increases in U.S. aid. We argue that this is because Western donors try to avoid the impression that their aid is used for political purposes. While the political role of leader birth regions seems obvious, the role of their spouses' birthplaces is less so. In line with this interpretation, we find the increase in aid to spousal regions to be particularly pronounced at election time, while aid to leader regions decreases at such times (though the latter coefficient is estimated imprecisely).

The pattern is different for China, which shows increases in aid flows to birth regions of around the same order of magnitude compared to what these regions received before they became birth regions, for leaders and their spouses alike. In line with previous work for Africa (Dreher et al. 2019), we also find that Word Bank aid does not flow more abundantly at times a region is politically influential.

These results put those of previous work in perspective. According to much of the literature, political motives make aid less effective (e.g., Dreher et al. 2018, 2013). Based on this, the literature expects aid from those donors to be more effective, that prevent the use of their aid for reasons of domestic politics (Minoiu and Reddy 2010). The results of this paper instead suggest that aid from donors who are careful to avoid their aid from being used for obvious political motives does not move to more effective purposes, but instead to projects that are similarly political, but where such motives are more difficult to detect. Our analysis of night light emissions and infant mortality indeed shows that birth regions of leaders develop more, but birth regions of their spouses develop less, compared to how these same regions develop at other times. As one channel we test the effects of aid there. We find that aid given to a region is less effective when it was given at times the leader of a spouse born there is in power compared to the effects of aid to that region at other times. This is in line with the interpretation that donors carefully weigh the benefits and drawbacks of giving aid to leader regions, making sure the aid that goes there promotes development. Donors might not exert similar scrutiny when aid goes to spousal regions—either because they do not realize the political importance of these regions or because they are less likely to be held accountable for an abuse of aid there.

References

Adida, C. L., N. Combes, A. Lo, and A. Verink (2016). The Spousal Bump: Do Cross-Ethnic Marriages Increase Political Support in Multiethnic Democracies? *Comparative Political Studies 49*(5), 635–661.

AidData (2016). Uganda AIMS Geocoded Research Release Level 1 v1.4.1. Williamsburg, VA and Washington, DC: AidData.

AidData (2017). AidData's Geocoded Global Chinese Official Finance, Version 1.1.1.

Asatryan, Z., T. Baskaran, C. Birkholz, and D. Gomtsyan (2022). Favoritism and Firms: Micro Evidence and Macro Implications.

Asatryan, Z., T. Baskaran, C. Birkholz, and P. Hufschmidt (2023). Favoritism by the governing elite. Technical report, Ruhr Economic Papers.

Asmus, G., V. Z. Eichenauer, A. Fuchs, and B. Parks (2023). Does India Use Development Finance to Compete With China? A Subnational Analysis. *Journal of Conflict Resolution*, forthcoming.

Barkan, J. D. and M. Chege (1989). Decentralising the State: District Focus and the Politics of Reallocation in Kenya. *Journal of Modern African Studies 27*(3), 431–453.

Bommer, C., A. Dreher, and M. Perez-Alvarez (2022). Home Bias in Humanitarian Aid: The Role of Regional Favoritism in the Allocation of International Disaster Relief. *Journal of Public Economics 208*, 104604.

Branigin, W. (1984). 'Crony Capitalism' Blamed for Economic Crisis. *The Washington Post*, 16 August.

Briggs, R. C. (2014). Aiding and Abetting: Project Aid and Ethnic Politics in Kenya. *World Development 64*, 194–205.

Burgess, R., R. Jedwab, E. Miguel, A. Morjaria, and G. Padró i Miquel (2015). The Value of Democracy: Evidence From Road Building in Kenya. *American Economic Review 105*(6), 1817–51.

Callaway, B. and P. H. Sant'Anna (2021). Difference-In-Differences With Multiple Time Periods. *Journal of Econometrics 225*(2), 200–230.

Chiu, A., X. Lan, Z. Liu, and Y. Xu (2023). What To Do (and Not to Do) with Causal Panel Analysis under Parallel Trends: Lessons from A Large Reanalysis Study. *Unpublished Manuscript*.

CIESIN (2018). Gridded Population of the World, Version 4 (GPWv4): Population Count, Revision 11. Center for International Earth Science Information Network, NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4JW8BX5.

CIESIN and ITOS (2013). Global Roads Open Access Data Set, Version 1 (gROADSv1). NASA Socioeconomic Data and Applications Center (SEDAC).

Coppedge, M., J. Gerring, C. H. Knutsen, S. I. Lindberg, J. Teorell, D. Altman, M. Bernhard, A. Cornell, M. S. Fish, L. Gastaldi, H. Gjerløw, A. Glynn, S. Grahn, A. Hicken, K. Kinzelbach, K. L. Marquardt, K. McMann, V. Mechkova, A. Neundorf, P. Paxton, D. Pemstein, O. Rydén, J. von Römer, B. Seim, R. Sigman, S.-E. Skaaning, J. Staton, A. Sundström, E. Tzelgov, L. Uberti, Y. ting Wang, T. Wig, and D. Ziblatt (2023). V-Dem Codebook v13. Varieties of Democracy (V-Dem) Project.

Cruzatti, J., A. Dreher, and J. Matzat (2023). Chinese Aid and Health at the Country and Local Level. *World Development 167*, 106214.

Custer, S., A. Dreher, T.-B. Elston, A. Fuchs, S. Ghose, J. Lin, A. Malik, B. C. Parks, B. Russell, K. Solomon, et al. (2021). Tracking Chinese Development Finance: An Application of AidData's TUFF 2.0 Methodology. *Williamsburg, VA: AidData at William & Mary*.

Dahl, R. (1961). *Who Governs? Democracy and Power in an American City*. Yale University

Press: New Haven.
De Chaisemartin, C., X. d'Haultfoeuille, and A. Deeb (2019). TWOWAYFEWEIGHTS: Stata Module to Estimate the Weights and Measure of Robustness to Treatment Effect Heterogeneity Attached to Two-way Fixed Effects Regressions. Technical report, HAL.
De Chaisemartin, C. and X. d'Haultfoeuille (2020). Two-way Fixed Effects Estimators With Heterogeneous Treatment Effects. *American Economic Review 110*(9), 2964–2996.
De Luca, G., R. Hodler, P. A. Raschky, and M. Valsecchi (2018). Ethnic Favoritism: An Axiom of Politics? *Journal of Development Economics 132*, 115–129.
Do, Q.-A., K.-T. Nguyen, and A. N. Tran (2017). One Mandarin Benefits the Whole Clan: Hometown Favoritism in an Authoritarian Regime. *American Economic Journal: Applied Economics 9*(4), 1–29.
Dolan, C. B. and K. K. McDade (2020). Pulling the Purse Strings: Are There Sectoral Differences in Political Preferencing of Chinese Aid to Africa? *PloS one 15*(4), e0232126.
Dreher, A., V. Z. Eichenauer, and K. Gehring (2018). Geopolitics, Aid and Growth: The Impact of UN Security Council Membership on the Effectiveness of Aid. *World Bank Economic Review 32*, 268–286.
Dreher, A., A. Fuchs, R. Hodler, B. C. Parks, P. A. Raschky, and M. J. Tierney (2019). African Leaders and the Geography of China's Foreign Assistance. *Journal of Development Economics 140*, 44–71.
Dreher, A., A. Fuchs, R. Hodler, B. C. Parks, P. A. Raschky, and M. J. Tierney (2021). Is Favoritism a Threat to Chinese Aid Effectiveness? A Subnational Analysis of Chinese Development Projects. *World Development 139*, 105291.
Dreher, A., A. Fuchs, B. Parks, A. Strange, and M. J. Tierney (2022). *Banking on Beijing: The Aims and Impacts of China's Overseas Development Program.* Cambridge University Press.
Dreher, A., S. Klasen, J. R. Vreeland, and E. Werker (2013). The costs of favoritism: Is politically driven aid less effective? *Economic Development and Cultural Change 62*(1), 157–191.
Goemans, H. E., K. S. Gleditsch, and G. Chiozza (2009). Introducing Archigos: A Dataset of Political Leaders. *Journal of Peace Research 46*(2), 269–283.
Gonnella-Platts, N. and K. E. Fritz (2017). *A Role Without a Rulebook: The Influence and Leadership of Global First Ladies.* George W. Bush Institute.
Hay, P. (1988). *All the Presidents' Ladies: Anecdotes of the Women Behind the Men in the White House.* Penguin.
Hijmans, R., N. Garcia, J. Kapoor, A. Rala, A. Maunahan, and J. Wieczorek (2020). Global Administrative Areas, Version 3.6.
Hodler, R. and P. A. Raschky (2014). Regional Favoritism. *Quarterly Journal of Economics 129*(2), 995–1033.
Jablonski, R. (2023). *Voting for Foreign Aid: How Donors Affect Political Behavior in Aid Dependent Democracies.* Unpublished book manuscript.
Kasara, K. (2007). Tax Me if You Can: Ethnic Geography, Democracy, and the Taxation of Agriculture in Africa. *American Political Science Review*, 159–172.
Kersting, E. and C. Kilby (2021). Do Domestic Politics Shape US Influence in the World Bank? *Review of International Organizations 16*, 29–58.
Kramon, E. and D. N. Posner (2013). Who Benefits from Distributive Politics? How the Studied Outcome Affects the Answer One Gets. *Perspectives on Politics*, 461–474.
Kramon, E. and D. N. Posner (2016). Ethnic Favoritism in Education in Kenya. *Quarterly Journal of Political Science 11*(1), 1–58.
Li, X., Y. Zhou, M. Zhao, et al. (2020). A Harmonized Global Nighttime Light Dataset 1992–2018. *Scientific Data 7*, 168.
Licht, A. A. (2022). Introducing Regular Turnover Details, 1960–2015: A Dataset on World

Leaders' Legal Removal from Office. *Journal of Peace Research 59*(2), 277–285.

Lujala, P., J. Ketil Rod, and N. Thieme (2007). Fighting Over Oil: Introducing a New Dataset. *Conflict Management and Peace Science 24*(3), 239–256.

McGuirk, E., N. Hilger, and N. Miller (2023). No Kin in the Game: Moral Hazard and War in the US Congress. *Journal of Political Economy 131*(9), 000–000.

Minoiu, C. and S. G. Reddy (2010). Development Aid and Economic Growth: A Positive Long-Run Relation. *Quarterly Review of Economics and Finance 50*(1), 27–39.

Moser, C. (2008). Poverty Reduction, Patronage, or Vote Buying? The Allocation of Public Goods and the 2001 Election in Madagascar. *Economic Development and Cultural Change 57*(1), 137–162.

Mu, R. and X. Zhang (2014). Do Elected Leaders in a Limited Democracy Have Real Power? Evidence from Rural China. *Journal of Development Economics 107*, 17–27.

O'Connor, K., B. Nye, and L. Van Assendelft (1996). Wives in the White House: The Political Influence of First Ladies. *Presidential Studies Quarterly 26*(3), 835–853.

Osei-Tutu, F. and L. Weill (2023). Regional Favoritism in Access To Credit: Just Believe It. BOFIT Discussion Papers 1.

Scartascini, C., C. Cruz, and P. Keefer (2021). The database of political institutions 2020 (dpi2020). *Washington, Banco Interamericano de Desarrollo*.

Stone, R. (2011). *Controlling Institutions: International Organizations and the Global Economy*. Cambridge University Press.

Stone, R. (2013). Informal Governance in International Organizations: Introduction to the Special Issue. *Review of International Organizations 8*(2), 121–136.

Strandow, D., M. Findley, D. Nielson, and J. Powell (2011). The UCDP-AidData Codebook on Geo-Referencing Foreign Aid. Version 1.1, Uppsala Conflict Data Program.

Troy, G. (2006). *Hillary Rodham Clinton: Polarizing First Lady*. University Press of Kansas.

Van Wyk, J.-A. K., A. Muresan, and C. Nyere (2018). African First Ladies, Politics, and the State. *Politeia 37*(2), 1–20.

Washington, E. L. (2008). Female Socialization: How Daughters Affect Their Legislator Fathers' Voting on Women's Issues. *American Economic Review 98*(1), 311–332.

Widmer, P. and N. Zurlinden (2022). Ministers Engage in Favoritism Too. *Journal of Public Economics 213*, 104707.

Wooldridge, J. M. (2021). Two-Way Fixed Effects, The Two-Way Mundlak Regression, and Difference-In-Differences Estimators. *Available at SSRN 3906345*.

Appendix

A Constructing PLAD

A.1 Introduction

The Political Leaders' Affiliation Database (PLAD) contains information on the birthplaces and ethnicities of the effective leaders of 176 countries around the world in the 1989–2020 period. The dataset is at the political leader level and reports information on 1,078 effective leaders, in office for 1,229 distinct terms in office. It follows the definitions in the Archigos database on Political Leaders (Goemans et al. 2009). PLAD codes ethnicities and georeferences birthplaces of the effective leader, with highest precision being the village or city level. We provide the same information for the birth regions of leaders' spouses, covering 276 spouses "in power" (32.8% percent of the leader-years in the database). The following Section A.2 describes the dataset and its variables.[55] Section A.3 contains a detailed description of the precision codes and the data collection procedure.

A.2 Variables and data set description

The following table describes the variables included in the dataset:

Table A.1 – Description of PLAD Variables

Variable	Description
idacr	Code of the country that the leader governs based on the Correlates of War Project (source until 2015: Archigos)
leader	Name of the leader (source until 2015: Archigos)
plad_id	Unique leader identification code
comments_spouse_general	General comments on the leader's relationships that do not refer to a specific spouse.
polygamous	A dichotomous variable, which indicates whether the leader had multiple spouses simultaneously.
nr_spouse	Total number of spouses that leaders had during their life.
year_spouse_change	This variable is the year a new spouse becomes the first gentleman/lady (either by government change or new marriage). If we do not know the exact year of marriage, we code the year when the president's tenure starts.

Continued on next page

[55] We link leaders' ethnicity to the definitions of the GeoEPR database (Vogt et al. 2015).

Table A.1 – continued from previous page

Variable	Description
name_spouse*	Name of the spouse of the leader.
birthdate_spouse*	Birthdate or year of birth of the spouse.
profession_spouse*	Profession of the spouse.
categorized_profession_spouse*	Category of the profession of the spouse.
categorized_education_spouse*	Spouse's subject of education.
education_level_spouse*	Spouse's level of education/degree.
education_level_isced_spouse*	Spouse's level of education/degree following the ISCED classification (Source: UNESCO).
marriage_start_spouse*	Year of marriage with the spouse.
marriage_end_spouse*	End of the marriage with the spouse (e.g., due to divorce or death of either spouse). If the couple is still together, "Present" is entered.
total_children_spouse*	Total number of children of the leader with the spouse. If there is no clear statement available on the number of children, the number of children is coded as ".".
female_children_spouse*	Total number of female children of the leader with the spouse.
male_children_spouse*	Total number of male children of the leader with the spouse.
adm1_spouse*	First-order administrative birth region of the spouse.
adm2_spouse*	Second-order administrative birth region of the spouse.
lat_spouse*	Latitude of the spouse's birthplace.
long_spouse*	Longitude of the spouse's birthplace.
geoname_spouse*	Name of the spouse's birthplace used to identify coordinates using geonames.org. The coordinates were retrieved using Google Maps when the place is not referenced on geonames.org. These rare cases are indicated in the comments.
geonamesid_spouse*	ID of the spouse's birthplace from geonames.org.
precision_spouse*	Precision of the spouse's birthplace.
foreignborn_spouse*	Dummy variable equal to 1 if the spouse is born abroad.
divorce_spouse*	Categorical variable equal to 1 if the couple divorced/separated/estranged, 2 if either the spouse or the leader died, 0 if the couple is still together, "." if no information could be obtained.
source1_spouse*	Source from which the information on the spouse was found.

Continued on next page

Table A.1 – continued from previous page

Variable	Description
source2_spouse*	Second source verifying and/or adding information on the spouse, when possible.
comments_spouse*	Further notable information on the spouse.
archigos_id	Unique identification code corresponding to Archigos variable "leadid" (available until 2015: source: Archigos).
startdate	Date of entry to office (source until 2015: Archigos).
enddate	Date of exit from office (source until 2015: Archigos).
startyear	Year of entry to office (source until 2015: Archigos).
endyear	Year of exit from office (source until 2015: Archigos). We code exit dates for leaders still in office on December 31, 2020 with this date.
adm0	Country in which the birthplace of the leader is located. Note: May differ from the country variable if the leader was born abroad. Countries are identified based on the GADM dataset (version 3.6, Hijmans et al. 2020.
adm1	First administrative division in which the birthplace of the leader is located. The first administrative division is identified based on the GADM dataset (version 3.6).
adm2	Second administrative division in which the birthplace of the leader is located. The second administrative division is identified based on the GADM dataset (version 3.6).
country	Name of the country that the leader governs.
continent	Continent in which the country governed by the leader is located.
latitude	Latitude of the leader birthplace.
longitude	Longitude of the leader birthplace.
geoname	Search term used in Geonames to geolocate birthplace.[56]
geo_precision	Precision of birthplace information on a scale of 1 - 6 (see Section 3.2).
foreign_leader	Indicator variable that is one if the leader was not born in the country of government; zero otherwise.

Continued on next page

[56]When the place is not referenced on Geonames.org, we retrieved coordinates using Google Maps. In this case, geoname corresponds to the search term used to locate the birthplace on Google Maps. These rare cases are indicated in the comments. For some rare cases (Bosnia and Herzegovina, Eritrea) we adapted some of the administrative regions provided in the GADM manually, as their data did not correspond to the actual administrative regions.

Table A.1 – continued from previous page

Variable	Description
ethnicity1	First ethnic group a leader belongs to as stated in the source.
ethnicity2	Second ethnic group a leader belongs to as stated in the source.
ethnicity_geoepr1	First ethnic group corresponding to the names of groups in the GeoEPR dataset.
ethnicity_geoepr2	Second ethnic group corresponding to the names of groups in the GeoEPR dataset.
ethnicitysource1	First source used to retrieve the leader's ethnicity.
ethnicitysource2	Second source used to retrieve the leader's ethnicity.
ethnicity_precision	Precision of the information on the leader's ethnicity on a scale of 1 - 4 (see Section 3.3).
entry	Type of entry (source: Archigos).
exit	Type of exit (source: Archigos).
gender	Leader's gender (source until 2015: Archigos).
yrborn	Year of leader birth (source until 2015: Archigos).
birthdate	Birthdate of leader (source until 2015: Archigos.)
uid	Object ID from GADM dataset (version 3.6).[57]
id_0	Numerical ID for country from GADM dataset (version 3.6) (ADM0 layer).
id_1	Numerical ID for the first administrative division from GADM dataset (version 3.6) (ADM1 layer).
id_2	Numerical ID for the second administrative division from GADM dataset (version 3.6) (ADM2 layer).
gid_0	String ID for country from GADM dataset (version 3.6). ISO 3166-1 alpha-3 country code when available (ADM0 layer).
gid_2	String ID for the second administrative division from GADM dataset (version 3.6) (ADM2 layer).
edu_name	Name of the education degree/training obtained by the leader and the field of study.
edu_r	Leader's level of education summarized in eight categories (source until 2012: Yu and Jong-A-Pin 2020).

Continued on next page

[57] Object IDs are from the GADM polygons. We only provide them for leaders whose birth region we could identify at least at the second administrative level (precision code 3), as we merged our data with the GADM dataset at the second administrative level. We advise users who would like to merge the dataset on the first administrative layer via object IDs to carry out an individual geo-merge with the GADM data via the latitude and longitude data.

Table A.1 – continued from previous page

Variable	Description
birthplace_comment	Notes on controversial cases and additional information on the leader birthplace.
ethnicity_comment	Notes on controversial cases and additional information on the leader's ethnicity.

* These variables are not yet included in the public version of our dataset.

A.3 Details on the data collection process

A.3.1 General information

The Political Leaders' Affiliation Database is a manually compiled dataset. We collected information by structured internet searches. As sources, we mainly rely on online libraries and databases such as CIA World Factbook, Munzinger, Encyclopaedia Britannica, and Ethnicity of Celebs, as well as on reports from popular news services such as BBC News, The Guardian, and The Washington Post.

The dataset provides information on the birthplaces of leaders and their spouses if at least two reliable sources report the same location of birth. In most cases, the determination of birthplaces is clear and uncontroversial. As information on ethnicity is not reported as frequently as birthplaces, we provide a leader's ethnicity if at least one source is available and state the sources used in the dataset. Given the scarcity of data, we do not report data on spouses' ethnicity.

In order to ensure the accuracy of the data, we apply the following three steps:

1. A second coder reviewed all information.

2. We further checked for consistency by comparing our data with those of Dreher et al. (2019), Bommer et al. (2022) and Berlin et al. (2023).

3. The lead researchers randomly selected and screened 25% of all leaders' entries again as quality assurance.

A.3.2 Information on birthplaces

This section describes the georeferencing of national leaders' and their spouses' birthplaces. The location of a person's origin is determined by biographical information on the place where they were born. If sources stated contradictory places, we noted this in the column birthplace_comment. For the identification of the birthplaces, we used the online encyclopedias Encyclopedia Britannica and Wikipedia[58] and complemented

[58]While we prefer other non-editable sources, Wikipedia offers comprehensive and well-structured information on leader characteristics.

this information with resources focusing on bibliographic accounts, for example, CIDOB, The Famous People, and Munzinger. After identifying the birthplaces, we obtained the respective coordinates using Geonames (https://www.geonames.org/). We provide the search term used to retrieve the coordinates in the variable geonames.[59] The variables longitude and latitude contain the coordinates of the leaders' birthplace. As the quality of information, i.e., the accuracy of coordinates, differs substantially between different leaders, the variable geo_precision contains information on how accurate the coordinates are. The categories of geo_precision are based on Strandow et al. (2011):

1 = The coordinates correspond to an exact location, such as a populated place (villages, cities) or a hill. The code is also used for locations that join a location which is a line (such as a road or railroad). We coded the points that connect lines, but not the lines between points (all points that are mentioned in the source are coded).[60]

2 = The location is mentioned in the source as being "near," in the "area" of, or up to 25 km away from an exact location. The coordinates refer to that adjacent, exact, location.

3 = The location is, or is analogous to, a second order administrative division (ADM2), such as a district, municipality or commune.

4 = The location is, or is analogous to, a first order administrative division (ADM1), such as a province, state or governorate.

5 = The location can only be related to estimated coordinates, such as when a location lies between populated places; along rivers, roads and borders; more than 25 km away from a specific location; or when sources refer to parts of a country greater than ADM1 such as a National Park which spans across several provinces (e.g., Foret Classee de Gongon in Benin).

6 = The location can only be related to an independent political entity, meaning the pair of coordinates that represent a country. This includes leaders that were born in larger areas that cannot be geo-referenced at a more precise level.

7 = Unclear. No coordinates are entered to reflect that sub-country information is unavailable. Furthermore, the names of the second and first administrative

[59]Oftentimes, the capital city has the same name as the administrative divisions in which it is located, on the first and second level. If we cannot identify whether leaders were born in the city or in the broader administrative division, we adopt a conservative approach and assume that they were born in the first or second order administrative division that contains the capital city and shares the same name.

[60]Sometimes we could identify a precise birthplace, but the GADM data did not provide information on ADM2 regions. In this case, we add a note in the variable birthplace˙comment.

region can be found in adm1 and adm2, respectively. The variable foreign_leader shows if a national leader was born in a foreign country. It takes a value of one if the leader was not born in the country they govern and is zero otherwise. The variable birthplace_comment gives further information on the birthplace and explains decision making in controversial cases.

A.3.3 Information on (leaders') ethnicity

The concept of ethnicities relates to a belief of a common culture and ancestry, being thus inherently subjective (Weber 1976). For our purpose, we draw on a commonly used classification by Vogt et al. (2015), who suggest a classification of such common culture and ancestry by features including (i) language, (ii) beliefs/religion, or (iii) phenotypical characteristics. The salience of those features may differ by world region, where ethnicities in Latin America separate by skin color, while African ethnicities are more separated along the lines of linguism. This also relates to the level of ethnic distinction, where some countries only know two ethnicities and in other countries ethnicity is a very complex concept with several sub-groups. We rely on the ethnic concepts as mentioned in the sources and the main groups given in Vogt et al. (2015). While the level of differentiation is not fixed across countries, it represents the salience within the country and is thus at the relevant level for questions in political economy.[61] Furthermore, similar ethnic groups may be named rather heterogeneously across countries, where, for instance, "Afro-American," "Afro-Haitian" or "black" refers to people with African descent. In those cases we stick to the ethnicity name used by the source or by Vogt et al. (2015), to minimize interpretation biases. We also draw from previous databases on leader characteristics (e.g., Fearon et al. 2007, Parks 2013), and encyclopedias such as CIDOB, The Famous People, Munzinger, Encyclopaedia Britannica, and Ethnicity of Celebs. If sources named contradictory ethnicities, we note this in column ethnicity_comment.

Note that the availability of information on leaders' ethnicity depends on the country context. Information is less likely to be available in countries where ethnicity is less salient, e.g., in less ethnically fractionalized societies. What is more, data availability depends on whether ethnicity is determined by language, phenotypical factors, or religion. Therefore, the quality and quantity of sources differ strongly across contexts. For countries that are less represented on major encyclopedias, we draw on country-specific resources like books or webpages of the parliament but also country-specific webpages like Afghan Bios for Afghanistan, Banglapedia for Bangladesh, or BiographyBD for India and Pakistan. What is more, some sources just provide indirect information on the ethnicity

[61] Due to differing relevance of ethnic affiliation across countries, sometimes it was harder to find data and we had to base information regarding ethnicity on related concepts (nationality, skin color or family ties). We coded the ethnicity_precision in those cases as 3 and added in column ethnicity_comment "Assessment of ethnicity is based on nationality/skin color" or provided a reference to the family's linkages (e.g., "parents were farmers with long ancestry in the region").

of leaders based on the individual's ancestry. Here, we also offer users the option to filter with the following precision codes (variable ethnicity_precision):

1 = Two sources state the ethnicity directly.

2 = Only one direct source or one of the sources is Wikipedia.

3 = No direct mentioning of ethnicity. Attribution via characteristics mentioned in the text or phenotypical factors in picture.

A.3.4 Information on (leaders') education

Data on leaders' education level are taken from Yu and Jong-A-Pin (2020) until 2012 and updated to include leaders up to the year 2020. Leaders' level of education is summarized in the categorical variable edu_r following the eight-way classification used in Ludwig (2002):

1 = Illiterate (no formal education)

2 = Literate (no formal education)

3 = Elementary/primary school education or tutors

4 = High/finishing/secondary/trade school

5 = Special training (beyond high school, such as mechanical, nursing, art, music, or military training)

6 = College-educated

7 = Qualifications from a graduate or professional school (e.g., master's degree)

8 = Doctorates (e.g., Ph.D.)

Military training programs that do not lead to a bachelor degree are considered to be category 5. When it is known that a leader attended college, but there is uncertainty on whether they graduated or the level at which they graduated, we code their education level as category 6. However, if the leader is known to be a lawyer or a medical doctor, we rank the education level as category 7, as these professions require at least a master's degree in most countries. We supplement the data from Yu and Jong-A-Pin (2020) with the variable edu_name, which reports the name of the highest degree obtained by the leader, complemented with their field of study, when such information is available.

We collected data on spouses' education level in a similar way.

A.3.5 Information on (spouses') profession

We categorized spouses' professions into ten categories: Art/Entertainment, Activism, Business, Education (e.g., professors, educators, teachers, lecturers, academics), Health (e.g., pediatricians, cardiologists, nurses, physicians), Politics, Sciences/Researcher, Social Work (e.g., philanthropists, charity workers, counselors, social project managers, humanitarian workers, and volunteers), Law, and Other. Some spouses could have more than one profession. In these cases, all categories are listed and separated by a comma.

A.3.6 Descriptive statistics

Tables A.2 and A.3 show summary statistics.[62] On average (in the 1990–2020 period), an ADM2 region is coded as spousal region in about 0.14%—0.044 years—across all regions. Conditional on being a spousal region, the average (uninterrupted) duration of a spouse region is 6.3 years. There is some variation across continents: an ADM2 region is coded on average as spousal region in about 0.25% (0.08%) of all years in Africa (the Americas). An average spousal region spell lasts 9.24 years (4.11 years) in Africa (Oceania).

500 ADM2 regions change their status as spousal birth region at least once. This value is lowest in Oceania, with only 17 changes, and highest in Europe (180 changes). Considering within-region changes, we observe the highest number of spouse region switching in Bagmati, Nepal (8 changes).

It is important to note the substantial overlap between the birth regions of leaders and their spouses (see Figure 1 in the main text). The raw correlation between the two variables is 0.226. Out of the 1,979 spousal-year observations in our sample, 482 are leader-year observations as well. Of the 500 ADM2 regions that change their status as birth region at least once, 1.8% correspond to a matching change in leader region—e.g., while both Kolinda and her husband Jakov Kitarović were born in Rijeka, their successors as president and first spouse of Croatia (Zoran Milanovic and Sanja Musić) both originated from Zagreb.

[62] Note that we provide statistics for the entire database here, rather than the estimation sample used in the main part of the paper.

Table A.2 – Descriptive Statistics Birth Regions

	Mean	Std.Dev.
Spouse		
Spouse birthregion dummy	0.14	3.75
Change spouse birthregion	0.04	1.88
Leader		
Leader birthregion dummy	0.31	5.55
Change leader birthregion	0.08	2.87
Both		
Spouse and leader birthregion dummy	0.03	1.87
Change of spouse and leader birthregions	0.01	0.96

Note: All dummies equal 100 if the person(s) was/were born in the region, 0 otherwise. Thus, values should be interpreted as percentage shares. Descriptive statistics refer to the full sample of the PLAD.

Table A.3 – Descriptive Statistics Birth Regions by Macro Region

	Spouses			***Leaders***			***Both***		
	Mean	Std.Dev.	Obs	Mean	Std.Dev.	Obs	Mean	Std.Dev.	Obs
Africa									
Birthregion dummies	0.2483	4.9768	192512	0.6856	8.2519	194272	0.0343	1.8513	192512
Americas									
Birthregion dummies	0.0795	2.8186	491779	0.1469	3.8305	482508	0.0267	1.6353	482276
Asia									
Birthregion dummies	0.1517	3.8922	339434	0.3073	5.5349	345922	0.0426	2.0636	338014
Europe									
Birthregion dummies	0.1634	4.0393	340221	0.2881	5.3597	356488	0.0421	2.0512	339738
Oceania									
Birthregion dummies	0.1432	3.7809	27244	0.4424	6.6366	28708	0.0000	0.0000	27244

Note: All dummies equal 100 if the person(s) was/were born in the region, 0 otherwise. Thus, values should be interpreted as percentage shares. Descriptive statistics refer to the full sample of the PLAD.

References

Berlin, M. P., R. M. Desai, and A. Olofsgård (2023). Trading Favors? UN Security Council Membership and Subnational Favoritism in Aid Recipients. *Review of International Organizations 18*(2), 237–258.

Bommer, C., A. Dreher, and M. Perez-Alvarez (2022). Home Bias in Humanitarian Aid: The Role of Regional Favoritism in the Allocation of International Disaster Relief. *Journal of Public Economics 208*, 104604.

Dreher, A., A. Fuchs, R. Hodler, B. C. Parks, P. A. Raschky, and M. J. Tierney (2019). African Leaders and the Geography of China's Foreign Assistance. *Journal of Development Economics 140*, 44–71.

Fearon, J. D., K. Kasara, and D. D. Laitin (2007). Ethnic Minority Rule and Civil War Onset. *American Political Science Review 101*(1), 187–193.

Goemans, H. E., K. S. Gleditsch, and G. Chiozza (2009). Introducing Archigos: A Dataset of Political Leaders. *Journal of Peace Research 46*(2), 269–283.

Hijmans, R., N. Garcia, J. Kapoor, A. Rala, A. Maunahan, and J. Wieczorek (2020). Global Administrative Areas, Version 3.6.

Ludwig, A. M. (2002). *King of the Mountain: The Nature of Political Leadership*. University Press of Kentucky.

Parks, B. (2013). *Brokering Development Policy Change: The Parallel Pursuit of Millennium Challenge Account Resources and Reform*. Ph. D. thesis, London School of Economics and Political Science.

Strandow, D., M. Findley, D. Nielson, and J. Powell (2011). The UCDP-AidData Codebook on Geo-Referencing Foreign Aid. Version 1.1, Uppsala Conflict Data Program.

Vogt, M., N.-C. Bormann, S. Rüegger, L.-E. Cederman, P. Hunziker, and L. Girardin (2015). Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Data Set Family. *Journal of Conflict Resolution 59*(7), 1327–1342.

Weber, M. (1976). *Wirtschaft und Gesellschaft: Grundriss der verstehenden Soziologie*, Volume 2. Mohr.

Yu, S. and R. Jong-A-Pin (2020). Rich or Alive? Political (In)stability, Political Leader Selection, and Economic Growth. *Journal of Comparative Economics 48*(3), 561–577.

B Constructing GODAD

B.1 Introduction

The Geocoded Official Development Assistance Dataset (GODAD) that we introduce with this paper provides newly georeferenced aid data on 18 European bilateral donors and the United States, for the period between 1973-2020. We geocode these data in two steps. First, we extract relevant geographic information from CRS projects. We then pass candidate locations for each project on to a geocoder that returns a matched location and coordinates. The purpose of this appendix is to describe the details of how we extract these data. We also evaluate the quality of the resulting dataset, whether and to what degree the resulting sub-set of projects is a representative sub-set of all projects, and compare them to existing (AIMS) data for Uganda.

Section B.2 describes the data pipeline for constructing this dataset, starting with a description of the original CRS data, and then explaining the different steps that we take to geolocate the projects. Section B.3 evaluates the final geocoded project data, reporting descriptive statistics and robustness tests aimed at assessing their accuracy and quality. We compare our data with CRS in Section B.4, and with existing geocoded data for Uganda in Section B.5.

B.2 The data pipeline

B.2.1 Description of raw data

We extract data for 19 bilateral donors from the OECD's Creditor Reporting System, which provides detailed information on individual aid activities at the project level from 1973-2021.[63] The CRS data provide information about projects along several dimensions, including donor and recipient countries, donor agencies, channel of delivery (government, NGO, institutes), flow type (grants, loans, other official flows), sectors and sub-sectors of aid, and commitment and disbursement amounts. For most projects, the data also include project titles and descriptions. It is these fields that we use to extract information to geolocate the projects. Table B.1 provides some examples of these raw data.

The CRS vintage we use includes 1,647,838 projects from the 19 donors combined, for the full 1973-2021 period. While we include all available years in our database, the statistics, figures, and tables below refer to the years 1990–2020, which are the years we included in the empirical analysis of this paper (1,605,748 bilateral projects from the CRS in total from 1990 to 2020). Project identifiers provided in the CRS do not identify unique projects and not all projects are committed to individual countries or subnational regions. The raw data thus require initial cleaning. First, we drop all projects which

[63] At the time of extraction (January 2023), data for 2021 were only partially complete. The files in raw text format are available at https://stats.oecd.org/DownloadFiles.aspx?DatasetCode=CRS1.

Table B.1 – CRS raw data example

DonorName	RecipientName	ProjectTitle
Germany	Afghanistan	Building a skateboarding facility in Kabul to engage youth throughout Afghanistan, building technical skills, confidence and life opportunities
Netherlands	Afghanistan	BBC Kunduz
United Kingdom	Afghanistan	Helmand Alternative Livelihoods Programme (HALP)
Italy	D.R.C	Renovation of the toilets of the Notre Dame College of Mbasa-Mboma
Norway	D.R.C	LIKATI AGRICULTURAL PROGRAM
Italy	D.R.C	Costruzione di una scuola nel quartiere periferico di Motumbe
Italy	D.R.C	Agricultural production in the territories of Aketi
Italy	D.R.C	Support for the medical faculty of the University of the Uele
Norway	D.R.C	DNB-Education Lower-Bas-Uele

Note: Table shows the sample of projects before and after geocoding, for the years 1990–2020. Values are in millions (constant 2014) US$.

are not clearly bilateral (i.e., the recipient is either a macro region ("Africa") or is not specified). Second, we drop projects that donors provide to the central government or that stay in the donor country, such as budget aid, donors' costs for administration or hosting refugees, or debt relief. We also exclude project entries that aggregate various smaller expenses and thus cannot be attributed to specific regions. Finally, we drop duplicate entries that refer to the same project more than once. These steps reduce the number of projects available for geocoding to 1,541,106 for the 1990–2020 period.[64]

B.2.2 Entity extraction

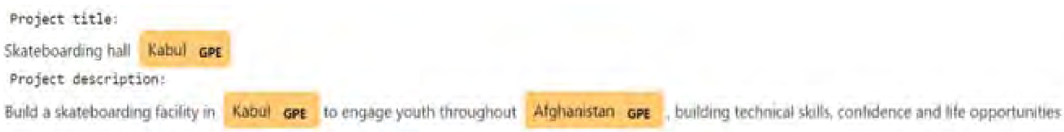
To extract geographic entities from the data, the first step consists in utilizing the Spacy library for natural language processing tools; specifically, the (pre-trained) Spacy core English transformer pipeline with the Named Entity Recognition (NER) model. These models are typically used to identify within text pieces of information such as names, actions, or geopolitical entities. The advantage of this specific pipeline is in its speed, flexibility, and method of processing text data. Transformer models process all inputs bidirectionally, unlike traditional recurrent neural networks which process sequentially. This allows for greater parallelization in computations and hence speed, and improved

[64]To provide additional input information for the entity extraction model, using Google Translate API to translate 122,000 of the total raw projects into English. These are predominantly Spanish (44,409), French (26,293), and German (21,855).

accuracy because the model learns to interpret sentences, or pieces of string, from multiple directions. Furthermore, this parallelization has allowed the models to be trained on massive datasets, resulting in greater accuracy. The underlying model is the RoBERTa-base model, trained on the entire English language Wikipedia and the online book corpus, a large online collection of digitalized books. The NER feature of this pipeline is run on the different sources of original text information (as well as translated information where available) for each project: the title, the short description, and the long description. Figure B.1 shows a stylized example of what the model would identify.

Finally, to partially correct some of the potential shortcomings of the NER model, we implement a fuzzy string matching approach.[65] In the first part, we split the text into chunks and filter out "noisy" words based on the frequency of words in the full dataset. The algorithm then matches the remaining candidate words with a hierarchical dataset of location names from the GeoNames database. This dataset consists of a set of organized text files, where there are lists of all administrative units, cities, and localities for each country. The procedure is essentially a record-linkage approach, which returns a closeness score for each matched candidate word. We keep only the match that is ranked as most precise and consider it as a candidate geo-entity for the geocoder when the precision is sufficiently high and the NER output did not return suitable candidate words. The advantage is that the algorithm always extracts at least one match for each string. This additional entity extraction method is particularly useful in cases where the project description is short but the geographical entity is clear (e.g., "Dam restoration, Cairo"). Of all the projects considered for geo-entity extraction, for 110,628 of them the fuzzy matching algorithm was used to supplement the NER output.

Figure B.1 – NER model geo-tag example



Note: Figure shows stylized example of Named Entity Recognition model output.

B.2.3 Geocoding and data cleaning

We geocode the extracted entities with the GoogleV3 geocoding API. The choice of the geocoder is relevant when systematically checking the coordinates for a large number of cross-country locations. First, although the performance of Google Maps may vary by country, overall, it provides results of consistent quality across the sample. Unlike other

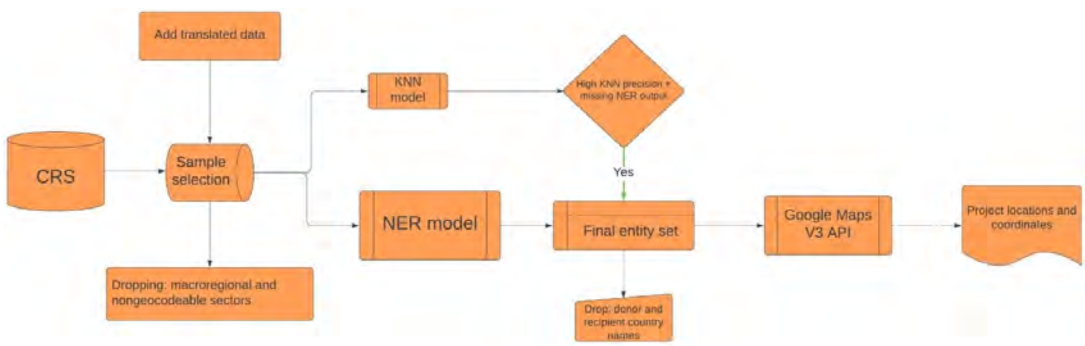
[65] Specifically, we use a Term Frequency–Inverse Document Frequency (TF–IDF) algorithm with k-Nearest Neighbor (KNN) matching.

popular (free) geolocating services, which essentially match locations with large databases that are in turn taken from multiple geodatabases of locations, the Google Maps API can leverage more sophisticated ranking algorithms in combination with an immense amount of lower-level GPS, WIFI, network and other data. Second, the geocoder can be set up to be restricted in its search to the recipient country. This ranking algorithm for matching locations results in substantially fewer errors with respect to geocoders which rely on fuzzy matching algorithms.

Following the geocoding, we implement a series of data cleaning procedures. First, project-locations with no subnational information are dropped.[66] When the model extracts more than one location, we check whether these entities are multiple project location or entities that are nested in each other, such as a city located in a larger ADM1 region. For example, a project description may contain references to both cities in a province as well as a reference to the province. We then exclude the more central administrative units to avoid double counting.[67]

Figure B.2 provides a schematic overview of the data pipeline in use. The following section evaluates this pipeline in its two main stages—the entity extraction and the geocoding of entities.

Figure B.2 – Entity extraction and geotagging pipeline



Note: Visualization of the steps in the geocoding pipeline, from raw CRS data to entity extraction to geocoding.

B.3 Evaluating the dataset

The final dataset consists of geographic information for 220,369 unique projects (and 285,558 project-location combinations). This corresponds to about 15% of the original,

[66] Google Maps API returns the country centroid when the geolocation precision is very low, meaning that it was unable to return anything more than the search country it was assigned.

[67] There are a number of examples. The project could reference both "Lashkargah, Helmand, Afghanistan" as well as "Helmand, Afghanistan," which would both be tagged in by the NER model. In this case we would only keep "Lashkargah." In other cases the project might reference "Lashkargah, Helmand province, and Kabul province." In this case, we would keep both Lashkargah and Kabul.

1,541,106 bilateral projects considered. However, it is important to note that this does not imply that the data pipeline failed to geocode the majority of projects. Because the CRS is itself a dataset compiled from primary data with different degrees of accuracy and compiling errors, many projects do not include the necessary information we require to geocode them. Furthermore, even among the sample of potentially geocodable project sectors (i.e., not budget support or administrative costs), the majority of aid projects are not earmarked for specific, subnational purposes. For example, aid projects in the health sector may consist of localized projects, such as a 2008, US$1.29 million grant from the Spanish Ministry of Foreign Affairs and Co-operation to provide health services in Afar and Amhara in Ethiopia. Such projects would be geocodable within our pipeline. However, they may often be large loans for sectoral budget support in the health sector (US$ 113 million loan to Bolivia from the French Development Agency during Covid) or country-wide grants under partnerships for broad projects (a US$ 129 million grant from the UK as part of an initiative to eradicate polio in India in 2002). These projects are both large and non-geocodable.

Table B.2 provides an overview of the final data by showing the main descriptive statistics both before and after the geocoding. On average, projects are worth US$8 million, but with substantial variance. Similarly to the examples above, very large loans and grants are typically country-wide aid flows or are part of larger initiatives, such as the Iraq Relief and Reconstruction programs under United States funding in 2004. This series of projects in key infrastructure was financed by large grants, with electricity and oil infrastructure projects requiring US$1.7 billion and US$900 million respectively. However, more precise information in the project descriptions which could indicate the specific region of these projects is missing. These cases drive much of the changes in the distribution of project sizes before and after geocoding. Finally, Figure B.3 shows the average commitments by aid sectors for the pre- and post-geocoded sample. In general, the distribution remains stable, indicating that the geocoding process did not introduce major distortions.

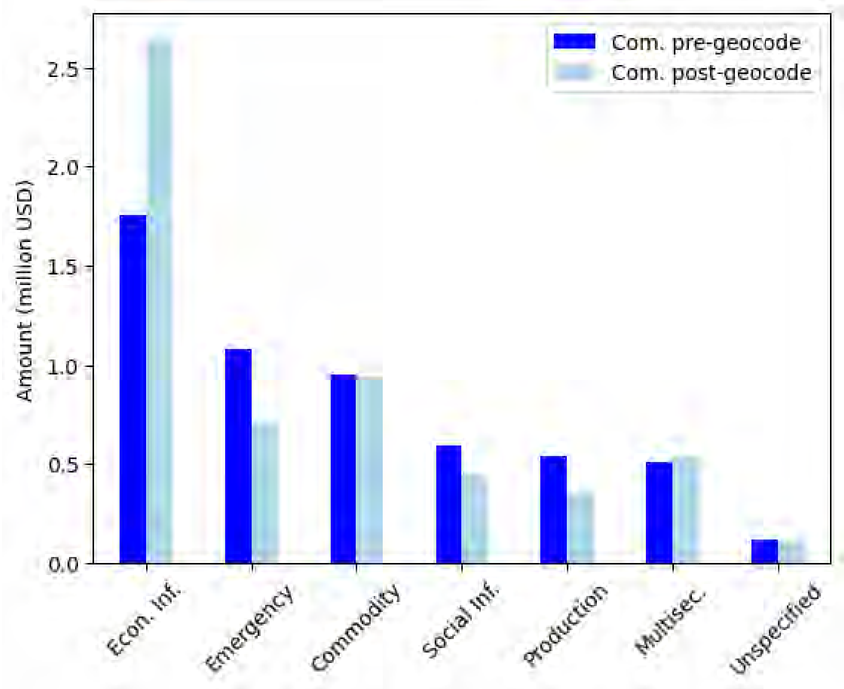
Table B.2 – GODAD EU and US sample description

	Pre-geocode			**Post-geocode**		
	N	Mean	S.D	N	Mean	S.D
Number of projects	1,541,106			220,480		
Commitments	1,221,102	8.5	7.7	194,127	0.7	5.9
Disbursements	1,344,909	5.5	4.6	203,687	0.5	3.1

Note: Table shows the sample of projects before and after geocoding, for the years 1990–2020. Values are in millions (constant 2014) US$.

In the following sections we describe potential limitations in the data pipeline in

Figure B.3 – Pre and post geocoding commitments by sector



Note: Figure shows the average size of project commitments across the main aid sectors; multisector aid, social infrastructure, economic infrastructure, and commodity and emergency aid.

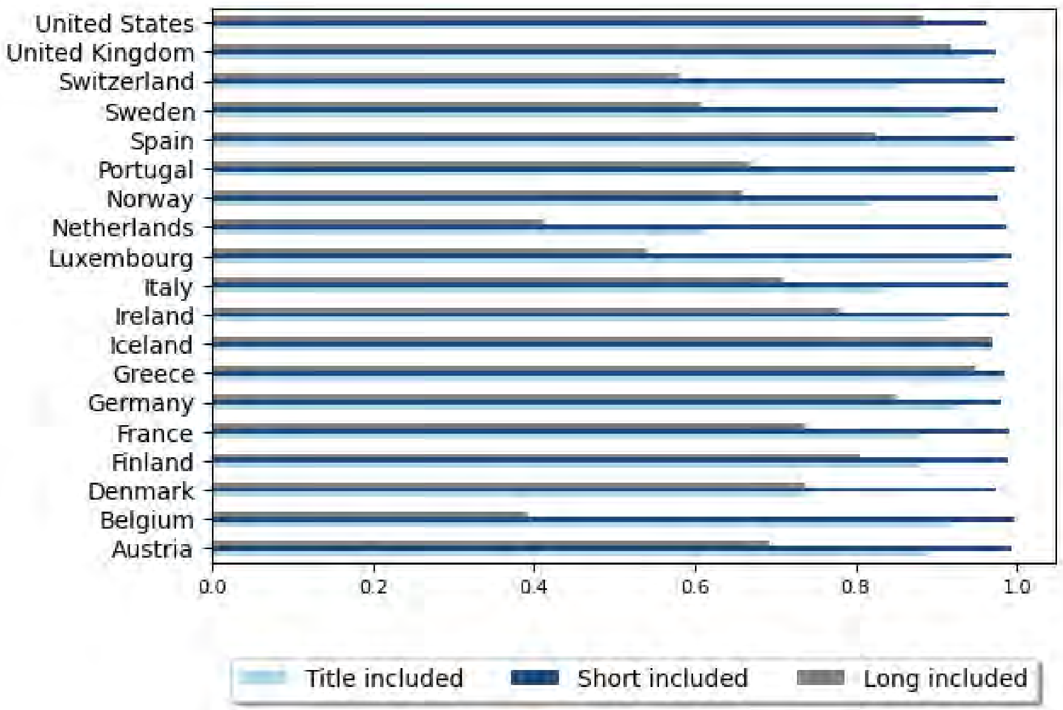
detail. We divide the evaluation of the dataset in three parts: Section B.3.1 describes the raw data—namely project titles and descriptions—from which the project locations are extracted. Section B.3.2 evaluates the first stage of the data pipeline—the entity extraction from the project titles and descriptions. Section B.3.3 evaluates the final geocoded dataset.

B.3.1 CRS raw data descriptives

The starting point for the data pipeline is the project titles and descriptions from the CRS data. Consequently, the quality of the geocoded data depends in large part on the quality of the raw data. This section provides additional descriptive statistics on the text used to extract and geolocate project locations. Figure B.4 shows the share of projects for which at least one data field—project titles, short description, or long descriptions—is available, for each donor separately. In general, between the project title or one of the short or long descriptions, the vast majority of projects contain sufficient textual information to run the NER model on.[68]

[68] The average project title includes 56 characters. Short and long project descriptions include 49 and 291 characters on average, respectively. The length of descriptions for projects with non-missing information varies substantially across donors, with Belgium, Finland or Spain being more detailed in their descriptions, while other donors are on average very similar. Certain small donors such as Greece

Figure B.4 – Share of missing values in raw text data



Note: Figure shows the share of project title, short, and long descriptions for which text data is available.

B.3.2 Evaluating the Named Entity Recognition (NER) model accuracy

To evaluate the accuracy of the first-stage entity extraction model, we use a random sample of 1,000 hand-coded projects to compare the model output with human coding. We run the project descriptions and titles of this random sample through the NER model and confront the model output with the hand-coded outcomes. Equation (B.1) shows how we compute an accuracy metric for this output, defined as the share of corresponding model and hand-coded locations over total model and true locations for each project. Accuracy can take values from 0 to 1, where 1 means that model and true outcomes perfectly match. This metric is therefore not a binary measure of whether the individual project geocoding was correct. Over the entire random sample of 1,000 projects, total accuracy—defined as the average of the project-level measures—reaches 89%.

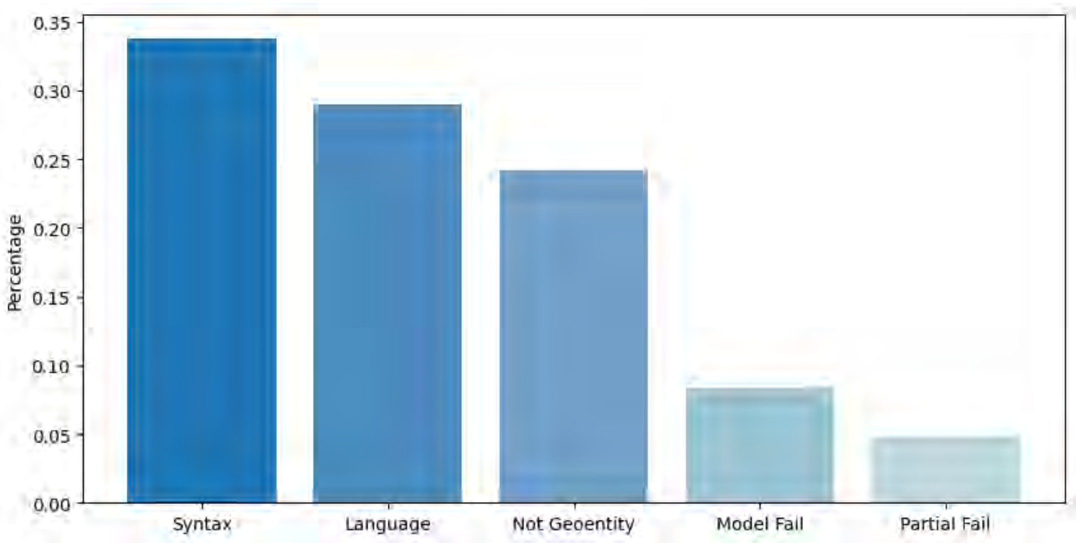
$$Accuracy = \frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|} \tag{B.1}$$

In terms of individual projects, the model produced errors in 145 cases out of 1,000. Figure B.5 shows the distribution of these specific errors. The largest sources of errors are syntax issues and non-English language of titles and descriptions, corresponding to around

or Luxembourg provide less wordy descriptions.

34% and 29% (49 and 42 total) of all errors, respectively. Syntax issues are errors deriving from poorly structured text, such as short sentences or improper grammar. Language models such as the one we employ here for entity extraction rely essentially on learned probabilities about the distribution of elements of a sentence (parts of speech, prefixes, suffixes, punctuation, etc.) to identify entities in a text. Poorly structured sentences, such as a project title which may not contain a verb or punctuation but for which the location can be inferred ("Hydroelectric Dam, Giza"), will increase the likelihood that the model fails to extract an entity. The abundance of available English language text suitable for training these models with respect to other languages implies that these models will perform more consistently on English text. In both these cases of syntax and language, the errors correspond almost always to instances of false negatives, where the model does not return a true location. 24% of the errors (35 projects) included cases where the extracted entity was not a real geographical entity (false positive). In these cases, the model may be "tricked" by phrases such as "Moving to Freedom," which refers to the name of an NGO project and not a city called Freedom. In other cases, the model may extract geopolitical qualifiers which are not true locations ("Kyrgyz administrative systems"). What is more, in some instances the model fails to extract either all locations (12 cases) or some of the locations in the text (7 cases), for no reason attributable to the above categories. These errors are marginal categories, only 8% and 4% of cases.

Figure B.5 – Share of missing values in raw text data



Note: Figure shows the distribution of NER model errors based on a hand coded dataset of 1,000 projects. Syntax errors relate to grammar or length. Language refers to issues of non-English text. Not geoentity errors are false positive model errors. Failure and partial failure are false negative errors.

B.3.3 Final output evaluation

This section analyses the quality of the final geocoded data. The goal is to evaluate the final output rather than the individual steps. There are two main assumptions that need to be made before evaluating the quality of the geocoding process. First and most importantly, we exclusively focus on feasibly geolocatable entities. Projects at times may reference numerous, extremely small localities (neighborhoods, small villages, etc.) as beneficiaries. However, to the best of our knowledge there is no geocoder which will return consistently correct coordinates, across all countries, for these small localities. Our geocoder of choice (GoogleV3) almost always identifies administrative regions (provinces, municipalities, regions) or cities. Finally, most project descriptions include locations that allow us to geocode at the ADM2 level, usually referencing large cities or municipalities. The data thus cannot be coded at the level of small administrative entities such as villages, isolated localities, or neighborhoods (which for some countries correspond to ADM3 or ADM4 regions).

Similarly, our goal is to geocode explicitly mentioned geographical entities, not those from which the location can be inferred. For example, some projects may reference an allocation of funds to the University of Guadalajara for a collaboration. While the university is a clearly geolocatable entity, the entity extraction in the first stage of the data pipeline has been tailored to identify strictly geographical entities and therefore may not tag the project to be placed in Guadalajara. The geocoded dataset provided reflects the allocation of funds for place-based development projects, as opposed to the transfer of funds to subnational non-geographic entities.

We evaluate the final dataset by checking a random sample from each CRS-defined recipient macro region in the CRS data (Caribbean and Central America, South America, North of Sahara, South of Sahara, Middle East, South and Central Asia, Far East Asia, and Oceania), which we sample under two conditions. First, each macro region's random sample contains at least 1,000 projects which mimic the share of projects by donors in the region. In other words, if in South America Spain accounts for 30% of the projects in the data, then the random sample is composed of 30% Spanish projects. However, a minimum of 100 projects are considered for each donor. We do the sampling at the project level and not the observation level as each project may contain multiple locations with different errors. Finally, the random sample goes through the standard data cleaning done in the pipeline.

Table B.3 presents a scheme for the main errors considered. Broadly, errors can be classified into two types: false positives and false negatives. False positives refer to locations which have been wrongly assigned to a project, while false negatives refer to locations missed by the geocoding process. These error types in turn can be separated into three main categories: common names in project descriptions, failure by the geocoder, and NER model errors.

"Common names" appear exclusively as false positive type errors. In these cases, the geolocated entity does not correspond to the intended project location. For example, if the project describes funds allocated from the municipality of Cordoba (Spain) to Colombia, the project may be associated to the location of Cordoba, Colombia. In the cases classified as "geocoder fail," most errors are also false positives, where the geocoder received the correct input but was not able to provide a specific location. These false positives however are easy to detect, as the geocoder returns the geometric centroid of the country for the majority of these cases. Other specific examples include projects which describe ethnic territories or historical names of locations. Humanitarian aid projects for the Sahawari refugee camps on the border between Western Sahara and Algeria are almost always described with Morocco being the recipient country, but the project description references the camps on the Algerian side. The geocoder, calibrated to search within Morocco, may not return anything for these cases. At times the geocoder may fail to return coordinates because of nomenclature differences for geographical locations between the donor and recipient country. This is separate from the case where the entity was not tagged in the string because of language limitations in the NER model. For example, the first stage entity extraction may correctly identify the Muskitia coast in Honduras as a location, but the geocoder may not recognize it because it expects the spelling "Mosquito Coast."

Finally, there are a set of errors attributable to the first stage of the data pipeline—the entity extraction phase. These specific issues were explained in detail in Section B.3.2. When the entity extraction fails to extract the location, typically because of syntax or language, then it results in false negatives. In these cases, the geocoder will receive no input and return no location. In other cases, the NER model may tag some arbitrary entities in the project descriptions which are not geographic entities, resulting in false positives.

Figure B.6 shows the error rates across different categories. In the random sample of over 10,000 projects, about 5% of the projects presented one of the errors described in Table B.3. Among all errors, 45% (313) derive from the "common names" category. In these instances, the geocoding pipeline returns locations which are not strictly related to the geographic scope of the project, as described above. "Geocoder fail" and "NER model fail" together are relatively less important error categories (around 25% each, or specifically 174 and 187 errors). When we further decomposes these error categories into the share of error type (false positive vs. false negative) we find that, as previously explained, common names correspond to false positive errors. Aside from this, false positives also occur when the geocoder fails due to nomenclature differences or contested territories. False negatives, or instances where the geocoding pipeline fails to extract and then geocode a location, are a small category in absolute value, totaling just 187.

Table B.3 – Geocoding pipeline errors

	False positives	False negatives
Common names	Similar names in donor and recipient Sub-national entity not target of project Mountains, lakes, etc.	
Geocoder fail	Antiquated nomenclature Contested or ethnic territories Error in reported recipient country Nomenclature differences	Geocoder fail (no output) Nomenclature differences
NER model fail	Not geo entity	Syntax Language NER model failure

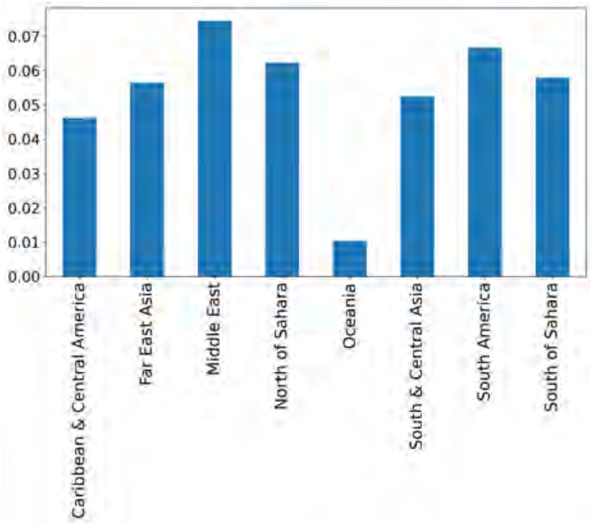
Note: Figure shows the error rates across different categories for a random sample of over 10,000 projects.

When we focus on individual donors rather than aggregates the error rate is below 10% for most countries. Exceptions are Finland, Iceland, and Spain with an error rate between 11% and 12%. When we disaggregate the error rate by recipient region rather than donors, no region stands out. The exception is the rather small region of Oceania, with an error rate of less than 1%.

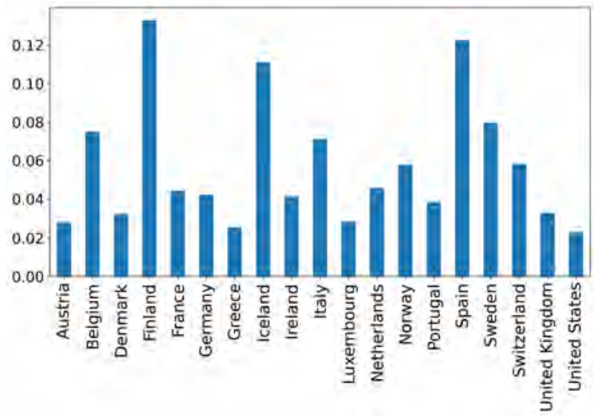
Figure B.7 shows the final decomposition of errors in the geocoding pipeline, given as the share of errors by category and donor over the total number of donor projects. The figure is useful to highlight the outliers in the error rates (measured by the number of projects of a donor with a given error over total donor projects) among donor countries. In the case of Finland, common names in project descriptions determine most of the errors. For Iceland, the geocoder failed more often than average, with respect to other donors. The error rate determined by geocoder failure in Iceland is almost 9%, compared with the mean of 2%. Both Iceland and Finland are relatively minor donors in terms of total commitments. NER model failures, which determine most missed cases of locations (false negatives), are rare across all donors, with an average error rate of 1% and with some countries like Ireland, Iceland, or Netherlands having none of these issues in the random sample.

Figure B.6 – Decomposition of pipeline errors

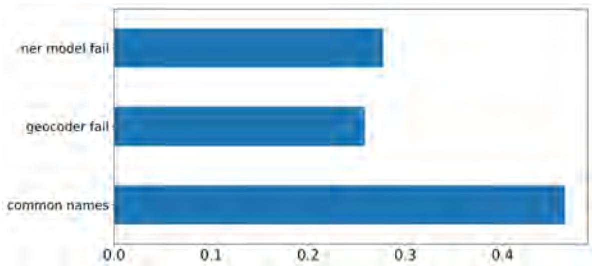
Panel A: Error rate by macro-recipient region



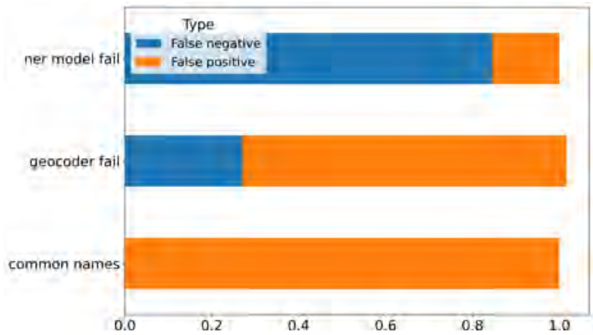
Panel B: Error rate by donor



Panel C: Error by type

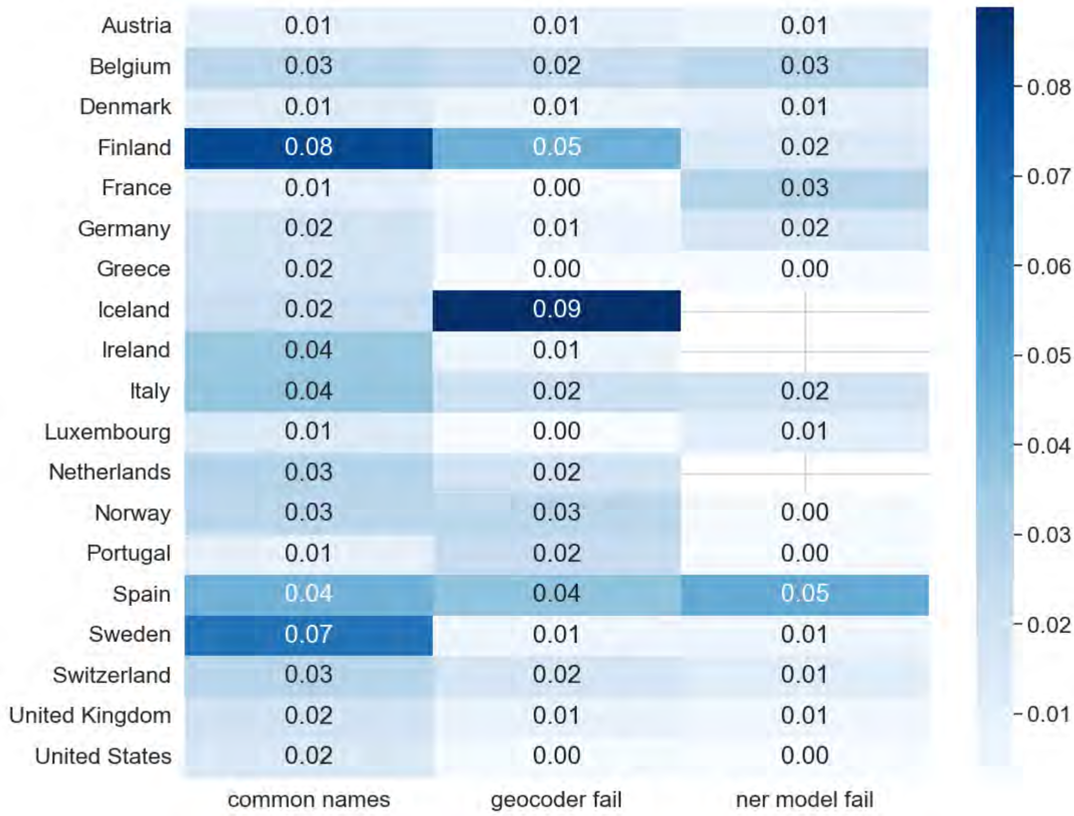


Panel D: False positives and negatives



Note: The figure shows a decomposition of the pipeline errors in different classifications; by recipient macro region, by donor, by type, and by type and category. We measure the error rate for projects over a random sample of 10,000 projects that we have hand-coded as the share of projects presenting at least one mistake in the geocoding output.

Figure B.7 – Donor-error heatmap



Note: The figure shows the error rate matrix for donors and error type. Cells show error rates, computed as the share of projects of a donor with at least one error, over the total number of donor projects. Darker colors indicate a greater error rate.

B.4 Comparison of geocoded projects with CRS data

In order to get a better understanding of in how far the around 15% of the projects we were able to geocode are a representative sample of the full CRS database, we compare project characteristics of the GODAD projects to non-geocoded projects reported in the CRS. As outlined above, certain criteria have to be fulfilled in order to be able to extract a geographic entity from the data. A key criteria is the including a location in the project title or its description that might depend on various factors, for instance on the type of project (an infrastructure project might be more location-specific compared to a finance project) or on the reporting quality of the donor and could restrict the representativeness of the geocoded sample.

In Table B.4 we correlate several project characteristics to an indicator variable that identifies the projects with geocodes (i.e., the dependent variable is an indicator for whether or not a project is geocoded, in a sample of all CRS projects). We consider five project characteristics: the year in which the project was reported, project size in terms of commitments, the donor country, the sector in which the project can be categorized, type of financial flow, and rough indicators for reporting quality. Starting with the regression specification in column 1 that includes recipient fixed effects to account for any time-invariant characteristic at the country level that might influence the likelihood of a project to be geocoded, e.g., the difficulty of the spelling of location names, we gradually add additional fixed effects and controls. Referring to column 1, the results show that projects reported in more recent years are more likely to be geocoded, suggesting a more precise, i.e., more location-specific reporting of projects over time. We find no statistically significant difference with respect to project size. The likelihood of a project to be geocoded, however, depends on the donor country, its sector, and type of financial flow. Relative to the U.S., the largest donor country, projects donated by most European countries are on average more likely to be geocoded except for projects from Denmark, Greece, Ireland, Sweden, and Switzerland. Projects from France are less likely to be geocoded. Relative to projects in the education sector, our baseline, infrastructure, health, governmental, agricultural, and other projects as well as emergency aid are more likely to be geocoded and projects in the sector "business, finance, and services" are less likely. These differences can be explained by the degree of location-dependence of a certain sector. For instance, infrastructure projects might be more likely to be linked to a specific location compared to finance projects that are often distributed throughout the country along other dimensions. Lastly, equity investments and other official flows are less and ODA loans more likely to be geocoded compared to ODA grants, the most frequent financial flow type. From column 2 to column 5 we include additional fixed effects and interactions among project characteristics to account for their interrelation. Specifically, we replace recipient fixed effects with recipient-year fixed effects in column 2, which does not change much of the results, and include sector-year fixed effects on top in column 3, absorbing some of the differences with respect to donors. We include a full set of donor-sector and donor-flow interactions in the regression models shown in columns 4 and 5. Once controlled for these factors, we find no statistically significant difference in the likelihood to be geocoded between ODA loans and ODA grants, and between Germany, Luxembourg, Netherlands, Portugal, and the UK relative to the United States. Lastly, in column 6 we account for differences in the overall reporting quality with the inclusion of two indicator variables that capture how many of the three columns (project title, short and long project description) considered for the geocoding process are filled-in in the CRS data. As expected projects that are reported more extensively are more likely to be geocoded. The reporting quality can explain the difference between ODA grants and other official flows. Differences in the likelihood to be geocoded remain by donor countries.

In summary, more recent projects and projects in more location-dependent sectors are over-represented in the GODAD; projects from some donor countries are more likely to be included in the GODAD compared to others. To the contrary, the GODAD represents the CRS data well with respect to project size. These insights need to be taken into account in the interpretation of statistical results using the GODAD. Given the lack of representativeness in some dimensions, results of statistical analyses do not necessarily hold for all aid and are more likely to do so for some donors than others.

Table B.4 – Factors correlated with geocoded projects

	Geo-coded project					
	(1)	(2)	(3)	(4)	(5)	(6)
Year	0.009***					
	(18.946)					
Commitment in USD, ln	0.000	0.001	0.001	0.004	0.004	0.003
	(0.066)	(0.363)	(0.405)	(1.353)	(1.216)	(0.906)
Donor						
Austria	0.134***	0.140***	0.141***	0.017	0.016	0.021
	(6.385)	(7.488)	(7.507)	(0.713)	(0.689)	(0.889)
Belgium	0.132***	0.141***	0.145***	0.102***	0.102***	0.124***
	(7.609)	(8.843)	(9.185)	(5.241)	(5.281)	(6.389)
Denmark	0.005	0.009	0.011	0.014	0.014	0.023
	(0.386)	(0.720)	(0.894)	(0.481)	(0.473)	(0.790)
Finland	0.262***	0.266***	0.268***	0.330***	0.331***	0.332***
	(14.887)	(16.789)	(16.829)	(12.471)	(12.523)	(12.618)
France	-0.023**	-0.018*	-0.017*	-0.065***	-0.069***	-0.055***
	(-2.033)	(-1.958)	(-1.791)	(-4.120)	(-4.481)	(-3.616)
Germany	0.110***	0.116***	0.116***	-0.006	-0.006	-0.002
	(6.926)	(8.385)	(8.391)	(-0.316)	(-0.308)	(-0.086)
Greece	-0.018	-0.013	-0.013	-0.064***	-0.065***	-0.073***
	(-1.041)	(-0.965)	(-0.913)	(-2.912)	(-2.957)	(-3.412)
Iceland	0.512***	0.516***	0.516***	0.581***	0.582***	0.586***
	(7.208)	(7.569)	(7.524)	(12.387)	(12.356)	(12.478)
Ireland	0.050	0.030	0.032	-0.009	-0.009	0.004
	(1.306)	(1.198)	(1.306)	(-0.458)	(-0.451)	(0.184)
Italy	0.232***	0.239***	0.240***	0.132***	0.132***	0.135***
	(12.693)	(14.912)	(15.057)	(6.780)	(6.768)	(6.927)
Luxembourg	0.048***	0.051***	0.053***	0.024	0.023	0.059**
	(2.967)	(3.460)	(3.558)	(0.942)	(0.930)	(2.309)
Netherlands	0.050***	0.054***	0.055***	0.034	0.035	0.026
	(3.512)	(4.243)	(4.305)	(1.451)	(1.484)	(1.105)
Norway	0.081***	0.086***	0.088***	0.082***	0.082***	0.083***
	(5.432)	(6.615)	(6.742)	(3.529)	(3.527)	(3.567)
Portugal	0.082***	0.095***	0.097***	-0.006	-0.007	-0.008
	(3.526)	(4.170)	(4.289)	(-0.234)	(-0.277)	(-0.343)
Sweden	-0.013	-0.007	-0.006	-0.073***	-0.074***	-0.067***
	(-0.964)	(-0.665)	(-0.497)	(-3.821)	(-3.844)	(-3.456)
Switzerland	0.018	0.024**	0.025**	-0.056***	-0.057***	-0.035*
	(1.384)	(2.091)	(2.126)	(-2.981)	(-3.045)	(-1.894)
United Kingdom	0.033*	0.034**	0.036**	-0.039	-0.039	-0.044
	(1.950)	(2.336)	(2.408)	(-1.229)	(-1.216)	(-1.370)
Sector						
Health	0.117***	0.115***				
	(17.393)	(17.741)				
Government	0.047***	0.047***				
	(5.945)	(6.016)				
Infrastructure	0.072***	0.068***				
	(13.111)	(13.515)				
Business, finance and services	-0.018**	-0.018**				
	(-2.497)	(-2.525)				
Agriculture	0.166***	0.163***				
	(16.136)	(16.274)				
Industry	0.018**	0.019**				
	(2.110)	(2.342)				
Food and commodity assistance	0.016	0.025**				
	(1.380)	(2.172)				
Emergencies	0.044***	0.050***				
	(4.079)	(5.410)				
Other	0.064***	0.062***				
	(7.379)	(7.742)				
Flow						
Equity Investment	-0.168***	-0.170***	-0.168***	-0.154***	-0.146***	-0.125***
	(-6.460)	(-6.313)	(-6.171)	(-5.672)	(-4.194)	(-3.640)
ODA Loans	0.060**	0.056**	0.057**	0.059**	0.013	0.033
	(2.180)	(2.050)	(2.109)	(2.256)	(0.553)	(1.436)
Other Official Flows (non Export Credit)	-0.198***	-0.184***	-0.171***	-0.166***	-0.130*	-0.120
	(-6.234)	(-5.756)	(-5.417)	(-5.516)	(-1.753)	(-1.631)
Reporting quality						
Two columns filled-in						0.052***
						(7.454)
All columns filled-in						0.145***
						(15.005)
Recipient FE	Yes					
Recipient-Year FE		Yes	Yes	Yes	Yes	Yes
Sector-year FE			Yes	Yes	Yes	Yes
Donor-sector Interactions				Yes	Yes	Yes
Donor-flow interaction					Yes	Yes
Observations	1,167,602	1,167,472	1,167,472	1,167,472	1,167,472	1,167,472
R^2	0.151	0.172	0.175	0.193	0.197	0.200

Note: The dependent variable is a binary indicator that is one if a project is geocoded and zero otherwise, using project-year level data. Regressions include year and country fixed effects as well as continent-specific time trends. Reference categories: Donor=US; Sector=Education; Flow: ODA Grants. Reporting quality refers to the reporting of project title, short- and long description that was used for the geo-coding process. The most frequent category was chosen for the base level. Fixed effects included as indicated in the table. Standard errors clustered at the country level.
t-statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01.

Based on these observations, it is important to discuss the extent to which differences in project characteristics between the GODAD and CRS database could bias our results related to birth regions of spouses and leaders. One possibility could be that the project location is more likely to be reported (highlighted) in the project title or its description when the project is located in the birth region of the current leader or spouse. Consequently, projects located in those regions are more likely to be geocoded and over-represented among the GODAD projects. If this is the case, we could falsely assume that more aid is distributed to the birth regions of leaders or spouses during their time in office, while in reality the increase in project size is driven by the over-representation of projects in those regions in our dataset.

First of all, note that the results in Table B.4 show no evidence for a difference in project size between GODAD and CRS data on average. Second, with the inclusion of region and country-year fixed effects in all our regressions we account for differences in several project characteristics that influence the likelihood of a project to be geocoded. While country-year fixed effects absorb improvements of reporting quality over time and changes in the donor-recipient composition, region fixed effects account for the average propensity of projects to be geocodable per region that might be related to the number of projects received, their sector, or difficulty of location names, among others.

Lastly, we can test whether a larger number of projects is geocoded at times of greater political scrutiny compared to other times. If that would be the case, our estimated positive effect of *Spouseregion* on aid might be spurious. Ideally, we would want to observe the share of geocoded projects at the regional level and compare those of spouses' birth regions during their "tenure" relative to other times. However, we are (by definition) unable to observe the number of non-geocodable projects at such a geographical level. Therefore, we aggregate the data to a country-year panel and analyze whether the share of geocoded projects in all projects varies relative to the year of election or—independent of how leaders assumed office—relative to the year the leader or spouse is "taking office." Arguably, political scrutiny is most severe at election time. If the share of geocoded projects would increase in election years, this would cast doubts on the causal interpretation of leader and spousal birth regions above.

Table B.5 presents the results. In column 1, we regress an indicator variable identifying the years of legislative elections and several lags and leads on the share of geocoded projects. In column 2 we consider executive elections and columns 3 and 4 replace the year of election with the year in which a leader or spouse enters office (independent of whether or not they were elected). All regressions include country and year fixed effects and control for continent-specific time trends.

According to our results, the share of geocoded projects in the year of election is lower rather than higher in election years and years the leaders' or their spouses' tenure start (with two of the four coefficients being significant at conventional levels).[69] Potentially, administrations are more occupied with other duties than the exact reporting of aid projects at election time or violence related to taking office exacerbates the gathering of information. Either way, there is no evidence that geocoded projects are more likely to be found in our data at times of heightened political scrutiny. While we cannot directly test whether the same is true in *Spouseregions* this result suggests that the positive effects on aid we report above rather represent a lower bound. What is more, recall that we do not find differences in the average project size between the CRS database and GODAD. Given that large (geocodable) projects are likely to receive scrutiny at any time, and small projects might be more likely to be scrutinized at times of political salience, it seems unlikely that our results can be explained by a larger share of geocoded projects rather than an increase in the number of projects.

Table B.5 – Share of geo-coded projects relative to the timing of elections and inauguration

Dependent:	Legislative election	Executive election	Leader taking office	Spouse taking office
Share of geocoded projects	(1)	(2)	(3)	(4)
5 years before	0.003	-0.001	0.001	0.007*
	(0.752)	(-0.196)	(0.231)	(1.812)
4 years before	0.002	-0.001	-0.001	0.003
	(0.344)	(-0.068)	(-0.202)	(0.562)
3 years before	0.002	0.004	-0.005	-0.003
	(0.275)	(0.373)	(-0.911)	(-0.612)
2 years before	-0.005	0.000	-0.002	0.000
	(-0.573)	(0.010)	(-0.381)	(0.003)
1 year before	-0.011	-0.001	-0.005	-0.008
	(-1.275)	(-0.049)	(-0.903)	(-1.594)
Year of election/office start	-0.012	-0.003	-0.008*	-0.010**
	(-1.612)	(-0.292)	(-1.727)	(-2.118)
1 year after	-0.015**	-0.004	-0.003	-0.003
	(-2.044)	(-0.322)	(-0.531)	(-0.680)
2 year after	-0.015**	-0.009	-0.003	-0.006
	(-2.385)	(-0.946)	(-0.500)	(-1.539)
3 year after	-0.013**	-0.010	-0.004	-0.005
	(-2.356)	(-1.349)	(-0.892)	(-1.350)
4 year after	-0.006	-0.010	0.001	-0.001
	(-1.278)	(-1.593)	(0.186)	(-0.195)
5 year after	-0.006	-0.008	0.003	0.001
	(-1.608)	(-1.582)	(0.643)	(0.217)
Country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Continent time trend	Yes	Yes	Yes	Yes
Observations	2,222	2,222	2,249	2,249
R^2	0.725	0.724	0.723	0.724

Note: The dependent variable is the share of geo-coded projects in all projects, in a country-year panel. Regressions include year and country fixed effects as well as continent-specific time trends. Standard errors are clustered at the country level.
t-statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01.

B.5 Comparison with existing geocoded data

The GODAD represents a significant milestone as the first comprehensive cross-country compilation of geocoded aid data, spanning across 175 recipient countries and 21 donors. To gauge the enhancements this dataset brings compared to existing data sources, this

[69] For legislative elections this correlation persists until the third year after election.

section compares our data to those reported in Uganda's Aid Management Platform. We take these data from AidData's Uganda AIMS Geocoded Research Release (AidData 2016), which includes more than US$12 billion in aid commitments for 565 geocoded projects in the 1978-2014 period.[70] Among the geocoded projects, 120 pertain to budget support and are not included in the following statistics.[71] Considering the donors we include in the GODAD, the Uganda AIMS geocoded dataset comprises 224 projects. We show this number in column 1 of Table B.6, where the additional rows report project numbers for individual donors. Column 2 shows the number of AIMS that can be merged to the CRS raw data and, for comparison, column 3 reports the number of projects in CRS. Not all projects included in AIMS projects can be merged to CRS projects.[72] In total, for the Western bilateral donors in the CRS sample, AIMS includes 197 projects, 144 of which are also included in the CRS.[73]

Column 4 shows data from the GODAD. As can be seen, when compared to AIMS the GODAD contains a substantially higher number of aid projects in Uganda (5,970 in total). When we restrict the sample period to those years that are covered in AIMS (in column 5), the GODAD still includes 3,777 projects, compared to the 224 projects (197 from Western donors plus 27 from China) included in AIMS (and 21,257 projects in the CRS, see column 6.). By definition, all (Western bilateral) GODAD projects are also included in the CRS, exceeding the number of projects reported in AIMS by an order of magnitude. Overall, the GODAD thus improves substantially on previously available data, even for the few recipient countries such data exist at all.[74]

References

AidData (2016). Uganda AIMS Geocoded Research Release Level 1 v1.4.1. Williamsburg, VA and Washington, DC: AidData.

Blair, R. A., R. Marty, and P. Roessler (2022). Foreign Aid and Soft Power: Great Power Competition in Africa in the Early Twenty-first Century. *British Journal of Political Science 52*(3), 1355–1376.

Briggs, R. C. (2019). Receiving Foreign Aid Can Reduce Support for Incumbent Presidents. *Political Research Quarterly 72*(3), 610–622.

[70] We focus on Uganda because these data have been used in a number of research articles (e.g., Briggs 2019, Blair et al. 2022) and are among the most comprehensive single-recipient datasets.

[71] AIMS codes 113 of these projects at the level of the country (i.e., their designated location is "Uganda"), so that they cannot be used for within-country analyses.

[72] The two datasets lack uniform project identifiers and may exhibit disparities in project titles attributable to the use of acronyms or different syntax. Consequently, we identified projects that appear in both datasets through keyword analysis within project titles and descriptions, start and end dates, as well as commitment volumes.

[73] AIMS includes 26 projects that are supported by multiple contributors that therefore cannot be attributed to single donors. An example is the following combination of donors: Denmark, United Kingdom, Austria, Netherlands, Ireland, as well as Denmark, European Union, United Nations Development Programme, Sweden, Austria, Ireland, Belgium, and Norway.

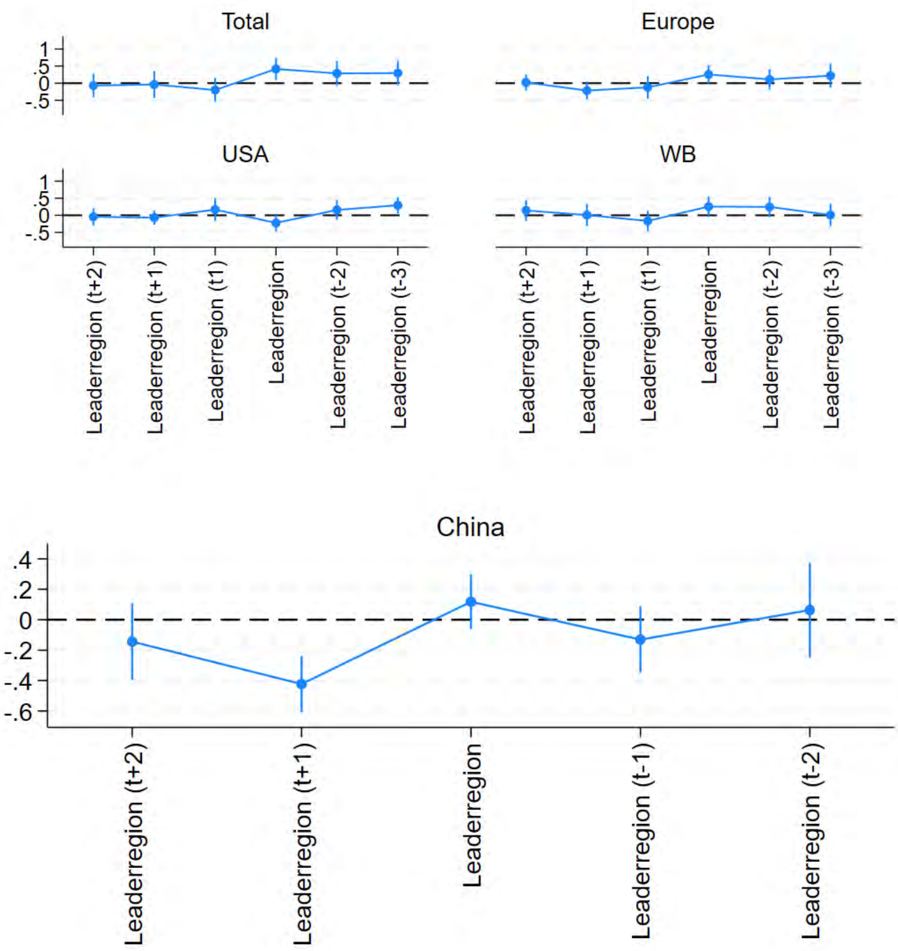
[74] Recall that the final data that we aim to report in GODAD will also include information from AIMS. This implies that the total number of projects we include for Uganda will further increase.

Table B.6 – Number of projects by donors, AIMS Uganda, CRS, GODAD

	(1) AIMS Uganda	(2) Merged projects	(3) CRS	(4) GODAD	(5) GODAD (1978-2014)	(6) CRS (1978-2014)
Total	224	144	30988	5970	3777	21257
Austria	35	25	1493	416	266	927
Belgium	4	3	1079	223	138	786
Denmark	31	18	1105	116	91	888
Finland	-	-	469	156	87	262
France	-	-	625	88	26	376
Germany	5	3	2404	679	368	1625
Greece	-	-	40	1	1	34
Iceland	1	-	71	50	15	23
Ireland	17	14	3877	551	409	3051
Italy	-	-	1254	509	361	966
Luxembourg	-	-	85	2	2	31
Netherlands	-	-	1028	119	112	971
Norway	60	45	2450	328	281	1917
Portugal	-	-	2	-	-	-
Spain	1	-	381	150	103	267
Sweden	11	9	2449	248	138	1741
Switzerland	-	-	417	33	11	180
United Kingdom	12	10	1881	236	145	1188
United States	20	17	9878	799	319	6024
World Bank	-	-	-	1122	796	-
China	27	-	-	144	108	-

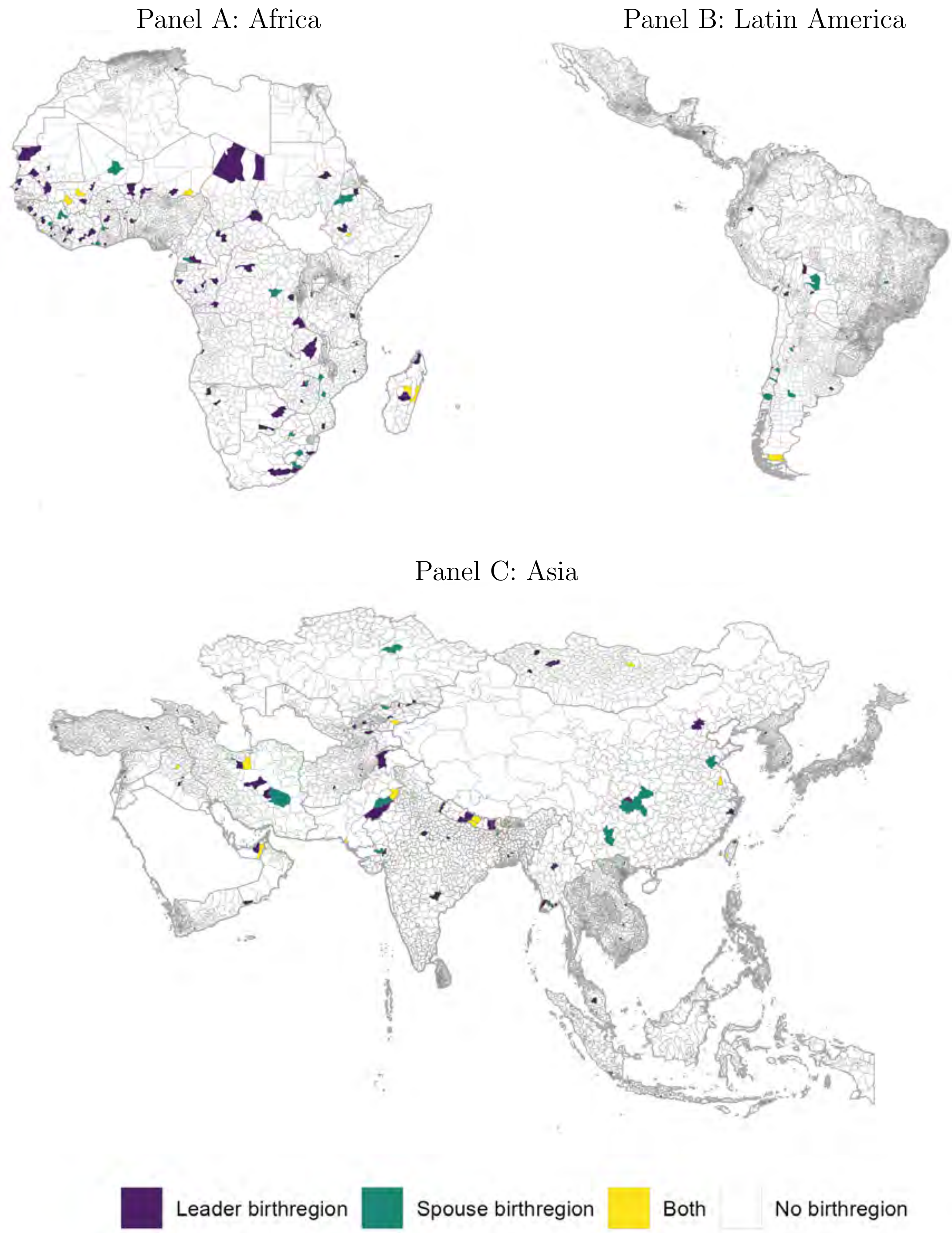
C Additional Tables and Figures

Figure C.1 – Effect of Leader Birthregions on Aid, ADM2



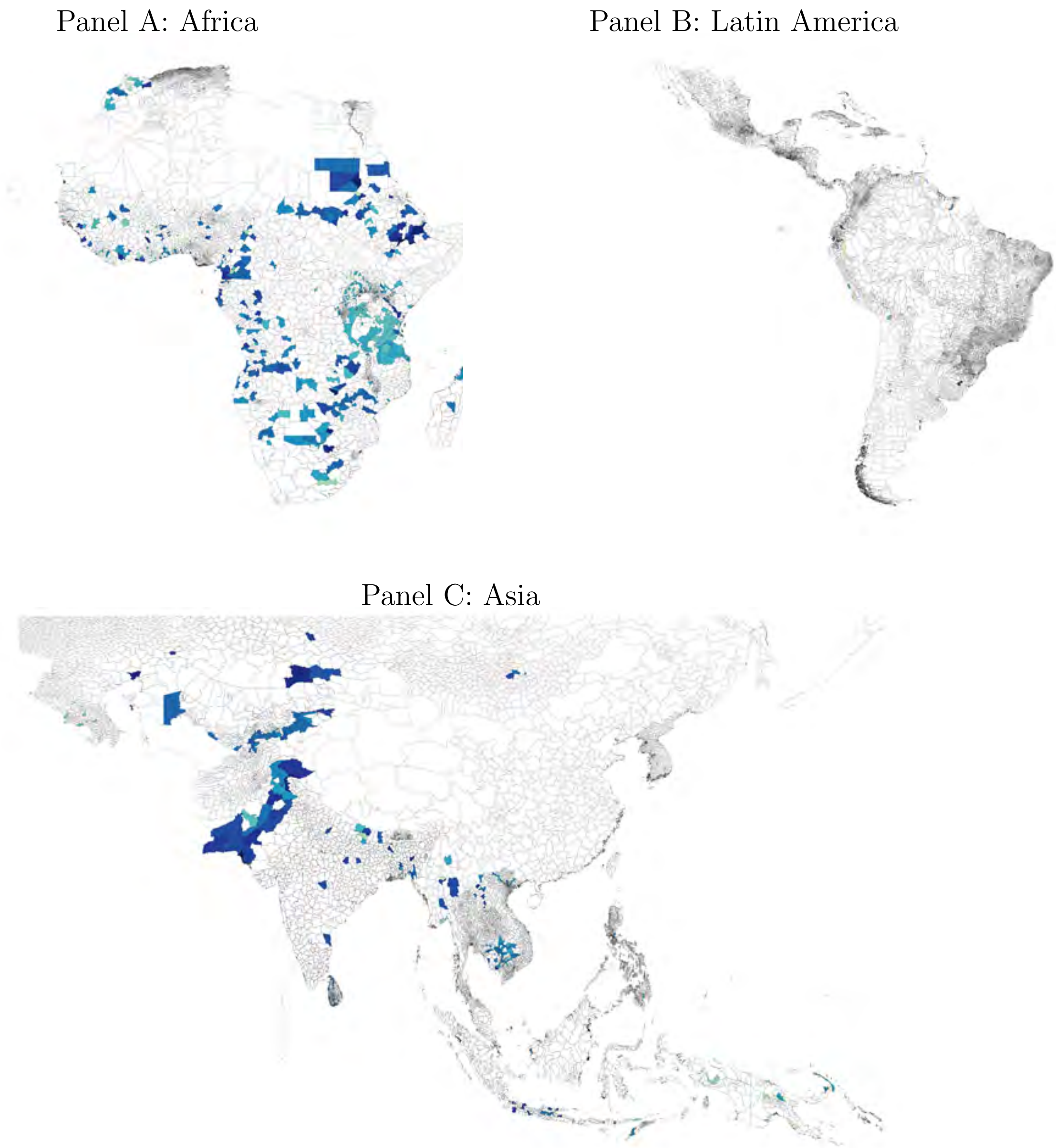
Note: The figure plots the coefficients of *Leaderregion* in concert with its 'leads' and 'lags' and 90% confidence intervals, corresponding to column 4 of Table 1 and columns 2, 4, 6, and 8 of Table 4.

Figure C.2 – Leader and Spouse Birthregion, ADM2, 1990–2020



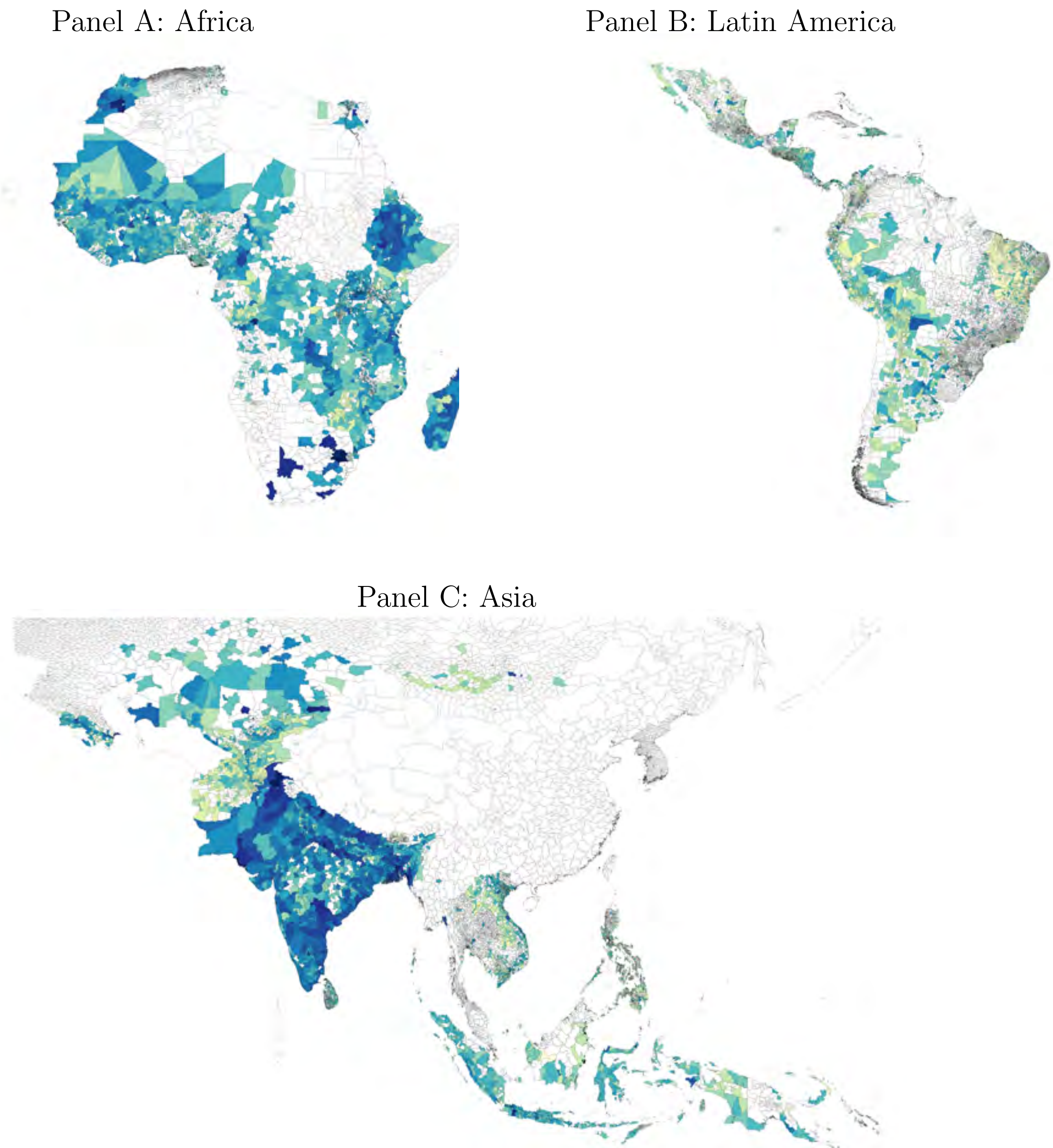
Note: The maps indicate whether an ADM2 region has been a leader birth region (in purple), spouse birth region (in green), both (in yellow), or none (in white) over the 1990–2020 period.

Figure C.3 – Chinese aid by macro region, ADM2, 1990–2020



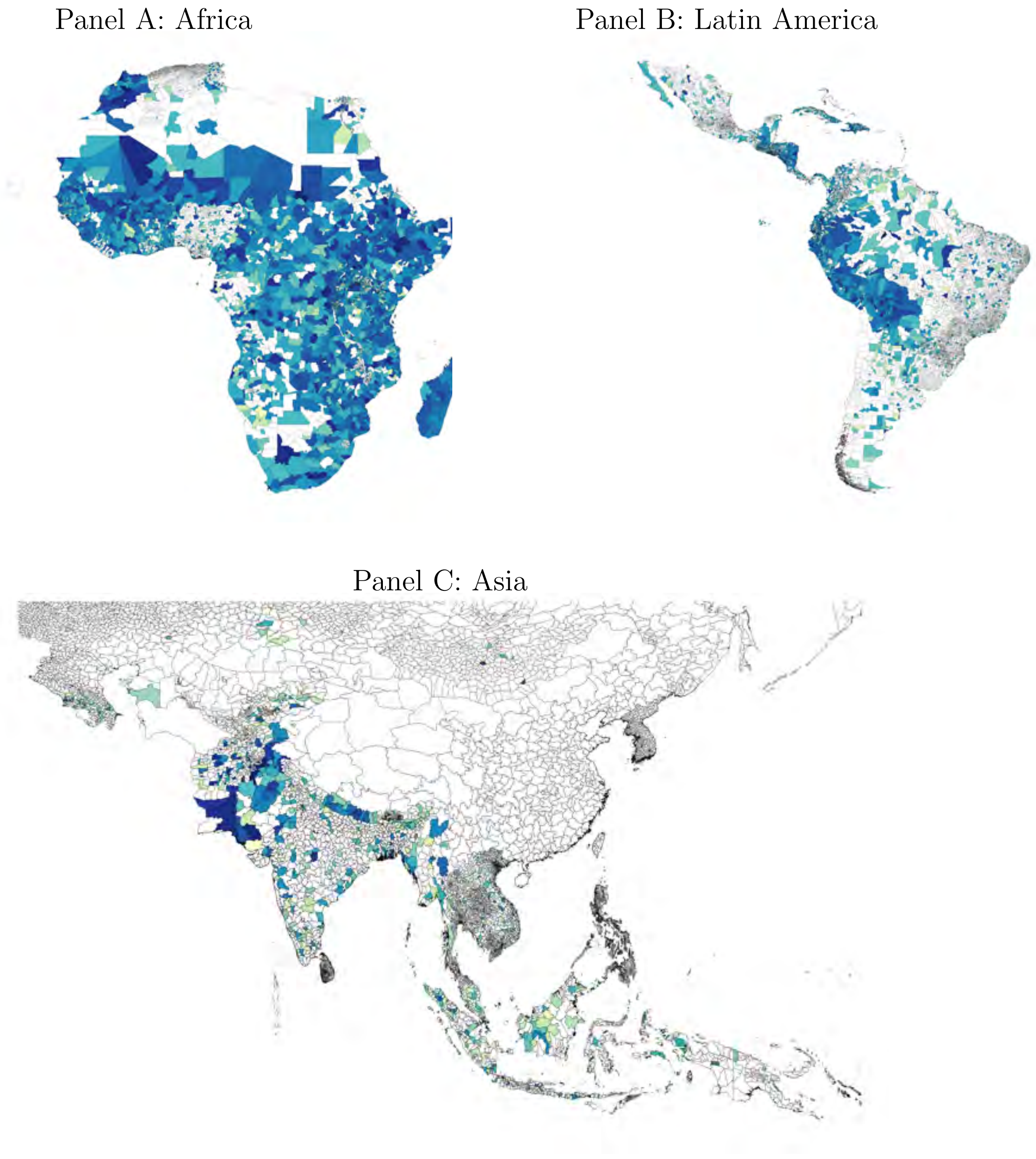
Note: The maps indicate the amount of log Chinese aid per ADM2 region over the 1990–2020 period.

Figure C.4 – World Bank aid by macro region, ADM2, 1990–2020



Note: The maps indicate the amount of log World Bank aid per ADM2 region over the 1990–2020 period.

Figure C.5 – EU aid by macro region, ADM2, 1990–2020



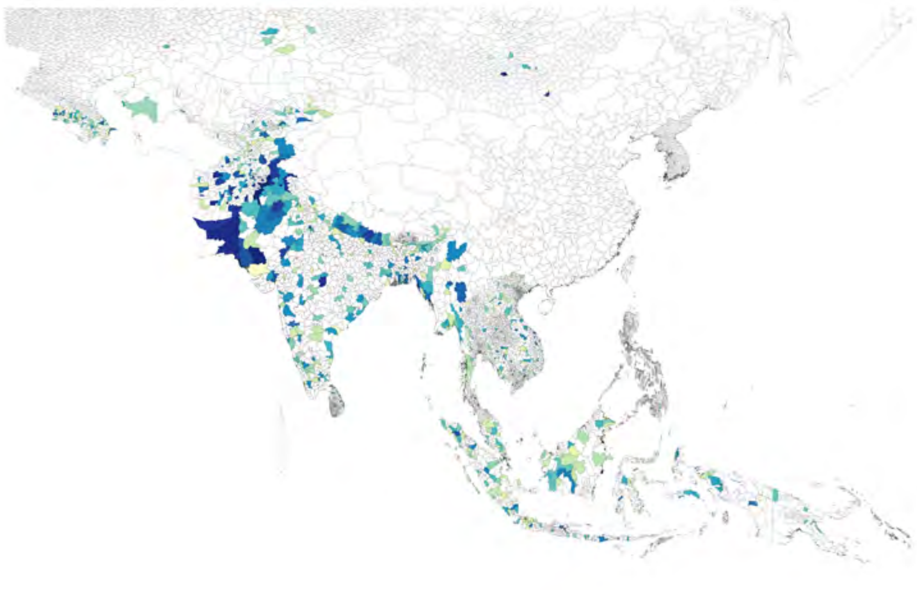
Note: The maps indicate the amount of log aid of 18 European donors per ADM2 region over the 1990–2020 period.

Figure C.6 – US aid by macro region, ADM2, 1990–2020

Panel A: Africa

Panel B: Latin America

Panel C: Asia



Note: The maps indicate the amount of log US aid per ADM2 region over the 1990–2020 period.

Table C.1 – Birth Regions and Total Aid with Control Variables, 1995-2020

	(1) ADM2	(2) ADM2	(3) ADM2	(4) ADM2	(5) ADM2	(6) ADM2	(7) ADM2	(8) ADM2	(9) ADM1	(10) ADM1
Spouseregion (b+2)		1.00**		0.42	0.43	0.25	0.29	0.26		0.11
		(2.00)		(0.97)	(0.98)	(0.40)	(0.69)	(0.50)		(0.39)
Spouseregion (b+1)		0.70		0.14	0.12	-0.34	-0.36	-0.61		-0.19
		(1.51)		(0.31)	(0.26)	(0.63)	(0.84)	(1.05)		(0.69)
Spouseregion (b1)		1.07*		1.41***	1.41***	0.78	1.49***	1.16**		0.28
		(1.89)		(3.00)	(3.00)	(1.35)	(2.98)	(2.14)		(0.86)
Spouseregion	1.99***	1.96***	1.66***	1.62***		0.48	1.42***	1.14***	0.62**	0.62**
	(4.64)	(4.51)	(3.97)	(3.85)		(0.90)	(3.77)	(2.94)	(2.41)	(2.36)
Spouseregion (b2)					1.13**					
					(2.31)					
Spouseregion (b3)					1.51***					
					(3.09)					
Spouseregion (b4)					2.40***					
					(4.62)					
Spouseregion (b5)					2.54***					
					(3.77)					
Spouseregion (b6)					2.41***					
					(3.59)					
Spouseregion (other year)					1.39**					
					(2.07)					
Spouseregion (b-1)					1.35***					
					(3.05)					
Spouseregion (b-2)		1.36***		0.97**	0.97**	-0.54	1.26**	1.30**		-0.01
		(2.63)		(2.29)	(2.28)	(0.83)	(2.62)	(2.03)		(0.03)
Spouseregion (b-3)		0.49		0.75*	0.77*	-0.87	0.78*	0.74		0.07
		(1.07)		(1.73)	(1.76)	(1.44)	(1.86)	(1.24)		(0.27)
Leaderregion	0.84***	0.78**	0.87***	1.02***	1.04***	1.57*	1.42***	1.89***	0.03	-0.09
	(3.43)	(2.39)	(3.57)	(3.59)	(3.61)	(1.90)	(2.75)	(3.04)	(0.16)	(0.45)
(log) Nightlight	0.06***	0.06***	0.06***	0.06***	0.06***	-0.40***	0.00	0.01	0.13***	0.13***
	(3.54)	(3.51)	(3.54)	(3.51)	(3.52)	(2.68)	(0.03)	(0.26)	(4.23)	(4.33)
(log) Population	0.27***	0.27***	0.27***	0.27***	0.27***	1.27***	0.32***	0.50***	0.63***	0.62***
	(5.78)	(5.73)	(5.79)	(5.73)	(5.62)	(4.29)	(5.75)	(7.88)	(5.12)	(5.04)
Capital city	4.40***	4.35***	4.43***	4.38***	4.37***	3.81***	3.66***	3.68***	2.80***	2.80***
	(12.75)	(12.65)	(12.81)	(12.76)	(12.67)	(3.84)	(5.66)	(6.57)	(11.09)	(11.01)
Mine	0.22***	0.22***	0.22***	0.22***	0.22***	-1.20	0.20*	0.19	0.26**	0.26**
	(3.04)	(3.00)	(3.04)	(3.00)	(3.00)	(1.16)	(1.71)	(0.97)	(2.17)	(2.19)
Oil/gas	-0.02	-0.02	-0.02	-0.02	-0.03	-0.76	0.01	0.05	0.03	0.03
	(0.35)	(0.34)	(0.35)	(0.35)	(0.44)	(0.85)	(0.06)	(0.25)	(0.24)	(0.22)
(log) Area	-0.00	-0.00	-0.00	-0.00	-0.00	0.61**	-0.01	-0.02	0.14**	0.14**
	(0.02)	(0.00)	(0.02)	(0.01)	(0.02)	(2.54)	(0.30)	(0.35)	(2.29)	(2.26)
Port	0.65***	0.65***	0.65***	0.65***	0.65***	-0.96	0.46**	0.73**	0.44***	0.44***
	(4.81)	(4.76)	(4.81)	(4.77)	(4.78)	(0.95)	(2.00)	(2.45)	(3.43)	(3.43)
Road density	0.16	0.17	0.16	0.17	0.17	3.90***	0.48*	0.66*	-0.06	-0.07
	(1.25)	(1.26)	(1.24)	(1.25)	(1.25)	(2.72)	(1.85)	(1.70)	(0.63)	(0.72)
Country-year FE	no	no	yes	yes	yes	yes	yes	yes	yes	yes
Number of countries	122	122	122	122	122	95	90	79	161	161
Number of regions	34100	34100	34100	34100	34100	190	3836	1390	2981	2981
Number of observations	778253	769651	778253	769651	773056	4209	80365	34133	66630	66052
Prob > F Spouse		0.01		0.00		0.11	0.00	0.00		0.00
Prob > F Leader		0.09		0.05		0.47	0.18	0.13		0.41
R squared (within)	0.0165	0.0168	0.0272	0.0274	0.0274	0.2021	0.0615	0.0632	0.0752	0.0754

Note: The dependent variable is *Aid*, i.e., the logarithm of aid (plus 1) given to region i of country c in year t: ODA disbursements of 18 European donors, IDA, and the U.S. and ODA commitments from China (which we set to zero for the years 1995-1999 and 2018-2020). *Spouseregion* and *Leaderregion* are lagged by one year. *Spouseregion (bx)* indicates the xth year into being a birth region. *Prob > F Spouse/Leader* tests whether the coefficients of *Spouseregion/Leaderregion* are larger than those of $Spouseregion_{b+1}/Leaderregion_{b+1}$. Except columns 1, 3, and 9, all regressions control for $Leaderregion_{b+2}$, $Leaderregion_{b+1}$, $Leaderregion_{b1}$, $Leaderregion_{b-2}$, $Leaderregion_{b-3}$. $(log)Nightlight$ is defined as the log of mean nightlight emissions in region i of country c in year t (+0.01). $(log)Population$ is the logarithm of region i's population. $Capitalcity$ is a dummy variable taking the value 1 if region i contains the capital city of country c. $Mine$ is a dummy variable taking the value 1 if region i contains an active mine. Oil/gas is a dummy variable taking the value 1 if region i contains a petroleum field. $(log)Area$ is the log of the size of region i in km^2. $Port$ takes the value 1 if the region contains at least one port. Finally, $Roaddensity$ is the road density in region i, in km/km^2. Columns 1–2 include country fixed effects. Columns 3–10 include country-year fixed effects (FE). Standard errors are clustered at the country level.
t-statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01.

Table C.2 – Birth Regions and Total Aid with Clustering Options, 1995-2020

	(1) ADM2	(2) ADM2	(3) c-y	(4) c-y	(5) HAC	(6) HAC	(7) DK	(8) DK
Spouseregion (b+2)		-0.21		-0.21		-0.21		-0.21
		(0.66)		(0.65)		(0.64)		(0.49)
Spouseregion (b+1)		-0.33		-0.33		-0.33		-0.33
		(0.97)		(0.96)		(0.96)		(0.90)
Spouseregion (b1)		0.36		0.36		0.36		0.36
		(0.82)		(0.83)		(0.83)		(0.59)
Spouseregion	0.84***	0.83**	0.84***	0.83***	0.84***	0.83***	0.84***	0.83**
	(2.88)	(2.52)	(4.78)	(4.30)	(3.99)	(3.58)	(2.95)	(2.26)
Spouseregion (b-2)		0.02		0.02		0.02		0.02
		(0.07)		(0.07)		(0.07)		(0.07)
Spouseregion (b-3)		0.06		0.06		0.06		0.06
		(0.14)		(0.15)		(0.15)		(0.17)
Leaderregion	0.25*	0.42**	0.25**	0.42**	0.25**	0.42**	0.25	0.42
	(1.68)	(1.99)	(2.29)	(2.09)	(2.05)	(2.06)	(1.60)	(1.34)
Number of countries	141	141	141	141	141	141	141	141
Number of regions	44506	44506	44506	44506	44506	44506	44506	44506
Number of observations	854418	842702	854418	842702	854418	842702	854418	842702
Prob > F Spouse		0.00		0.00		0.00		0.00
Prob > F Leader		0.07		0.06		0.06		0.10
R squared (within)	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003

Note: The dependent variable is *Aid*, i.e., the logarithm of aid (plus 1) given to region i of country c in year t: ODA disbursements of 18 European donors, IDA, and the U.S. and ODA commitments from China (which we set to zero for the years 1995-1999 and 2018-2020). *Spouseregion* and *Leaderregion* are lagged by one year. *Spouseregion (b1)* indicates the first year into being a birth region. *Prob > F Spouse/Leader* tests whether the coefficients of *Spouseregion/Leaderregion* are larger than those of $Spouseregion_{b+1}/Leaderregion_{b+1}$. Columns 2, 4, 6, and 8 control for $Leaderregion_{b+2}$, $Leaderregion_{b+1}$, $Leaderregion_{b1}$, $Leaderregion_{b-2}$, $Leaderregion_{b-3}$. Columns 1–2 include country fixed effects. Columns 3–10 include country-year fixed effects (FE). Standard errors are clustered at the level indicated at the column header: ADM2, country-year, Heteroskedasticity-consistent (HAC), and Driscoll-Kraay (DK).
t-statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01.

Table C.3 – Birth Regions and Aid by Sector, 1995-2020

	(1) Econ	(2) Econ	(3) Social	(4) Social	(5) Prod	(6) Prod
Spouseregion (b+2)		0.11		-0.21		0.29
		(0.39)		(0.72)		(1.01)
Spouseregion (b+1)		0.52*		-0.03		0.00
		(1.82)		(0.08)		(0.01)
Spouseregion (b1)		0.50		0.25		-0.37
		(1.37)		(0.63)		(1.10)
Spouseregion	0.90***	0.98***	0.72***	0.73**	0.45*	0.45*
	(3.75)	(3.72)	(2.83)	(2.50)	(1.95)	(1.70)
Spouseregion (b-2)		0.33		0.34		-0.04
		(0.95)		(0.85)		(0.11)
Spouseregion (b-3)		-0.00		-0.01		0.16
		(0.01)		(0.02)		(0.40)
Leaderregion	0.05	0.18	0.21	0.35*	-0.12	-0.07
	(0.40)	(0.90)	(1.38)	(1.74)	(0.89)	(0.32)
Number of countries	141	141	141	141	141	141
Number of regions	44506	44506	44506	44506	44506	44506
Number of observations	854418	842702	854418	842702	854418	842702
Prob > F Spouse		0.10		0.03		0.12
Prob > F Leader		0.09		0.02		0.28
R squared (within)	0.0001	0.0002	0.0002	0.0003	0.0001	0.0001

Note: The dependent variable is *Aid*, i.e., the logarithm of aid (plus 1) given to region i of country c in year t: ODA disbursements of 18 European donors, IDA, and the U.S. and ODA commitments from China (which we set to zero for the years 1995-1999 and 2018-2020). *Spouseregion* and *Leaderregion* are lagged by one year. *Spouseregion (b1)* indicates the first year into being a birth region. *Prob > F Spouse/Leader* tests whether the coefficients of $Spouseregion/Leaderregion$ are larger than those of $Spouseregion_{b+1}/Leaderregion_{b+1}$. Columns 2, 4, and 6 control for $Leaderregion_{b+2}$, $Leaderregion_{b+1}$, $Leaderregion_{b1}$, $Leaderregion_{b-2}$, and $Leaderregion_{b-3}$. Our sectoral definitions follow the OECD's DAC: Social Infrastructure & Services includes Education, Health, Population Pol./Progr. & Reproductive Health, Water Supply & Sanitation, Government & Civil Society, and Other Social Infrastructure & Services. Economic Infrastructure & Services includes Transport & Storage, Communications, Energy, Banking & Financial Services, and Business & Other Services. The Production Sector includes Agriculture, Forestry, Fishing, Industry, Mining, Construction, Trade Policies & Regulations, and Tourism. All regressions include ADM2 fixed effects, country-year fixed effects, and the logarithm of a region's population size. Standard errors are clustered at the country level.
t-statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01.

Table C.4 – Birth Regions and Total Aid, 1995-2020, Additional Interactions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
INTER:		Corr	Law	PRS	Democ	Control	Censor
Spouseregion	1.35**	1.86**	1.90**	1.74	1.09	0.84***	0.83**
	(2.51)	(2.58)	(2.47)	(1.13)	(1.55)	(2.98)	(2.50)
Leaderregion	0.03	0.50	0.31	0.53	0.88*	0.24*	0.43**
	(0.10)	(1.39)	(0.81)	(0.67)	(1.88)	(1.66)	(2.18)
Spouseregion*Africa	-0.88						
	(1.27)						
Spouseregion*Asia	-0.35						
	(0.47)						
Spouseregion*Europe	-1.37						
	(1.54)						
Spouseregion*Oceania	-1.34**						
	(2.50)						
Leaderregion*Africa	-0.03						
	(0.09)						
Leaderregion*Asia	0.84*						
	(1.74)						
Leaderregion*Europe	-0.07						
	(0.20)						
Leaderregion*Oceania	0.03						
	(0.09)						
Spouseregion*INTER		-0.44*	-0.34	-0.01	-0.50	0.05	-0.01
		(1.89)	(1.45)	(0.60)	(0.42)	(0.21)	(0.03)
Leaderregion*INTER		-0.12	-0.03	-0.01	-1.21*	0.17*	-0.24**
		(1.05)	(0.27)	(0.46)	(1.66)	(1.74)	(2.01)
Number of countries	141	119	119	119	141	141	141
Number of regions	44506	42502	42502	42502	44506	44506	44506
Number of observations	854418	804736	804736	804736	854418	854418	854418
R squared (within)	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003

Note: The dependent variable is *Aid*, i.e., the logarithm of aid (plus 1) given to region i of country c in year t: ODA disbursements of 18 European donors, IDA, and the U.S. and ODA commitments from China (which we set to zero for the years 1995-1999 and 2018-2020). *Spouseregion* and *Leaderregion* are lagged by one year. All regressions control for the logarithm of a region's population size. The omitted category in column 1 is the Americas. *Corr* measures Control of Corruption, *Law* the Rule of Law, *PRS* the Political Risk Score. These variables are taken from the PRS Group's International Country Risk Guids, ICRG. They range from 0–6, 0–6, and 0–100, with higher values indicating better governance. *Democ* measures to what extent the ideal of electoral democracy is achieved, on a score from 0–1. *Control* measures whether party control of the national government is unified, ranging from 0–2. *Censor* measures whether the government attempts to censor the media, on an ordinal 0–4 scale. These variables are taken from the Varieties of Democracy Project—higher values measure more democracy, unified control, and less censorship. All regressions include ADM2 fixed effects and country-year fixed effects. Standard errors are clustered at the country level.
t-statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01.

Table C.5 – Birth regions and aid, ADM2, individual European donors

	(1) AUT	(2) BEL	(3) DEN	(4) FIN	(5) FRA	(6) GER	(7) GRE	(8) ICE	(9) IRE	(10) ITA	(11) LUX	(12) NED	(13) NOR	(14) POR	(15) ESP	(16) SWE	(17) CHE	(18) UK
Spouseregion (b+2)	-0.002	0.145	0.048	0.222	0.287	-0.147	0.036	-0.011	0.040	0.003	-0.039	-0.108	-0.051	-0.039	-0.082	-0.094	-0.164	-0.354*
	(0.01)	(0.66)	(0.39)	(1.21)	(1.17)	(0.53)	(0.67)	(1.25)	(0.35)	(0.01)	(0.25)	(0.89)	(0.24)	(0.68)	(0.42)	(0.63)	(1.08)	(1.78)
Spouseregion (b+1)	-0.016	-0.042	0.061	0.170	0.070	-0.332	-0.053*	-0.034**	-0.086	0.095	0.006	-0.077	0.047	-0.040	0.020	-0.116	-0.231	0.017
	(0.10)	(0.23)	(0.52)	(1.16)	(0.29)	(1.11)	(1.91)	(2.25)	(1.12)	(0.40)	(0.04)	(0.60)	(0.22)	(0.74)	(0.09)	(0.64)	(1.28)	(0.10)
Spouseregion (b1)	-0.066	0.254	0.078	0.486**	0.268	0.124	-0.086	0.137	-0.048	0.518	-0.068	0.002	0.080	0.053	0.195	0.061	-0.047	-0.209
	(0.40)	(1.19)	(0.58)	(2.31)	(0.88)	(0.39)	(1.62)	(1.24)	(0.41)	(1.51)	(0.35)	(0.01)	(0.27)	(0.61)	(0.62)	(0.30)	(0.21)	(0.86)
Spouseregion	0.168	-0.019	0.134	0.191	0.264	0.930***	0.014	0.030	0.181**	0.239	0.188	0.123	0.242	0.039	0.428*	-0.078	0.347*	0.248
	(1.16)	(0.11)	(1.31)	(1.12)	(1.19)	(2.76)	(0.52)	(1.32)	(1.99)	(1.00)	(1.07)	(0.85)	(1.26)	(0.95)	(1.85)	(0.52)	(1.88)	(1.21)
Spouseregion (b-2)	0.072	-0.228	0.197	-0.026	-0.022	0.312	-0.060	0.059	0.003	-0.088	-0.080	-0.097	0.195	0.074	0.140	0.125	0.224	0.038
	(0.43)	(1.33)	(1.26)	(0.12)	(0.11)	(0.73)	(1.61)	(0.80)	(0.03)	(0.32)	(0.55)	(0.59)	(0.87)	(0.95)	(0.62)	(0.61)	(0.80)	(0.21)
Spouseregion (b-3)	-0.103	-0.109	0.095	-0.033	0.144	0.415	-0.077	0.080	-0.041	0.044	-0.085	-0.124	0.132	0.026	0.621***	-0.029	0.169	0.321
	(0.64)	(0.58)	(0.81)	(0.17)	(0.62)	(1.21)	(1.43)	(0.96)	(0.47)	(0.16)	(0.61)	(0.77)	(0.56)	(0.46)	(2.65)	(0.14)	(0.73)	(1.58)
Leaderregion	0.139	0.018	-0.010	-0.040	-0.038	-0.131	-0.050	-0.021*	-0.031	-0.216**	-0.118	0.011	0.065	0.102*	0.090	0.251**	-0.024	0.108
	(1.52)	(0.16)	(0.15)	(0.39)	(0.28)	(0.90)	(1.14)	(1.79)	(0.99)	(2.07)	(1.56)	(0.15)	(0.51)	(1.82)	(0.72)	(2.20)	(0.22)	(1.09)
First year	1990	1990	1990	1990	1990	1990	1990	1990	1990	1990	1990	1990	1990	1990	1990	1990	1990	1990
Last year	2020	2020	2020	2020	2020	2020	2020	2020	2020	2020	2020	2020	2020	2020	2020	2020	2020	2020
Number of countries	142	142	142	142	142	142	142	142	142	142	142	142	142	142	142	142	142	142
Number of regions	44799	44799	44799	44799	44799	44799	44799	44799	44799	44799	44799	44799	44799	44799	44799	44799	44799	44799
Number of observations	1009949	1009949	1009949	1009949	1009949	1009949	1009949	1009949	1009949	1009949	1009949	1009949	1009949	1009949	1009949	1009949	1009949	1009949
Prob > F Spouse	0.15	0.44	0.27	0.45	0.13	0.00	0.05	0.02	0.00	0.29	0.14	0.10	0.21	0.09	0.08	0.41	0.00	0.13
Prob > F Leader	0.10	0.03	0.21	0.42	0.35	0.25	0.09	0.04	0.09	0.03	0.21	0.37	0.33	0.19	0.22	0.11	0.42	0.31
R squared (within)	0.0002	0.0003	0.0003	0.0002	0.0003	0.0007	0.0002	0.0002	0.0002	0.0003	0.0004	0.0003	0.0004	0.0002	0.0004	0.0002	0.0004	0.0003

Note: The dependent variable is the logarithm of ODA disbursements from one of 18 European donors (plus 1) given to region i of country c in year t, 1990–2020: (1) Austria, (2) Belgium, (3) Denmark, (4) Finland, (5) France, (6) Germany, (7) Greece, (8) Ireland, (9) Iceland, (10) Italy, (11) Luxembourg, (12) Netherlands, (13) Norway, (14) Portugal, (15) Spain, (16) Sweden, (17) Switzerland, and (18) United Kingdom. *Spouseregion* and *Leaderregion* are lagged by one year. *Prob > F Spouse/Leader* tests whether the coefficients of *Spouseregion*/*Leaderregion* are larger than those of $Spouseregion_{b+1}/Leaderregion_{b+1}$. All regressions control for $Leaderregion_{b+2}$, $Leaderregion_{b+1}$, $Leaderregion_{b1}$, $Leaderregion_{b-2}$, $Leaderregion_{b-3}$, and the logarithm of a region's population size. All columns include ADM2 and country-year fixed effects. Standard errors are clustered at the country level.
t-statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01.

Table C.6 – Birth Regions and Aid, ADM2, Event Study, Alternative Definition of Aid

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Europe		USA		World Bank		China	
	Commitm	Dummy	Commitm	Dummy	Commitm	Dummy	Commitm	Dummy
Spouseregion (b+2)	-0.245	0.002	-0.509*	-0.050**	0.076	-0.009	0.343	0.035**
	(0.72)	(0.08)	(1.73)	(2.22)	(0.32)	(0.42)	(1.45)	(1.98)
Spouseregion (b+1)	-0.230	-0.017	0.172	0.009	-0.136	-0.057***	0.316	0.027*
	(0.65)	(0.65)	(0.65)	(0.39)	(0.58)	(2.74)	(1.53)	(1.67)
Spouseregion (b1)		0.018		0.018		-0.009		
		(0.58)		(0.68)		(0.36)		
Spouseregion	0.558*	0.055**	0.643**	0.079***	0.161	-0.003	0.232*	0.021*
	(1.81)	(2.25)	(2.07)	(2.79)	(0.99)	(0.21)	(1.86)	(1.87)
Spouseregion (b-1)	0.261		-0.019		0.141		-0.015	-0.007
	(0.80)		(0.07)		(0.68)		(0.12)	(0.52)
Spouseregion (b-2)	0.192	0.039	0.286	0.037	-0.364	-0.001	-0.162	0.014
	(0.51)	(1.22)	(1.04)	(1.53)	(1.23)	(0.02)	(0.95)	(0.83)
Spouseregion (b-3)		0.028		0.032		-0.002		
		(1.02)		(1.35)		(0.07)		
Leaderregion	0.098	0.011	-0.061	-0.005	0.017	0.010	0.118	0.006
	(0.79)	(0.78)	(0.40)	(0.40)	(0.18)	(0.71)	(1.09)	(0.80)
First year	1990	1990	1990	1990	1995	1995	2000	2000
Last year	2020	2020	2020	2020	2020	2020	2017	2017
Number of countries	142	142	142	142	141	141	137	137
Number of regions	44799	44799	44799	44799	44506	44506	40921	40921
Number of observations	1028291	1009949	1028291	1009949	854232	842702	568950	568950
Prob > F Spouse	0.02	0.01	0.06	0.00	0.11	0.01	0.32	0.35
Prob > F Leader	0.47	0.26	0.30	0.36	0.20	0.44	0.00	0.00
R squared (within)	0.0004	0.0006	0.0006	0.0009	0.0000	0.0000	0.0006	0.0007

Note: The dependent variables are *Aid*, i.e., the logarithm of aid (plus 1) given to region i of country c in year t: ODA commitments or a binary variable indicating at least one new project, of 18 European donors, IDA, the U.S. and China. *Spouseregion* and *Leaderregion* are lagged by one year in column 2, 4, and 6, which are based on disbursements. *Spouseregion (bx)* indicates the xth year into being a birth region. *Prob > F Spouse/Leader* tests whether the coefficients of *Spouseregion/Leaderregion* are larger than those of $Spouseregion_{b+1}/Leaderregion_{b+1}$. Columns 2, 4, and 6 control for $Leaderregion_{b+2}$, $Leaderregion_{b+1}$, $Leaderregion_{b1}$, $Leaderregion_{b-2}$, $Leaderregion_{b-3}$, all other regressions control for $Leaderregion_{b+2}$, $Leaderregion_{b+1}$, $Leaderregion_{b-1}$, and $Leaderregion_{b-2}$. All columns include the logarithm of a regions's population size, ADM2 fixed effects and country-year fixed effects. Standard errors are clustered at the country level.
t-statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01.