IMR

Toward an Early Warning System for Monitoring Asylum-Related Migration Flows in Europe

Joanna Napierała
Joint Research Centre (JRC), European Commission, Brussels;
European Asylum Support Office (EASO), Malta

Jason Hilton
ESRC Centre for Population Change, Centre for Population Change,
University of Southampton, UK

Jonathan J. Forster
ESRC Centre for Population Change, University of Southampton, UK;
Department of Statistics, University of Warwick, UK

Marcello Carammia
European Asylum Support Office (EASO), Malta;
Department of Political and Social Sciences, University of Catania, Italy

Jakub Bijak (iD)
ESRC Centre for Population Change, University of Southampton, UK

Abstract

Asylum-related migration is highly complex, uncertain, and volatile, which precludes using standard model-based predictions to inform policy and operational decisions. At the same time, asylum's potentially high societal impacts on receiving countries and the resource implications of asylum processes call for more proactive approaches for assessing current and future migration flows. In this article, we

Corresponding Author:
Jakub Bijak, University of Southampton, UK.
Email: j.bijak@soton.ac.uk

propose an alternative approach to asylum modeling, based on the detection of early warning signals by using models originating from statistical control theory. Our empirical analysis of several asylum flows into Europe in 2010–2016 demonstrates the approach's utility and potential in aiding the management of mixed migration flows, while also shedding more light on the work needed to make better use of the "big data" and scenario-based methods for comprehensive and systematic examination of risk, uncertainty, and emerging trends.

Introduction

It is a truism to state that contemporary migration processes are very complex and driven by a vast array of interrelated factors and drivers. One implication of this complexity is that the predictability of migration flows is very limited, much more so than is the case for other demographic processes (Bongaarts and Bulatao 2000). The limits of migration's predictability are strongly pronounced even in the case of relatively regular flows, such as labor or student migration, and much more so for various forms of forced migration, including asylum flows, which can have large impacts on both sending and receiving countries (Bijak et al. 2019).

Asylum migration is so unpredictable because it is highly dependent on specific underlying geopolitical events and processes, which themselves are difficult to foresee, and because these political events and policy developments interact with migration flows in a feedback loop (Castles 2004; Pijpers 2008; Bijak 2010; Bijak and Czaika 2020). The inherent uncertainty associated with asylum migration is further compounded by the agency of the various actors involved, from migrants to policy-makers (Castles 2004). Still, from a policy and planning viewpoint, there is a need for at least some form of quantitative foresight into the future, or even current, magnitude of migration flows. The armed conflicts that took place in recent decades, such as those in the former Yugoslavia, Iraq, Afghanistan, or Syria, the political crisis in Venezuela, and the associated flows of asylum-seekers migrating toward Europe, have generated an increased desire among recipient countries' governments to have access to predictions of the nature, magnitude, and composition of future flows (Sohst et al. 2020).

Migration's high volatility means that traditional forecasting attempts, based on "orderly" assumptions implying stable trends, are problematic (Pijpers 2008). Simple time-series-based forecasts can yield reasonable predictions in the perspective of a few years ahead, but with uncertainty rapidly increasing with the forecast horizon (Bijak and Wiśniowski 2010), even more so for asylum migration (Bijak

et al. 2019). Even though some migration indicators may exhibit long-term regularities, detectable over a span of a few decades (Azose and Raftery 2015), this time perspective is too coarse to enable responding to the immediate policy needs in such rapidly changing areas as an asylum.

As many of the recent asylum flows were sizeable and long-lasting, there was an increasing need for national and European Union (EU) policy-makers to facilitate planning and to ensure that asylum claims are processed quickly and efficiently and that relief is provided promptly to those in need (Kegels 2016). The utility of asylum forecasts, even with short horizons, is that they can aid planning to ensure that the necessary support, as well as administrative and welfare capacity, is in place ahead of time, enabling more proactive responses and planning. For asylum-related migration, preparedness is often related to securing basic operational needs: host countries and relief organizations need to provide food and shelter to people seeking asylum to the right extent and in the right places; administrations need adequate resources for registering asylum-seekers, verifying their identity, issuing documentation, processing applications, and resettling refugees into their receiving societies. As became apparent during the 2015–2016 peak of asylum migration into Europe, without an early warning system in place, the building of additional capacity often must be done on a short notice, especially as by law, asylum claims must be processed within six months (ECRE 2016).

This article aims to respond to the policy and operational needs for forward-looking information on asylum-related migration by offering a tool for early detection of changes in migration flows based on higher-frequency data (e.g., monthly or weekly). To that end, we advance existing quantitative approaches for migration modeling by introducing an early warning model based on change-point detection, inspired by statistical control theory, and then assessing the feasibility of its use. We demonstrate this approach's utility on weekly and monthly data of migrant arrivals and asylum applications in Europe and on detections of irregular border crossings, and assess its practical feasibility for monitoring and managing asylum-related migration into and within the EU. We focus on Europe because of the presence of unique regulated regional cooperation in establishing responsibility for examining asylum applications and the exchange of information and creation of a joint system for early warnings, preparedness, and management of asylum crises. Our proposed methodology, however, has broader applicability to the global problem of forecasting and responding to forced displacements, subject to the availability of appropriate data.

From a theoretical standpoint, the tools we propose aim to help nation-states and supranational bodies, such as the EU, better discharge their legal and moral obligations toward refugees and asylum-seekers by providing them with a more efficient system of humanitarian protection. In particular, the approach we propose focuses on the supply of immediate humanitarian needs—offering *refuge* to those in need (Owen 2020)—and, thus, takes the perspective of receiving states or entities, while recognizing these states' role in constructing the legal notion of refugeehood and the practice of granting asylum.

Our article's structure is as follows. After this introduction, the section "Modeling Asylum-Related Migration: From Concepts to Operational Needs" provides a brief discussion of the conceptual difficulties with defining the analytical categories and variables of interest and offers pragmatic solutions related to their operationalization for modeling and early warnings. One such model, based on statistical control theory, is introduced in the section "Early Warning Models of Asylum Migration: State of the art and Application", and the results of its application to weekly and monthly data on asylum applications and detections of irregular border crossings in Europe are summarized and discussed in the section "Detecting Changes in Asylum Flows Into the EU". Section "Sensitivity Checks and Perspectives for Early Warnings for Asylum Migration" presents the results, as well as a range of sensitivity checks and perspectives for using early warnings for asylum migration. The article finishes with a discussion of practical recommendations and possible extensions of the approach presented in the section "Conclusions and Recommendations". In addition, an Online Appendix provides a review of formal, quantitative methods and models applied to asylum migration, as well as data availability for the purpose of building models providing early warnings related to asylum migration and the computer code for simulations presented in the section "Detecting Changes in Asylum Flows Into the EU".

Modeling Asylum-Related Migration: From Concepts to Operational Needs

Formal quantitative work on asylum-related migration faces four key challenges, related to the ways in which the asylum processes are conceptualized, defined, explained, and measured (e.g., Singleton 2016; Bijak, Forster and Hilton 2017; Erdal and Oeppen 2018). In this section, these four challenges are discussed in turn. First, at the conceptual level, contemporary migration scholarship has moved away from simple dichotomies, such as "forced" versus "voluntary" migration, and toward a multidimensional spectrum in which individual migration motives, including different levels of human agency, interact, overlap, and are difficult to isolate (King 2002; Foresight 2011; Erdal and Oeppen 2018; Hatton 2020).

Second, the very concept of "asylum migration" similarly evades precise analytical definition and must remain relatively vague to accommodate different forms of international mobility, the common element of which is related to the need for international protection (Böcker and Havinga 1997). As recognized in the theoretical literature in the field of refugee studies, there is an important conceptual gap between the legal (political, constructivist) status of refugees and asylum-seekers and the practical (humanitarian, realist) situation of people in need of protection (FitzGerald and Arar 2018; Owen 2020). Still, the state's role in constructing the legal and practical aspects of the status of refugees and asylum-seekers, who find themselves as outsiders "between sovereigns" (Haddad 2008), cannot be overlooked. Equally important are state *responses* to the challenges of asylum, or migration more generally, through changes in the

spatial and organizational features of the different types of borders, be they intra-European or external (e.g., Geddes 2005; Geddes and Scholten 2016).

Third, the known theoretical gaps and limitations of migration studies (Arango 2000) are exacerbated in the case of asylum. In particular, the caveats regarding weak and fragmented theories of migration are further amplified by the theoretical disconnection between refugee scholarship and mainstream migration studies (FitzGerald 2015). As a result, theoretical work on refugee or asylum processes is still relatively rare, with the seminal overview of migration theories (Massey et al. 1993) mentioning refugees only in the context of the world systems theory and global impacts of military interventions.

Notable examples of existing theoretical studies on refugee and asylum processes date back at least to the comprehensive framework offered by Zolberg (1989). Building on the world systems theory, this framework added important elements from state theory and highlighted nation-states' pivotal role in shaping asylum policies and, hence, flows. This approach has recently been revisited, reviewed, and revised by FitzGerald and Arar (2018), who advocate for a multiperspective view on refugee and asylum flows and draw attention to distinct developments in the Global South. Other theoretical perspectives of refugee and asylum migration highlight the role of international coordination (or competition) between the countries hosting refugees (for a review, see Suriyakumaran and Tamura 2016), as well as the interplay between asylum migration, policy, and public attitudes toward migrants (e.g., Hatton 2020), both studied typically from an economic viewpoint.

For the purpose of empirical modeling, a range of political, policy, and socioeconomic proxy variables can be used as predictors (see the Online Appendix for details), an approach dating back to the *push-and-pull factors* framework proposed in Lee's (1966) classic paper. The push-and-pull approach, which has inspired several updates and refinements in more contemporary literature (e.g., Arango 2000; Carling and Collins 2018; Van Hear, Bakewell and Long 2018), can be used as a practical guide for discussing migration factors and drivers. One of its important extensions, crucial from the viewpoint of asylum flows' inherent uncertainty, is *chance* (Böcker and Havinga 1997), which of itself reduces the importance of other factors and drivers of asylum. In this context, Öberg's (1996) distinction between "soft" and "hard" factors is relevant, the latter including drivers more associated with the involuntary end of the migration spectrum and implying less choice and self-selection among prospective migrants. A recent survey of asylum migration's factors and drivers is available in the EASO (2016a) report.

At the same time, migration processes are known to self-perpetuate through a variety of mechanisms (for a review, see Massey et al. 1993). Asylum flows are no exception—in their case, the "cumulative causation" results from the existence of migrant networks and the presence of formal family reunion routes. Such self-perpetuation mechanisms result in high inertia of migration processes. This inertia can be formally described and modelled, for example, by using time series approaches, such as autoregressive models (e.g., Bijak 2010). Many existing models of asylum

migration include autoregressive features, aimed to capture the self-perpetuating nature of migration (see the review in the Online Appendix for details on selected examples of how various migration theories could be operationalized in models).

Fourth, in addition to the theoretical considerations presented above, the measurement of asylum-related migration is problematic. Not only is population-level data collection on asylum migration heavily politicized and subject to external pressures (Bakewell 1999; Crisp 1999), but it also often focuses on operational humanitarian data or administrative procedures, such as registrations (NASEM 2019). Administrative records, especially those of different countries, are likely to contain duplications for those who have applied for international protection more than once in different countries, either after initial rejections or due to different lengths of the asylum processes (Singleton 2016). Even within the EU, the development of a harmonized system of asylum statistics remains an ongoing process involving Eurostat, the European Asylum Support Office (EASO), and other agencies, with much current work centered on the use of biometric data, such as fingerprints, held at the EURODAC database.

A pragmatic solution to the dilemmas posed by the difficulties associated with conceptualizing, defining, and measuring asylum-related migration is to adopt the perspective of potential users of the early warning models and to focus on the indicators that are of relevance for them. The main indicator used to analyze and manage asylum processes is the number of *applications lodged* in a given country (or reception center) in a unit of time: day, week, month, or year (e.g., EC 2016). Given this indicator's availability for most European countries, from national statistical and migration authorities, Eurostat, or EASO Early warning and Preparedness System data,[1] and its importance for operational reasons, the empirical analysis presented in this article is based on the numbers of applications.

In this way, at the strictly operational level, the conceptual and theoretical challenges related to the quantification of asylum migration become less relevant, as do other definitional or measurement criteria, such as the duration of stay. What matters, instead, is the purely administrative fact of lodging an asylum application by a migrant at a specific time and place. This approach, relying on the analysis of crude numbers of asylum applications is, of course, a reductionist view of the asylum process but, nonetheless, enables a forward-looking analysis that can have a direct policy use through helping with preparedness planning.

Early Warning Models of Asylum Migration: State-of-the-art and Application

At a general level, the main purpose of an early warning model is to detect changes in the trend of a variable or variables early enough to allow for adaptation or

---

[1] See the latest asylum trends, https://www.easo.europa.eu/latest-asylum-trends (as of July 5, 2021).

undertaking remedial action at least sometime before the consequences of the changes become too overbearing. Early warning modeling is frequently used in the financial context, chiefly in central banking and macroprudential regulation (Lang, Peltonen and Sarlin 2018). Such models gained popularity after the 2008–2009 economic downturn, as they are designed to help detect the first signs of an upcoming economic slowdown or recession (see Lang, Peltonen and Sarlin 2018 for a recent overview).

To that end, the purpose of early warning *modeling*—defined as "identifying vulnerable states prior to… crises, which can also be viewed as a standard two-class classification task, where the key objective is to separate the vulnerable from non-vulnerable states" (Lang, Peltonen and Sarlin 2018: 5)—can be easily generalized beyond finance. Depending on whether an indicator being monitored crosses a certain threshold, the trend is classified as either exhibiting regular behavior ("in control") or generating a warning signal indicating vulnerability of the process ("out of control"). In addition, in this article, we understand an early warning *system* as a collection of interlinked early warning models and the underpinning sources of information and data, serving the same purpose, only for a broader range of indicators and sources.

Statistical models for monitoring data to identify change points in series have a long history in the area of industrial process control, where measures of process quality are monitored to minimize losses on production lines (Page 1954, 1957). The literature on early warnings and statistical control theory offers many choices of modeling approaches, levels of the warning thresholds, and model calibration and validation methods (e.g., Zeileis 2004; Li et al. 2013). Of particular importance are the tradeoffs between the frequency of "false alarms" (false positives), unwarranted complacency (false negatives), and the correct classification of changes in trends. The presence of these tradeoffs necessitates the direct involvement of migration policy-makers in the design of early warning models so that they can describe the possible *losses* which can occur under different circumstances, for a range of possible policy decisions (Bijak, Forster and Hilton 2017; Lang, Peltonen and Sarlin 2018). An all-encompassing methodological framework combining all desirable elements is offered by statistical control theory and change-point models.

In terms of existing applications, early warning macroeconomic models are usually based on large general equilibrium approaches, the equivalents of which do not yet exist in the context of migration (Barker and Bijak 2020). Still, in the migration literature, the use of early warnings has been proposed for the most volatile and weakly predictable flows, such as asylum migration (see Shellman and Stewart 2007; Regehr 2014), ideally in conjunction with a coherent description of uncertainty and impact of individual decisions for the purpose of managing the associated risk (Bijak et al. 2019).

For a single early warning model, as mentioned above, using exogenous variables as predictors does not seem useful, as such predictors' values would need to be modelled and forecast separately, increasing the overall prediction uncertainty (Bijak

2010). On the other hand, if models are to be deployed as a part of a wider system for several countries or flows, some exogenous variables—most notably, migration or asylum policies for destinations (see Beine et al. 2016) or conflict intensity measures for origin countries or regions—can help model the interactions between the different elements of such a system.

In the context of asylum migration, in common with the general settings of statistical control theory, the aim of an early warning analysis is to identify as early as possible when the process governing a series has changed, while minimizing the number of false alarms (Page 1954). To achieve this goal, a set of rules must be devised governing when action is to be taken. In the case of asylum, such actions might involve planning and preparation for the arrival of more asylum-seekers. These decision rules should account for the variability of the process under "normal" conditions and should be sensitive enough to detect subtle shifts at an early stage.

The cumulative sum (or "Cusum") model set out by E.S. Page (Page 1954) and developed by a range of subsequent authors (Barnard 1959; Harrison and Davies 1964) provides a setting for such rules. As the name suggests, this approach monitors the cumulative sum of some function of the observed data. In the original formulation of this method by Page (1954), which is summarized below, this function is chosen to produce a score that is negative when the data are behaving as expected and positive when unexpectedly large values are observed.

Through monitoring the cumulative sum of this score, an action can be recommended when the difference between the cumulative sum's current value and minimum value passes a certain limit (threshold), which must be set in accordance with knowledge about what change in the data is deemed important. This rule has the effect of detecting change both in the face of a very large single observation and in the case of many moderately large observations in a relatively short space of time, both of which would indicate an increase in the mean of the process.

More formally, assuming some series of data observed across time t, $\boldsymbol{x} = x_0, x_1, \ldots, x_t$, either raw or generated by some model (e.g., a constant or autoregressive one), the tth observation can be given a numerical score $y_t = f(x_t)$. With a starting value of zero $S_0 = 0$, the cumulative sum of the first n scores $S_n = \sum_{t=0}^{n} y_t$ triggers action if $(S_n - \min_{k<n}(S_k)) > h$, where h is the chosen limit (threshold). The limit implies the value of the scoring function for a single observation that would immediately demand action, irrespective of the series' history. Because values that precede the minimum of the cumulative sum series are not relevant for determining when action should be taken, the rule can be simplified by restarting the sum whenever a new minimum is reached:

$$V_n = \max(V_{n-1} + y_n, \ 0),$$
$$V_0 = 0.$$

(1)

Action is now demanded whenever $V_n > h$.

A simple example is given in Figure 1 to aid understanding. Here, a fictitious series of simulated observation points x_t, $t > 0$, with an initial mean of $\mu = 5$, is being monitored, as shown in Figure 1A. After thirty observations (dotted line), the series changes so that observations come from a distribution with a higher mean and variance. For this example, we use a score of $y_t = x_t - \mu$ so that we
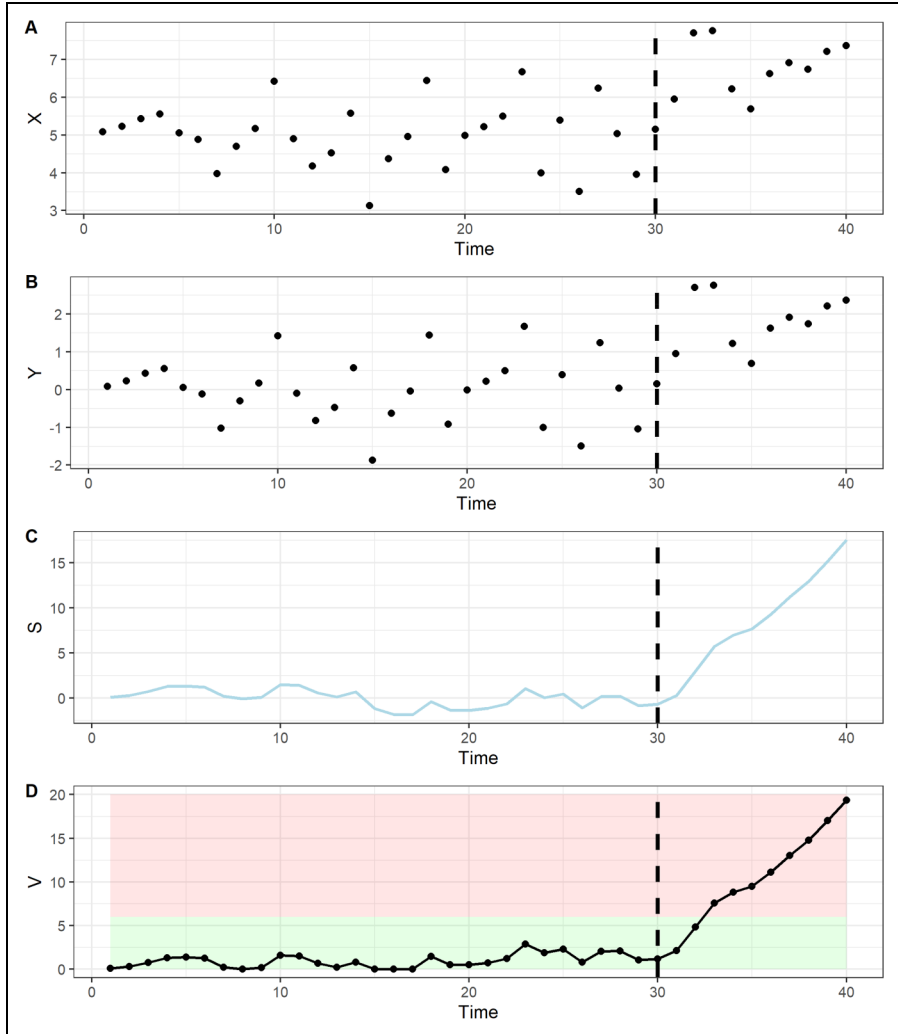

Figure 1. Cusum Control Chart Applied to Simulated Data with a Change Point at Time 30. (A) Simulated Observations *x*; (B) Score Function *y*; (C) Cumulative Sum *S*; (D) Cumulative Sum *V*, Restarted at Minimum Values.

consider observations higher than the initial mean as potentially troublesome and assume that the true initial value (5) of the time series' mean is known. Figure 1B plots the values of y_t, which consists of the series x_t, which is shifted so that it is centered about 0.

We can observe in Figure 1C that the individual values of the vector of cumulative sums, S, initially remain roughly horizontal, as the observed values fluctuate around the mean. After the change-point, the process shifts, and the average score is greater than 0; thus, the cumulative sum starts to increase. By observing the values of the vector V of the corresponding indicators V_n, Figure 1D shows that the series restarts at 0 as new minimums of the cumulative sum are reached. The limit h is arbitrarily chosen to be 6 in this case and is represented by the boundary between the red and green areas in Figure 1D. After the change-point occurs at $t = 30$, only three additional observations are needed to trigger the action rule, at which point the series V crosses the action threshold by entering the red zone.

Figure 1D is an example of a *control chart*, used to provide a simple graphic illustration of when a series exceeds expected limits. We might also be interested in cases when a series shows a substantial decline. In this case, a two-sided control chart can be used (Page 1957; Barnard 1959). Two-sided charts would involve constructing an equivalent scoring function for declines in the series with an opposite sign and monitoring changes in the cumulative sum of this sequence relative to its maximum, as an equivalent to the series V.

Furthermore, if the series to be monitored is assumed to follow a particular distribution or model, then both the scoring function and the control limits can be chosen to reflect the probability of observing particular sequences of values of the series, conditional on some model parameters (Page 1957). Alternatively, the likelihood of a model describing the system's normal behavior can be compared with an alternative model with, for example, a shifted mean (Page 1957).

In the example above, by using the score $y_t = x_t - 5,$ we are monitoring for any shift in the process to a higher mean. By using a score of the form $y_t = x_t - \mu - k$, only values of x_t that are of a distance of more than k from the mean contribute to the Cusum. Setting k to some multiple of the standard deviation of the process, thus, reduces the alarm rule's sensitivity to small shifts in the mean. The discussion of the score to this point assumed that the mean parameters of the process were known and fixed, which need not always be the case. The series' parameters can be estimated either from some historical calibration data or "on the fly," with the estimates being updated as more observations become available. Table 1 summarizes all the variables and parameters used in the Cusum model.

The sequential updating of parameters has a natural Bayesian statistical interpretation. West and Harrison (1986) introduced a Bayesian approach to the problem of change-point detection by comparing the relative evidence for a "standard" model against one which assigns a greater probability to extreme observations, using cumulative Bayes factors. If the evidence strongly indicates the latter model, a change-point can be considered to have occurred (see also Tartakovsky and Moustakides

Table 1. Variables and Parameters of the Cusum Model

Variable/parameter	Description
x_t	Observations (e.g., number of lodged applications, the number of people apprehended trying to cross the border/detections of irregular border crossings, and simulated values of the process)
y_t	Numerical score based on the observations, $y_t = f(x_t)$
S_n	Cusum: Cumulative sum of the first *n* scores, $S_n = \sum_{t=0}^{n} y_t$
V_n	Alternative Cusum, with values, reset to zero after every minimum (alarm): $V_n = \max(V_{n-1} + y_n, \ 0)$
h	Constant, a chosen limit (threshold)
k	Constant, a multiplier (e.g., the number of standard deviations calculated based on the observed historical data)
μ	Mean value of observations
t	Units of time (e.g., weeks and months)

2010 for a recent review of Bayesian approaches to change-point detection). The identification of change points can also be posed as a decision theory problem, providing a theoretical framework within which to set control limits *h* based on the losses and gains associated with particular outcomes (Harrison and Veerapen 1994).

A separate question concerns the possible methods of evaluating the performance of such early warning models. Two main perspectives for assessing model performance are (1) internal, looking at how well the model performs by using either analytical techniques or real or simulated datasets, and (2) external, comparing with other benchmark approaches. The internal approach involves measuring the frequency of false positives and false negatives and combining them in a quality indicator, possibly allowing for asymmetry in impact, cost, or other consequences between false alarms and false complacency. A special case, focusing just on false positives, as measured by using *p*-values, is well documented in the literature on statistical process control and Cusum models (e.g., Zeileis 2004; Li et al. 2013). Judgment on whether the alarm was true or false, to a large extent, rests with the user, which is a subjective element of the internal assessment.

A mathematical analysis of alarm schemes also helps assess model performance from an internal perspective. Work by Lorden (1971) and Moustakides (1986) show that the Cusum procedure described by Page (1954) for choosing a stopping time has some desirable qualities. More specifically, for any chosen probability of false alarms, it is a so-called *minimax decision rule*, as it minimizes the expected delay in detecting a change in the level under the worst possible conditions, in terms of the run of observations preceding the change in level (Moustakides 1986). As Yashchin (1993) notes, minimax rules have desirable properties in monitoring schemes, as in general, an observer is not able to make assumptions about the typicality of observations preceding a change in level. The precautionary nature of

minimax approaches is certainly helpful in the case of monitoring asylum applications, where worst-case scenarios can form a basis of planning.

In the external approach to evaluation, a selected quality indicator can be compared with the corresponding measure obtained for the benchmark model. At the same time, there is no "gold standard" for selecting the right benchmark for such models. In the next two sections, we expand this discussion by illustrating the framework with a few examples related to recent asylum migration flows into Europe.

Detecting Changes in Asylum Flows Into the EU

To illustrate the Cusum model's potential application for detecting changes in asylum trends, this section presents a selection of examples based on data about the number of asylum applications lodged in the EU. In these examples, Cusum is applied to raw application data, without any additional modeling or smoothing.[2] As for the data used, although the administrative practice of evaluating an asylum application may vary between EU member-states, to access the procedure, the applicant usually must file their asylum application. This procedure is based on three steps, which consists of *making*, *registering*, and *lodging* the application. The process of *making* an application starts as soon as the person expresses a wish to receive international protection from authorities, which in theory usually happens after crossing the border of the first safe country and lasts until all necessary documents are completed. By the EU law, the application must be *registered* by representatives of the national authority responsible for refugees within a certain time limit, which depends on the host country (EC 2016). This process of *registering* a claim should normally take place within three working days from the time the migrant crossed the border and informed the authorities about their wish to apply for international protection (EC 2016). *Lodging* the application is the final step required to access the procedure to obtain international protection and is an official acknowledgment of the start of the determination process when the asylum-seeker receives a certification of acceptance of their application (EC 2016).

Ideally, a Cusum model for asylum migration would be based on the monitoring of the number of individuals who are in the process of making an application, since this quantity is most relevant to decision-makers for planning and operational purposes. However, in most EU countries, the national asylum authorities only register the last step of the procedure—the *lodging* of an application formally accepted by the relevant authorities, while the start of the process of making an application is not recorded. The main disadvantage of setting up a Cusum model for data on the number of lodged asylum applications is that as asylum flow increases, these

---

[2] A natural extension of the proposed approach would be to apply Cusum models to autoregressive or other time series or trend models fitted to the raw data series, if such models were found to better reflect the true data-generating process.

numbers at some point may start to reflect, not the size of the flow, but the relevant authorities' maximum capacity to process applications. Reflecting the processing capacity, rather than the real size of asylum-seeker flows, is an acute problem when the number of applicants is higher than the number of available staff to process applications. As a result, in such periods, the number of asylum applications observed each week or month will approximately correspond to the number of available staff multiplied by the average number of applications they can process in a given time period, which in turn will vary (i.e., depending on the time it takes for each application to be processed).

For these reasons, any variation observed in applications between time periods could be explained by the increase or decrease in available staff's working hours, not by real changes in inflows. For instance, analysis of variation in the weekly number of lodged applications (see Figure 2) shows that during the weeks with more than one day of public holidays, the numbers of lodged applications dropped significantly, increasing by almost exactly the same quantity in the following week, as authorities tried to catch up with the backlog. Sometimes, the lodging of applications could also be postponed for other reasons such as when, for example, interpreters are not available. In such cases, a Cusum model might trigger alarms both in the week of the holidays and in subsequent weeks. These alarms could become problematic, especially when inaccurately interpreted as a change in trends of inflows, but such an issue is easy to address by monitoring the situation together with the calendar of public holidays and employees' leave schedules and adjusting the data accordingly.

When setting up the early warning system based on a Cusum model, the choice of process parameters is critical to the system's level of sensitivity. For instance, establishing a Cusum model to provide warnings about abnormal values using an upper limit h set at one historical standard deviation from the observed average values will be far more sensitive than deciding on a threshold based on two standard deviations. As an example, when systems based on these specifications are applied to the same real data series from Italy, the former system triggered nine alerts while the latter only four (see Figure 2).

To choose the parameters h and k and determine how often various combinations of these parameters triggered alarms in cases where there had been no "real" change in level, we conducted simulations, using the statistical programming language R (R Core Team 2020). More specifically, we conducted 1,000,000 simulations of repeated observations drawn from a standard normal distribution for a Cusum model with $k = 1$ and $h = 2.5$. The R code to reproduce these simulation experiments is provided in the Online Appendix. With this specification, the median time for a false positive to occur (the median "run length") was 498 observations (about once every 41 years in a monthly scheme), and 95 percent of simulations ran for at least 39 observations before a false positive occurred (corresponding to more than three years). This rate of false positives is deemed to be sufficiently conservative to guard against the excessive costs of very frequent alarms, while still
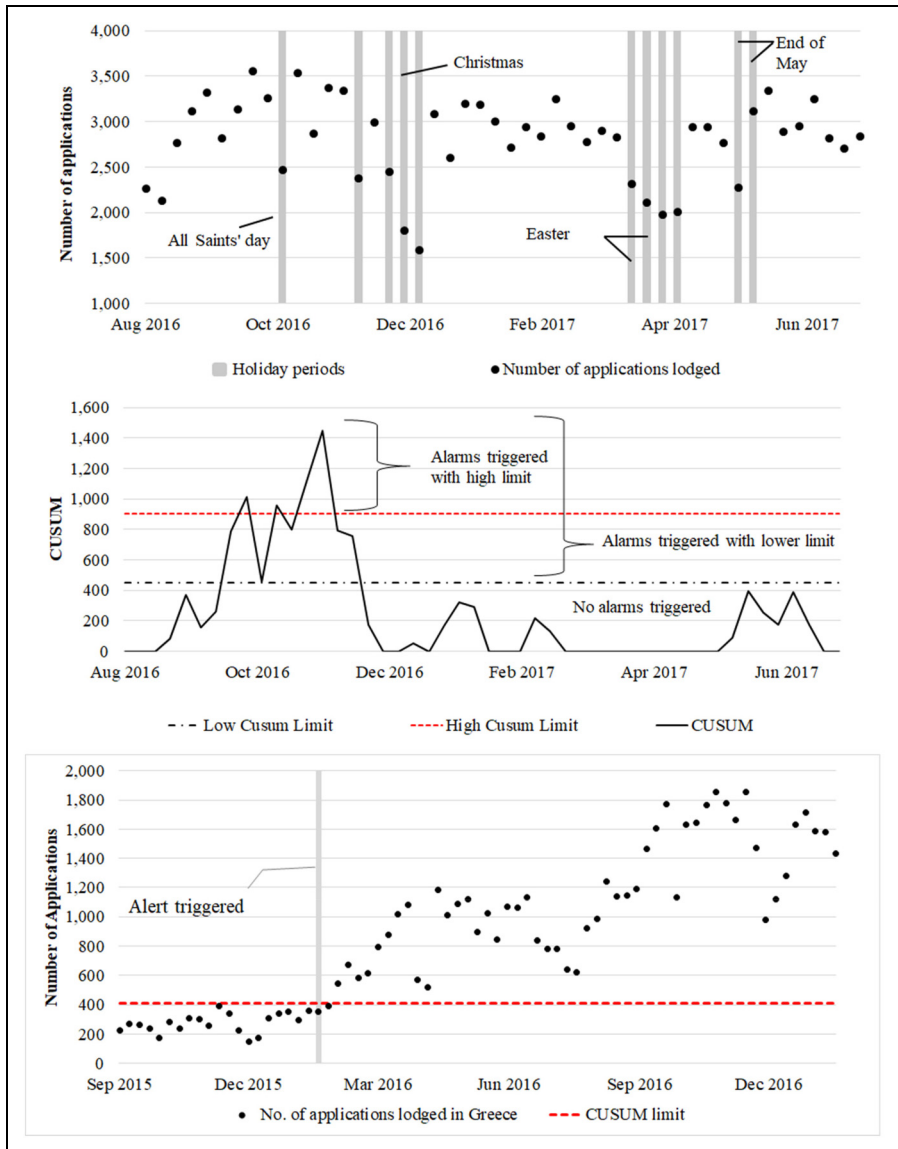
Figure 2. Holiday Periods (Both Religious and Public) Influence the Numbers of Applications Lodged in Italy (End of 2016 and the First Half of 2017) (Upper Panel), the Impact of the Choice of Thresholds Boundaries on the Sensitivity of the Early Warning System Based on Cusum (Middle Panel), Cusum Applied to the Numbers of Applications Lodged in Greece (End of 2015 to the Beginning of 2017) (Bottom Panel).
Source: EASO.

being sensitive to real changes in level; thus, these parameters are used in results presented in Figures 3 to 5.

Another aspect to be considered when building an early warning system based on a Cusum model is the calibration period. Ideally, the information used for calibration should come from the periods prior to the asylum crisis. Longer data series will not only allow for a better understanding of past developments but also help detect specific features, such as seasonality in data. The weekly data on asylum applications lodged in Greece provide an example of how the choice of calibration period may impact the system's functionality (see Figure 2, bottom panel). The alarm triggered by the Cusum model using these data occurred only in the third week of February 2016, when the crisis had started in 2015. For comparison, Pew Research Center analysis based on big data, which looked at the Arabic-language Google searches in Turkey for the word "Greece," peaked in August 2015—two months before the increased inflow of asylum-seekers (Connor 2017). Yet other Cusum models, based on different data sources (such as numbers of apprehensions reported by Frontex) started triggering alerts already in 2014 (see Figure 3).

In this case, the choice of the calibration period was limited, as the collection of these data started only from the fortieth week of 2015, making the calibration period relatively short. Moreover, the information used for calibration came from the period just before the crisis, during which the number of asylum applications was already elevated. Using too short a calibration timeframe from the precrisis period resulted in the alert's delay to the moment of the next significant increase in the number of applications lodged, which did not occur before February. Besides, even this increase could to some extent reflect the higher processing capacity related to the new staff hired in response to the crisis observed already during 2015.
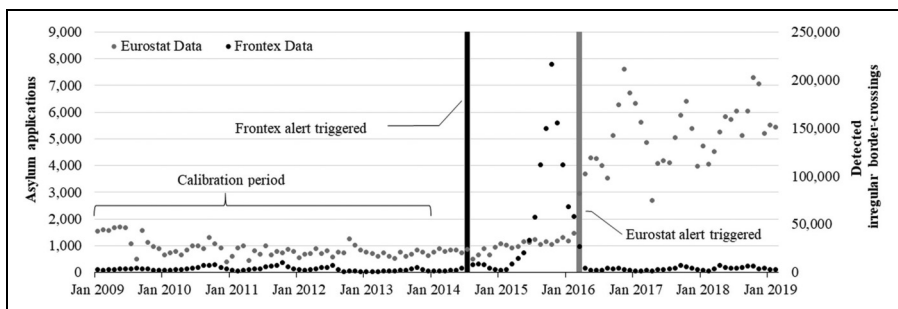

Figure 3. Examples of Application of Cusum to Greek Data on the Monthly Number of Applications Lodged (Eurostat Data) and the Monthly Number of People Apprehended Trying to Cross the Border (Eastern Mediterranean Route) (Frontex Data). The Upper Limit is Set Up on one Standard Deviation.
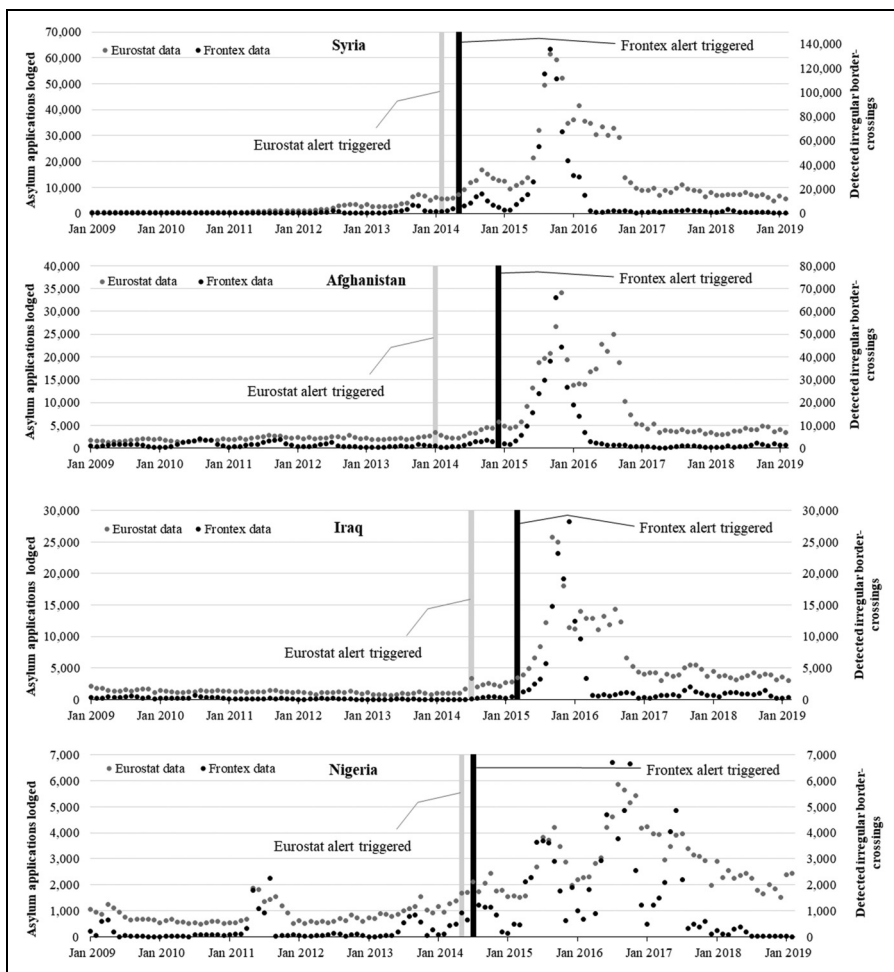*Source*: Eurostat, EASO, and Frontex data.

Figure 4. Early Warning Alerts Generated by a Cusum Model Applied to Applications Lodged by Asylum-Seekers (Eurostat Data) and Detections of Irregular Border Crossings (Frontex Data) to Flows of Afghans, Iraqis, Syrians, and Nigerians to EU28 Between January 2009 and February 2019. The Color of the Bar Corresponds to the Alarm Triggered by Cusum Based on Respective Data.
Source: Eurostat and Frontex.

For these reasons, another potential limitation of using a Cusum model based on registration data on lodged applications is that the alarms triggered by observed changes may happen too late. In particular, such warnings may not be useful whenever the decisions to increase the processing capacity by national authorities must be urgently taken earlier, in response to the observed increase in inflows of migrants or
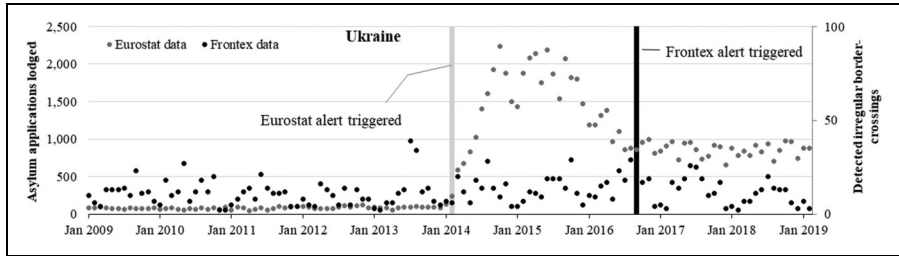
Figure 5. Early Warning Alerts Generated by a Cusum Model Applied to Applications Lodged by Asylum-Seekers (Eurostat Data) and Detections of Irregular Border Crossings (Frontex Data) to Flows From Ukraine to EU28 Between January 2009 and February 2019. The Color of the Bar Corresponds to the Alarm Triggered by Cusum Based on Respective Data.
Source: Eurostat and Frontex.

existing long-lasting backlogs. While recruitment and training of new staff usually last at least a few weeks,[3] any alarm triggered by Cusum based on the monitoring of registry data of lodged applications would alert about the changes in asylum flows post factum.

The comparison of Greek data from Summer 2016 shows this problem clearly. When the asylum crisis started, many asylum-seekers were waiting in reception centers in Greece to lodge their applications.[4] In response to the crisis, the Hellenic Asylum Services progressively increased the number of staff responsible for processing asylum applications, from 218 people at the end of 2014 to 290 staff a year later and 650 workers by January 1, 2017.[5] Looking at the developments in asylum migration trends in this period, we observe that on average 4,000 applications per month were lodged in the first half of 2016, with a peak of 8,000 registered in November 2016 (see Figure 3). That notwithstanding, a large-scale exercise launched in June 2016 to preregister asylum-seekers in mainland Greece, in only the first month of activity, revealed that an additional 15,500 asylum-seekers were present but had not yet had a chance to register their applications.[6]

Looking retrospectively at data on monthly applications lodged in Greece, an early warning system based on the Cusum model with a calibration period of

---

[3] See, for example, the EASO training programme in: EASO Training Newsletter_5th edition. pdf, accessed on July 5, 2021.

[4] See, for example, information provided by United Nations High Commissioner for Refugees: UNHCRGreeceFactSheet01-31August2016.pdf (reliefweb.int); accessed on July 5, 2021.

[5] Hellenic Asylum Service Press Release: greece-ministry-prel.pdf (via: statewatch.org), accessed on July 5, 2021.

[6] See https://www.unhcr.org/news/latest/2016/7/577687af4/15500-asylum-seekers-pre-registered-mainland-greece.html, accessed on July 5, 2021.

2008–2013 would have triggered an alarm in January 2015. In contrast, for a Cusum model based on the number of people apprehended when trying to cross the Eastern Mediterranean EU borders by land or sea, an alert would have been issued a few months earlier (see Figure 3). The latter data are collected by Frontex, the EU external border management agency, and may be better suited for reflecting the temporal pattern of real inflows, although they do not allow users to distinguish between asylum-seekers and other migrant groups (for a data discussion, see Vespe, Natale and Pappalardo 2017). Furthermore, the relationship between the number of apprehensions and the number of successful crossings is not known and may vary over time.

Sensitivity Checks and Perspectives for Early Warnings for Asylum Migration

To better prepare for future asylum inflows, many countries are in the process of creating, or have already built, monitoring systems for asylum flows (Carammia and Dumont 2018). Some countries have created predictive systems utilizing machine learning algorithms (i.e., Sweden; see Berggren and Al-Talibi 2017), while other use historic data and expert knowledge (i.e., Switzerland, see Bijak, Forster and Hilton 2017 for a review). In this section, we show results of tests of the Cusum model's utility in detecting shifts in the level of both applications for asylum and border crossings. Results of this test are shown in Figure 4 for four groups of asylum-seekers: Syrians, Afghans, Iraqis, and Nigerians, who jointly accounted for around 60 percent of the total asylum applications lodged in Europe between 2015 and 2018.

In all four cases, the alarms triggered by early warning systems based on Cusum models set up on Eurostat data happened a few months earlier than for the systems based on the number of detections of irregular border crossings (see Figure 4). In part, the early alarms could be explained by some applications not being filed by applicants for the first time. For example, people who were already present in the EU used the worsening situation in their country to reapply for asylum, even though the new waves of refugees did not arrive until a few months after the situation deteriorated. This explanation is supported by the high shares of non-first-time applicants among Iraqis or Afghans. Another explanation of the alarms triggered too early is that not all asylum-seekers enter the EU irregularly. For instance, the Cusum model based on Frontex data failed to issue any alerts for groups of migrants who may have entered the EU with a visa or under a visa-free scheme (see the example of Ukraine in Figure 5).

Venezuela is another recent example of a situation in which the presence of a visa-free travel regime may hinder the detection of changes in migration trends in some data sources. Because Venezuelan citizens do not need a visa to enter most EU countries,[7] between 2009 and 2019, only eight were detained for trying to cross the border

---

[7] See https://www.etiasvisa.com/, accessed on July 5, 2021.

in an irregular way. Consequently, setting up a Cusum model on the Frontex dataset to detect changes in asylum inflows would be of no use for groups of citizens with similar entry rights to the EU, such as Venezuelans or Ukrainians. By contrast, basing the Cusum model on Eurostat asylum data worked quite well. Prior to 2016, very few applications were lodged, and due to this low average during the calibration period, the first alert was issued in May 2015. This alert reflects the start of Venezuelans' increasing mobility in reaction to their home country's intensifying economic crisis in early 2015, when thousands of people started fleeing their country (UNHCR 2018). Later, new alerts were triggered as the numbers of lodged applications doubled every quarter. In such a case, to avoid multiple alerts generated one by one, it is worth considering (1) the use of a version of the Cusum model in which the upper limit is calculated based on a 12-month moving average and (2) whether to reset the Cusum value to zero every time a new alert is triggered (see Figure 6).

The decision on whether to reset the Cusum values to zero after an alarm is triggered raises substantive questions worth exploring in further investigations. If the variable being monitored is, for example, the logarithm of the daily (or weekly) rate of change, resetting is not necessary. On the other hand, resetting is particularly warranted if the analyst believes that the process based on migration counts moved to a new equilibrium and that its past ceases to matter. This is the case for *nonstationary* processes, whereby the characteristics governing the underlying random process change over time—a feature shared by many migration flows (Bijak et al. 2019). Nonstationarity does not have to signify a complete change of the whole underlying migration *system*, which is rarer (e.g., de Haas 2010), but just a shift of the volume of flows, for example, in response to the changing political, policy, or economic environment. One promising line of future enquiry could look into running two Cusum models in parallel—one for migration counts and one for rates of change—to get a better understanding of the characteristics of the underlying processes.

To externally validate the Cusum model's performance for detecting changes in levels of asylum flows, we compare it to a competing method based on the use of exponentially weighted moving averages (EWMAs) (e.g., Sonesson 2003; Frisén and Sonesson 2006). This method calculates the EWMA at time t, z_t, as follows:

$$z_t = \alpha\, x_t + (1 - \alpha)z_{t-1}, \tag{2}$$

where α is the weight parameter and x_t is the observation of the data at time t (Sonesson 2003). The parameter α takes values between (0, 1]: if it is near zero, all observations have approximately the same weight, while if $\alpha = 1$, only the last observation is used in the alarm statistic (Sonesson 2003, p. 1117). In addition, when the underlying distribution of x_t does not change, the variance of this process can be calculated. Thus, an alarm can be triggered whenever the value of z_t falls above some chosen multiple L of the expected standard deviation at time t (Frisén and Sonesson 2006).
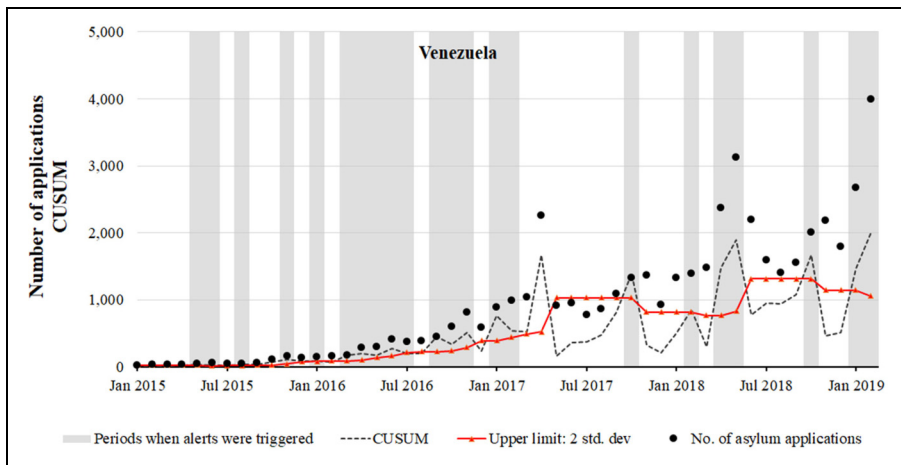
Figure 6. Early Warning Alerts Generated by a Cusum Model Based on a 12-Month Moving Average and Upper Limited Set Up at Two Standard Deviations Applied to Applications Lodged by Asylum-Seekers (Eurostat Data) to Flows from Venezuela to EU28.
Source: Eurostat.

To provide a fair comparison to the Cusum model, the parameters of EWMA, α, and L must be set to reasonable values. Here, these parameters are chosen using tables provided by Lucas and Succucci (1990), which present ranges of optimal parameters for various combinations of the desired false alarm frequency and the size of the mean shift to be detected. To be roughly consistent with the Cusum model, we allow for optimal detection of a mean shift of two standard deviations, while maintaining an average run length of 500 observations for series with no change of level. The estimation results in parameter values of $L = 3.0455$ and $\alpha = 0.365$.

Table 2 gives the month in which an alarm was triggered by the EWMA and Cusum models based on the data from five countries, looking at both application and border apprehension series for each country. Highlighted cells indicate the earliest data for each combination of country and data type. In eight of 10 cases, both monitoring schemes trigger alarms at the same time. In the case of Ukraine, the alarm is one month earlier in the EWMA scheme, although this alarm is triggered in the very first month of the monitoring period. For the case of Afghanistan, using the Frontex data, we see that the Cusum triggers an alarm four months before the EWMA-based scheme. Overall, however, the general agreement between the two approaches gives reasons for confidence in the use of either scheme.

Overall, our experience with the use of the Cusum, or alternative approaches such as EWMA, shows the potential of simple methods as elements of an early warning system. More advanced versions should ideally be based on multiple data sources and require cross-checking of alerts between them. Such analyses would require

Table 2. Dates at Which Alarms Were Triggered by Exponentially Weighted Moving Average (EWMA) and Cusum Monitoring Schemes Using Both Eurostat and Frontex Data for Five Countries.

	Eurostat application data		Frontex border crossing data	
	EWMA alarm date	Cusum alarm date	EWMA alarm date	Cusum alarm date
Ukraine	January 2014	February 2014	September 2016	September 2016
Nigeria	May 2014	May 2014	July 2014	July 2014
Iraq	July 2014	July 2014	March 2015	March 2015
Syria	February 2014	February 2014	May 2014	May 2014
Afghanistan	January 2014	January 2014	April 2015	December 2014

harmonizing the data in a systematic way to ensure consistency by mapping them onto common concepts and definitions.

Conclusions and Recommendations

The results presented in the previous section reveal very encouraging prospects for applying early warning systems as a policy and operational support tool in the context of asylum migration. Depending on the choice of variables and the definition of thresholds, the models can be useful for a range of applications, from simple monitoring of flows to increasing preparedness and enhancing contingency planning. The proposed approach's key feature enables a shift of perspective, from purely reactive toward proactive, in an environment when any advance warning is very important from both humanitarian and resource perspectives. The built-in interactive features necessitate proper exchange and coproduction of knowledge between decision-makers and modelers, whereas the model design and specification can be left to analysts, as setting a threshold at a certain level is ultimately a policy, or even political, decision.

Of course, there are some clear limitations to the proposed approach. Most importantly, in the context of uncertain and volatile asylum migration, the early warning models do not generate proper migration predictions; their main aim is to provide advice to analysts and decision-makers, rather than to predict future migration. Likewise, migration's uncertainty is not reduced by the early warning models, but they do allow for better management of flows. One implication of these limitations of early warnings is that the outcomes of the modeling process, and their inherent uncertainty, must be communicated clearly and honestly, also with respect to what the models cannot achieve. Another consideration is that formal monitoring of trends to identify early warnings is never a finished job, but rather a continuously evolving process of successive model adjustments. Finally, as mentioned before,

statistical modeling cannot be the only component of a wider early warning system, even though it can be very important and offer unique insights.

Another set of challenges related to early warning models concerns data availability, comparability, quality, and completeness. As mentioned above, data limitations, such as the use of the lodging of applications, rather than arrivals, possible double-counting of applicants, the presence of applications from people already in the country, and so on, require additional care when setting up early warning models and interpreting their results. To that end, a statistical description of the data collection processes and their errors can be beneficial, providing additional insights into the uncertainty of the underlying processes.

Despite these limitations, the approach presented in this article is flexible and can be extended in several ways. In terms of exploring the potential of alternative data sources, there may be additional mileage in using "big data," for example, from a mobile phone or social media "digital footprints," in addition to more traditional data sources, such as administrative records. Migration-related applications of "big data" and "digital traces" are currently gaining traction, with several available examples, from a feasibility study related to labor migration by Hughes et al. (2016) to an analysis of flows driven by a natural disaster (Puerto Rican migration following Hurricane Maria, Alexander, Polimis and Zagheni 2019) or to migration with a substantial asylum component (Venezuelan "exodus" in Palotti et al. 2020). The "new data" situation relevant for Europe has been recently reviewed by Spyratos et al. (2018), and a general overview of the big data perspective on migration is provided in Sîrbu et al. (2021).

Besides, some ongoing projects aim to gain more understanding of refugee mobility by monitoring social media (e.g., Twitter or Instagram posts by refugees; Iacus 2017), conflict events (e.g., based on the GDELT data—https://www.gdeltproject.org—as done at EASO (2016b) or mobile phone Call Details records to better understand refugees' secondary movements, see Sterly et al. 2019), or satellite imagery (e.g., Satellite Sentinel Project—http://www.satsentinel.org—or Migration Radar 2.0). Other projects look at collecting information on web searches for asylum/migration-related topics. For example, Böhme, Gröger and Stöhr (2020) recently found that georeferenced internet search data can effectively predict bilateral migration flows. Reconciling the differences between various sources remains a key challenge for future model development.

Even though alternative sources on their own are unlikely to enable estimation of the *levels* of migration flows and eliminate the need for traditional data, their potential should not be underestimated. First, asylum-seekers' propensity to rely on mobile technologies is well documented (Kingsley 2016), so their journeys are likely to leave many "digital traces," enabling an analysis of flows along different routes, subject to privacy constraints. Second, these sources offer unique and timely data, which may be able to detect changes in trends at much finer-grained time scales than is the case with traditional data. Indeed, a uniquely appealing feature of "digital traces" is their potentially finer time-granularity, in comparison with even

the highest-frequency registrations (e.g., Spyratos et al. 2018). As mentioned before, current developments at EASO aim to include some of those data to boost the EU's asylum preparedness system (for a prototype work, see Carammia, Iacus and Wilkin 2020).

Subject to ensuring that the ethical, privacy, and data protection standards are appropriately maintained and that the information is safeguarded against "dual use" by malevolent actors, there is much potential in developing formal methods for linking traditional and nontraditional data sources. In the case proposed here, a natural starting point would be to feed the digital trace data into a change-point-detection model, such as (1) or (2), to identify and predict possible changes in patterns. As many such data sources can have arbitrarily fine time-granularity, an important caveat is that calibrating the warning thresholds depending on the data frequency would require attention so that the models do not "overreact." After all, a threshold of "one-in-a-hundred-observations" would randomly trigger an unwarranted alert on average once every two years for weekly data, once every three to four months for daily data, and every four days for hourly data.

A second natural extension of the proposed approach would be to move from country-specific early warning models to an interlinked multicountry and multi-source early warning system so that the advantages of a more general migration mod-eling framework could be utilized more fully (see the Online Appendix for examples). In particular, the elements of such a system could additionally include information on policy changes and some external variables, such as intensity of con-flict in the origin regions, some of which may be available instantaneously or only with a short delay. In this context, the input of policy or country-specific experts into the modeling process becomes even more important: detailed domain knowledge can not only help fine-tune the system but also provide more holistic insights that are not available from quantitative data alone.

Yet another extension could consist of combining early warning models with computer simulations to provide an interactive framework for stress-testing various policy and operational responses to changes in migration flows. Even though the design and implementation of such models can be resource-intensive, simulation approaches can offer unique insights into systems with many interacting elements and a prominent role of human agency. Prototype simulation models for other forms of migration exist, including environmental (Smith et al. 2008) and labor migration (Klabunde and Willekens 2016), which could serve as starting points for developing a similar model specifically tailored for asylum.

In terms of recommendations for users of the models and migration management practitioners, three main lessons emerge from the modeling exercise presented here. First, uncertainty in migration can be managed to some extent by using appropriate tools. Second, making model building work must be a continuously updated process, rather than a one-off event. Third, honest, open, and transparent two-way communi-cation between analysts and model users is a prerequisite for their successful practical application. Still, given the high volatility of asylum flows, the very high stakes

involved in asylum migration, and the need for rapid humanitarian, operational, and policy responses, a methodological movement toward a proactive early warning system enabling more informed contingency planning is already long overdue.

Data Availability Statement

The publicly available Frontex data used in our analysis can be obtained from https://frontex. europa.eu/along-eu-borders/migratory-map/ and the Eurostat statistics on asylum are at https:// ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Application_for_ international_protection (as of August 1, 2020). Data from Greece and Italy can be obtained from the respective authorities upon request.

Declaration of Conflicting Interests

JN is a former staff member of EASO, at the time of writing employed at the EC Joint Research Centre, Brussels, and currently at Cedefop, Thessaloniki. MC is also a former staff member of EASO, currently working at the University of Catania. JB, JJF, and JH declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ORCID iD

Jakub Bijak https://orcid.org/0000-0002-2563-5040

Supplemental Material

Supplemental material for this article is available online at https://journals.sagepub.com/ home/mrx.

References

Alexander, M., K. Polimis, and E. Zagheni. 2019. "The Impact of Hurricane Maria on Out-Migration From Puerto Rico: Evidence From Facebook Data." *Population and Development Review* 45(3):617–30.

Arango, J. 2000. "Explaining Migration: A Critical View." *International Social Science Journal* 52(165):283–96.

Azose, J., and A. E. Raftery. 2015. "Bayesian Probabilistic Projection of International Migration." *Demography* 52(5):1627–50.

Bakewell, O. 1999. "Can we Ever Rely on Refugee Statistics?" *Radical Statistics* 72, article 1. Accessed June 1, 2019. http://www.radstats.org.uk/no072/article1.htm.

Barker, E., and J Bijak. 2020. "Conceptualisation and Analysis of Migration Uncertainty: Insights from Macroeconomics." QuantMig Project Deliverable D9.1. Southampton: University of Southampton. Accessed July 1, 2021. http://quantmig.eu/res/files/QuantMigD9.1AnalysisofMigrationUncertaintyV1.1.pdf.

Barnard, G. A. 1959. "Control Charts and Stochastic Processes." *Journal of the Royal Statistical Society* 21(2):239–71.

Beine, M., A. Boucher, B. Burgoon, M. Crock, J. Gest, M. Hiscox, P. McGovern, H. Rapoport, J. Schaper, and E. Thielemann. 2016. "Comparing Immigration Policies: An Overview From the IMPALA Database." *International Migration Review* 50(4):827–63.

Berggren, S., and H. Al-Talibi. 2017. "Long-term and short-term migration predictions." Presentation during *The Early Warning and Preparedness System (EPS) Network Advisory Group—Early Warning and Forecasting Methods in Migration*, September 21–22, 2017, EASO, Malta (Mimeo).

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

Bijak, J., and M. Czaika. 2020. "Black Swans and Grey Rhinos: Migration Policy Under Uncertainty." *Migration Policy Practice* X(4):14–20. Accessed April 1, 2021. https://publications.iom.int/system/files/pdf/mpp-43.pdf.

———, G. Disney, A. M. Findlay, J. J. Forster, P. W. F. Smith, and A. Wiśniowski. 2019. "Assessing Time Series Models for Forecasting International Migration: Lessons From the United Kingdom." *Journal of Forecasting* 38(5):470–87.

———, J.J. Forster, and J. Hilton. 2017. "Quantitative assessment of asylum-related migration: A survey of methodology." Report for the European Asylum Support Office. Malta: EASO. Accessed February 1, 2021. https://www.easo.europa.eu/sites/default/files/publications/Quantitative_assessment_of_asylum_related_migration_1.pdf.

———, and A. Wiśniowski. 2010. "Bayesian Forecasting of Immigration to Selected European Countries by Using Expert Knowledge." *Journal of the Royal Statistical Society A* 173(4):775–96.

Böcker, A., and T. Havinga. 1997. "*Asylum Migration to the European Union. Patterns of Origin and Destination*." Luxembourg: Office for Official Publications of the European Communities. ISBN: 92-828-3306-2.

Böhme, M. H., A. Gröger, and T. Stöhr. 2020. "Searching for A Better Life: Predicting International Migration with Online Search Keywords." *Journal of Development Economics* 142(January):102347.

Bongaarts, J., and R. Bulatao, eds. 2000. *Beyond Six Billion: Forecasting the World's Population*. Washington, DC: National Academy Press. ISBN: 978-0-309-06990-4.

Carammia, M., and J.C. Dumont. 2018. "Can We Anticipate Future Migration Flows?" *Migration Policy Debates*, OECD Migration Policy Debates, No. 16. Accessed August 6, 2020. https://www.oecd.org/els/mig/migration-policy-debate-16.pdf.

———, S. M. Iacus, and T. Wilkin. 2020. "Forecasting Asylum Applications in the European Union with Machine Learning and Data at Scale." Preprint, ArXiv:2011.04348. Accessed April 15, 2021. http://arxiv.org/abs/2011.04348.

Carling, J., and F. Collins. 2018. "Aspiration, Desire and Drivers of Migration." *Journal of Ethnic and Migration Studies* 44(6):909–26.

Castles, S. 2004. "Why Migration Policies Fail." *Ethnic and Racial Studies* 27(2):205–27.

Connor, P. 2017. "The Digital Footprint of Europe's Refugees." Pew Research Center, June. Accessed July 24, 2020. https://www.pewresearch.org/global/wp-content/uploads/sites/2/2017/06/Pew-Research-Center_Digital-Footprint-of-Europes-Refugees_Full-Report_06.08.2017.pdf.

Crisp, J. 1999. "Who has counted the refugees? UNHCR and the politics of numbers." New Issues in Refugee Research, No. 12, Geneva: UNHCR. ISSN 1020-7473. Accessed July 24, 2020. https://www.refworld.org/pdfid/4ff58e4b2.pdf.

De Haas, H. 2010. "The Internal Dynamics of Migration Processes: A Theoretical Inquiry." *Journal of Ethnic and Migration Studies* 36(10):1587–617.

EASO (European Asylum Support Office). 2016b. "EASO Single Programming Document. Multiannual Programming 2017–-2019, Work Programme 2019." Accessed June 1, 2019. https://www.easo.europa.eu/sites/default/files/easo-spd-2017-2019_WP%202017%20-%20am1.pdf.

EASO (European Asylum Support Office). 2016a. *The Push and Pull Factors of Asylum-Related Migration. A Literature Review. Report prepared by Maastricht University and the Global Migration Data Analysis Centre (GMDAC) for the European Asylum Support Office*. Luxembourg: Publications Office of the European Union. Accessed July 24, 2020. https://www.easo.europa.eu/sites/default/files/publications/The%20Push%20and%20Pull%20Factors%20of%20Asylum%20-%20Related%20Migration.pdf.

ECRE (European Council on Refugees and Exiles). 2016. "The Length of Asylum Procedures in Europe." Report, European Council on Refugees and Exiles—Asylum Information Database (aida). Accessed April 1, 2021. https://www.ecre.org/wp-content/uploads/2016/10/AIDA-Brief-DurationProcedures.pdf.

Erdal, M. B., and C. Oeppen. 2018. "Forced to Leave? The Discursive and Analytical Significance of Describing Migration as Forced and Voluntary." *Journal of Ethnic and Migration Studies* 44(6):981–98.

European Commission. 2016. "*European Commission—Fact Sheet. Reforming the Common European Asylum System: Frequently Asked Questions*." Brussels: European

Commission, 13 July 2016. Accessed June 1, 2019. http://europa.eu/rapid/press-release_MEMO-16-2436_en.htm.

FitzGerald, D. S. 2015. "The Sociology of International Migration." In *Migration Theory, Talking Across Disciplines*, third edition, edited by C. B. Brettell and J. F. Hollifield, 115-47. New York: Routledge. ISBN: 978-0-415-74203-0.

————, and R. Arar. 2018. "The Sociology of Refugee Migration." *Annual Review of Sociology* 44:387–406.

Foresight. 2011. "Migration and Global Environmental Change. Future Challenges and Opportunities." Final Project Report. London: Government Office for Science. Accessed July 24, 2020. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/287717/11-1116-migration-and-global-environmental-change.pdf.

Frisén, M., and C. Sonesson. 2006. "Optimal Surveillance Based on Exponentially Weighted Moving Averages." *Sequential Analysis* 25(4):379–403.

Geddes, A. 2005. "Europe's Border Relationships and International Migration Relations." *JCMS: Journal of Common Market Studies* 43(4):787–806.

————, and P. Scholten. 2016. *The Politics of Migration and Immigration in Europe*. London: Sage. ISBN: 978-1-84920-467-5.

Haddad, E. 2008. *The Refugee in International Society: Between Sovereigns*. Cambridge: Cambridge University Press. ISBN: 978-0-511-49135-1.

Harrison, P. J., and O. L. Davies. 1964. "The Use of Cumulative Sum (Cusum) Techniques for the Control of Routine Forecasts of Product Demand." *Operations Research* 12(2):325–33.

Harrison, J. P., and P. Veerapen. 1994. "A Bayesian Decision Approach to Model Monitoring and Cusums." *Journal of Forecasting* 13(1):29–36.

Hatton, T. J. 2020. "Asylum Migration to the Developed World: Persecution, Incentives, and Policy." *Journal of Economic Perspectives* 34(1):75–93.

Hughes, Ch., E. Zagheni, G.J. Abel, A. Wiśniowski, A. Sorichetta, I. Weber, and A.J. Tatem. 2016. *Inferring Migrations: Traditional Methods and New Approaches Based on Mobile Phone, Social Media, and other Big Data. Feasibility Study on Inferring (Labour) Mobility and Migration in the European Union from Big Data and Social Media Data. Report for the European Commission*. Luxembourg: Publications Office of the EU. ISBN: 978-92-79-59742-8.

Iacus, S. M. 2017. "Big Data and Migration: Limits and Perspectives". Presentation delivered at the conference on Big Data and Alternative Data Sources on Migration: from case studies to policy support, JRC-ISPRA, 30 November 2017. Mimeo.

Kegels, M. 2016. *Getting the Balance Right: Strengthening Asylum Reception Capacity at National and EU Levels*. Brussels: Migration Policy Institute Europe. Accessed April 1, 2021. https://www.migrationpolicy.org/research/getting-balance-right-strengthening-asylum-reception-capacity-national-and-eu-levels.

King, R. 2002. "Towards A New Map of European Migration." *International Journal of Population Geography* 8(2):89–106.

Kingsley, P. 2016. *The New Odyssey: The Story of Europe's Refugee Crisis*. London: Faber & Faber.

Klabunde, A., and F. J. Willekens. 2016. "Decision-making in Agent-Based Models of Migration: State of the art and Challenges." *European Journal of Population* 32(1):73–97.

Lang, J.H., T.A. Peltonen, and P. Sarlin. 2018. "A Framework for Early-Warning Modeling with an Application to Banks." European Central Bank Working Paper No. 2182 (October). Frankfurt: ECB. Accessed July 24, 2020. https://www.ecb.europa.eu/pub/pdf/scpwps/ecb.wp2182.en.pdf.

Lee, E. S. 1966. "A Theory of Migration." *Demography* 3(1):47–57.

Li, Zh, P. Qiu, S. Chatterjee, and Zh. Wang. 2013. "Using P Values to Design Statistical Process Control Charts." *Statistical Papers* 54:523–39.

Lorden, G. 1971. "Procedures for Reacting to A Change in Distribution." *Annals of Mathematical Statistics* 42(6):1897–908.

Lucas, J. M., and M. S. Saccucci. 1990. "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements." *Technometrics* 32(1):1–12.

Massey, D. S., J. Arango, G. Hugo, A. Kouaouci, A. Pellegrino, and J. E. Taylor. 1993. "Theories of International Migration: Review and Appraisal." *Population and Development Review* 19(3):431–66.

Moustakides, G. V. 1986. "Optimal Stopping Times for Detecting Changes in Distributions." *The Annals of Statistics* 14(4):1379–87.

NASEM [National Academies of Sciences, Engineering, and Medicine]. 2019. *Forced Migration Research: From Theory to Practice in Promoting Migrant Well-Being: Proceedings of A Workshop*. Washington, DC: The National Academies Press.

Öberg, S. 1996. "Spatial and Economic Factors in Future South-North Migration." In *The Future Population of the World: What Can We Assume Today?* edited by W. Lutz, 361-85. London: Earthscan. ISBN: 1-85383-239-1. Accessed July 24, 2020. http://pure.iiasa.ac.at/id/eprint/4762/1/XB-96-003.pdf.

Owen, D. 2020. *What Do We Owe to Refugees?* Cambridge: Polity Press. ISBN: 978-1-509-53974-1.

Page, E. S. 1954. "Continuous Inspection Schemes." *Biometrika* 41(1):100–15.

———— 1957. "A Test for a Change in a Parameter Occurring at an Unknown Point." *Biometrika* 44(1/2):248–52.

Palotti, J., N. Adler, A. Morales-Guzman, J. Villaveces, V. Sekara, M. Garcia Herranz, M. Al-Asad, and I. Weber. 2020. "Monitoring of the Venezuelan Exodus Through Facebook's Advertising Platform." *PLoS ONE* 15(2):e0229175.

Pijpers, R. 2008. "Problematising the 'Orderly' Aesthetic Assumptions of Forecasts of East–West Migration in the European Union." *Environment and Planning A* 40(1):174–88.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, Accessed July 1, 2020. https://www.R-project.org.

Regehr, E. 2014. "Armed Conflict: Trends and Drivers." Background Paper for the Workshop 'Civil Society Views on Next Generation Peacebuilding and Conflict Prevention Policy and Programming Issues and Responses', Ottawa, 14.03.2011. Accessed July 24, 2020. https://www.peacebuild.ca/Microsoft%20Word%20-%20TRENDS%20IN%20GLOBAL%20CONFLICT.170511.docx%20-%20Ernie%20Regehr.pdf.

Shellman, S. M., and B. M. Stewart. 2007. "Predicting Risk Factors Associated with Forced Migration: An Early Warning Model of Haitian Flight." *Civil Wars* 9(2):174–99.

Singleton, A. 2016. "Migration and Asylum Data for Policy-Making in the European Union. The Problem with Numbers." *CEPS Paper* No. *89*, Centre for Europe and Policy Studies, Brussels. ISBN 978-94-6138-509-3. Accessed July 24, 2020. https://www.ceps.eu/download/publication/?id=9357&pdf=LSE%2089%20AS%20Migration%20and%20Asylum%20Data.pdf.

Sîrbu, A., G. Andrienko, N. Andrienko, C. Boldrini, M. Conti, F. Giannotti, and R. Guidotti, 2021. "Human Migration: The Big Data Perspective". *International Journal of Data Science and Analytics* 11(4):341–60. DOI: 10.1007/s41060-020-00213-5.

Smith, Ch., D. Kniveton, S. Wood, and R. Black. 2008. "Predictive Modelling." *Forced Migration Review* 31:58–9. Accessed July 24, 2020. https://www.files.ethz.ch/isn/139362/FMR31full.pdf

Sohst, R., J. Tjaden, H. de Valk, and S Melde. 2020. "The Future of Migration to Europe: A Systematic Review of the Literature on Migration Scenarios and Forecasts." Geneva: IOM —International Organization for Migration, and The Hague: Netherlands Interdisciplinary Demographic Institute. Accessed April 1, 2021. https://publications.iom.int/system/files/pdf/the-future-of-migration-to-europe.pdf.

Sonesson, C. 2003. "Evaluations of Some Exponentially Weighted Moving Average Methods." *Journal of Applied Statistics* 30(10):1115–33.

Spyratos, S., M. Vespe, F. Natale, I. Weber, E. Zagheni, and M. Rango. 2018. "Migration data using social media: a European perspective." Joint Research Centre Technical Report. Luxembourg: Publications Office of the European Union. ISBN 978-92-79-87989-0. Accessed July 1, 2020. https://publications.jrc.ec.europa.eu/repository/bitstream/JRC112310/facebook_study_v3.pdf.

Sterly, H., B. Etzold, L. Wirkus, P. Sakdapolrak, J. Schewe, C.-F. Schleussner, and B. Hennig. 2019. "Assessing Refugees' Onward Mobility with Mobile Phone Data—A Case Study of (Syrian) Refugees in Turkey". In *Guide to Mobile Data Analytics in Refugee Scenarios*, edited by A. Salah, A. Pentland, B. Lepri, and E. Letouzé, 251-64. Cham: Springer. ISBN 978-3-030-12553-0.

Suriyakumaran, A., and Y. Tamura. 2016. "Asylum Provision: A Review of Economic Theories." *International Migration* 54(4):18–30.

Tartakovsky, A. G., and G. V. Moustakides. 2010. "State-of-the-Art in Bayesian Changepoint Detection." *Sequential Analysis* 29(2):125–45.

UNHCR (United Nations High Commissioner for Refugees) 2018. "Situation Overview IV: Venezuelan asylum seekers and migrants living outside of shelters, Boa Vista city— October 2018." UNHCR, European Union Humanitarian Aid and Civil Protection, and REACH. Accessed June 1, 2019. https://reliefweb.int/sites/reliefweb.int/files/resources/69194.pdf.

Van Hear, N., O. Bakewell, and K. Long. 2018. "Push-pull plus: Reconsidering the Drivers of Migration." *Journal of Ethnic and Migration Studies* 44(6):927–44.

Vespe, M., F. Natale, and L. Pappalardo. 2017. "Data sets on irregular migration and irregular migrants in the European Union." *Migration Policy Practice* VII(2):26–33. http://www.eurasylum.org/wp-content/uploads/2017/08/MPP-30.pdf.

West, M., and P. J. Harrison. 1986. "Monitoring and Adaptation in Bayesian Forecasting Models." *Journal of the American Statistical Association* 81(395):741–50.

Yashchin, E. 1993. "Statistical Control Schemes: Methods, Applications and Generalizations." *International Statistical Review/Revue Internationale de Statistique* 61(1):41.

Zeileis, A. 2004. "Alternative Boundaries for CUSUM Tests." *Statistical Papers* 45:123–31.

Zolberg, A. R. 1989. "The Next Waves: Migration Theory for A Changing World." *International Migration Review* 23:403–30.