Emily R. Barker, Jakub Bijak

Could we have seen it coming? Towards an early warning system for asylum applications in the EU

Deliverable 9.3

QuantMig has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870299.

History of changes

Version	Date	Changes
1.0	28 July 2022	Issued for Consortium Review
1.1	25 October 2022	First version submitted as official deliverable to the EC

Suggested citation

Barker E. R. and Bijak J (2022) Could we have seen it coming? Towards an early warning system for asylum applications in the EU. QuantMig Project Deliverable D9.3. Southampton: University of Southampton.

Dissemination level

PU Public

Key words

Asylum, Big data, Early Warning Systems, International migration, Macroeconomic drivers, Scenarios, Syria, Ukraine

Acknowledgments

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870299 QuantMig: Quantifying Migration Scenarios for Better Policy. We are very grateful to Marta Bivand Erdal, Rainer Münz and Teddy Wilkin for their comments that helped improve an earlier draft, and to Peter WF Smith and Jason Hilton for methodological suggestions. All the remaining errors and inaccuracies are ours. This document reflects the authors' views, and the Research Executive Agency of the European Commission are not responsible for any use that may be made of the information it contains.

Cover photo: iStockphoto.com/Guenter Guni

Technical Report:
D9.3: Could we have seen it coming? Towards an early warning system for asylum applications in the EU

Emily R. Barker[*] Jakub Bijak[†]

25 October 2022

Abstract

Forecasting large changes in the number of asylum applications, which are seen as one constituting element of so-called asylum 'crises', is incredibly challenging. Demographers, migration scholars, analysts, and policy makers have tried to build and apply early warning systems for that purpose at least since the large inflow of asylum seekers into Europe in 2015–16, which at the time was relatively unforeseen. Recently, another humanitarian crisis occurred after the Russian invasion of Ukraine in February 2022. In this case, temporary protection offered to Ukrainian nationals in all EU member states superseded asylum applications, and the number of applications would not have necessarily raised alarms for early warning systems based on approaches using a single dependent variable. In this report, we present a model that shows that the warning signs of a crisis could appear in some publicly-available data sources, including some 'big data' collections. Our aim is to propose and test an early warning system for asylum applications in the EU that would be easy to use, effective and interoperable for policy makers, and that would give sufficient advance warning that authorities can be prepared for an increase in the number of asylum applications or asylum case load. We look to see if a model can give a warning signal up to six months in advance for two of the most prominent asylum flows from the recent decade, involving people fleeing the wars in Syria and in Ukraine. In addition, we provide an empirical reflection on the definition of a 'crisis', and offer some practical recommendations for further work in the area of early warning modelling for migration-related applications.

Keywords: Asylum applications, Big data, Early Warning Systems, International migration, Macroeconomic drivers, Scenarios, Syria, Ukraine
JEL Classification: C52, C53, E27, E32, F22, F42, J11

---

[*]University of Southampton, UK. Email: E.R.Barker@soton.ac.uk
[†]University of Southampton, UK. Email: J.Bijak@soton.ac.uk

Contents

[1]We are very grateful to Marta Bivand Erdal, Teddy Wilkin and Rainer Münz for their suggestions for interpretations of some of the results, related caveats, and possible directions of future work. Needless to say, all the responsibility for the content of this section remains ours.

List of Figures

List of Tables

1 Introduction

There is large uncertainty around international migration flows in general; particularly for the numbers of asylum seekers related to high-impact events, including several experienced in the last decade. These 'crises', as they are often perceived, involve large numbers of asylum seekers fleeing their home countries due to politically-motivated factors, such as war, state violence, or persecution. The magnitude and impact of these flows vary in size and intensity; they can be related to single countries of origin, such as Afghanistan, Bosnia and Herzegovina, Kosovo or Ukraine, or whole regions, e.g. many Middle Eastern and North African countries since the Arab Spring. In the latter group, while individual countries were affected by different political events, including military interventions (Libya), it was the war in Syria that has become almost synonymous with asylum[2].

Recently, there has been a considerable discussion whether the 'crisis' as experienced by host countries is related to migration *per se* or rather to its governance (Crawley, 2016; Carastathis et al., 2018; Bijak and Czaika, 2020; Sahin-Mencutek et al., 2022). The surge in asylum applications in 2015–16 resulted in many practical challenges, which could have arisen from both the number of people seeking protection, and from how the governments dealt with the sudden increases in the inflows. In this report, we use the term 'crisis' to refer to an unforeseen arrival of large numbers of asylum seekers in the European Union (EU) and its neighbourhood. For practical reasons, related to the availability of data, we use the number of asylum applications as an imperfect proxy measure – and also as a variable of interest in its own right, related to the management of migration processes.

Throughout this report, we also use the term 'crisis' in a generic sense, as possibly related both to the asylum processes and their governance aspects, while realising that the governance response can be pivotal in either addressing or exacerbating challenges related to sudden increase in the numbers of migrants, including asylum seekers. We discuss in detail the issues related to operationalising the notion of a 'crisis' in the context of early warning modelling, so that it reflects extraordinary events, the onset of which should ideally be detected by a model and flagged to decision-makers for action.

[2]Notably, North African countries – such as Libya or Morocco – remain indirect contributors to the numbers of asylum seekers, acting mainly as transit countries for people from the Horn of Africa (such as Somalia and Eritrea) and Western Africa trying to reach Europe. (With thanks to Rainer Münz for drawing our attention to this important distinction.)

The early warnings systems for asylum applications are thus intended to alert policy makers – in our case, in the EU – about future increases in the numbers of asylum seekers, which are not typically foreseeable with traditional methods of analysis. The main aim of warning decision makers early is to give them time for a support network to be put in place to cope with an increase in the number of asylum seekers (and asylum applications), and if required to be prepared for a change in policy. It is worth recalling here that an asylum *application* constitutes an early stage of the process of seeking asylum: as per definition, *asylum seekers* are people who have left their home country for reasons such as violence or human rights violations, and are *in the process* of applying for asylum in a safe third country. Even though asylum-related migration cannot be seen purely as *forced* migration (Erdal and Oeppen, 2018), with a mix of economic and other drivers being in play, political factors, including war, violence and persecution, still have a dominant role.

By the end of the process, those applicants successful in their asylum applications become *refugees* and receive full protection from the host country, having been granted refugee status. Clearly, not every asylum seeker becomes an refugee, as applications can be denied, but the number of asylum applications is likely to be closely related to the caseload of asylum authorities, and to have impact on the numbers of the decisions, appeals and ultimately the number of people granted refugee status. The concept of a refugee is strongly related to offering high level of international protection, however, there are other types of protection that may be offered, on a temporary or semi-permanent basis. This can range from humanitarian assistance to provide food and basic amenities, to some forms of subsidiary protection for those people who fail to qualify for formal refugee status, yet can still count on some level of protection in the host society[3].

To meet the challenges related to uncertain and often high-impact migration flows, effective planning and preparedness in the area of migration policy is essential. This can be especially seen in the light of the political battles fought at the EU level over how to best manage the 'migration crisis' of 2015–16 (Hampshire, 2015), or the instrumentalisation of migrants seeking asylum for political aims, such as in the 2021–22 Belarus-EU border crisis. Other political challenges are related to the application of the EU Dublin Reg-

[3]For formal definitions of terms related to asylum and international protection, see e.g. the UNHCR Master Glossary of Terms, https://www.unhcr.org/glossary/ (as of 20 September 2022).

ulation, which was supposed to determine with which country was responsible for each asylum seeker. As a result of the events of 2015–16, the Dublin system was recognised as being in need of being updated or replaced with a more suitable one[4].

In practice, in the light of high uncertainty, before even considering the questions of migration governance, it is understandable that migration 'crises' seem ill-prepared for. Researchers and policy makers alike have, however, learned some lessons from the 2015–16 asylum crisis driven by influxes from Afghanistan, Iraq and notably Syria, in particular with research shifting to the early signal detection (e.g. Napierała et al., 2022; Carammia et al., 2022), and with policy responses moving towards preparedness. The latter involves the EU Blueprint on preparedness and crisis management mechanism in the area of migration[5]. The key principles of the Blueprint include anticipation, coordination, timely and flexible reaction, as well as responsibility sharing. Importantly from the point of view of this work, the anticipation principle is intended to be realised through monitoring and preparedness, two key objectives of which include situational awareness and early warning/forecasting, as detailed in the Annex to the Blueprint (*idem*).

The insights from an early warning system (hereafter, EWS) can hopefully allow policy makers to improve the management of asylum-related migration at different decision levels by providing a more timely and potentially better targeted operational response at subsequent stages of the process. Ideally, the models used should also be able to describe, or perhaps even reduce, some of the associated uncertainty. Migration policies can be slow to change, but as has been seen during the 2015–16 crisis, and even more so in the (rapid) policy shift following the 2022 invasion of Ukraine, granting Ukrainian citizens temporary protection in the EU, swift action can be put in place when needed.

A policy for effective managed migration also requires an understanding of the drivers of migration (de Haas, 2018; Czaika and Reinprecht, 2020), which applies to all types of migration, including secondary movements. Still, economic migration, particularly intra-European and within the Schengen area, is *relatively* simpler – if still not simple – to model (see, Barker and Bijak, 2021) through a series of underlying drivers (push and

[4]Updates to the Dublin Regulation were agreed in 2020, see: European Commission, DG HOME.

[5]Commission Recommendation (EU) 2020/1366 of 23 September 2020 on an EU mechanism for preparedness and management of crises related to migration, OJ L 317, 1.10.2020, p. 26–38, `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2020:317:TOC`.

pull factors, the data for many of which come from official statistics and open sources), as well as fitting in with a number of existing theories outlined e.g. in Massey et al. (1993). Family migration is also better predictable, as it is largely driven by marriage migration, related to the migrant stocks from origin countries and also to some extent to the composition of migrant families already present at the destination. As such, family migration is less likely to be susceptible to shocks, typically not being linked to sudden policy changes or previous high-impact events. At the other end of the spectrum, environment-related migration is of a small scale at an international level now, but likely to have higher impact in the future due to increasing effects of climate change. These flows can be short-distance and lead to returning to the original places of residence once possible, which is typically not captured by data (see, Vestby et al., 2022).

The research presented in this report builds on the existing work by Napierała et al. (2022) and Carammia et al. (2022) who provided examples of early warning systems and models for migration, with the former looking at detection of change-points in asylum series, and the latter focused on forecasting of migration flows with 'big data' with a four-week horizon. In addition, Melachrinos et al. (2020) also employed 'big data' to identify migration push factors from Africa. This report aims to contribute to the existing literature by exploring longer horizons, for which early warnings aimed at the implementation of any necessary preparations may be realistic, and also by including insights from early warning and signal detection models existing in macroeconomic practice, such as central banking. The models we present are designed to trigger alerts up to six months prior to the surge in asylum applications, which should enhance the decision making and preparedness for dealing with the surge in inflows. The applications of our models are illustrated for two case studies: the Syrian 'asylum crisis' of 2015–16 and the recent (and ongoing) migration implications of the war and humanitarian crisis in Ukraine.

The remainder of this report is organised as follows: Section 2 discusses the existing literature and context of early warnings of unforeseen events; Section 3 introduces the methodology behind our early warning model and the data employed, as well as operationalising the definition of a 'crisis'; Section 4 presents the applications of the model for the Syrian and Ukraine case studies; and Section 5 discusses the practical and policy implication of the results, and offers a conclusion, with recommendations for future work.

2 Selected Existing Work on Early Warning Models

The literature on early warning systems in the context of arrival of asylum seekers remains relatively limited. The research interest in this area was piqued following the rise of Syrian migration to Europe of 2015–16, since a need for such a system proved to be potentially useful for policy makers. Yet, there has not been a large development or optimisation of models which could be tested; instead, the formal work in this area remains largely in the prototype phase. Still, there are examples of early warning models in other strands of literature, such as macroeconomics, which can be useful for modelling migration as well. We discuss these approaches in turn.

Selected Early Warning Models in Macroeconomics

Before discussing the application of early warning models to asylum applications, we look at similar models in finance and macroeconomics, which mainly focus on predicting recessions. Such models, with a strong basis in macroeconomic theory, came to prominence in light of the Great Financial Crisis of 2008–09. In economic applications, various input variables may include interest rate spreads along with other macroeconomic indicators, which can jointly determine the probability of a crisis occurring. If this probability exceeds a certain threshold, a warning signal is triggered. We describe this approach in more details in Section 3.1, with special attention paid to defining what can be classified as a 'crisis' in the context of migration, as discussed further in Section 3.3.

A significant number of early warning models in macroeconomics use (multinomial) logit or similar binary-response models. Such models have been operationalised in a number of software packages, one example of which being the EWS [Early Warning System] software package in R[6]. The EWS package was originally designed with macroeconomic applications in mind, and employs a binary indicator for whether a recession (economic crisis event) is occurring or not. One example of an application that looks at the emerging economies is the study by Bussiere and Fratzscher (2006), focusing on high volatility of such economic systems, particularly those in Latin America. For developed economies, the key focus of applied early warning studies is the US economy.

Central banks, and international organisations such as the IMF had used early warning

[6] See `https://CRAN.R-project.org/package=EWS`.

models, particularly to consider economic crises, yet these have (largely) failed to predict or have any impact on the financial crisis of 2008–09 and the following recession. The reasons for that are not directly relevant for this study, but in a nutshell, the origins of the crisis were either financial and thus largely excluded from some of the models, or the models failed to give adequate warning, being used only as diagnostic tools (see for example, Edison, 2003). Notably though, the crises within economics are easier to quantify according to commonly-adopted definitions, such as in the case of a recession, which is typically defined as two consecutive quarters of negative growth.

One common difficulty found across different early warning studies is that the inclusion of post-crisis data or observations makes the performance of the model worse (Filippopoulou et al., 2020). In terms of the predictive horizons, Filippopoulou et al. (2020) found in their study of 11 Eurozone countries that most of the European Central Bank's macroprudential risk indicators are important in forecasting banking crises with a lead time of one to four years ahead. We use some macroeconomic variables in our study as well, albeit with much shorter warning horizons of a maximum of six months ahead. Besides, whilst the number asylum applications are not related to the same types of macroprudential indicators as in macroeconomic models, we still include exchange rates, which strongly feature in non-US-centric early warning models.

Existing Early Warning Models for Migration and Asylum

Unlike in macroeconomics, the underpinnings of asylum and other types of migration shaped predominantly by political drivers are often not addressed to the sufficient extent in the theoretical literature of migration, which can constitute a challenge for modelling. The seminal paper on migration theories, by Massey et al. (1993), mentions refugees and asylum seekers mainly in the context of the world systems theory and global impacts of military interventions. Despite the humanitarian importance and political salience of refugee and asylum flows, refugee studies often remain disconnected from the broader field of migration research (FitzGerald, 2015), itself often mired in disciplinary silos.

More recently, many other studies have looked at various *drivers* of migration, such as Raleigh (2011); Van Hear et al. (2018); Cummings et al. (2015), or De Haas (2021), with focus on different aspects, such as aspirations and motivations, reasoning and decision making, or push and pull factors for different types of migration. Still, given the prevailing high levels of uncertainty about such drivers, modelling, 'nowcasting' or forecasting the numbers of asylum seekers remains particularly challenging (Bijak et al., 2019). Further reviews of the literature on migration and its modelling are offered in Bijak (2010) and Sohst et al. (2020), with the main conclusion that operationalisation of drivers for forecasting purposes remains a challenge – and that such drivers would need to be predicted as well to be useful in forward-looking migration models. For this reason, the work presented in the current report is largely based on contemporaneous variables that can provide some signal for early warning models about upcoming increase in the number of asylum applications, rather than being based on theoretically-informed migration or asylum drivers or complex models of migration mechanisms.

As mentioned in Section 1, this work draws inspiration from Napierała et al. (2022) and Carammia et al. (2022). The papers differ in their approach: the former uses a cumulative sum (*Cusum*) approach to presaging asylum 'crisis' of 2015–16 based on the statistical control theory, whereas the latter offers an early warning system for asylum-related migration, largely focusing on forecasting with a four-week lead-in time, using 'big data' from a range of sources for 2016 onward. There are also other pieces of recent work looking into similar topics (e.g. Melachrinos et al., 2020). Initial research on the 2022 Ukraine crisis, presented in Juric (2022), used Google Trends search data focusing on migration planning terms[7], while Avramescu and Wiśniowski (2021) used data on Google searches to now-cast Romanian migration to the United Kingdom. Milliff and Christia (2021) used mobile phone data to analyse micro-level drivers of displacement in Yemen.

Using 'big data' sources such as Google Trends on their own when modelling patterns of people moving to other countries during the war is contentious also because of different levels of Internet access. In 2020, an estimated 75% of the Ukrainian population had access to the Internet, which was already a steep increase from 18% in 2009. For comparison, in Syria – with Arabic as a single dominant language – in 2009, a similar share (17%) of the population had Internet access, increasing to just 36% in 2020[8] Clearly,

[7]The author noted one important limitation related to searches made by Ukrainians both in the Russian language as well as in the Ukrainian language. Using data from the 2001 census, Young (2015) showed that the eastern Ukrainian regions of Donetsk and Luhansk had 75% and 69% of people respectively, with Russian as their primary language.

[8]Source: World Bank Data Portal.

people who are at the risk of persecution often use virtual private networks (VPNs) to obscure their identity and location, as was common in Syria (Eissa and Cho, 2013). The use of VPNs is becoming more accessible, making the numbers of some searches underestimated, however, for what does remain, it can provide at least some indication as to what might be happening in a country. The use of Google Trends can be country-specific too, as some countries host people of multiple nationalities seeking asylum, such as in the case of Turkey, which in some instances would make a more targeted analysis difficult.

In other examples of using 'big data' to monitor flows, recently, Leasure et al. (2022) proposed using data on social media activity to monitor ('nowcast') the number of displaced persons – in this case, within Ukraine – at a very high temporal granularity (daily). Still, despite the remarkable advancements in modelling, particularly the Ukraine crisis, due to its recency, following the full-scale Russian invasion in February 2022, is one for which the early warning models are yet to be tested in scholarly reviewed models.

Reflection on the Choice of the Case Studies

The two case studies for the current report have been selected due to their recency, high impact, and heterogeneity. The Ukraine crisis differs from the Syrian asylum crisis of 2015–16 in a number of important ways. First, the corridors of asylum seekers are different: the Mediterranean was the primary route for the asylum seekers from the Middle East, while Ukraine directly borders (overland) four European Union countries (Poland, Slovakia, Hungary, and Romania) and another EU candidate country, Moldova[9]. Second, the magnitude and time scales (and thus also intensity) of flows also vary for both case study countries: in 2015–16, 2.39 million asylum applications were lodged by Syrians across the whole EU27[10], compared to 5.26 million Ukrainians arriving in the EU only in the first two months of the 2022 war based on regular border crossing data[11].

Third, due to lack of 'regular' options, Syrian migrants into Europe have been faced with navigating dangerous irregular channels of migration, involving often risky maritime crossings (even if the Eastern Mediterranean route is not as perilous as the Central and

[9] Moldova formally applied to join the European Union on 3 March 2022, along with Georgia, following Ukraine's application. Source: Tanas (2022) and Gehrke (2022). Ukraine and Moldova were already granted candidate status, while Georgia has been granted a potential candidate status.

[10] Source Eurostat MIGR_ASYAPPCTZA. EU27 definition as of 2020. Data accessed on 15 July 2022

[11] Source: UNHCR (2022) Ukraine Situation Report (as of 22 July 2022).

Western routs are). Conversely, important legal channels for Ukrainian migration were open already before the outbreak of the war, and the temporary protection status route was opened soon after the Russian invasion. This was partially due to different levels of political will: as noted e.g. by Drazanova (2022), back in 2016–17 the Central and Eastern European EU members strongly opposed the decision of the EU Council of Ministers of Justice and Home Affairs (JHA) to distribute the asylum applicants across the EU, with some even proposing to close the borders altogether. Besides, already before the Russian invasion, Ukrainians could enter the EU / Schengen area without a visa for 90 days, whereas similar option was not – and is not – available for Syrians, and citizens of a vast majority of other Middle Eastern and African countries.

Since March 2017, Ukrainians with a biometric passport had been granted visa-free access to the EU[12] for up to 90 days in any 180-day period. Since the conflict in Ukraine intensified in 2014, a significant number of visas have been issued to Ukrainians, especially by Poland, providing other legal channels for Ukrainians to leave the country. However, these legal channels were not sufficient to cope with the flows of Ukrainians arriving following the recent invasion, which led to the emergency application of the temporary protection status granted to Ukrainian citizens across the EU in 2022. The status will last for up to three years (one year in the first instance, subject to renewal), by which point it is hoped that those who have taken up this offer will have either returned to Ukraine or found means to be granted a regular visa or residence status in the host country.

As a result, the differences between the legal situation of the two groups – Syrians and Ukrainians – are pronounced. Ukrainians have been collectively granted temporary protection status rather than having to undergo the process of individually applying for asylum, as Syrians do. The temporary protection process was easier to manage in the short term, even though in the long term, successful asylum applicants would have more rights and could secure more stable outcomes, such as the formal refugee status, permanent residence, and ultimately a possibility of naturalisation[13]. In this report, we will look into both case studies to see how, and to what extent, early warning methods could have helped predict the migration events with any lead time.

[12]Source: European Commission (as of 15 September 2022).

[13]Many thanks to Rainer Münz for drawing our attention to this aspect.

3 Methodology and Data

In this section, we explain the model employed for the early warning analysis and summarise the data sources for the input variables and other observables of interest, which were used in preliminary investigations. The case study-specific analysis is detailed in Section 4. We also discuss problems with defining defining a 'crisis' for use in early warning models, and offer practical recommendations for further research in this area.

3.1 Methodology and Model Design

Let us first consider an early warning model with a binary response variable indicating the presence or absence of a crisis. Our work is similar to Carammia et al. (2022) as we would be also using 'big data', though with an addition of macroeconomic variables.

For each period in the observation window, the binary variable takes a value of 0 to indicate no crisis or 1 for a crisis. Defining a crisis in the context of any form of migration, especially asylum-related migration due to its high volatility, is a very complex task, which we discuss in more detail in Section 3.3. For each period, the model estimates a probability, $\widehat{Pr}$, that a crisis will occur, which will trigger an early warning if this probability is greater than some threshold value, c. There are two types of error when a crisis is misidentified: a false positive (type I error) occurs when the probability of a crisis is greater than the threshold level (signal), but a crisis does not occur. The occurrence of a false negative (type II error), happens when the probability does not meet the threshold level (no signal) but a crisis occurs, as summarised in Table 1.

Table 1: Matrix for models signals and crises

		Crisis	**No crisis**
Signal	$\widehat{Pr} > c$	True Positive	False Positive *Type I Error*
No Signal	$\widehat{Pr} \leq c$	False Negative *Type II Error*	True Negative

Adapted from Lang et al. (2018) and Lajaunie (2021)

Kauppi and Saikkonen (2008) used a dynamic binary probit approach as a general framework for four dichotomous models to evaluate historical data and to forecast potential crises. In its general form, such dichotomous model can be written as:

$$P_{t-1}(y_t = 1) = F(\pi_t) = F(\beta' X_{t-1} + \Sigma_{j=1}^{p} \gamma_j y_{t-j} + \Sigma_{j=1}^{p} \eta_j \pi_{t-j}), \tag{3.1}$$

where y_t is the dichotomous variable at time t, π_t is the corresponding latent variable, X_{t-1} is a kx1 vector of k exogenous variables, and β is a kx1 vector that responds to the coefficients associated with X_{t-1} (Lajaunie, 2021). In equation (3.1), γ_j and η_j are the vectors of coefficients for lagged values of y_{t-i} and π_{t-j} respectively, with p and q being the number of lags for the dichotomous and index variables. Kauppi and Saikkonen (2008) distinguished four specifications of the model (3.1): (1) a static model with $\gamma = \eta = 0$; (2) a dynamic model with $\eta = 0$; (3) a dynamic model with $\gamma = 0$, and (4) a dynamic model with lagged binary and index variables, with $\gamma, \eta \neq 0$. We consider these models, summarised in Table 2, in our applications to the case studies presented in Section 4.

Table 2: EWS Model Equations

Model	**Equation**
1	$P_{t-1}(y_t^1 = 1) = F(\pi_t^{(1)}) = F(\beta' X_{t-1})$
2	$P_{t-1}(y_t^2 = 1) = F(\pi_t^{(2)}) = F(\beta' X_{t-1} + \Sigma_{j=1}^{p} \gamma_j y_{t-1}^2)$
3	$P_{t-1}(y_t^3 = 1) = F(\pi_t^{(3)}) = F(\beta' X_{t-1} + \Sigma_{j=1}^{p} \eta_j \pi_{t-1}^3)$
4	$P_{t-1}(y_t^4 = 1) = F(\pi_t^{(4)}) = F(\beta' X_{t-1} + \Sigma_{j=1}^{p} \gamma_j y_{t-1}^4 + \Sigma_{j=1}^{p} \eta_j \pi_{t-1}^4)$

The probability $\widehat{Pr}$ of a crisis occurring is subsequently calculated for each period. Whether or not a crisis is identified, is determined by whether the probability exceeds certain thresholds. In this methodology, we employ three threshold criteria, each of which optimises the cut-off periods in a different way. The values for all three thresholds range between 0 and 1, and are determined by the vector of probabilities calculated from the logistic estimation using equation (2). The accuracy measure (AM) criterion aggregates the number of periods in which a crisis occurs and does not occur to give an optimal value that maximises the number of correctly identified periods (Hasse and Lajaunie, 2021). The credit-scoring approach (CSA) criterion looks at the sensitivity and specificity of correctly-identified crises: *sensitivity* gives the proportion of correctly identified crisis periods, whereas *specificity* is the proportion of correctly identified calm (non-crisis) periods (Hasse and Lajaunie, 2021), with the threshold value minimising the absolute difference between the two. The Noise-to-Signal Ratio (NSR) criterion (proposed by Kaminsky et al. (1998)) represents the ratio of false alarms relative to correct alarms, with the threshold value minimising such relative errors. In general, the AM and CSA criteria are relatively similar and identify more crises than the NSR criterion, which has higher values and is thus more conservative. As such, AM and CSA are more likely to give Type 1 errors (false positives), while NSR gives more Type 2 errors (false negatives).

3.2 Main Data Sources for Modelling

The model (3.1) has two main non-latent empirical components: the binary variable y_t indicating a crisis and the vector of exogenous explanatory variables, X_t. As the definition of a crisis requires a separate reflection, first we present the data sources for the explanatory variables, before defining what may constitute a migration crisis in Section 3.3. Overall this study is based on a variety of data sources, listed in Table 3, some of which we discuss in more detail below.

Table 3: Data Sources

Variable	Source and URL
Asylum Applications	Eurostat - MIGR_ASYCTZM & MIGR_ASYAPPCTZM https://ec.europa.eu/eurostat/data/database
Irregular border crossings	Frontex https://frontex.europa.eu/we-know/migratory-map
Price of wheat	Humanitarian Data Exchange https://data.humdata.org
Fragile States Index	The Fund for Peace https://fragilestatesindex.org
Global event data	GDELT https://www.gdeltproject.org
Google searches	Google Trends https://trends.google.com/trends
Ukraine Inflation	State Statistics Service of Ukraine https://ukrstat.gov.ua
US Exports and Imports	FRED St Louis https://fred.stlouisfed.org
Exchange Rates	IMF - International Financial Statistics https://data.imf.org

Details on data and their sources used in this report and model estimation, accessed on 15 June 2022.

The main variable of interest, underpinning the binary crisis response, is the number of asylum applications. We use first-time applications, as they are more likely to capture newly-arrived applicants. The relevant data are available at a monthly frequency for 1999:01-2007:12 in the Eurostat database, table MIGR_ASYCTZM. The countries reporting to Eurostat for the period are the EU27 (as of 2007), Iceland and Norway. For the period 2008:01 to present, table MIGR_ASYAPPCTZM is used. The reporting countries include 'EU+' countries: EU28 (with UK until December 2020), Iceland, Norway, Switzerland, Lichtenstein and Montenegro. The aggregation of the asylum applications for all these countries together yields the overall values used in our analysis.

Identifying countries that could face political instability can be done by using a variety of sources. One such accessible variable, collected on an annual basis, is the Fragile States Index, established by The Fund for Peace, that as of 2021 evaluated 179 countries[14]. Produced since 2006, this index studies the stability, or fragility of countries on 12 cohesion-related, economic, political and social indicators. In Figure 1, we present the aggregate values of the index for five selected countries: Afghanistan, Syria, Russia, Ukraine, and for comparison of a large and stable EU country – Germany. The maximum (negative) score is 120, with war-torn countries regularly coming on top of the ranking, whilst traditional peaceful, socio-economically developed countries, such as Finland, scoring the lowest. The impact of the Syrian civil war can be seen demonstrated by the change from a similar score to Russia in 2006 to scoring more than Afghanistan from 2016 onward. Similarly, the earlier period of the conflict in Ukraine is illustrated by a peak in 2015 – note that the data 2022 were not yet available at the time of writing.

Tracking summary indicators, such as the Fragile State Index, can alert the user to any significant changes that may lead to an increase in asylum applications from the countries in question. For this research, we use the Fragile States Index to show that countries can be either extremely fragile or unstable enough that any significant political event or shock could result in an increase in asylum applications. Conversely, a relatively stable country, such as Germany, exhibits a downwards trend of this index, i.e. with increasing levels of stability, indicating that such countries need not necessarily be of interest for early warning purposes.

[14] Available from: https://fragilestatesindex.org, accessed on 15 June 2022.

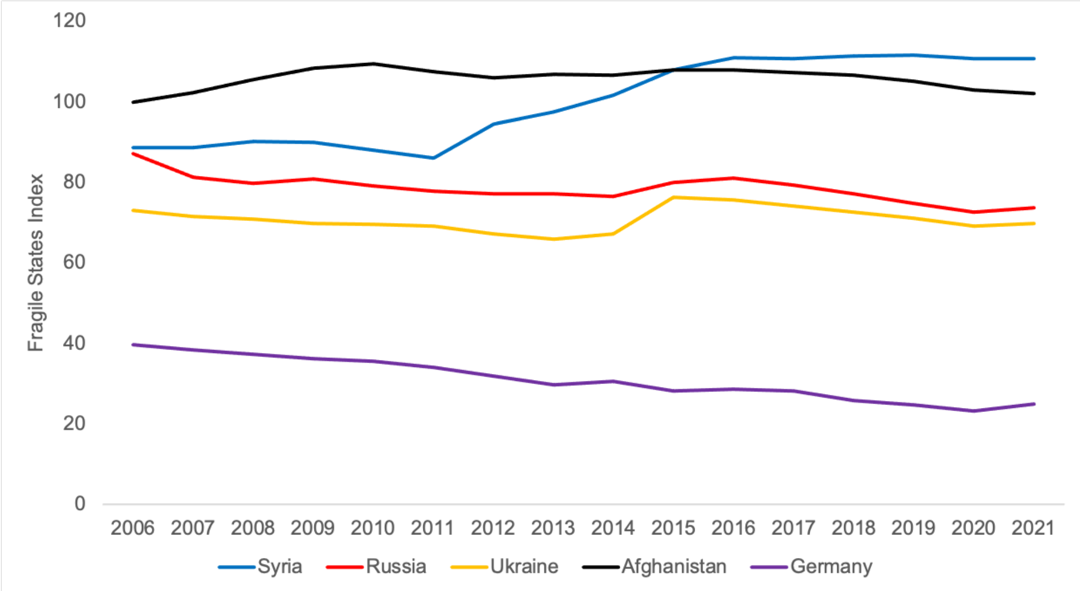


Figure 1: Fragile States Index (selected countries)

The total value of the fragile states index for selected countries 2006-2021. The Fragile States Index, is an annual index compiled by The Fund for Peace. The value is determined by twelve conflict risk indicators based on cohesion, economics, political and social indicators. Specific examples include: economic inequality, state legitimacy, and the number of refugees and internally displaced persons. The maximum score is 120, which indicates a highly fragile state. In 2021, the most fragile country, Yemen, scored 111.7, whilst the most stable country, Finland, scored 16.2 (*idem*).

Recently, attempts to forecast civil unrest have been also made, in particular with reference to the Arab Spring, using data from social media, and (protest) events using the GDELT [Global Database of Events, Language and Tone] data by Wu and Gerber (2018). To that end, Leetaru (2014) looked at the protest intensity using GDELT, defined as the percentage of all GDELT events that could be categorised as protest events. Using this analysis, we could see how the Arab Springs resulted in a surge of protests across the Arab world. Some of these protests were controlled by government forces with not much change, e.g. in Bahrain, whereas some leaders, such as Muammar Gaddafi of Libya or two Egyptian governments, lost power, while Yemen has since descended into a civil war and a large scale humanitarian crisis. Some governments made concessions though, leading to the different outcomes of these events. The use of GDELT in early warnings for the numbers of asylum seekers was pioneered by Carammia et al. (2022).

At a general level, GDELT provides an invaluable source of information about events that occur in each country, categorising them using the extended codebook using the CAMEO format (for CAMEO codes, see website). There are many available resources, particularly in GDELT 2.0, however for our purposes, we focus on the events database, which includes the numbers of events, average tone of reporting, the so-called *Goldstein Scale*, summarising the severity of events and the tone, and a number of other ways to analyse the events. We explore these measures in more detail in Section 4.

In addition, in our initial investigations, we also considered weather data, in particular precipitation (rainfall), as Selby et al. (2017) suggested that the drought prior to 2011 may have contributed to the causes of the Syrian civil war. As there have been reporting issues since conflicts began, we looked at the situation in the neighbouring Turkey, in particular, Gaziantep – a city close to the border with Syria. However, as this indicator did not prove to be of sufficiently high predictive potential, we have not included it in further analysis. Due to the complexity of droughts, further research could look at alternative data sources such as data on water reserves or data related to food dependence. However, these data sources are often not available from official sources, yet there are examples of attempts made to estimate the volume of water in reservoirs or lakes. For example, the Database for Hydrological Time Series of Inland Waters (DAHITI)[15] estimate the water level of the Addad, Reservoir (Tabqa, Syria) by using oceanographic techniques.

For Google Trends data, we looked at the relative frequencies of internet searches for migration-related terms, in the spirit of Böhme et al. (2020) and Avramescu and Wiśniowski (2021), adjusting them to more specific circumstances and using the sending country's language – Arabic and Ukrainian respectively. The trends for the same topic in English are, generally, similar. In June and July 2014, we were cautious of using such search terms as *Germany*, which would show up strongly in Google Trends because German national team won the FIFA World Cup in July. Isolated, this would provide a false signal. Noteworthy, Juric (2022) found Russian language searches in Ukraine to be more effective that Ukrainian language ones. This is not surprising, given that most of the areas in Eastern Ukraine that were directly attacked and invaded first (some of them already in 2014) are inhabited by a Russian-speaking population, as opposed for example to the predominantly Ukrainian-speaking areas of Western Ukraine, which were not invaded, and where (presumably) the need to leave was not as pressing.

Gathering timely data can be also challenging as some series are no longer reported to

[15] Available at TUM (as of 31 August 2022)

the international organisations for the countries which would be large senders of asylum seekers, such as Syria. In the analysis of the case studies presented in this report we only include the variables which were still accessible as of 2022. This can be of concern, and an obstacle to employing macroeconomic data. Gathering data on GDP is possible, but as these largely rely on government departments or ministries that could be influenced by political narratives, some of these datasets can be unreliable or incomplete.[16]

3.3 What is a Migration Crisis?

A common feature shared by many, if not all early warning models is that their response is typically a binary variable identifying a *migration crisis*. This, however, sidesteps the question, how a crisis should be defined. In macroeconomics, for example, a conventional definition of recession is two consecutive quarters of negative growth. For migration, in particular asylum-related migration, this is much less clear. In existing work, Napierała et al. (2022) used migration exceeding one standard deviation above the historical average (or trend) for the distribution of asylum applications lodged in the EU as one example of such a definition, which turned out to be a rather sensitive threshold, as well as two standard deviations, as an upper limit. Approach based on standard deviations is not always guaranteed to work, though. Our investigation focuses on the period 2008–2022, which contains a large variation in the number of asylum applications. For Syria, before 2013 migration was relatively low, and even after the 2015–16 'asylum crisis', the asylum applications in the EU have continued to exceed one standard deviation from the trend.

In terms of analysing the numbers of asylum applications, several important caveats need to be made. First, the number of asylum *applications*, as reported to Eurostat, does not need to be equal to the number of asylum *seekers* – the number of *first applications* is a better, if still imperfect, approximation. Second, in some countries, children born to asylum seeking parents, or new arrivals via family reunion routes, are registered as asylum

[16]The issue of data being manipulated to suit political narratives can be recently seen in the case of Russia. Since the invasion on Ukraine, heavy sanctions from Western countries were imposed, which have severely limited Russian ability to trade. A study by Sonnenfeld et al. (2022), explains how the official data put out in press releases are hiding the true state of the Russian economy, including GDP. What is of relevance to this research, the Russian government no longer publishes (macro) economic indicators that used to be available at a monthly frequency prior to the invasion. Notably, this includes trade data (commodities, and non-commodities), monetary base data, capital flows and other financial data. Outside of the economic scope, even airline and passenger data are no longer published by the aviation authority, *Rosaviatsiya*.

seekers, which limits the usefulness of such statistics for early warning modelling[17].

Besides, this type of definition only tells the user that asylum applications are high compared to the benchmark distribution. Given that their level might have reached a *new steady state*, the expected level of asylum applications, might not give useful information with respect to large changes in the numbers of first-time asylum applicants. The focus of this research is on significant increases in asylum applications, although clearly identifying a significant decrease could also aid policy makers in countries that experience volatile number of applications. The users could find such definitions of a crisis more useful, if they were to be accompanied by large percentage change in the numbers of applications.

We demonstrate the issue in Figures 2a and 2b, which show examples of permanent changes in an average or expected monthly asylum applications[18]. For both Syria and Ukraine, applying the threshold of one standard deviation (for the period 2008–2021 inclusive), for the absolute values would imply warnings issued before the respective 2015 peaks, and for Ukraine the average values would remain above the threshold post 2015. The less sensitive value of two standard deviations, as used by Napierała et al. (2022), would at the same time identify the peak periods of the respective asylum processes as 'crises': 2014–15 for Ukraine and 2015–16 for Syria. Figures 3a and 3b show the application of different thresholds for Syria and Ukraine respectively, with possible crisis warnings triggered at 0.5, 1, 1.5 and 2 standard deviations.

It is noteworthy that a lot of displacement can be internal: at the time of writing, the most recent estimates for Syria indicate 6.7 million internally displaced persons at the end of 2021, and for Ukraine nearly 7.0 million at the end of August 2022, the latter based on a survey carried out by the International Organization for Migration[19]. In particular, the displacement of a large number of Ukrainian citizens during 2014–15 occurred internally, such that the scale of asylum applications abroad is not necessarily proportional to the overall scale of displacement. There was a significant rise of work permit applications in Poland too, which is not reflected in the asylum application figures.[20] The Ukrainian

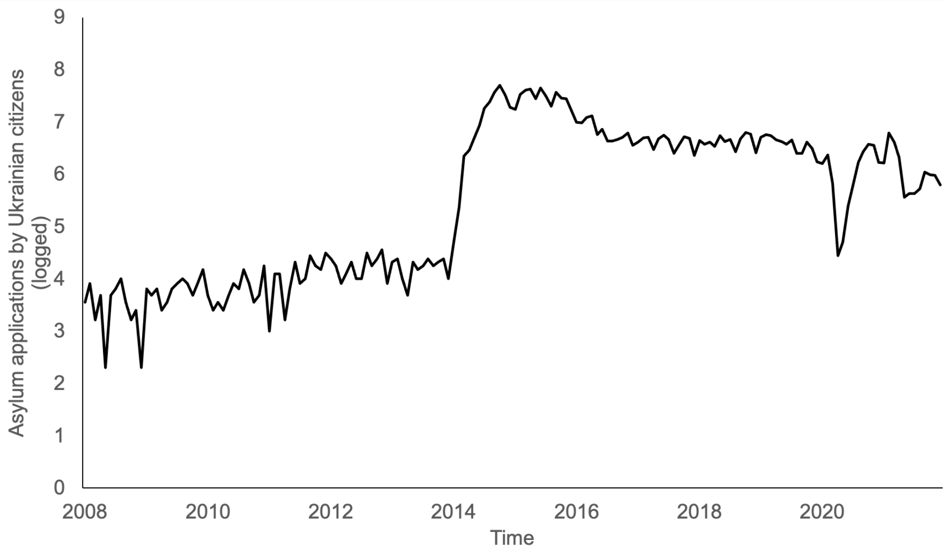
[17] Many thanks to Rainer Münz for drawing our attention to this aspect.

[18] We plot the log-transformed numbers in both figures due to the large change in 2015 and 2014, respectively, which (somewhat) mask the true scale of changes afterwards.

[19] Source: IDMC - Internal Displacement Monitoring Centre, Syria and Ukraine country profiles, accessible via `https://www.internal-displacement.org` (as of 25 September 2022)

[20] Over 2 millions Ukrainians settled in Poland between 2015-2019. Source: Politico.eu. The statistics for Europe wide are also available via Eurostat and European Commission (as of 31 August 2022).

(a) Syrian citizens



(b) Ukrainian citizens

Figure 2: Number of monthly asylum applications (logged)

Log-transformed numbers of monthly asylum applications made by Syrian and Ukrainian citizens in EU+ (EU, EFTA, UK, ME) 2008:01–2021:12. The fall in early 2020 is caused by the COVID-19 pandemic.

government reported that at the height of military operations in that period, 1.5 million people had been internally displaced (Jaroszewicz, 2019), which is more that one order of magnitude more than the asylum applications shown in Figure 3b. At the same time, including internal displacement in early warning models of displacement is challenging, with data required to be high-frequency and available at high spatial granularity within countries experiencing conflict, which may be unreliable. One potentially promising avenue of research involves social media data, which were used to estimate the extent of the displacement of Venezuelans by using Facebook's advertising platforms (Palotti et al., 2020) or Twitter-based sampling methods (Hausmann et al., 2018).

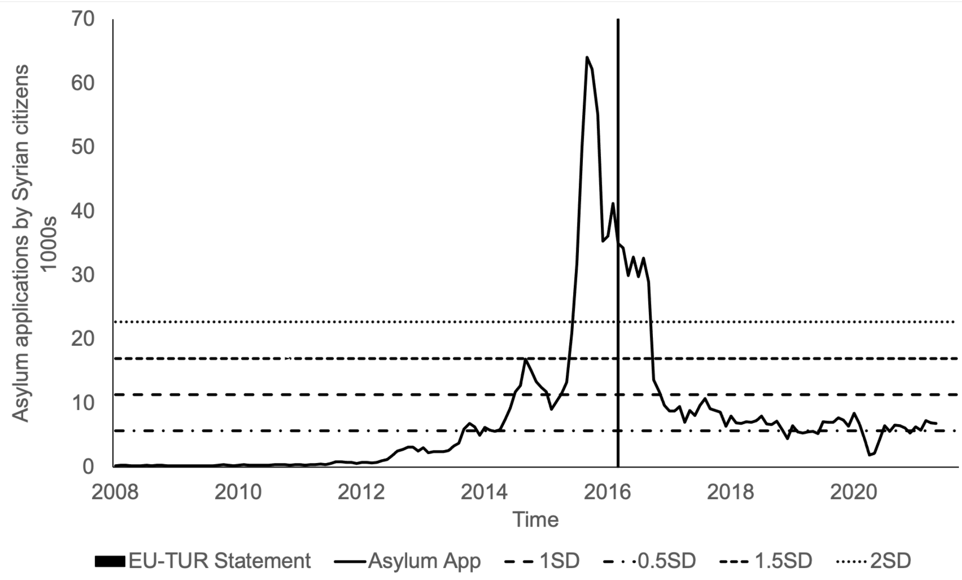
From Figure 3, we can also see that for Syria, from December 2016, the monthly number of asylum application falls below one standard deviation, yet the absolute values still exceeded 100,000 in 2017, 83,045 in 2018, 77,755 in 2019, 66,460 in 2020, and 99,810 in 2021. Are these years a crisis? In one sense, yes, since there is a continued high inflow of asylum seekers. In another sense, no, since the numbers are smaller number than the ones previously recorded, especially at the 2015–16 peak.

The Ukrainian case is similar, but not the same. Firstly, from March 2014, except for the first COVID-19 lockdown and a handful of other months post the original COVID-19 outbreak, asylum applications exceed the one standard deviation threshold. Secondly, since the Russian invasion of Ukraine in February 2022, the EU policy allowed Ukrainian citizens similar rights to EU citizens (initially for a year, with possible renewal for up to three years), by applying the Temporary Protection Directive based on citizenship and residence grounds, removing the need to individually claim asylum[21]. The temporary protection status is therefore easier to obtain, and comes with many rights (residence, work, choice of an EU country, access to services) but is also time-limited and potentially less stable than a refugee status, which is more difficult to secure.

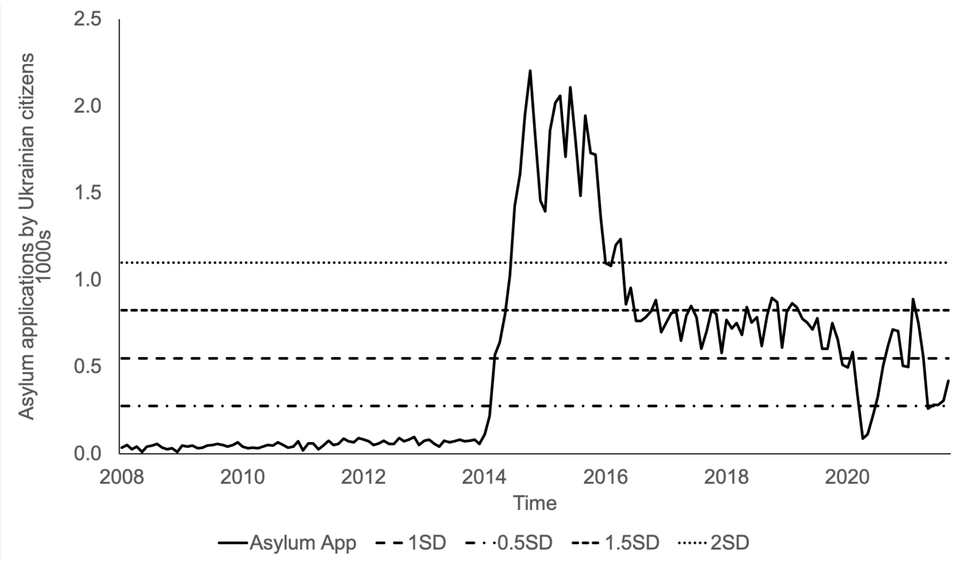
As a result, the numbers of Ukrainian nationals seeking protection in the EU will not show in the asylum application data. Relying on the same signal would hide the true scale of migration for humanitarian protection, and of the related challenges. For instance, Poland reported only 150 formal asylum applications by Ukrainian nationals in February 2022, which only relates to the period prior to the Russian invasion[22]. Still,

[21]The Temporary Protection Directive mechanism was in place since 2001 - see Council Directive 2001/55/EC of 20 July 2001 *on minimum standards for giving temporary protection in the event of a mass influx of displaced persons and on measures promoting a balance of efforts between Member States in receiving such persons and bearing the consequences thereof*, OJ L 212/2001, 7 August 2001. The mechanism was used before the Russian invasion of Ukraine, mainly for people fleeing the wars in Bosnia and Herzegovina and Kosovo, but only for a selection of receiving countries (Beirens et al., 2016), and never on such a scale.

[22]In the period of 24-28 February 2022 alone, the Polish Border Guard reported 355,000 crossings from Ukraine (Source: https://twitter.com/Straz_Graniczna, data collected on 13 March 2022). Even though there is some delay in reporting the number of asylum claims, this particular example shows that the true numbers of asylum seekers will not show in the asylum data.



(a) Syrian citizens



(b) Ukrainian citizens

Figure 3: Number of monthly asylum applications

The number of monthly asylum applications made by Syrian and Ukrainian citizens in EU+ (EU, EFTA, UK, ME) 2008:01–2021:12. The vertical axis shows the number of monthly asylum applications in 1000s. The fall in early 2020 is explained by the COVID-19 pandemic. Standard deviations of 0.5, 1, 1.5 and 2.0 are shown by the horizontal lines. The vertical line on Figure 3a highlights the March 2016 statement between the EU and Turkey (Source European Council).

for most countries with population or migration registers, the numbers of people under temporary protection would show up in the official statistics, albeit with some delay.

Defining a crisis is a complex endeavour. The option to use multitudes of standard deviations, as discussed above – in the Ukrainian case, for the pre-2014 and post-2016 periods – needs not a be viable long-term solution, as the dynamics of asylum flows can fluctuate very rapidly. In addition, there are other features of migration from Ukraine that may mask the asylum signal: the largely visa-free character of migration, coupled with many work permits issued to Ukrainian citizens already before the Russian invasion[23]. An alternative approach would be to use a dichotomous modelling of the growth rate to exceed a certain percentage. However, this would then also hide any decreases (e.g. in 2016–17 following the peak in asylum applications, especially pronounced for Syrian asylum seekers).

In the ensuing analysis, we therefore try three options, with crisis period defined respectively as: (i) a binary indicator of when asylum applications exceed a function (e.g. a certain multitude) of the standard deviation threshold; (ii) a binary indicator triggered when growth rates exceed a set percentage; and (iii) a combination of the two indicators, set at higher sensitivity levels. This raises further questions as to the aim of the modelling exercise, namely is it to provide an alert to an *increase* in the number of asylum applications, or provide an alert that asylum applications will continue to remain high? Taking each question individually, they both have merits. The first one is clearly sensitive to the probability threshold (the value of c in Table 1), and a less sensitive analysis would require that the user would need to determine by how much the number of asylum applications would have to increase before the warning is triggered. For the Syrian case, prior to the sudden increases in 2012–2013, the user of an EWS system then would have seen the increases exceeding the standard deviations, whether a rolling 12-month sample or a total from the start of the series. Using a function of the standard deviation for the total sample in the present day only (as shown in Figure 3a)), a factor of 0.5 of the standard deviation would provide potential triggers from 2017 onwards. Using a rolling 12-month sample could lead to information loss, offering less information on the magnitude of any change in asylum applicants. The latter measure would, however, provide an indication that asylum applications would continue to remain high, which can be also useful – signal not present in the approach based on growth rates alone.

The second approach, based on rates of change, can help making preparations to

[23]In Poland alone, in 2021, some 325 thousand work permits were issued to Ukrainian nationals (source: Statistics Poland, https://stat.gov.pl/obszary-tematyczne/rynek-pracy/opracowania/zezwolenia-na-prace-cudzoziemcow-w-2021-roku,18,5.html),

accommodate the increase in the numbers of asylum seekers – but would does not tell the user whether there is a prolonged period of high flows, e.g. as in post-2017 Syria or today's Ukraine. A combination of both approaches, combining an increase in growth rates with the high flow relative to the standard deviation means that an alarm would be triggered less frequently relative to the two first approaches separately.

To examine the sensitivity of each approach, we present the number of warnings that would be triggered at different threshold levels. Table 4, in the spirit of Napierała et al. (2022), uses the multitudes of standard deviations. On top of their proposals of $1SD$ and $2SD$, we include $0.5SD$ and $1.5SD$. There are 170 periods observed in the sample 2008:01–2022:02. The first crisis indicator is for Syria in September 2013 at $0.5SD$, and for Ukraine in March 2014 at both $0.5SD$ and $1SD$. Table 4 presents therefore a descriptive frequency analysis of the data shown in Figure 3. This analysis can help identify any issues that might later arise from the modelling: while there is a steady increase with in the asylum applications for Syria, clearly visible in breaking the successive standard deviation-based thresholds, the same is not the case for Ukraine.

Table 4: Standard Deviation Based Analysis of the Numbers of Asylum Application

	2008:01-2022:02		**2008:01-2013:12**		**2014:01-2014:12**	
	Syria	***Ukraine***	***Syria***	***Ukraine***	***Syria***	***Ukraine***
Number of periods with the respective standard deviation thresholds exceeded (out of 170)						
$x \geq 0.5SD$	90	92	3	0	12	10
$x \geq 1.0SD$	28	77	0	0	6	10
$x \geq 1.5SD$	17	35	0	0	1	7
$x \geq 2.0SD$	15	21	0	0	0	6
Number of periods falling into different standard deviation intervals (out of 170)						
$0.0SD \leq x < 0.5SD$	80	78	69	72	0	2
$0.5SD \leq x < 1.0SD$	62	15	3	0	6	0
$1.0SD \leq x < 1.5SD$	11	42	0	0	5	3
$1.5SD \leq x < 2.0SD$	2	14	0	0	1	1
$2.0SD \leq x$	15	21	0	0	0	6

The top panel of the table shows the number of times that the standard deviation SD of asylum applications, x, exceeds the given multitude of the standard deviation for the respective country, as identified by the row. The lower panel of the table has the count broken down by specific intervals. There were 170 periods included in the analysis, which covers the period 2008:01-2022:02.

The lack of crisis indicators in asylum applications before 2014 for Ukraine based on the standard deviation approach requires us to reconsider the definition required for a binary crisis variable. The breakdown into different periods shown in Table 4 shows that no crises of asylum application were triggered pre-2014, primarily because in none of these months did applications rise above 100 asylum applications. The three alerts of $0.5SD$ for Syria in the whole period 2008–2013 were seen in September–November 2013, though there had been large growth rates in asylum applications observed since 2011.

Hence, the approach we examine in Table 5 focuses the number of crisis indicators based on growth rates, r, exceeding certain thresholds, with a reference time lag for calculating rates ranging from 3 to 12 months, and at different sensitivity levels, from 10% to 100%. Additionally, the upper panel of the table details how many times there growth rate has been positive, i.e. $r > 0\%$. As in Table 4, we include the number of indicators at each growth rate threshold, with the intervals presented in the lower panel.

Table 5: Growth Rate Based Analysis of the Numbers of Asylum Applications

	2008:01-2022:02								**2008:01-2013:12**							
	Syria				***Ukraine***				***Syria***				***Ukraine***			
	Number of months				Number of months				Number of months				Number of months			
	12	9	6	3	12	9	6	3	12	9	6	3	12	9	6	3
Number of periods with growth rate r exceeding selected limits (out of 170)																
$r \geq 10\%$	92	93	73	62	62	66	70	70	47	47	33	29	34	30	37	32
$r \geq 25\%$	86	66	62	43	44	50	53	48	43	33	32	19	20	24	26	21
$r \geq 50\%$	70	54	40	25	36	31	26	26	32	24	21	13	15	11	7	11
$r \geq 100\%$	49	35	22	10	26	24	19	14	24	17	12	4	6	6	5	3
r positive	100	101	89	86	77	75	86	86	51	49	42	41	35	32	39	39
Number of periods with r falling into different intervals (out of 170)																
$0 \leq r < 10\%$	8	8	16	24	15	9	16	16	4	2	9	12	1	2	2	7
$10\% \leq r < 25\%$	6	27	11	19	18	16	17	22	4	14	1	10	14	6	11	11
$25\% \leq r < 50\%$	16	12	22	18	8	19	27	22	11	9	11	6	5	13	19	10
$50\% \leq r < 100\%$	21	19	18	15	10	7	7	12	8	7	9	9	9	5	2	8
$100\% \leq r$	49	35	22	10	26	24	19	14	24	17	12	4	6	6	5	3

The top panel of the table shows the number of times that the growth rate r of the number of asylum applications, relative to the time lag listed in the column heading, exceeds the given percentage as identified by the row. The lower panel of the table has the count broken down by specific intervals. There were 170 periods included in the analysis, which covers the period 2008:01-2022:02.

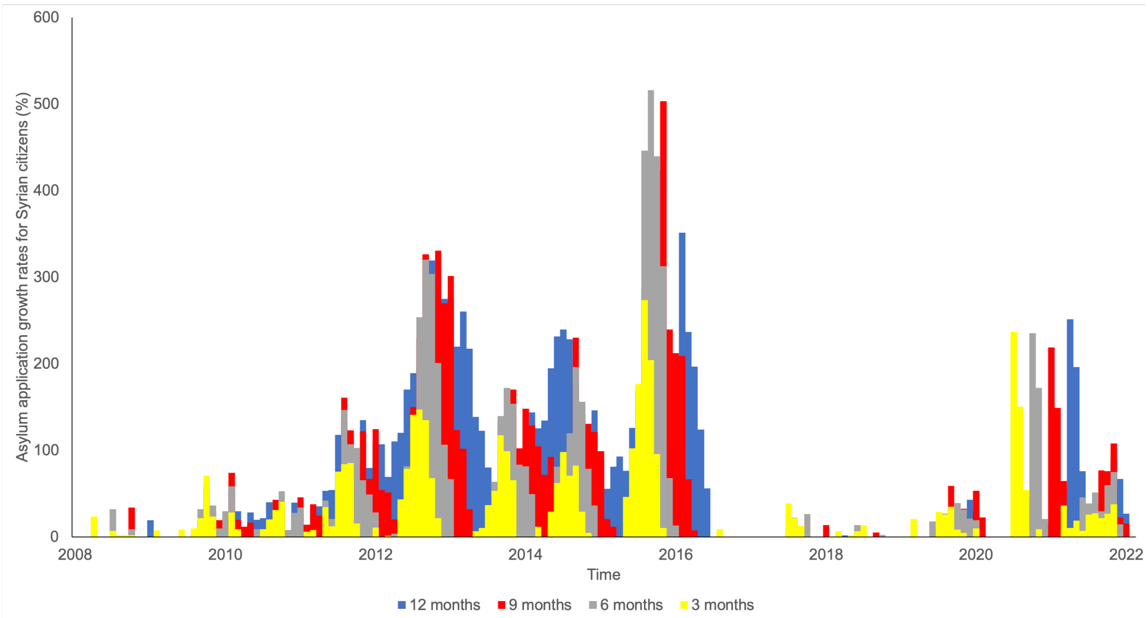
The growth rate-based method allows for a greater balance of alarms to be triggered along the whole time series, with some alerts for Syria also triggered prior to 2013. This approach also helps to provide a context to whether there is an increase or continued high levels of asylum applications. As we see in Figures 4a and 5a, the shift towards triggers prior to 2013 is remarkable, although very few alarms are triggered for Syria, after 2016.

In analysing Table 5 we can also prioritise the most suitable lag lengths and intervals. For Syria, the most frequent category of positive growth rates was in excess of 100%, except when using three-month lags, for which the most common rates were smaller than 10%. In contrast, for Ukraine, only at 9 and 12-month lags, were the most common occurrences of a growth rate greater than 100%. For the six-month lag, the most frequent category was $25\% \leq r < 50\%$, and for the three-month lag, the second and third categories ($10\% \leq r < 25\%$ and $25\% \leq r < 50\%$) were equal in size. The twelve-month lag accounts for seasonality but fails to capture quickly-moving crises. At the shortest length, 3 months, large growth rates are less common but if they occur, they are more likely to denote large changes.

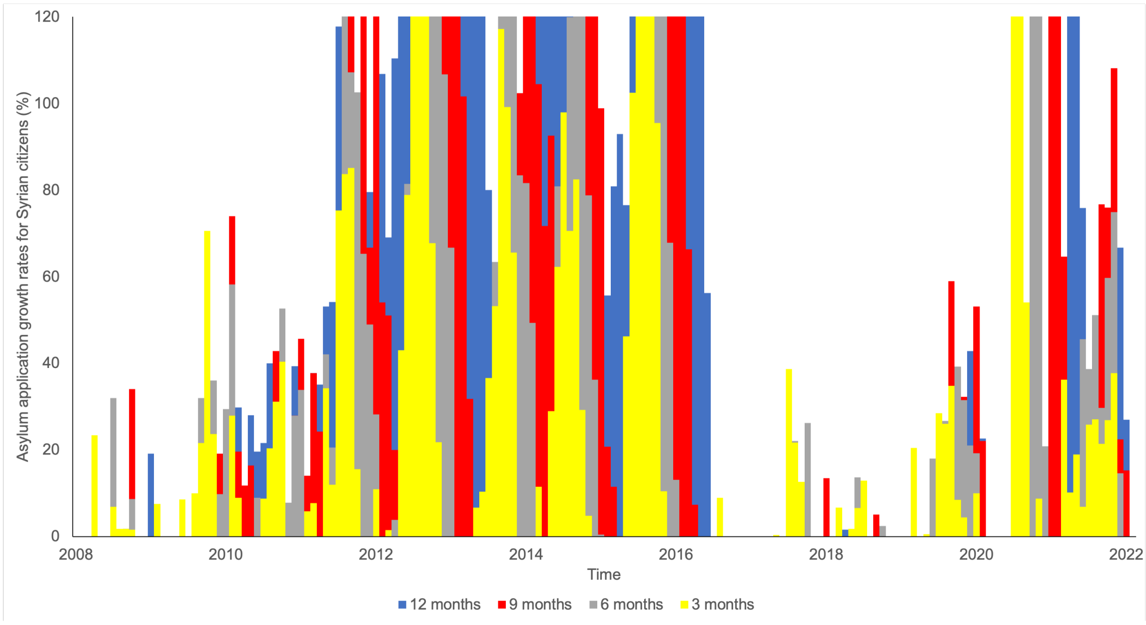
As a consequence of reporting delays, there is a lag at which migration data can be called current. For instance, in May 2022 we were only able to obtain an incomplete report of February 2022 asylum applications via Eurostat. Of course, for internal uses within government institutions, the lags would be shorter. To account for this problem in our analysis, we try models that change the reference time of asylum applications to reflect the *most recent data*, which we assumed to correspond to a three-month lag.

For these reasons, in an early warning model, we propose using a combination of identifiers to incorporate standard deviations, growth rates, and reducing the impact of small values before 2011. If we were modelling this in 2013, for example, the standard deviation for both series would be a small fraction of the overall value for 2008–2021. We thus propose using rolling standard deviation, covering the previous 12 months of asylum application, which also accounts for some seasonality. In addition, we include a growth rate for the current period of 20%, with a requirement for at least one of the last twelve periods to have a growth rate of 25%. Finally, we introduce a minimum requirement, such that the number of current applications is greater than 300 for Syria and 50 for Ukraine. This significantly reduces the number of triggers before 2011, which values are considered "normal" under the proposed approach.

Table 6 shows the number of times that such a warning is triggered for Syria and Ukraine, in 2008–10, 2011, 2012, 2013, 2014, and 2021. The reasons for selecting these particular years is that by analysing these periods we can pick up a preferred indicator of a crisis in advance of the *actual* crisis happening. For example, some of the indicators



(a) Growth rates of the number of asylum applications by Syrian citizens in EU+ countries (EU, EFTA, UK, ME), per cent

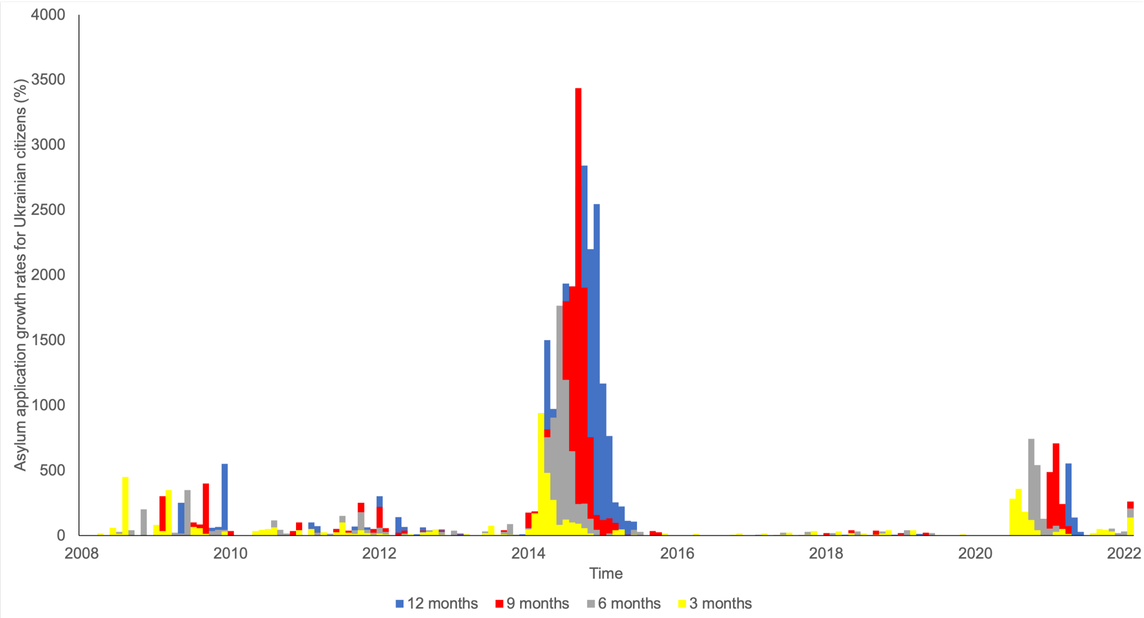


(b) Growth rates of the number of asylum applications by Syrian citizens in EU+ countries (EU, EFTA, UK, ME), zoomed (maximum value of 120%)

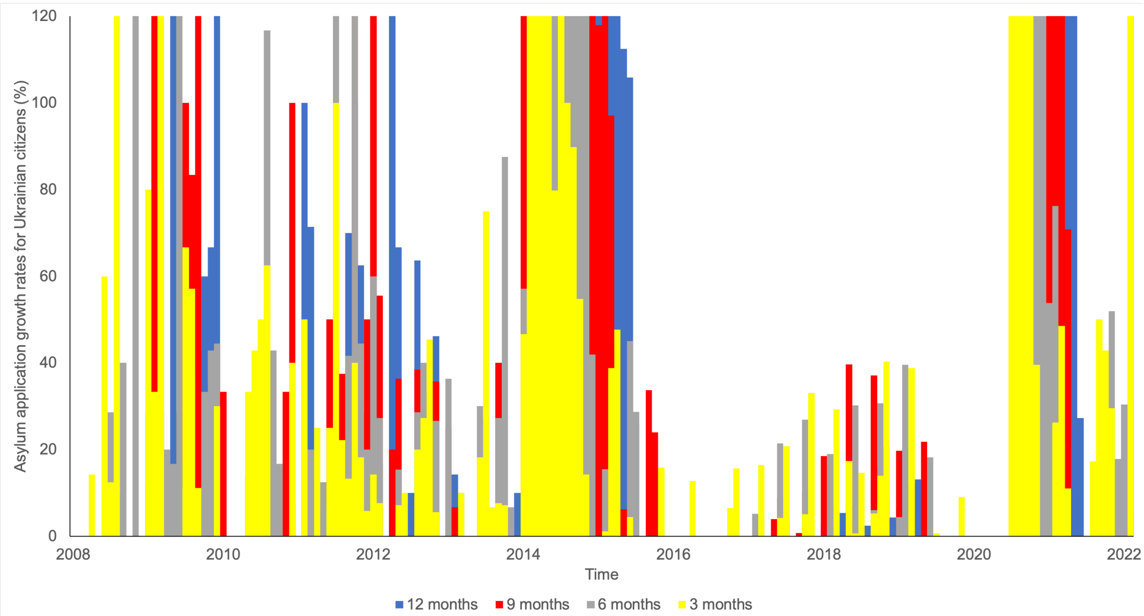
Figure 4: Growth rate of asylum applications by Syrian citizens in Europe

Growth rates of the number of monthly asylum applications made by Syrian citizens. Figure 4b has a maximum value of 120% so that the smaller growth rates are clearer. A 12-month growth rate is given by the blue bars, 9-month by the red bars, 6-month by the grey bars, and 3-month by the yellow bars.

for Ukraine that are too conservative, would not show up in 2013, e.g. requiring a 50% growth rate in the last 12 months would only trigger a response in January 2013, whilst



(a) Growth rates of the number of asylum applications by Ukrainian citizens in EU+ countries (EU, EFTA, UK, ME), per cent



(b) Growth rates of the number of asylum applications by Ukrainian citizens in EU+ (EU, EFTA, UK, ME), zoomed (maximum value of 120%)

Figure 5: Growth rate of asylum applications by Ukrainian citizens in Europe

Growth rates of the number of monthly asylum applications made by Ukrainian citizens. Figure 5b has a maximum value of 120% so that the smaller growth rates are clearer. A 12-month growth rate is given by the blue bars, 9-month by the red bars, 6-month by the grey bars, and 3-month by the yellow bars.

25% growth rates at various lag lengths provide some signal throughout the year. In selecting these years to examine, we can see the effect of the start of the Arab Spring, the Syrian Civil War, the Euromaidan protests in Ukraine, the lead up to the 2015–16 crisis, as well as the build-up to the Russian invasion of Ukraine in 2021. Introducing the minimum limit is less effective for Syria, but allows us to drastically reduce the number of triggers for Ukraine pre-2011. When we introduce the minimum magnitudes of 300 for Syria, and 50 for Ukraine, the number of warnings removed for Syria is only one at the 3 month lag in 2008–10, while for Ukraine, 5, 7, 7, and 8 warnings are removed respectively for the 12-, 9-, 6- and 3-month growth rates. These all removed warnings concern the period 2008–11, which is desirable.

The threshold values for Syria in the upper section of Table 6 provide a large number of signals, which in isolation is not helpful particularly when asylum applications did not exceed 500 per month until July 2011. In the 168 months covering 2008:01–2021:12, the numbers of asylum applicants from Syria exceeded the threshold value in nearly half of the periods which, suggests that a higher minimum limit would be required. With the minimum requirements for the growth rate of at least 50% in a given period and a least once over 100% in the last 12 months, along with exceeding rolling two standard deviations, there is a significant reduction of the number of warnings triggered. The introduction of a rolling standard deviation has a similar absolute effect in reducing the number of warnings, as was the case with using lower thresholds. Higher thresholds eliminate all the triggers in 2008–10 for Syria, and reduce their numbers significantly in other period.

Still, even under this approach, some problems with the Syrian early warnings persist, in that from when the first warning is triggered in 2011–12, with only a few exceptions, it continues to be triggered until 2016, and then gets switched off until 2020, where the numbers of asylum applications decline, yet remain at high levels. With growth rates of 100% in the current period, it is easy to argue that such rapidly-growing numbers of applications can be classified as a 'crisis'. As an even more sensitive alternative, we can set the minimum growth rate to 50% in the present period, with another 50% growth rate to have occurred in the previous 12 months. This definition provides a small increase to the number of triggers for Syria, most notably in 2011. At the same time, this measure is unsuitable for Ukraine as there are no triggers in 2013, which is when we would expect a signal. In Section 4, we run the model for Syria with the higher and lower thresholds,

Table 6: Number of EWS triggers in selected years

	Syria				***Ukraine***				***Syria***				***Ukraine***			
	Number of months				Number of months				Number of months				Number of months			
	12	9	6	3	12	9	6	3	12	9	6	3	12	9	6	3
	20% ≤ r and 25% ≤ r in 12 months								*20% ≤ r; 25% ≤ r and 2SD in 12 months*							
2008-10	8	7	7	6	6	10	13	13	8	7	7	6	6	10	10	11
2011	11	10	9	4	6	8	6	6	11	10	9	4	6	8	6	5
2012	12	12	9	7	5	6	6	3	12	12	9	7	5	6	6	3
2013	12	9	6	5	1	2	4	1	12	9	6	5	1	2	4	1
2014	12	12	9	6	12	12	12	10	12	12	9	6	12	12	12	10
2021	9	9	6	6	4	4	4	5	9	9	6	6	2	4	4	4
2008-21	84	74	61	46	41	53	53	47	81	71	60	45	39	53	50	43
	50% ≤ r and 100% ≤ r in 12 months								*50% ≤ r; 100% ≤ r and 2SD in 12 months*							
2008-10	0	0	0	0	3	3	2	6	0	0	0	0	3	3	1	4
2011	5	4	3	0	5	4	2	0	5	4	0	0	5	4	2	0
2012	12	9	7	3	4	2	1	0	12	9	5	3	4	2	1	0
2013	11	7	6	4	0	0	0	0	11	7	6	4	0	0	0	0
2014	12	12	7	4	10	11	9	8	12	12	7	4	10	11	9	8
2021	6	5	3	0	2	3	3	1	6	5	2	0	1	3	3	1
2008-21	64	47	34	17	30	26	19	18	61	44	26	17	29	26	18	16
	50% ≤ r and 50% ≤ r in 12 months								*50% ≤ r; 50% ≤ r and 2SD in 12 months*							
2008-10	0	1	2	0	4	3	2	8	0	1	1	0	4	3	1	6
2011	7	5	5	2	5	4	2	2	7	5	5	2	5	4	2	2
2012	12	9	7	5	4	2	1	0	12	9	7	5	4	2	1	0
2013	11	7	6	4	0	0	0	0	11	7	6	4	0	0	0	0
2014	12	12	7	4	10	11	11	9	12	12	7	4	10	11	11	9
2021	7	6	3	0	3	3	3	1	7	6	2	0	2	3	3	1
2008-21	67	51	38	22	32	26	22	23	64	48	34	22	31	26	21	21

The top panel of the table shows the number of times that the a trigger for an EWS is given when the value for the period growth rate exceeds 20% and in the last 12 months there has been one period that exceeded 25%. The right-hand section uses the same measure, with the addition of a requirement for the number of asylum applications to exceed 2 standard deviations *SD*. The centre panel has a minimum growth rate for the period at 50% and a 12 month requirement for 100%. The right-hand section in the centre panel has an additional requirement for the number of applications to be greater than two standard deviations in a rolling 12 month sample. The final (bottom) panel has a current period growth rate requirement of greater than 50%, and another value of 50% in the previous 12 months. The second section has the rolling standard deviation that requires the number of asylum applications to exceed 2*SD*. The growth rate, r of the number of asylum applications is relative to the time lag shown in the column headings, exceeding the given percentage as identified by the panel heading. There were 168 periods included in the analysis, which covers the period 2008:01–2021:12.

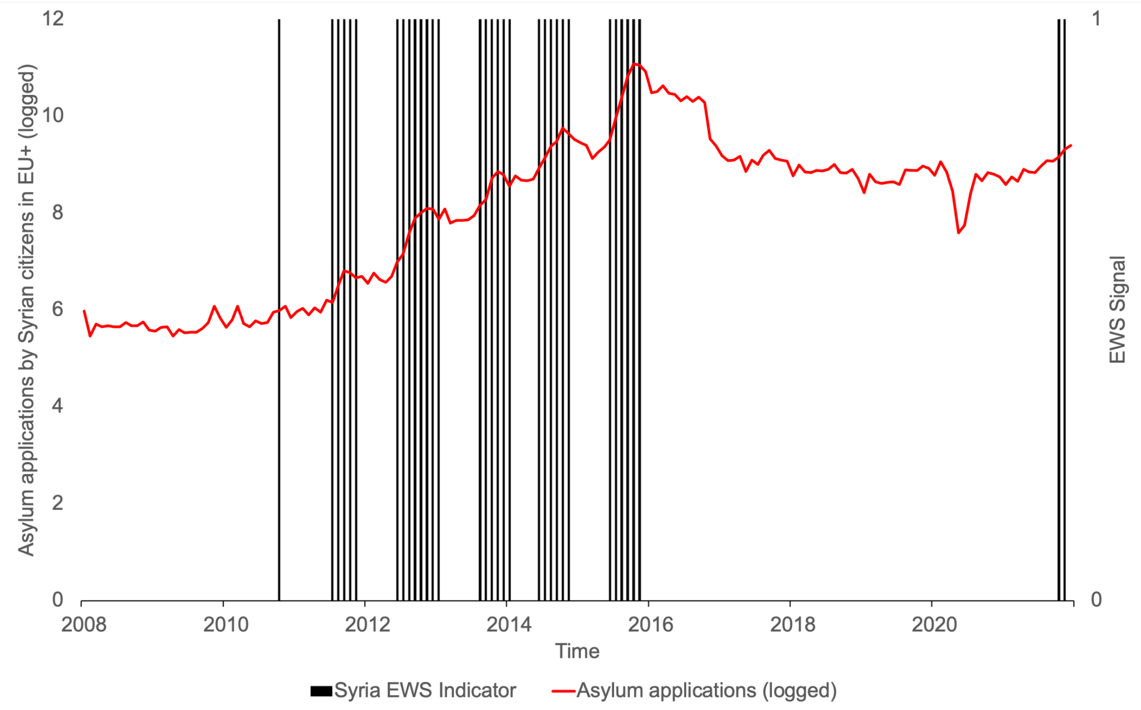
while for Ukraine only at the lower threshold. The reason is that for Syria we want to learn as much from the available data as we can, across different sensitivity levels.

Figure 6 shows the occurrence of a crisis, or indicators employed for Syria and Ukraine in the model introduced in Section 4, along with the (log of) asylum applications lodged by citizens of each country. Figure 6a, for Syria, uses the present period of minimum growth rate of 50%, as well as at least one value at 50% or more in the previous 12 months, with the value of asylum applications exceeding $2SD$ of the previous 12 months in order to trigger a warning. Figure 6b, for Ukraine, shows the more sensitive approach, with the present period growth rate of at least 20%, at least one value greater than 25% in the previous 12 months, and the number of applications greater than $2SD$ from the preceding year. For greater visibility of important changes, the log-transformed numbers of asylum applications are shown.

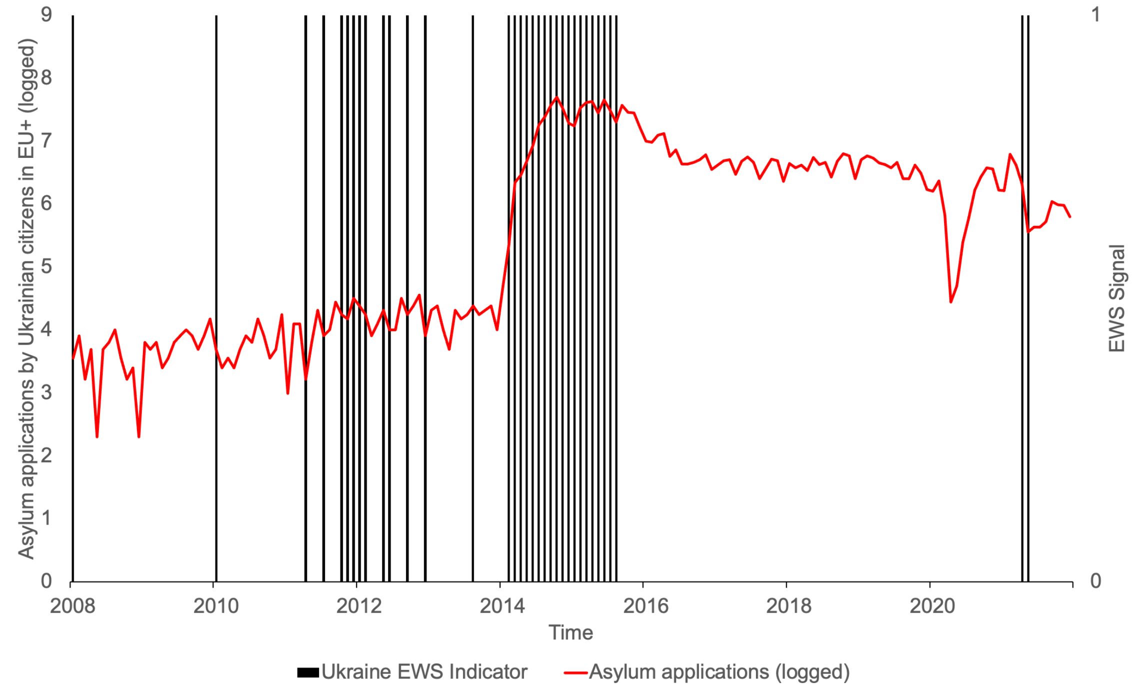
In this section, we have discussed the advantages and disadvantages of the different methods and indicators used to define a 'crisis' in the context of asylum application. Two different, yet not mutually exclusive, questions are important here: (i) is there asignificant stochastic movement in the number of asylum applications, or (ii) is there a high volume of asylum applications. We have shown that for methods using the whole period of analysis (2008:01–2021:12) standard deviation-based measures are unable to ensure the required accuracy pre-2014. Secondly, while growth rates of asylum applications offer an indicator of whether there is a an increase of asylum applications, these can be particularly sensitive to changes at lower number of applications, which are manageable from a policy perspective, but hide crises if the numbers are too high but not growing compared to previous levels. There no single formula that can answer both questions, and a degree of subjectivity is inevitably required in defining the 'crisis'. More generally, the question of identifying a crisis remains to some extent open. A promising avenue of future work in this area could include automation via formal machine learning or statistical models, such as LASSO regression, for selecting the optimum variables for a given purpose[24].

By comparing different definitions, we see that for Syria a tighter one is needed than for Ukraine, to enable focus on periods of high growth in the numbers of applications. In this report, by prioritising preparedness for large increases in asylum applications, we specifically focus on the first question above, which is the detection of *changes* in the numbers. In our case studies presented in Section 4, the aim of our EWS models is to see if there was any signal in the series related to from Syria (in 2014) and Ukraine (in the second half of 2013 and in 2021), presaging the sharp increases of the numbers of asylum seekers in the subsequent periods (2015–16 and 2014 as well as 2022, respectively). Specifically, we are looking at identifying warnings with (ideally) a 3-6 months lead time.

[24]We are grateful to Peter W F Smith and Jason Hilton for this suggestion.



(a) Syria



(b) Ukraine

Figure 6: EWS triggers and monthly asylum applications (logged)

The number of monthly asylum applications (log-transformed) made by the nationals of Syria (top) and Ukraine (bottom) in EU+ (EU, EFTA, UK, ME), 2008:01–2021:12, shown as the red line, measured on the left axis. The presence of EWS triggers, defined in the text above, is shown by black bars.

4 Case Studies

In order to test the methodology proposed in this report, we look into two of the most challenging examples of asylum flows into Europe from the recent decade, that is, those related to the war in Syria, especially in 2015–16, and the Russian political conflict with Ukraine since 2013, culminating in the full-scale war and invasion in 2022.

4.1 Syria 2015–2016

4.1.1 Historical Background and Data

The first case study designed to test our model is the related to asylum applications of Syrian nationals (or people claiming Syrian citizenship) in the EU and neighbouring countries, with focus on the events of 2015–16, which are well documented in the wider migration literature. The Arab Spring protests across the Arab world, led to political unrest across the region and to a full-scale civil war in Syria. Since then, increasingly more asylum seekers, especially Syrians, were trying to reach Europe. With some family members having left earlier on in this period, family reunification followed suit.

The reasons that have caused the civil war, which in turn triggered large migration flows, are predominantly political – related to violence and persecution, but the decisions to flee are largely mediated by the existence of opportunities, social and human capital, and access to information – and by chance (e.g. Schon, 2019; Belabbas et al., 2022). Some of the background, macro-level explanations can also include e.g. environmental factors – for instance, Selby et al. (2017) sought to link climate change to the Syrian civil war. According to this argument, extreme drought (a short-term, high-impact event) that proceeded the civil war can be attributed to climate change (long-term process), with problems arising from the drought causing and exacerbating civil unrest in Syria.

The changes in the number of asylum applications in Europe ('EU+': EU and EFTA countries plus the United Kingdom and Montenegro) lodged by people who have fled Syria are shown in Figures 2a and 3a, with most pronounced changes following the start of the Syrian civil war in 2011. The conflict continued to worsen into 2014–15, with asylum applications in EU+ peaking in September 2015 at 64,040 persons per month according to Eurostat. A standard deviation for this period is 11,350 persons per month.

4.1.1.1 The Syrian crisis in the light of the GDELT data

In an attempt to identify possible explanatory variables that could serve as early warnings, first, we examine the number of protests and total number of political events in Syria in the period 2007:01-2022:04[25], as shown in Figure 7. The spike in the number of protests in Figure 7a, is replicated in the number of total events, and once the civil war starts, there is a further increase in the number of events. Notably, the protests as such have largely decreased by the end of 2012. In Figure 8, we see the protest *intensity*, defined as the number of protests divided by the number of total recorded events. The protest intensity peaked in May 2021, however, during the Arab Spring the protest intensity was at its maximum in April 2011, which coincided with the maximum number of protest events per month registered in Syria in that period[26].

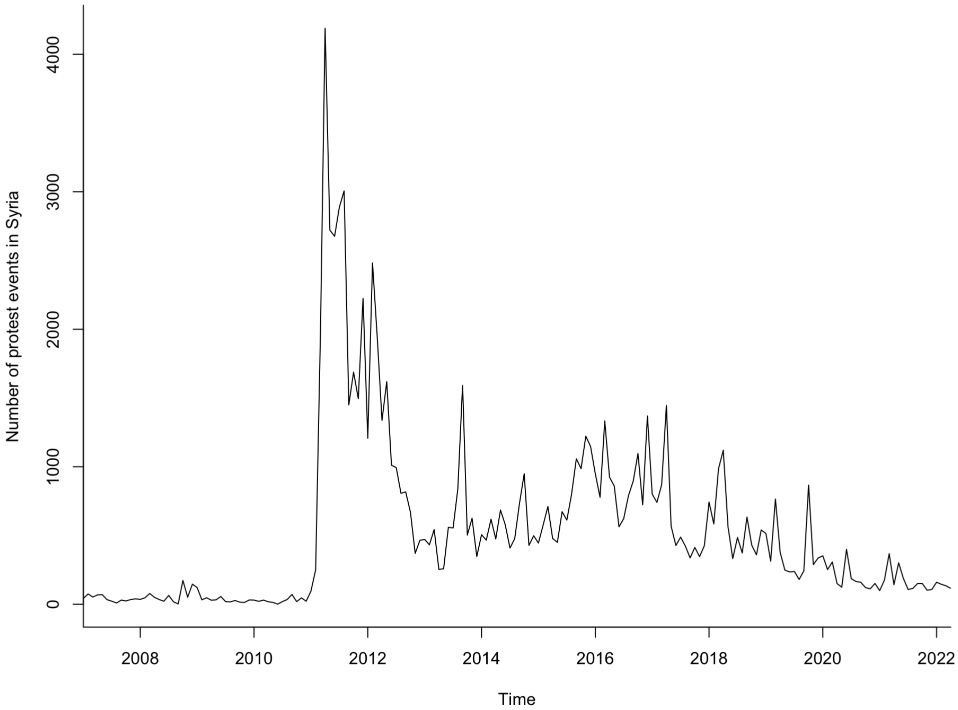
Even though the number of protest events can tell a profound story, another important variable available in the GDELT database is the average *tone* of the reporting of the event in the media. The average tone, as defined by the GDELT collection, is how positively or negatively the reports of the event are. The tone scores range from −100 (extremely negative) to +100 (extremely positive), though more commonly they range between −10 to +10, with 0 being considered neutral. We cannot identify the actual events or reports using this approach of GDELT, however, through a combination of actors, date and location we are able to isolate them and approximate the likely events. The average tone is useful in determining the context and seriousness of protests or conflict. Small riots are likely to receive a small negative score, whilst large negative scores would indicate more serious events.[27]

For Syria, there are three notable changes in the tone of reporting: in early months of 2011, at the end of 2012 and, most pronounced, in winter 2014/15. Figure 9 shows the average tone of reporting of protests and all political events in Syria, overlaid with the number of asylum applications. There is a clear time lag between the changes in the average tone of reporting and the number of asylum applications, but at the same time,

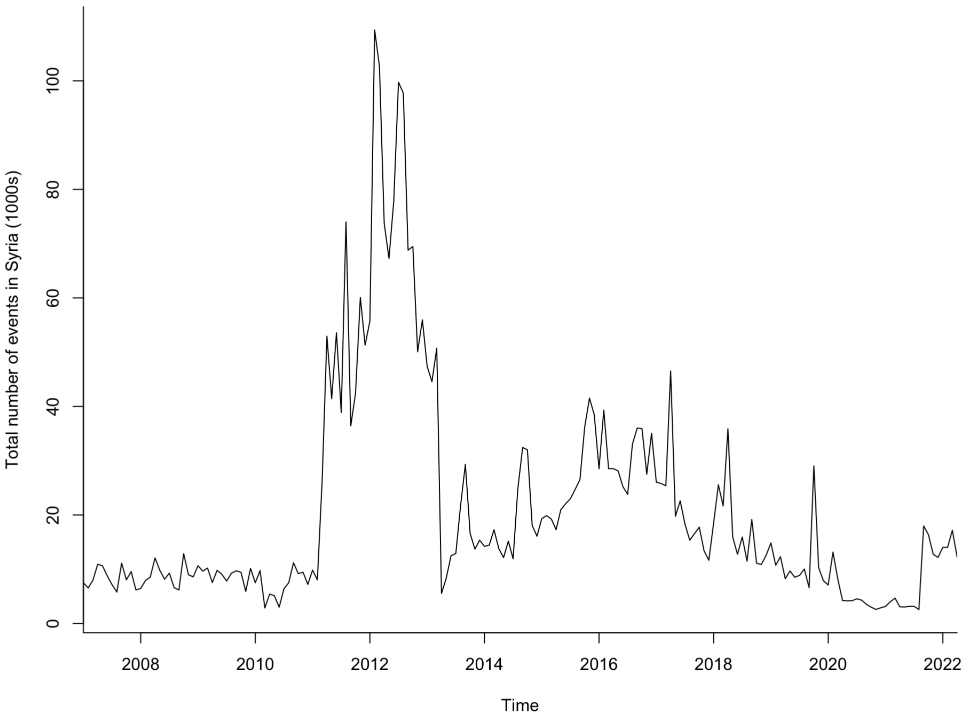
[25] GDELT data is unavailable for extraction for 23-25 January 2014 and 19 March 2014.

[26] Given the large numbers of Syrians seeking asylum, it is noted that the number of protests could be negatively correlated with asylum applications. Even though the fall in the number of protests does not not perfectly coincide with the increase in asylum applications, there is a clear relationship between the events of the Arab Spring (Figure 7a and the start of the increase in asylum applications. There is, however, a change in other variables available in the GDELT database, as later shown in Figure 9.

[27] For further details on fields of the GDELT definitions, see the GDELT Codebook.



(a) Number of protests in Syria



(b) Number of political events in Syria (1000s)

Figure 7: Number of protests and all political events in Syria

The first figure shows the number of protests in Syria for 2007:01-2022:04. The second figure shows the number of all political events in Syria for 2007:01-2022:04. Protests are filtered by using GDELT code $EventRootCode ==$ "14", and the location by $ActionGeo\_CountryCode ==$ "SY"

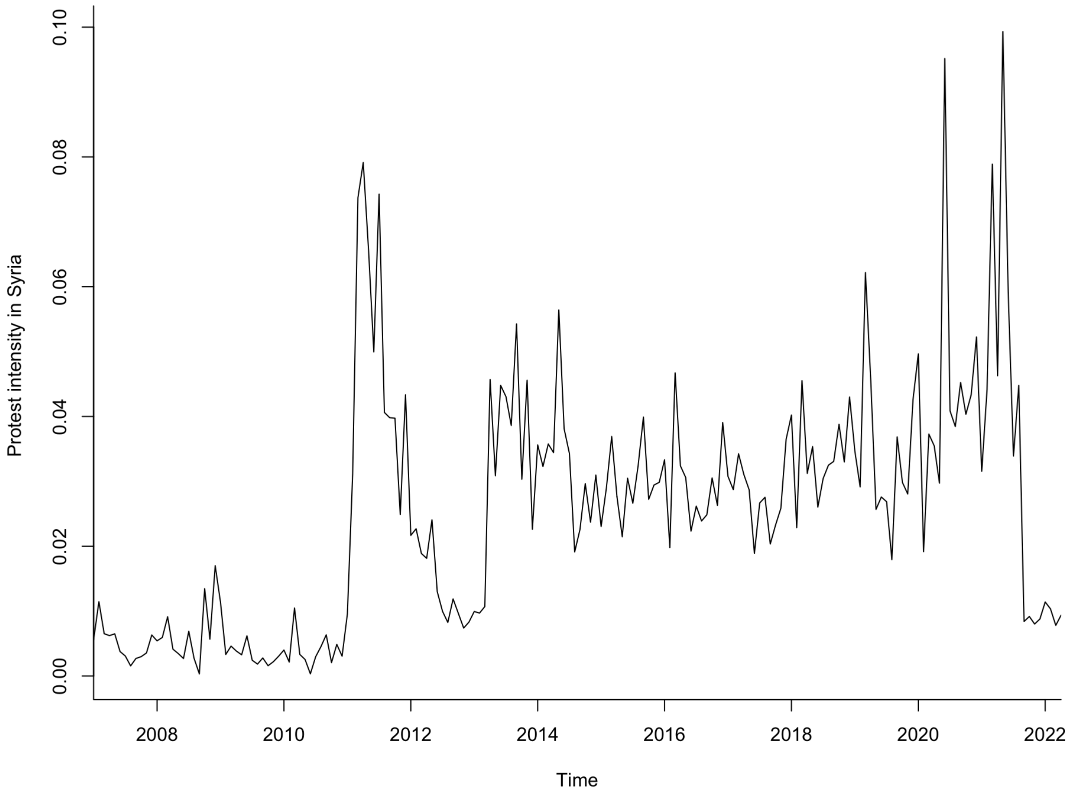


Figure 8: Protest intensity in Syria

The relative protest intensity in Syria using events reported on the GDELT database 2007:01-2022:04. The protest intensity is calculated by summing the total number of protests in each month, divided by the total number of events. These are identified using GDELT codes $ActionGeo\_CountryCode == SY$, with protests filtered using $EventRootCode == 14$.

there seems to be a visible relationship between these variables as well. This information can help build our model, since the presence of any lags can offer an opportunity for policy makers to prepare for a possible change in the numbers of asylum applications.

4.1.1.2 The Syrian crisis in the light of the macroeconomic data

Within macroeconomics, key parameters of macro-financial stability include exchange rate and inflation[28]. However, for some less developed countries, and in particular, conflict strewn or under economic sanctions from e.g western economies, data are often unavailable or unreliable. A key example of data being reduced in quality (and thus its

[28] High rates of inflation, particularly hyperinflation, can be a push factor for would-be asylum applicants. High rates of inflation weakens the currency and can make it hard to access essential supplies. There are a number of examples of this, both in conflict or war-torn countries but also in politically-unstable countries such as Venezuela, which has seen over 3 million refugee since 2015 (Source: UNCHR Global Trends 2018, accessed 1 September 2022, as well as Barker (2020).) The high rates of inflation can be thus an early indication of an increase of asylum applications.

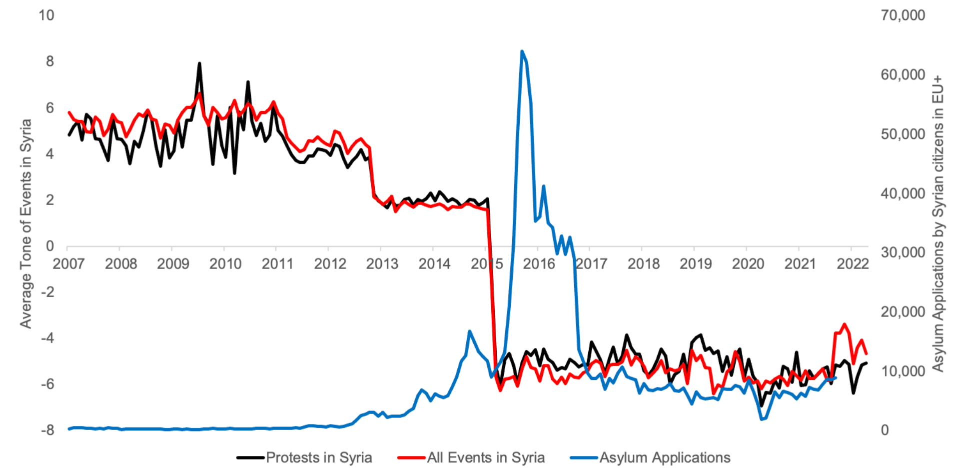


Figure 9: Average tone of events in Syria

The average tone for protests in Syria is given by the black line, the average tone for all events in Syria is given by the red line (both shown on the left-hand axis). The number of asylum applications made by Syrian citizens in EU+ is shown by the blue line (right-hand axis). Protest events are filtered by using GDELT code EventRootCode == "14", and the location by ActionGeo_CountryCode == "SY"

information value) is the official exchange rate: an example for the Syrian Pound (SYP) exchange rate against the USD is shown in Figure 10. Notably, the series of data available from the International Monetary Fund (IMF), and a number of other financial data outlets, stops towards the end of 2018.

From a macroeconomic perspective, we can also see very clear effects of the Syrian civil war by looking at the US imports and exports from/to Syria, as shown in Figure 11. This data are available to 2022, with some reporting lag to be expected. There is a sharp decrease in trade in winter 2011 – exports peaked in September 2010 at $ 89.6 million, a year later they were down to just $3.4 million. Imports peaked in June 2011 at $ 116 million, falling to $1.9 million only six months later. Only a handful of months since, trade flows barely exceeded $2 million per month. This was predominantly caused by the sanctions placed on Syria by the US Government, in particular Executive Order 13582 signed on 18th August 2011 (The United States Government Publishing Office, 2011).

4.1.1.3 The Syrian crisis in the light of Google Trends

There have already been some attempts to employ Google Trends data in detecting signal of increasing asylum applications, as mentioned in Section 2. In Table 7 we present the variables we have prioritised for modelling, which can differ from those used before. We

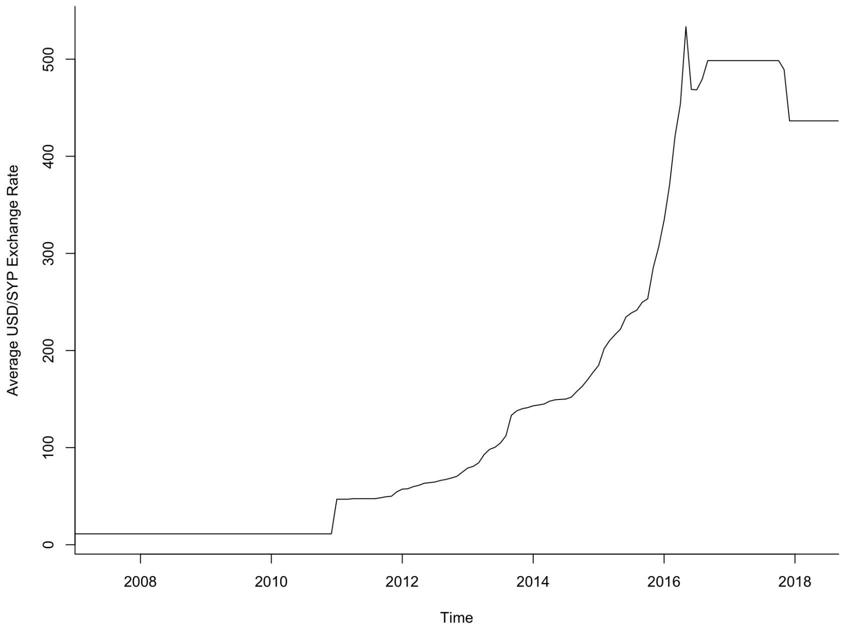


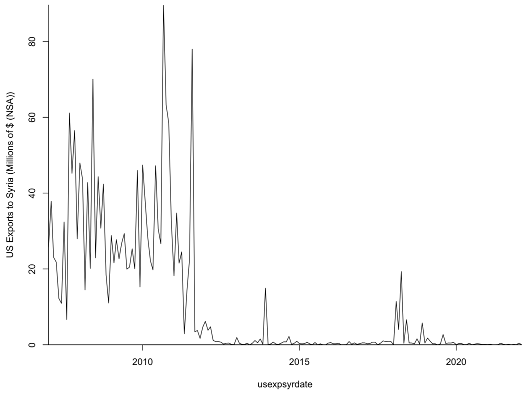
Figure 10: Nominal exchange rate of USD-SYP.

The nominal exchange rate for the currency pair US Dollar and Syrian Pound. The value is given as the number of Syrian pounds per one USD. An increase in the exchange rate, is an appreciation of the USD, equivalently, a depreciation of the SYP. Source: IMF - International Financial Statistics.

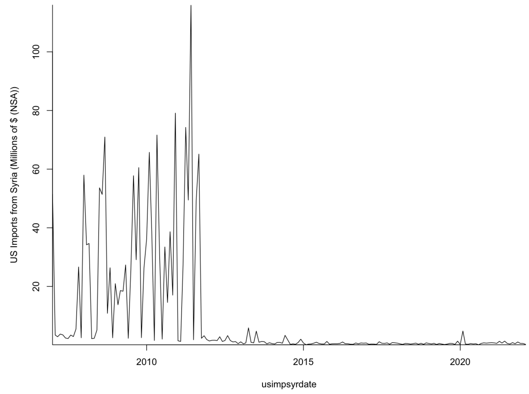
included searchers on various border towns to the south and to the west of Syria, and Lebanon, and on some of the largest refugee camps. For asylum seekers coming from southern Syria the nearest border would be one of towns listed in Table 7, while the eastern route, to Iraq via a desert in south-east Syria, would not be a primary choice for most people. Noteworthy, both Damascus and Homs, two of the largest cities in Syria, are located within 40km of the Lebanese border, whilst the Turkish border is just in excess of 150km from Homs and 300km for Damascus. Regarding those who decides to reach Turkey, though, Google searches for border towns and refugee camps in Turkey did not return meaningful results, or the signal was too weak to show any patterns.

In Google Trends, we searched through a large number of terms[29], both in Arabic and English, though in our dataset we have only ended up included those listed in Table 7. The impact of the language varies, though to differing extent for different terms.

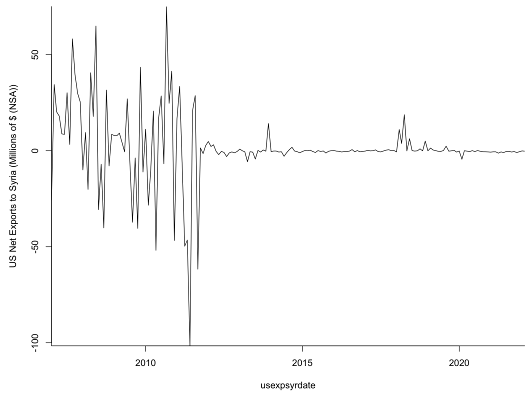
[29]When searching for terms on the Google Trends platform, the user is able to select words, search terms or topics (groups of terms related to the same concept). For example, *Goodbye*, can be selected as a search term for only 'Goodbye', or the song 'Goodbye' by David Guetta and Jason Derulo. Equivalently, *Asylum* can be a search term or a topic – here we used it exclusively as a search term. *Turkey* used on its own as a search term could refer to many other events, such as the collapse of Turkish Lira in 2018.



(a) US exports to Syria



(b) US imports from Syria



(c) US net exports to Syria

Figure 11: US trade with Syria

The figures show the US exports, imports and net exports of goods to/from Syria at a monthly frequency. The figures are given in millions of US$ not seasonally adjusted (NSA). Source: U.S. Bureau of Economic Analysis (via FRED: Federal Reserve Bank of St. Louis); codes EXP5020, IMP5020, and own calculations for net exports.

An example is shown in Figure 12 for 'Tell Abyad' – a town in northern Syria on the border with Turkey, and an important border crossing. Especially prior to 2012, there are significant numbers of months in which the search intensity in one language equals 0, while being greater than zero in the other language. Post-2012 though, the main differences occur in the size of the relative index values for both languages.

Figure 13 shows a selection of other Google Trends indicators that we looked at. There are a few points worth noting here. 'Reunification' increases in significance from August 2014 onwards, reaching a maximum in September 2015. This is in contrast to 'Goodbye', which peaks in October 2012, before reducing significantly. There is a similar pattern for

Table 7: Google Trends

Google Trends search terms	
Airport (A)	Lebanon (A)
Asylum (A)	Mafraq, Jordan (A)
Diaspora (A)	Refugee (A)
Duolingo (E)	Reunification (A)
Emigration (A)	Smuggler (A)
German Embassy (E)	Smuggling (A)
Germany (A)	Tell Abyad (A)
Goodbye (A)	Tell Abyad (E)
Greece (A)	Turkey Border (A)
Immigration (A)	

A list of Google Trends terms included in the exploratory analysis for the EWS model. *(A)* identified a search term in Arabic, whilst *(E)* denotes searches in English. The differences in language searches vary, with some being quite similar whilst others not. NB: Mafraq is a border town in Jordan. Data exported 9 March 2022.

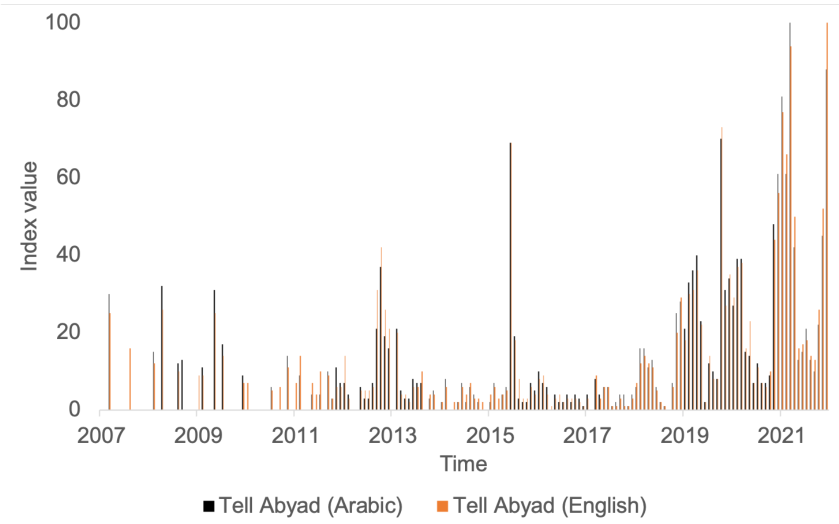
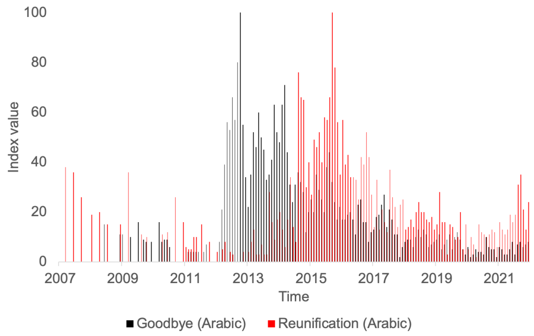


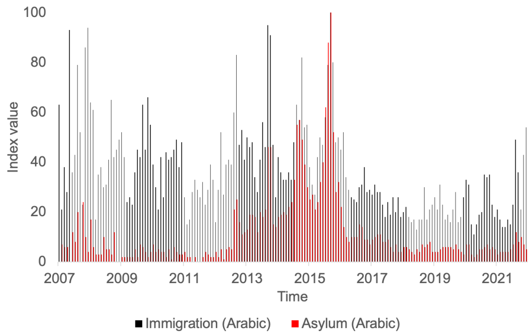
Figure 12: Google Trends for Tell Abyad in Arabic and in English

The Google Trend index for searches in Syria for ‘Tell Abyad’ in English and Arabic, showing small differences. Source: Google Trends

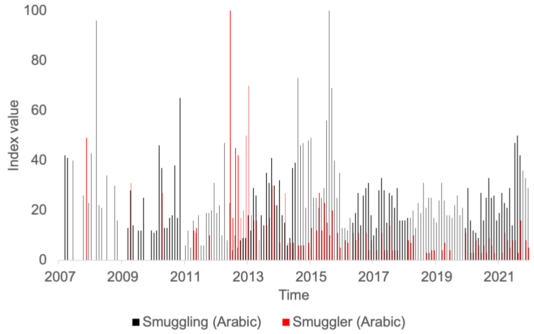
‘immigration’ and ‘asylum’, as shown in Figure 13c, with ‘immigration’ returning higher search values before 2015, before ‘asylum’ started to dominate in 2014–16. To analyse the difference of similar words, Figure 13c additionally looks as ‘smuggling’ vs ‘smuggler’. ‘Smuggling’ generally has a higher search intensity score, but there are distinct changes observed for ‘smuggler’, for which high levels are observed from June 2012 to April 2014,



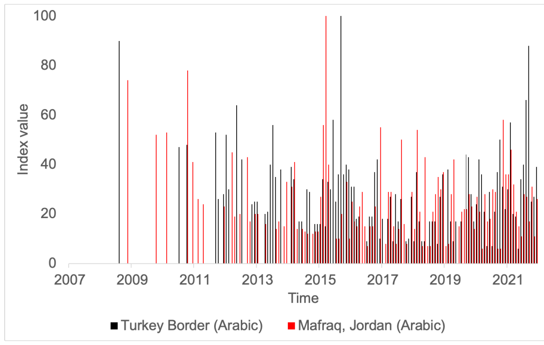
(a) Goodbye (black) Reunification (red)



(b) Immigration (black) Asylum (red)



(c) Smuggling (black) Smuggler (red)



(d) Turkey Border (black) Mafraq (red)

Figure 13: A selection of Google Trends values

The figures show the values from January 2007 to December 2021 for Google Trend searches made from Syria. The maximum value is 100. The searches shown in black include (in sequence), 'Goodbye' (Arabic), 'Immigration' (Arabic), 'Smuggling' (Arabic), and 'Turkey Border' (Arabic), whilst the terms depicted in red are 'Reunification' (Arabic), 'Asylum' (Arabic), 'Smuggler' (Arabic), and 'Mafraq, Jordan' (Arabic).

followed by small increases in 2015. Whereas 'smuggl*ing*' exhibits a variety of peaks, with the highest between August 2014 and October 2015. Finally, Figure 13d looks at the searches for 'Turkish border', and a border city Mafraq, Jordan. Since 2015, the value of search indicators for either term rarely equalled zero, and the peaks in the searches for Mafraq tended to come before those observed for Turkey.

Still, on balance, for early warning purposes we cannot only rely exclusively on Google Trends, given insufficient number of people who had access to internet in Syria, and thus Google[30]. On the other hand, the GDELT and US trade data are ones that bear more promise in terms of the information potential that they may carry. We explore them further in the modelling exercise presented in the next section.

[30] In 2009, 17% of the Syrian people had access to the internet, 28% in 2014, and 36% in 2020. Though there have been significant population changes since then, these measures show that Google Trends only can tell part of the story.

4.1.2 Early Warning Modelling Results

The data employed in the early warning model for Syrian asylum applications cover January 2009 to March 2022 – determined by the length of the available Frontex series on irregular border crossings. Due to low levels of Syrian asylum applications in 2007 and 2008, we do not believe there would be a significant number of irregular border crossings in those years. Due to the availability of data, we were not able to use all possibly helpful macroeconomic series: for instance, the exchange rate reports for Syria cease in 2018, and inflation is poorly reported to create a consistent series.[31] Food price data are not fully available pre-2011 for Syria (as well as for other countries, such as Afghanistan). Therefore, in our model, we focus on GDELT, Google Trends and US trade data.

In this section, we present three early warning models, each with slightly different response and explanatory variables. For each model, there are three threshold levels that use different criteria to determine whether a warning is triggered, as defined in Candelon et al. (2012) and used in the EWS R Package (Hasse and Lajaunie, 2021). These are discussed in Section 3.1, with each model using a different method to determine the optimal cut-off. As mentioned before, the NSR threshold is typically the highest and as such it is less likely to identify a crisis, whilst the lower AM and CSA criteria are more likely to flag a crisis when there is none.

Model 1

In our first model, the binary response variable used is a minimum 50% growth rate in the current period, with a requirement for at least one 50% growth rate in the previous 12 months, and the current period asylum applications exceeding 2 standard deviations from the previous 12-month rolling sample. The vector of explanatory variables is given by equation (4.1), and includes: the change in asylum applications over the last 12 months, $AsyApp_{t-12}$, the change in average tone of all events in Syria (GDELT Event Codes 01-20), $SyriaERCAll_{t-12}^{AvgTone}$, Frontex crossings, $Frontex_t$, US Net exports to Syria, $USNX_t$, number of protests in Syria (GDELT Code 14), $SyriaERC14_t^{Count}$, and Google Trends change in 12 months related to searches for 'Asylum' in Arabic, $GT_{t-12}^{Asylum(A)}$. The left

[31] There are price indices available from the Syrian national statistics office, however, they change the base year throughout the sample required.

axis shows the estimated probability of the threshold being exceeded, while the right axis measures the monthly asylum applications (in 1000s).

$$X_t = \left[AsyApp_{t-12}, SyriaERCAll_{t-12}^{AvgTone}, Frontex_t \right.$$
$$\left. USNX_t, SyriaERC14_t^{Count}, GT_{t-12}^{Asylum(A)}\right] \quad (4.1)$$

In our example, for Model 1, the calculated values of the different threshold criteria are: AM = 0.3007, CSA = 0.446 and NSR = 0.9917, with a model R^2 of 0.679.

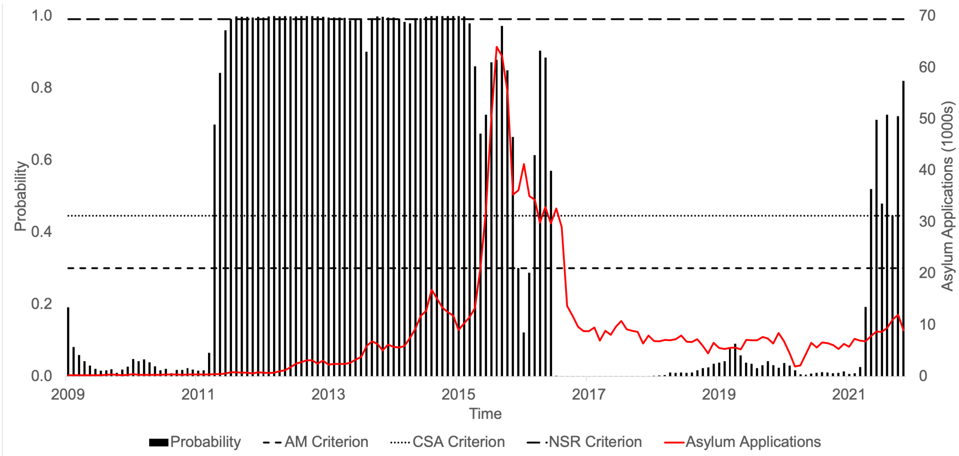


Figure 14: EWS model for Syria (Model 1)
Source: Eurostat (asylum data); own elaboration

Model 2

In the second model, the same binary response variable is used, with the vector of explanatory variables given by equation (4.2) – in comparison with the previous model, it has been additionally expanded to include Google Trends searches for 'Mafraq' in Arabic for the current period, $GT_t^{Mafraq(A)}$. In this case, the calculated values of the threshold criteria are: AM = 0.1603, CSA = 0.3496 and NSR = 0.9429, with R^2 of 0.738.

$$X_t = \left[AsyApp_{t-12}, SyriaERCAll_{t-12}^{AvgTone}, Frontex_t \right.$$
$$\left. USNX_t, SyriaERC14_t^{Count}, GT_{t-12}^{Asylum(A)}, GT_t^{Mafraq(A)}\right] \quad (4.2)$$

As we already saw in the raw Google Trends data, the patterns for different indicators, albeit similar, can provide different modelling results – in our example 'smuggling' provided marginally more information (as measured by the R^2) than 'smuggler', and so did 'Mafraq' relative to 'Lebanon', with 'Tell Abyad' marginally more informative in Arabic than in English. 'Reunification', given hardly any presence prior to 2015, is characterised – unsurprisingly – by relatively poor explanatory power. Finally, 'emigration' searches are more informative than 'immigration', with 'asylum' more informative then either.

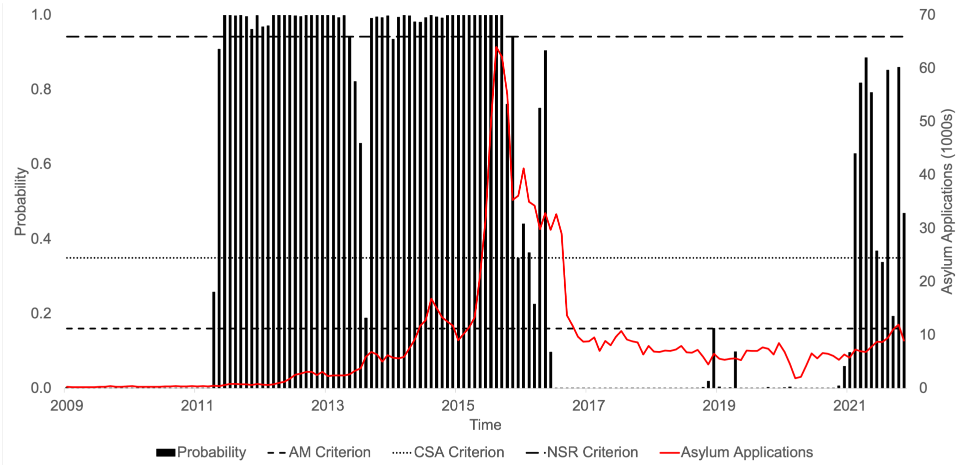


Figure 15: EWS model for Syria (Model 2)
Source: Eurostat (asylum data); own elaboration

Model 2 yields a slightly lower NSR, allowing for more periods in which the criterion can exceed the NSR limit, although the warnings are triggered in similar periods, largely irrespectively of the criterion, indicating a high likelihood of a crisis occurring. The model again provides warnings from 2011 onwards, which is in advance of the observed increase in asylum applications. There is a brief time in 2013 in which the crisis probability decreases, but then it rises significantly again in advance of the events of 2014 and 2015.

One problem for Syria is a prolonged increase in the number of asylum applications, which Models 1 and 2 struggle to fully acknowledge. From 2011, except for a few months, there is hardly any decrease in the observed numbers. In the seasonally-adjusted numbers of asylum applications, there is an increase in the 12-month differences between March 2011 and July 2016.Similarly, for nine-month differences, increases in the numbers of applications can be observed for the whole period from December 2010 to April 2016, and for six-month differences, from December 2010 to February 2016, whereas the three-month differences display only six negative changes in the same period. These patterns are similar, with a few minor exceptions, in the data that are not seasonally adjusted.

Model 3

To account for the consistently high crisis probability levels, in our third model, the response variable is stricter, with a minimum 50% growth rate in the current period and a requirement for at least one 100% growth rate in the previous 12 months, and the current period asylum applications above 2 standard deviations from the previous 12-month rolling sample. The vector of explanatory variables is the same as in Model 2, except for using the past asylum applications and other lagged explanatory variables from three, not 12 months before. The equation for Model 3 is given by (4.3).

$$X_t = \big[AsyApp_{t-3}, SyriaERCAll_{t-3}^{AvgTone}, Frontex_t \\ USNX_t, SyriaERC14_t^{Count}, GT_{t-3}^{Asylum(A)}, GT_t^{Mafraq(A)}\big] \quad (4.3)$$

The estimated threshold criteria for this model are AM = CSA = 0.0422, and NSR = 0.9939, with R^2 of 0.426, as shown in Figure 16, with the estimated probabilities generally smaller than 0.04 (with a few exceptions). The main reason for the very low values of AM and CSA is that there are now fewer crisis periods, and in this model, the pre-crisis triggers are much clearer to identify. There are groups of alerts given in the mid to end 2011, late 2012, later 2014, and mid-2015. Before the pandemic, only single months exceed the lower thresholds. The decrease in applications due to COVID lockdowns, and the subsequent increase, impact the definition of a crisis is somewhat; nevertheless, there are again groups of alerts that precede the increase in applications to above the 2019 levels. These warnings can provide researchers and policy makes with a large window to prepare for large increases, in contrast to the previous models which were too sensitive.

The models presented in this section, in particular the contrasting number of alarms triggered by Models 1 and 2 relative to Model 3, highlight the issue discussed in Section 3.3. The definition of a crisis varies: at a macro level, this may be a significant change in the number of asylum applications which Model 3 is most suitable for. At the operational level, related to dealing with the practicalities of handling asylum applications, Models 1 and 2 may be better suited. Different models show that since the early months of 2011, there were signals of a potential surge in asylum applications which took another 12–18 months to fully materialise. Near-constant probabilities exceeding the AM and CSA

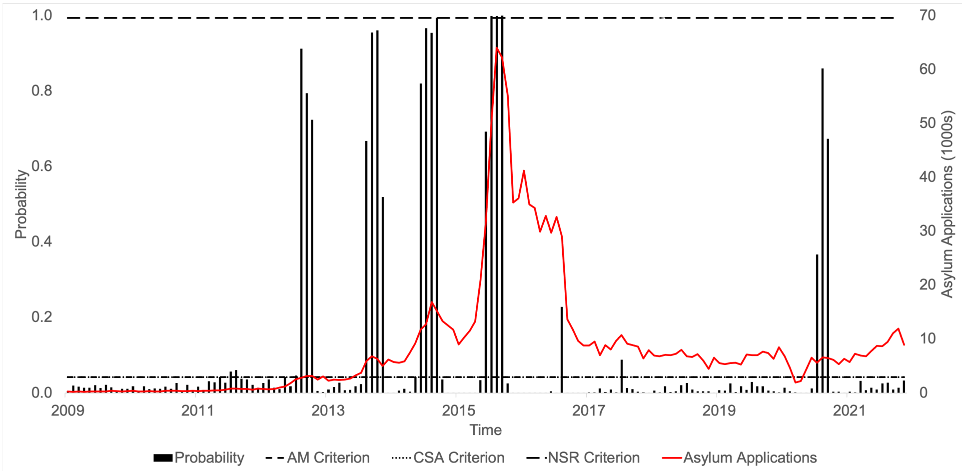


Figure 16: EWS model for Syria (Model 3)
Source: Eurostat (asylum data); own elaboration

thresholds provide a signal that the numbers of asylum applications remain high. In particular, Model 2, with the highest R^2 of the three, identifies a large number of periods for which the NSR criterion is significantly exceeded. One downside of Model 3 is that the NSR criterion is close to 1 with AM and CSA being relatively low, generating the need for substantive user input to interpret the differences between different thresholds.

Syria: Is there another migration crisis coming?

In each set of results, particularly Model 3 (Figure 16), we can see periods of concern, indicated by higher probabilities, preceding the large increase in migration flows. Still, the models featured in this section may suffer from some of the issues found by Filippopoulou et al. (2020), such as the inclusion of post-crisis data. Examining similar models in future research remains a possibility, also in dealing with e.g. COVID-19 pandemic data.

In particular, the increases in asylum applications towards the end of 2021, showed levels not seen since 2016. Of course, one consequence of the COVID-19 pandemic is that some asylum applications are delayed. However, the alerts given by other indicators suggest that at least some of this increase may be sustained. Still, in comparison with the magnitude of the Ukraine crisis caused by the Russian invasion in 2022, these increases in the numbers of new Syrian asylum applications would be much easier manageable, even if the legal treatment of these two groups of asylum seekers is different, as discussed above. The Ukraine case study in its own right is discussed in the next section.

4.2 Ukraine 2013–2022

The second case study presented in this report concerns Ukraine, with focus on two periods: the run-up to 2022, with its mass-scale population displacement caused by the war and Russian invasion, but also to the longer-term buildup of the crisis since 2013–14.

4.2.1 Historical Background and Data

The current stage of the war in Ukraine, which started with the full-scale Russian invasion on 24 February 2022, can be linked to the the Russian annexation of Crimea in 2014, and other political events that preceded it. As summarised by Zelinska (2017), the key original trigger was the pro-European *Euromaidan* protests, which started on 21 November 2013, after the government, headed by the-then president Viktor Yanukovych, did not sign the Association Agreement with the EU as a result of Russian pressure.

What started out as civil protests, escalated into wider political instability: on 22 February 2014, Yanukovych was ousted as president of Ukraine and has since lived in exile in Russia. The instability that followed granted Russia an opportunity to send troops into Crimea, which was illegally annexed to Russia on 18 March 2014 (AFP, 2015). In the intervening period since, there has been intermittent fighting in the eastern Ukrainian regions of Donetsk and Luhansk. The presence of a low-intensity conflict in these areas was a constant feature of the period 2014–2021.

After the full-scale Russian invasion on 24 February 2022, the UNHCR estimated that as of 15 June 2022, some 7.5 million Ukrainians crossed the borders into neighbouring countries, of whom around 4 million entered into Poland alone – while in the same period 2.5 million have crossed the border back into Ukraine[32]. One important issue characterising the recent situation, in contrast to the Syrian 'crisis' of 2015–16, is a different legal framework, under which Ukrainian nationals can access temporary protection in the EU, as discussed in Section 1, but also the welcoming nature of Central and Eastern European (CEE) countries. A more co-operative stance among EU members has enabled an effective policy – by implementing the provisions of the Temporary Protection Directive – that otherwise, due to the scale, would make the inflows incredibly hard to manage.

[32]Source: UNHCR, Ukraine Situation Report, https://data2.unhcr.org/en/situations/ukraine, accessed on 15 June 2022.

The legal ramifications have also impact on the available data, as discussed in Section 3.2, with numbers of asylum applications only available for the period before the 2022 Russian invasion. For the earlier period of the conflict, the numbers of asylum applications from Ukraine in EU+ countries are shown in Figures 2b and 3b in Section 3.3, with visible changes in November 2013–February 2014 following the start of the Euromaidan protests and the annexation of Crimea. The increase in asylum applications continued following the worsening conflict in eastern Ukraine through 2014 and 2015. One standard deviation for this period corresponds to approximately 550 persons per month.

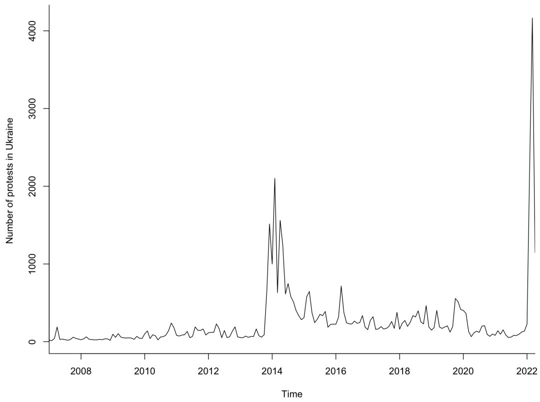
4.2.1.1 The Ukraine crisis in the light of the GDELT data

As mentioned before, GDELT is a large resource of *global* events, but an important caveat is that the indicators that might have worked for Syria, do not necessarily have to work as well for Ukraine, although there have been studies that looked at both at the same time – notably Leetaru (2014), who examined the protest intensity for the Arab Spring and in Ukraine around 2014. The numbers of protests and other relevant political events in Ukraine recorded in the GDELT database are shown in Figure 17, both in terms of their absolute values, and their logs, to help detect changes in the protest intensity.[33]

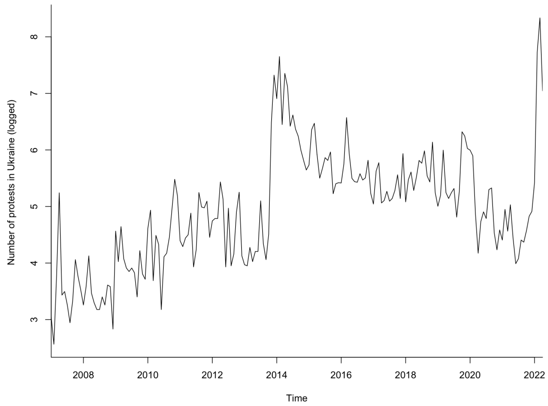
In the context of Ukraine, analysing protests and number of other political events *separately* is particularly important. In Figure 18, there is a visible spike in the relative protest *intensity* for the 2013–14 crisis. There is a also large spike in the *number* of protests, as seen in Figure 17a. Still, the 2021–22 crisis does not have the same distinction due to the protests being relatively dwarfed by the number of other political events. For this reason, the protest intensity is not a variable that we can use reasonably use to provide signal in an early warning model.

Similarly to the Syrian example, an alternative approach relies on looking at the *average tone* of reporting of all political events. Figures 19a and 19b show the average tone of all events and average tone of coverage of all protests in Ukraine, with Figure 19b overlaid with the numbers of asylum applications in the EU+ countries. Figure 19a highlights the key crisis periods of November 2013–February 2014, and November 2021–April 2022 with grey shading. However, even with reasonable lags, there does not appear

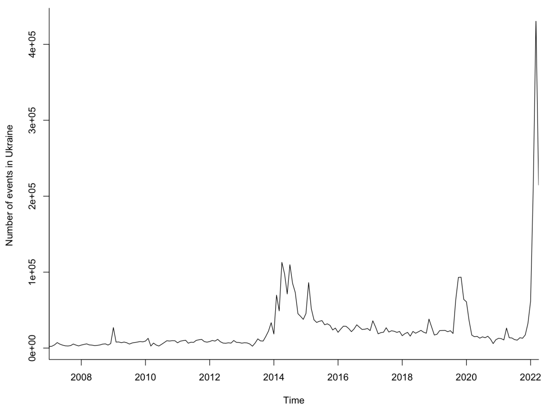
[33]Note that no GDELT are available for 23-25 January 2014 and 19 March 2014.



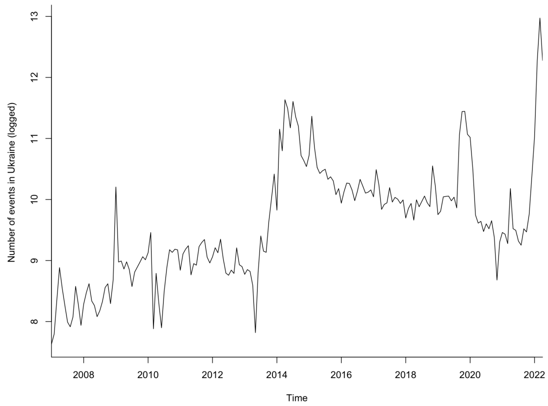
(a) Number of protests in Ukraine



(b) Number of protests in Ukraine (logged)



(c) Number of political events in Ukraine



(d) Number of political events in Ukraine (logged)

Figure 17: Number of protests and political events in Ukraine

The first two figures show the number of protests in Ukraine for 2007:01-2022:04 (up to 26 April 2022), the latter on the log scale. The second two figures show the number of all political events in Ukraine for 2007:01-2022:04 (up to 26 April 2022), the latter on the log scale. Protests are filtered by using GDELT codes $EventRootCode ==$ "14", and the location by $ActionGeo\_CountryCode ==$ "UP"

to be a significant drop in tone to coincide with the events that trigger an increase in asylum applications represented in this data series.

One possible explanation here is that the average tone is not well estimated relatively to the political scale of events, or there are a series of distortions in the series. The two significant drops are in November 2012 and February 2015, which do not have obvious interpretation. In these cases, the changes in the average tone might be artefacts of the way the data were compiled or the tone indicators constructed. If these changes were real (despite their difficult to explain magnitude), some tentative interpretations might include a general election on 28 October 2012, the results of which were disputed, with allegations of fraud and international observers declaring it a backwards step for

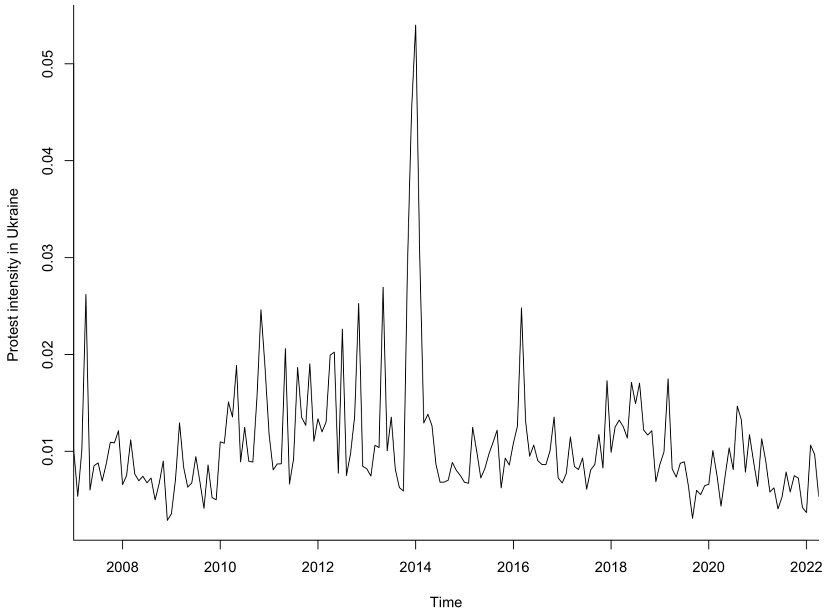


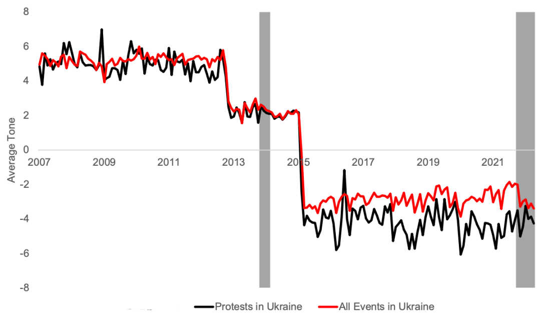
Figure 18: Protest intensity in Ukraine

The protest intensity in Ukraine using events reported on the GDELT database 2007:01-2022:04. The protest intensity is calculated as the sum of the total number of protests in each month, divided by the total number of political events. These events are identified using GDELT codes $ActionGeo\_CountryCode == "UP"$, with protests filtered using $EventRootCode == "14"$

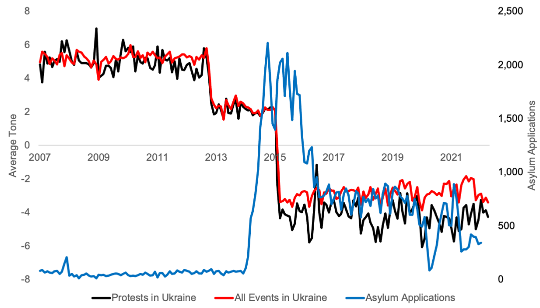
democracy in Ukraine (BBC News, 2012). Similarly, the change in 2015 might coincide with the escalation in conflict, and the attempt at a ceasefire and peace talks with the new Minsk Agreement (BBC News, 2015). However, even assuming that the change itself is real, we see from the overlap with the asylum applications that the change in tone does not reflect the increase in asylum applications. Still, the magnitude of shifts in the average tone *outside* of the periods of the highest intensity of conflict and tensions in Ukraine gives reasons to be cautious about this indicator.

4.2.1.2 The Ukraine crisis in the light of the macroeconomic data

There are a number of possible *leading* macroeconomic indicators that can signal upcoming crises weeks or months ahead, in particular, exchange rates, as discussed in Section 3. It is long believed that exchange rates can be approximated by a random walk process or, in forecasting terms, it is rare that another model would be able able to perform better than a random walk in terms of its predictive capacity (Kilian and Taylor, 2003). This



(a) Average tone with crises highlighted



(b) Average tone overlayed with asylum applications

Figure 19: The average tone of the article for protests and events in Ukraine

The figures show the average tone of reporting on all events located in Ukraine, and on protest events in Ukraine. The black line identifies the average tone for protest events in Ukraine, while the red line identifies all events in Ukraine. The first figure has vertical bars highlighting November 2013–February 2014, and November 2021–April 2022. The right hand axis on figure (b) identifies the number of asylum applications, as shown by the blue line. Protests are filtered by using GDELT codes $EventRootCode ==$ "14", and the location by $ActionGeo\_CountryCode ==$ "UP"

is partially due to the sensitivity of exchange rates to what is happening in the news[34].

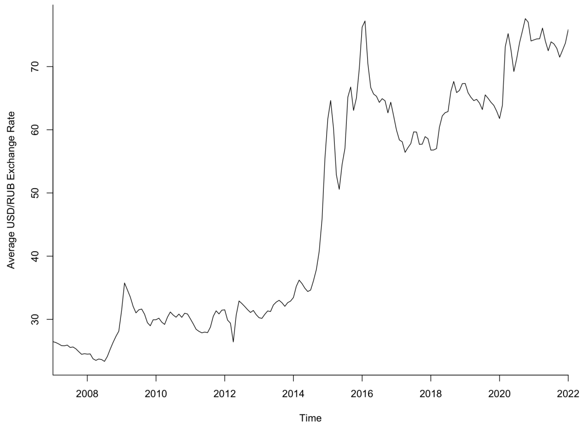
The exchange rates of some countries are fixed (or *pegged*), quite often to the USD, as the Ukrainian hryvnia was before becoming a managed float. Following any type of fixed exchange rate can lead to large increases in inflation – these often cause re-evaluations of exchange rates. To counter this, we also analyse the real effective exchange rate based on the CPI index, which is not affected by pegging or interest rates.

Figures 20a and 20b present the exchange rates expressed as the amount of USD per Russian rouble (RUR) and Ukrainian hryvnia (UAH)[35] An increase in the value, represents an appreciation of the USD, or a devaluation of the RUR or UAH. The recent history of the Ukrainian hryvnia is outside the scope of the paper, however, we can see the fixed exchange rate levels that changed in 2009, 2014, and 2015.

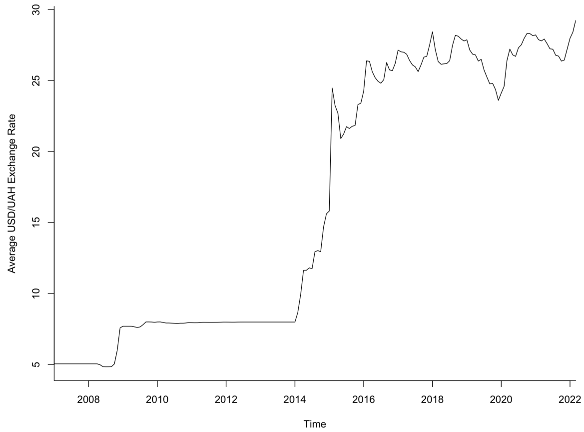
The real effective exchange rate (REER), presented here as an index, is defined by the IMF as "... a measure of the value of a currency against a weighted average of several foreign currencies) divided by a price deflator or index of costs". An increase in the REER thus indicates a decline in trade competitiveness. The REER is shown for Russia in Figure 20c, and for Ukraine in 20d. It is important to note, that the Russian economy

[34]In the UK Brexit referendum, the exit poll suggested a Remain win, strengthening the pound, but when the results started tilting towards Leave, the pound dramatically weakened (BBC News, 2016).

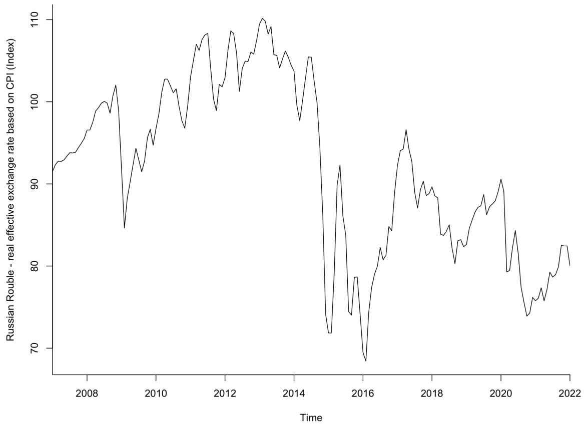
[35]Note that, as of 26 April 2022, the IMF data for the Russian Rouble stopped in January 2022, whilst for Ukraine the rates were updated to March 2022.



(a) RUR nominal exchange rate to USD



(b) UAH nominal exchange rate to USD



(c) RUR real effective exchange rate (index)

(d) UAH real effective exchange rate (index)

Figure 20: Nominal and real effective exchange rates

The figures show the nominal USD exchange rate (period average) and real effective exchange rate based on CPI (index), data code $EREER\_IX$ sourced from the IMF eLibrary, International Financial Statistics Database.

is strongly related to commodity prices, in particular fossil fuels, whilst Ukraine is a large exporter of grains, as well as iron and steel.

Inflation rates can be also strongly impacted by exchange rate pressures, particularly for pegged rates. To illustrate this, Figure 21 shows the inflation rate of food and non-alcoholic drinks, based on the nominal UAH to USD exchange rate, an increase of which would indicate a depreciation of UAH. Domestic and international pressures, the effect of the financial crisis, and inflationary spikes led to re-evaluations respectively in 2008, 2009, 2014, and 2015 before the peg was finally abandoned.[36]

Inflation is a promising alternative to the USD exchange rate as it is not fixed, and

[36] An example of a western European country leaving an exchange rate peg due to inflationary pressure is the UK exiting the European Exchange Rate Mechanism (ERM) in 1992.

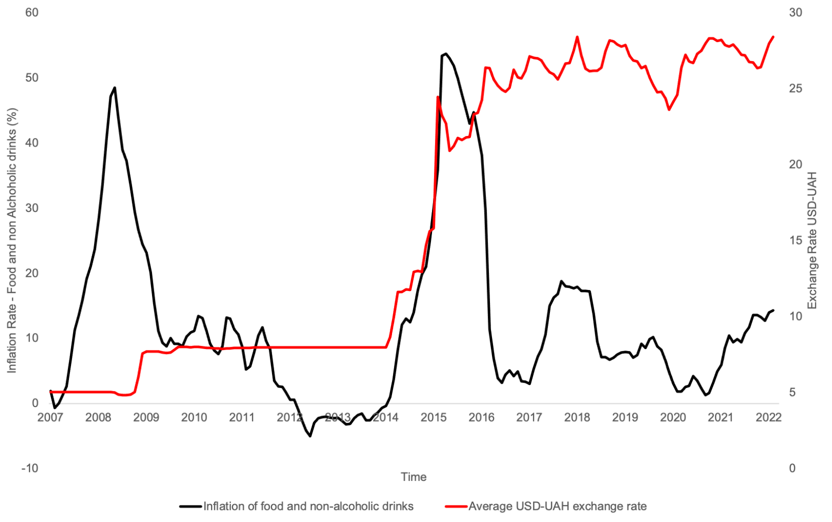
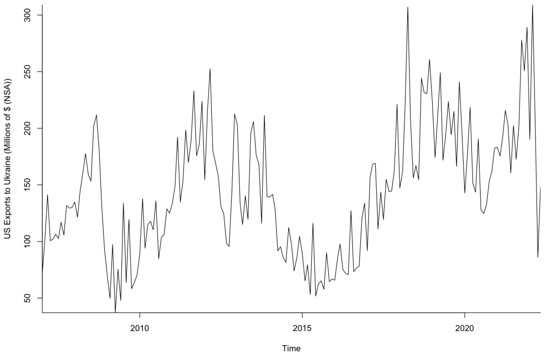


Figure 21: Inflation rate in Ukraine of food and non-alcoholic drinks

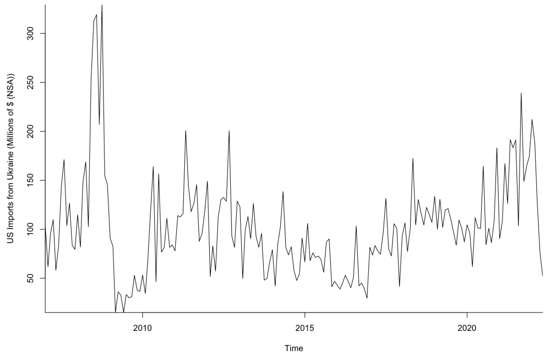
The inflation rate of food and non-alcoholic drinks is measured with the left-hand axis and denoted by the black line. The nominal average for the month exchange rate of UAH with the USD is shown in red and uses the right-hand axis. Source: State Statistics Service of Ukraine

takes into account price changes over a year. In Ukraine, the Euromaidan protests were preceded by a period of disinflation, followed by deflation. The latter is a particular problem because it leads to slowing down of the economy: wages fall, and people postpone spending as they believe goods will be cheaper, hence consumer confidence falls.

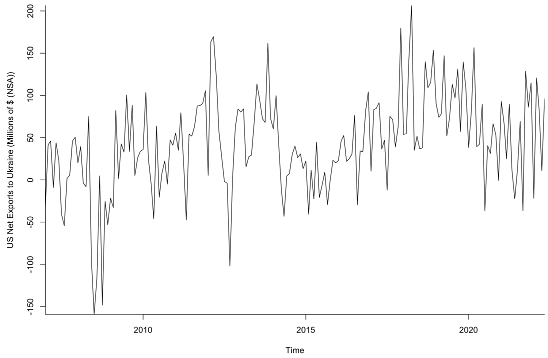
As for the trade balance, the decline in trade concerning Ukraine was not externally imposed, as was the case with the US ban on trade with Syria, but rather the political (and consequentially economic) situation in Ukraine has dictated the change in trade flows. The US has been shifting between a trade surplus and trade deficit with Ukraine, with a general trend towards a trade surplus as shown in Figure 22c. Since 2007, net US exports have remained in the range from –159.3 to +206.4 million US$. Following the start of both phases of the conflict, there was a general decline in trade both for imports (Figure 22b) and exports (Figure 22a).



(a) US exports to Ukraine



(b) US imports from Ukraine

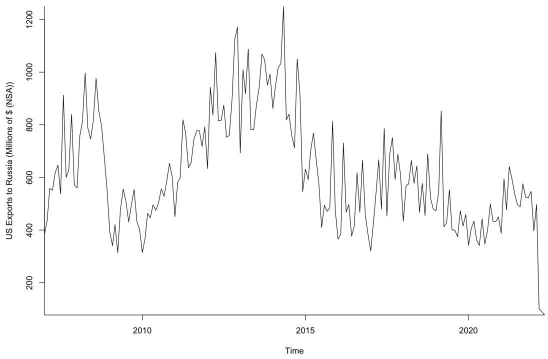


(c) US net exports to Ukraine

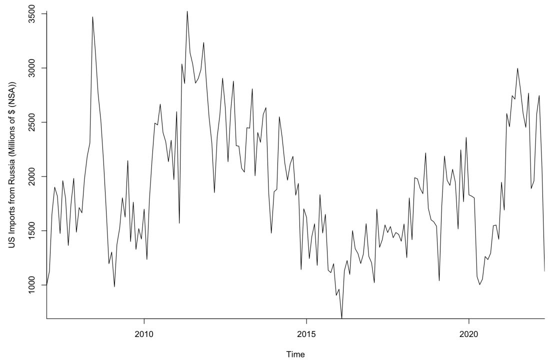
Figure 22: US trade with Ukraine

The figures show the US exports, imports and net exports of goods to/from Ukraine at a monthly frequency. The figure is given in millions of US$ not seasonally adjusted (NSA). Source: U.S. Bureau of Economic Analysis (via FRED: Federal Reserve Bank of St. Louis). Codes EXP4623, IMP4623, and own calculations for net exports.

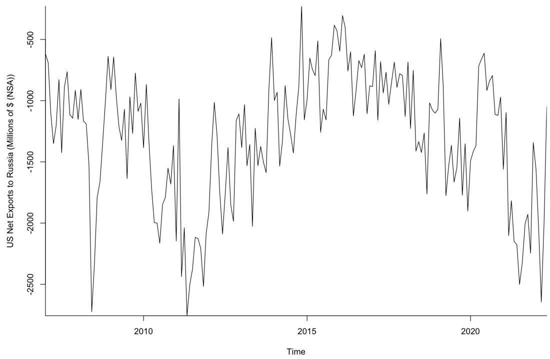
At the same time, for Russia, recent figures from the US Bureau of Economic Analysis show a significant decline in trade since the invasion of Ukraine, due in part to the sanctions imposed by the West, as shown in the trade components in Figure 23. There was also a decrease in trade flows after the 2014 annexation of Crimea but not to the same extent. The US has consistently run a large trade deficit with Russia, with net *exports* in the range of –228.6 million to –2.757 billion US$ in a month since 2007. In summary, the changes in trade flows can indicate issues or conflicts that arise within an economy, or in this case, are imposed by external forces.



(a) US exports to Russia



(b) US imports from Russia



(c) US net exports to Russia

Figure 23: US trade with Russia

The figures show the US exports, imports and net exports of goods to/from Russia at a monthly frequency 2007:01 - 2022:05. The figure is given in millions of US$ not seasonally adjusted (NSA). Source: U.S. Bureau of Economic Analysis (via FRED: Federal Reserve Bank of St. Louis). Codes EXP4621, IMP4621, and own calculations for net exports.

4.2.2 Early Warning Modelling Results

The data used in modelling cover the same period as in the Syrian case study, January 2009–March 2022. The particularly important dates we are looking at in Ukraine for the first phase of the crisis are late 2012 onwards, with disputed elections in 2012, Euromaidan protests in 2013, and the Russian annexation of Crimea in 2014, with particular focus on 2013–14. For the most recent crisis, we focus on the period April–December 2021, with an increase in the number of significant political events presaging the 2022 invasion, including the deployment of Russian forces close to the border.

Since the average GDELT tone indicators for events *in* Ukraine do not visually appear to provide strong signals (see earlier in Section 4.2), and can be linked to possible data problems, as discussed above, we additionally employ constraints on the "actors" involved, as enabled by the GDELT database. In our case, for modelling we define two such actors: Actor 1, who is causing the event (Russia, or Russian actors, such as Vladimir Putin), and Actor 2, the receiver of the event (Ukraine). As before, we estimate three models.

Model 1

We begin our estimation with the model based on 6-month growth rates in asylum applications, with the binary response variable based on the growth rates of at least 20% in the current period, over 25% in the last 12 months, and asylum applications being above two standard deviations in a 12-month rolling period, with the minimum value of 50 in the last 3 months. We use a vector of explanatory variables X_t, listed in equation (4.4).

$$X_t = \left[AsyApp^{UKR}_{t-(t-3)}, Frontrex_t, COVID_t, \pi^{UKR}_t, A1RGA2U^{NegGS}_{t-(t-12)}\right]' \tag{4.4}$$

This vector consists of: asylum applications, $AsyApp$, Frontex-recorded border crossings, $Frontex$, a $COVID$ indicator, inflation rate in Ukraine for food and drink, π^{UKR}_t, and 12-month change in the number of negative Goldstein Scale events (Actor 1: Russian Government, Actor 2: Ukraine), $A1RGMA2U^{NegGS}_{t-12}$, with 3 lags[37]. The results for Model 1 are shown in Figure 24. The right axis shows the number of asylum applications by Ukrainian citizens up to February 2022, for the reasons related to change of the measurement instrument from asylum applications to border crossings and registrations.

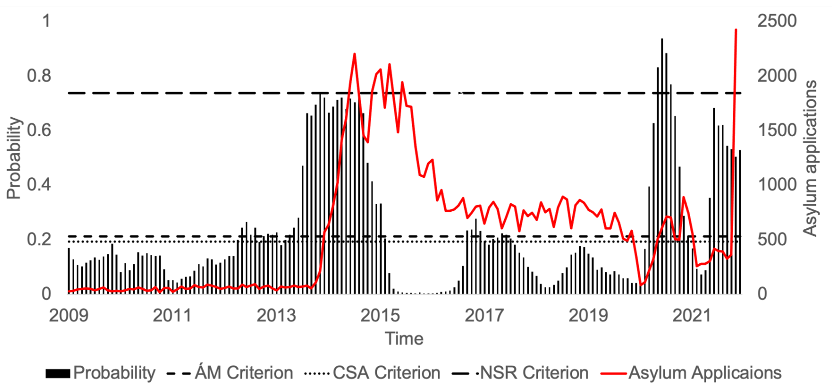


Figure 24: EWS model for Ukraine (Model 1)
Source: Eurostat (asylum data); own elaboration

[37]As discussed in Section 3.2, the Goldstein Scale gives each event a number in the range of –10 to +10 which identifies how an event could impact the stability of a country. Unlike the average tone, the Goldstein score does not put the events in context, so the event magnitudes are not shown in this figure.

The estimated threshold criteria for Model 1 are: AM$=0.2131$, CSA $=0.1936$, and NSR $=0.7388$, with $R^2 = 0.1290$. In comparison to the models presented for Syria, such a low R^2 score indicates much poorer model fit. From a visual perspective, the model does first exceed the first AM and CSA thresholds in August and July 2012 respectively. This was after Ukraine jointly hosted the European football championships with Poland, and in the lead-up to the disputed elections. The probability remains around similar levels of these criteria until October 2013, when it doubles, and by February 2014, the NSR criterion is nearly met (with the estimated probability of 0.7386 vs NSR of 0.7388). This high crisis probability is sustained until 2015, followed by a period of low probability. More recently, the probability has risen again, coinciding with the escalation of the political situation and tensions, in June to December 2020, and again since August 2021. These trends can provide indicators for in advance of the increase in asylum applications.

Model 2

In Model 2, the binary response variable used is bases on 6-month growth rate in asylum applications of at least 20% in the current period, exceeding 25% in the last 12 months, the number of applications above two standard deviations from a 12-month rolling period, and the minimum value of 50 in the last six months. The input variables are the same ones as in Model 1, additionally including US Net Exports to Ukraine with a 12-month difference ($USNX_{t-12}$), as shown in equation (4.5). The model results are shown in Figure 25. The right axis shows the number of asylum applications until February 2022 (as collected on 18 May 2022). For this model, R^2 is still low, at, 0.214, and the estimated warning thresholds are as follows: AM $=0.3267$, CSA $=0.2475$, and NSR $=0.7322$.

$$X_t = \left[AsyApp_{t-3}^{UKR}, Frontrex_t, COVID_t, \pi_t^{UKR}, A1RGMA2U_t^{NegGS}, USNX_{t(t-12)}\right]' \tag{4.5}$$

As shown in Figure 25, the crisis probabilities are higher than in the previous model, though show some similar patterns. There are a number of spikes higher than in Model 1 in 2012, though the first time that the NSR criterion is exceeded is April 2014. The addition of difference in US net exports to Ukraine is key to these differences between two models, as seen also in the raw trade statistics (e.g. in Figure 22).

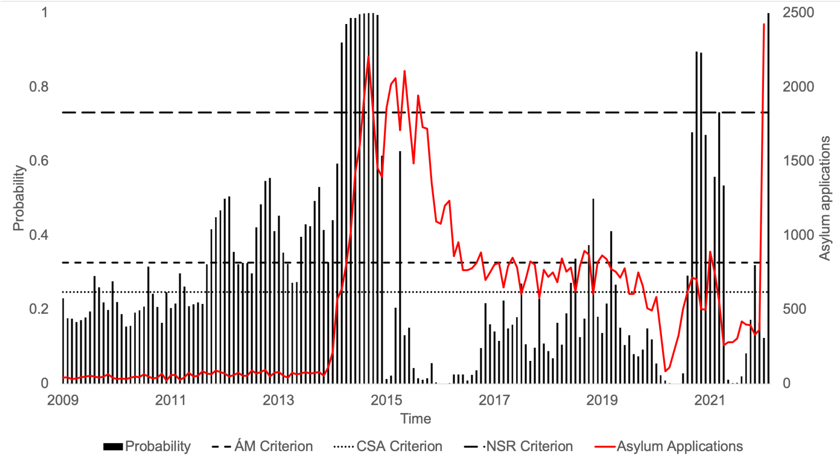


Figure 25: EWS model for Ukraine (Model 2)
Source: Eurostat (asylum data); own elaboration

Model 3

In Model 3, the binary response variable is based on 6-month growth rates in asylum applications exceeding 20% in the current period, with at least 25% in the last 12 months, and the threshold of two standard deviations. The vector of explanatory variables includes the same variables as in Model 2, with an addition of real exchange rates (3 month difference, $UKR - RealXR_{t-3}$), see equation (4.6)). The model results are shown in Figure 26, with the number of asylum applications shown on the right axis. The estimated threshold criteria are: AM $= 0.2882$, CSA $= 0.2476$, and NSR $= 0.8439$, with $R^2 = 0.18$.

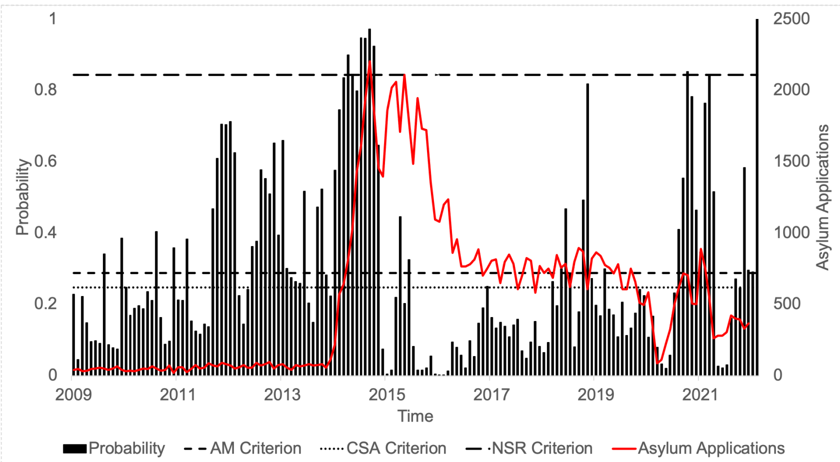


Figure 26: EWS model for Ukraine (Model 3)
Source: Eurostat (asylum data); own elaboration

$$X_t = \big[AsyApp^{UKR}_{t-3}, Frontrex_t, \pi^{UKR}_t, A1RGMA2U^{NegGS}_{t-3}, A1RGMA2U^{AvgTone}_{t-3},$$
$$USNX_{t-(t-3)}, UKR - RealXR_{t-3}\big]' \quad (4.6)$$

Across the three models for Ukraine, there are visible early warning triggers in 2012, 2013, and 2014, which get near or exceed the NSR thresholds. There are also a few ones in later years, with visible significant increases in the estimated crisis probability in 2020 and 2021. Unlike for Syria, there are clear periods of relative calm, but the models accept that there has been a ‘new’ steady state reached following the highs of 2014–16. One of the key differences between the models for the two countries is that to identify the crisis, and make it useful to researchers and policy makers, the Syrian situation would require a higher change in flows for the growth rates to trigger an early warning than for Ukraine, as discussed in Section 3.3. This is the main reason, why in the Ukrainian examples it was more desirable to look at models with relatively sensitive response variables. In any case, this also confirms the dependence of the model and the results on the context, and the need for tailoring the model accordingly – in our case, this was done by the analysts, but in future versions of the models, some aspects of the process could be automated.

A generalised conclusion, as in the Syrian case study, is that the models generally fail to fully predict the scale of the crisis. Still, there is some signals in the data: a weak indicator is that the probability of a crisis, as identified by the bars in the figures, increases sharply. This is where a human factor is required to interpret the results. Future research can look at improving the inference about the scale of the possible crisis. At the same time, it is notable that – especially under Model 3 – there are some strong early warning indicators present in 2013, as well as in Spring 2021, late 2021 and early 2022. One issue we came across is that the R^2 value is not a helpful measure in finding the most useful model. For instance, Model 3 has a higher R^2 than Model 1 but the difference is due to varying explanatory power between the models. This strengthens the case for future work on the methodology for formal model choice, as suggested in Section 3.3.

5 Discussion and Conclusions[38]

In this report, we have presented an early warning modelling system that managed to successfully trigger some of the expected warnings for the two case studies under observation, related to the number of Syrian and Ukrainian asylum applications from the last decade. The models for these differing crises used varied datasets, including some 'big data' indicators, with modelling techniques adopted largely from finance and macroeconomics to generate warnings with a lead-time of up to six months before the increases in the numbers of applications could be observed.

One important question that was raised in this research, which requires a follow-up in future investigations is, *what is a migration crisis*? Macroeconomics has a simple definition of a recession with two consecutive quarters of economic contraction. Having an equivalent operational definition for migration would be important, although it needs to be recognised that different users (and uses) of early warning models may have different requirements in that regard. However, without formally defining and justifying what a 'crisis' is in substantive terms – or at least establishing a foundation from which such a definition can be built, such as the one proposed in this study – analysts might be trying to solve a problem that is not necessarily meaningful for the users.

At the same time, defining what *is* meaningful – and for whom – remains a challenge. Different users may require information not only about asylum applications, but also, for example, such politically salient topics as irregular border crossings. Despite being related in some contexts, especially where legal channels of asylum migration barely exist, the underlying *processes* are empathically not the same. Bridging the gap between the available data that can be used for modelling, as in the case of asylum applications, and the underlying processes, is far from trivial, and raises additional questions, both scientific and ethical, especially if the line between adequate preparedness for possible humanitarian catastrophes and politicisation of migration becomes blurred.

As an example of potential ethical challenges, a possibility of malicious use of early warning tools by hostile actors, possibly in order to generate or strengthen a crisis, or to

[38]We are very grateful to Marta Bivand Erdal, Teddy Wilkin and Rainer Münz for their suggestions for interpretations of some of the results, related caveats, and possible directions of future work. Needless to say, all the responsibility for the content of this section remains ours.

sabotage a humanitarian response, cannot be excluded. With such actors, one possible response would be to employ strategic tools of risk management, such as for example Adversarial Risk Analysis (Banks et al., 2015, 2022). Similarly, using the tools by authorities of the receiving countries for political gains may not be ethical or legal, especially from the human rights perspective, such as forcible 'pushback' actions[39]. There is no perfect safeguard here, although transparency in the use of models as decision aiding tools can go some way towards managing this challenge, even if not eliminating it completely.

There are also several other caveats related to defining and operationalising the crisis indicators. Using a static definition for countries such as Syria before 2011 would not tell us much on what might be coming in the near future. Growth rates are not a perfect substitute, either, due to their volatility and seasonality. Using 12-month growth rates reduces the impact of seasonality, but can miss some of the seasonal events that can affect the flows. In this report, we proposed a combined solution, looking at growth rates in addition to a standard deviation threshold calculated on a 12-month rolling basis. In our examples, this helped reduce anomalies, although the usual caveat applies: no measure is perfect, and each one needs to come with a clear definition and health warnings. More generally, an indicator of a 'crisis' could be a composite measure, involving for example illegal border crossing, asylum applications as well as data on various permits, thus encompassing different indicators that may be relevant for the decision makers.

Having defined the intended use of an early warning model, and having established a question for the model to answer, as well as a matching indicator of a 'crisis' event, the next step is to assemble a vector of explanatory variables potentially carrying sufficient signal for early warnings. In this research we have looked at the causes and drivers of migration using an array of various factors and corresponding variables. In our examples, the data came from a range of traditional (economic, geopolitical) and 'big data' sources. The model analysis presented in this report has demonstrated that no single model can be useful in every context, with different variables being preferable for different applications, situations and countries. While macroeconomic data might not be the first choice for migration scholars, there are also important insights that can be learned from them.

[39]Pushback can be defined as "state measures aimed at forcing refugees and migrants out of their territory while obstructing access to applicable legal and procedural frameworks", according to the European Center for Constitutional and Human Rights (https://www.ecchr.eu/en/glossary/push-back/).

Exchange rates, for example, are quick to react to news. Even though fixed exchange rates can be an issue, there are measures that allow analysts to see past those, in particular looking at real effective exchange rates. On the news themselves, 'big data' sources such as GDELT provide an excellent, large-scale resource that allows investigating specific actors and circumstances, such as in Ukraine and its relationship with Russia. Google Trends introduce another aspect – Internet searches – however, due to differences in internet access in origin countries, and language limitations, this may only be useful to a certain extent. Compiling a set of EWS models on a per country basis would provide more accurate results than employing the same model to all situations.

One improvement for future research would involve examining broader classes of econometric methods beyond those presented in the report, as well as trying to identify the best-possible vector of explanatory variables, both from traditional and 'big data' sources. In particular, there are further explanatory variables that could have been useful, which were not fully available. They include, for example, digital or crypto currencies, as has been seen with the increase of purchases by Afghan citizens in 2021 (Chipolina, 2021) – with sanctions being placed on Afghanistan by the US Government, including freezing of assets and delivery of USD, Afghanis turned to cryptocurrency as a way to store money, and with additional cash shortages of the domestic currency (afghani), payments were made in the currencies of neighbouring countries (Bautista-González, 2022)

Based on the findings from the models presented in this study, one particularly promising direction of early warning modelling for asylum and policy support is to build 'model toolboxes', which could flexibly include relevant variables and indicators, which may be different for different countries and contexts. The explanatory variables could start from a mix of available high-frequency macroeconomic indicators (especially exchange rates and trade), 'big data' (such as GDELT, search trends, or mobile phone locators), as well as various indicators of conflict and political (in)stability. At the same time, further work can be done in the area of selecting the warning thresholds – their choice can formally incorporate different costs of false alarms and missed warnings, for example through employing the statistical decision analysis (see, e.g. Bijak and Czaika, 2020). In this way, balancing the costs of overreacting to false positives *versus* underpreparing in the case of false negatives, could be also done openly and transparently.

To increase their predictive potential, early warning models could also look beyond the origin countries, and additionally look at the situation in the whole regions, as well as in transit countries. One important caveat here is that no single model will be even applicable to all possible situations – but a flexible approach can help with automating the task of identifying possible signals of an upcoming crisis, to be later looked into more closely. This is all the more important as early warnings may be even more relevant for other, non-European countries of destination and transit for asylum seekers – especially given that these countries may be much less able to cope with humanitarian crises caused by population displacement than European ones.

To that end, one crucial element of building an EWS model involves desk research on the causes of each of the crises, to identify background and context to find why, and how, these conflicts occurred and escalated. Of course, with hindsight, it is always easy to say that there may have been clear signals at the time. Still, while the same or similar geopolitical conditions persist, these signals can help establish warnings for any future crisis, which the results for Syria and Ukraine were able to show to differing extents. In this research, we have tried to strike a balance between human input and data-driven modelling results. Fully grasping the magnitude of the crisis was not something that could be reliably explained by the model alone. In such applications, only human input would be ultimately able to fully confirm the seriousness of the challenge, whilst remaining cognisant of all the ethical and legal aspects involved in relying on models for helping shape the political or humanitarian responses to the crisis of displacement.

Bibliography

AFP (2015), 'Putin describes secret operation to seize Crimea'. Agence France-Presse (AFP) (via Yahoo News), Accessed: 28 April 2022.
URL: *https://news.yahoo.com/putin-describes-secret-operation-seize-crimea-212858356.html*

Avramescu, A. and Wiśniowski, A. (2021), 'Now-casting Romanian migration into the United Kingdom by using Google Search engine data', *Demographic Research* **45**(40), 1219–1254.

Banks, D., Gallego, V., Naveiro, R. and Ríos Insua, D. (2022), 'Adversarial Risk Analysis: An overview', *WIREs Computational Statistics* **14**(1), e1530.

Banks, D., Ríos Aliaga, J. M. and Ríos Insua, D. (2015), *Adversarial Risk Analysis*, CRC Press, Boca Raton, FL.

Barker, E. R. (2020), The Resource Curse and Migration, Technical report, The University of Sheffield.

Barker, E. R. and Bijak, J. (2021), Uncertainty in Migration Scenarios, QuantMig Project Deliverable D9.2, University of Southampton, Southampton.

Bautista-González, M. A. (2022), 'Revisiting the Cash Shortage in Afghanistan'. CashEssentials, Accessed: 9 March 2022.
URL: *https://cashessentials.org/revisiting-the-cash-shortage-in-afghanistan/*

BBC News (2012), 'Ukraine election 'reversed democracy', OSCE says'. Accessed: 29 April 2022.
URL: *https://www.bbc.co.uk/news/world-europe-20120888*

BBC News (2015), 'Ukraine ceasefire: New Minsk agreement key points'. Accessed: 29 April 2022.
URL: *https://www.bbc.co.uk/news/world-europe-31436513*

BBC News (2016), 'Pound plunges after Leave vote'. Accessed: 28 April 2022.
URL: *https://www.bbc.co.uk/news/business-36611512*

Beirens, H., Maas, S., Petronella, S. and van der Velden, M. (2016), Study on the Temporary Protection Directive. Final report, Technical report, European Commission - Directorate General for Migration and Home Affairs.

Belabbas, S., Bijak, J., Modirrousta-Galian, A. and Nurse, S. (2022), 'From conflict zones to europe: Syrian and afghan refugees' journeys, stories, and strategies', *Social Inclusion* **10**(4).

Bijak, J. (2010), *Forecasting International Migration in Europe: A Bayesian View*, Springer, Dordrecht.

Bijak, J. and Czaika, M. (2020), 'Black swans and grey rhinos: Migration policy under uncertainty', *Migration Policy Practice* **X**, 14–20.

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

Bussiere, M. and Fratzscher, M. (2006), 'Towards a new early warning system of financial crises', *journal of International Money and Finance* **25**(6), 953–973.

Böhme, M. H., Gröger, A. and Stöhr, T. (2020), 'Searching for a better life: Predicting international migration with online search keywords', *Journal of Development Economics* **142**(C), 102347.

Candelon, B., Dumitrescu, E.-I. and Hurlin, C. (2012), 'How to evaluate an early-warning system: Toward a unified statistical framework for assessing financial crises forecasting methods', *IMF Economic Review* **60**(1), 75–113.

Carammia, M., Iacus, S. M. and Wilkin, T. (2022), 'Forecasting asylum-related migration flows with machine learning and data at scale', *Scientific Reports* **12**(1), 1–16.

Carastathis, A., Spathopoulou, A. and Tsilimpounidi, M. (2018), 'Crisis, What Crisis? Immigrants, Refugees, and Invisible Struggles', *Refuge: Canada's Journal on Refugees* **34**(1), 29–38.

Chipolina, S. (2021), 'Afghanistan's Pivot to Crypto: Will It Work?, url=https://decrypt.co/84071/afghanistan-bitcoin'. Decrypt, Accessed: 9 March 2022.

Crawley, H. (2016), 'Managing the Unmanageable? Understanding Europe's Response to the Migration 'Crisis'', *Human Geography* **9**(2), 13–23.

Cummings, C., Pacitto, J., Lauro, D. and Foresti, M. (2015), 'Why people move: understanding the drivers and trends of migration to Europe', *London: Overseas Development Institute* .

Czaika, M. and Reinprecht, C. (2020), Drivers of migration: A synthesis of knowledge, IMI Working Paper No. 163, International Migration Institute, University of Amsterdam, Amsterdam.

de Haas, H. (2018), European migrations: Dynamics, drivers, and the role of policies, Technical Report JRC109783, Publications Office of the European Union, Luxembourg. EUR 29060 EN.
URL: *https://publications.jrc.ec.europa.eu/repository/handle/JRC109783?mode=full*

De Haas, H. (2021), 'A theory of migration: the aspirations-capabilities framework', *Comparative Migration Studies* **9**(1), 1–35.

Drazanova, L. (2022), 'Why are Ukrainian refugees welcomed in Central and Eastern Europe?'. EUI - Migration Poilcy Centre Blog, Accessed: 1 April 2022.
URL: *https://blogs.eui.eu/migrationpolicycentre/why-are-ukrainian-refugees-welcomed-in-central-and-eastern-europe/*

Edison, H. J. (2003), 'Do indicators of financial crises work? An evaluation of an early warning system', *International Journal of Finance & Economics* **8**(1), 11–53.

Eissa, T. and Cho, G.-h. (2013), Internet Anonymity in Syria, Challenges and Solution, *in* 'IT Convergence and Security 2012', Springer, pp. 177–186.

Erdal, M. B. and Oeppen, C. (2018), 'Forced to leave? The discursive and analytical significance of describing migration as forced and voluntary', *Journal of Ethnic and Migration Studies* **44**(6), 981–998.

Filippopoulou, C., Galariotis, E. and Spyrou, S. (2020), 'An early warning system for predicting systemic banking crises in the Eurozone: A logit regression approach', *Journal of Economic Behavior & Organization* **172**, 344–363.

FitzGerald, D. S. (2015), The sociology of international migration, *in* C. B. Brettell and J. F. Hollifield, eds, 'Migration Theory, Talking Across Disciplines', Routlege, New York, pp. 115–146.

Gehrke, L. (2022), 'Georgia, Moldova follow Ukraine in applying to join EU'. Politico, Accessed: 28 April 2022.
URL: *https://www.politico.eu/article/georgia-and-moldova-apply-for-eu-membership/*

Hampshire, J. (2015), 'Europe's migration crisis', *Political Insight* **6**(3), 8–11.

Hasse, J.-B. and Lajaunie, Q. (2021), Package 'EWS', Technical Report EWS, CRAN Repository.

Hausmann, R., Hinz, J. and Yildirim, M. A. (2018), Measuring Venezuelan emigration with Twitter, Kiel Working Papers 2106, Kiel Institute for the World Economy (IfW Kiel).
URL: *https://ideas.repec.org/p/zbw/ifwkwp/2106.html*

Jaroszewicz, M. (2019), 'Years After Crimea's Annexation, Integration of Ukraine's Internally Displaced Population Remains Uneven'. Migration Policy Institute, Accessed: 28 April 2022.
URL: *https://www.migrationpolicy.org/article/fyears-after-crimea-annexation-integration-ukraine-internally-displaced-population*

Juric, T. (2022), 'Predicting refugee flows from Ukraine with an approach to Big (Crisis) Data: a new opportunity for refugee and humanitarian studies', *medRxiv* .

Kaminsky, G., Lizondo, S. and Reinhart, C. M. (1998), 'Leading indicators of currency crises', *Staff Papers* **45**(1), 1–48.

Kauppi, H. and Saikkonen, P. (2008), 'Predicting u.s. recessions with dynamic binary response models', *The Review of Economics and Statistics* **90**(4), 777–791.

Kilian, L. and Taylor, M. P. (2003), 'Why is it so difficult to beat the random walk forecast of exchange rates?', *Journal of International Economics* **60**(1), 85–107.

Lajaunie, Q. (2021), Nonlinear Impulse Response Function for Dichotomous Models, LEO Working Papers / DR LEO 2852, Orleans Economics Laboratory / Laboratoire d'Economie d'Orleans (LEO), University of Orleans.

Lang, J. H., Peltonen, T. A. and Sarlin, P. (2018), A framework for early-warning modeling with an application to banks, Working Paper Series 2182, European Central Bank.

Leasure, D. R., Kashyap, R., Rampazzo, F., Elbers, B., Dooley, C., Weber, I., Fatehkia, M., Bondarenko, M., Verhagen, M. D., Frey, A. and et al. (2022), Ukraine Crisis: Monitoring population displacement through social media activity, Technical report.
URL: *osf.io/preprints/socarxiv/6j9wq*

Leetaru, K. (2014), 'Did the Arab Spring Really Spark a Wave of Global Protests?'. Accessed: 26 April 2022.
URL: *https://foreignpolicy.com/2014/05/30/did-the-arab-spring-really-spark-a-wave-of-global-protests/*

Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A. and Taylor, J. E. (1993), 'Theories of International Migration: A Review and Appraisal', *Population and Development Review* **19**(3), 431–466.

Melachrinos, C., Carammia, M. and Wilkin, T. (2020), Using big data to estimate migration "push factors" from Africa, *in* P. Fargues, M. Rango, E. Borgnäs and I. Schöfberg, eds, 'Migration in West and North Africa and across the Mediterranean Trends, risks, development and governance', IOM, Geneva, pp. 98–116.

Milliff, A. and Christia, F. (2021), Who Flees Conflict? A Big-Data Approach to the Determinants of Forced Migration, Preprint.
URL: *https://aidanmilliff.com/publication/who-flees-big-data-migration/WhoFlees.pdf*

Napierała, J., Hilton, J., Forster, J. J., Carammia, M. and Bijak, J. (2022), 'Toward an early warning system for monitoring asylum-related migration flows in Europe', *International Migration Review* **56**, 33–62.

Palotti, J., Adler, N., Morales-Guzman, A., Villaveces, J., Sekara, V., Garcia Herranz, M., Al-Asad, M. and Weber, I. (2020), 'Monitoring of the venezuelan exodus through facebook's advertising platform', *PLoS One* **15**(2), e0229175.

Raleigh, C. (2011), 'The search for safety: The effects of conflict, poverty and ecological influences on migration in the developing world', *Global Environmental Change* **21**, S82–S93.

Sahin-Mencutek, Z., Barthoma, S., Gökalp-Aras, N. E. and Triandafyllidou, A. (2022), 'A crisis mode in migration governance: comparative and analytical insights', *Comparative Migration Studies* **10**.

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

Selby, J., Dahi, O. S., Fröhlich, C. and Hulmee, M. (2017), 'Climate change and the Syrian civil war revisited', *Political Geography* **60**, 232–244.

Sohst, R. R., Tjaden, J. D., de Valk, H. and Melde, S. (2020), *The future of migration to Europe: A systematic review of the literature on migration scenarios and forecasts*, International Organization for Migration, Geneva and The Netherlands Interdisciplinary Demographic Institute, The Hague.

Sonnenfeld, J., Tian, S., Sokolowski, F., Wyrebkowski, M. and Kasprowicz, M. (2022), Business Retreats and Sanctions Are Crippling the Russian Economy, Technical report, Yale University. Available at SSRN: https://ssrn.com/abstract=4167193.

Tanas, A. (2022), 'With war on its doorstep, Moldova applies for EU membership'. Accessed: 28 April 2022.
URL: *https://www.reuters.com/world/europe/moldovan-president-says-moldova-applies-eu-membership-2022-03-03/*

The United States Government Publishing Office (2011), ‘Executive Order 13582 of August 17, 2011 Blocking Property of the Government of Syria and Prohibiting Certain Transactions With Respect to Syria’. Accessed: 2 May 2022.
URL: *https://www.govinfo.gov/app/details/CFR-2012-title3-vol1/CFR-2012-title3-vol1-eo13582*

Van Hear, N., Bakewell, O. and Long, K. (2018), ‘Push-pull plus: reconsidering the drivers of migration’, *Journal of ethnic and migration studies* **44**(6), 927–944.

Vestby, J., Tollefsen, A. F. and Buhaug, H. (2022), Climate and international migration flows: A sensitivity analysis of gravity model specifications, QuantMig Project Deliverable D2.6, PRIO, Oslo.

Wu, C. and Gerber, M. S. (2018), ‘Forecasting civil unrest using social media and protest participation theory’, *IEEE Transactions on Computational Social Systems* **5**(1), 82–94.

Young, T. (2015), ‘10 maps that explain Ukraine’s struggle for independence’. Accessed: 27 April 2022.
URL: *https://www.brookings.edu/blog/brookings-now/2015/05/21/10-maps-that-explain-ukraines-struggle-for-independence/*

Zelinska, O. (2017), ‘Ukrainian Euromaidan protest: Dynamics, causes, and aftermath’, *Sociology Compass* **11**(9), e12502.