International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

Address: IIASA, Schlossplatz 1, A-2361 Laxenburg, Austria
Email: permissions@iiasa.ac.at

Working paper

The Short-Term Dynamics of Conflict-Driven Displacement: Bayesian Modeling of Disaggregated Data from Somalia

Gregor Zens[1,2] (zens@iiasa.ac.at)
Lisa Thalheimer[3] (thalheimer@ehs.unu.edu)

[1] Population and Just Societies (POPJUS) Program, International Institute for Applied Systems Analysis (IIASA)
[2] Wittgenstein Centre for Demography and Global Human Capital (VID/ÖAW, IIASA, University of Vienna)
[3] United Nations University (UNU)

WP-24-002

Approved by:

Name Anne Goujon
Program: Population and Just Societies (POPJUS) Program
Date: 17 January 2024

ZVR 524808900

Disclaimer, funding acknowledgment, and copyright information:
IIASA Working Papers report on research carried out at IIASA and have received only limited review. Views or opinions expressed herein do not necessarily represent those of the institute, its National Member Organizations, or other organizations supporting the work.

The authors gratefully acknowledge funding from IIASA and the National Member Organizations that support the institute (The Austrian Academy of Sciences; The Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES); The National Natural Science Foundation of China (NSFC); The Academy of Scientific Research and Technology (ASRT), Egypt; The Finnish Committee for IIASA; The Association for the Advancement of IIASA, Germany; The Technology Information, Forecasting and Assessment Council (TIFAC), India; The Indonesian National Committee for IIASA; The Iran National Science Foundation (INSF); The Israel Committee for IIASA; The Japan Committee for IIASA; The National Research Foundation of Korea (NRF); The Mexican National Committee for IIASA; The Research Council of Norway (RCN); The Russian Academy of Sciences (RAS); Ministry of Education, Science, Research and Sport, Slovakia; The National Research Foundation (NRF), South Africa; The Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS); The Ukrainian Academy of Sciences; The Research Councils of the UK; The National Academy of Sciences (NAS), USA; The Vietnam Academy of Science and Technology (VAST).

The Short-Term Dynamics of Conflict-Driven Displacement: Bayesian Modeling of Disaggregated Data from Somalia*

GREGOR ZENS[a,b] LISA THALHEIMER[c]

[a] *International Institute for Applied Systems Analysis (IIASA)*

[b] *Wittgenstein Centre for Demography and Global Human Capital (WIC)*

[c] *United Nations University (UNU)*

January 17, 2024

Abstract

Understanding the short-run dynamics of conflict and forced displacement is crucial for policy intervention, yet quantitative analyses in this realm are sparse. This is primarily due to the scarcity of high-frequency displacement data and methodological challenges arising when modeling imperfect data collected in conflict zones. Addressing both issues, we develop a Bayesian panel regression model to assess the short-term impact of conflict on displacement in Somalia, utilizing weekly panel data that encompasses 8 million displacements and 19,000 conflict events from 2017 to 2023. Results suggest a rapid and non-linear displacement response post-conflict, with significant heterogeneity in effects dependent on the nature of conflict events. In a displacement forecasting exercise, our model outperforms standard benchmarks, underscoring its relevance for informed decision-making in crisis scenarios.

Keywords: Internal Displacement, Distributed Lag Model, Factor Model, State Space Model.

*Correspondence to Gregor Zens, International Institute for Applied Systems Analysis. Schlossplatz 1, 2361 Laxenburg, Austria. E-mail: zens@iiasa.ac.at.

1 Introduction

In 2022, the United Nations High Commissioner for Refugees (UNHCR) reported 108 million forcibly displaced persons worldwide, surpassing 1% of the global population.[1] This displacement crisis is primarily attributed to conflict and generalized violence (Schmeidl, 1997, Schmeidl, 2001, Davenport et al., 2003, Hatton, 2009, Conte and Migali, 2019, Abel et al., 2019). A detailed understanding of the dynamics of conflict and displacement is crucial for policymakers, NGOs, and other stakeholders involved in humanitarian assistance. This paper contributes to this understanding by focusing on the immediate effects of conflict events on forced displacement, with a specific focus on Somalia. The situation in Somalia is especially severe, with an estimated 3.8 million people, or about 20% of the population, internally displaced in 2022 (DRC, 2022).

According to theoretical models of displacement and qualitative evidence, the impact of conflict on displacement materializes rapidly and in the immediate spatial proximity of the conflict event.[2] This suggests that empirical models examining conflict and displacement should ideally focus on small-scale geographical areas and short-term effects. However, this presents quantitative research in this field with two serious challenges. First, fine-grained data on displacement is only rarely available. This is largely due to inadequate data infrastructure in areas where this issue is most pertinent. As a result, most quantitative literature on the topic focuses on aggregate data, for instance on the country-year level.[3] Although such data can reveal broad patterns, it fails to capture the immediate impact of conflict, thereby limiting the direct applicability of these studies to policy formulation. Second, the task of modeling disaggregated data on conflict-related displacement presents significant challenges in itself. Such data usually stems from volatile conflict environments, leading to issues such as aggravated sampling noise and partially missing data (Sarzin, 2017). Moreover, the generally limited availability of data in conflict-affected areas makes it difficult to statistically account for spatial and temporal dependencies that may characterize fine-grained displacement data.

In this article, we aim to overcome both of these challenges. We compile weekly panel data covering more than 8 million internal displacements and 19,000 conflict events in 74 districts of Somalia between 2017 and 2023. We highlight several empirical issues that potentially arise when working with disaggregated data from conflict zones. To address these issues, we propose a specialized Bayesian statistical framework. This framework is grounded in both theoretical considerations on the dynamics of human displacement in Somalia and in stylized empirical facts derived from the data. The model integrates Bayesian factor models (Conti et al., 2014), dynamic linear models (Kim and Nelson, 1999), distributed lag models (Schwartz, 2000) as well as Bayesian shrinkage priors (Piironen and Vehtari, 2017). Combining these components in a single panel

[1] https://www.unhcr.org/global-trends-report-2022

[2] A review of empirical and theoretical literature on conflict and displacement is provided in Sec. A1.

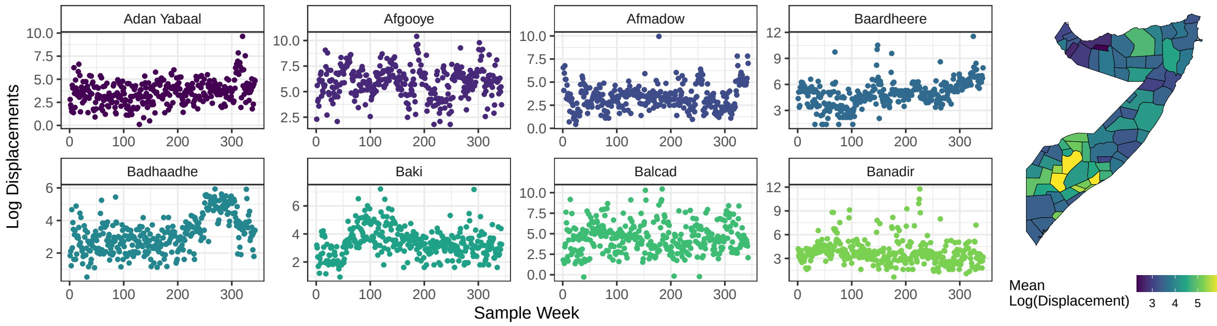
[3] Exceptions include studies using mobile phone tracking (Tai et al., 2022) or analysis of refugee flows from Ukraine in 2022 (Wycoff et al., 2023), which underscore the rapid impact of conflict on displacement.

regression framework allows us to flexibly account for latent spatio-temporal dependencies while ensuring robustness against overfitting noisy data.

Applying the model to displacement data from Somalia allows us to provide an in-depth analysis of the short-run impact of different types of conflict events on displacement. Our findings reveal a rapid and non-linear displacement response post-conflict, with significant heterogeneity in effects dependent on the nature of conflict events. In addition, we demonstrate the utility of the approach in the context of displacement forecasting, where our framework outperforms several benchmark models in producing short-run displacement predictions.

This article hence makes three key contributions. First, we introduce a statistical framework that can be used to explore and analyze the underlying drivers of forced displacement and has a broader application in Bayesian regression analysis of high-resolution panel data. The model can be used to evaluate predictions obtained from theoretical models on forced displacement or to conduct scenario analyses. In addition, models that can credibly attribute displacement to conflict have applications in the human rights context, and have for instance been used in war crime trials in The Hague (Ball and Asher, 2002). Second, we provide policy-relevant evidence on the impact of conflict on internal displacement in Somalia, contributing directly to the literature on short-run drivers of displacement. Given the challenges in data collection and modeling, previous research in this area has partially yielded inconsistent results, see the literature review in Sec. A1. The obtained impact estimates can further be used to inform policymakers, NGOs, and complementary theoretical and simulation-based works, relying for instance on agent-based modeling approaches. Third, our evaluation of the model's predictive power contributes to the literature on displacement forecasting. This evaluation is especially relevant considering the previously demonstrated shortcomings of black-box machine learning tools in predicting displacement in Somalia (Pham and Luengo-Oroz, 2022).

The remainder of this article is structured as follows. Sec. 2 introduces the data we use and highlights challenges when modeling detailed displacement data. In Sec. 3, we propose a specialized Bayesian statistical framework for modeling conflict and displacement. Sec. 4 presents our results on the short-term dynamics of conflict-driven displacement and the results of the displacement forecasting exercise. Sec. 5 concludes with key insights for empirical studies on conflict and displacement, strategic considerations for policy development, and pathways for future research.



(a) **(b)**

Figure 1: Patterns of observed internal displacements. (a) shows weekly time series for eight example districts across the sample period from January 2017 to July 2022. y-axis is on a log scale. x-axis indicates weeks in the sample period. (b) shows average log displacements in each of the 74 districts of Somalia.

2 Data Description

In this section, we describe the two main data sets from which we source the variables used for the empirical analysis in detail, highlighting potential methodological challenges. Limitations and potential shortcomings of the data are discussed in Sec. A3.

2.1 Data on Internal Displacement

The dependent variable of interest is the number of internal displacements reported in the 74 districts of Somalia for each week from January 2017 to July 2023, a period of $T = 343$ weeks. The data is obtained from the *Protection and Return Monitoring Network (PRMN) Somalia* led by the UNHCR. This monitoring network documents the number of internally displaced persons (IDPs) every week (UNHCR, 2017). The data is collected by the Norwegian Refugee Council (NRC) and local partners on behalf of the UNHCR. Specifically trained monitors measure the movements of displaced populations at strategic points, including transit sites and IDP settlements. The data are obtained based on individual or group interviews of displaced persons and make use of standardized forms to collect information. Quality control procedures are in place to ensure a reasonable level of data quality after the partners upload the results of these interviews onto an online platform.

Fig. 1 visualizes the average log displacement in each district as well as eight example time series of observed log displacements. A preliminary investigation of these series makes some stylized facts of the data apparent. First, the unconditional means of the time series vary strongly across districts, reflecting both differences in population size and resilience to factors leading to displacement. Furthermore, the displacement series often follow district-specific and complex trending patterns. These slow-moving trends potentially reflect slow-onset developments such as droughts or changing economic conditions. While the variance around these trends is relatively small in some districts, the behavior of the displacement flow series in other districts is much more volatile.

In addition to these within-district temporal dependency patterns, cross-district spatial dependencies are to be expected as well. For instance, several districts in Southern Somalia are connected via the major rivers *Juba* and *Shebelle*. Flash floods caused by rising river levels can lead to simultaneous displacement fluctuations in these districts. Similarly, droughts or economic crises are likely to affect displacement in multiple districts jointly. A major modeling challenge in this context is that no explicit measurements of river levels, flood depth, and economic or drought developments are available. Moreover, for around a third of the data points, no displacement counts are reported. We aim to address all of these stylized facts and modeling issues explicitly in the statistical framework and estimation algorithm outlined in Sec. 3.

2.2 Data on Conflict and Conflict-Related Events

The main independent variable of interest is the occurrence of conflict and conflict-related events, which we measure using data obtained from the *Armed Conflict Location & Event Data Project* (ACLED). This event-based data set collects occurrences of political violence and protest, including for instance military battles, suicide bombings, and riots (Raleigh et al., 2010). ACLED includes information on events by date, location, agent, and event type. It is one of the highest quality data sets on violent events on a disaggregated scale and is widely used in research analyzing the causes and effects of conflict (Eck, 2012; Thalheimer et al., 2023; Oh et al., 2024). Data is collected via traditional media, reports of international institutions and NGOs, local partners, and social media channels such as Twitter or Telegram. The reliability of the data is ensured via several layers of quality control, including source control and peer review mechanisms, involving academic researchers, policy and practitioner communities as well as country experts.

ACLED events are categorized into several event *types*, allowing us to investigate the impact of different classes of conflict and conflict-related events on displacement. Specifically, we will base our investigation on four ACLED categories. First, *battles* include events that are defined as a violent interaction between two organized armed groups at a particular time and location. For instance, this category includes clashes of armed government forces and rebel groups. Second, *explosions and remote violence* events are defined as 'one-sided violent events in which the tool for engaging in conflict creates asymmetry by taking away the ability of the target to respond' and include suicide bombings or artillery shelling. Third, *violence against civilians* events capture any event 'where an organized armed group deliberately inflicts violence upon unarmed non-combatants', such as shootings, sexual violence, or kidnapping of civilians. Fourth, the event type *strategic development* captures events 'regarding the activities of violent groups that is not itself recorded as political violence, yet may trigger future events or contribute to political dynamics within and across states'. Events

(a) Battles. **(b)** Explosive / Remote. **(c)** Strategic Developments. **(d)** Viol. against civilians.

Figure 2: Average weekly occurrence of various types of conflict and conflict-related events from January 2017 to July 2023. Data is presented on a log scale, and scales differ across panels. In case no events are recorded during the sample period, the corresponding districts are shown in grey.

that fall in this category include the formation of new rebel groups, the establishment of new headquarters, or non-violent transfers of territory. All of the above definitions are taken from the ACLED codebook[4] and are repeated here for completeness. Fig. 2 plots the average number of different types of conflict events across districts in Somalia. In total, in the sample period between January 2017 and July 2023, around 19,000 relevant events are recorded in the ACLED database, where 52.6% are coded as battles, 23.6% are coded as explosive and remote violence, 17.4% fall in the category violence against civilians and the remaining 6.4% are strategic developments. More context on the civil war in Somalia is given in Sec. A2.

3 Statistical Framework

Let y_{it} denote log displacements in district i $(i = 1, \ldots, N)$ in week t $(t = 1, \ldots, T)$. We model y_{it} as the sum of a component C_{it}, representing conflict-driven displacement, and O_{it}, representing displacement due to non-conflict events:

$$y_{it} = C_{it} + O_{it}. \tag{3.1}$$

Conflict-driven displacement is further decomposed as $C_{it} = A_{it} + I_{it} + F_{it}$, where A_{it} are anticipation effects, I_{it} are immediate effects on impact and F_{it} are aftermath effects of conflict events. We assume that these anticipation, impact, and aftermath components can be modeled using a distributed lead and lag structure

[4] ACLED Codebook Version 1 from January 2021, available online at the ACLED website https://acleddata.com/.

$$
\begin{aligned}
A_{it} &= \sum_{s\in\mathcal{C}}\sum_{j=1}^{4} c_{s,i,t+j}\beta_{sj}^{-} \\
I_{it} &= \sum_{s\in\mathcal{C}} c_{s,i,t}\beta_{s} \\
F_{it} &= \sum_{s\in\mathcal{C}}\sum_{j=1}^{8} c_{s,i,t-j}\beta_{sj}^{+}
\end{aligned}
\tag{3.2}
$$

where $\mathcal{C}$ is the set containing the four conflict event types and $c_{s,i,t}$ denotes the count of conflict events of type s in district i and week t. This specification allows us to investigate concurrent conflict impacts (via β_s), pre-emptive relocation efforts in the month before (via β_{sj}^{-}), and aftermath effects in the two months after conflict events (via β_{sj}^{+}).

The specification of O_{it} aims to capture latent spatio-temporal dependencies in displacement that are not captured by the conflict event indicators. We assume that displacement due to events other than conflict can be represented as

$$
O_{it} = \alpha_{it} + \boldsymbol{\lambda}_i'\boldsymbol{f}_t + \boldsymbol{x}_{it}'\boldsymbol{\delta} + \varepsilon_{it} \qquad \varepsilon_{it} \sim \mathcal{N}(0,\sigma_i^2) \tag{3.3}
$$

where α_{it} is a latent trend component, responsible for capturing slow-moving within-district variables, such as droughts and economic factors. This dynamic component is modeled using a random walk structure $\alpha_{it} = \alpha_{it-1} + \eta_{it}$ where $\eta_{it} \sim \mathcal{N}(0,\theta_i)$. A factor structure with Q latent factors $\boldsymbol{f}_t = (f_{1t},\dots,f_{Qt})'$ and corresponding loadings $\boldsymbol{\lambda}_i = (\lambda_{1i},\dots,\lambda_{Qi})'$ accounts for latent cross-district spatial dependencies. The factors $\boldsymbol{f}_t$ represent unmeasured variables relevant to multiple districts, including for example river levels or supra-regional drought developments. The linear component $\boldsymbol{x}_{it}'\boldsymbol{\delta}$ captures the impact of available control variables collected in $\boldsymbol{x}_{it}$. Here, we include two variables measuring the UNHCR-reported share of displacements due to drought and floods in district i and week t, respectively. Broadly speaking, this allows us to control for average differences between conflict-related displacement and displacement due to droughts and floods. Finally, ε_{it} represents a heteroskedastic error term that captures potentially varying quality of measurement and heterogeneous error volatilities across districts. Importantly, the factor structure and the heteroskedastic error term marginally imply a full $N \times N$ cross-district error covariance matrix $\boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma}$ where $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1,\dots,\boldsymbol{\lambda}_N)'$ and $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_i^2)$.

3.1 Prior Elicitation

We pursue a Bayesian approach to parameter estimation. The Bayesian paradigm allows for a fully probabilistic and therefore intuitive interpretation of the obtained parameter estimates and forecasts, facilitating communication of results to potential stakeholders. Bayesian estimation requires soliciting suitable prior distributions on all relevant parameters. Weakly informative $N(0, 100)$ priors are chosen for each β_s and $\boldsymbol{\delta}$. The variance parameters σ_i^2 are assumed to follow weakly informative inverse gamma distributions $\sigma_i^2 \sim \mathcal{IG}(2.5, 1.5)$. For the state equation variances, we choose $\theta_i \sim \mathcal{IG}(1, 0.005)$, which implies relatively smooth paths of the latent trend components α_{it}.

The factors are assumed to arise from independent Gaussian densities with unit variances, i.e., $\boldsymbol{f}_t \sim \mathcal{N}(0, \boldsymbol{I}_Q)$ where $\boldsymbol{I}_Q$ is the Q-dimensional identity matrix. This implies that all the information regarding the error correlations across districts is summarized in the factor loadings $\boldsymbol{\lambda}_i$. Here, we choose an informative prior distribution that reflects the idea that the error correlation matrix across districts is potentially sparse and that the latent shocks to displacement are uncorrelated for certain district pairs. Specifically, we rely on the horseshoe prior of Carvalho et al. (2009) due to its well-documented ability to regularize noisy signals and the ease of implementation due to the algorithm proposed in Makalic and Schmidt (2015). This prior places most probability mass on zero. Estimates of $\boldsymbol{\lambda}_i$ will therefore only deviate significantly from zero in case the data is informative enough, encouraging in turn sparse estimates of $\boldsymbol{\Lambda\Lambda'}$. Importantly, this prior mitigates the issue of overfitting when attempting to estimate the $N \times N$ covariance matrix of $\boldsymbol{\varepsilon}_i$ from potentially noisy and incomplete data.

Finally, a prior on the distributed lead and lag coefficients $\boldsymbol{\beta}^+$ and $\boldsymbol{\beta}^-$ needs to be chosen. Here, we superimpose the a priori assumption that the impact of conflict on displacement plays out relatively smoothly over time. In a setting with noisy data from conflict zones, this assumption provides further structure and mitigates overfitting. To enforce smooth distributed lag functions for the anticipation and aftermath effects, we assume a second-order random walk on the distributed lag coefficients

$$
\begin{aligned}
\beta^+_{s,j} &= 2\beta^+_{s,j-1} - \beta^+_{s,j-2} + \zeta^+_{s,j} \qquad \zeta^+_{s,j} \sim \mathcal{N}(0, \tau^+_s) \\
\beta^-_{s,j} &= 2\beta^-_{s,j-1} - \beta^-_{s,j-2} + \zeta^-_{s,j} \qquad \zeta^-_{s,j} \sim \mathcal{N}(0, \tau^-_s),
\end{aligned}
\tag{3.4}
$$

with heavy-tailed innovations based on standard Cauchy priors $\tau^+_s, \tau^-_s \sim \mathcal{C}(0, 1)$, as suggested in Lang and Brezger (2004).

3.2 Parameter Estimation and Model Identification

The proposed model can be estimated efficiently using Markov Chain Monte Carlo (MCMC) methods. In particular, the static factor structure enables the parallelization of otherwise computationally costly sampling steps, such as updating the N latent state space components $\boldsymbol{\alpha_i}$. Parallel updates of factors and loadings further streamline the computational workload. Efficient estimation of the model via MCMC is convenient, as this facilitates the handling of missing data points via imputation during estimation.

A sketch of the Gibbs sampling algorithm we use to sample iteratively from the conditional posterior distributions of the parameters is as follows. First, missing data points are imputed from $y_{it} \sim \mathcal{N}(C_{it} + O_{it}, \sigma_i^2)$. Second, the latent states α_{it} are updated for all i in parallel, conditional on the factors $\boldsymbol{f}_t$. These updates can be conducted highly efficiently using simulation smoothing techniques as per Chan and Jeliazkov (2009) or McCausland et al. (2011). The regression coefficients $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are then jointly simulated via one large Bayesian regression update, conditional on the states α_{it} and the factor structure. The prior precision matrix for $\boldsymbol{\beta}$, implied by Eq. (3.4), can be constructed using the ideas outlined in Chan and Jeliazkov (2009). Factors $\boldsymbol{f}_t$ for all t can be updated in parallel, followed by parallel Bayesian regression updates to obtain samples for the loadings $\boldsymbol{\lambda}_i$ for all i. Finally, the variance parameters are updated, using data augmentation for the horseshoe parameters as in Makalic and Schmidt (2016), data augmentation for the Cauchy prior updates as in Lang and Brezger (2004), and relying on standard conjugate updates for the remaining parameters.

Identification of the factors $\boldsymbol{f}_t$ and the corresponding loadings collected in $\boldsymbol{\Lambda}$ requires a set of restrictions to resolve identification issues regularly appearing in the context of latent factor models (Frühwirth-Schnatter et al., 2023). However, the cross-district covariances encoded in $\boldsymbol{\Sigma} + \boldsymbol{\Lambda}\boldsymbol{\Lambda}'$ are identified under mild conditions. As we are not directly interested in inference on $\boldsymbol{f}_t$, we do not restrict the loadings and factors beyond the a priori assumptions of independence across factors and unit variances for all factors.

For selecting an appropriate number of factors Q, we let the data decide which factors are deemed relevant and which are eliminated from the model via the horseshoe prior on $\boldsymbol{\lambda}_i$. We found that increasing the number of factors beyond five does not considerably change results in our setting and therefore use five factors as the standard setting in what follows. An alternative, but computationally more involved approach that treats the number of factors as a random quantity to be estimated, is discussed in Frühwirth-Schnatter et al. (2023).

4 Results

The presented results are based on 6,000 posterior iterations, where the first 2,000 draws are discarded, and every 2nd draw is saved to reduce the storage demand of the results. One estimation run using `R` takes around

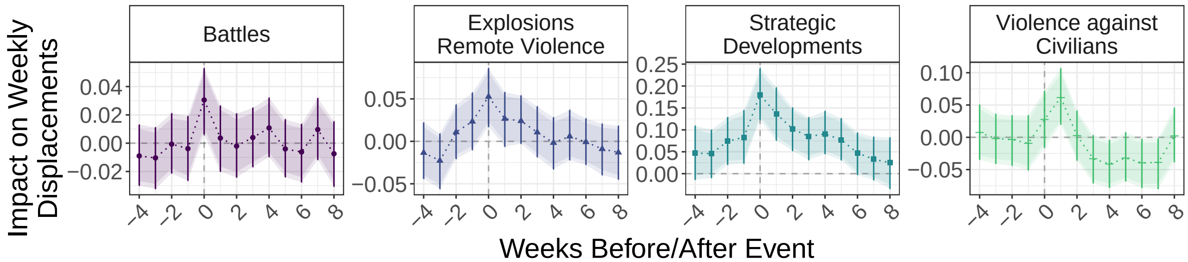


Figure 3: Posterior distributions of the estimated distributed lead and lag coefficients $\boldsymbol{\beta}$, $\boldsymbol{\beta}^{+}$ and $\boldsymbol{\beta}^{-}$ for the four considered types of conflict events. Shaded areas correspond to 95% credible intervals.

10 minutes on a laptop with an AMD Ryzen 5 5500U CPU. In general, convergence of the Markov chains is rapid and the Gibbs sampler is mixing well.

4.1 The Short-Run Impact of Conflict on Displacement in Somalia

The main results of the estimation exercise are summarized in Fig. 3 which gives point estimates and estimated uncertainty bounds for $\boldsymbol{\beta}$, $\boldsymbol{\beta}^{+}$ and $\boldsymbol{\beta}^{-}$, the distributed lead and lag coefficients of the four conflict event indicators. These estimates indicate that, on average, a single battle event is associated with a 3% increase in displacements in the same week, with no detectable aftermath or anticipation effects. The effect of explosions and remote violence events is slightly stronger, with a 5% increase on average in the same week, and more pronounced aftermath effects. These stronger and more persistent effects are potentially due to the destruction of buildings, which is likely following this type of event. Significant anticipation, impact, and aftermath effects are observed for strategic developments, with a peak increase of between 15% and 20% within the same week, and detectable effects up to more than a month after the event occurred. Strategic developments often correspond to high-visibility events with substantive information flow, such as land transfers or troop movements, signaling shifts in the balance of power, and potential for future violence.[5] These events appear to prompt both pre-emptive relocation and strong aftermath, potentially in groups that are now exposed to new threats. Evidence for the important role of structural changes in conflict-driven displacement is also reported in Schon (2015). The pattern after events involving violence against civilians is more intricate, with a positive but insignificant point estimate in the week of the event, a stronger effect in the week after, and then some decreases in displacement occurrences in the following weeks. In general, these results illustrate the complex and varied nature of displacement dynamics in the context of different types

[5] Within the considered sample period, the composition of 'strategic developments' according to the ACLED sub-event categories is as follows. 'Agreement' accounts for 4%, 'Arrests' for 7%, 'Change to group/activity' for 14%, 'Disrupted weapons use' for 22%, 'Headquarters or base established' for 1%, 'Looting/property destruction' for 12%, 'Non-violent transfer of territory' for 34%, and 'Other' for 5% of the events. Developing methods that can robustly disentangle the effect of conflict by many, potentially very rare sub-event types is a promising avenue for future research.

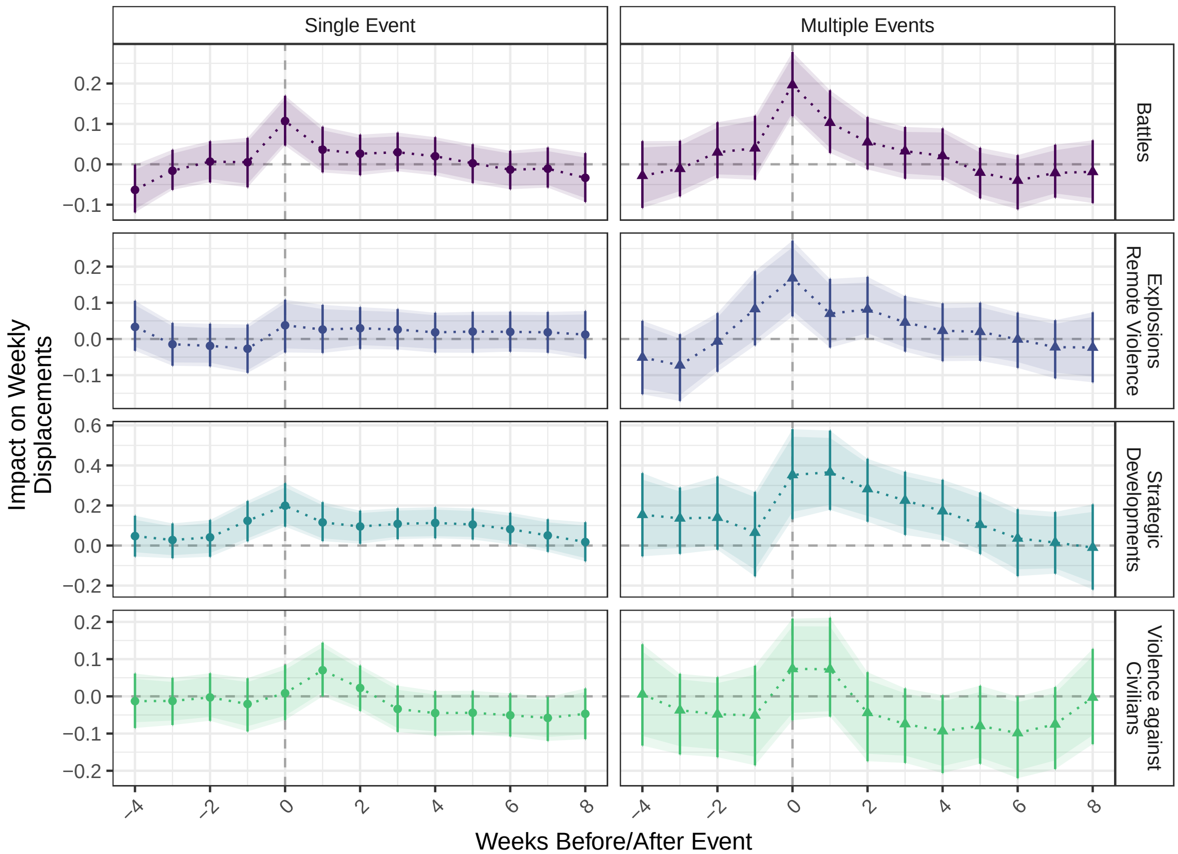


Figure 4: Posterior distributions of the estimated distributed lead and lag coefficients $\boldsymbol{\beta}$, $\boldsymbol{\beta}^{+}$ and $\boldsymbol{\beta}^{-}$ in the non-linear specification. Shaded areas correspond to 95% credible intervals.

of conflict events. While we can speculate what drives these empirical patterns, these hypotheses cannot be conclusively evaluated using the data at hand, warranting further research in the future.

We provide some additional results on non-linear impacts across the number of conflict events per week in Fig. 4. The left columns show results from an estimation run including leading and lagging binary indicators for weeks with a single conflict event. The right column shows results from an estimation run including binary indicators for weeks with multiple conflict events. We find that the average effects reported in Fig. 3 are likely driven by weeks in which more than one conflict event occurs. Specifically, the results point towards a non-linear relationship between displacement and conflict, where we observe larger displacement increases in weeks when many conflict events take place. This highlights the cumulative and compounding psychological and physical toll of frequent conflict occurrences within short periods. Similar evidence of non-linear effects in the sense that higher intensity of conflict leads to more displacement while this not being necessarily the case for low conflict intensity is for instance reported in Bohra-Mishra and Massey (2011).

Overall, our findings therefore indicate that displacement as a response to conflict materializes very rapidly, particularly following battles and explosions, with the majority of the impact typically observed

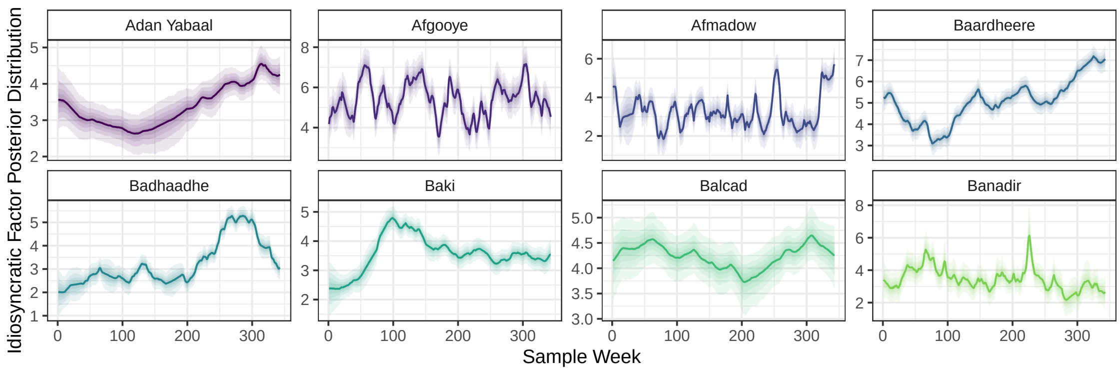


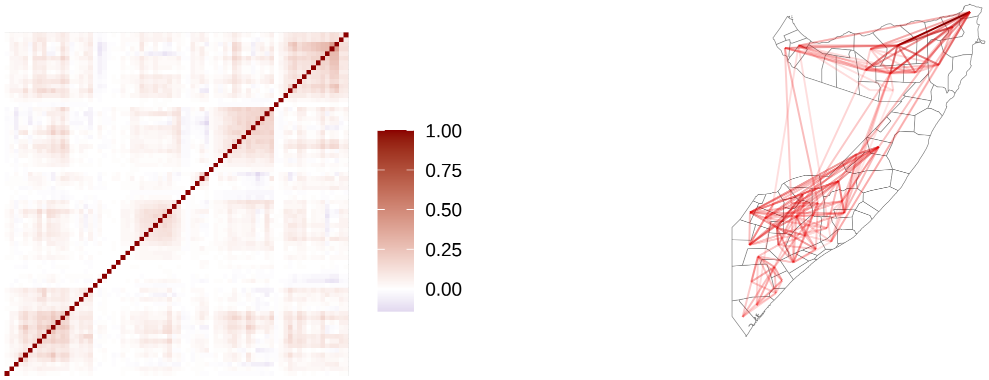
Figure 5: Posterior distributions of the estimated idiosyncratic trend components α_{it} for eight example districts across the sample period from January 2017 to July 2023. Shaded areas correspond to 95% credible intervals.

within 0-2 weeks after the event. Increases in displacement may last longer, up to several weeks, following strategic developments. A significant observation is the heterogeneous nature of displacement reactions to different types of conflict events, where both the magnitude and the temporal pattern of the effect vary. Furthermore, the impact of conflict on displacement is characterized by a non-linear pattern. Weeks with multiple conflict events exhibit a significantly stronger impact and more pronounced aftermath than weeks with a single event. A discussion on the internal and external validity of these results is provided in Sec. A4.

4.2 Results on Spatio-Temporal Dependencies

In Fig. 5, we provide the estimated posterior distributions of α_{it} for eight example districts. The estimation results reveal several distinct patterns concerning the trending behavior of the displacement series. In certain districts, the data is informative on highly intricate latent trends α_{it}. These trend components capture all sufficiently slow-moving, district-specific developments that correlate with displacement and are not accounted for by the cross-district factor structure, the included control variables, and the leads and lags of the conflict measures included in the model. The complex nature of the trend components α_{it} highlights the importance of such unmeasured within-district factors when modeling displacement patterns. The estimated paths of α_{it} further demonstrate that a conceptually simpler, classical panel specification based on district fixed effects – implying constant α_{it} for all i – is likely under-specified. Additional evidence for this claim is provided in Sec. 4.3.

Next, we focus on posterior mean estimates of the covariance structure across districts implied by the estimates of the idiosyncratic variances σ_i^2 and the factor loadings $\boldsymbol{\lambda}_i$. The implied correlation matrix is visualized in Fig. 6. It reveals a significant amount of cross-district correlation in the unobserved shocks to displacement. Blocks of correlated districts are visible in panel (a). Almost all non-zero correlations are estimated to be positive, implying predominantly positive co-movements in the latent shocks to displacement



(a) Error Correlation Estimates.

(b) Partial Correlation Network.

Figure 6: Estimated spatial dependency structure. (a) shows posterior mean estimates of error correlations across districts. Ordering of rows and columns of the matrices is based on a clustering of highly correlated districts for visualization purposes. (b) shows an estimated partial correlation network derived from the posterior distribution of $\mathbf{\Sigma} + \mathbf{\Lambda\Lambda'}$. Stronger partial correlations correspond to less transparent edges.

$\boldsymbol{\varepsilon}_i$. Suspected unmeasured factors such as flash floods or political developments concerning more than one district are in line with this finding. The correlation matrices are further estimated to be pronouncedly sparse, i.e., including many zero elements, which is a consequence of the shrinkage prior specified on the elements of $\boldsymbol{\lambda}_i$.

Panel (b) of Fig. 6 shows a partial correlation network, based on the estimated error covariance matrix. Posterior mean estimates of the partial correlations are used and partial correlations where zero is included in the 95% credible interval are dropped. More transparent lines correspond to smaller partial correlations. The resulting network structure is broadly clustered into two geographically distinct blocks of districts. The first cluster, in Southern Somalia, is characterized by districts close to the capital district, connected by rivers and subject to flash flooding. The second cluster in the North corresponds to Somaliland and Puntland, two larger regions that have long tried to formally declare independence from the Southern part of Somalia. Somaliland and Puntland are further characterized by repeated armed clashes over disputed districts and provinces in their bordering region. It is worth noting that our model does detect these geo-clusters and dependency structures in the absence of any information on the spatial structure of the districts.

Finally, in Fig. 7, we visualize the fitted values of the model together with the raw data. Focusing on the previously showcased eight example districts again, we find that the model is able to follow the general trends in $y_{i,t}$ relatively well. To give a sense of overall in-sample fit, the coefficient of determination of the Bayesian approach is around $R^2 = 0.64$. The results nonetheless demonstrate the difficulty of designing empirical models that can extract useful information from potentially noisy and highly fluctuating data in

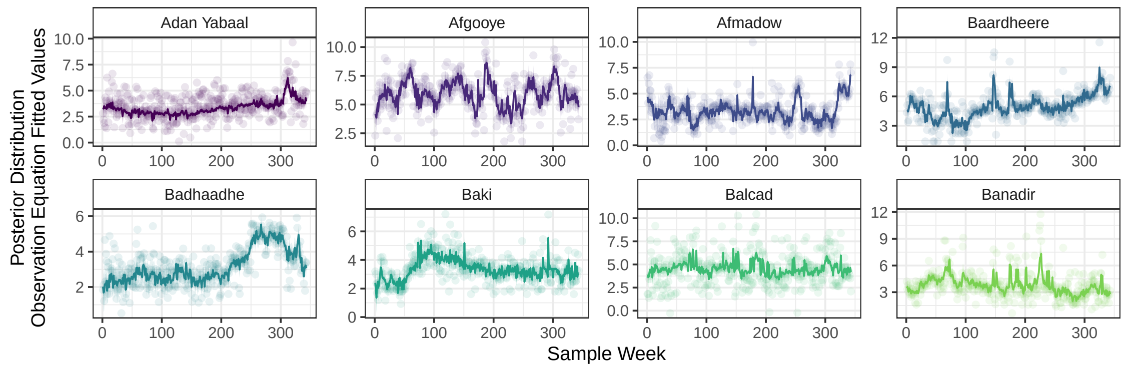


Figure 7: Observed log displacement $y_{i,t}$ together with posterior means of fitted values $\hat{y}_{i,t}$ for eight example districts across the sample period from January 2017 to July 2023.

the context of displacement modeling, as also noted in Pham and Luengo-Oroz (2022). In general, empirical modeling of human mobility patterns over time is a highly challenging task (Bijak et al., 2019).

4.3 Humanitarian Forecasting Exercise

Recent contributions have shown that the widely and traditionally used gravity models of migration are not well suited for forecasting migratory movements, and that well-designed time series approaches can outperform gravity models by a significant margin (Welch and Raftery, 2022; Beyer et al., 2022). Moreover, even highly flexible machine learning frameworks tend to perform poorly when forecasting displacement on a disaggregated level (Pham and Luengo-Oroz, 2022). At the same time, forecast-based humanitarian efforts are attracting growing attention (Thalheimer et al., 2021). For these reasons, we conduct an out-of-sample forecasting exercise to evaluate how well the proposed model performs compared to several benchmarks. In addition, the model specified in Sec. 3 is relatively flexible, warranting some evaluation of its out-of-sample performance to mitigate concerns about overfitting.

We focus on obtaining one-week ahead predictions for displacements for all 74 districts under consideration. For this, we repeatedly split the data set into a training sample with a length of 291 weeks and a single one-step ahead hold-out week. The parameters of the competing modeling frameworks are estimated using the training sample and are subsequently used to predict displacement in the test week. This exercise is repeated 52 times, shifting the time window of the training sample by one week each time, until eventually all weeks in the last year of the full sample period have been used as holdout weeks once.

We compare ten different forecasting methods. First, we estimate several simple benchmarks, including a random walk model, rolling averages with varying window sizes, and an AR(1) model. Long-run average and long-run median displacement in each district, computed across the whole sample, are included as simplest benchmarks. Second, two fixed effects panel regression models leveraging a full set of week and district

Table 1: Results of Forecasting Exercise.

Model	RMSE	MAE	MAPE	Corr.
Random Walk	1.000	1.000	1.000	0.593
AR(1)	0.775	0.948	0.863	0.636
Long Run Average	1.114	1.171	1.104	0.465
Long Run Median	1.195	1.204	1.184	0.434
Rolling Average (4 Weeks)	0.946	0.980	0.958	0.606
Rolling Average (8 Weeks)	0.940	0.980	0.949	0.605
Rolling Average (12 Weeks)	0.931	0.976	0.943	0.607
OLS 2FE DL (incl. zeros)	1.527	1.359	1.935	0.390
OLS 2FE DL (excl. zeros)	0.970	1.113	0.900	0.456
Bayesian DL-DLM	**0.737**	**0.907**	**0.804**	**0.639**

Note: 'RMSE' is the root mean squared error, 'MAE' is the mean average error, and 'MAPE' is the mean average percentage error of the point predictions, relative to the true values. 'RMSE', 'MAE', and 'MAPE' are reported relative to the random walk specification. 'Corr.' is the correlation of the point predictions and the true values. Results are averaged across 52 one-step-ahead hold-out samples.

fixed effects are included in the exercise. The first panel model treats unobserved displacement as zero observations, while the second one drops unobserved displacement from the data set. Finally, the Bayesian model introduced in Sec. 3 is estimated. In terms of covariates, all panel models include lagged conflict indicators from $t-1$ to $t-8$.

The results of the forecasting exercise are summarized in Tab. 1, where we report the root mean squared error (RMSE), the mean absolute error (MAE), and the mean absolute percentage error (MAPE) of the point forecasts, as well as the correlation of the point predictions and the true values. All criteria are averaged across the 52 hold-out periods. RMSE, MAE, and MAPE are reported relative to the random walk benchmark model. We find that forecasts based on simple rolling averages already outperform the random walk model as well as the panel fixed effects regression considerably. This is in line with the observation that the simplest benchmarks perform relatively well in displacement forecasting, as reported in Pham and Luengo-Oroz (2022). The fixed effects panel models also give considerably worse results than the simple AR(1) model. The Bayesian approach introduced in Sec. 3 can improve over all benchmark models in all four forecast quality criteria.

It is important to note that the results of this small forecasting exercise are no conclusive evidence that the proposed model will show any kind of 'optimal' forecasting performance in a given displacement forecasting setting. After all, the model is not explicitly designed with forecasting performance in mind. However, these results are a good indication that the additional complexity we introduce to capture latent spatio-temporal dependencies does not lead to severe overfitting concerns. Overall, the results in Tab. 1 are indicative of a certain strength of the proposed Bayesian framework when it comes to short-term displacement predictions. Further investigations and extensions towards fully-fledged early warning systems (Martin and Singh, 2019) or a forecasting tool for forced displacement flow matrices in the style of Welch and Raftery (2022) are promising future research avenues.

5 Discussion & Concluding Remarks

Drawing on unique and fine-grained displacement data from Somalia, this paper empirically investigates the short-run dynamics of conflict and displacement. We highlight several challenges that are likely to complicate working with similar data sets and develop a Bayesian statistical approach that overcomes these challenges. We point out that our considerations apply more generally to the analysis of large N large T panel data sets. The utility of the approach is demonstrated in two empirical exercises, focused on impact evaluation and humanitarian forecasting. We find that the response of displacement to conflict is rapid, non-linear in the number of conflict events, and heterogeneous by conflict type. Finally, the model shows good forecasting performance relative to benchmark models. Besides potential use cases for empirical analyses and for humanitarian forecasting purposes, the modeling framework and derived impact estimates are of relevance for policymakers, non-governmental organizations, and complementary theoretical and simulation-based literature.

In terms of policy insights, we acknowledge that any real and definite solution concerning the root causes of displacement in Somalia will require ending the ongoing state of civil war. However, this goal will remain unlikely for some time to come, not only in Somalia, but in many other parts of the world that experience prolonged periods of conflict. As long as conflict situations are not fully resolved, a thorough understanding of the determinants of displacement is a necessary prerequisite to alleviate the manifold issues that displaced populations face and to design appropriate policies that focus on assistance, prevention, and relocation (Engel and Ibáñez, 2007; Dirikgil, 2022). In this context, our findings highlight the importance of rapid response teams, as displacement tends to occur swiftly following conflict. The relevance of daily and weekly time frames in the context of conflict-driven displacement cannot be overstated. Another notable observation is the apparent lack of anticipation effects in the case of battles, which suggests that civilians either lack adequate forewarning, choose not to relocate preemptively, or lack the resources to do so. Disseminating information on battle-related developments, if available, could facilitate preemptive relocation efforts. Our findings further underscore the importance of strategic developments. It appears that such events hold significant predictive power for displacement. In addition, local populations seemingly anticipate some of these events. Training and empowering local communities to report anticipated events could therefore support anticipatory action.

Concerning quantitative research on conflict-driven displacement, this study highlights the importance of a fine-grained perspective. Our findings suggest that relying on aggregate or stock data may obscure important details, limiting the depth and accuracy of empirical analysis. In general, a more nuanced approach to the broad concept of 'conflict' is likely to unveil patterns and dynamics that are otherwise overlooked. Hence, there is a significant need for better and more granular data, not only on displacement but also on environmental drivers such as droughts and floods.

In concluding the article, several pathways for future research and modeling approaches emerge. One key area is the exploration of differential vulnerabilities of populations, which could be explored potentially through hierarchical random effects models. Another avenue is the investigation of spatial spillovers, which can illuminate the broader regional impacts of conflict. Predictive modeling of inflowing IDPs, which focuses on where people move to, is another promising direction and could greatly enhance our ability to anticipate and respond to displacement events.

Acknowledgements

The authors gratefully acknowledge helpful comments from the participants of the *eXplore! Workshop 'Migration Modelling and Climate Change Research'* in Vienna, the *'GENPOP/POPCLIMA Research Seminar'* at the University of Bologna, the *16th International Conference on Computational and Methodological Statistics* in Berlin, and research seminars at the *Center for Policy Research on Energy and the Environment* at Princeton University and the International Institute for Applied Systems Analysis.

Disclosure Statement

The authors report there are no competing interests to declare.

REFERENCES

Abel, G. J., Brottrager, M., Cuaresma, J. C., & Muttarak, R. (2019). Climate, conflict and forced migration. *Global Environmental Change*, *54*, 239–249.

Adhikari, P. (2013). Conflict-induced displacement, understanding the causes of flight. *American Journal of Political Science*, *57*(1), 82–89.

Admasu, Y., Alkire, S., Ekhator-Mobayode, U. E., Kovesdi, F., Santamaria, J., & Scharlin-Pettee, S. (2021). A multi-country analysis of multidimensional poverty in contexts of forced displacement.

Akee, R. K., Basu, A. K., Chau, N. H., & Khamis, M. Ethnic fragmentation, conflict, displaced persons and human trafficking: An empirical analysis. In: In *Migration and culture*. Emerald Group Publishing Limited, 2010.

Ball, P., & Asher, J. (2002). Statistics and Slobodan: Using data analysis and statistics in the war crimes trial of former President Milosevic. *Chance*, *15*(4), 17–24.

Barro, R. J. (2001). Human capital and growth. *American economic review*, *91*(2), 12–17.

Benhabib, J., & Spiegel, M. M. (1994). The role of human capital in economic development: Evidence from aggregate cross-country data. *Journal of Monetary economics*, *34*(2), 143–173.

Beyer, R. M., Schewe, J., & Lotze-Campen, H. (2022). Gravity models do not explain, and cannot predict, international migration dynamics. *Humanities and Social Sciences Communications*, *9*(1), 1–10.

Bijak, J., Disney, G., Findlay, A. M., Forster, J. J., Smith, P. W., & Wiśniowski, A. (2019). Assessing time series models for forecasting international migration: Lessons from the United Kingdom. *Journal of Forecasting*, *38*(5), 470–487.

Bohra-Mishra, P., & Massey, D. S. (2011). Individual decisions to migrate during civil conflict. *Demography*, *48*(2), 401–424.

Carvalho, C. M., Polson, N. G., & Scott, J. G. Handling sparsity via the horseshoe. In: In *Artificial intelligence and statistics*. PMLR. 2009, 73–80.

Cassanelli, L. V. (2016). *The shaping of Somali society*. University of Pennsylvania Press.

Chan, J. C. C., & Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, *1*, 101–120.

Cohen, D., & Soto, M. (2007). Growth and human capital: Good data, good results. *Journal of economic growth*, *12*(1), 51–76.

Conte, A., & Migali, S. (2019). The role of conflict and organized violence in international forced migration. *Demographic Research*, *41*, 393–424.

Conti, G., Frühwirth-Schnatter, S., Heckman, J. J., & Piatek, R. (2014). Bayesian exploratory factor analysis. *Journal of econometrics*, *183*(1), 31–57.

Crippa, A., d'Agostino, G., Dunne, P., Pieroni, L., et al. (2022). *Conflict as a cause of migration* (tech. rep.). University Library of Munich, Germany.

Czaika, M., & Kis-Katos, K. (2009). Civil conflict and displacement: Village-level determinants of forced migration in Aceh. *Journal of peace research*, *46*(3), 399–418.

Davenport, C., Moore, W., & Poe, S. (2003). Sometimes you just have to leave: Domestic threats and forced migration, 1964-1989. *International Interactions*, *29*(1), 27–55.

Devakumar, D., Birch, M., Osrin, D., Sondorp, E., & Wells, J. C. (2014). The intergenerational effects of war on the health of children. *BMC medicine*, *12*(1), 1–15.

Dirikgil, N. (2022). Addressing the prevention of internal displacement: The right not to be arbitrarily displaced. *Journal of International Migration and Integration*, 1–26.

DRC. (2022). Global displacement forecast 2022. *Danish Refugee Council Technical Report*.

Eck, K. (2012). In data we trust? A comparison of UCDP GED and ACLED conflict events datasets. *Cooperation and Conflict*, *47*(1), 124–141.

Eklöw, K., & Krampe, F. (2019). *Climate-related security risks and peacebuilding in Somalia*. Stockholm International Peace Research Institute (SIPRI).

Engel, S., & Ibáñez, A. M. (2007). Displacement due to violence in Colombia: A household-level analysis. *Economic development and cultural change*, *55*(2), 335–365.

Fiala, N. (2015). Economic consequences of forced displacement. *The Journal of Development Studies*, *51*(10), 1275–1293.

Frühwirth-Schnatter, S., Hosszejni, D., & Lopes, H. F. (2023). Sparse Bayesian factor analysis when the number of factors is unknown. https://doi.org/10.48550/ARXIV.2301.06459

Gele, A. A., & Bjune, G. A. (2010). Armed conflicts have an impact on the spread of tuberculosis: the case of the Somali Regional State of Ethiopia. *Conflict and health*, *4*(1), 1–6.

George, J., & Adelaja, A. (2022). Armed conflicts, forced displacement and food security in host communities. *World Development*, *158*, 105991.

Harpviken, K. B. (2009). *Social networks and migration in wartime Afghanistan*. Springer.

Hatton, T. J. (2009). The rise and fall of asylum: What happened and why? *The Economic Journal*, *119*(535), F183–F213.

Healy, S., & Bradbury, M. (2010). Somalia: International and local peace processes: Bad situation, getting worse. *The World Today*, *66*(1), 18–19.

Ibáñez, A. M. (2014). Growth in forced displacement: Cross-country, sub-national and household evidence on potential determinants. *International handbook on migration and economic development*, 350–387.

Im, H., George, N., & Swan, L. E. (2022). Born in displacement: Psychosocial and mental health impacts of country of birth among urban Somali refugee youth. *Global Public Health*, 1–14.

Kim, C.-J., & Nelson, C. R. (1999). *State-space models with regime switching: Classical and Gibbs-sampling approaches with applications*. MIT Press.

Kovac, D., Efendic, A., & Shapiro, J. N. (2022). Forced displacement, exposure to conflict and long-run education and income inequality.

Lang, S., & Brezger, A. (2004). Bayesian P-Splines. *Journal of Computational and Graphical Statistics*, *13*, 183–212.

Liu, W. T., Lamanna, M., & Murata, A. (1979). Transition to nowhere: Vietnamese refugees in america.

Lucas Jr, R. E. (2015). Human capital and growth. *American economic review*, *105*(5), 85–88.

Mahamud, A., Burton, A., Hassan, M., Ahmed, J. A., Wagacha, J. B., Spiegel, P., Haskew, C., Eidex, R. B., Shetty, S., Cookson, S., et al. (2013). Risk factors for measles mortality among hospitalized Somali refugees displaced by famine, Kenya, 2011. *Clinical Infectious Diseases*, *57*(8), e160–e166.

Makalic, E., & Schmidt, D. F. (2015). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters*, *23*(1), 179–182.

Makalic, E., & Schmidt, D. F. (2016). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters*, *23*, 179–182.

Martin, S. F., & Singh, L. (2019). Big data and early warning of displacement. *Mobilizing global knowledge: Refugee research in an age of displacement*, 129–150.

Maystadt, J.-F., Hirvonen, K., Mabiso, A., Vandercasteelen, J., et al. (2019). Impacts of hosting forced migrants in poor countries. *Annual Review of Resource Economics*, *11*(1), 439–459.

McCausland, W. J., Miller, S., & Pelletier, D. (2011). Simulation smoothing for state space models: A computational efficiency analysis. *Computational Statistics and Data Analysis*, *55*, 199–212.

McEniry, M., Samper-Ternent, R., & Cano-Gutierrez, C. (2019). Displacement due to armed conflict and violence in childhood and adulthood and its effects on older adult health: the case of the middle-income country of Colombia. *SSM-population health*, *7*, 100369.

Melander, E., & Öberg, M. (2007). The threat of violence and forced migration: Geographical scope trumps intensity of fighting. *Civil Wars*, *9*(2), 156–173.

Menkhaus, K. (2004). Vicious circles and the security development nexus in Somalia. *Conflict, Security & Development*, *4*(2), 149–165.

Moore, W. H., & Shellman, S. M. (2004). Fear of persecution: Forced migration, 1952-1995. *Journal of Conflict Resolution*, *48*(5), 723–745.

Moore, W. H., & Shellman, S. M. (2006). Refugee or internally displaced person? To where should one flee? *Comparative Political Studies*, *39*(5), 599–622.

Morina, N., Akhtar, A., Barth, J., & Schnyder, U. (2018). Psychiatric disorders in refugees and internally displaced persons after forced displacement: A systematic review. *Frontiers in psychiatry*, *9*, 433.

Murphy, M., Bingenheimer, J. B., Ovince, J., Ellsberg, M., & Contreras-Urbina, M. (2019). The effects of conflict and displacement on violence against adolescent girls in South Sudan: the case of adolescent girls in the Protection of Civilian sites in Juba. *Sexual and reproductive health matters*, *27*(1), 181–191.

Nilsson, J. E., Brown, C., Russell, E. B., & Khamphakdy-Brown, S. (2008). Acculturation, partner violence, and psychological distress in refugee women from Somalia. *Journal of Interpersonal Violence*, *23*(11), 1654–1663.

Oh, W. S., Muneepeerakul, R., Rubenstein, D., & Levin, S. (2024). Emergent network patterns of internal displacement in Somalia driven by natural disasters and conflicts. *Global Environmental Change*, *84*, 102793.

Pham, K. H., & Luengo-Oroz, M. (2022). Predictive modeling of movements of refugees and internally displaced people: Towards a computational framework. *arXiv preprint arXiv:2201.08006*.

Piironen, J., & Vehtari, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Stasitistics*, *11*, 5018–5051.

Polonsky, J. A., Ronsse, A., Ciglenecki, I., Rull, M., & Porten, K. (2013). High levels of mortality, malnutrition, and measles, among recently-displaced Somali refugees in Dagahaley camp, Dadaab refugee camp complex, Kenya, 2011. *Conflict and health*, *7*(1), 1–9.

Raleigh, C., Linke, A., Hegre, H., & Karlsen, J. (2010). Introducing ACLED: An armed conflict location and event dataset: Special data feature. *Journal of peace research*, *47*(5), 651–660.

Salami, B., Iwuagwu, S., Amodu, O., Tulli, M., Ndikom, C., Gommaa, H., Lavin, T., & Kariwo, M. (2020). The health of internally displaced children in sub-Saharan Africa: A scoping review. *BMJ Global Health*, *5*(8), e002584.

Sarzin, Z. I. (2017). Stocktaking of global forced displacement data. *World Bank Policy Research Working Paper*, (7985).

Schmeidl, S. (1997). Exploring the causes of forced migration: A pooled time-series analysis, 1971-1990. *Social Science Quarterly*, 284–308.

Schmeidl, S. (2001). Conflict and forced migration: A quantitative review, 1964-1995. *Global migrants, global refugees: Problems and solutions*, 62–94.

Schon, J. (2015). Focus on the forest, not the trees: A changepoint model of forced displacement. *Journal of Refugee Studies*, *28*(4), 437–467.

Schon, J. (2016). The centrality of checkpoints for civilians during conflict. *Civil Wars*, *18*(3), 281–310.

Schon, J. (2019). Motivation and opportunity for conflict-induced migration: An analysis of Syrian migration timing. *Journal of Peace Research*, *56*(1), 12–27.

Schwartz, J. (2000). The distributed lag between air pollution and daily deaths. *Epidemiology*, *11*(3), 320–326.

Searle, C, & van Vuuren, J. H. (2021). Modelling forced migration: A framework for conflict-induced forced migration modelling according to an agent-based approach. *Computers, Environment and Urban Systems*, *85*, 101568.

Sokolowski, J. A., Banks, C. M., & Hayes, R. L. Modeling population displacement in the Syrian city of Aleppo. In: In *Proceedings of the winter simulation conference 2014*. IEEE. 2014, 252–263.

Spagat, M., Mack, A., Cooper, T., & Kreutz, J. (2009). Estimating war deaths: An arena of contestation. *Journal of Conflict Resolution*, *53*(6), 934–950.

Suleimenova, D., Bell, D., & Groen, D. (2017). A generalized simulation development approach for predicting refugee destinations. *Scientific reports*, *7*(1), 1–13.

Tai, X. H., Mehra, S., & Blumenstock, J. E. (2022). Mobile phone data reveal the effects of violence on internal displacement in Afghanistan. *Nature human behaviour*, *6*(5), 624–634.

Thalheimer, L., Schwarz, M. P., & Pretis, F. (2023). Large weather and conflict effects on internal displacement in Somalia with little evidence of feedback onto conflict. *Global Environmental Change*, *79*, 102641.

Thalheimer, L., Simperingham, E., & Jjemba, E. W. (2021). The role of anticipatory humanitarian action to reduce disaster displacement. *Environmental Research Letters*, *13*.

Thalheimer, L., & Webersik, C. Climate change, conflicts and migration. In: In *Environmental conflicts, migration and governance*. Bristol University Press, 2020, pp. 59–82.

UNESCO. (2016). *The impacts of internal displacement on education in Sub-Saharan Africa*. UNESCO. Internal Displacement Monitoring Centre.

UNHCR. (2017). Internal displacements recorded by Protection and Return Monitoring Network Somalia. Notes on methodology.

Verme, P., & Schuettler, K. (2021). The impact of forced displacement on host communities: A review of the empirical literature in economics. *Journal of Development Economics*, *150*, 102606.

Welch, N. G., & Raftery, A. E. (2022). Probabilistic forecasts of international bilateral migration flows. *Proceedings of the National Academy of Sciences*, *119*(35), e2203822119.

Wirtz, A. L., Pham, K., Glass, N., Loochkartt, S., Kidane, T., Cuspoca, D., Rubenstein, L. S., Singh, S., & Vu, A. (2014). Gender-based violence in conflict and displacement: qualitative findings from displaced women in Colombia. *Conflict and health*, *8*(1), 1–14.

Wycoff, N., Arab, A., Donato, K., Singh, L., Kawintiranon, K., Liu, Y., & Jacobs, E. (2023). Forecasting Ukrainian refugee flows with organic data sources. *International Migration Review*, 01979183231203931.

Yuen, A., Warsame, A., & Checchi, F. (2022). Exploring the temporal patterns and crisis-related risk factors for population displacement in Somalia (2016-2018). *Journal of Migration and Health*, *5*, 100095.

SUPPLEMENTARY MATERIAL

A1 Theoretical and Empirical Literature on Conflict and Forced Displacement

Numerous adverse outcomes of forced displacement populations have been documented. Displaced populations are often living in extreme poverty (Admasu et al., 2021), have higher mortality rates (Polonsky et al., 2013) and increased rates of childhood malnutrition (Polonsky et al., 2013). The hardships that displaced populations endure lead to trauma and other mental health issues such as post-traumatic stress disorder (Morina et al., 2018). Often, the living situation in displacement camps leads to deteriorating physical health (Salami et al., 2020), for instance via exposure to diseases such as measles (Polonsky et al., 2013; Mahamud et al., 2013) and tuberculosis (Gele and Bjune, 2010). Displaced women frequently experience gender-based violence, including domestic violence (Nilsson et al., 2008) and sexual violence, both from partners (Murphy et al., 2019) and non-partners (Wirtz et al., 2014). Moreover, displaced populations are exposed to criminal groups involved in human trafficking (Akee et al., 2010). In addition to the impacts on the displaced communities, hosting communities can be negatively impacted as well. Inflows of displaced populations put local health infrastructure systems to test (Maystadt et al., 2019), can lead to food insecurity (George and Adelaja, 2022), and correlate with wage compression in local labor markets (Verme and Schuettler, 2021).

Apart from the humanitarian crisis, forced displacement poses a severe challenge for long-run economic development. In addition to the direct negative effect on the economy due to forcing people out of their productive activities, displacement interrupts children's education paths, often for prolonged periods (UNESCO, 2016). Kovac et al. (2022) document that adverse educational outcomes due to forced displacement are persistent and do not vanish in the long run. Similarly, experiencing displacement and conflict in early childhood is correlated to poor health outcomes at older ages (Devakumar et al., 2014; McEniry et al., 2019). Moreover, several studies report intergenerational transmission of mental health issues such as trauma and post-traumatic distress disorder, especially for children born in displacement (Im et al., 2022). All of these channels combined lead to a persistent negative long-run impact of displacement experiences on economic outcomes on the household level (Fiala, 2015). As a result, *scarring effects* have a severe negative impact on a country's long-run human capital stock, which in turn limits the potential for economic development and poverty reduction (Benhabib and Spiegel, 1994; Barro, 2001; Cohen and Soto, 2007; Lucas Jr, 2015).

Theoretical analysis of forced displacement and conflict usually leverages modeling frameworks focused on individual-level or household-level decision-making problems. This 'choice-centered' approach is based on the idea that individuals choose to either leave or stay, even in difficult habitual circumstances such as violent conflict. Agents observe information in their environment and use this information to evaluate their security and the threat to their lives. The choice between leaving and staying is then made based on whether the perceived physical threat to life in a conflict situation surpasses a threshold that represents an 'opportunity cost of fleeing' (Davenport et al., 2003; Moore and Shellman, 2004; Moore and Shellman, 2006; Adhikari, 2013). This opportunity cost is usually thought of being a function of several factors, including regional and supra-regional social networks, transportation costs, economic opportunities and emotional attachment to home, personal risk aversion, and the quality of the available information set (Engel and Ibáñez, 2007; Harpviken, 2009; Adhikari, 2013; Ibáñez, 2014; Schon, 2019). Many of these factors are themselves a function of conflict. For instance, ongoing violent conflict may reduce economic opportunities at home and increase transportation costs via potential encounters of 'road violence' at checkpoints (Schon, 2016).

Generally speaking, theory therefore suggests that a higher intensity of conflict events can lead to an increased perceived threat to the life of individuals and hence implies an increase in forced displacement. From an empirical perspective, this relationship is well-established on the aggregate level. Numerous studies base their analysis on country-year data and document that higher levels of conflict intensity are at least correlated with an increased number of refugees and internally displaced persons (Schmeidl, 1997; Schmeidl, 2001; Davenport et al., 2003; Hatton, 2009, Conte and Migali, 2019; Abel et al., 2019). Conflict measures have also been identified as the most important predictor variables in the modeling framework used by the Danish Refugee Council to predict stocks of displaced persons for their 'Global Displacement Forecasts' reports (DRC, 2022).

However, the usual aggregate perspective characterizing most existing studies tells only part of the story. Focusing on the country-level masks that a given conflict situation is unlikely to affect the full geographic scope of a country similarly. The impact of conflict may well be restricted to certain parts of a country, having close to zero impacts in other regions. Melander and Öberg (2007) argue that the perceived threat of potential forced migrants is mostly affected by conflict that happens in close geographical proximity. Engel and Ibáñez (2007) further stress the localized nature of conflict impacts, mentioning that the perceived level of threat of an individual may well be influenced by violent events affecting neighbors or members of the local community. Apart from the aggregation in the geographical dimension, aggregation to annual or monthly flows of migrants is likely to mask some of the often rapidly fluctuating population movements that characterize the displacement process due to conflict events. Liu et al. (1979) report in a study on Vietnamese refugees that more than 80% of surveyed displaced persons fled between two hours and two days after experiencing violent events. Similarly, Schon (2015) mentions that, once the perceived threat threshold has been surpassed, displacement is expected to occur within a few days of the conflict event taking place.

Overall, analyzing the nexus of conflict and displacement on a granular spatio-temporal scale remains a serious empirical challenge. As a result of this, the few existing empirical contributions that analyze disaggregated time series or panel data are not always conclusive or in line with theoretical predictions. For instance, Yuen et al. (2022) do not detect a significant effect of conflict on displacement using monthly data from Somalia. On the contrary, Schon (2015) employs a Bayesian change point model, finding that changes in the conflict's geographical scope and balance of power are more relevant drivers of displacement than daily conflict intensity measures. Ball and Asher (2002) find significant effects of conflict on forced displacement in Kosovo, in the context of a war crime trial in The Hague. Schon (2016) focuses on daily displacement data collected in Somalia and analyses the impact of violence at home and violence on the road during fleeing, finding various significant relationships between population movements and the occurrence as well as the uncertainty surrounding events of violence. Tai et al. (2022) circumvent the use of survey or administrative data and leverage daily data on human mobility derived from cell phone networks for analysis. They evaluate the short-run impact of conflict on displacement in Afghanistan and find significant effects of conflict events on the probability of leaving the district where the event occurred.

Several empirical studies have focused on static, household survey data sets instead of panel or time series data. Czaika and Kis-Katos (2009) use survey data to identify the determinants of displacement patterns in Aceh, Indonesia, and find that conflict is one of the major drivers of population movements on the village level. Engel and Ibáñez (2007) design a microeconomic model and analyze household-level survey data set to explore various aspects of the decision-making process potentially leading to displacement during or after conflict. Adhikari (2013) uses survey data from Nepal to explore individual decision-making processes in the context of conflict and displacement and reports that violence is the strongest predictor of displacement during conflict.

Finally, there is a rich literature on agent-based modeling in the context of forced displacement, see for instance Sokolowski et al. (2014), Suleimenova et al. (2017) and Searle and van Vuuren (2021). In these computational frameworks, various assumptions, e.g., how displaced individuals choose their destination and how fast and far individuals can travel per day are combined with conflict data and terrain data to predict the movements of displaced populations. Importantly, one key ingredient in these models is an a priori estimate of how much displacement a given conflict day or event will cause. One aim of this article is to inform these complementary approaches by providing impact estimates, both in terms of average and heterogeneous effects.

A2 Study Context: Conflict in Somalia

Somalia gained independence from the United Kingdom and Italy in 1960. It is one of the most ethnically homogeneous countries in sub-Saharan Africa, but the population is divided into more than 500 clans and sub-clans, though four major clans dominate the population (Menkhaus, 2004). Partially armed resistance against Siad Barre's post-independence regime during the 1980s finally led to a collapse of the military junta in 1991, sparking civil war between local power factions. It also dismantled the central state completely, with Somaliland self-declaring independence (Thalheimer and Webersik, 2020, Eklöw and Krampe, 2019). After the warfare from 1991-1992, UN and US engagement began after a ceasefire was negotiated and the UN Operation in Somalia (UNOSOM) mission was launched, dominating the political scene until the UN withdrawal in 1995 (Healy and Bradbury, 2010). Multiple clan-based groups, typically led by local warlords, merged into two major political coalitions by the late 1990s with differing approaches to state building: one backing strong central rule and containing the subgroups of the Hawiye clan, and the other an anti-Islamist group dominated by the Darood clan and backing federal rule. These clan-based groups were largely divided among regional lines, with the Hawiye mainly inhabiting central and southern Somalia and the Darood largely in the north, with the largest clans and sub-clans dominated by warlords. From 1995 to 2000, regional administrations emerged across the country, as security improved and economic development accelerated (Menkhaus, 2004).

Since 1991, Islamist movements seeking to establish an Islamic state have been gaining influence, leading to the formation of the Islamic Courts Union (ICU) in 2000, an umbrella organization of Islamic courts from various clans. A Transitional Federal Government (TFG) was formed in 2004 as the result of a negotiation process led by the Intergovernmental Authority on Development (IGAD) to build a functioning government representative of the various clans. In 2006, however, the ICU took Mogadishu in control before the TFG could negotiate its arrival with the alliance of warlords. When mediation efforts failed, Ethiopian forces ousted the ICU at the request of TFG, establishing the TFG in the capital and sparking the formation of Al-Shabaab, a more radical offshoot of ICU (Healy and Bradbury, 2010). With the approval of the UN Security Council, the African Union Mission in Somalia (AMISOM) was deployed to the capital in 2007 to protect the TFG, sparking further violence on the part of Al-Shabaab. Armed groups, Al-Shabaab being the most prominent among these, remain the major group fighting against the internationally recognized Federal Government of Somalia (FGS) which replaced the TFG in 2012. Today, Somalia continues to be fragile despite improvements in political stabilization. The population of Somalia faces recurring episodes of natural disasters, conflict, and mass internal displacement.

A3 Data Limitations & Measurement Issues

The UNHCR data covers displacement at a high spatial and temporal resolution over a long period, providing a unique opportunity to study the relationship between conflict and displacement. However, it is characterized by several issues that are likely to plague any disaggregated analysis of displacement patterns in conflict zones. First, it is unlikely that the monitoring network is able to capture all relevant displaced individuals at all times. Hence, the data needs to be seen as merely indicative of larger movements of displaced populations (UNHCR, 2017). Undercounting can for instance occur due to people avoiding common travel routes (Schon, 2016), due to parts of the monitoring network not being "active" at certain times, or due to chaotic and large population movements that 'overload' the monitoring network (Ball and Asher, 2002). Second, the coverage and quality of the obtained data are likely to vary due to changes in external partners, varying access to a sufficient number of well-trained local field staff, or the local security situation (UNHCR, 2017). However, as the data we use captures only displaced persons who have contact with the UNHCR, increasing displacement represents an increased strain on humanitarian organizations and hosting communities and therefore an event of interest. Limitations of the ACLED database include potential media fatigue during conflict (Eck, 2012) and that data quality may not be evenly distributed within countries, for instance, because certain regions have better coverage than others, which is, for instance, to be expected in urban versus rural settings. Both channels may lead to underreporting of conflict events.

In our analysis, we measure conflict intensity as the number of conflict events per week, as opposed to alternative measures such as the number of conflict-related fatalities. We give two reasons in favor of this measure of conflict, one being related to data reliability and one being related to the expected relationship between the perceived threat to life and conflict intensity. In terms of data reliability, it is well known that fatality counts are based on estimates that are notoriously hard to obtain (Spagat et al., 2009). In addition, the credibility of sources regarding fatality counts is hard to establish in events of violent conflict. In ACLED, fatality numbers are hence often based on somewhat arbitrary imputation techniques,[6] potentially leading to bias in later stages of analysis. Compared to that, the occurrence of a violent event can be measured with a much higher level of certainty and consistency. In terms of the relationship between conflict and displacement, literature has shown that the quantity of violent events matters more than their intensity. For instance, Melander and Öberg (2007) document that the proximity of conflict to the own livelihood is more important than the actual intensity of the event. In addition, fatality counts do not account for non-deadly acts of violence such as rapes or the destruction of property, which are potentially important triggers of

[6] For instance based on decision-tree type imputation using questions like, 'Is this an attack outside of a warzone? If yes, fatalities are estimated at 3.'. More information can be found in the ACLED Fatality Methodology Notes in Version 1 from February 2020, available from the ACLED website https://acleddata.com/.

displacement. As a result, we focus on event *counts* as the independent variable of interest. Compared to alternative conflict event databases, ACLED does not apply fatality thresholds when recording events, which effectively avoids missing events with 'too little' intensity and allows for an analysis that is more in line with theoretical literature.

A4 Discussion of Internal and External Validity of the Results

In terms of internal validity, a limitation of the presented analysis is that we do not include additional controls for other factors triggering displacement, such as droughts, floods, and economic hardship. This is because these variables are unavailable on a week-district level in Somalia. Systematic positive co-movements between conflict and these confounders would potentially bias our estimates upwards if unaccounted for in the model. However, both drought periods and economic hardships are likely to affect several districts jointly and are relatively slow-moving variables. Hence, we expect that a large part of the variation is captured in the common factor structure and the district-specific trend components. Similarly, flash floods usually affect more than one district at a time and are therefore likely captured by the latent factor structure.

In terms of external validity, it is further important to keep in mind that the effect of conflict on displacement may well be different in countries other than Somalia. One potential reason is that general living standards are likely a determinant of the individual decision to flee or not. In Somalia, living standards are often poor, potentially leading to less attachment to the place of residence. In countries with higher living standards, populations may be more resilient and less likely to flee during conflict events. Along those lines, Crippa et al. (2022) provide empirical evidence that conflict is a relevant driver of migration, but mostly in low-income economies. Moreover, the estimates we obtain are based on a sample period of long-term and ongoing civil war, furthermore undermining living conditions and exposing the population to violence-related trauma. These factors may also impact the resilience of the population. In addition, the life of a large share of the population of Somalia includes nomadic and pastoral elements irrespective of the conflict situation (Cassanelli, 2016), which may imply a generally higher likelihood of migrating after shocks, including conflict and conflict-related events.