MANAGING MIGRATION: FEMALE MAYORS AND THE INTAKE OF REFUGEES

Sebastian Schirner*

Published version in the *Journal of Economic Behavior & Organization*:

https://doi.org/10.1016/j.jebo.2025.107306

Abstract

This paper studies whether political leaders' gender matters for crisis management. I examine female mayors in the German state of North Rhine-Westphalia during the intake of Ukrainian refugees in 2022/23. I use granular data on fulfillment of the municipal refugee allocation quota and 2020 municipal election data. I use a two-way fixed effects specification to compare quota fulfillment of female and male mayors before and after the intake of Ukrainian refugees. I supplement this approach by using close mixed-gender mayorship races. Female mayors fulfill the allocation quota less than male mayors in response to the crisis. This difference is not driven by mayor or municipality characteristics. Instead, electoral competition is a plausible mechanism for female mayors' lower quota fulfillment.

Keywords: Political selection, female mayors, local policy choices, refugee migration
JEL codes: D72, J16, J15, H70, F22

---

*ifo Institute, Poschingerstraße 5, 81679 Munich, Germany. E-mail: schirner@ifo.de. I thank participants at the 2024 Meeting of the European Public Choice Society in Vienna, the workshop on the Economic, Social and Political Effects of Migration in Leiden, the 3rd IEB Workshop on Public Policies in Barcelona, the 2024 Annual Congress of the IIPF in Prague, the 14th ifo Dresden Workshop on Regional Economics, the 2024 VfS Conference in Berlin, the 14th Annual Conference on Immigration in OECD Countries in Paris, and in seminars at the Ruhr University Bochum. I am grateful for funding from the German Research Foundation (DFG) for the project "Elections and Public Policy in Times of Crisis" (Grant no. DFG-506165485). Wim Riecken provided excellent research assistance.

1 Introduction

Cooperation of local authorities is key in managing refugee migration. They organize accommodation, provide local public services, and support the integration of refugees into local communities. Local policymakers are in charge of implementing these policies. Thus, it is crucial to understand how their personal characteristics shape migration policy at the local level (Besley, 2005). This paper studies the effect of policymakers' gender on crisis management, focusing on a migration crisis. Specifically, I examine differences in migration policy of female and male mayors in in the German state of North Rhine-Westphalia (NRW) during the intake of Ukrainian refugees after the Russian invasion in February 2022.[1]

An emerging literature examines female leaders in times of crisis (Bruce *et al.*, 2022; Eslava, 2024; Chauvin and Tricaud, 2024). These studies find substantial differences in how female and male policymakers respond to crises. There is evidence that differences in crisis management reflect inherent gender-based preferences but also arise from different electoral incentives for female policymakers. Beyond crisis management, there are gender differences in policy outcomes. For example, female policymakers increase the provision of social public services, such as child care, education, and health care (see Hessami and Lopes da Fonseca (2020) for an overview). These public services are also relevant when hosting refugees. It is thus conceivable that policymakers' gender matters for migration policy at the local level.[2]

To examine female mayors during a migration crisis, I use municipal data from NRW, the most populous German state. Its distinct refugee allocation procedure creates a unique setting within Germany to study female mayors and the intake of refugees. In NRW, refugees are allocated to municipalities instead of districts via an administrative quota based on municipal population and area. The allocation quota produces weekly data on municipal refugee alloca-

[1]The Russian invasion of Ukraine on 24. February 2022 created an unexpected and large intake of refugees to Germany. In the subsequent months, more than 1.1 million Ukrainians arrived in Germany (DESTATIS, 2023). This intake is comparable in size to the previous episode of refugee migration in 2015/16 (BAMF, 2017).

[2]From a theoretical point of view, citizen-candidate models provide arguments why personal characteristics, such as gender, matter for policy. Citizens endogenously run for office and implement their individual policy preferences once elected. In such settings, candidates' own policy preferences are the only policies they can credibly commit to (Osborne and Slivinski, 1996; Besley and Coate, 1997). These models' predictions differ from the median voter theorem, where policymakers' characteristics do not matter for policy (Downs, 1957).

tion and quota fulfillment between 2020 and 2023. The availability of refugee data at a fine-grained spatial and temporal level makes NRW an ideal setting to examine the recent intake of Ukrainian refugees. Studying quota fulfillment is relevant as there is imperfect compliance with this administrative rule. The aim of the allocation mechanism is to evenly distribute refugee-related responsibilities among municipalities. Yet, deviations from the quota are widespread and frequent. All municipalities deviated by at least 20 ppts relative to complete fulfillment at least once during the sample period. To design effective allocation mechanisms it is crucial to understand the reasons for the variation in fulfillment.

I combine the refugee data with data on the municipal election in September 2020. As of this election, 14% of mayors are female (57 out of 396 municipalities). During the legislative period, the Russian invasion of Ukraine caused an unexpected increase in refugees entering Germany and NRW. I compare refugee quota fulfillment of female and male mayors in response to this migration crisis. I use a standard two-way fixed effects (TWFE) model in my baseline specification. To complement this approach, I use close mixed-gender races for the mayorship in a local difference-in-differences setting. The latter approach leverages plausibly exogenous variation in mayor gender while comparing quota fulfillment before and after the crisis.

I find substantial differences between female and male mayors. In the TWFE specification, the monthly quota fulfillment is about 5 ppts lower for female mayors than for their male counterparts in response to the crisis (average fulfillment: 95.6%). This corresponds to about 26 fewer refugees. Importantly, as monthly state-level refugee numbers are fixed, the gap in fulfillment is the relative difference between female and male mayors. Female and male mayors do not differ in the quota fulfillment prior to the intake of Ukrainian refugees. The gap in fulfillment arises immediately after February 2022 and remains stable thereafter. Most Ukrainian refugees arrived immediately after the Russian invasion. This required a swift reaction by mayors. Thus, the stable difference in fulfillment over time is due to path dependence after the initial refugee shock. The difference in fulfillment is larger in the local difference-in-differences setting (10–12 ppts). Coefficients are less precisely estimated, however. A battery

of robustness checks, e.g., using weekly data, excluding large cities, and excluding municipalities with central asylum facilities, confirms the baseline result.

I argue that the difference in quota fulfillment is due to supply-side factors, that is, mayor and municipality behavior. To substantiate this, I exclude demand-side factors. I show that the gap in fulfillment does not reflect refugees' sorting into municipalities. Observable municipality characteristics are strongly balanced between municipalities with female and male mayors. In particular, economic characteristics and preexisting Eastern European networks do not differ. Municipalities are also comparable in refugee numbers during the European refugee crisis in 2015/16. Thus, demand-side factors do not explain the gap in fulfillment.

Evidence from 2,000 hand-collected council minutes reveals an increase in discussions about refugees in response to the crisis. This underscores their relevance for policymakers at the local level. This pattern does not differ between female and male mayors, however. To examine the mayors' agenda, I classify topics in council discussions about refugees using a Large Language Model (LLM). The most relevant refugee topics are accommodation and municipal finances. There are no differences in topics between female and male mayors in response to the crisis.

I examine several mechanisms to shed light on reasons for the quota fulfillment gap. The gap does not reflect differences in sentiment of council discussions about refugees. That is, councils chaired by female and male mayors use positive and negative terms with similar frequency. Furthermore, the baseline gap is not driven by other personal characteristics, such as incumbency, party affiliation, or prior experience in public administration. Also, municipal finances, the support by female councilors, or labor market competition for women do not explain the difference in quota fulfillment in response to the crisis.

Instead, differential electoral incentives are a plausible explanation for the difference in fulfillment. Both, female and male mayors typically strife for reelection. As female politicians find it harder to get reelected, the electoral incentive to deviate from costly fulfillment is stronger for female mayors. To test this mechanism, I use variation in electoral incentives. I use mayors' age to proxy for the probability of running in the 2025 election as there are no

term limits (Foremny *et al.*, 2018). The difference in quota fulfillment is more pronounced among younger mayors (< 57 years) who are more likely to recontest in the upcoming election. In addition, I follow Chauvin and Tricaud (2024) and use the closeness of the 2020 election as a proxy for future electoral competition. The difference in fulfillment is visible only for mayors that were elected by a narrow margin in 2020. That is, female mayors fulfill the quota less to secure reelection when they expect strong competition in the upcoming election. Taken together, the difference in quota fulfillment is not rooted in refugee sentiment or personal and municipality characteristics. Instead, concerns for reelection and electoral incentives explain the behavior of female mayors in times of a migration crisis.

This research relates to the literature on women as policymakers.[3] Evidence from industrialized countries is mixed. Lippmann (2022) finds that female legislators in France focus on topics that are more relevant for women. Baskaran and Hessami (2025) show that female councilors in Bavaria increase the provision of child care. Conversely, Ferreira and Gyourko (2014), Geys and Sørensen (2019), Casarico *et al.* (2022), and Carozzi and Gago (2023) find no evidence for substantial effects of women on policy outcomes. In particular, my paper relates to the literature on female policymakers in times of crisis. Eslava (2024) shows that female mayors led to a reduction in violence during the civil war in Colombia. Bruce *et al.* (2022) show that female mayors in Brazil mitigate the health consequences of COVID-19 and increase enforcement of non-pharmaceutical interventions. Congruently, Chauvin and Tricaud (2024) find that female mayors in Brazil reduce deaths from COVID-19. They establish that gender-specific electoral incentives matter in times of crisis. Danzer *et al.* (2024) show that a higher share of female politicians at the national level is associated with fewer school closures during COVID-19 in Europe.[4]

[3]There is a large literature on female policymakers in developing countries. Chattopadhyay and Duflo (2004) find that female leaders provide public services that reflect female preferences in India. Also, female policymakers increase spending on education (Clots-Figueras, 2012) and health care (Bhalotra and Clots-Figueras, 2014) in India. Brollo and Troiano (2016) find that female mayors in Brazil are less corrupt than male mayors.

[4]My paper also relates to the literature on gender differences in compliance behavior. Apesteguia *et al.* (2013) find no gender difference in compliance with rules in public libraries. D'Attoma *et al.* (2020) show that women are more tax compliant than men. Galasso *et al.* (2020) show that women exhibited higher compliance with rules imposed during COVID-19. Croson and Gneezy (2009) argue that women are more responsive to social and environmental cues, a possible explanation for these heterogeneous results

In addition, there is a large literature on various dimensions of refugee migration. Typically, this literature focuses on the consequences of refugees on political outcomes (Dustmann *et al.*, 2019; Steinmayr, 2021; Bredtmann, 2022), public spending (Ruist, 2020), local population movements (Batut and Schneider-Strawczynski, 2021; Zuchowski, 2025), or labor markets (Dustmann *et al.*, 2016). There is substantially less evidence on the political determinants of refugee allocation.[5] Fratesi *et al.* (2018) show that Italian municipalities with high social capital oppose hosting refugees to retain community cohesion. Bracco *et al.* (2018) show that Lega Nord mayors reduce the intake of immigrants in Northern Italy. Ferwerda and Riaz (2023) document that the placement of refugee facilities in Germany is endogenous to political deliberations. In particular, they show that facilities are concentrated in left-leaning municipalities. Gamalerio and Negri (2023) show that the timing of local elections affects the probability that Italian municipalities apply for a refugee facility.

I contribute to these strands of the literature in three distinct ways. First, I show that mayor characteristics like gender matter for the allocation of refugees at the local level. This opens up a new perspective on the political determinants of refugee allocation, as previous research has focused on municipality characteristics. Specifically, I show that gender-specific incentives matter even in the presence of an administrative allocation quota. Second, the intake of Ukrainian refugees represents a substantially different context compared to the European refugee crisis in 2015/16, which has been examined extensively. Refugees from Ukraine and Syria differ for example, in their cultural proximity to Germans, their socio-economic characteristics, and their perception by natives. Third, I examine the role of female leaders during a specific type of crisis. This is the first paper examining the role of female policymakers during a migration crisis. Managing the intake of refugees gives rise to different challenges relative to health crises, such as COVID-19. In particular, the provision of suitable housing and child care or the support of integration into local communities are unique to migration crises. Examin-

[5]There is evidence on where refugees and migrants settle if allowed to choose their location. Previous settlement patterns of co-ethnics predict current migration flows. This finding provides the foundation for shift-share type instruments (Card, 2001). In addition, Damm (2007) shows that the availability of housing, education, and labor market characteristics matter for location choices.

ing migration crises is important as e.g., climate change potentially increases migration in the future (Cattaneo *et al.*, 2019).

2 Institutional background

2.1 Mayors and municipal councils in NRW

The mayorship is the most important municipal office in NRW. Mayors are directly elected and serve a term of five years. Candidates need an absolute majority of votes to win. If no candidate obtains the majority of votes in the first round, runoff elections take place. All mayors in NRW are full-time public employees. Candidates must be older than 23 years, citizen of the European Union, and have their place of residence in Germany. Starting with the election in September 2020, mayors are elected on the same day as the municipal council. Councilors in NRW are elected in a personalized proportional representation system with closed lists. Municipal councils have between 20 and 90 seats based on the population of the municipality. The mayor is the chairperson of the council. She schedules the meetings, assembles the agenda, and has a vote in the council. In addition, the mayor is responsible for the municipal administration and the execution of council decisions (Kost, 2022). Mayors advocate the policies of the municipal administration towards citizens and other tiers of government. Thus, mayors are crucial in managing the intake of refugees.

2.2 Refugees in NRW

Refugees are allocated to the sixteen German states based on the *Königsteiner Schlüssel*.[6] Those allocated to NRW arrive in the states' initial reception facility in Bochum (*Landeserstaufnahmeeinrichtung* – LEA). They are subsequently transferred to one of five reception facilities (*Erstaufnahmeeinrichtungen* – EAE) where they formally apply for asylum and a medi-

[6]The *Königsteiner Schlüssel* is not unique to the allocation of refugees. It is used for the allocation of financial funds and responsibilities among German states. It is calculated based on a states' tax revenue (two-thirds) and population (one-third).

cal examination takes place. After about one week, refugees are transferred to accommodation facilities run by the state of NRW (*Zentrale Unterbringungseinrichtung* – ZUE). After three months, refugees are allocated to the 396 municipalities in NRW.

Refugees are allocated to municipalities based on an administrative quota. All municipalities are obliged to host refugees. NRW differs from other German states that allocate refugees to districts rather than to municipalities. The local authority of Arnsberg is in charge of the allocation. The allocation quota is based on the population of a municipality (90%) and its area (10%). There are deductions for municipalities hosting the states' asylum facilities. Municipalities can suspend hosting refugees for a limited time due to unforeseeable circumstances. There are no sanctions for municipalities that do not fulfill the quota. Refugees that are not hosted by a municipality are transferred to a different municipality or stay in a state facility. Municipalities are obliged to report current refugee numbers on a weekly basis, so that quotas can be updated frequently (SGV, 2023).

Municipalities are compensated for hosting refugees by the state government. They receive a lump sum transfer for each refugee living in the municipality (2023: 875/1,125 Euros per refugee in municipalities/independent cities). Municipalities are required to provide suitable accommodation for refugees, either in shared accommodation facilities or in regular flats. In addition, the intake of refugees requires the provision of local public services, such as child care, schooling, or health care.

There are special regulations for Ukrainian refugees. The German federal government waived Ukrainian refugees from obtaining a German residence permit immediately after the start of the Russian invasion. Thus, they are not obliged to apply for asylum. In addition, Ukrainian refugees can obtain regular social security instead of refugee benefits. In principle, Ukrainian refugees are part of the allocation quota. There is evidence that Ukrainians partly settled in without government support and thus may not be included in the administrative allocation process, however.

The socio-economic characteristics of Ukrainian refugees differ from those of previous refugee groups. 80% of Ukrainian refugees are female. Many arrive with their children. 72%

have a higher education degree. Most Ukrainians live in a private accommodation (74%), partly with friends or family. Only about 9% live in shared refugee facilities. About one third of Ukrainians intend to return to Ukraine after the war (Brücker *et al.*, 2023).

3 Data

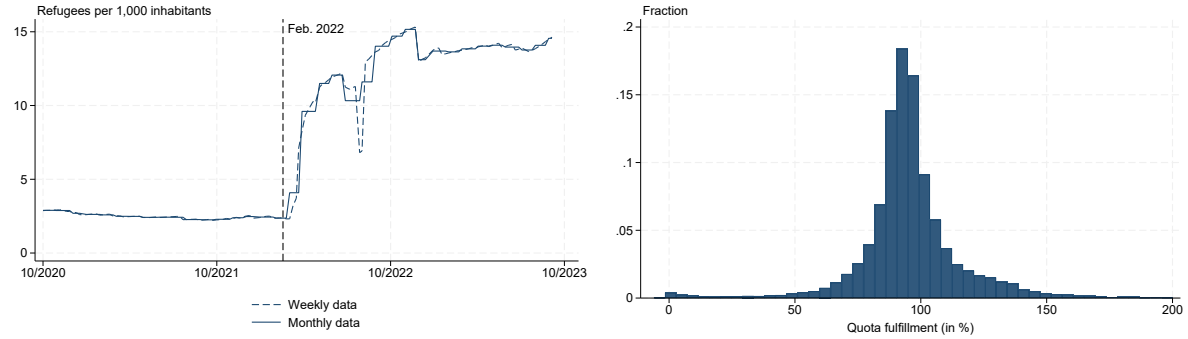
3.1 Election data

Data on the 2020 municipal election is provided by the Statistical Office of NRW. Data on mayor elections includes candidates' name, list, and individual vote shares plus information on electorate size, valid votes, invalid votes, and turnout. A subset of municipalities had an extraordinary election prior to 2020. Information on these elections was hand-collected from publicly available sources.[7] I infer a mayor's gender from his/her first name using the R-package *gender*. Ambiguous names were checked manually. Mayors in NRW are predominately male. Only about 14% of mayors in the election period of 2020 are female. Subfigure (a) of Figure A.8 shows that female mayors are distributed evenly across NRW. In addition, I hand-collect information on year of birth, birthplace, education, occupation, marital status, immigrant origin, and incumbency (see Section A.1 in the online appendix). To complement information on mayors, I obtain data on the share of female councilors from the Statistical Office of NRW (average: 27.34%).

3.2 Refugee data

NRW is an ideal setting to examine the intake of Ukrainian refugees after February 2022. Due to the allocation quota, there are weekly reports on the stock of refugees at the municipality level. Data is publicly available from the regional authority of Arnsberg in charge of refugee

[7]Reasons for extraordinary elections are for example, the resignation of a mayor or a recall from office. There were extraordinary elections in 16 municipalities. Extraordinarily elected mayors remain in office until the next regular municipal election in 2025. I exclude these municipalities in a robustness check to ensure that longer tenure does not drive the results. There were two extraordinary elections between the regular election in September 2020 and September 2023. In both cases the mayors' gender did not change.



(a) Refugees over time

(b) Distribution of quota fulfillment

Figure 1: Refugees over time and quota fulfillment

Notes: Subfigure (a) shows the stock of refugees per 1,000 inhabitants over time. The graph shows weekly and monthly data between October 2020 and September 2023. The complete time-series from 2018 to 2023 can be obtained in Figure A.9 in the online appendix. The vertical line indicates the beginning of the Russian invasion of Ukraine. Subfigure (b) shows the distribution of quota fulfillment in percent. The graph shows weekly data between October 2020 and September 2023. Values above 200% are excluded for readability.

allocation. This includes information on the number of refugees currently living in the municipality, the municipality's allocation quota, the number of refugees allocated by the quota, and quota fulfillment.[8] Details on the refugee data are described in Section A.2 in the online appendix. Data is available from April 2018 to September 2023.[9] Note that data is not available for the previous period of increased refugee migration in 2015/16. Instead, I obtain data on refugees in 2015 from a survey among municipalities by a public TV channel (Westdeutscher Rundfunk Köln, 2015).

Data does not include information on the origin of refugees. It covers refugees from Ukraine, but also from all other countries. Subfigure (a) of Figure 1 shows the intake of refugees per 1,000 inhabitants over time. The descriptive pattern indicates that the intake of refugees is mainly driven by Ukrainians. Before February 2022, the number of refugees was stagnating at about 2.5 refugees per 1,000 inhabitants. Refugee numbers strongly increased

[8]Ukrainian refugees do not formally apply for asylum. Some may go unregistered as they settle in with friends or family. Thus, the refugee numbers in the data are a lower bound. A back-of-the-envelope calculation sheds light on the extent of underestimation. Between March and August 2022, Germany hosted about 1,000,000 Ukrainian refugees. Following the *Königsteiner Schlüssel* about 210,000 were assigned to NRW. The actual number of refugees between March and August 2022 in NRW is 185,019. Assuming almost no other refugee migration, this would indicate that the data captures 88% of Ukrainian refugees.

[9]This corresponds to all weeks available publicly until the time of data collection (15.09.2023). While data is still provided on a weekly basis, I refrain from collecting additional data after conducting the analysis.

after February 2022. By the end of 2022, municipalities in NRW hosted about 14 refugees per 1,000 inhabitants. Subfigure (a) shows a large drop in the number of refugees per 1,000 inhabitants in the summer of 2022. A legal change for Ukrainian refugees regarding social security benefits resulted in uncertainty among municipalities if Ukrainians are still part of the allocation process. From September 2022 onward, all Ukrainian refugees are included in the allocation quota again. To obtain a smoother time series, I aggregate weekly data per month for the baseline specification. In addition, I use linear interpolation to obtain hypothetical data for these weeks for a robustness check.

The main dependent variable is the fulfillment of the refugee allocation to the municipality. This measures the local administration's ability or willingness to organize and manage the intake of refugees. The quota fulfillment is calculated as follows:

$$Y_{m,t} = \left(\frac{\text{Refugees}_{m,t}}{\text{Refugees allocated}_{m,t}} \right) \times 100. \tag{1}$$

There is substantial variation in the quota fulfillment. Subfigure (b) of Figure 1 shows the distribution of weekly fulfillment which is symmetric with a mode close to 100%. On average, municipalities in NRW host 95.6% of the refugees that they are requested to host (see Table A.7 in the online appendix). There are municipalities that host substantially fewer or more refugees than allocated and deviations are widespread. All 396 municipalities deviate from the quota by more than 20 ppts at least once during the sample period. Figure A.10 shows the distribution separately for female and male mayors and before and after the Russian invasion. While the distributions overlap, a comparison of averages points at a fulfillment gap between female and male mayors in response to the crisis.[10] Subfigure (b) of Figure A.8 in the online appendix shows the spatial distribution of the quota fulfillment as of March 2022. There are no apparent spatial clusters.

[10] Figure A.11 in the online appendix shows the distribution of the dependent variable after taking municipality averages across the 2020 election period. There are municipalities that on average do not fulfill substantially. Conversely, there are municipalities that on average host more refugees than required by the quota.

3.3 Additional municipality data

Data on municipal elections and refugees is complemented with municipality-level information on population structure, area, municipal finances, unemployment, and vote shares of major parties in the 2017 federal election. Data is provided by the Statistical Office of NRW. As hosting refugees relies on the availability of housing, I obtain data on housing construction. In addition, the German censuses of 2011 and 2022 provide information on the share of empty flats. The RWI-GEO-GRID provides the population share by ethnic background which is derived from the names of household heads (RWI and Microm, 2023).

4 Empirical strategy

I examine the difference in refugee quota fulfillment between female and male mayors in times of a migration crisis. In essence, I compare municipalities with female and male mayors before and after the intake of Ukrainian refugees. In practice, I estimate the following dynamic two-way fixed effects (TWFE) specification for a balanced panel of municipalities and months:

$$Y_{m,t} = \sum_{t=2020.10}^{2023.9} \beta(\textit{Female mayor}_{m,2020.9} \times T_t) + \gamma_m + \zeta_t + \varepsilon_{m,t}. \tag{2}$$

$Y_{m,t}$ is quota fulfillment in municipality m in month t in percent. $\textit{Female mayor}_{m,2020.9}$ is an indicator that is one if the municipality elected a female mayor in September 2020 (57 municipalities). This variable is interacted with month indicators T_t. The reference period is February 2022. I also estimate a reduced version of this specification in which I interact the indicator for female mayors with an indicator for time periods after February 2022.[11] The specification includes municipality fixed effects γ_m and month fixed effects ζ_t. The sample covers all 396

[11] The Russian invasion of Ukraine started on 24. February 2022. It caused more than two million Ukrainian citizens to flee their country immediately (UNHCR, 2022). Weekly data on refugee allocation in NRW shows that the increase in refugees in the municipalities starts in the end of March 2022 (Figure 1). This is because it likely takes time until municipalities report all refugees.

municipalities between October 2020 and September 2023. Standard errors are clustered at the municipality level.

In this specification, the treatment, i.e., being governed by a female mayor in times of crisis, is assigned prior to the unexpected increase in Ukrainian refugees. The identification assumption is that the quota fulfillment of female mayors in times of crisis would have evolved similarly if the municipality were governed by a male mayor. I corroborate this parallel trends assumption by examining the quota fulfillment of female and male mayors before the intake of refugees after February 2022. In addition, Subfigure (a) of Figure A.12 in the online appendix shows that municipalities with female and male mayors do not significantly differ in observable pre-treatment characteristics. In particular, municipalities were not differently affected by the refugee crisis in 2015.

The election of a female mayor is not exogenous. Potentially, unobservable time-varying municipality characteristics influence both the election of a female mayor and quota fulfillment. To alleviate such concerns, I use exogenous variation from close mixed-gender races for the mayorship. I follow Bruce *et al.* (2022) and compare quota fulfillment by mayor gender before and after February 2022 limiting the sample to municipalities with a close mixed-gender race in the election of 2020 (local difference-in-differences specification). This approach leverages plausibly exogenous variation in mayor gender while comparing quota fulfillment before and after the crisis. Details on this approach are presented in Section A.3 in the online appendix.

5 Results: Female mayors and quota fulfillment

5.1 TWFE results

First, I examine the difference in quota fulfillment between female and male mayors descriptively. Figure 2 shows quota fulfillment for female and male mayors over time. Prior to the intake of refugees from Ukraine, fulfillment was similar for female and male mayors. If anything, fulfillment was higher for female mayors in some months. Quota fulfillment drops considerably for female and male mayors after the intake of Ukrainian refugees. This drop is substantially

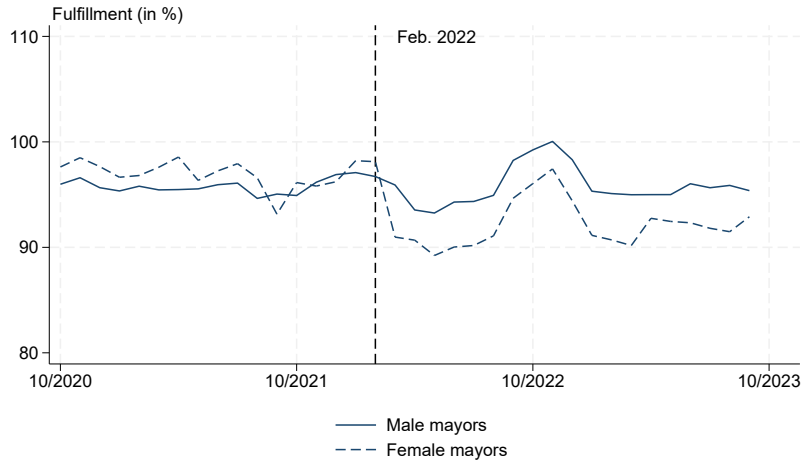


Figure 2: Female mayors and refugee quota fulfillment

Notes: This graph shows quota fulfillment in percent over time for male and female mayors. The graph shows monthly data between October 2020 and September 2023. The vertical line indicates the beginning of the Russian invasion of Ukraine in February 2022.

larger for municipalities with a female mayor. Fulfillment remains lower for female mayors until the end of the sample period.

Next, I analyze this pattern in a regression framework following Equation (2). Figure 3 shows the baseline results of the dynamic TWFE specification. The effect of female mayors on quota fulfillment is insignificant and close to zero prior to the intake of refugees from Ukraine.[12] This lends credibility to the parallel trends assumption: Municipalities with female mayors were not on a different trend in fulfillment prior to the intake of refugees from Ukraine. Coefficients are consistently negative after the intake of refugees from Ukraine. Female mayors fulfill the quota substantially less in response to the migration crisis. In most months, coefficients are significant at the 5%-level.[13] The difference in fulfillment arises immediately after February 2022. It remains stable thereafter. Most refugees fled Ukraine immediately after the

[12]I test this formally with an F-test on the joint significance of all coefficients prior to the intake of refugees from Ukraine. The null hypothesis that all coefficients are jointly zero cannot be rejected at common levels ($p = 0.69$). Rambachan and Roth (2023) provide an approach to assess the credibility of the parallel trends assumption. They use deviations of pre-treatment coefficients to examine the robustness of post-treatment coefficients to violations of parallel pre-trends. Results are presented in Figure A.13 in the online appendix. There is evidence that parallel trends may not be robust to large deviations in pre-trends.

[13]In settings with few treated units, inference using cluster-robust standard errors can lead to over-rejection of the null hypothesis of no effect. Thus, I also calculate the standard error using a randomized inference procedure (MacKinnon and Webb, 2020). The baseline coefficient remains significant at the 5%-level ($p = 0.03$). Figure A.14 in the online appendix shows the result of the randomized inference procedure.

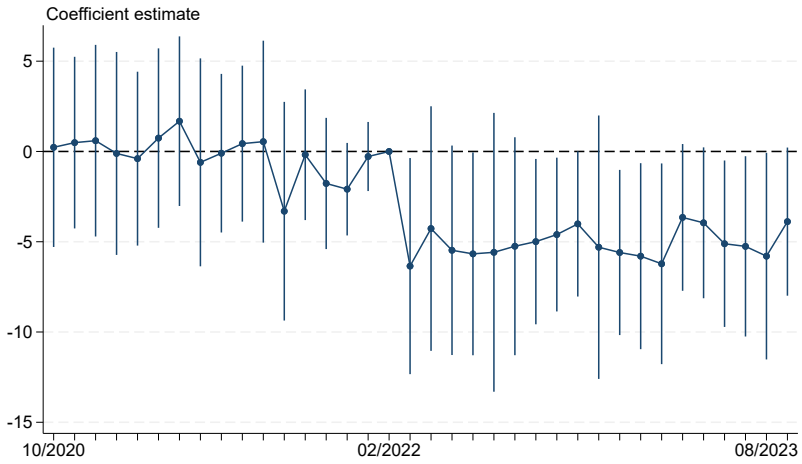


Figure 3: Female mayors and refugee quota fulfillment – TWFE

Notes: This graph shows coefficient estimates of interactions between month indicators and an indicator for female mayors (N: 14,256). The dependent variable is monthly aggregates of refugee quota fulfillment in percent. The regression includes municipality and month fixed effects. The reference period is February 2022. Standard errors are clustered at the municipality level. The full sample of 396 municipalities is used. 95% confidence intervals are indicated in the graph.

Russian invasion. Substantially fewer Ukrainian refugees arrived as the war went on (compare Figure 1). Also, Ukrainian refugees tend to stay in the municipality they initially arrived (Siegert *et al.*, 2023). Thus, the stable difference in fulfillment reflects path dependence after the initial refugee shock.

To assess the size of the gap in fulfillment I estimate a reduced model. Model (1) of Table 1 shows the result of a model in which female mayor indicator is interacted with an indicator for months after February 2022. The difference in quota fulfillment is quantitatively meaningful. Municipalities with female mayors have 4.9 ppts lower fulfillment relative to municipalities with male mayors after February 2022. Evaluated at the average fulfillment of 95.6%, this amounts to a 5.1% difference. On average, municipalities in NRW are supposed to accommodate about 540 refugees after February 2022. That is, 4.9 ppts lower fulfillment corresponds to 26 fewer refugees.[14] An alternative outcome is the number of refugees per 1,000 inhabitants. Model (2) of Table 1 shows the difference between female and male mayors using

[14]Monthly averages of fulfillment represent the intensive margin. I also examine differences in the probability to fulfill or overfulfill the quota (extensive margin). I run a linear probability model with a binary outcome that is one if fulfillment in a given month is higher or equal to 100%. Results are presented in Figure A.15 in the online appendix. There is no significant difference in the extensive margin in response to the crisis.

Table 1: BASELINE – FEMALE MAYORS AND FULFILLMENT

				Local diff-in-diff (Fulfillment)		
	(1) Fulfillment	(2) Refugees p.c.	(3) Alloc. quota	(4)	(5)	(6)
Female mayor × Post	-4.851**	-0.502*	-0.013	-11.454*	-12.284*	-9.909**
	(2.291)	(0.286)	(0.065)	(6.748)	(6.815)	(4.544)
Mean (SD)	95.63 (19.14)	7.90 (5.87)	25.25 (46.62)	99.92 (24.95)	99.92 (24.95)	96.62 (21.11)
Time FE	✓	✓	✓	✓	✓	✓
Municipality FE	✓	✓	✓	✓	✓	✓
Bandwidth				20%	20%	All
Polynomial				Linear	Quadratic	Linear
Municipalities	396	396	396	40	40	110
N	14,256	14,255	14,256	1,440	1,440	3,960

Notes: This table reports results from regressions that relate an indicator for female mayor (*Female mayor*) interacted with an indicator for the time after the intake of refugees from Ukraine (*Post*) to fulfillment of the refugee allocation quota (in %). Model (1) reports results from a TWFE model. Model (2) reports results from a TWFE model using refugees per 1,000 inhabitants as dependent variable. Model (3) reports results from a TWFE model using the allocation quota as dependent variable. Model (4) reports results from a local difference-in-differences model using a linear polynomial of the running variable and a bandwidth of 20%. Model (5) reports results using a quadratic polynomial of the running variable and a bandwidth of 20%. Model (6) includes a linear polynomial and all mixed-gender races. The row entitled Mean (SD) reports the mean and standard deviation of the dependent variable. Stars indicate significance levels at 10%(*), 5%(**), and 1%(***). Standard errors in parentheses are clustered at the municipality level.

this outcome. Municipalities with female mayors host fewer refugees per 1,000 inhabitants in response to the crisis.

The above findings show that mayors can influence allocation quota fulfillment. They cannot influence the underlying allocation quota, however. Thus, there should be no difference in the quota between female and male mayors. To test this empirically, I use the allocation quota as dependent variable. Model (3) in Table 1 shows the result of this placebo exercise. Reassuringly, the coefficient is insignificant and virtually zero.

Monthly state-level refugee numbers are fixed. This affects the interpretation of the difference in fulfillment. As municipalities with female mayors host fewer refugees, other municipalities must host more refugees. Thus, the coefficients capture the relative difference in allocation between female and male mayors in a post-crisis equilibrium. They do not reflect an absolute reduction of refugees in NRW. This interpretation acknowledges spillovers between municipalities with female and male mayors by construction. This potentially violates the stable unit treatment value assumption (SUTVA): There must be no interference between treatment and control units. I examine the role of SUTVA in more detail in Section 5.3.

Taken together, the TWFE results show substantially lower quota fulfillment by female mayors in response to the crisis. Note that I do not infer normative claims from quota fulfill-

ment. Underfulfillment can be reasonable in economically strained municipalities. Overfulfillment may have adverse consequences for adequate housing and integration of refugees. I will examine the mechanism behind the gender difference in quota fulfillment in Section 7.

5.2 Local difference-in-differences results

To support the findings of the TWFE approach, I next show the results of the local difference-in-differences specification. Graphical evidence reveals a visible discontinuity at the threshold: Female mayors elected in close mixed-gender races fulfill the quota less in response to the crisis (see Figure A.4 in the online appendix). Regression results are presented in Table 1. Model (4) shows the results of the local difference-in-differences approach using a linear polynomial and limiting the sample to close races (20% bandwidth). Indeed, female mayors elected in close mixed-gender races fulfill the quota less than male mayors in response to the crisis. The coefficient is significant at the 10%-level. The effect size is considerably larger than in the TWFE specification. The effect is similar when controlling for a quadratic specification of the margin of victory in Model (5) and when using all 110 mixed-gender races in Model (6). I show robustness to different bandwidths and validity tests in Section A.3 in the online appendix. As there is only one election and thus a limited number of mixed-gender races, I cannot use arbitrarily narrow bandwidths. This limits local randomization and results in imprecise estimates. Also, this local approach raises questions on external validity. Thus, I consider these results as supporting evidence only.

5.3 Robustness

In this section, I assess the robustness of the TWFE approach to different specifications and estimation samples. Table 2 shows the main robustness checks. I discuss additional robustness checks in Section A.4 in the online appendix. First, I show robustness to the choice of temporal aggregation in Models (1) and (2). The corresponding dynamic plot is Figure A.16 in the online appendix. Using weekly instead of monthly data does not change the main finding. As

Table 2: ROBUSTNESS – FEMALE MAYORS AND FULFILLMENT

	Dep. var.: Quota fulfillment (in %)					
	(1) Weekly	(2) Interpol.	(3) Excl. *Krfr.*	(4) Excl. < 2020	(5) EAE/ZUE	(6) < 09/2022
Female mayor × Post	-4.929**	-5.012**	-4.541*	-5.007**	-4.120*	-5.190*
	(2.309)	(2.300)	(2.331)	(2.388)	(2.398)	(2.855)
Mean (SD)	95.59 (21.10)	95.74 (20.29)	95.16 (19.04)	95.72 (19.05)	93.46 (19.36)	95.44 (19.53)
Time FE	✓	✓	✓	✓	✓	✓
Municipality FE	✓	✓	✓	✓	✓	✓
Municipalities	396	396	374	381	362	396
N	59,002	59,002	13,464	13,716	13,031	9,108

Notes: This table reports results from regressions that relate an indicator for female mayor (*Female mayor*) interacted with an indicator for the time after the intake of refugees from Ukraine (*Post*) to fulfillment of the refugee allocation quota (in %). Model (1) uses weekly instead of monthly data. Model (2) uses weekly data and values of a linear interpolation during the summer of 2022. Model (3) excludes the 22 independent cities. Model (4) excludes municipalities that had an extraordinary election before September 2020. Model (5) excludes municipalities that host state-run asylum seeker facilities (LEA/EAE/ZUE). Model (6) limits the sample to the months before September 2022. The row entitled Mean (SD) reports the mean and standard deviation of the dependent variable. Stars indicate significance levels at 10%(*), 5%(**), and 1%(***). Standard errors in parentheses are clustered at the municipality level.

explained in Section 3.2, weekly data exhibits a large deviation during the summer of 2022 for purely administrative reasons. Using a linear interpolation instead of the original values in the summer of 2022 does not change the main finding.[15]

Next, I examine if spatial or temporal patterns drive the results. In Model (3), I exclude all 22 independent cities (*Kreisfreie Städte*). Mayors in these cities likely face different possibilities and constraints relative to mayors in more rural areas. I exclude all municipalities that had an extraordinary election before the election in September 2020 in Model (4). In Model (5), I exclude 34 municipalities that host a state-run refugee facility (LEA/EAE/ZUE, see Section 2.2). In Model (6), I limit the time period to the first six months after the Russian invasion of Ukraine. No model yields a coefficient that is substantially different from the baseline finding.

Finally, I examine the consequences of SUTVA violations explicitly. As the monthly number of refugees in NRW is fixed, municipalities hosting fewer refugees may create spillovers to other municipalities that host additional refugees. To examine the role of spillovers empirically, I alter the composition of the control group (Batut and Schneider-Strawczynski, 2021). Results are shown in Table A.8 in the online appendix. In Models (1) and (2), I keep only

[15] I use linear interpolation due to its simplicity and because the time period to interpolate is short. There are more complex methods, such as piecewise cubic Hermite interpolation. Figure A.17 in the online appendix shows that both methods yield virtually identical interpolated values.

municipality-time pairs with lower than 100% fulfillment. That is, I compare female and male mayors within a subgroup of municipalities that do not completely fulfill the quota. This rules out that differences in fulfillment are solely driven by over-fulfillment of other municipalities. Coefficients are somewhat smaller than the baseline but remain significant at the 10%-level.

Spillovers can also be spatial. Municipalities in close geographical proximity may know about their fulfillment and are more likely to coordinate. To rule out that lower fulfillment of female mayors affects their neighbor's fulfillment I drop all municipalities that directly border a municipality with a female mayor in Model (3) of Table A.8. The coefficient is larger than the baseline and significant at the 5%-level. In addition, I follow Clarke (2017) and control for distance explicitly.[16] Results are presented in Model (4) of Table A.8. The coefficient is similar to the baseline and significant at the 10%-level. Taken together, spatial SUTVA violations do not drive my baseline findings.

6 Extensions

6.1 Selective migration of refugees

I can identify lower quota fulfillment of female mayors only if there is no selective migration of Ukrainian refugees. That is, deviations from the allocation quota must be solely due to choices by mayors not due to location choices of refugees. I next provide evidence that selective migration does not explain my findings.[17]

First, refugees may move to municipalities because of favorable economic conditions. This includes employment opportunities, housing availability, or the quality of public services,

[16] In practice, I calculate the Euclidean distance between all municipalities and all municipalities with a female mayor. For each municipality, I take the average distance to all female led municipalities. I interact average distance with a linear time trend and control for it in the regression.

[17] Some Ukrainian refugees settled without government support and are not included in the allocation data. If there are systematic unobserved differences between municipalities with female and male mayors in refugees arriving independently, this could affect the main finding. To alleviate such concerns, I benchmark the allocation data against official refugee numbers from the central register of foreigners (*Ausländerzentralregister*) where registration is mandatory after 90 days. This data is available only at the district level. The correlation between official refugee data and the allocation data is plotted in Figure A.18 in the online appendix. Official district-level refugee shares align closely with refugee shares using the allocation data.

such as schooling (Damm, 2007). If municipalities with female mayors offer systematically worse conditions, causing refugees to avoid these municipalities, this could explain the difference in fulfillment. Figure A.12 in the online appendix shows balance checks for economic conditions, such as municipal spending, unemployment, and the number of schools. Characteristics are strongly balanced between municipalities with female and male mayors. There is little scope for selective migration due to local economic conditions.

Second, refugees may move to municipalities with existing networks of immigrants from similar origins (Bartel, 1989; Card, 2001). In particular, they may select into municipalities with many Ukrainians or Germans with an Ukrainian immigrant origin.[18] Prior to the Russian invasion, there were few Ukrainians in NRW. Between 2001 and 2021, only 1.4% of non-Germans in NRW were Ukrainian citizens (Figure A.19 in the online appendix). Also, there are few Germans with an Ukrainian immigrant origin. About 1.1% of Germans with an immigrant origin in NRW have an Ukrainian origin.

To ensure that selection into pre-existing immigrant networks does not explain lower fulfillment, I use information on ethnic backgrounds available in the RWI-GEO-GRID data. Results are presented in Figure A.20 in the online appendix. There are no significant differences between municipalities with female and male mayors for any ethnic background.[19] In particular, the coefficient for the share of Eastern Europeans, the culturally most similar ethnic background to Ukrainians, is insignificant and virtually zero. In summary, pre-existing Ukrainian networks are limited and the ethnic composition of municipalities is balanced.

[18] Sauer *et al.* (2023) document a correlation between pre-existing Ukrainian networks and the intake of Ukrainian refugees across German districts. At the same time, they find a correlation between the location of earlier refugees and Ukrainian refugees. Thus, they argue that administrative allocation continues to be relevant for location choices.

[19] The Statistical Office provides data on the municipality-level share of non-Germans. The share of non-Germans is balanced between municipalities with female and male mayors (Figure A.12). Administrative data on Ukrainians is available only at the district level. The district-level share of Ukrainians in 2021 is balanced among municipalities with female and male mayors.

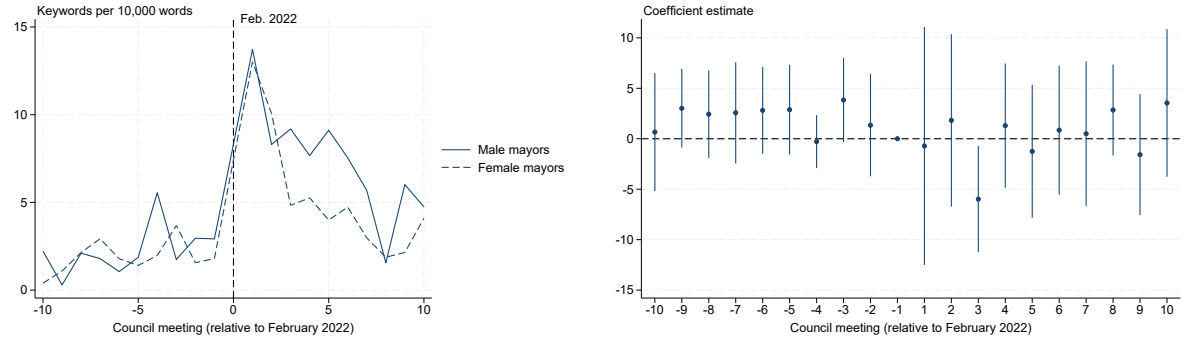
6.2 Municipal finances

I next examine if the difference in quota fulfillment translates into tangible policy differences. Annual data on municipal spending categories between 2020 and 2023 is available from the Statistical Office of NRW. I run models that examine the difference in municipal spending between female and male mayors in response to the crisis. I define 2022 and 2023 as post-crisis years. All outcomes are relative to population and transformed using the inverse hyperbolic sine. Table A.9 in the online appendix shows the results. First, I examine the overall trend in municipal spending. There is no difference in total spending between female and male mayors in response to the crisis. Next, I examine social transfers, which include the social benefits for Ukrainian refugees. Female mayors spend 8.9% less on social transfers per capita in response to the crisis. The coefficient is significant at the 10%-level. The average municipality spends about 180€ per capita on social transfers. Thus, the monetary difference is small. To exclude differences in non-refugee policies, I examine spending categories arguably unrelated to refugees. There is no difference in spending on municipal roads and no difference in municipal transfers to districts. Thus, the difference in quota fulfillment does not translate into broader differences in municipal spending.

6.3 Refugee topics and agenda setting

The local council is a municipalities' most important deliberative body. I examine how differences in managing migration materialize in council discussions. To examine changes in refugee topics and agenda setting, I hand-collected about 2,000 council minutes.[20] Council minutes are publicly available for most municipalities. They include the agenda of the meeting, key lines of discussion, and council decisions. Minutes are not ad verbatim, however. Meetings of the local council take place about once every two months. The mayor prepares the agenda and chairs

[20] I limit the collection of council minutes to municipalities with a mixed-gender race. Unfortunately, I am not able to obtain minutes for all 110 municipalities. Details on the collection of council minutes are provided in Section A.6 in the online appendix. Minutes are pre-processed following common practice (Gentzkow *et al.*, 2019).



(a) Descriptive results **(b)** Dynamic TWFE

Figure 4: Female mayors and agenda setting

Notes: Subfigure (a) shows the number of times refugee-terms are mentioned in council minutes. This number is normalized by the total number of words in the council minutes. The share is shown separately for councils chaired by female and male mayors. The stemmed refugee-terms are *asylbewerb*, *fluechtling*, *asylsuch*, *gefluechtet*, *migrant*, *kriegsfluechtling*, *fluechtlingskind*, *asylant*, *schutzsuch*, *fluechtlingsfamili*, *zuwand*, *fluechtlingshelf*, *asylwerb*. As municipalities choose the timing and the number of meetings individually, I code the time relative to February 2022. For example, the number "five" on the x-axis indicates the fifth meeting since February 2022. Subfigure (b) shows coefficients of regressions following Equation (2) in which the share of refugee-terms relative to all words in council minutes is the dependent variable. 95% confidence intervals are indicated in the graph.

the meeting. The corpus of council minutes is a unique source to examine refugee topics and agenda setting at the local level.

I first examine the frequency of keywords related to refugees in council minutes over time. I obtain keywords automatically from pre-trained German word embeddings. For this, I use German word vectors provided by fastText (Bojanowski *et al.*, 2016). I use the 20 words that appear most often in connection with the term *Fluechtlinge* (i.e., the German word for refugees) measured by cosine-similarity.[21] I count these terms as well as all terms that share the same stem (referred to as refugee-terms in the following). Term counts are normalized by the total number of terms in a council minute to account for their exhaustiveness.

Subfigure (a) of Figure 4 shows the frequency of refugee-terms in council meetings chaired by female and male mayors. The intake of refugees from Ukraine is associated with a strong increase in refugee-terms. This underscores the importance of refugee migration in local politics. The increase is transitory, however. About seven meetings after February 2022, refugee-

[21] Words are stemmed and thereafter duplicates are dropped. The stemmed keywords are *asylbewerb*, *fluechtling*, *asylsuch*, *gefluechtet*, *migrant*, *kriegsfluechtling*, *fluechtlingskind*, *asylant*, *schutzsuch*, *fluechtlingsfamili*, *zuwand*, *fluechtlingshelf*, *asylwerb*. Dictionary approaches are complicated by words having several meanings (polysemy). This does not apply for these keywords, however.

terms are mentioned as often as before. Refugee-terms evolve broadly similar for female and male mayors. To substantiate the descriptive pattern, I rerun the TWFE model described in Equation (2) using the normalized count of refugee-terms as dependent variable. Results are presented in Subfigure (b) of Figure 4. There is no significant difference in the frequency of refugee-terms between female and male mayors in response to the crisis.

Differences in managing migration can also materialize in different policy priorities of mayors. To test for differences in policy, I examine topics discussed in council meetings using a LLM (Korinek, 2023). I limit the corpus to relevant paragraphs, i.e., paragraphs containing refugee-terms. I access *GPT-3.5 turbo* and *GPT-4* via the API of OpenAI to identify topics in an automated way. Based on the paragraphs, I ask the model if accommodation, education, child care, crime, municipal finance, or integration are discussed in the council meeting (yes/no).[22] From the models' answers, I define indicator variables and use them as dependent variables in linear probability models. That is, I analyze if the probability of a given topic differs between female and male mayors after the intake of Ukrainian refugees.

Results are presented in Figure A.21 in the online appendix. Descriptively, the most frequent topic is accommodation. 74% of council meetings that include discussions about refugees deal with accommodation. Other frequent topics are municipal finance and integration. Education, child care, and crime receive relatively little attention. There is no significant change in topics between female and male mayors in response to the crisis. Coefficients are close to zero and insignificant. Results are almost identical for *GPT-3.5 turbo* and *GPT-4*.[23]

Taken together, there is no evidence that refugee topics and agenda setting differ for female and male mayors in response to the crisis. These findings are not consistent with inherently different policy priorities of female mayors in times of crisis.

[22]Several topics per council minute are possible. Using the API is subject to a fee. To ensure reproducibility, I set the temperature parameter to zero. The model thus provides close to deterministic answers. According to OpenAI, there is still some variability left, however. Details on the procedure are described in Section A.7 in the online appendix.

[23]I validate the LLM classification against human coding. I draw a 10% random sample of relevant paragraphs. The sample is classified by a German speaking student assistant. The instruction is the same prompt provided to the LLM. The overlap is generally large, but varies by topic. On average, the overlap is 81.9% for *GPT-4* and 83.3% for *GPT-3.5 turbo*. Overlap is lowest for integration, a relatively vague topic. Details are discussed in Figure A.22 in the online appendix.

7 Mechanisms

In this section, I examine mechanisms explaining lower refugee quota fulfillment of female mayors in response to the crisis. I use the TWFE specification described in Section 4. It provides a larger sample and thus more variation. In addition, using the TWFE model increases external validity relative to using municipalities with close mixed-gender races.

7.1 Differences in electoral incentives

Both, female and male mayors typically strife for reelection. Yet, a traditional literature in political science argues that gender stereotypes may shape voters' perception of female politicians (Huddy and Terkildsen, 1993; Lawless, 2004; Dolan, 2010). More recent findings indeed show that female politicians are held to higher standards by voters and that voters are biased against women (Iversen and Rosenbluth, 2010; Baskaran and Hessami, 2018; Baltrunaite *et al.*, 2019; Le Barbanchon and Sauvagnat, 2021). As a consequence, it is harder for female mayors to secure reelection. This results in differential electoral incentives for female and male mayors. Assuming that quota fulfillment is financially and politically costly, female mayors may reduce fulfillment to increase their chances of electoral success.[24] While the incentive to reduce fulfillment is small when few refugees arrive, it becomes larger during a crisis. Thus, female and male mayors differ in fulfillment only after the intake of Ukrainian refugees. To substantiate this mechanism, I examine the heterogeneity of the difference in quota fulfillment along four dimensions.[25]

[24] Several previous studies examine the political cost – measured by electoral outcomes – of refugee migration (Dustmann *et al.*, 2019; Steinmayr, 2021; Bredtmann, 2022). There is no direct evidence on the effect of Ukrainian refugees on mayoral elections in Germany, however. A financial cost-benefit assessment of refugee migration at the local level is highly context dependent. The financial burden for municipalities is highlighted e.g., by the association of German cities and municipalities (Deutscher Städte- und Gemeindebund, 2024). Taken together, refugee migration is likely perceived as politically and financially costly by mayors, at least in the short run.

[25] In the traditional literature on political budget cycles policymakers adjust fiscal policy close to elections (e.g., Rogoff, 1990; Foremny and Riedel, 2014). The prerequisite is that policymakers can regularly adjust these measures. In contrast, crises occur unexpectedly. Even in the middle of the electoral cycle strategic behavior is rational for forward-looking mayors. It is plausible that the professional full-time mayors in NRW take electoral consequences into account also outside of election years. In addition, voters must remember the mayors' behavior in times of crisis to make electoral incentives credible. Ukrainian refugees tend to stay in the municipality they arrived such that they are still visible to voters close to the next election (Siegert *et al.*, 2023).

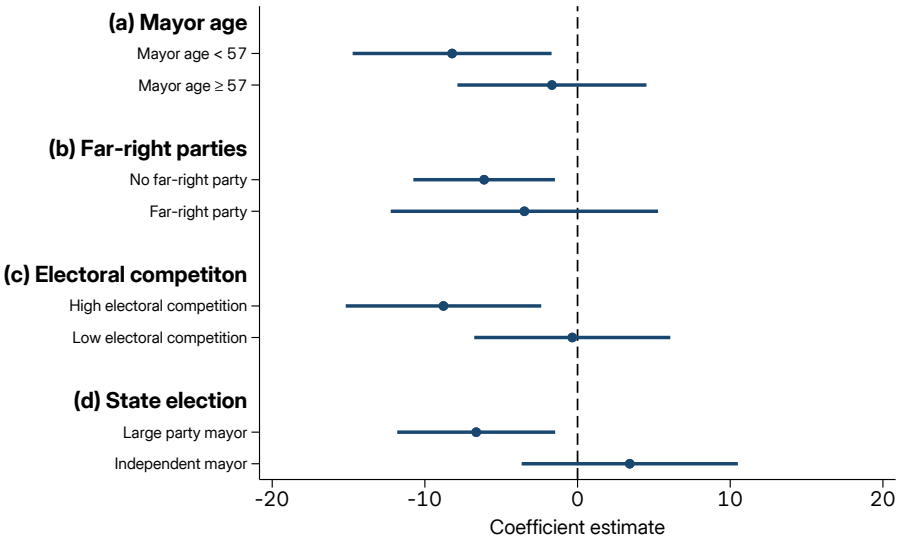


Figure 5: Mechanisms – Electoral incentives

Notes: This figure shows coefficient estimates of a version of the model described in Equation (2). Subfigure (a) shows coefficient estimates limiting the sample to below and above or equal to median age mayors (57 as of 2025). This proxies for the probability to recontest in the election of 2025. Subfigure (b) shows coefficient estimates limiting the sample to municipalities with and without far-right party (AfD, REP, and NPD) participating in the election of 2020. Subfigure (c) shows coefficient estimates limiting the sample to races for the mayorship with below and above or equal median closeness of the election. This proxies for the electoral competition mayors face. Subfigure (d) shows coefficient estimates for mayors from one or more of the large parties (CDU, SPD, Green Party, FDP) and for independent mayors. The latter have no stakes in the state election of May 2022. 95% confidence intervals are indicated in the graph.

First, seeking reelection in the first place results in different electoral incentives. Mayors planning to recontest in the 2025 municipal election may be more concerned about voters' perception of their refugee policy. Ideally, one would use institutional term limits to compare mayors who stand for reelection with those who do not (Besley and Case, 1995). There are no term limits for mayors in NRW, however. Without term limits, age and pension eligibility proxy for seeking reelection (Foremny *et al.*, 2018). In NRW, pension eligibility depends on age and time in office, but also on previous civil servant status, or university training periods. Thus, a mayors' true pension eligibility is unobservable. I have information on mayors' age, however.

I examine reelection concerns in Subfigure (a) of Figure 5. I split the sample at the median mayor age in 2025 (57 years). I compare female and male mayors within both subsamples. The difference in fulfillment is large and significant for younger mayors. It is close to zero and insignificant for older mayors that are arguably less likely to seek reelection. This pattern

suggests that whether a mayor seeks reelection affects the quota fulfillment. Results are similar when using 55 and 60 as alternative age thresholds (Figure A.23 in the online appendix).

Second, female mayors planning to recontest may reduce fulfillment to regain voters sympathizing with far-right parties. These parties have a negative stance on refugees and emphasize the burden and challenges they pose on municipalities. To test this empirically, I spit the sample into municipalities in which one of the far-right parties AfD, REP, or NPD participated in 2020 and those in which they did not. These parties favor stricter regulation of immigration and use anti-refugee rhetoric. At least one of the far-right parties participated in 34% of municipalities. Subfigure (b) of Figure 5 shows the results. The difference in fulfillment is only somewhat smaller in municipalities with a far-right party. Thus, the electoral pressure from far-right parties is not a plausible mechanism.

Third, female mayors may face different incentives because of strong electoral competition. Mayoral elections in NRW are typically contested.[26] Even incumbent mayors likely face one or more competitors in the next election. There is heterogeneity in the extent of electoral competition, however. Female mayors may reduce quota fulfillment to increase their electoral chances if they expect stronger electoral competition. Female mayors' expectations about the next election are unknown. Instead, I use the absolute margin of victory in races for the mayorship in 2020 to proxy future electoral competition (Chauvin and Tricaud, 2024). If mayors perceive competition in the previous election a good predictor for competition in future elections, the difference in quota fulfillment should be larger among mayors with strong electoral competition in 2020.[27]

To examine electoral competition empirically, I split the sample at the median absolute margin of victory in 2020. I compare fulfillment of female and male mayors before and after

[26] In 43 municipalities (10.9%) the election of the mayor was uncontested in 2020. These municipalities are ideal to study the role of electoral competition and the absence thereof. Only one of these municipalities has a female mayor, however. Thus, I cannot examine the difference in fulfillment in municipalities with uncontested elections.

[27] To test this crucial assumption, I obtain data on the local election of 2014/15. I calculate the margin of victory and estimate its correlation with the margin of victory in the 2020 election. Results are presented in Table A.10 in the online appendix. Indeed, the closeness of the mayoral race in 2014/15 correlates significantly with the closeness in 2020 – conditional on municipality and mayor controls. Thus, it is conceivable that mayors make predictions based on past electoral competition.

February 2022 in both subsamples. Results are presented in Subfigure (c) of Figure 5. The difference in fulfillment is significant and larger than the baseline among mayors with high electoral competition in 2020. It is insignificant and virtually zero for elections with low electoral competition. The difference between the two coefficients is significant at the 10%-level. Results are unchanged when excluding large cities, which tend to have closer races for the mayorship (Figure A.23 in the online appendix). Also, results are similar when interacting the treatment with the continuous absolute margin of victory (Figure A.24 in the online appendix). Taken together, only female mayors more likely to face strong electoral competition in 2025 reduce quota fulfillment.

Fourth, state elections create electoral incentives. In particular, there were state elections in NRW in May 2022. The proximity of this election to the intake of Ukrainian refugees results in immediate electoral incentives. Mayors from large parties (CDU, SPD, Green Party, FDP) have the incentive to reduce quota fulfillment to shield their party in state elections, even if typically not running themselves. About 20% of mayors run independently or for local lists. These mayors have no electoral stakes in the state election. Consequently, there should be no difference in quota fulfillment among these mayors. The result of the respective sample split is presented in Subfigure (d) of Figure 5. Indeed, the gap in quota fulfillment is large and significant only among mayors from large parties. This complements the relevance of electoral incentives in other elections.

In summary, standing for reelection, the prospect of strong electoral competition, and electoral stakes in the state election can explain the difference in fulfillment between female and male mayors. The relevance of gender-specific electoral incentives as opposed to gender-specific characteristics is in line with Chauvin and Tricaud (2024).[28] In particular, lower fulfillment does not contradict the finding that women tend to be more risk-averse in some settings (Croson and Gneezy, 2009). For one, there is evidence that female politicians are in fact less risk-averse relative to men (Magalhães and Pereira, 2024). Also, it is unclear if lower fulfill-

[28] Pulejo and Querubín (2021) also show that electoral incentives matter for the stringency of COVID-19 measures. They do not consider the gender dimension in their analysis, however.

ment is risky as there is no enforcement. On the contrary, hosting refugees can be risky as its fiscal and political consequences may be substantial. In addition, professional mayors are a selected group that is unlikely to differ in personal characteristics (see Table 4). While Chauvin and Tricaud (2024) find that female leaders implement more stringent measures to reduce the risks of COVID-19, they also argue that risk-aversion is unlikely to explain this behavior.

7.2 Alternative mechanisms

7.2.1 Council discussion sentiment

An alternative explanation for the difference in fulfillment are inherently different preferences for migration policy of female mayors. Female mayors may view refugees less favorably and consequently do not fulfill the quota. This aversion may surface only if refugee numbers are high, i.e., in times of crisis. There is no information on mayors' opinion on refugees at the local level. As mayors prepare the agenda and chair the meetings, the sentiment in council discussions on refugee topics may reveal their preferences, however.

To test this mechanism, I use the corpus of council minutes described in Section 6.3. I focus on relevant paragraphs, i.e., paragraphs that include refugee-terms. I compare the terms in the paragraphs against the sentiment dictionary on political language by Rauh (2018). This dictionary approach results in a binary classification of positive and negative terms. The paragraph-level sentiment score is calculated as the logarithm of the ratio of positive to negative terms. The setting at hand has two limitations. First, the small corpus of council minutes makes – arguably more flexible – supervised machine learning approaches unfeasible. Second, council minutes are summaries by an administrative clerk. This limits the variation in sentiment as minutes are written in an abbreviated and factual style.

I use the sentiment score of paragraphs about refugees as dependent variable in a TWFE model. Table 3 shows the results. Sentiment in refugee discussions is not different for female and male mayors in response to the crisis. The coefficient is small and far from conventional significance levels. The coefficient remains insignificant when controlling for minute and para-

Table 3: MECHANISM – FEMALE MAYORS AND REFUGEE SENTIMENT

	Dep. var.: Sentiment score (logistic)			
	(1)	(2)	(3) Meeting average	(4) Mayor mentioned
Female mayor × Post	0.038	0.044	-0.005	0.050
	(0.109)	(0.108)	(0.136)	(0.406)
Mean (SD)	0.57 (0.91)	0.57 (0.91)	0.59 (0.72)	0.58 (0.97)
Time FE	✓	✓	✓	✓
Municipality FE	✓	✓	✓	✓
Minutes controls		✓	✓	✓
Municipalities	100	100	98	63
N	2,944	2,944	775	349

Notes: This table reports results from regressions that relate an indicator for female mayor (*Female mayor*) interacted with an indicator for the time after the intake of refugees from Ukraine (*Post*) to sentiment scores derived from council minutes' paragraphs. The sentiment score is the logarithm of the ratio of positive to negative words. In Model (1), the dependent variable is the sentiment score of all relevant paragraphs. Model (2) includes controls for paragraph and minutes length. In Model (4), the dependent variable is the meeting average of paragraphs' sentiment score. Model (4) limits the sample to paragraphs in which the mayor is mentioned. The row entitled Mean (SD) reports the mean and standard deviation of the dependent variable. Stars indicate significance levels at 10%(*), 5%(**), and 1%(***). Standard errors in parentheses are clustered at the municipality level.

graph length and when averaging sentiment scores at the council meeting. So far, all paragraphs are included, irrespective of who "speaks" in the paragraph. To exclude the possibility that female mayors view refugees less favorably, the mayors' sentiment is most informative, however. I limit the sample to paragraphs that include the mayors' full name, last name, or the denotation *(Ober-)Bürgermeister/(Ober-)Bürgermeisterin* (the German term for mayor). These are paragraphs in which the mayor speaks or is mentioned by the speaker. There is no difference in sentiment between female and male mayors in response to the crisis in this subset of paragraphs. Thus, the difference in fulfillment is not driven by mayors' perception of refugees.

7.2.2 Mayors' party affiliation

Ferwerda and Riaz (2023) show that partisanship matters for the opening of refugee centers in Germany. Thus, party affiliation may explain the difference in quota fulfillment between female and male mayors. Figure A.25 in the online appendix shows the party affiliation of mayors. Most mayors in NRW are from a large party or are the joint candidate for several parties. About 20% run as individuals or for a local list. Table 4 shows the results of t-tests comparing male and female mayors' party affiliation. Female mayors run significantly more

Table 4: BALANCE OF MAYOR CHARACTERISTICS

Variable	Female	Male	Difference	Std. Errors	N
Age	50.74	50.46	-0.279	1.183	393
University degree	0.82	0.83	0.007	0.055	383
Public administration	0.18	0.39	0.206***	0.069	388
Incumbent	0.47	0.54	0.066	0.072	396
# Children	1.86	2.16	0.299*	0.176	298
Immigrant origin	0.02	0.01	-0.006	0.016	396
SPD or Green Party	0.35	0.24	-0.106*	0.063	396
CDU or FDP	0.18	0.46	0.282***	0.069	396
Independent	0.18	0.19	0.010	0.056	396

Notes: This table shows t-tests for differences in personal characteristics between female and male mayors. Stars indicate significance levels at 10%(*), 5%(**), and 1%(***).

often for the center-left SPD and the Green Party. In turn, male mayors are significantly more often in the conservative CDU and the liberal FDP.

To alleviate concerns that party affiliation drives the difference in quota fulfillment, I examine mayors' party affiliation explicitly. I separate the sample into mayors that run for the SPD, the Green Party, or for both jointly. Models (1) and (2) of Table A.11 show a gender gap in fulfillment among both, left and right/independent mayors. If anything, the gap is slightly larger, but imprecisely estimated, for left mayors. The difference in the gender gap is insignificant, however. Thus, party affiliation does not explain the gender gap in quota fulfillment.[29]

7.2.3 Experience in public administration

The fulfillment gap may reflect differences in experience in the public administration. To test this mechanism, I use hand-collected information on mayors' education, prior occupation, and incumbency. Details are presented in Section A.1 in the online appendix. Table 4 shows that female and male mayors do not differ significantly in incumbency, which proxies experience in office. They also do not differ in the probability of having a university degree (83% of mayors have a university degree). Female mayors significantly less often have an occupational background in public administration, however. I examine if differences in fulfillment are due to less

[29]This exercise alleviates concerns that political alignment between mayors and the head of the local authority in charge of refugee allocation drives the results. During the sample period, the most senior politician in the local authority of Arnsberg is a CDU member.

experience in managing the administration. I examine the difference in fulfillment separately for mayors with and without a public administration background. Results are shown in Models (3) and (4) of Table A.11. The difference in fulfillment is similar for mayors with and without public administration background. If anything, the difference is larger for female mayors with a public administration background. In summary, there is no evidence that experience explains female mayors' lower fulfillment.

7.2.4 Labor market competition

The intake of Ukrainian refugees has consequences for local labor markets. Contrary to regular asylum seekers, Ukrainian refugees are allowed to work relatively soon after their arrival. As of October 2022, 17% of Ukrainian refugees were employed. A vast majority intends to seek employment. Most Ukrainian refugees are female (Brücker *et al.*, 2023). Thus, female mayors may be concerned about an increase in labor market competition for women due to Ukrainian refugees. As a consequence, they may reduce the number of refugees to protect native female employees.

To test this mechanism, I use municipality-level data on female unemployment as of 2019. I use unemployed women relative to overall population as a measure for tight labor markets.[30] I construct an indicator that is one if female unemployment is above or equal to the median. The difference in fulfillment should be larger for municipalities with tight labor markets for women. Results are presented in Models (1) and (2) of Table A.12 in the online appendix. The difference in fulfillment is somewhat larger for tight labor markets. The coefficient is not statistically different from the coefficient for less tight markets, however. Thus, female labor market competition is not a plausible mechanism for the difference in fulfillment in response to the crisis.

[30] I use this measure as there is no municipality-level data on the unemployment rate. In particular, there is no information on the potential labor force. Instead, I scale unemployment by overall population.

7.2.5 Local fiscal capacity

Hosting refugees creates a substantial financial burden. In particular, municipalities need to adjust the provision of local public services. At the same time, female and male mayors may handle the scarcity of resources differently in response to the crisis. If the difference in quota fulfillment is larger in financially strained municipalities this suggests that female mayors adjust their behavior and use available leeway to avoid the financial burden of hosting refugees.

To examine this mechanism, I use data on tax revenues and municipal spending provided by the Statistical Office of NRW. I use the share of tax revenues relative to total spending as proxy for local fiscal capacity. It captures the spending share that municipalities can cover with their own revenues. I use data from 2019, i.e., the last year before the 2020 election. Thereafter, fiscal capacity is endogenous to female mayors. I split the sample at the median share of tax revenue relative to total spending. Models (3) and (4) of Table A.12 in the online appendix shows the results. There is a difference in quota fulfillment between female and male mayors irrespective of fiscal capacity. The coefficients are similar in size to the baseline. Thus, differences in fiscal capacity do not explain the lower fulfillment of female mayors.

7.2.6 Female councilors

Mayors need the support of councilors to enact policies. I next examine the role of female councilors in shaping migration policies in times of crisis. The female councilor share can be interpreted in two ways. First, female councilors are natural allies of female mayors seeking support for their policies. Second, the female councilor share can proxy for gender bias in the electorate. More gender equal municipalities tend to have a higher share of female councilors (Chauvin and Tricaud, 2024). To examine this mechanism, I split the sample at the median share of female councilors (26.92%) and examine if quota fulfillment of female mayors varies with the share of female councilors. Results are presented in Models (5) and (6) of Table A.12 in the online appendix. There is no evidence for heterogeneous effects due to the female councilor share.

8 Conclusion

I examine the role of female leaders during a migration crisis. To study policy differences between female and male mayors, the intake of refugees from Ukraine after February 2022 provides a compelling setting. I use monthly data on allocation quota fulfillment in the German state of NRW. Results from TWFE specifications show that female mayors fulfill the quota substantially less. Fulfillment is about 5 lower for female mayors. The difference is not driven by other personal characteristics, municipality characteristics, or refugee preferences. Electoral incentives matter for fulfillment, however. Female mayors that face strong electoral competition and are more likely to seek reelection exhibit lower quota fulfillment.

To what extent do these results generalize to other refugee waves? Unfortunately, limited data availability does not allow me to study the intake of refugees in 2015/16. Refugees arriving during this time differ substantially in their demographic characteristics and cultural similarity to Germany. External validity is further limited by the unique context of the first major war in Europe for decades. Indeed, attitudes are more positive towards Ukrainian refugees than to previous refugees from e.g., Syria (Bansak *et al.*, 2023). This has important consequences in light of electoral incentives as key mechanism. Future refugees that are perceived less favorably by natives may widen the gap in quota fulfillment by female mayors seeking reelection.

The strategic adjustment of quota fulfillment in response to electoral incentives has important implications for the institutional design of refugee allocation. As refugee migration is likely an ongoing issue, the adequate allocation of refugees is key in obtaining natives' support. As highlighted in this paper, administrative quotas cannot accomplish adequate allocation if enforcement is absent and electoral incentives are relevant. Instead, policymakers should create an incentive-compatible allocation process. That is, local policymakers facing electoral constraints should have a compensating incentive to host refugees. Such incentives include but extend beyond adequate financial support for municipalities.

References

APESTEGUIA, J., FUNK, P. and IRIBERRI, N. (2013). Promoting rule compliance in daily-life: Evidence from a randomized field experiment in the public libraries of Barcelona. *European Economic Review*, **64**, 266–284.

BALTRUNAITE, A., CASARICO, A., PROFETA, P. and SAVIO, G. (2019). Let the voters choose women. *Journal of Public Economics*, **180**.

BAMF (2017). *Das Bundesamt in Zahlen 2016 – Asyl, Migration und Integration*. Bundesamt für Migration und Flüchtlinge Nürnberg.

BANSAK, K., HAINMUELLER, J. and HANGARTNER, D. (2023). Europeans' support for refugees of varying background is stable over time. *Nature*, **620**, 849–854.

BARTEL, A. P. (1989). Where do the new U.S. immigrants live? *Journal of Labor Economics*, **7** (4), 371–391.

BASKARAN, T. and HESSAMI, Z. (2018). Does the election of a female leader clear the way for more women in politics? *American Economic Journal: Economic Policy*, **10**, 95–121.

— and — (2025). Women in political bodies as policymakers. *Review of Economics and Statistics*, forthcoming.

BATUT, C. and SCHNEIDER-STRAWCZYNSKI, S. (2021). Rival guests or defiant hosts? The local economic impact of hosting refugees. *Journal of Economic Geography*, **22** (2), 327–350.

BESLEY, T. (2005). Political selection. *Journal of Economic Perspectives*, **19** (3), 43–60.

— and CASE, A. (1995). Does electoral accountability affect economic policy choices? Evidence from gubernatorial term limits. *Quarterly Journal of Economics*, **110** (3), 769–798.

— and COATE, S. (1997). An economic model of representative democracy. *Quarterly Journal of Economics*, **112**, 85–114.

BHALOTRA, S. and CLOTS-FIGUERAS, I. (2014). Health and the political agency of women. *American Economic Journal: Economic Policy*, **6** (2), 164–197.

BOJANOWSKI, P., GRAVE, E., JOULIN, A. and MIKOLOV, T. (2016). Enriching word vectors with subword information. *arXiv: 1607.04606*.

BRACCO, E., DE PAOLA, M., GREEN, C. P. and SCOPPA, V. (2018). The effect of far right parties on the location choice of immigrants: Evidence from Lega Nord mayors. *Journal of Public Economics*, **166**, 12–26.

BREDTMANN, J. (2022). Immigration and electoral outcomes: Evidence from the 2015 refugee inflow to Germany. *Regional Science and Urban Economics*, **96**, 103807.

BROLLO, F. and TROIANO, U. (2016). What happens when a woman wins a close election? Evidence from Brazil. *Journal of Development Economics*, **122**, 28–45.

BRUCE, R., CAVGIAS, A., MELONI, L. and REMÍGIO, M. (2022). Under pressure: Women's leadership during the COVID-19 crisis. *Journal of Development Economics*, **154**, 1027–61.

BRÜCKER, H., ETTE, A., GRABKA, M. M., KOSYAKOVA, Y., NIEHUES, W., ROTHER, N., SPIESS, C. K., ZINN, S., BUJARD, M., CARDOZO, A., DÉCIEUX, J. P., MADDOX, A., MILEWSKI, N., NADERI, R., SAUER, L., SCHMITZ, S., SCHWANHÄUSER, S., SIEGERT, M., TANIS, K. and STEINHAUER, H. W. (2023). Geflüchtete aus der Ukraine in Deutschland: Ergebnisse der ersten Welle der IAB-BiB/FReDA-BAMF-SOEP Befragung. *IAB-Forschungsbericht*.

CALONICO, S., CATTANEO, M. D. and TITIUNIK, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, **82**, 2295–2326.

CARD, D. (2001). Immigrant inflows, native outflows, and the local market impacts of higher immigration. *Journal of Labor Economics*, **19** (1), 22–64.

CAROZZI, F. and GAGO, A. (2023). Who promotes gender-sensitive policies? *Journal of Economic Behavior & Organization*, **206**, 371–405.

CASARICO, A., LATTANZIO, S. and PROFETA, P. (2022). Women and local public finance. *European Journal of Political Economy*, **72**, 102096.

CATTANEO, C., BEINE, M., FRÖHLICH, C. J., KNIVETON, D., MARTINEZ-ZARZOSO, I., MASTRORILLO, M., MILLOCK, K., PIGUET, E. and SCHRAVEN, B. (2019). Human migration in the era of climate change. *Review of Environmental Economics and Policy*, **13** (2), 189–206.

CHATTOPADHYAY, R. and DUFLO, E. (2004). Women as policy makers: Evidence from a randomized experiment in India. *Econometrica*, **72**, 1409–43.

CHAUVIN, J. P. and TRICAUD, C. (2024). Gender and electoral incentives: Evidence from crisis response. *NBER Working Paper No. 32410*.

CLARKE, D. (2017). Estimating difference-in-differences in the presence of spillovers. *MPRA Paper No. 81604*.

CLOTS-FIGUERAS, I. (2012). Are female leaders good for education? Evidence from India. *American Economic Journal: Applied Economics*, **4**, 212–44.

COLONNELLI, E., PREM, M. and TESO, E. (2020). Patronage and selection in public sector organizations. *American Economic Review*, **110** (10), 3071–3099.

CROSON, R. and GNEEZY, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, **47** (2), 448–474.

DAMM, A. P. (2007). Determinants of recent immigrants' location choices: Quasi-experimental evidence. *Journal of Population Economics*, **22** (1), 145–174.

DANIELE, G. and GIOMMONI, T. (2021). Corruption under austerity. *CESifo Working Paper No. 8909*.

DANZER, N., GARCIA-TORRES, S., STEINHARDT, M. F. and STELLA, L. (2024). Women in political power and school closure during COVID times. *Economic Policy*, **39** (120), 765–810.

DE CHAISEMARTIN, C. and D'HAULTFŒUILLE, X. (2024). Difference-in-differences estimators of intertemporal treatment effects. *Review of Economics and Statistics*, pp. 1–45.

DESTATIS (2023). Starker Zuwachs an ukrainischen Staatsbürgern seit Ende Februar 2022. Retrieved last on 30.09.2025 from https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/ukraine.html.

DEUTSCHER STÄDTE- UND GEMEINDEBUND (2024). Kommunen fordern mehr Unterstützung bei Integration und Begrenzung der illegalen Migration. Retrieved last on 30.09.2025 from https://www.dstgb.de/publikationen/pressemitteilungen/kommunen-fordern-mehr-unterstuetzung-bei-integration-und-begrenzung-der-illegalen-migration/.

DE CHAISEMARTIN, C. and D'HAULTFŒUILLE, X. (2022). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey. *The Econometrics Journal*, **26** (3), 1–30.

DOLAN, K. (2010). The impact of gender stereotyped evaluations on support for women candidates. *Political Behavior*, **32** (1), 69–88.

DOWNS, A. (1957). *An economic theory of democracy*. Reading, MA: Addison-Wesley.

DUSTMANN, C., SCHÖNBERG, U. and STUHLER, J. (2016). The impact of immigration: Why do studies reach such different results? *Journal of Economic Perspectives*, **30** (4), 31–56.

—, VASILJEVA, K. and DAMM, A. P. (2019). Refugee migration and electoral outcomes. *Review of Economic Studies*, **86** (5), 2035–2091.

D'ATTOMA, J. W., VOLINTIRU, C. and MALÉZIEUX, A. (2020). Gender, social value orientation, and tax compliance. *CESifo Economic Studies*, **66** (3), 265–284.

ESLAVA, F. (2024). Conflict and female leadership: Evidence from Colombia. *Mimeo*.

FERREIRA, F. and GYOURKO, J. (2014). Does gender matter for political leadership? The case of U.S. mayors. *Journal of Public Economics*, **112**, 24–39.

FERWERDA, J. and RIAZ, S. (2023). Shielding voters? How partisanship shapes the placement of refugee housing facilities. *Mimeo*.

FOREMNY, D., FREIER, R., MOESSINGER, M.-D. and YETER, M. (2018). Overlapping political budget cycles. *Public Choice*, **177** (1–2), 1–27.

— and RIEDEL, N. (2014). Business taxes and the electoral cycle. *Journal of Public Economics*, **115**, 48–61.

FRATESI, U., PERCOCO, M. and PROIETTI, P. (2018). Who welcomes them? Determinants of the spatial distribution of asylum seekers in Italy. *Regional Studies*, **53** (1), 118–130.

FREIER, R. and THOMASIUS, S. (2015). Voters prefer more qualified mayors, but does it matter for public finances? Evidence for Germany. *International Tax and Public Finance*, **23** (5), 875–910.

GALASSO, V., PONS, V., PROFETA, P., BECHER, M., BROUARD, S. and FOUCAULT, M. (2020). Gender differences in COVID-19 attitudes and behavior: Panel evidence from eight countries. *Proceedings of the National Academy of Sciences*, **117** (44), 27285–27291.

GAMALERIO, M. and NEGRI, M. (2023). Not welcome anymore: the effect of electoral incentives on the reception of refugees. *Journal of Economic Geography*, **23** (4), 901–920.

GENTZKOW, M., KELLY, B. and TADDY, M. (2019). Text as Data. *Journal of Economic Literature*, **57** (3), 535–574.

GEYS, B. and SØRENSEN, R. J. (2019). The impact of women above the political glass ceiling: Evidence from a Norwegian executive gender quota reform. *Electoral Studies*, **60**, 102050.

GREMBI, V., NANNICINI, T. and TROIANO, U. (2016). Do fiscal rules matter? *American Economic Journal: Applied Economics*, **8** (3), 1–30.

HESSAMI, Z., HÄCKER, T. and THOMAS, M. (2025). Public administrators as politicians in office. *International Tax and Public Finance*, forthcoming.

— and LOPES DA FONSECA, M. (2020). Female political representation and substantive effects on policies: A literature review. *European Journal of Political Economy*, **63**.

HUDDY, L. and TERKILDSEN, N. (1993). Gender stereotypes and the perception of male and female candidates. *American Journal of Political Science*, **37** (1), 119.

IVERSEN, T. and ROSENBLUTH, F. (2010). *Women, Work, and Politics: The Political Economy of Gender Inequality.* New Haven, CT: Yale University Press.

KORINEK, A. (2023). Generative AI for economic research: Use cases and implications for economists. *Journal of Economic Literature*, **61** (4), 1281–1317.

KOST, A. (2022). Bürgermeister in Nordrhein-Westfalen. In *Karrierechance Bürgermeisteramt – Leitfaden für die erfolgreiche Kandidatur und Amtsführung*, Paul Witt.

KREITMEIR, D. (2023). The political economy of socioenvironmental conflict: Evidence from Peru. *Mimeo*.

LAWLESS, J. L. (2004). Women, war, and winning elections: Gender stereotyping in the post-September 11th era. *Political Research Quarterly*, **57** (3), 479.

LE BARBANCHON, T. and SAUVAGNAT, J. (2021). Electoral competition, voter bias, and women in politics. *Journal of the European Economic Association*, **20** (1), 352–394.

LIPPMANN, Q. (2022). Gender and lawmaking in times of quotas. *Journal of Public Economics*, **207**, 104610.

MACKINNON, J. G. and WEBB, M. D. (2020). Randomization inference for difference-in-differences with few treated clusters. *Journal of Econometrics*, **218** (2), 435–450.

MAGALHÃES, P. C. and PEREIRA, M. M. (2024). Women running for office are less risk averse than men: Evidence from Portugal. *The Journal of Politics*, **86** (3), 1093–1097.

MARSHALL, J. (2024). Can close election regression discontinuity designs identify effects of winning politician characteristics? *American Journal of Political Science*, **68** (2), 494–510.

MCCRARY, J. (2008). Manipulation of the running variable in the regression discontinuity design: a density test. *Journal of Econometrics*, **142**, 698–714.

OSBORNE, M. J. and SLIVINSKI, A. (1996). A model of political competition with citizen-candidates. *Quarterly Journal of Economics*, **111**, 65–96.

PICCHETTI, P., PINTO, C. C. X. and SHINOKI, S. T. (2024). Difference-in-discontinuities: Estimation, inference, and validity tests. *Mimeo*.

PULEJO, M. and QUERUBÍN, P. (2021). Electoral concerns reduce restrictive measures during the COVID-19 pandemic. *Journal of Public Economics*, **198**, 104387.

RAMBACHAN, A. and ROTH, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, **90** (5), 2555–2591.

RAUH, C. (2018). Validating a sentiment dictionary for German political language – a workbench note. *Journal of Information Technology & Politics*, **15** (4), 319–343.

ROGOFF, K. (1990). Equilibrium political business cycles. *American Economic Review*, **80** (21-36).

RUIST, J. (2020). The fiscal aspect of the refugee crisis. *International Tax and Public Finance*, **27** (2), 478–492.

RWI and MICROM (2023). Sozioökonomische Daten auf Rasterebene - Scientific Use File (Welle 13).

SAUER, L., ETTE, A., STEINHAUER, H. W., SIEGERT, M. and TANIS, K. (2023). Spatial patterns of recent Ukrainian refugees in Germany: Administrative dispersal and existing ethnic networks. *Comparative Population Studies*, **48**.

SGV (2023). Gesetz über die Zuweisung und Aufnahme ausländischer Flüchtlinge (Flüchtlingsaufnahmegesetz - FlüAG). Retrieved last on 30.09.2025 from https://recht.nrw.de/lmi/owa/br_text_anzeigen?v_id=10000000000000000407.

SIEGERT, M., TANIS, K., ETTE, A. and SAUER, L. (2023). Entwicklung der Wohnsituation ukrainischer Geflüchteter in Deutschland. *BAMF-Kurzanalyse 03/2023*.

STEINMAYR, A. (2021). Contact versus exposure: Refugee presence and voting for the far right. *Review of Economics and Statistics*, **103** (2), 310–327.

UNHCR (2022). Ukraine situation flash update #1. Retrieved last on 30.09.2025 from https://data.unhcr.org/en/documents/details/91208.

VOIGTLÄNDER, N. and VOTH, H.-J. (2012). Persecution perpetuated: The medieval origins of anti-semitic violence in Nazi Germany. *Quarterly Journal of Economics*, **127** (3), 1339–1392.

WESTDEUTSCHER RUNDFUNK KÖLN (2015). WDR-Umfrage zu Flüchtlingen: Weniger Turnhallen mit Flüchtlingen belegt. Retrieved last on 30.09.2025 via Wayback Machine from https://web.archive.org/web/20151202114026/http://www1.wdr.de/themen/aktuell/fluechtlinge-karte-100.html.

ZUCHOWSKI, D. (2025). Migration response to an immigration shock: Evidence from Russia's aggression against Ukraine. *Journal of Economic Geography*, **25** (1), 21–40.

Online appendix

A.1 Election data and mayor characteristics

Data for the election of September 2020 was provided by the Statistical Office of NRW. Data was provided in an Excel file instead of the publicly available PDF upon request. Data includes only the results for the regular election in September 2020. Information on additional 16 municipalities with an extraordinary election prior to 2020 were hand-collected from publicly available sources. Information was available for all 16 municipalities. Details on the extraordinary elections are listed in Table A.1. Overall, 1,395 candidates competed in the mayoral election of 2020.

The election data contains the names of all candidates and elected mayors. They do not contain information on gender. To avoid manual classification, I use the R–package *gender*. This package obtains a persons' gender from their first name. Before classification, I remove all titles (e.g., *Dr.* or *Prof.*). I split double names and use the first part for classification (e.g., *Hans* for *Hans-Jürgen*). This results in 411 unique candidate names to be classified (1,395 total candidates). The package offers different data sources for classification. I use the default data source, i.e., the U.S. Social Security Administration baby name data ("ssa"). As German names can have a different gender connotation, I check ambiguous cases manually and confirm gender using photographs on municipal or private homepages (e.g., *Sasha*). For some rare names the algorithm is not able to provide a solution. Again, I manually classify these names. For the 16 municipalities with extraordinary elections, I code gender manually. Visual inspection of the subsample of elected mayors shows that the adjusted classification provides 100% accuracy.

There is no official data on mayors' personal characteristics. Thus, I hand-collected information for the 396 mayors in NRW. Sources are typically personal or municipality homepages, newspaper articles, or social media profiles. Some mayor characteristics are available for all mayors. This includes immigrant origin, incumbency, and party membership. I infer immigrant origin from candidates' names. For other personal characteristics, coverage is limited. I ob-

Table A.1: LIST OF EXTRAORDINARY ELECTIONS

Municipality	Election date	Reason
Arnsberg, Stadt	18.02.2018	Promotion
Bergheim, Stadt	25.06.2017	Health
Gronau (Westf.), Stadt	24.03.2019	Asymmetric end of term
Heiligenhaus, Stadt	24.09.2017	Promotion
Hürtgenwald	28.08.2022	Recall from office
Isselburg, Stadt	14.01.2018	Asymmetric end of term
Kall	24.09.2017	Health
Kamen, Stadt	01.07.2018	Private reasons
Krfr. Stadt Duisburg	24.09.2017	Asymmetric end of term
Lage, Stadt	16.06.2019	Private reasons
Lindlar	04.03.2018	Asymmetric end of term
Nachrodt-Wiblingwerde	15.04.2018	Asymmetric end of term
Rietberg, Stadt	16.09.2018	Asymmetric end of term
Schleiden, Stadt	18.11.2018	Asymmetric end of term
Schwerte, Stadt	04.03.2018	Promotion
Stolberg (Rhld.), Stadt	16.06.2019	Asymmetric end of term
Wesseling	13.11.2022	Health
Windeck	18.11.2018	Asymmetric end of term

tain year of birth for most mayors (393 out of 396). Education and occupation are also widely available. Coverage is sparser for the number of children and marital status.

From their occupation before the mayorship I infer if mayors have a background in public administration (Hessami *et al.*, 2025). The indicator variable is coded one if the occupation contains the stem *verwaltungs* (≈ administrative). This is typically a part of the job description in occupations linked to public administration. In addition, I inspect occupations visually and code the indicator as one for the following occupations: *Leiter des Ordnungsamtes*, *Kämmerer*, *Stadtdirektor*, *Städtischer Baurat*, *Teamleiter Kommunales Rechenzentrum*, *Sachbearbeiter im Bauamt*, *Leiter Fachbereich Ordnung, Soziales, Jugend und Sport*, *Beamter Kommunales Rechenzentrum Niederrhein*, and *Gemeindekämmerer*.

A.2 Refugee data

Data on the number of refugees and the allocation quota is provided by the local authority of Arnsberg (*Bezirksregierung Arnsberg*). Data is publicly available from April 2018 on the website of the local authority.[31] Data is provided weekly in PDF format. I extract the tables from the PDF files and transform them to datasets in an automated way. Data collection of PDF

[31] https://www.bra.nrw.de/integration-migration/fluechtlinge-nrw/informationen-fuer-kommunen/zuweisung-nach-dem-fluechtlingsaufnahmegesetz

Table A.2: REFUGEE MEASUREMENT – FEMALE MAYORS AND FULFILLMENT

				Dep. var.: Quota fulfillment (in %)	
	(1) I(Missing)	(2) I(Div. 10)	(3) I(Div. 100)	(4) Drop missing	(5) First months
Female mayor × Post	-0.001	0.012	0.002	-4.663**	-4.800**
	(0.010)	(0.012)	(0.004)	(2.292)	(2.259)
Mean (SD)	0.02 (0.13)	0.10 (0.30)	0.01 (0.09)	96.48 (19.22)	95.81 (18.26)
Time FE	✓	✓	✓	✓	✓
Municipality FE	✓	✓	✓	✓	✓
Mun	396	396	396	396	396
N	59,004	59,004	59,004	58,003	13,068

Notes: This table reports additional robustness checks on the quality of refugee measurement. It reports results from regressions that relate an indicator for female mayor (*Female mayor*) interacted with an indicator for the time after the intake of refugees from Ukraine (*Post*) to different outcomes. Weekly data is used in all models. In Model (1), the dependent variable is an indicator that is one for missing or delayed reports. In Model (2), the dependent variable is an indicator that is one if the reported refugee number is divisible by 10 without remainder. In Model (3), the dependent variable is an indicator that is one if the reported refugee number is divisible by 100 without remainder. In Model (4), all missing or delayed reports are dropped from the sample. The dependent variable is quota fulfillment (in%). Model (5) excludes the months March 2022, April 2022, and May 2022. Again, the dependent variable is quota fulfillment (in%). The row entitled Mean (SD) reports the mean and standard deviation of the dependent variable. Stars indicate significance levels at 10%(*), 5%(**), and 1%(***). Standard errors in parentheses are clustered at the municipality level.

files took place on 15.09.2023. Thus, I have data between the last week of April 2018 and the first week of September 2023. In the analysis, I focus on data between the first week of October 2020 and the first week of September 2023. This corresponds to the election period of 2020.

Municipalities are required to report the number of refugees to the local authority of Arnsberg on a weekly basis. Based on these reports and the underlying quota refugees are allocated and fulfillment is calculated. The quota is reported explicitly in the data. The quota is updated once a year when new population data is available. Area is constant for all municipalities throughout the sample period. The quota is internally consistent and adds up to 100% across municipalities in a given week. Because the quota is based on population (90%) and area (10%), it is strongly correlated with these variables and almost perfectly for population (see Figure A.26).

Upon request, the local authority of Arnsberg states several factors limiting the accuracy of data in some instances. Numbers are calculations that do not always reflect the precise number of refugees, which is typically a bit lower. Accuracy is lower in situations of an unexpected increase of refugee numbers in a short time. This was likely the case immediately after the start of the Russian invasion. Also, deductions for municipalities that host a central refugee facility,

may vary depending on capacities in these facilities. In some cases, spouses and children are allocated irrespective of the quota or there are similar humanitarian reasons for extraordinary allocations. Finally, some municipalities report data with delay or not at all in some weeks.

To ensure that the above caveats and measurement issues do not affect my findings, I run additional robustness checks in Table A.2. The PDF files provide information on missing or delayed reports from municipalities. In the sample period, information is missing or delayed in 1.7% of cases. Unsurprisingly, the share of missing or delayed reports is higher after the intake of refugees in the aftermath of the Russian invasion. There is no significant difference in the probability to report in time between female and male mayors, however (Model 1). Measurement error in the refugee numbers could also materialize in the precision of numbers. Typically, refugee numbers are precise up to the last digit. In some cases, numbers are reported as round numbers. These numbers may reflect estimates rather than precise numbers. Thus, I use an indicator that is one if the refugee number is divisible by 10 and 100 without remainder as an outcome. Models (2) and (3) show that there is no difference between female and male mayors in the precision of the refugee numbers in response to the crisis. When dropping missing or delayed reports, the main finding remains unchanged (Model 4). Also, the main finding remains unchanged when excluding the first three months after the Russian invasion from the estimation sample (Model 5). Thus, errors in refugee number reporting do not affect the difference in fulfillment between female and male mayors.

A.3 Local difference-in-differences approach

In this section, I present details on the local difference-in-differences approach. This strategy relies on the subset of municipalities that have a close mixed-gender race. As such, I cannot use arbitrarily small bandwidths that are key for the credibility of the quasi-random assignment of female mayors. Also, while internal validity is more plausible, this approach estimates the local effect near the threshold. This limits external validity. Thus, I refrain from using this approach as my main specification. Instead, its results are presented as supporting evidence.

A.3.1 Empirical strategy – Local difference-in-differences approach

In most municipalities, multiple candidates compete for the mayors' office. On average, 3.5 candidates per municipality participated in the 2020 election. If the two candidates with the highest vote share are of different gender, a mixed-gender race takes place. For the 2020 election, there were 110 such races.[32] The local randomization by close and thus unpredictable races provides plausibly exogenous allocation of the treatment.

Following Bruce *et al.* (2022), I estimate the specification described in Section 4 while limiting the sample to close mixed-gender races for the mayorship (local difference-in-differences specification):

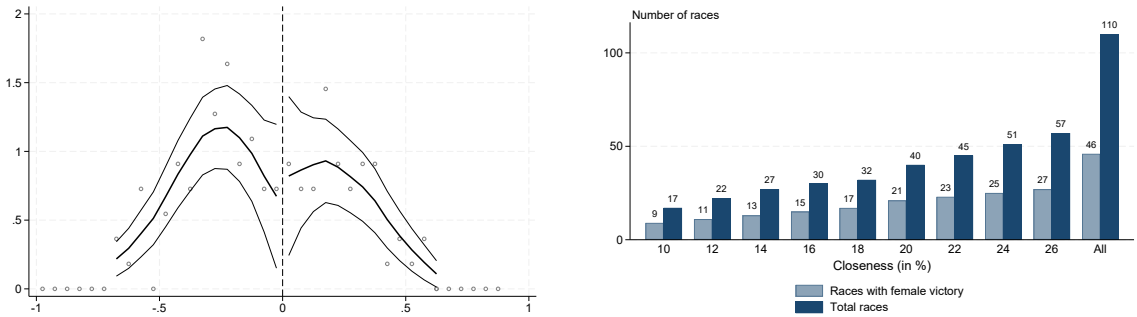
$$Y_{m,t} = \beta(Female\ mayor_{m,2020.9} \times Post_t) + f_t(FemaleMargin_m) + \gamma_m + \zeta_t + \varepsilon_{m,t}. \quad (3)$$

In this specification, I compare quota fulfillment by mayor gender before and after February 2022 limiting the sample to municipalities with a close mixed-gender race in the election of 2020.[33] I control for linear and quadratic polynomials of the running variable, i.e., the female margin of victory $f_t(FemaleMargin_m)$. The margin of victory is the difference in votes between the two candidates in the mixed-gender race relative to all valid votes (in %). The polynomial of the margin of victory is allowed to be different on either side of the 50% vote share threshold. In addition, I interact the polynomial with a linear time trend. That is, I can include municipality fixed effects γ_m although the margin of victory is time-invariant. I use a rectangular kernel and limit the sample to observations close to the threshold.[34] I show robustness to different degrees of closeness. Standard errors are clustered at the municipality level.

[32] If no candidate receives the absolute majority of votes in the first round, there are runoff elections. This is the case for about 30% of elections. In these cases, I consider the closeness of the race in the second round.

[33] This approach is similar to a difference-in-discontinuities strategy (Grembi *et al.*, 2016). The key distinction is the use of unit and time fixed effects akin to traditional difference-in-differences settings. Hence, I use the term local difference-in-differences in line with Daniele and Giommoni (2021) and Bruce *et al.* (2022). See also Colonnelli *et al.* (2020) for a similar strategy.

[34] Results are qualitatively similar when using a triangular kernel. The coefficients are slightly larger (results available upon request).



(a) McCrary plot **(b)** Number of mixed-gender races

Figure A.1: Validity of the local difference-in-differences approach

Notes: Subfigure (a) shows the McCrary plot to test for manipulation of the running variable. The log difference in height is 0.26 with a standard error of 0.69. Subfigure (b) shows the number of mixed-gender races for different levels of closeness in the 2020 election. The closeness of the race is defined as the difference in votes between the two candidates in the mixed-gender race relative to all valid votes (in %). The graph also shows the number of races won by a female candidate.

Isolating the effect of personal characteristics with close races is challenging. Marshall (2024) argues that compensating differentials may ensure the closeness of a race. That is, if female mayors are discriminated against by voters, they may end up in a close race due to counterbalancing personal characteristics. As a consequence, I can estimate the effect of female mayors, but not the effect of gender per se. This interpretation includes all compensating differentials.

A.3.2 Validity of the local difference-in-differences approach

To estimate a causal effect using local randomization from close mixed-gender races the following assumptions must hold. First, there must be no discontinuity in the density of the running variable at the threshold. German states are stable democracies, thus electoral fraud is unlikely. Reassuringly, the McCrary (2008) test shows no evidence for manipulation of the running variable. In line with this, Subfigure (a) of Figure A.1 shows that there is no discontinuity in the margin of victory at the threshold.

Second, female candidates must not be more or less likely to win close elections. Subfigure (b) of Figure A.1 shows the total number of mixed-gender races and the number of female victories. Considering all races, the victory rate of female candidates is 42%. For increas-

ingly closer races the victory rate oscillates around 50%. Thus, there is no difference in the probability of female and male victories in close races.

Third, all other municipality characteristics must vary smoothly at the threshold. That is, municipalities with female mayors must not systematically differ from municipalities with male mayors. Figure A.2 shows balance checks for municipalities with close mixed-gender races (using a bandwidth of 20%). There are no discontinuities at the threshold for municipal revenues, spending, construction activity, unemployment, share of foreigners, and population density. Also, mayor characteristics are balanced between female and male mayors elected in close races (see Table A.3).

Fourth, Picchetti *et al.* (2024) argue that potential confounding effects must be constant over time. Thus, differences between municipalities with female and male mayors other than those due to the intake of Ukrainian refugees must be similar before February 2022. They suggest to estimate RDD-coefficients for pre-treatment periods within the same regression and to test their equality using a Wald-test. Indeed, almost all coefficients for the periods before February 2022 are very similar in size. I cannot reject the null-hypothesis that all pre-treatment coefficients are equal ($F = 1.55$, $p = 0.13$).

Table A.3: BALANCE OF MAYOR CHARACTERISTICS (CLOSE RACES)

Variable	Female	Male	Difference	Std. Errors	N
Age	52.05	51.32	-0.732	2.213	40
University degree	0.86	0.89	0.038	0.107	40
Public administration	0.10	0.26	0.168	0.120	40
Incumbent	0.33	0.37	0.035	0.155	40
# Children	1.87	1.93	0.067	0.356	30
Immigrant origin	0.00	0.00	0.000	0.000	40
SPD or Green Party	0.38	0.32	-0.065	0.155	40
CDU or FDP	0.14	0.58	0.436***	0.138	40
Independent	0.10	0.00	-0.095	0.069	40

Notes: This table shows t-tests for differences in personal characteristics between female and male mayors. The sample is limited to close mixed-gender races. The bandwidth is 20%. Stars indicate significance levels at 10%(*), 5%(**), and 1%(***).

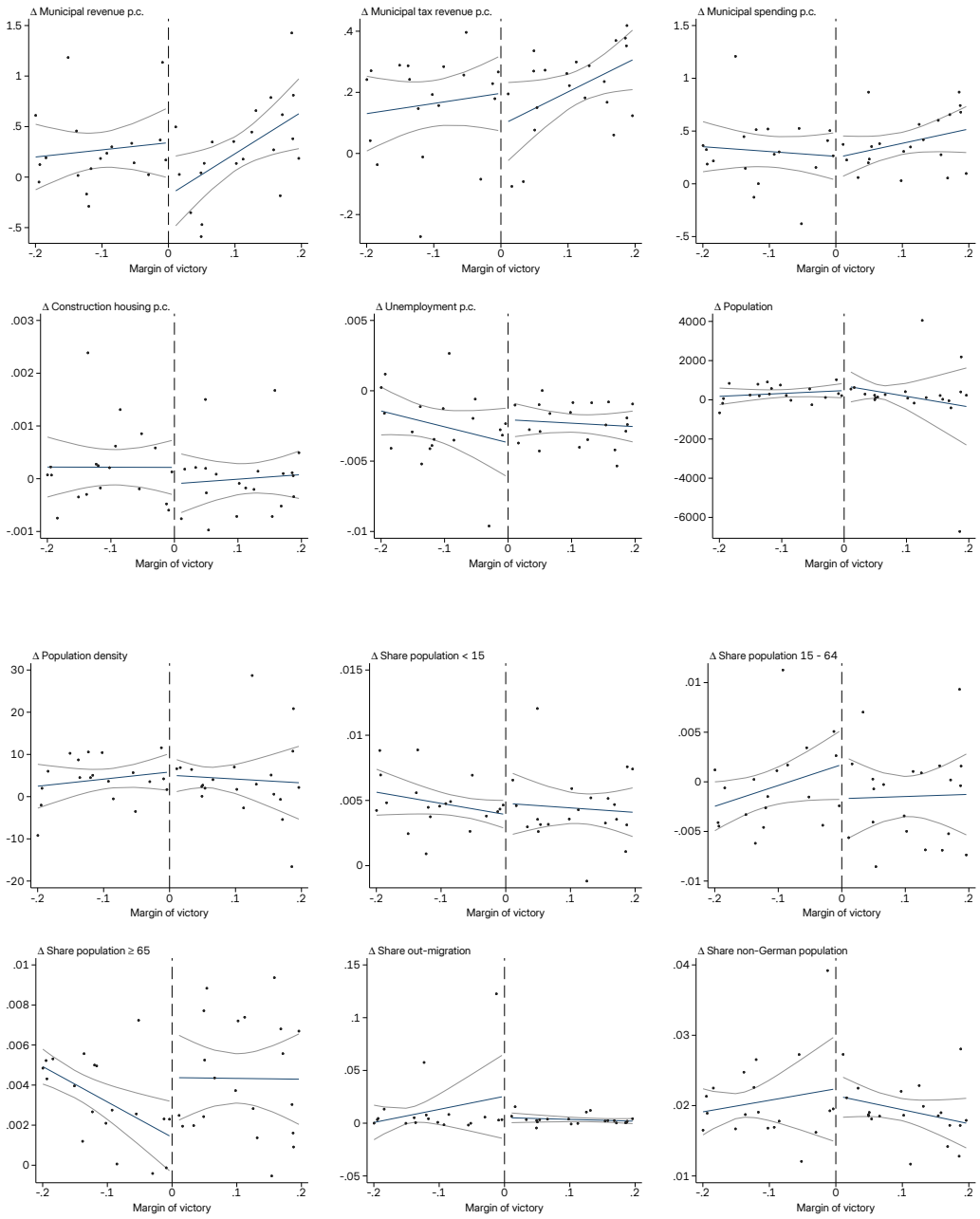
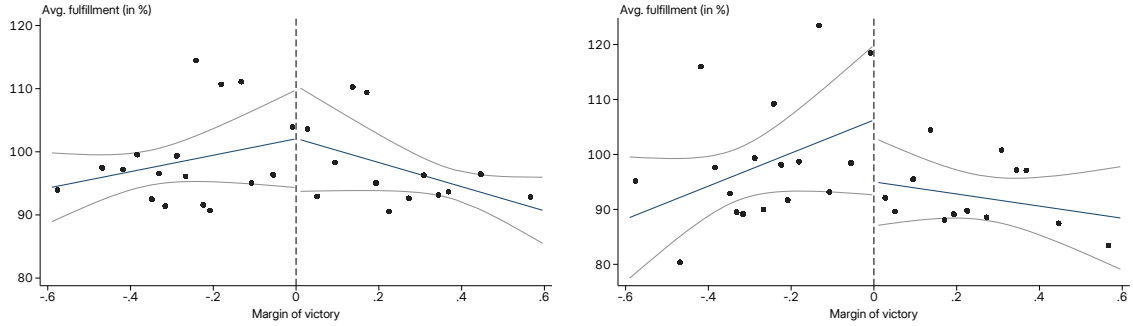


Figure A.2: Balance of municipality characteristics (II)

Notes: This graph shows coefficient estimates of local difference-in-differences specifications that relate female mayors to municipality outcomes before and after the intake of Ukrainian refugees. The y-axis shows the difference in average dependent variables between before and after February 2022. Observations within the 20% bandwidth akin to the baseline local difference-in-difference approach are used for estimation. Linear regression lines are fitted on both sides of the threshold. 95% confidence intervals are indicated in the graph.



(a) Before February 2022 **(b)** After February 2022

Figure A.3: First stage discontinuities

Notes: This graph shows regression discontinuity plots separately for before (a) and after (b) February 2022. The y-axis shows average fulfillment (in %). The x-axis shows the margin of victory for female mayors. Negative values of the margin of victory indicate a male victory, positive values indicate a female victory in the 2020 election. Data is pooled into 40 bins. Linear regression lines are fitted on both sides of the threshold. A rectangular kernel is used. 95% confidence intervals are indicated in the graph.

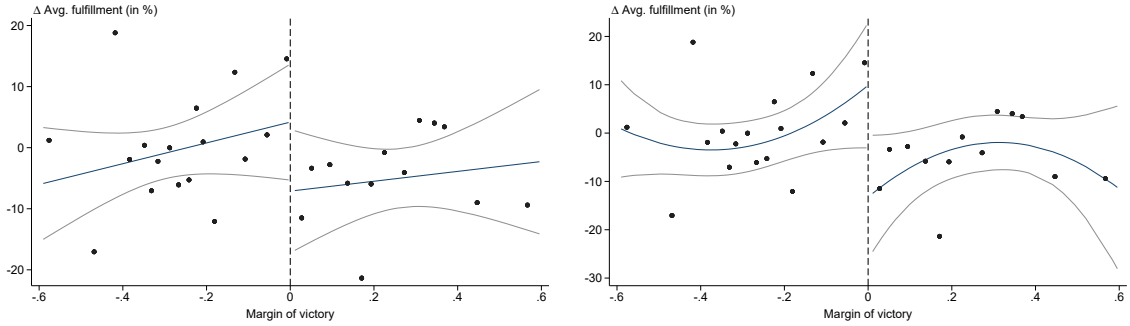
A.3.3 Local difference-in-differences results

Next, I examine the results for the local difference-in-differences specification.[35] Graphical evidence is presented in Figure A.4. There is a visible discontinuity at the threshold: female mayors fulfill the quota less in response to the crisis. Regression results are presented in Models (4) – (6) of Table 1 in the main text. Indeed, female mayors elected in close mixed-gender races fulfill the quota less relative to male mayors in response to the crisis. The coefficient is significant at the 10%-level. The effect size is considerably larger than in the TWFE specification. The effect is similar when controlling for a quadratic specification of the margin of victory and when using all 110 mixed-gender races.[36]

I show robustness to different bandwidths in Figure A.6. Focusing on mixed-gender races limits the variation available for estimation considerably. Even when using a relatively large bandwidth of 20%, there are only 40 municipalities – and thus elections – with mixed-gender

[35] Figure A.3 shows regression discontinuity plots separately for the periods before and after February 2022. There is no discontinuity at the threshold before the intake of Ukrainian refugees. There is a visible discontinuity after the intake.

[36] Similar to the TWFE specification, there must be no differential pre-trends between municipalities with and without a female mayor. This assumption is more credible because of the exogenous allocation of the female mayor treatment. Reassuringly, Figure A.5 shows that there are no systematic differences in fulfillment before the intake of refugees.



(a) Linear **(b)** Quadratic

Figure A.4: Local difference-in-differences plot

Notes: This graph shows local difference-in-differences plots in the spirit of Grembi *et al.* (2016). The y-axis shows the difference in average fulfillment (in %) between before and after February 2022. The x-axis shows the margin of victory for female mayors. Negative values of the margin of victory indicate a male victory, positive values indicate a female victory in the 2020 election. Data is pooled into 30 bins. Linear (a) and quadratic (b) regression lines are fitted on both sides of the threshold. A rectangular kernel is used. 95% confidence intervals are indicated in the graph.

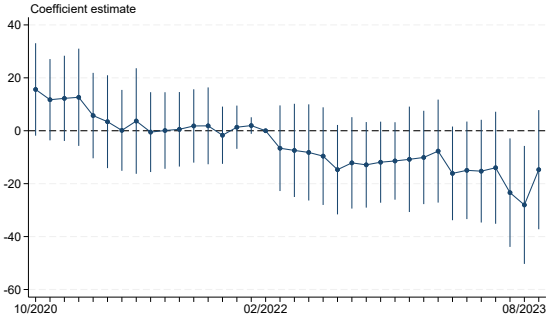
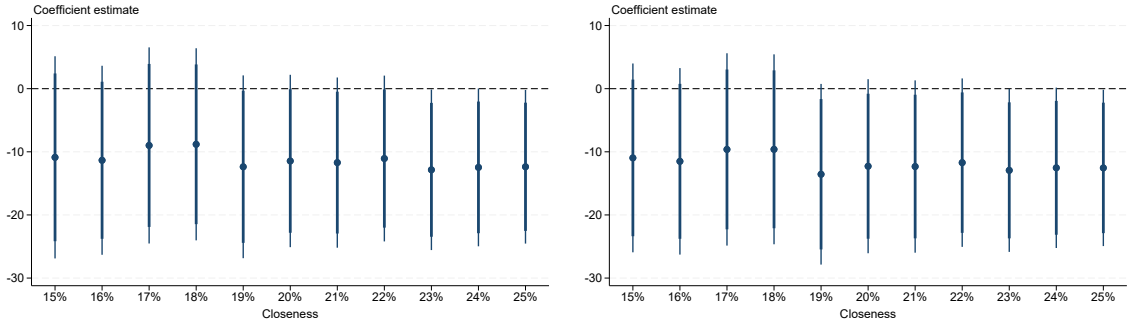


Figure A.5: (Local-) difference-in-differences

Notes: This graph shows coefficient estimates of interactions between month indicators and an indicator for female mayors. The sample is limited to close mixed-gender races using the 20% bandwidth (N: 1,440). The dependent variable is monthly aggregates of refugee quota fulfillment in percent. The regression includes municipality and month fixed effects. The reference period is February 2022. Standard errors are clustered at the municipality. 95% confidence intervals are indicated in the graph.

races.[37] As a consequence, I am not able to use arbitrarily narrow bandwidths. This limits local randomization. Still, point estimates remain relatively stable across bandwidths and qualitatively align with the TWFE results. Thus, local difference-in-differences results provide additional support for the lower fulfillment of female mayors.

[37] In the median mixed-gender race, there are a total of 10,704 votes. Thus, a bandwidth of 20% corresponds to 2,141 votes. While the baseline bandwidth is relatively large, it is similar to other settings using data at the municipality level in Germany (Freier and Thomasius, 2015; Baskaran and Hessami, 2018) and Italy (Bracco *et al.*, 2018). The optimal bandwidth algorithm of Calonico *et al.* (2014) results in a bandwidth of 10% with only 17 municipalities in the sample. Thus, I refrain from using the optimal bandwidth.



(a) Linear **(b)** Quadratic

Figure A.6: Robustness to different bandwidths

Notes: This graph shows coefficient estimates of the linear (a) and quadratic (b) local difference-in-differences specification following Equation (3). The respective sample is limited to different bandwidths of the margin of victory of female candidates around the threshold. 90 and 95% confidence intervals are indicated in the graph.

While the local difference-in-difference approach is internally valid, its reliance on close mixed-gender races – and thus on a small subset of municipalities – limits its external validity. By construction, the local difference-in-differences approach estimates the local treatment effect near the threshold. Municipalities with close mixed-gender races may be systematically different to other municipalities, however. In Table A.4, I compare pre-treatment municipality characteristics of the 40 municipalities with close mixed-gender races (20% bandwidth) to the remaining municipalities. Close mixed-gender race municipalities are larger and more densely populated. Also, some related characteristics are different, such as the CDU vote share in the 2017 federal election or more municipal employees per capita. Overall, this suggests that municipalities with close mixed-gender races are somewhat different to the average NRW municipality.

Table A.4: EXTERNAL VALIDITY – CLOSE MIXED-GENDER RACES

Variable	Mixed-gender races	Other races	Difference	Std. Errors	N
Municipal employees p.c.	0.02	0.01	-0.009***	0.002	396
Municipal revenue p.c.	3.05	2.89	-0.160	0.145	396
Municipal tax revenue p.c.	1.47	1.47	0.005	0.082	396
Municipal spending p.c.	3.02	2.84	-0.176	0.150	396
Number schools p.c.	0.00	0.00	-0.000	0.000	396
Construction housing p.c.	0.00	0.00	0.000	0.000	395
Share empty flats (2011)	0.04	0.04	-0.001	0.002	396
Share empty flats (2022)	0.03	0.03	0.001	0.002	396
Unemployment p.c.	0.03	0.02	-0.006***	0.002	396
Business registration p.c.	0.01	0.01	-0.000*	0.000	396
Population	83023.02	41044.19	-41978.840***	14970.682	396
Population density (per km^2)	808.49	475.53	-332.958***	88.790	396
Share population < 15	0.14	0.14	0.000	0.002	396
Share population 15 - 64	0.64	0.64	-0.000	0.003	396
Share population $\geq$ 65	0.22	0.22	-0.001	0.003	396
Share out-migration	0.06	0.06	-0.000	0.003	396
Share non-German population	0.11	0.10	-0.016**	0.007	396
Share refugees (2015)	0.02	0.02	0.002	0.002	292
Vote share CDU (2017)	0.33	0.38	0.042***	0.011	396
Vote share SPD (2017)	0.26	0.25	-0.012	0.008	396
Vote share AfD (2017)	0.10	0.09	-0.008**	0.004	396
Turnout (2017)	0.75	0.77	0.012*	0.006	396
Vote share NSDAP (1928)	0.01	0.01	0.003	0.004	187

Notes: This table shows t-tests for differences in municipality characteristics. The 40 municipalities with a close mixed-gender race (20% bandwidth) are compared to the remaining municipalities. If not indicated differently, municipal characteristics are as of 2019. See Figure A.12 for additional explanation on the variables. Stars indicate significance levels at 10%(*), 5%(**), and 1%(***).

A.4 Additional robustness checks

To substantiate the baseline finding, I conduct additional robustness checks. Results are presented in Table A.5. First, I show robustness to different specifications of the dependent variable. I transform quota fulfillment using the inverse hyperbolic sine. In addition, I rerun the model excluding outliers. That is, I exclude municipality-month combinations with above or equal 200% fulfillment. There are 20 such combinations. Both specifications result in very similar coefficients relative to the baseline.

Second, I show robustness to changes in the sample composition. Traditionally, the industrialized and urbanized Ruhr Area was one of the main destinations of migration flows in Germany. Also, this region underwent extensive structural change after the end of coal mining (starting in the 1960s). This may affect the willingness of these municipalities to host refugees. I proxy the Ruhr Area using members of the municipal initiative *Metropole Ruhr* and exclude these 53 municipalities from the sample. I also exclude 49 cities that are part of a German pro-refugee initiative on municipality level (*Städte Sicherer Häfen*). Results remain unchanged for both alternative samples.[38]

Table A.5: ROBUSTNESS – FEMALE MAYORS AND FULFILLMENT

	Dep. var.: Quota fulfillment (in %)					
	(1) IHS	(2) Excl. outliers	(3) Excl. *Ruhr*	(4) Excl. *Häfen*	(5) Excl. runoff	(6) New elec.
Female mayor × Post	-0.049**	-4.563**	-5.251*	-5.307**	-7.279***	-5.019**
	(0.024)	(2.274)	(2.705)	(2.296)	(2.596)	(2.295)
Mean (SD)	4.55 (0.26)	95.46 (18.58)	95.47 (19.29)	94.99 (19.16)	94.79 (18.59)	95.68 (19.22)
Time FE	✓	✓	✓	✓	✓	✓
Municipality FE	✓	✓	✓	✓	✓	✓
Municipalities	396	396	343	347	279	390
N	14,256	14,236	12,348	12,492	10,044	14,040

Notes: This table reports results from regressions that relate an indicator for female mayor (*Female mayor*) interacted with an indicator for the time after the intake of refugees from Ukraine (*Post*) to fulfillment of the refugee allocation quota (in %). Model (1) uses the inverse hyperbolic sine of the dependent variable. Model (2) excludes outliers (≥ 200% fulfillment). Model (3) excludes the 53 municipalities of the Ruhr Area. Model (4) excludes municipalities that are a safe haven for refugees. Model (5) excludes municipalities that had a runoff election in 2020. Model (6) excludes municipalities that had an election after 2020. The row entitled Mean (SD) reports the mean and standard deviation of the dependent variable. Stars indicate significance levels at 10%(*), 5%(**), and 1%(***). Standard errors in parentheses are clustered at the municipality level.

[38]The list of municipalities that are a member of *Metropole Ruhr* can be obtained from https://metropole.ruhr/metropole. The list of municipalities that are a member of *Städte Sicherer Häfen* can be obtained from https://staedte-sicherer-haefen.de/.

Third, I examine electoral peculiarities. I exclude all municipalities that had a runoff election in the 2020 mayoral election. In addition, I exclude all municipalities that had an extraordinary election after the regular election in 2020. This was typically due to the resignation of a mayor. Again, results remain unchanged for both specifications.

Fourth, I examine the robustness of the TWFE approach to alternative empirical strategies. To ensure that treatment and control municipalities are similar in observable characteristics I use a matched TWFE model as alternative specification. I calculate propensity scores using a logistic regression predicting the probability of a female mayor from various pre-treatment municipality trends (for the variables see Table A.6). I rerun the TWFE model on a sample of matched municipalities (with replacement). Results are presented in Table A.6. The coefficient is somewhat larger than the baseline and significant at the 5%-level. There is no variation in treatment timing, i.e., there is no staggered treatment. Still, heterogeneous or dynamic treatment effects could affect the main finding. Thus, I use an alternative estimator that is robust to heterogeneous treatment effects in non-staggered settings with a binary treatment (de Chaisemartin and D'Haultfœuille, 2024). The implementation follows de Chaisemartin and D'Haultfœuille (2022). I calculate standard errors using 1,000 bootstrap repetitions. The average treatment effect is shown in Model (3) of Table A.6. It is slightly larger than the baseline and statistically significant at the 5%-level. Thus, heterogeneous treatment effects do not affect the main finding much.

Fifth, I exclude that the difference in quota fulfillment is driven by particular municipalities. This is important as only few municipalities have a female mayor. Figure A.7 shows estimations akin to the baseline excluding one of the 57 female mayor municipalities at a time. The main finding is not driven by influential municipalities. Across all iterations the coefficient remains stable and similar to the baseline. In few instances, coefficients are insignificant at the 5%-level.

Table A.6: ALTERNATIVE ESTIMATORS – FEMALE MAYORS AND FULFILLMENT

	(1) Logit	(2) TWFE	(3) C&H 2024
Female mayor × Post		-6.960**	-5.124**
		(3.200)	(2.104)
Δ Construction p.c.	0.872		
	(0.795)		
Δ Unemployment p.c.	2.181		
	(2.292)		
Δ Municipal employees p.c.	22.341***		
	(8.168)		
Δ Business registration p.c.	-1.624*		
	(0.877)		
Δ Business deregistration p.c.	1.682**		
	(0.659)		
Δ Municipal revenue p.c.	-0.146		
	(2.779)		
Δ Municipal spending p.c.	-1.943		
	(2.607)		
Δ Municipal tax revenue p.c.	0.739		
	(1.911)		
Δ Population	0.001		
	(0.000)		
Δ Foreign population	-0.000		
	(0.000)		
Mean (SD)	0.14 (0.35)	96.52 (20.74)	95.63 (19.14)
Time FE		✓	
Municipality FE		✓	
Municipalities	396	103	396
N	396	3,708	6,732

Notes: This table reports results from a matched difference-in-differences approach and an alternative estimator. First, a propensity score is estimated. Model (1) reports results from a logistic regression that uses pre-treatment municipality trends of several variables to predict the victory of a female mayor. The propensity scores are then used to identify control municipalities that are similar to treatment municipalities. Control municipalities are drawn with replacement. Model (2) reports results from regressions that relate an indicator for female mayor (*Female mayor*) interacted with an indicator for the time after the intake of refugees from Ukraine (*Post*) to fulfillment of the refugee allocation quota (in %). The regression is run on the matched sample. Model (3) uses an estimator that is robust to heterogenous treatment effects in non-staggered and binary designs proposed by de Chaisemartin and D'Haultfœuille (2024). The row entitled Mean (SD) reports the mean and standard deviation of the dependent variable. Stars indicate significance levels at 10%(*), 5%(**), and 1%(***). Models (2) and (3) include standard errors that are clustered at the municipality level.

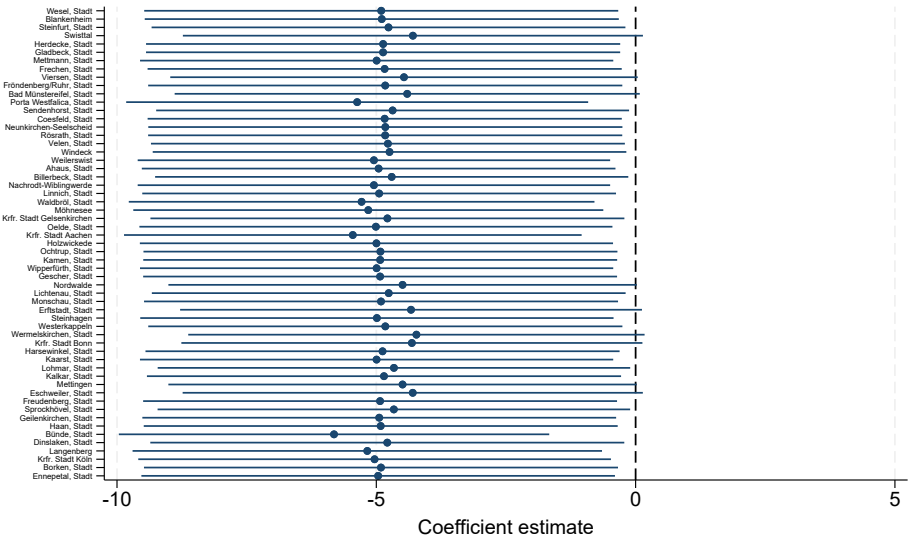


Figure A.7: Female mayors and refugee quota fulfillment – Jackknife resampling

Notes: This graph shows coefficient estimates from regressions that relate an indicator for female mayor (*Female mayor*) interacted with an indicator for the time after the intake of refugees from Ukraine (*Post*) to fulfillment of the refugee allocation quota (in %). In each iteration, one municipality with a female mayor is excluded from the estimation sample. The respective municipality is shown on the y-axis. Regressions include municipality and month fixed effects. Standard errors are clustered at the municipality level. 95% confidence intervals are indicated in the graph.

A.5 Ukrainians in NRW

In this section, I briefly describe the number of Ukrainian citizens and Germans with an Ukrainian immigrant origin in NRW using data from the Statistical Office. There is no data on population by nationality or immigrant origin at the municipality level. Thus, I present data at the state level. First, Figure A.19 shows the number of Ukrainian citizens in NRW over time. Before the Russian invasion, there were few Ukrainian citizens in NRW. On average, 28,729 Ukrainian citizens lived in NRW per year. 1.4% of non-Germans in NRW were Ukrainian citizens. In 2022, the number and share of Ukrainian citizens in NRW increased considerably due to the intake of refugees from Ukraine. Also, there are relatively few Germans with an Ukrainian immigrant origin in NRW. They themselves or at least one of their parents (also) have Ukrainian citizenship. About 1.1% of Germans with an immigrant origin in NRW have a Ukrainian background.

A.6 Collection of council minutes

To examine refugee topics and agenda setting in more detail, I hand-collect official minutes of council meetings. Data collection took place in December 2023. I examine the election period of 2020, i.e., council meetings between November 2020 and December 2023. There is no mandatory frequency of council meetings. The municipal code of NRW recommends at least one meeting every two months. There is considerable variation in meeting frequency. There are municipalities that had more than 30 meetings during this period. On average, municipalities had 21.7 meetings between October 2020 and December 2023.

Typically, municipalities have an online information platform for citizens and council members (*Ratsinformationssystem*) from which I manually downloaded council minutes. In some cases, no such platform was available. In these cases, I contacted municipalities via e-mail and phone. In total, I obtain minutes for 103 of the 110 municipalities that had a mixed-gender race (93.6%). This corresponds to 2,025 unique council minutes. Note that for some municipalities not all minutes are publicly available. Minutes are in PDF format. They include

information on participation, duration, the agenda of the meeting, key lines of discussion, and decisions of the council. Minutes are not recorded ad verbatim. Instead, they provide an abbreviated version of council discussions. Council meetings can include parts that are not public. This is possible for topics that concern individuals' right for privacy, such as property purchases. The discussions and decisions in these parts of the council meetings are not recorded for public use. Minutes are written by a municipal clerk who is selected at the beginning of the council period.

To improve comparability across municipalities, I limit minutes to their core part. In particular, I remove all appendices. Appendices can be large and include personal statements, sketches of infrastructure projects, proposals for public contracts, or tables for budgetary planning. They typically do not include additional information on the discussion in council. I inspect minutes visually to determine the extent of the core part and cut them accordingly. After removing the appendix, the average minutes have 19 pages.

The corpus of minutes is pre-processed for automated analysis following common practice (Gentzkow *et al.*, 2019). I remove numbers, special characters, and lines. All words are transformed to lowercase. Typical German stop-words are removed and words are reduced to their stem using the Porter stemmer.

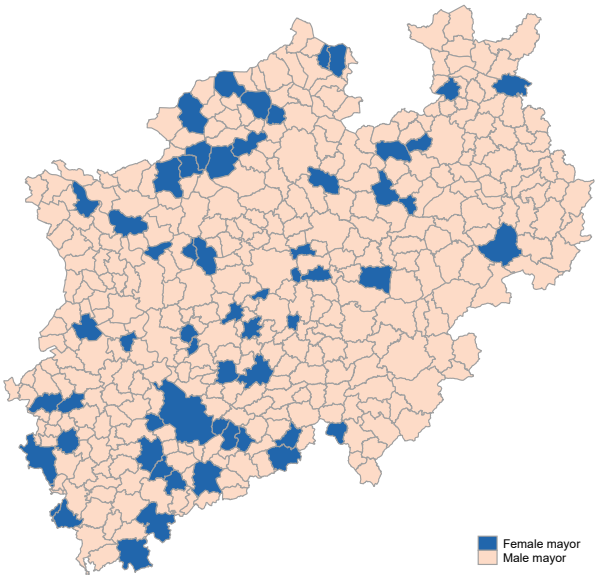
A.7 Topic analysis using LLM

I analyze topics discussed in council minutes using a LLM. The approach is similar to Kreitmeir (2023). First, I limit the corpus to paragraphs that contain the refugee-terms described in Section 6.3. I combine all relevant paragraphs in minutes. Focusing on relevant paragraphs limits the number of tokens, i.e., the length of text to be analyzed by the LLM. Using the LLM is subject to a fee which is a function of the number of tokens. Also, the length of the text must be below the limit of the model's context window. The combined paragraphs are provided to the LLM. I access the LLM using OpenAI's API. I use *GPT-3.5 turbo* and *GPT-4*. I set the temperature parameter to zero to obtain close to deterministic answers. Answers are reported in JSON format. I use the following prompt:

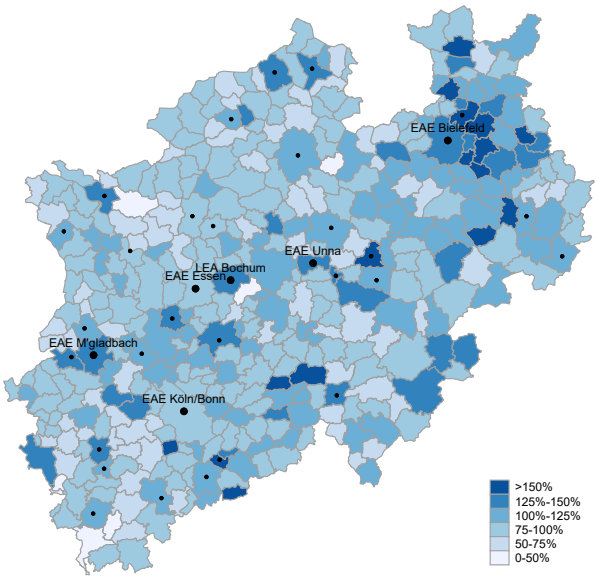
[paragraphs]

Based on the above text please complete the following three tasks. (1) Rate the mayor's attitude towards refugees based on the text on a scale from 1 (negative) to 5 (positive). Please also state the mayor's name. (2) Indicate Yes/No if the following topics are discussed in the text in relation to refugees: Municipal finance, housing, childcare, education, integration, crime. (3) Identify one additional relevant refugee topic in the text. Answer in English. Report results in JSON format with the following columns: attitude; mayor's name; municipal finance; housing; child care; education; integration; crime; open topic.

A.8 Additional figures



(a) Female mayors in NRW



(b) Refugee quota fulfillment

Figure A.8: Female mayors and refugee quota fulfillment across municipalities

Notes: Subfigure (a) shows the municipalities with female and male mayors as of the September 2020 election. Subfigure (b) shows the quota fulfillment of the municipalities in NRW. The average weekly fulfillment in March 2022 is depicted. Different types of refugee facilities are indicated by the black dots. LEA is the state's initial reception facility. EAE are the five state reception facilities. The unlabeled dots are the 28 state accommodation facilities. The projection of the maps is UTM32N.

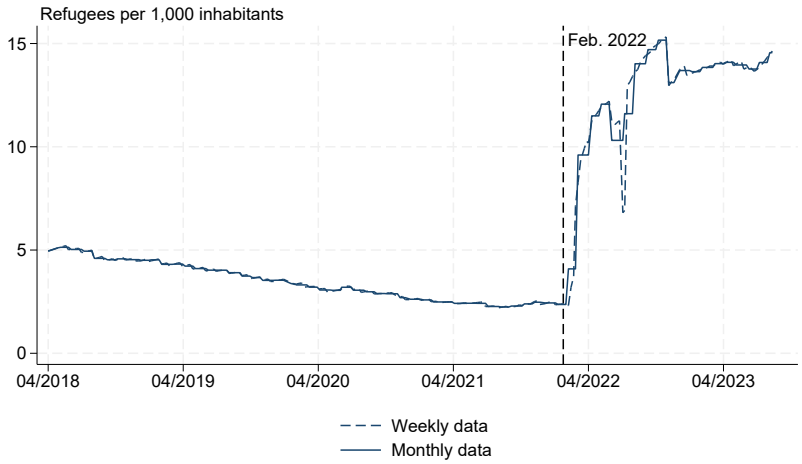
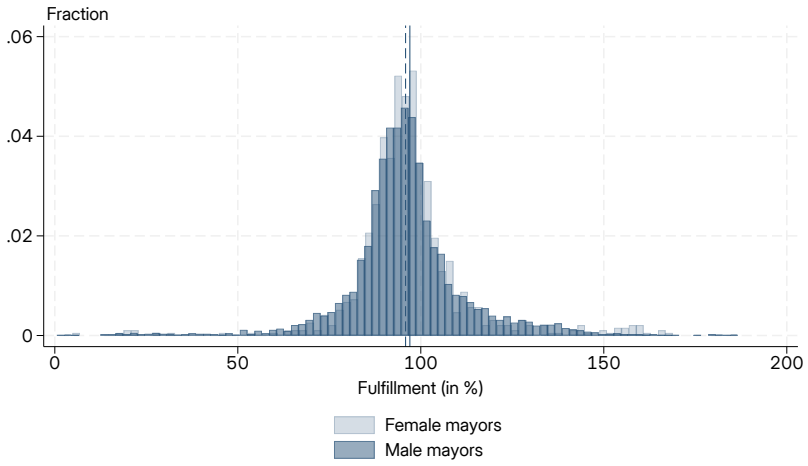
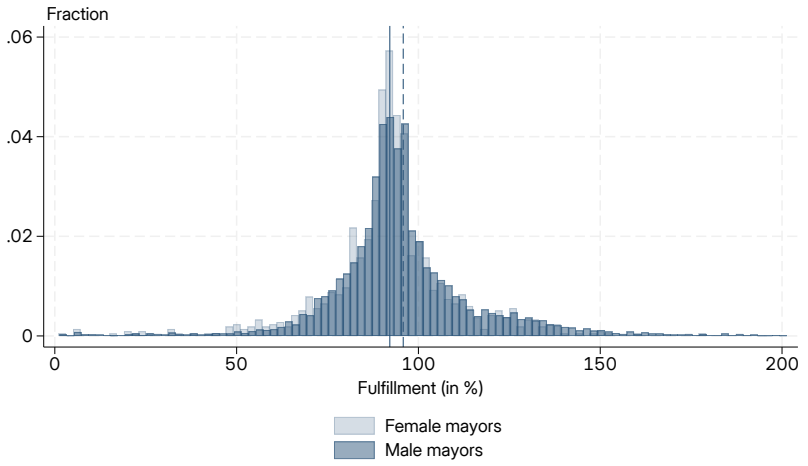


Figure A.9: Refugees over time (full time-series)

Notes: This graph shows the number of refugees per 1,000 inhabitants over time. The graph shows weekly and monthly data between March 2018 and September 2023. The vertical line indicates the beginning of the Russian invasion of Ukraine.



(a) Before February 2022



(b) After February 2022

Figure A.10: Distribution of quota fulfillment

Notes: Subfigure (a) shows the distribution of quota fulfillment in percent for male and female mayors between October 2020 and February 2022. Subfigure (b) shows the distribution of quota fulfillment in percent for male and female mayors between March 2022 and September 2023 (i.e., after the Russian invasion of Ukraine). The solid vertical lines are the averages for female mayors. The dashed vertical lines are the averages for male mayors. Values above 200% are excluded for readability.

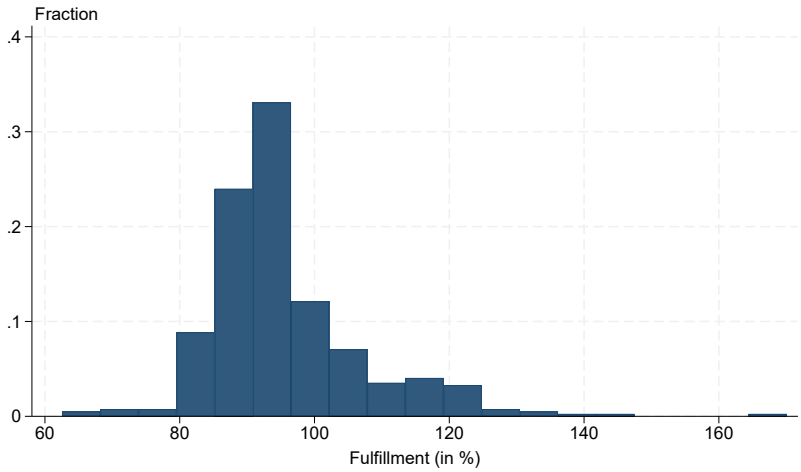
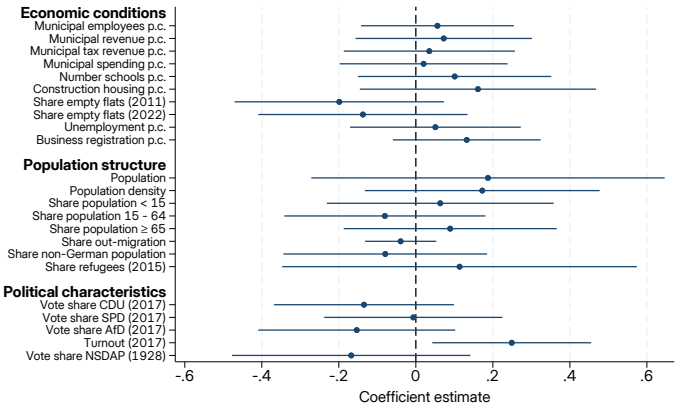
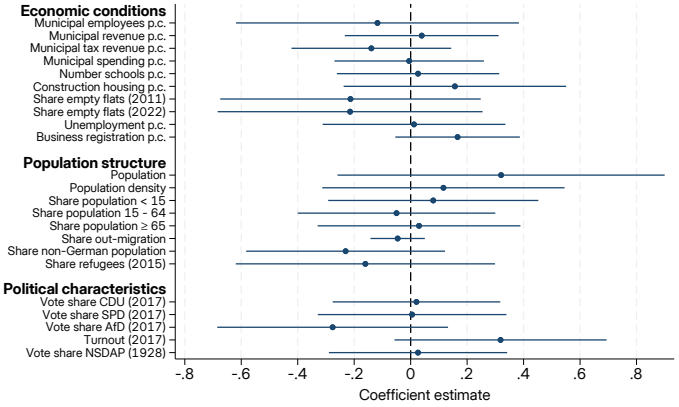


Figure A.11: Distribution of municipality average refugee quota fulfillment

Notes: This graph shows the distribution of the average refugee quota fulfillment for the 396 municipalities. The average is calculated across all weeks between October 2020 and September 2023.



(a) Full sample



(b) Mixed-gender races

Figure A.12: Balance of municipality characteristics (I)

Notes: This graph shows coefficient estimates of regressions that relate an indicator for female mayors to the respective municipality-level dependent variable. If not indicated differently, data is from 2019. That is, data is from before the regular municipal election of 2020. All dependent variables are standardized. The share of empty flats is from the German censuses of 2011 and 2022. Data on refugees in 2015 is from Westdeutscher Rundfunk Köln (2015). Election data refers to the federal election of 2017. Data on the NSDAP vote share is from Voigtländer and Voth (2012). It is available for a subset of municipalities only. Other than that, data is from the Statistical Office of NRW. Subfigure (a) shows estimates for all 396 municipalities. Subfigure (b) shows estimates for the 110 municipalities with a mixed-gender race. 95% confidence intervals are indicated in the graph.

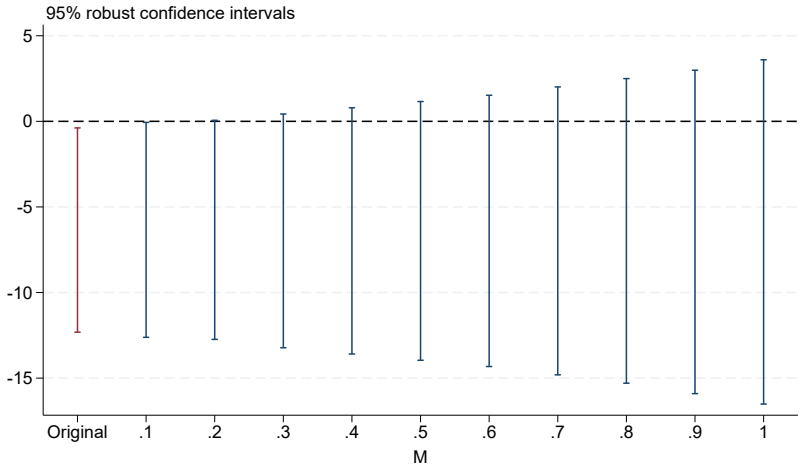


Figure A.13: Honest difference-in-differences

Notes: This graph shows the result of the approach by Rambachan and Roth (2023) to test the plausibility of the parallel trends assumption. I use the relative magnitudes restriction. The result labeled "Original" corresponds to the results of the baseline estimation. M refers to the extent of 95% robust confidence intervals are indicated in the graph.

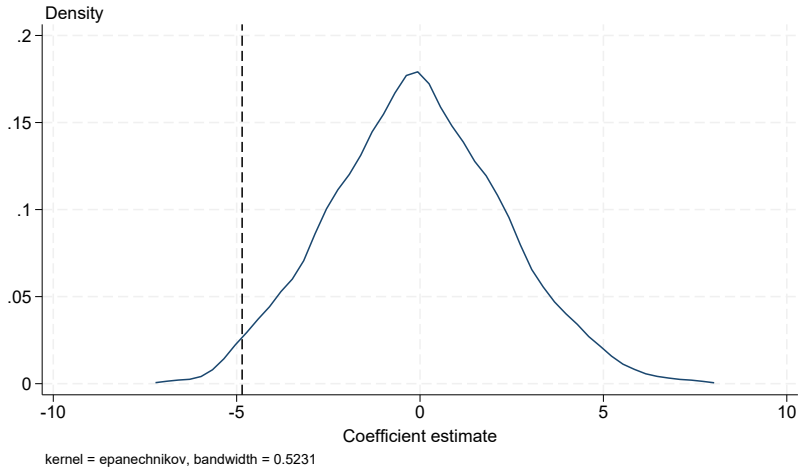


Figure A.14: Randomized inference

Notes: This graph shows the results of a randomized inference procedure. The female mayor treatment is randomly allocated to municipalities (999 repetitions). The graph shows a kernel density plot of the estimated coefficients. The dashed line indicates the actual coefficient from model (1) of Table 1. The estimated p-value is 0.030 with a standard error of 0.0054. Its 95% confidence interval excludes the 0.05 and 0.10 thresholds.

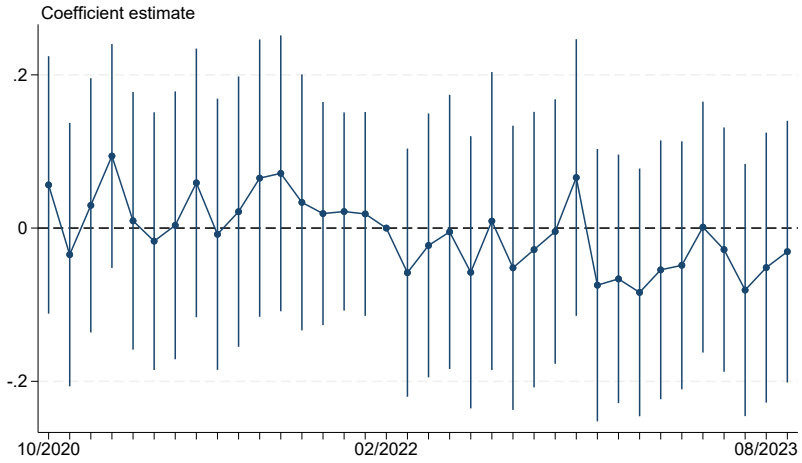


Figure A.15: Female mayors and refugee quota fulfillment – Extensive margin

Notes: This graph shows coefficient estimates of interactions between month indicators and an indicator for female mayors (N: 14,256). The dependent variable is an indicator that is one if quota fulfillment is higher or equal to 100%. The regression includes municipality and month fixed effects. The reference period is February 2022. Standard errors are clustered at the municipality. The full sample of municipalities is used. 95% confidence intervals are indicated in the graph.

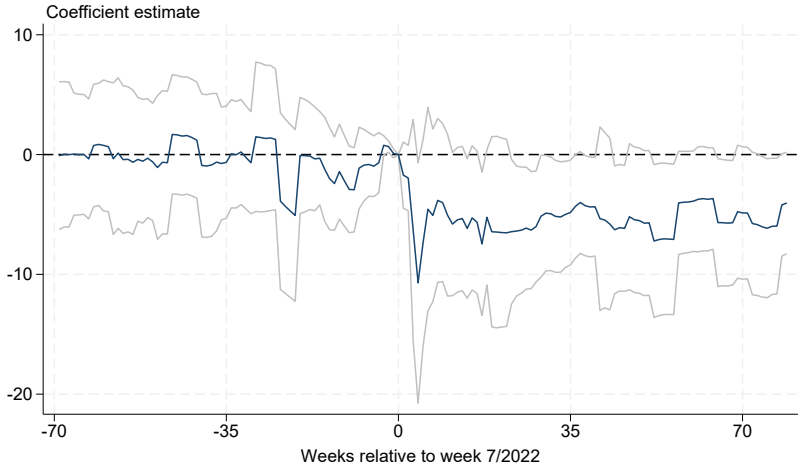
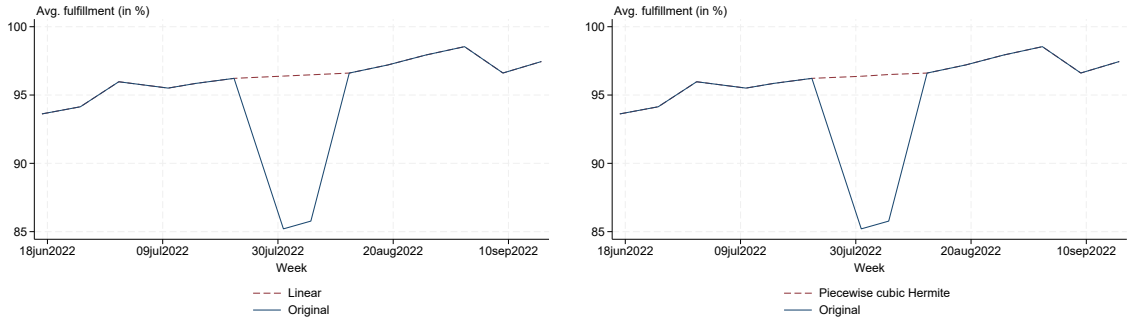


Figure A.16: Alternative difference-in-differences plot

Notes: This graph shows coefficient estimates of interactions between week indicators and an indicator for female mayors (N: 59,004). The dependent variable is weekly fulfillment of the refugee quota. Values are imputed during two month in summer 2022 (see Section 5.3). Regressions include municipality and week fixed effects. The reference period is week seven in 2022. Standard errors are clustered at the municipality. The full sample of municipalities is used. 95% confidence intervals are indicated in the graph.



(a) **(b)**

Figure A.17: Interpolation of refugee quota fulfillment

Notes: This graph shows the interpolated values during the drop in fulfillment in summer 2022 for administrative reasons. The time series is limited to a window around the interpolation of data points. Subfigure (a) shows a linear interpolation. Subfigure (b) shows a piecewise cubic Hermite interpolation.

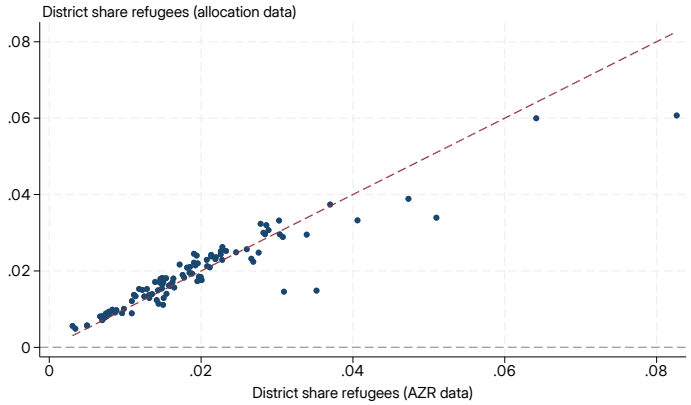


Figure A.18: Correlation between official refugee data and the allocation data

Notes: This graph shows the correlation between official refugee data from the central register of foreigners (*Ausländerzentralregister – AZR*) and the allocation data. Both variables are the share of refugees in the district relative to the total number of refugees in NRW in 2022 and 2023. Allocation data is aggregated at the district and year level. The dashed line is the 45°- line.

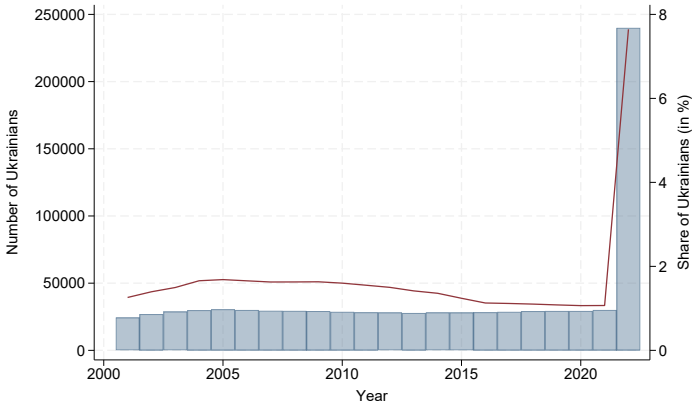


Figure A.19: Number of Ukrainian citizens over time

Notes: This graph shows the number of Ukrainian citizens (left axis) and their share relative to all non-Germans (right axis) over time. Data is for NRW in total.

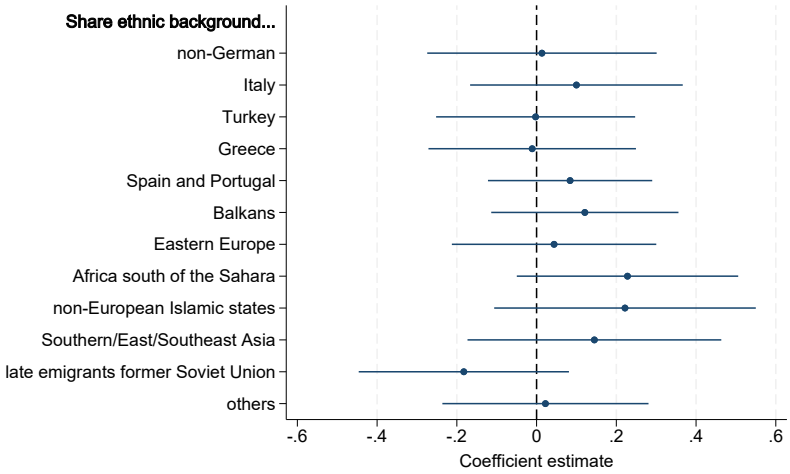


Figure A.20: Balance test of ethnic backgrounds

Notes: This graph shows coefficient estimates of regressions that relate an indicator for female mayors to the respective municipality-level dependent variable. Dependent variables are the share of residents with the specified ethnic background in 2019. Ethnic background is derived from the first and second name of household heads. Data is provided by RWI and Microm (2023). All dependent variables are standardized. The full sample of municipalities is used. 95% confidence intervals are indicated in the graph.

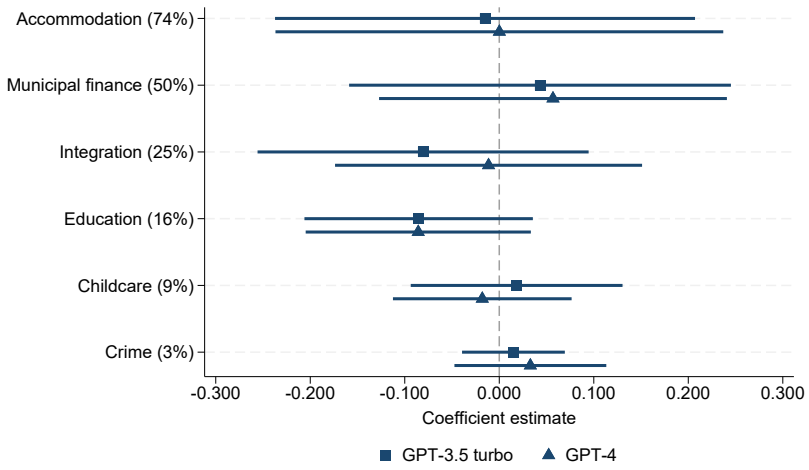


Figure A.21: Changes in topics discussed in council meetings

Notes: This graph shows coefficient estimates of interactions between an indicator for observations after February 2022 and an indicator for female mayors. The dependent variables are indicators that are one if the topic is discussed in the council meeting. The share of council meetings discussing the topic relative to all meetings about refugee topics is indicated in brackets. Results are shown for *GPT-3.5 turbo* and *GPT-4*. 95% confidence intervals are indicated in the graph.

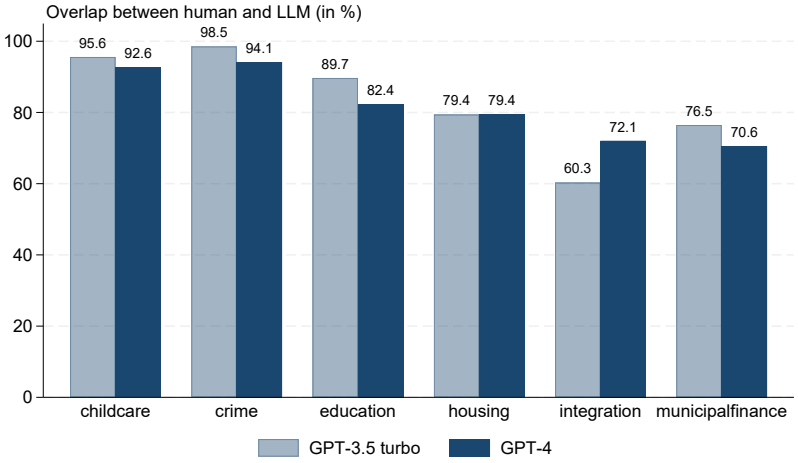


Figure A.22: Overlap between human and LLM classification

Notes: This graph shows the share of overlap in topic classification between a student assistant and the two LLM. The overlap is shown for all refugee topics. The test was conducted on a 10% random sample of relevant council minute paragraphs.

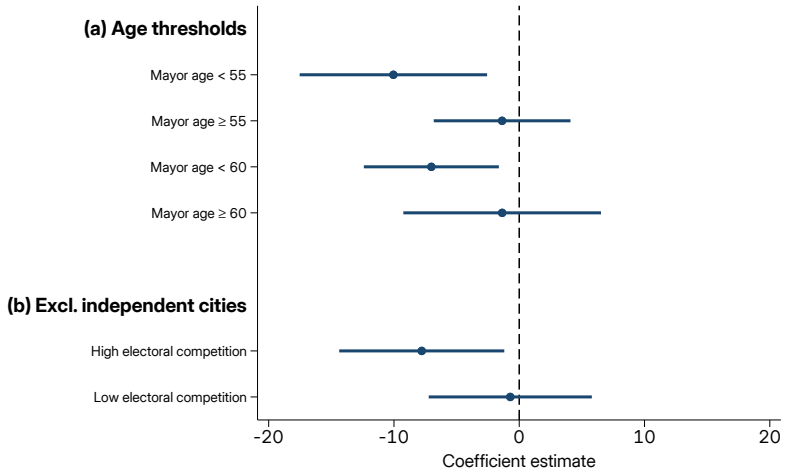


Figure A.23: Mechanism – Electoral competition (robustness)

Notes: This figure shows coefficient estimates of a version of the model described in Equation (2). Subfigure (a) shows coefficient estimates limiting the sample to younger and older mayors using alternative age thresholds (below and above or equal 55/60 years as of 2025). Subfigure (b) shows coefficient estimates limiting the sample to races for the mayorship with below and above or equal median closeness of the election in 2020. The 22 independent cities are excluded from the estimation sample. This ensures that the pattern is not driven by large cities. 95% confidence intervals are indicated in the graph.

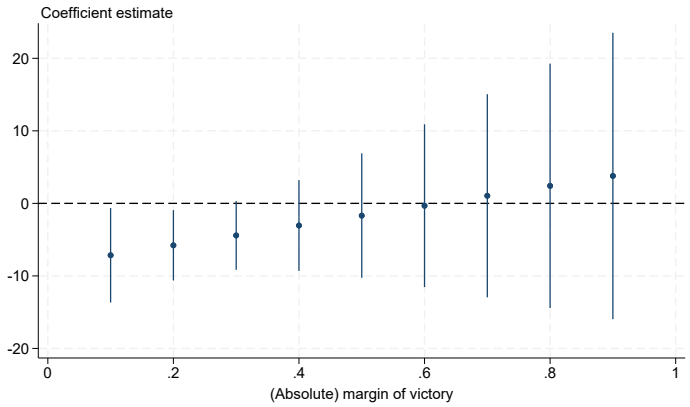


Figure A.24: Mechanism – Electoral competition (marginsplot)

Notes: This graph shows the effect of female mayors on fulfillment in response to the intake of Ukrainian refugees. The effect is shown for different levels of the absolute margin of victory, i.e., a measure for electoral competition in the local election of 2020. 95% confidence intervals are indicated in the graph.

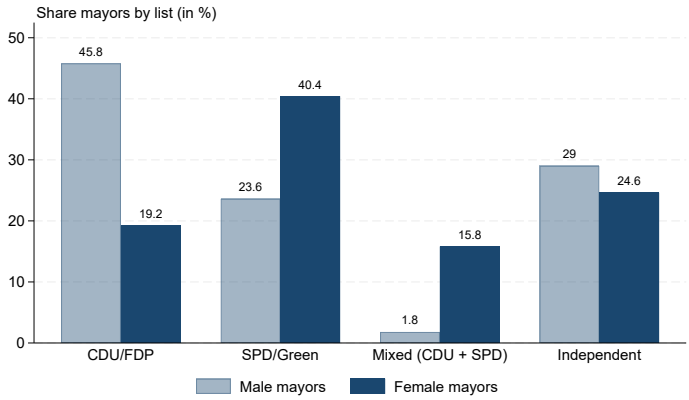
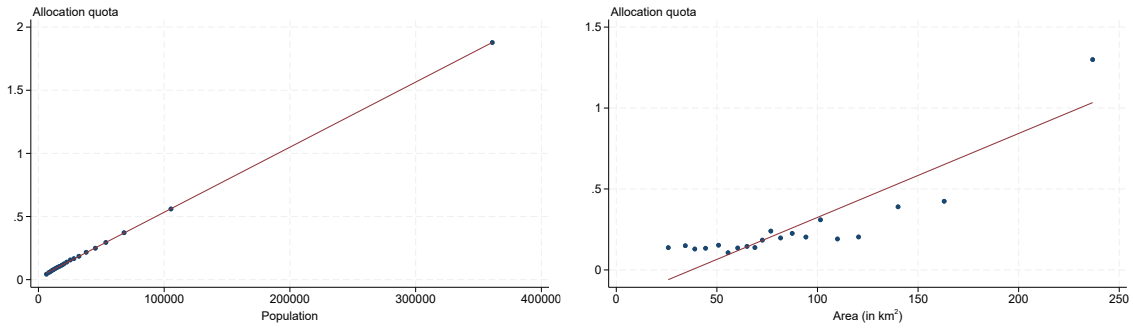


Figure A.25: Mechanisms – Mayors' party affiliation

Notes: This graph shows the share of mayors that run for the different parties (in %). Shares are depicted separately for female and male mayors. Categories group together different combinations of parties. For example, CDU/FDP contains mayors for CDU, FDP, CDU and FDP combined, and CDU and an independent voter list combined. Mixed mayors are mayors supported by large parties from both sides of the political spectrum (e.g., CDU/SPD). Independent candidates run without the support of a large party.



(a) **(b)**

Figure A.26: Components of the allocation quota

Notes: This graph shows binned scatterplots of the allocation quota and population (a) and area in km^2 (b). The straight line shows the linear fit.

A.9 Additional tables

Table A.7: SUMMARY STATISTICS: MUNICIPALITY CHARACTERISTICS

Variable	Mean	SD	Min.	Max.	N
Panel A: Full sample					
Quota fulfillment (in %)	95.63	19.14	0	275	14256
Allocation quota (in %)	0.25	0.47	0	5.62	14256
Refugees p.c.	7.90	5.87	0	35	14255
Female mayor	0.14	0.35	0	1	14256
Post Feb. 2022	0.53	0.50	0	1	14256
Margin of victory (in %)	-2.13	33.04	-81	88	12708
Panel B: Mixed-gender races					
Quota fulfillment (in %)	96.62	21.11	0	229	3960
Allocation quota (in %)	0.28	0.57	0	5.62	3960
Refugees p.c.	7.76	5.70	0	27	3959
Female mayor	0.42	0.49	0	1	3960
Post Feb. 2022	0.53	0.50	0	1	3960
Margin of victory (in %)	-6.69	30.70	-69	60	3960
Share keywords	5.37	14.43	0	207	2025

Notes: This table reports summary statistics on key variables for the full sample (Panel A) and the mixed-gender races sample (Panel B). All variables are at the municipality level and represent monthly aggregates. Quota fulfillment (in %) is the number of refugees relative to the allocated number of refugees following Equation (1). Allocation quota (in %) is the administrative quota based on municipal population and area. Refugees p.c. is the number of refugees per 1,000 inhabitants. Female mayor is one in municipalities governed by a woman. Post Feb. 2022 is one for months after February 2022. Margin of victory (in %) is the difference in votes between the two candidates in the mixed-gender race relative to all valid votes (in %). Share keywords is the number of refugee terms relative to 10,000 words in council minutes.

Table A.8: SUTVA – FEMALE MAYORS AND FULFILLMENT

	Dep. var.: Quota fulfillment (in %)			
	(1) Non compl. weekly	(2) Non compl. monthly	(3) Drop neighbors	(4) Distance control
Female mayor × Post	-2.392*	-2.122*	-6.744**	-4.288*
	(1.272)	(1.278)	(2.635)	(2.270)
Mean (SD)	86.82 (14.71)	87.50 (12.54)	95.82 (19.53)	95.63 (19.14)
Time FE	✓	✓	✓	✓
Municipality FE	✓	✓	✓	✓
Municipalities	387	379	205	396
N	41,558	10,157	7,380	14,256

Notes: This table reports results from regressions that relate an indicator for female mayor (*Female mayor*) interacted with an indicator for the time after the intake of refugees from Ukraine (*Post*) to fulfillment of the refugee allocation quota (in %). Model (1) reports results keeping only municipality-week pairs with lower than 100% fulfillment. Model (2) reports results keeping only municipality-month pairs with lower than 100% fulfillment. Model (3) reports results dropping all contiguous neighbors of municipalities governed by female mayors. Model (4) reports results controlling for average distance to municipalities governed by female mayors interacted with a linear time trend. The row entitled Mean (SD) reports the mean and standard deviation of the dependent variable. Stars indicate significance levels at 10%(*), 5%(**), and 1%(***). Standard errors in parentheses are clustered at the municipality level.

Table A.9: ALTERNATIVE OUTCOMES – FEMALE MAYORS AND MUNICIPAL SPENDING

	Dep. var.: Spending in IHS and per capita			
	(1) Total spending	(2) Social transfers	(3) Road construction	(4) District transfers
Female mayor × Post	-0.009	-0.089*	-0.072	-0.007
	(0.013)	(0.048)	(0.136)	(0.008)
Mean (SD)	8.73 (0.25)	5.16 (1.24)	4.01 (1.60)	6.57 (1.61)
Time FE	✓	✓	✓	✓
Municipality FE	✓	✓	✓	✓
Municipalities	396	396	396	396
N	1,584	1,584	1,584	1,584

Notes: This table reports results from regressions that relate an indicator for female mayor (*Female mayor*) interacted with an indicator for the time after the intake of refugees from Ukraine (*Post*) to municipal spending outcomes between 2020 and 2023. All outcomes are per capita and transformed using the inverse hyperbolic sine (IHS). Model (1) reports results for total spending. Model (2) reports results for social tranfers (including e.g., unemployment benefits). Model (3) reports results for spending on road construction. Model (4) reports results for municipal transfers to districts. The row entitled Mean (SD) reports the mean and standard deviation of the dependent variable. Stars indicate significance levels at 10%(*), 5%(**), and 1%(***). Standard errors in parentheses are clustered at the municipality level.

Table A.10: CORRELATION MARGIN OF VICTORY OVER TIME

	Dep. var.: Margin of victory in 2020	
	(1)	(2)
Margin of victory 2014/15	0.205***	0.151***
	(0.067)	(0.051)
Mean (SD)	0.28 (0.18)	0.28 (0.18)
Controls		✓
N	329	329

Notes: This table reports results from regressions that relate the absolute margin of victory in the municipal election of 2014/15 to the absolute margin of victory in the municipal election of 2020. Model (1) reports results without controls. Model (2) reports results when controlling for population, turnout, female mayor, runoff elections, and incumbent mayor. The row entitled Mean (SD) reports the mean and standard deviation of the dependent variable. Stars indicate significance levels at 10%(*), 5%(**), and 1%(***). Standard errors in parentheses are clustered at the district level.

Table A.11: MECHANISM – FEMALE MAYORS AND FULFILLMENT (I)

	Dep. var.: Quota fulfillment (in %)			
	Partisanship		Public administration background	
	(1) Left parties	(2) Other parties	(3) Yes	(4) No
Female mayor × Post	-6.242	-4.169*	-7.000	-3.992
	(4.749)	(2.502)	(5.540)	(2.604)
p-value (odd) - (even)		0.70		0.62
Mean (SD)	98.91 (19.79)	94.48 (18.78)	94.98 (19.53)	96.17 (18.85)
Time FE	✓	✓	✓	✓
Municipality FE	✓	✓	✓	✓
Municipalities	103	293	139	249
N	3,708	10,548	5,004	8,964

Notes: This table reports results from regressions that relate an indicator for female mayor (*Female mayor*) interacted with an indicator for the time after the intake of refugees from Ukraine (*Post*) to fulfillment of the refugee allocation quota (in %). Model (1) shows results for mayors from the SPD, the Green Party, or both. Model (2) shows results for mayors from right parties (CDU, FDP) and independent mayors. Models (3) and (4) show results for mayors with and without an occupational background in public administration. The row entitled *p*-value (odd) - (even) shows the *p*-value for the difference between models with odd numbers and models with even numbers. The row entitled Mean (SD) reports the mean and standard deviation of the dependent variable. Stars indicate significance levels at 10%(*), 5%(**), and 1%(***). Standard errors in parentheses are clustered at the municipality level.

Table A.12: MECHANISM – FEMALE MAYORS AND FULFILLMENT (II)

	Dep. var.: Quota fulfillment (in %)					
	Female unemployment		Fiscal capacity		Female councilor share	
	(1) Below median	(2) Above median	(3) Below median	(4) Above median	(5) Below median	(6) Above median
Female mayor × Post	-2.663	-6.514*	-5.467	-4.232	-4.612	-5.048
	(2.778)	(3.549)	(3.662)	(2.734)	(3.177)	(3.250)
p-value (odd) - (even)		0.39		0.79		0.92
Mean (SD)	93.51 (17.82)	97.75 (20.16)	97.26 (20.54)	94.00 (17.48)	94.02 (18.41)	97.12 (19.68)
Time FE	✓	✓	✓	✓	✓	✓
Municipality FE	✓	✓	✓	✓	✓	✓
Municipalities	203	202	198	198	190	206
N	7,125	7,131	7,128	7,128	6,840	7,416

Notes: This table reports results from regressions that relate an indicator for female mayor (*Female mayor*) interacted with an indicator for the time after the intake of refugees from Ukraine (*Post*) to fulfillment of the refugee allocation quota (in %). Models (1) and (2) show results for below/above median female unemployment p.c. in 2019. Models (3) and (4) show results for below/above median share tax revenue relative to total municipal spending in 2019. Models (5) and (6) show results for below/above median share of female councilors in the 2020 election cycle. The row entitled *p*-value (odd) - (even) shows the *p*-value for the difference between models with odd numbers and models with even numbers. The row entitled Mean (SD) reports the mean and standard deviation of the dependent variable. Stars indicate significance levels at 10%(*), 5%(**), and 1%(***). Standard errors in parentheses are clustered at the municipality level.

Table A.13: SOFTWARE-PACKAGES

Package	Author(s)	Software
did_multiplegt	Clément de Chaisemartin and Xavier D'Haultfoeuille	Stata
reghdfe	Sergio Correia	Stata
coefplot	Ben Jann	Stata
spmap	Maurizio Pisati	Stata
shp2dta	Kevin Crow	Stata
estout	Ben Jann	Stata
rdrobust	Sebastian Calonico, Matias D. Cattaneo, Max H. Farrell, and Rocio Titiunik	Stata
gender	Lincoln Mullen	R
quanteda	Kenneth Benoit, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, Akitaka Matsuo, and William Lowe	R
camelot	Vinayak Mehta	Python
pandas	NumFOCUS	Python
numpy	Travis Oliphant	Python
ritest	Simon Heß	Stata

Notes: This table shows the most important software packages used in the analysis.