National Bank of Ukraine

NBU Working Papers

03/2024

The Warcast Index: Estimating Economic Activity without Official Data during the Ukraine War in 2022

Mihnea Constantinescu

Kalle Kappner

Nikodem Szumilo

The NBU Working Papers

This publication is available on the NBU's website at www.bank.gov.ua.

Address:

9 Instytutska Street

01601, Kyiv, Ukraine

research@bank.gov.ua

The Warcast Index:
Estimating Economic Activity without Official Data during the Ukraine War in 2022[1]

Mihnea Constantinescu[2], Kalle Kappner[3], Nikodem Szumilo[4]

Abstract

We introduce the Warcast Index, an approach for estimating regional economic activity during periods of extreme uncertainty using publicly available data. We show that combining widely used correlates of economic activity – nightlight intensity, Google Trends, and Twitter activity – can improve the tracking of economic performance and even allow the approximation of monthly economic activity after extreme structural breaks, like war or occupation. We apply this approach to Ukraine during the 2022 war. Our findings show that combining multiple data sources not only improves tracking accuracy compared to single-correlate models, but also provides timely, transparent and flexible data for policy-making in situations where conventional economic data is unavailable or unreliable. We also contribute to the literature on wartime economics by providing a novel analysis of the economic effects of armed conflict with high frequency (monthly) and spatially granular (regional) data.

Keywords: Estimating GDP, Nowcasting GDP, Wartime economics, Nightlights, Google Trends, Twitter data

JEL Codes: B41, C82, E01, O11

[1] We are grateful to Martin Cesnak, Peter Toh, Jan Klasco, Artem Kochnev, Augustin Indaco, Nikhil Ray, Zsoka Koczan, Pawel Ponsko, Pawel Baranowski, Jaroslaw Cwiek-Karpowicz, Robert Beyer, Alexander Plekhanov, Alisa Gnydyuk, Yiqi Huang, Florain Blum, Sanaa Nadeem and Mariia Sydorovych as well as participants of the 90th East Jour Fixe workshop at the National Bank of Austria, the National Bank of Slovakia seminar, the Atlanta Federal Board of Reserve seminar and the EBRD seminar and the National Bank of Ukraine Seminar for their thoughts and comments on this work.
[2] National Bank of Ukraine and University of Amsterdam, m.constantinescu@uva.nl, https://orcid.org/0000-0002-2700-2589.
[3] Ludwig-Maximilians-Universität München, kallekappner@googlemail.com.
[4] University College London., n.szumilo@ucl.ac.uk.

1. Introduction

Violent conflicts shroud economic data not just in areas of direct impact, but also in adjacent regions where the rule of law weakens and reporting and record keeping are unlikely to be prioritized. The concept of “the fog of war”, originating in military terminology, encapsulates the uncertainty and lack of clarity in situational awareness experienced during violent conflicts, when crucial information is unavailable. Initially used to describe the complexities in military decision-making without fundamental information, this concept equally applies to the realm of economic decision-making. The unavailability of even the most basic data poses a significant challenge during any crisis (Qadir et al., 2016), but it becomes particularly problematic during a war. Effective policies can make a substantial difference, not only in terms of the long-term impact of the war on the economy and society, but also in contributing to the overall military effort. In this paper, we propose the development of an economic activity index that allows for monthly tracking at a regional level using publicly available data (nightlight intensity, Google Trends and Twitter activity) that remains accessible even during armed conflicts.

Our main contribution is to the growing literature on tracking economic activity when official data is unavailable or unreliable (because it is manipulated, not reported, or published with a lag). These methods rely on tracking variables correlated to economic activity, such as nightlights or social media activity. We develop this literature in several ways. First, we expand existing methods to track activity at higher frequency for smaller geographical units. Second, we show that different variables used by the literature yield predictions with independent prediction errors.[5] In other words, we find that prediction errors based on different correlates of economic activity are not strongly correlated to each other. This means that combining different predictors can make predictions more accurate. Finally, we show that even in the presence of likely structural breaks caused by war, economic activity measured using our mixture of non-traditional methods offers similar predictions to conventional methods (see Caverley (2018) or Chupilkin and Koczan (2022) for examples) but is faster, more transparent and more flexible.

We also contribute to the literature on the economic effects of armed conflict. While this literature focuses heavily on long and very long term effects of war (EBRD, 2022; Nordhaus, 2002; Bellows and Miguel, 2006; Davis and Weinstein, 2002), this is the first paper to study the first months of a conflict. Indeed, during the first months of the conflict in Ukraine in 2022, our index provided critical information for policy decisions. It offers three important advantages over alternative methods.

First, it is transparent and can be easily replicated. This is especially useful for nongovernment or international organizations, which may have limited access to local information. Since we share our code and use data that is publicly available even during an armed conflict, the results are more

[5] Throughout the paper we refer to values that are calculated on the basis of input data and model parameters as “predictions”. “Prediction errors” are therefore the difference between realized outcomes and model output in the same area and time period. Different literatures we draw upon often use inconsistent definitions of key technical terms. For clarity, we provide a glossary of our definitions (based mainly on applied microeconomics research) in Appendix C.1.

transparent than institutional statistics (from governments or NGOs) that are hard to verify. This matters especially in countries that struggle with corruption. In contrast, traditional methods used by large NGOs often rely on expert judgments. Moreover, NGOs do not usually publish detailed methodologies.

Second, our index can be produced with a considerably shorter time lag than alternatives. Our preliminary results were available just over two months after the full scale war in Ukraine began. Once the index is set up, it can be updated with a lag of two weeks, which is the frequency of the data we use being updated. This offers a considerable advantage over traditional methods that during a war may publish (preliminary) results with a lag of several months or longer.

Third, our index is flexible in assumptions about the territory considered in the analysis. It can be flexible about the boundaries of the country or region it focuses on. Importantly, the index can still be created if an area (or a part of it) is under occupation.

While our approach offers considerable benefits over existing alternatives, it suffers from several important issues typical to this literature, such as the lack of established micro foundations, susceptibility to structural breaks, technical data collection challenges, or favoring more developed regions. We explore these issues in the paper. It is also worth noting that the mix of indicators we use in our index may not be a perfect match for all conflict areas.[6] However, our key point – that combining different indicators is better than using a single one – is likely to remain valid.

The starting point for our index is the work on the correlation between GDP and nightlight intensity pioneered by Henderson et al. (2011). For over two decades, economists who were unable to track annual GDP using conventional metrics, have been using nightlights as a proxy. We extend this approach to a higher frequency following approaches developed during the COVID-19 pandemic (Stokes and Román, 2022). However, we find that during a war nightlights are likely affected by curfews, intentional manipulation, or fires, and it is unclear how these relate to economic activity.

To address this problem, we combine nightlight-based predictions with other models. Specifically, we use recent developments in GDP tracking and focus on the relationship between online behavior and economic activity. We begin with using Twitter (X) activity to track GDP as suggested by Indaco (2020). Indeed, we find that before the war, the number of tweets from a location correlated very highly to its economic activity and that there was a statistically significant relationship between changes in local economic activity and Twitter behavior. The second internet-based measure we use is tracking Google Trends. Google's virtual monopoly in the

[6] We acknowledge that the coverage of nightlights intensity data is not perfect for many countries due to satellite paths not being optimized for constant coverage (Beyer et al., 2018), Google data is not useful in places where access to Google is restricted, such as China, and Twitter (X) data differs in popularity and ease of access over time and space (technically current use of the data should refer to X data, but our agreement was with Twitter). However, alternative indicators that have been shown to correlate to GDP that were not available in Ukraine are available elsewhere, including pollution data (Bricongne et al., 2021) or mobility data (Caverley, 2018).

search market and the tendency of modern populations to seek information about most of their choices on the internet results in population-level behavior and sentiments being reflected in Google searches. Woloszko (2020) tracks national GDP on a weekly basis, and shows that this approach performed very well during the pandemic. We extend this idea, and develop regional indexes.

Naturally, combining predictions of different models will only give a better prediction of economic activity than a single model if their prediction errors have a low correlation. This is a testable assumption and we show that it holds in data from Ukraine for both in-sample and out-of-sample predictions.

Our empirical evidence from Ukraine allows us to demonstrate many important features of the index and validate modeling assumptions, but it also offers unprecedented evidence from an economy suffering from a full-scale invasion. The case we study provides not only data from a relatively modern economy, but also during two stages of an invasion (a partial invasion in 2014 and a full-scale invasion attempt in 2022). This means that in our data we have two relatively similar structural breaks. We use the earlier one to validate our out-of-sample predictions. We show that out-of-sample predictions match historical outcomes even despite the structural changes that occurred in 2014. Out-of-sample predictions in 2022 are more difficult to validate, as no alternative monthly and regional GDP data exists. However, our results are very close to national quarterly accounts and correlate very highly (negatively) to a regional index of war destruction.

The final benefit of the results we obtain in Ukraine is that we can estimate changes in individual elasticities between economic activity and nightlights, Twitter or Google Trends. This could be helpful in settings where only one of these indicators is available.

We do not claim that the index is a perfect proxy of economic activity. While the true level of economic activity during a war at our adopted frequency and geographical level will likely never be measured with great precision, our index is the first attempt to measure it at all. It is therefore a critical tool for policymakers who need to make decisions during an armed conflict or in the run up to one.

The rest of this paper is structured as follows: Section 2 reviews the relevant literature, Section 3 introduces the institutional setting of the war in Ukraine, Section 4 outlines our data, Section 5 describes our methodology, Section 6 presents and discusses our results, while Section 7 provides concluding remarks.

2. Tracking GDP Using Its Correlates

Our starting point is the well-established literature linking measures of light at night to human activity, as pioneered by Croft (1973, 1978). Motivated by the relationship between electricity consumption, employment, income and output, several authors have identified a close relationship at the sub-national level between selected economic variables and luminosity, as found for example in Elvidge et al. (1997), Doll et al. (2006), Sutton (2007), Ghosh et al (2010). Aggregate, national level studies, for example the work of Elvidge et al. (2009), Chen and Nordhaus (2011), and Henderson et al. (2012), further support these findings.

These studies show a high correlation between various measures of nighttime lights (such as average luminosity, sum of measured radiance, and other transformations of the raw nighttime light intensity data) and key economic variables. They also make an important distinction between the nature of the relationship at regional and national levels.

When the scale of analysis is national, most authors employ time-series or panel specifications relating *growth rates* in luminosity and GDP. A robust correlation emerges between luminosity and output growth rates in a large cross-section of countries, including both developed and developing ones. When considering finer geographical details, such as regions or large urban areas, the relationship in growth rates loses its stability and strength, with regression parameters being orders of magnitude smaller. At higher spatial resolutions, economic activity and nightlights show a strong correlation in *levels*, but not in growth rates.

Recent studies investigate this empirical (in)stability for a number of economic variables as a function of scale, as indicated for example in Addison and Stewart (2015), and Bickenbach et al. (2016). A further relevant recent publication by Asher et al. (2021), presenting a detailed database of nighttime lights for India, reinforces this point. The reason for the lack of robustness at different scales is also investigated by Nordhaus and Chen (2014), who attribute it to a mix of sampling and measurement errors in GDP and nighttime lights, varying in proportion between developed and developing economies. We address this issue by modeling levels of GDP and adopting a very general specification that assumes the same elasticity in a cross-section and over time in all regions. While this specification fits large urban regions relatively well, this approach comes at the cost of being less accurate in smaller regions (see Section 5.1 for details).

We also rely on studies that study the relevance of social media activity in predicting levels of income, consumption, investment and employment. Llorente et al. (2015), analyzing the content of over 146 million geo-located Twitter messages, provide evidence that the correctness of language used in tweets and diversity in mobility patterns across Spanish regions strongly correlate with unemployment. More specifically, higher shares of misspellings in a region (which are not abbreviations), are strongly correlated with unemployment in that region.

In a similar vein, Ortega-Bastida et al. (2021) predict regional GDP using NLP (Natural Language Processing) algorithms to distill textual information contained in tweets. These become inputs in a neural-network used to predict regional GDP levels.

Considering country-level studies, Indaco (2020) combines geo-located tweets containing a picture and nighttime lights to estimate national GDP. This study highlights the two alternative data sources, tweets containing an image and nighttime radiance, contain different statistical information relevant for estimating aggregate income. Bokányi et al. (2017), using the time of the day in which tweets are published, show novel measures of employment and unemployment in U.S. counties – a strategy employed also by Indaco (2020) to investigate the potential mechanism explaining why tweets containing a picture may act as proxies of income.

Finally, we build on the literature founded by Ettredge et al. (2005) and further expanded by Bartlett and Varian (2009) revealing the usefulness of Google search data in calculating, nowcasting and near-term forecasting of a wide range of economic variables. Google Trends provide a numeric range from 1 to 100, for a selected period in time and geographical area, of the relative ranking of specific keywords (for example the exact phrase “car”) or categories of keywords (for example all search phrases that refer to a car such as “car”, “automobile”, “clunker”, “truck” etc.). The widespread coverage of topics and categories, along with their very high frequency, led to a range of new methodological and policy applications. Early applications include McLaren and Shanbhogue (2011) who use the volume of online searches to track labor and housing markets in the United Kingdom, and Askitas and Zimmermann (2009), who validate the use of Internet searches during the 2008 Great Recession to better, pinpoint the contraction in the German labor market.

Goetz and Knetsch (2019) investigate the information content of Google Trends in well-established bridge equation models and outline their utility in improving the forecast of both GDP and disaggregated production-side components at various forecasting horizons. The authors show that Google Trends time-series may be an alternative to surveys when used in conjunction with hard data. They rely on LASSO and a suite of alternative variable selection algorithms to identify relevant trends according to the considered target.

Woloszko (2020) employs both Google categories and topics to develop real-time estimates of GDP for OECD countries. He shows that Google search data was an essential ingredient in improving GDP tracking and nowcasts during the onset of the COVID-19 pandemic and the recovery that followed.

It is worth noting that all of the above literature relies on the strong correlations found between relevant economic outcomes and unconventional proxies. While many arguments exist to explain this correlation, there are no established micro-foundations that would link economic choices directly to input data and outcomes. This is in contrast to the indirect high frequency nowcasting literature, which attempts to find alternative measures of variables that determine the nowcasted variables (Bok et al., 2018; Mazzi et al., 2014; Proietti et al., 2021).

3. Theoretical Framework

Consider a time series of economic activity measurements Y_t. If direct measures of economic activity (e.g. GDP) are published with a substantial lag or are unavilable, it is useful to approximate it using a more timely or available correlate X_t, e.g. nightlight activity. Following Henderson et al. (2011, 2012), the literature has focused on the case where economic activity and its correlate have a constant elasticity, i.e.

$$y_t = \beta x_t + \epsilon_t, \quad E[\epsilon_t] = 0,$$

where $y_t = \ln(Y_t)$ and $x_t = \ln(X_t)$. In this case, a regression of historically observed y_t on x_t yields an unbiased estimator $\hat{\beta}$ that can be used to predict economic activity in the absence of directly measured time series. Naturally, this empirically observed correlation could be spurious. To explore that notion, in appendix Section C.2 we formalize the econometric processes that could drive $\hat{\beta}$, and find three possibilities. First, something that drives y could also drive x, so changes in the two variables would be correlated in a systematic way (which we formalize in the appendix). Second, y and x could be independent but randomly follow the same trends in the data. Third, y and x could be measured with an error and there could be a correlation between the measurement errors of the two variables. Almost all papers we cite assume that the first explanation is driving the results without formalizing this assumption. Having considered the possible alternatives and their structure, we find the two other explanations would be unlikely to hold across different geographies, time periods and econometric approaches used in the many studies we cite above.

While the literature has focused on the case of a constant elasticity, this assumption is problematic if major shocks permanently alter the structural relationship between economic activity and its correlate measure. An example of such a shock is a war.[7] It can be conceptualized as a permanent shift in the elasticity by an unknown quantity γ.[8] Then,

$$y_t = \beta x_t + 1_t^{POST} \gamma x_t + \epsilon_t,$$

where 1_t^{POST} represents an indicator variable for the post-break periods. As new data points y_t from the post-shock period come with a lag (if their production does not cease altogether), no unbiased estimator for γ can be derived from the data. In particular, the use of $\hat{\beta}$ to approximate the post-shock elasticity implies an unknown, and potentially large, bias of γ.

[7] As discussed in Section 2, several papers have documented circumstances under which the relation between nighttime lights and economic activity becomes unstable, either in levels or in changes. To the best of our knowledge, the (in)stability of the relevant elasticities has not been investigated for the case of military conflicts.
[8] Additionally, one could also consider a permanent shift in the error term.

Intuitively, one can obtain a more accurate estimate when multiple correlates x^i (with $i = 1,2,\ldots,N$) for y are available and their elasticity shifts γ^i are not perfectly correlated with each other. Indeed, the intuition that predictions based on multiple proxies can improve on predictions based on any individual proxy has been formalized in different literatures, e.g. in the context of nighttime lights as a proxy for economic activity (Henderson et al., 2011, 2012), the discussion of so-called shrinkage estimators and related Bayesian frameworks (Fay III and Herriot, 1979; Efron and Morris, 1972; Chaney, 2020), and in more practically oriented discussions of methods to optimize the use of multiple proxies (Armstrong, 2001; Lubotsky and Wittenberg, 2006).

A standard result implied by this literature is that a linear combination of the N separately estimated predictions $\hat{y}^i$ (from each correlate x^i) can yield a more accurate prediction than that obtained from any individual model. More formally, consider the linear combination

$$\hat{y}_t = \sum_{i=1}^{N} w^i\, \hat{y}_t^i = \sum_{i=1}^{N} w^i \left(\hat{\beta}^i x_t^i\right) \quad with \quad \sum_{i=1}^{N} w^i = 1, \tag{1}$$

where $\hat{\beta}^i$ denotes the coefficients obtained from the respective single-correlate models defined above. By choosing an optimal set of weights $w^1, w^2, w^3 \ldots w^N$ such that

$$\min_{w^i}\left[\sum_{i=1}^{N} w^i w^i Var\left(\hat{\beta}^i\right) + \sum_{i\neq j} w^i w^j Cov\left(\hat{\beta}^i,\hat{\beta}^j\right) + \left(-\sum_{i=1}^{N} w^i\,\gamma^i\right)^2\right], \tag{2}$$

the mean squared error (MSE) of the prediction is minimized, and may be smaller than the MSE of any single-correlate model. Ex ante, the optimal weights are unknown and depend on unknown variances and covariances between the error terms of each pair of correlates. However, the optimal weights are recovered without any prior assumptions through an intuitive multiple regression approach (Lubotsky and Wittenberg, 2006). Thus, we use regressions of the form

$$y_t = \sum_{i=1}^{N} \theta^i x_t^i + \varepsilon_t\,, \tag{3}$$

to generate predictions $\hat{y}_t = \sum_{i=1}^{N} \hat{\theta}^i x_t^i + \varepsilon_t$ for the post-break period.

The critical assumption underlying this framework is that the biases induced by the structural change within each single-correlate model are unrelated to each other. In Section 5.1 (specifically in Table 1), we show empirical evidence that this assumption is indeed likely to be satisfied in our case. However, it should be noted that it is theoretically possible that a single-correlate model could yield more accurate predictions than a model using a combination of predictors. This could occur if one predictor is significantly better than the others. However, in a scenario where such a predictor is known, there is no need to develop a new model. In the more common situation, when it is not clear which (if any) single predictor outperforms the others, a model using a combination of predictors will be the rational choice. Moreover, this choice can be informed by data. In Section 5.1, we find no evidence that any single-predictor model outperforms the multi-predictor model in either the estimation or validation samples.

An intuitive approach to testing our argument empirically could be to use data from a developed country and try to show that our favorite model outperforms single-predictor specifications in generating accurate out-of-sample predictions. The biggest advantage of this approach would be high-quality input and validation data. However, it is unlikely that such results would generalize to the extreme uncertainty cases where our method is likely the most useful. Indeed, even during the COVID-19 pandemic, good data was available in most advanced economies, while those interested in less advanced economies had to rely on alternative indicators (Stokes and Román, 2022; Chetty et al., 2020). We, therefore, focus on data from Ukraine, as it offers not only good estimation sample data but also two similarly extreme shocks to economic activity and uncertainty. The fact that the shock occurs twice allows us to use the first shock to validate our results.

The intensification of the russian invasion of Ukraine in February 2022 had significant repercussions on data gathering and processing by state statistical agencies, including the SSSU (State Statistical Service of Ukraine), the official statistics office. Military hostilities severely limited the availability of timely economic data, affecting the operations of both the SSSU and private surveying companies. Consequently, the collection of both hard and soft data came to a halt until June-July of 2022.

COVID-19 lessons, where online and phone-based surveys were expanded to encompass broader economic topics at the household and company levels, did not help in this context. A military invasion essentially freezes the sampling and processing of economic activity-related information. Security concerns led to the suspension of access to mobility data (such as Apple and Google services), electricity consumption records, and other variables that had proved useful during the COVID-19 period.

In this challenging environment, the use of data-intensive models by central banks became constrained, with limited use beyond scenario analysis and comparative exercises. Without any measurements of macro aggregates or their components, assessing the scale and pace of changes in the Ukrainian economy required a new methodology.[9]

To navigate this situation, an important research avenue emerged, focusing on replacing traditional hard and soft data with alternative measures that exhibit strong correlations with the variables of interest – specifically GDP. This necessitates a shift in the data creation process,

[9] See the National Bank of Ukraine Inflation Report from July 2022 (NBU, 2022) for an initial policy application.

seeking suitable alternative inputs and models that demonstrate a robust relationship with GDP. In essence, it entails the development of GDP tracking models utilizing alternative data sources.

The nature of the shock posed challenges to the application of established forecasting techniques. The location, duration, and intensity of the shocks varied over time and across different regions. While uncertainty prevailed throughout the country, the physical damage and loss of human life were initially concentrated in the northern and southeastern oblasts. As time progressed, the eastern and southern regions experienced the most significant impact. These nuanced developments may be obscured when examining aggregate measures of output, as cross-sectional economic surprises may average out, assuming such measures are accessible.

Given the regional nature of the shock and the lack of timely statistics, we build on recent advances in development economics linking big data to changes in economic activity at the sub-national scale. While big data are not gathered with an economic surveying purpose in mind, they are widely used to complement, improve, and even replace conventional measures of economic activity when these are lacking (see Section 2 for a review of the literature).

4. Data

We focus on Ukrainian oblasts – the smallest sub-national administrative unit for which we have historic GDP data from the SSSU. Regional GDP data is available at annual frequency between 2012 and 2020. For presentation only (we do not use this for estimation) we extrapolate regional GDP figures in 2021 using the national GDP growth rate reported by NBU.

4.1. Regional Economic Activity

The first challenge of economic measurement during a war is the fact that areas of effective government control change dynamically and it is unclear what areas should contribute to the national figure. Being able to track regional changes allows the construction of different national aggregates for different purposes (such as comparing total GDP within a boundary with GDP generated only in regions under government control). This is also important because economic activity is not uniformly distributed across space. In fact, it is likely that more economically active areas will be military targets.

Indeed, the data from the SSSU shows large disparities between economic activity across regions, which we refer to as regional GDP. In 2020, Kyiv and its surrounding region accounted for 30% of the country's GDP (excluding occupied regions). Out of the 25 regions we have data for, the biggest five account for 50% and biggest 10 for 70% of the total GDP figure in 2020. This highlights the importance of regional dis-aggregation during a war, as economic activity is not equally distributed across space within countries. It is also worth noting that we find surprisingly little spatial correlation between regional GDP figures and their growth rates. We attribute this to the large distances and long travel times between cities in different oblasts (Fingleton and Szumilo, 2019).

It is also unclear what the appropriate measure of economic activity during a war should be. One of the biggest issues with measuring economic activity during dramatic shocks is the fact that the type and nature of economic activity is likely to change. For example, anecdotal evidence suggests that during a crisis volunteering work (which is not included in GDP) is significantly more popular. We use regional GDP denominated in 2004 hryvnias as our measure of activity. This means that all our predicted outcomes should be interpreted as the level of GDP that the input data (discussed below) corresponded to over our estimation period (2014-2020).

4.2. Nightlights

We construct measures of annual and monthly nightlight intensity (NLI) at the level of Ukrainian regions using NASA's Black Marble product suite. The Black Marble suite provides processed, geo-referenced raster images recording nighttime radiance measured from space at a resolution of approximately 500 meters per pixel (Román et al., 2018).[10]

Following the literature, we define a region's NLI as the sum of all encompassed pixels' radiance values over the region's area (Henderson et al., 2012). Given that GDP is available at a yearly frequency, we investigate the linear log-log relationship between annual NLI and GDP in a simple regression that takes the following form:

$$\ln(GDP_{it}) = a + \beta^{NLI} \ln(NLI_{it}) + \epsilon_{it}, \tag{4}$$

Guided by the literature, our specification has the same coefficient for all regions. We find that this simple specification explains around 65% of the variation in the outcome. It performs similarly well in a cross section, which is illustrated in panel A of Figure 1. Comparing panel A with a time series relationship in Kyiv, presented in panel B, shows that the relationship holds within regions. In Kyiv, nightlights can capture around 54% of changes in GDP over time, but this value changes across regions. With a region-fixed effect, NLI explains around 22% of the within-region variation in GDP over time. This is consistent with findings of the literature, which suggest that this approach works better in urban regions (Gibson et al., 2021). Notably, the regression coefficients estimated on the data represented in panels A and B of Figure 1 are statistically indistinguishable from each other.

[10] The raw data comes from the Visible Infrared Imaging Radiometer Suite Day/Night Band (VIIRSDNB) onboard the Suomi National Polar-orbiting Platform (SNPP). While VIIRS-DNB-generated data is only available from 2012 on, it removes many of the well-known measurement problems and interpretational ambiguities implied by longer-running time series from the Defense Meteorological Satellite/Operational Linescan System (DMPS/OLS), e.g. top-censored radiance measures (Min et al., 2021). Within the Black Marble suite, NASA corrects the VIIRS-DNB raw data for remaining disturbances, such as cloud-cover or lunar light effects, and converts daily records into monthly and annual composites.

This data begins in 2012, which marks the beginning of our annual data sample. In Ukraine, this data is also not very reliable for June and July (due to satellites' orbits in those months), which provides a natural finishing point for our analysis of the first months of the war – May 2022.

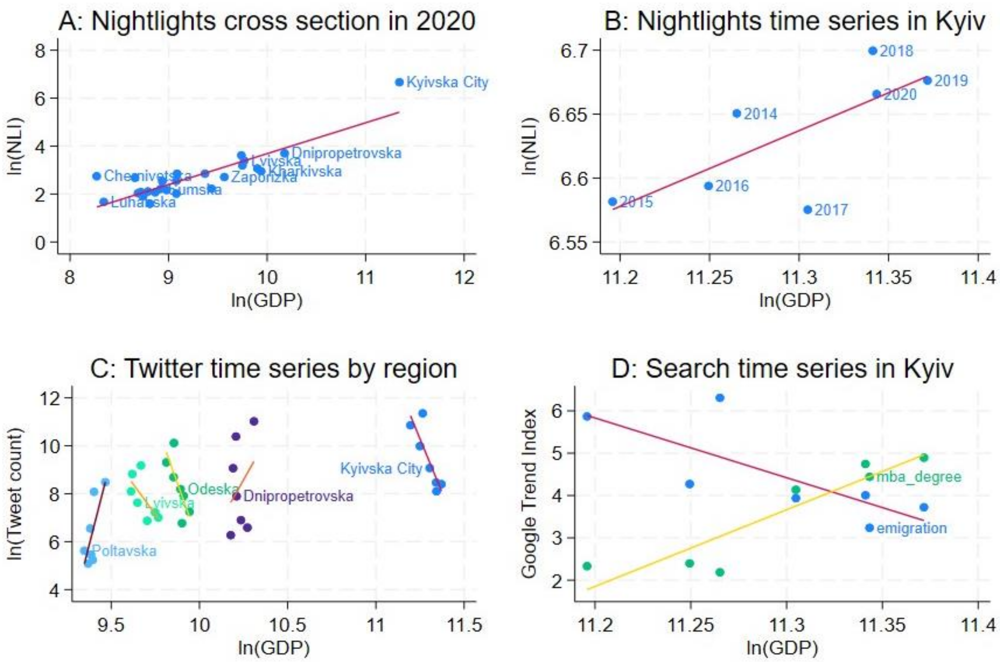

Figure 1. Correlation between Regional GDP and Economic Activity Indicators

Notes: *Panel A* shows a cross section of the log of GDP in 2020 against the log of nightlight intensity. *Panel B* shows logs of GDP and nightlight intensity in Kyiv over time between 2014 and 2020. *Panel C* plots the log of GDP against the log of the count of geolocated tweets containing media in five selected (for visual clarity) regions of Ukraine over time between 2014 and 2029. *Panel D* plots the log of GDP against values of Google Trends search index values over time between 2014 and 2020 in Kyiv.

4.3. Twitter Counts

Considering the mechanism put forward by Indaco (2020), tweets containing a picture may represent a proxy of conspicuous consumption. Using Twitter's Tweet Counts API, we generate measures of monthly and annual social media activity for Ukraine's regions. Following Indaco (2020), we only consider media tweets (with an image or video attachment) containing a user-provided location tag.[11]

Since taggable locations cannot be directly aggregated to Ukraine's regions, we rely on the API's functionality to provide reliable count measures for circles characterized by a centroid and a radius. In particular, we query tweets separately for Ukraine's largest 1,500 settlements by drawing a circle[12] around each settlement's centroid, with a radius that implies an area equivalent to the settlement's administrative extent. Finally, we aggregate our settlement level estimates by region

[11] Indaco (2020) reports that, as of 2018, roughly 23% of tweets contain media and 5% are geo-tagged.
[12] Note that our circles do not overlap so there is no double counting.

and desired temporal frequency.[13] Twitter had around 8% of the social media market in Ukraine in 2022, and this value was relatively constant in our sample period (data from StatCounter). However, social media was not popular enough to produce meaningful data in smaller regions until around 2014. This can be illustrated by looking at the count of tweets in the region with the fewest tweets that meet our criteria in 2012, which stood at only 96. By 2014, this figure had increased to 565, far outpacing the national growth in tweets.

Aware of the fact that the relationship between tweets and GDP at the regional level has not been explored and that there are no established micro-foundations, we considered different specifications of the model. We found that in a simple regression similar to Equation 4, Twitter explains around 28% of the variation in GDP, and the correlation is positive. However, when tested for individual regions, we found the relationship to be positive in some regions but negative in others (statistically significant in all but three regions). We demonstrate this in panel C of Figure 1, which plots the relationship between the count of tweets and economic activity in selected regions.[14] We conclude that at the regional level the relationship between tweets and economic activity is area-specific, and use the following model:

$$\ln(GDP_{it}) = a + \beta_i^T R_i \ln(Tw_{it}) + \epsilon_{it}, \quad (5)$$

This specification explains around 95% of the variation in GDP. Adding region-fixed effects increases the overall R^2 to 99% and results in "within R^2" of 47%. This suggests that Twitter can explain a significant proportion of variation in GDP within regions over time. Indeed, in Kyiv, changes in tweets explain 75% of changes in GDP over time, but the average for the large regions is lower (52% within-region R^2). In general, we find Twitter performs much better in tracking GDP in areas with higher economic activity and more tweets. This is not overly surprising, given the fact that the small regions in the early years of our sample had very few tweets.

[13] To determine the extent of Ukraine's settlements, we rely on an up-to-date shapefile provided by the Humanitarian Data Exchange project. Our approach misses tweets because of its restriction to the 1,500 largest settlements and the imperfect approximation of settlement areas via a circle. While we find that tweet counts quickly approach 0 with sufficiently small settlement size, our measures should be interpreted as lower-bound estimates of the true regional tweet media- and geo-tagged count.
[14] We selected regions of different sizes and with different slopes, which facilitated good visualization of the data. This panel figure is for illustration purposes only.

4.4. Google Trends

We measure online search activity using monthly Google Trends Indices (GTIs) for a battery of 30 topics that signify interest in consumption activities and macro-economic trends, e.g. "Washing machine" or "Unemployment".[15] We experimented with expanding our set of topics by another 30, but found that it did not have a big influence on the results due to the way we chose to collapse them (see below). The data is provided by Google and comes separately per region, at a monthly frequency. GTIs are scaled to integers between 0 and 100, where a value of 100 denotes the maximum search volume over the queried period and smaller values stand for the respective percentage share of that maximum. This means that each region needs to be modelled individually, since the data across regions is not directly comparable. Moreover, data provided by Google is sampled from its many servers, so the reported figures may be affected by a sampling bias. We address this issue by repeatedly sampling the data and averaging the results for each region. We find that this does not make a big difference to the national results, but can affect region-specific results for smaller regions – especially at the beginning of our sample period.

We find that GDP changes within each region can be explained very well even with simple models. For example, search trends for terms related to the topic "MBA degree" are positively correlated to GDP and explain 84% of the variation in changes in economic activity over time in Kyiv. At the same time "emigration" explains 58% and is negatively correlated. We visualize these relationships in panel D of Figure 1. However, including search trends for both MBA and emigration in a regression does not offer a better fit for the data as the two trends are highly correlated. This shows that selecting the right combination of trends is key to choosing a good model.[16] Intuitively, it is possible to find a combination of topics that give the highest R^2. Indeed, the standard approach taken in the literature is to use estimation and variable selection algorithms such as LASSO or random forest to find these combinations (Woloszko, 2020). However, we are severely constrained by estimation data availability, and with only seven time periods the use of these approaches results in overfitting (Bartlett et al., 2020). We opt for a simpler method and reduce the dimensionality of our GTI data by using Principal Component Analysis, as suggested by Woo and Owen (2019).

Inspired by Stock and Watson (2002) and Woo and Owen (2019), we represent the 30 standardized Google Trends as principal components (PCs) and regress them on GDP in each region using the following model:

[15] We list these 30 topics in Appendix 6.3. Note that the focus on topics rather than search terms ensures that linguistic variation across Ukraine does not distort our measure (Woloszko, 2020). We use the opensource PyTrends package to query the data from Google (PyTrends contributers, 2022).

[16] Note that because data is always region-specific, all coefficients have to be estimated separately for each region.

$$\ln(GDP_{it}) = a + \sum_{n=1}^{N} \beta_i^G PC_{itn} + \epsilon_{it}, \tag{6}$$

In-sample fit indicates that the first principal component alone captures almost 78% of the sample variance of GTIs at the national level. However, since principal components are region-specific, this differs across regions. To ensure that we do not miss any important factors, we use the first three PCs (n=3) in our favored regression. The within-region R^2 of a model with region-fixed effects is 64% for three PCs (28% for just one PC). Regional R^2 values vary from 17% to 92%.

5. Empirical Strategy and Results

In practice, we estimate the following equation:

$$\ln(GDP_{it}) = a + R_i \times (\sum(\theta_i PC_{it}) + \theta_i \ln(Tw_{it})) + \theta \ln(NLI_{it}) + \epsilon_{it}, \tag{7}$$

where i denotes a region, t a time period (year), R is a vector of region-specific dummy variables, a is a regression constant, $\sum PC$ are the first three principal components of Google trends, NLI are nightlights intensity, and Tw is the count of media Tweets.

It is worth noting that the θ parameters from Equation 7 are essentially a combination of weights W and individual parameters β for each predictor (as defined in Section 3). This links our estimating equation to the previous section. However, we make no claim that the parameters estimated by the model can be interpreted or treated as individual transferable “elasticities.” The endogeneity highlighted in the appendix Section C.2 implies that our estimated parameters may be biased (Szumilo, 2021). While this is not an issue for making predictions in our setting, it means that external comparisons require caution (Pesaran and Timmermann, 1995).

While the estimating equation has to be specified in years (as dictated by data availability), we need to generate predictions with higher (monthly) frequency. To do so, we will draw on the mixed-frequency predictions literature summarized by Ghysels (2016). Specifically, to estimate parameters in Equation 7, we will use annual GDP data and predictor data averaged at annual frequency. To generate predictions in 2022 we will use predictors averaged at monthly frequency. Since data for our predictors is generated at much higher frequency than a month (Twitter and Google are continuous while nightlights are daily), monthly and annual aggregations will have the same scale. Therefore, the interpretation of our monthly prediction is simply that they are predictions of annual GDP assuming that the whole year is the same as a given month. We adopted this approach because expressing monthly economic activity in annual terms makes comparing different periods in our data considerably easier. We note that our monthly predictions are likely more volatile due to the seasonality of economic activity. We experimented with removing seasonality from monthly predictors, and the impact of this on the results was very small (due to the magnitude of the exogenous shock), so the results presented in this paper are not adjusted for seasonality.

We divide our sample period into three parts, as illustrated in Figure 2. First, we set the estimation sample that runs from 2014 to 2020. This a period in which borders were stable following the first invasion in March 2014. It is also a period of significant economic, social and military reforms. Since the russian invasion in 2014 was a shock to Ukraine, it spurred important changes and reforms. It also altered the country's borders. In many ways, it is a structural shock similar to the full-scale invasion in 2022. We will therefore use the coefficients estimated in the period between invasions to generate out-of-sample predictions.

The second period we consider is 2021 and the first months of 2022 (specifically January to May, for which we use annualized monthly data). This is our main period of interest, as we generate out-of-sample predictions for 2021 and the first months of the war in 2022.

The third period includes only two years: 2012 and 2013. This is our validation sample, as we use coefficients estimated between the invasions to generate out-of-sample predictions before the first invasion. Since both invasions generated structural shocks, we can compare our out-of-sample predictions with realized outcomes before the first invasion to validate some of our modelling assumptions. We do not claim that the two invasions created the same effect, but simply that in both cases there are concerns about using coefficients estimated in a period of stability. The key assumption behind combining different predictors of GDP is that their prediction errors are uncorrelated. While it is straightforward to test this assumption in our estimation sample, using the validation sample allows us to test if our assumption holds even with structural breaks. One limitation of this approach is that the coverage of Internet activity data for small regions before 2014 is not very good, so our predictions for those areas and periods are less accurate.

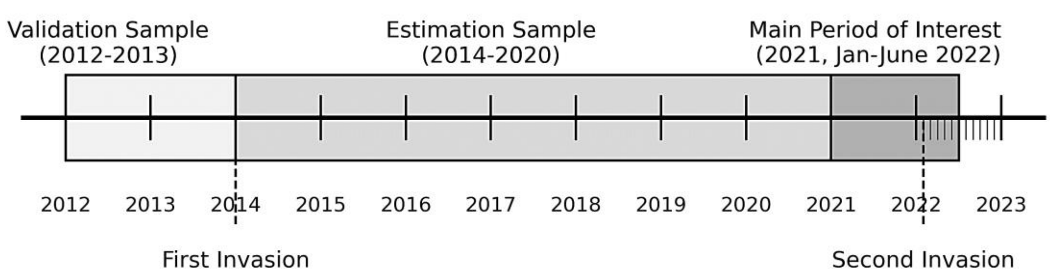

Figure 2. Timeline and Samples Used

To further validate our results we can also compare them against other data sources. First, although we do not have regional GDP data for 2021, we do have a national GDP growth figure from NBU. Our aggregated regional predictions for 2021 should match the national figure estimated by NBU. Second, given the common expectation that violence reduces economic activity, a higher level of war-related destruction should correlate (negatively) to our GDP predictions in 2022. We show this by using data on the number of fires caused by war per region from the Economist and Solstad (2023).

While nightlights and Twitter data are available at a fine geographical level, Google trends are aggregated to regions. This is a problem for two regions in our estimation sample that are under partial occupation. We therefore estimate a model in those regions that excludes Google Trends and generate separate predictions for the occupied and unoccupied parts. We also allow the Autonomous Republic of Crimea to have a different θ.

Indeed, it is not obvious what geographic area should be considered as Ukraine for the purpose of tracking economic activity during an invasion. While the conventional approach to calculating GDP is to only include areas under direct government control, during an invasion GDP from controlled areas may not be the most important indicator, and government control is not easily defined. While military control can change quickly, many government services can continue to be offered in areas occupied by foreign soldiers, and economic activity in those regions can be of interest to policy makers. Because we predict economic activity for individual regions (or their parts) regardless of whose control they are under, our results can be used to produce different figures. Our analysis follows data available from the SSSU, which offers data for all regions before 2014, but excludes Crimea and the occupied parts of Donetsk and Luhansk afterwards. We therefore use data from before 2014 to estimate our model for Crimea (including its largest city Sevastopol, for which data is separately reported) and parts of Donetsk and Luhansk that have been occupied since 2014.[17]

5.1. Results: Model Comparison

We begin presenting our results with an overview of the different models in Table 1. We present models for individual predictors described in Equations 4, 5 and 6, a combination of those predictors from Equation 7, and an alternative version of Equation 7 but with region-fixed effects. We included the last model for comparison as it could potentially have some advantages over a random effects model. However, those advantages come at a cost of an additional assumption – that the fixed effects do not change due to the structural shock. Given that it is unlikely that this assumption holds in our case, we prefer the simpler model.

In panel A we report R^2 values for the estimation sample (2014 to 2020). In single-predictor models, both Twitter and Google offer better fits than nightlights, but combining the three is clearly better than any single predictor. Overall, the R^2 values we report are very high, as our main model estimates a large number of parameters (four regional and one national one) compared to the number of observations. We therefore use panel A only for comparison between models estimated on our data, rather than comparing them externally. It is also worth noting that the model with

[17] To generate a GDP figure for pre-2014 for the parts of Donetsk and Luhansk that were occupied in 2014, we apportioned the total GDP reported for those years in each region to parts that later became occupied, based on nightlights.

regional fixed effects provides the same fit as our main model, so there seems to be little evidence that fixed effects help fit the data in our case.

In panel B we estimate changes in national GDP between 2020 and 2021 to compare them to the 3.5% reported by NBU. To do this, we aggregate our out-of-sample predictions for each region's GDP and calculate the growth rate of this sum between 2021 and 2020. Single-predictor models are relatively far from the target[18] , while multi-predictor models are significantly closer to the official figure. This suggests that combining predictors could be useful not only during extreme events such as wars, but also when no accurate data is available. The prediction of the fixed-effects model is slightly closer to the benchmark, but both multi-predictor models underestimate.

In the next row, we report changes in national GDP between 2021 and March 2022. While we do not have an official benchmark for this period, we expect a big drop in economic activity in the first month of the war. This is clearly the case in the nightlights model, but the Twitter model seems to suggest that the economy was growing, while the Google model suggests a modest decline in activity. As we already mentioned and discussed in Section 3, we expect each of those predictions to be biased. Indeed, it seems that while nightlights are likely underestimating, Twitter and Google seem to be overly optimistic (but to different degrees). As expected, the nightlights model offers the most pessimistic prediction of the economy declining by over 54%. Our multi-predictor models are a little more optimistic and suggest a decline of around 30%.

In panel C, we consider "goodness of fit" statistics of the out-sample-predictions in the validation sample. Once again, combining predictors is clearly better than using a single one. However, it is not clear if fixed effects improve prediction accuracy. While the model with fixed effects seems to be predicting GDP levels more accurately, our favorite model is better at predicting changes.

Finally, we turn to panel D to validate the key assumption that prediction errors of single predictor models are not highly correlated with each other. Indeed, in-sample error correlation seems to be low or negative. The most difficult test of our assumption comes from comparing correlations of prediction errors in our validation sample. If a structural break caused all predictors to be biased in the same way, the correlation would be high and our approach would not offer any improvement over using a single predictor. We find that in our validation sample the correlation increases. However, it is only high between prediction errors of Twitter and Google, while the other correlations are reassuringly low.

[18] This is not surprising, as none of these predictors has ever been claimed to be a very accurate tracker of economic activity. Instead, the literature shows that they are useful when no alternative data is available.

5.2. Results: Favorite Model

We present the predictions of our favorite model at two levels: national and regional.

First, we present the national aggregation (based on areas under government control in 2021) in Figure 3. The left hand side panel shows a very good fit of our regional predictions aggregated to national GDP in both estimation and validation periods. It shows a dramatic decline in economic activity in March 2022, but a partial recovery in April and May. This is consistent with changes in the military assessment of how the invasion was progressing. While in March most analysts predicted a swift russian victory, Ukrainian successes on the battlefield (including lifting the siege of Kyiv) resulted in far more optimistic outlooks in April and May (Masuhr and Zogg, 2022). In May, levels of economic activity are predicted to be only around 15% smaller than in 2021. This may seem small compared to the 30% drop in March, but it is large in absolute terms and does not reflect any loss of territory.

Focusing on the national results requires assumptions about the territory included in the analysis and obscures regional heterogeneity. Our preferred approach is to focus on individual regions. We present regional results for May 2022 in Figure 4. We omit an in-depth interpretation of this map, as it would require information on Ukrainian political and physical geography, industrial structure and the evolution of the conflict in 2022. Nonetheless, even with minimal background information, some high-level conclusions that are obscured when looking at national figures emerge. For example, western regions see their economic activity decline less (or even increase), while eastern regions experience dramatic declines. This pattern is consistent with an intuitive reaction to the invasion, which directly affected mainly the eastern part of the country. Indeed, early evidence seems to suggest that workers and firms from regions affected by fighting relocated to western parts of the country (Mykhnenko et al., 2022). The map also shows that regions that were under russian occupation since 2014 are affected to a smaller extent than regions that were invaded in 2022, and suggests that regions affected by the most fighting and destruction saw their economic activity shrink the most. However, the last point is better illustrated in the lower panel of Figure 4, which compares the level of change in economic activity to the level of destruction, as indicated by the number of war fires. The strong negative correlation between war fires and economic activity holds both in a cross-section and within regions over time.

The analysis of regional results can be further extended by considering a time series for each region. In the appendix Figure 8, we show observed and predicted levels of economic activity for each region of Ukraine. This includes regions (and their parts) under occupation since 2014.

Finally, coefficient estimates of the main model are available in Appendix (Table 2). We report them for information only, as interpreting individual coefficients is difficult due to collinearities in the estimation data.

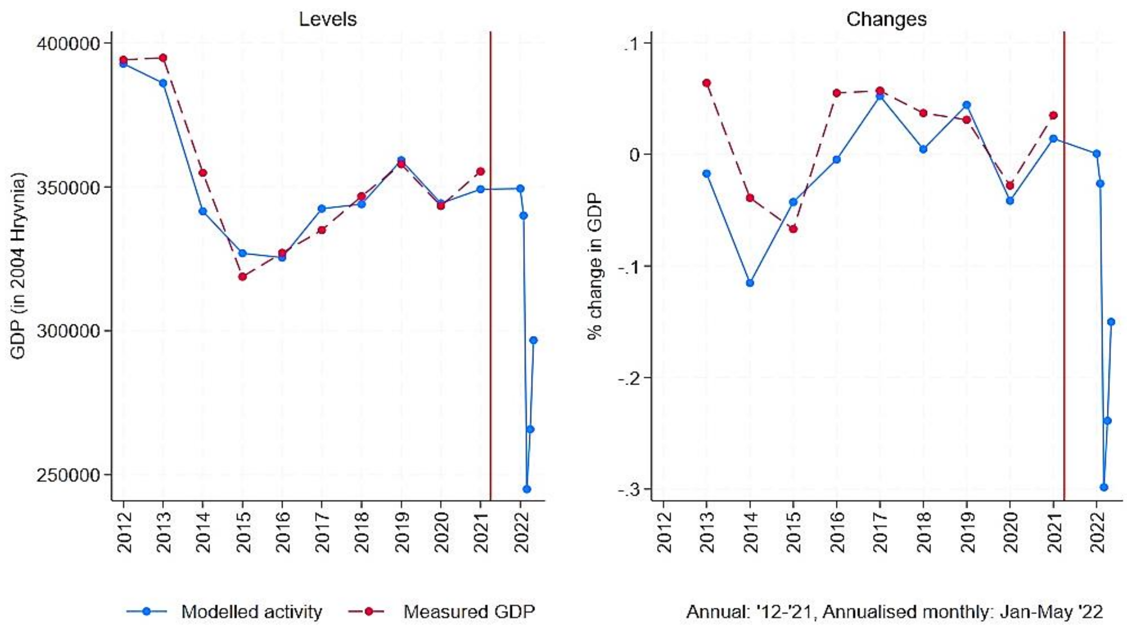

Figure 3. Actual and Predicted GDP of Ukraine

Notes: The panel on the left gives the national GDP figures in Ukraine as aggregated from regions for which data is available. Measured GDP refers to the sum of all GDP figures reported by SSSU. “Modelled activity” is the sum of all regional GDP figures predicted by our model. In-sample predictions are for 2014 to 2020, while all other predictions are out-of-sample. The panel on the right plots changes in the national aggregates over time. Data is annual for 2012 to 2021, and annualized monthly in 2022.

5.3. Main Model: Robustness

A possible concern could be that our panel model works in levels while the majority of the GDP tracking and nowcasting literature uses time-series data focusing on first differences or growth rates. While we do not estimate a first-difference model, we show a numerically equivalent fixed-effects specification. Overall, we find our favorite specification from Equation 7 and the fixed-effects model offer very similar results, but we prefer the simpler model with fewer assumptions. To show that our approach is useful in predicting changes, we convert our regional predictions into national growth rates and compare them to data from the SSSU in Figure 3.

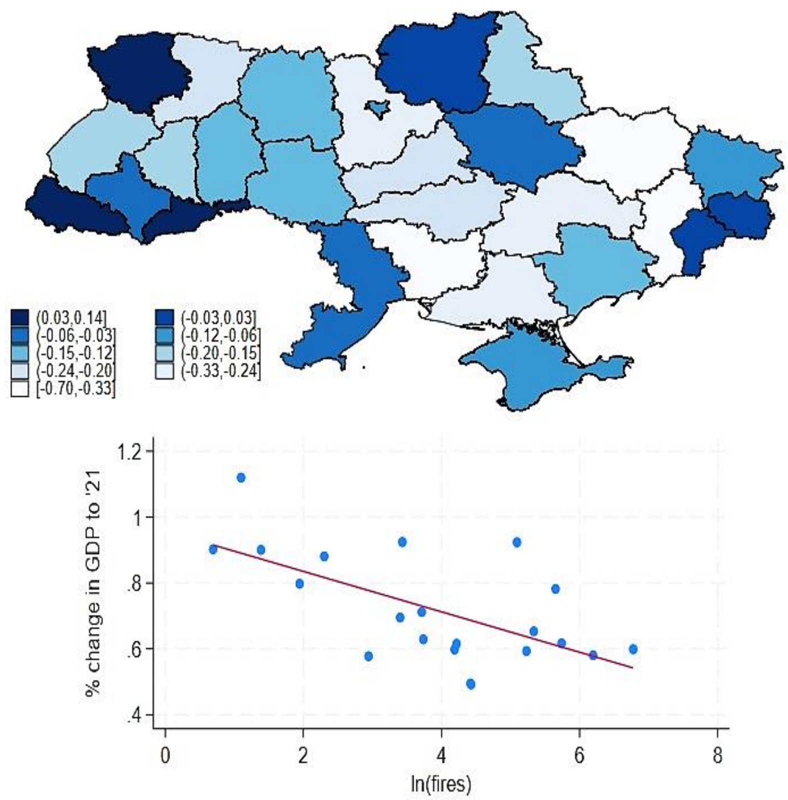

Figure 4. Regional Changes in Economic Activity in Ukraine (2021 to May 2022).

Notes: The figure uses changes in economic activity predicted by our model between the annual GDP in 2020 and the annualized monthly equivalent for May 2022. The top panel presents a map of Ukraine using country borders from 2010 and regional borders from 2020. Since 2014, Ukraine did not have effective government control over the Crimean Peninsula (the southernmost region of the country) and parts of Donetsk and Luhansk (the two easternmost regions of the country). The lower panel plots the log of the number of war fires reported by the Economist and Solstad (2023) against our monthly predictions of changes in economic activity between 2021 and the month in which fires are reported, for regions where war fires were reported in 2022.

The figure shows that our results match official data relatively well both in levels and changes. While changes in the early years of our sample are predicted less well than in the later years, this is mainly because of poor coverage of our Internet usage data in small regions. More detail on this can be seen in Figure 8 in the appendix, which shows that changes in larger regions in the early years are predicted much more closely than in smaller regions.

Another potential concern for our model that is a common issue in data-driven types of models is overfitting. In our case, a concern could be that we have too few observations and too many parameters per region. This is most likely to occur with the principal components of Google Trends, as they require three region-specific coefficients. To check if this indeed affects our results, we consider a model with only one principal component. This model yields predictions that are qualitatively very similar to our favorite specification, but fit the data less well in most regions. We conclude that overfitting is unlikely to be an issue for our results.

5.3.1. Machine Learning Results

For comparison and further validation of our main results, we also generated predictions for a model estimated using a random forest algorithm and all predictor variables in our dataset. Although machine learning approaches are favored in the literature using Google Trends, they are less reliable with limited training data. Surprisingly, we find that predictions based on the random forest model outperform our favorite predictions in some ways. In evaluating performance we follow metrics from Table 1 and note that the machine learning model gives an R^2 of 99.6%, national growth 2020 to 2021 of 1.2%, 2021 to March 2022 of -17.42%, Pre 2014 RMSE for levels of 0.0297 and 0.0174 for changes. A graphical representation of the national level results is available in the appendix Figure 7. As expected, random forest offers a better overall in-sample fit than our favorite specification, but the average low error comes at the cost of large errors at the beginning of the estimation sample (as can be seen from Figure 7). It is therefore surprising that it outperforms our favorite specification in predicting out-of-sample GDP before 2014, but not in 2021. Given our limited training data, we do not rely heavily on machine learning results. However, we find it reassuring that in many ways they are qualitatively similar to our favorite specification. For example, they both estimate a reduction in economic activity in March 2022 and subsequent partial recovery.

6. Conclusion

Although unconventional indicators of economic activity such as nightlight intensity, Twitter activity or Google Trends are frequently used by researchers and policymakers in places where no reliable data is available, combining them is not very common. Our results show that there are potential benefits of including more than one data source in models based on such data. In periods of extreme uncertainty (such as a full-scale military invasion), a model combining different unconventional indicators is likely to be more reliable than a single-indicator model.

Our results also offer an unprecedented insight into the short-term changes in economic activity in an economy under a full-scale invasion and in occupied regions. Our findings showing that local fighting reduces local economic activity, or that attacking regions with high GDP affects the national economy more than attacking regions with small GDP, are novel but not overly surprising. The more interesting of our findings include showing that: 1) economic activity in regions that receive people and firms who flee from fighting can increase, 2) regions that are overrun quickly[19] experience a lower short-term decrease in activity than regions where fighting persist, 3) changes in expectations of how the conflict will evolve correspond to changes in economic activity at the country level.

Finally, our experience of working with Ukrainian officials and the international organizations supporting them suggests that preparing for making economic decisions under extreme

[19] Note that this includes regions that are overrun and re-taken quickly without a significant amount of heavy fighting.

uncertainty should be a key element of war preparedness. This can include developing methods that, like our approach, help improve data availability or designing policies that are flexible and can adopt to changing circumstances.

References

Addison, D. M., Stewart, B. P. (2015). Nighttime lights revisited: the use of nighttime lights data as a proxy for economic variables. Policy Research Working Paper, 7496. Washington, D.C.: World Bank. Retrieved from http://documents.worldbank.org/curated/en/613891468195568648/Nighttime-lights-revisited-the-use-of-nighttime-lights-data-as-a-proxy-for-economic-variables

Armstrong, J. S. (2001). Principles of Forecasting: A Handbook for Researchers and Practitioners. Dordrecht: Kluwer Academic Publishers.

Asher, S., Lunt, T., Matsuura, R., Novosad, P. (2021). Development research at high geographic resolution: An analysis of night-lights, firms, and poverty in India using the SHRUG open data platform. The World Bank Economic Review, 35(4), 845–871. https://doi.org/10.1093/wber/lhab003

Askitas, N., Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting. Discussion Paper, 4201. Bonn: IZA. Retrieved from https://docs.iza.org/dp4201.pdf

Bartlett, P. L., Long, P. M., Lugosi, G., Tsigler, A. (2020). Benign overfitting in linear regression. Proceedings of the National Academy of Sciences, 117(48), 30063–30070. https://doi.org/10.1073/pnas.1907378117

Bellows, J., Miguel, E. (2006). War and institutions: New evidence from Sierra Leone. American Economic Review, 96(2), 394–399.https://doi.org/10.1257/000282806777212323

Beyer, R. C. M., Chhabra, E., Galdo, V., Rama, M. G. (2018). Measuring districts' monthly economic activity from outer space. Policy Research Working Paper, 8523. Washington, D.C.: World Bank. Retrieved from http://documents.worldbank.org/curated/en/835491531401292135/Measuring-districts-monthly-economic-activity-from-outer-space

Bickenbach, F., Bode, E., Nunnenkamp, P., Söder, M. (2016). Night lights and regional GDP. Review of World Economics, 152(2), 425–447. https://doi.org/10.1007/s10290-016-0246-0

Bok, B., Caratelli, D., Giannone, D., Sbordone, A. M., Tambalotti, A. (2018). Macroeconomic nowcasting and forecasting with big data. Annual Review of Economics, 10, 615–643. https://doi.org/10.1146/annurev-economics-080217-053214

Bokányi, E., Lábszki, Z., Vattay, G. (2017). Prediction of employment and unemployment rates from Twitter daily rhythms in the US. EPJ Data Science, 6, 14. https://doi.org/10.1140/epjds/s13688-017-0112-x

Bricongne, J. C., Meunier, B., Pical, T. (2021). Can satellite data on air pollution predict industrial production? Working Paper, 847. Paris: Banque de France. Retrieved from https://www.banque-france.fr/system/files/2023-03/wp847_0.pdf

Caverley, J. D. (2018). Economics of war and peace. The Oxford Handbook of International Security, p. 304.

Chaney, E. (2020). Modern library holdings and historic city growth. CEPR Discussion Paper, 14686. CEPR Press, Paris & London. https://cepr.org/publications/dp14686

Chen, X., Nordhaus, W. (2011). Using luminosity data as a proxy for economic statistics. Proceedings of the National Academy of Sciences, 108(21), 8589–8594. https://doi.org/10.1073/pnas.1017031108

Chetty, R., Friedman, J. N., Stepner, M., The Opportunity Insights Team, T. (2020). The economic impacts of COVID-19: Evidence from a new public database built using private sector data. NBER Working Paper, 27431. https://doi.org/10.3386/w27431

Choi, H., Varian, H. (2009). Predicting the present with Google Trends. Economic Record, 88(1), 2–9. https://doi.org/10.1111/j.1475-4932.2012.00809.x

Chupilkin, M., Koczan, Z. (2022). The Economic Consequences of War: Estimates Using Synthetic Controls. EBRD Working Paper, 271. European Bank for Reconstruction and Development. Retrieved from https://www.ebrd.com/publications/working-papers/the-economic-consequences-of-war#

Croft, T. A. (1973). Burning waste gas in oil fields. Nature, 245, 375–376. https://doi.org/10.1038/245375a0

Croft, T. A. (1978). Nighttime images of the Earth from space. Scientific American, 239(1), 86–101. https://doi.org/10.1038/scientificamerican0778-86

Davis, D. R., Weinstein, D. E. (2002). Bones, bombs, and break points: the geography of economic activity. American Economic Review, 92(5), 1269–1289. https://doi.org/10.1257/000282802762024502

Doll, C. N. H., Muller, J.-P., Morley, J. G. (2006). Mapping regional economic activity from night-time light satellite imagery. Ecological Economics, 57(1), 75–92. https://doi.org/10.1016/j.ecolecon.2005.03.007

EBRD (2022). Transition Report 2022-23. European Bank for Reconstruction and Development. Retrieved from https://2022.tr-ebrd.com/

Efron, B. Morris, C. (1972). Stein's estimation rule and its competitors – an empirical Bayes approach. Journal of the American Statistical Association, 68(341), 117–130. https://doi.org/10.1080/01621459.1973.10481350

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., Davis, E. R., Davis, C. W. (1997). Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. International Journal of Remote Sensing, 18(6), 1373–1379. https://doi.org/10.1080/014311697218485

Elvidge, C. D., Sutton, P. C., Ghosh, T., Tuttle, B. T., Baugh, K. E., Bhaduri, B., Bright, E. (2009). A global poverty map derived from satellite data. Computers and Geosciences, 35(8), 1652–1660. https://doi.org/10.1016/j.cageo.2009.01.009

Ettredge, M., Gerdes, J., Karuga, G. (2005). Using web-based search data to predict macroeconomic statistics. Communications of the ACM, 48(11), 87–92. https://doi.org/10.1145/1096000.1096010

Fay III, R. E., Herriot, R. A. (1979). Estimates of income for small places: an application of hames-stein procedures to census data. Journal of the American Statistical Association, (74), 269–277. https://doi.org/10.1080/01621459.1979.10482505

Fingleton, B., Szumilo, N. (2019). Simulating the impact of transport infrastructure investment on wages: a dynamic spatial panel model approach. Regional Science and Urban Economics, 75, 148–164. https://doi.org/10.1016/j.regsciurbeco.2018.12.004

Ghosh, T., Powell, R. L., Elvidge, C. D., Baugh, K. E., Sutton, P. C., Anderson, S. (2010). Shedding light on the global distribution of economic activity. The Open Geography Journal, 3, 147–160. http://doi.org/10.2174/1874923201003010147

Ghysels, E. (2016). Macroeconomics and the reality of mixed frequency data. Journal of Econometrics, 193(2), 294–314. https://doi.org/10.1016/j.jeconom.2016.04.008

Gibson, J., Olivia, S., Boe-Gibson, G., Li, C. (2021). Which night lights data should we use in economics, and where? Journal of Development Economics, 149, 102602. https://doi.org/10.1016/j.jdeveco.2020.102602

Goetz, T., Knetsch, T. (2019). Google data in bridge equation models for German GDP. International Journal of Forecasting, 35(1), 45–66. https://doi.org/10.1016/j.ijforecast.2018.08.001

Henderson, J., Storeygard, A., Weil, D. (2012). Measuring economic growth from outer space. American Economic Review, 102(2), 994–1028. https://doi.org/10.1257/aer.102.2.994

Henderson, V., Storeygard, A., Weil, D. (2011). A bright idea for measuring economic growth. American Economic Review, 101(3), 194–199. https://doi.org10.1257/aer.101.3.194

Indaco, A. (2020). From Twitter to GDP: Estimating economic activity from social media. Regional Science and Urban Economics, 85, 103591. https://doi.org/10.1016/j.regsciurbeco.2020.103591

Llorente, A., Garcia-Herranz, M., Cebrian, M., Moro, E. (2015). Social media fingerprints of unemployment. PLOS ONE, 10(5), e0128692. https://doi.org/10.1371/journal.pone.0128692

Lubotsky, D., Wittenberg, M. (2006). Interpretation of regressions with multiple proxies. The Review of Economics and Statistics, 88(3), 549–562. https://doi.org/10.1162/rest.88.3.549

Masuhr, N., Zogg, B. (2022). The war in Ukraine: first lessons. CSS Analyses in Security Policy, (301). Retrieved from https://ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/pdfs/CSSAnalyse301-EN.pdf

Mazzi, G. L., Mitchell, J., Montana, G. (2014). Density nowcasts and model combination: nowcasting euro-area GDP growth over the 2008–09 recession. Oxford Bulletin of Economics and Statistics, 76(2), 233–256. https://doi.org/10.1111/obes.12015

McLaren, N., Shanbhogue, R. (2011). Using Internet search data as economic indicators. Quarterly Bulletin, Q2, 134–140. Bank of England. Retrieved from https://www.bankofengland.co.uk/-/media/boe/files/quarterly-bulletin/2011/using-internet-search-data-as-economic-indicators.pdf?la=en&hash=CE738D37B3A15477EA70225F302BC94E63EA68BD

Min, B., Baugh, K., Monroe, T., Goldblatt, R., Stewart, B., Kosmidou-Bradley, W., Crull, D. (2021). Light every night – new nighttime light data set and tools for development. Technical report. World Bank.

Mykhnenko, V., Delahaye, E., Mehdi, N. (2022). Understanding forced internal displacement in Ukraine: insights and lessons for today's crises. Oxford Review of Economic Policy, 38(3), 699–716. https://doi.org/10.1093/oxrep/grac020

NBU (2022). Inflation Report, July 2022. Retrieved from https://bank.gov.ua/en/news/all/inflyatsiyniy-zvit-lipen-2022-roku

Nordhaus, W. D. (2002). The economic consequences of a war in Iraq. NBER Working Paper, 9361. https://doi.org/10.3386/w9361

Nordhaus, W., Chen, X. (2014). A sharper image? Estimates of the precision of nighttime lights as a proxy for economic statistics. Journal of Economic Geography, 15(1), 217–246. https://doi.org/10.1093/jeg/lbu010

Ortega-Bastida, J., Gallego, A. J., Rico-Juan, J. R., Albarrán, P. (2021). A multimodal approach for regional GDP prediction using social media activity and historical information. Applied Soft Computing, 111, 107693. https://doi.org/10.1016/j.asoc.2021.107693

Pesaran, M. H., Timmermann, A. (1995). Predictability of stock returns: Robustness and economic significance. The Journal of Finance, 50(4), 1201–1228. https://doi.org/10.1111/j.1540-6261.1995.tb04055.x

Proietti, T., Giovannelli, A., Ricchi, O., Citton, A., Tegami, C., Tinti, C. (2021). Nowcasting GDP and its components in a data-rich environment: The merits of the indirect approach. International Journal of Forecasting, 37(4), 1376–1398. https://doi.org/10.1016/j.ijforecast.2021.04.003

Qadir, J., Ali, A., Zwitter, A., ur Rasool, R., Zwitter, A., Sathiaseelan, A., Crowcroft, J. (2016). Crisis analytics: big data-driven crisis response. Journal of International Humanitarian Action, 1(1), 1–21. https://doi.org/10.1186/s41018-016-0013-9

Román, M. O., Wang, Z., Sun, Q., Kalb, V., et al. (2018). NASA's Black Marble nighttime lights product suite. Remote Sensing of Environment, 210, 113–143. https://doi.org/10.1016/j.rse.2018.03.017

Stock, J. H., Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association, 97(460), 1167–1179. https://doi.org/10.1198/016214502388618960

Stokes, E. C., Román, M. O. (2022). Tracking COVID-19 urban activity changes in the Middle East from nighttime lights. Scientific Reports, 12(1), 8096. https://doi.org/10.1038/s41598-022-12211-7

Sutton, P. C. (2007). A scale-adjusted measure of "Urban sprawl" using nighttime satellite imagery. Remote Sensing of Environment, 86(3), 353–369. https://doi.org/10.1016/S0034-4257(03)00078-6

Szumilo, N. (2021). Prices of peers: Identifying endogenous price effects in the housing market. The Economic Journal, 131(639), 3041–3070. https://doi.org/10.1093/ej/ueaa129

The Economist, Solstad, S. (2023). The economist war-fire model. First published in the article "A hail of destruction". The Economist, 23 February 2023. Retrieved from https://github.com/TheEconomist/the-economist-war-fire-model

Woloszko, N. (2020). Tracking activity in real time with Google Trends. OECD Economics, 1634. https://doi.org/10.1787/6b9c7518-en

Woo, J., Owen, A. L. (2019). Forecasting private consumption with Google Trends data. Journal of Forecasting, 38(2), 81–91. https://doi.org/10.1002/for.2559

APPENDICES

APPENDIX A. Tables

Table 1. Model Comparison and Performance

Model	(1) Nightlights	(2) Twitter	(3) Google	(4) Combined	(5) Com. FE
A: Model fit					
R^2 in sample	64.60%	95.20%	99.20%	99.60%	99.60%
B: Main predictions					
Nat. growth ’20 to ’21	14.33%	-2.1%	-0.65%	1.4%	2.04%
Nat. growth ’21 to Mar ’22	-54.05%	3.60%	-5.43%	-29.82%	-30.14%
C: Validation sample performance					
Pre ’14 RMSE for levels	0.341	0.124	0.155	0.0734	0.065
Pre ’14 RMSE for changes	0.059	0.056	0.096	0.042	0.047
D: Prediction error correlation					
Correlation of residuals in the estimation sample:					
Nightlights model	1				
Twitter model	0.12	1			
Google model	-0.03	0.04	1		
Correlation of prediction errors in the validation sample:					
Nightlights model	1				
Twitter model	0.24	1			
Google model	0.26	0.69	1		

Notes: *Column (1)* is based on Equation 4, *(2)* on Equation 5, *(3)* on Equation 6, *(4)* on Equation 7 and *(5)* is the same as *(4)* but with region-fixed effects. *Panel A* presents R^2 statistics for models estimated on annual data between 2014 and 2020. *Panel B* presents out-of-sample predictions for years after the end of the estimation sample. *Panel C* presents model out-of-sample performance for the 10 biggest regions before the start of the estimation sample. It presents a goodness of fit statistic for levels predicted by the model, and changes in levels against actual changes. *Panel D* presents correlations of residuals of single-predictor models.

Table 2. Main Model Coefficients

	GTI P1		GTI P1		GTI P1		ln(Tw)	
Region	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Vinnytsia	-0.14	0.15	0.02	0.71	0.22	0.42	-0.03	0.04
Volyn	0.16	0.25	0.00	0.58	-0.09	0.28	-0.02	0.04
Luhansk	1.40	0.32	3.57	3.86	-3.99	3.10	0.09	0.04
Dnipropetrovsk	0.14	0.08	0.14	0.51	-0.13	0.33	0.06	0.04
Donetsk	0.98	0.34	1.52	0.73	-1.71	0.67	-0.03	0.06
Zhytomyr	0.12	0.31	-0.04	0.74	0.25	0.72	0.01	0.10
Zakarpattia	0.14	0.24	0.92	1.01	-0.83	0.92	-0.04	0.09
Zaporizhzhia	-0.03	0.23	-0.36	0.40	0.34	0.31	0.06	0.04
Ivano-Frankivsk	0.41	0.36	0.14	0.95	-0.28	0.58	0.05	0.07
Kyiv City	0.03	0.02	-0.09	0.07	0.04	0.04	0.12	0.03
Kyiv	-0.24	0.26	-0.57	2.16	1.88	2.93	0.00	0.04
Kirovohrad	0.23	0.19	0.82	0.80	-0.43	0.49	-0.01	0.04
Lviv	0.07	0.08	0.28	0.33	-0.19	0.28	0.00	0.04
Mykolaiv	0.01	0.07	0.06	0.19	-0.01	0.11	-0.02	0.04
Odesa	0.03	0.08	-0.10	0.29	0.12	0.25	0.06	0.05
Poltava	0.21	0.35	0.72	1.05	-0.49	0.88	0.04	0.04
Rivne	-0.22	0.13	-0.33	0.19	0.29	0.19	0.07	0.07
Sumy	0.24	0.16	0.73	0.73	-0.31	0.41	0.01	0.05
Ternopil	0.29	0.28	0.26	1.09	-0.01	0.58	0.02	0.07
Kharkiv	0.14	0.09	0.20	0.48	-0.21	0.42	0.01	0.07
Kherson	0.14	0.17	1.23	1.93	-0.90	1.28	-0.06	0.08
Khmelnytskyi	0.06	0.19	0.19	0.99	-0.06	0.88	-0.02	0.07
Cherkasy	0.09	0.21	-0.12	1.40	0.11	1.17	0.00	0.10
Chernihiv	-0.09	0.29	-1.65	1.92	1.61	1.55	-0.02	0.04
Chernivtsi	0.37	0.16	0.23	1.01	-0.12	0.35	-0.01	0.05
ln(NLI): Coeff.: 0.13, S.E.: 0.03 (All regions)								

Notes: The table presents coefficients of the main model estimated using Equation 7 on a sample of 175 observations for all regions with no russian presence in the years 2014 – 2020. GTI P1, P2 and P3 refer to the first three principal components of Google Trends. ln(Tw) refers to the natural logarithm of the count of tweets, and ln(NLI) is the log of nightlight intensity.

Appendix B. Figures

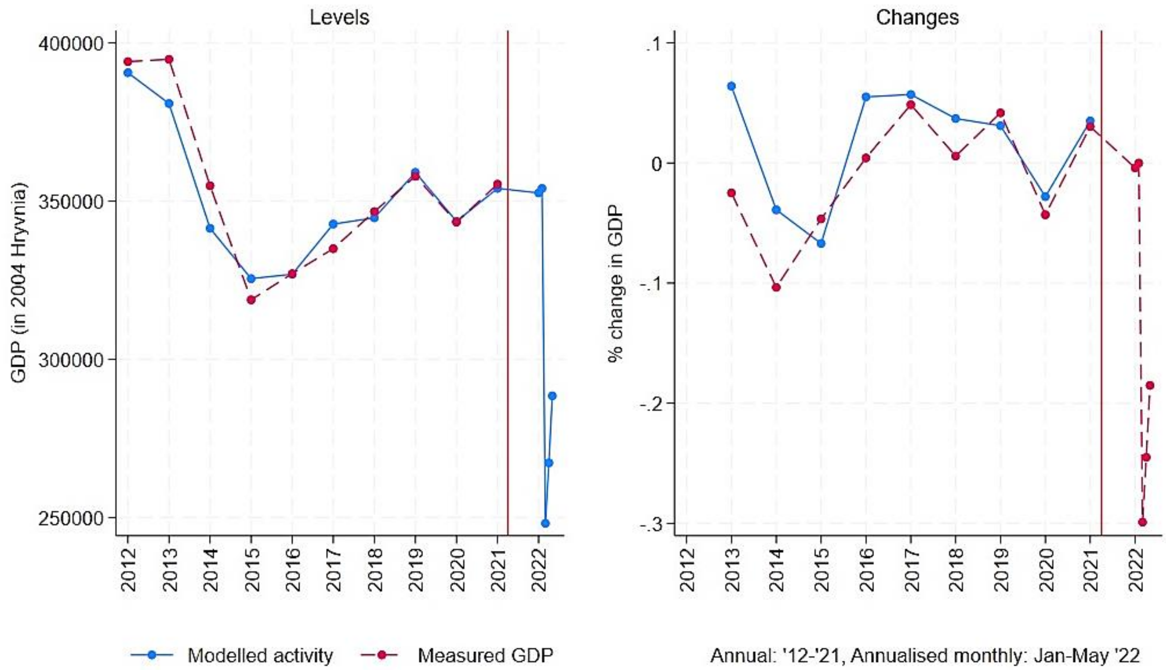

Figure 5. Actual and Predicted GDP in Ukraine – FE Model

Notes: The panel on the left gives the national GDP figures in Ukraine as aggregated from regions for which data is available. Measured GDP refers to the sum of all GDP figures reported by the SSSU. “Modelled activity” is the sum of all regional GDP figures predicted by our model. In-sample predictions are for 2014 to 2020, while all other predictions are out-of-sample. The panel on the right plots changes in the national aggregates over time. Data is annual for 2012 to 2021 and annualized monthly in 2022.

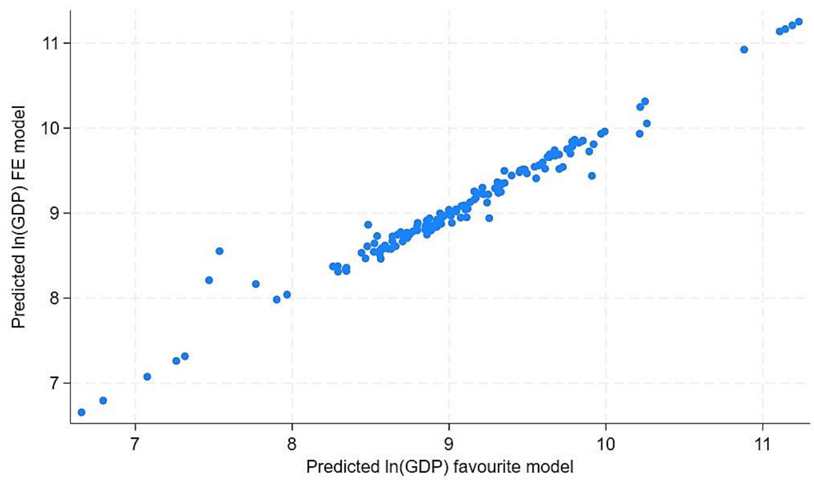

Figure 6. Monthly Regional Predictions in 2022 Favorite and FE Models

Notes: The figure plots monthly predictions in 2022 of our favorite specification against predictions of a model with region-fixed effects. There are five observations for each region - one for each month of our sample in 2022.

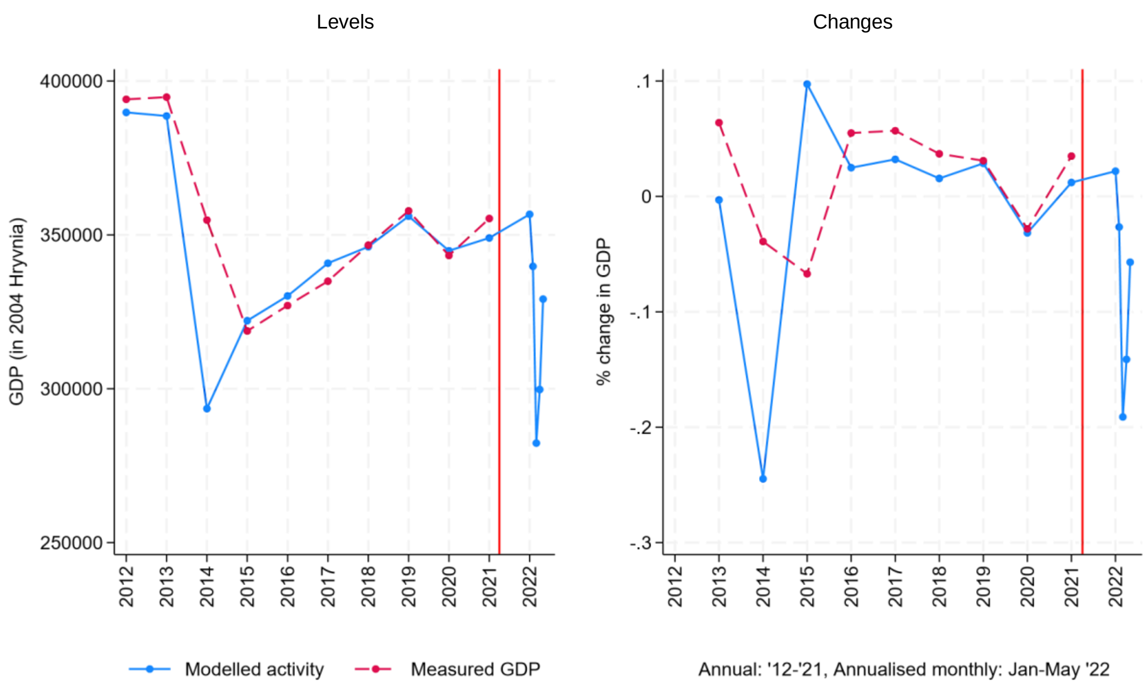

Figure 7. Actual and Predicted GDP of Ukraine – Random Forest

Notes: The panel on the left gives the national GDP figures in Ukraine as aggregated from regions for which data is available. Measured GDP refers to the sum of all GDP figures reported by the SSSU. "Modelled activity" is the sum of all regional GDP figures predicted by our model. In-sample predictions are for 2014 to 2020, while all other predictions are out-of-sample. The panel on the right plots changes in the national aggregates over time. Data is annual for 2012 to 2021 and annualized monthly in 2022.

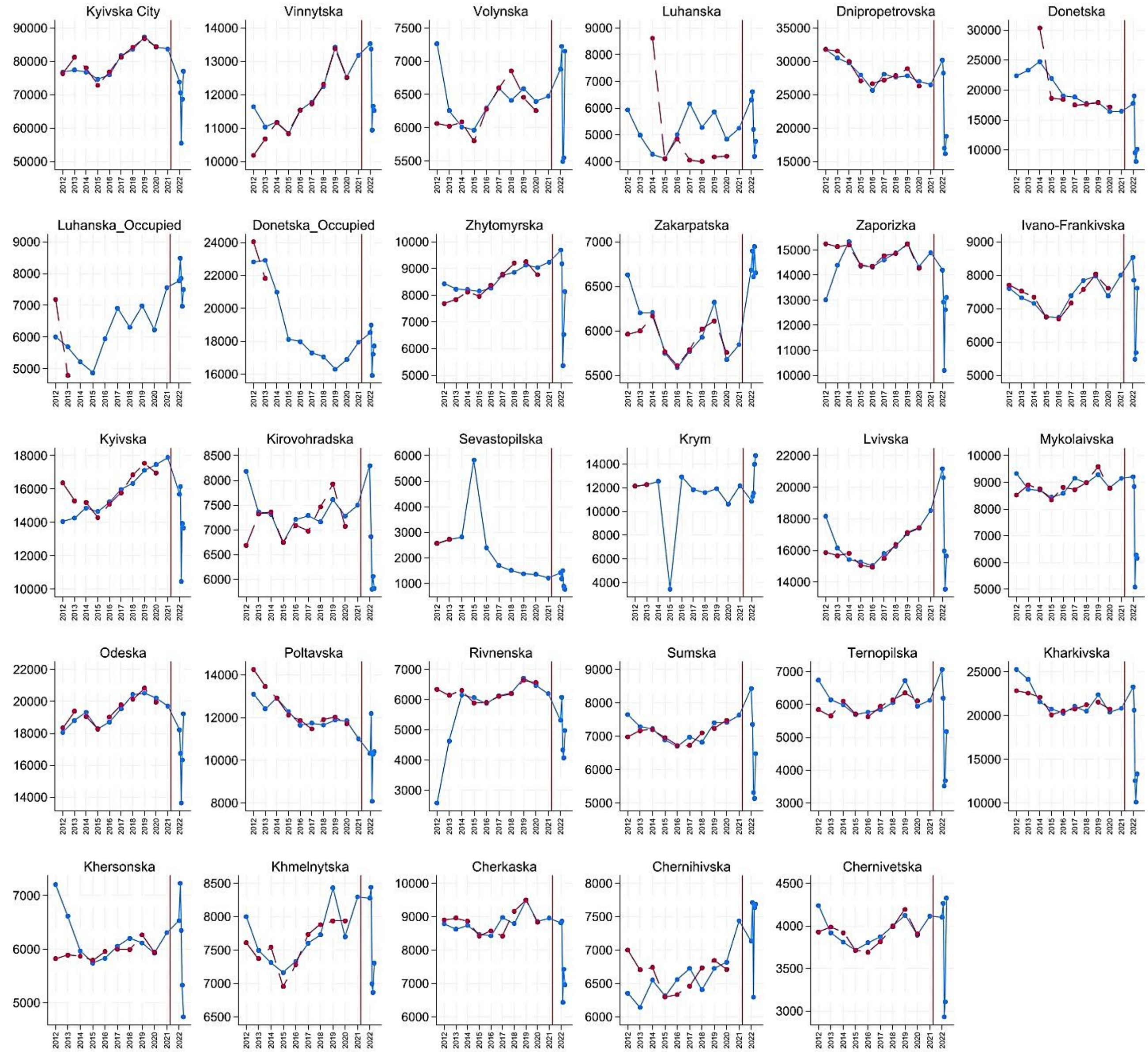

Figure 8. Actual and Predicted Regional GDP

Notes: Predicted activity is the continuous blue line and the actual GDP is the dashed red line. Areas occupied since 2014 are Crimea, Sevastopol and parts of Luhansk and Donetsk oblasts. We therefore use 2012 and 2013 data to estimate our model for Crimea and Sevastopol and the occupied parts of Donetsk and Luhansk oblasts. To generate a GDP figure for 2012 and 2013 for the parts of Donetsk and Luhansk oblasts that became occupied in 2014, we apportioned the total GDP reported for those years in each region to parts that later became occupied, based on nightlights.

Appendix C

C.1. Glossary of Key Terms

Predictions - The outputs of the models based on the predictor variables, model structure and model coefficients. For example, the predicted GDP levels based on nightlights data, the structure of the model that relates this data to GDP, and model coefficients.

Out-of-sample prediction - Making predictions on data not used in model estimation.

Used to evaluate predictive performance.

In-sample prediction - Predictions made on the data that was used to estimate the model.

Estimates - The coefficient values or parameters estimated in the regression models. For example, the β values relating nightlights to GDP.

Tracking - Approximating levels of a variable without being able to measure it directly. In our context, tracking refers to approximating GDP or economic activity on an ongoing basis.

Correlate - A variable that is correlated or associated with another variable of interest. For example, nighttime lights are a correlate of economic activity because areas with more lights tend to have higher GDP levels. However, correlation does not necessarily imply causation between the variables.

Nowcasting - Estimating the current economic situation in real-time, before official data is released using time-series methods. The difference between nowcasting and tracking is that the latter uses time-series models typical in macroeconomics, and the former relies on methods typical for applied microeconomics literature.

Structural break - A major change in the underlying relationship between variables, often due to an external shock or crisis.

Elasticity - The percent change in the dependent variable associated with a one percent change in the independent variable.

Endogeneity - When a predictor variable is correlated with the error term, biasing the estimates.

Measurement error - Discrepancies between the true values and measured values of variables.

Prediction error - The difference between the predicted value from a model and the actual observed value of the variable being predicted. For example, if a model predicts GDP will increase by 3% but a 5% increase is actually observed, the prediction error is -2% (the model under-predicted by 2 percentage points).

Microfoundations - Explaining relationships between variables like GDP and nightlights based on individual behavior and incentives.

Measured GDP - GDP reported by the authorities calculated using traditional methods such as surveys and sampling.

C.2. Theoretical Framework Continued

As discussed in Section 3, our framework builds on the common assumption that the relationship between economic activity Y and its correlation X is captured by an elasticity β (Henderson et al., 2011). Then, these two variables are proportional to each other in their natural logarithms, i.e. $y = \beta x$.[20]

In principle, this proportional relationship could arise from three sources or any combination of them. First, a common driving factor *c* could induce a spurious correlation. Second, y and x could directly affect each other. Finally, both variables might be measured in a way that implies a correlation between their measurement errors ϵ_y and ϵ_x. These notions are formalized in the following two equations:

$$y = a s_y + a_1 c + a_2 x + \epsilon_y \tag{8}$$

$$x = \mu s_x + \mu_1 c + \mu_2 y + \epsilon_x \tag{9}$$

where s_y and s_x refer to specific (and potentially uncorrelated) factors driving y and x, respectively, and a, a_1, a_2 and μ, μ_1, μ_2 are estimable parameters.

Economic activity can therefore be expressed by:

$$y = \frac{a}{1 - a_2\mu_2} s_y + \frac{a_2\mu}{1 - a_2\mu_2} s_x + \frac{a_1 + a_2\mu_1}{1 - a_2\mu_2} x + \epsilon_y + \frac{a2}{1 - a_2\mu_2} \epsilon_x \tag{10}$$

With measurement errors, the empirical relationship between measured values of Y and X is represented by:

$$y = \beta x + \epsilon \tag{10}$$

where ϵ is the difference between measured and predicted values of Y and has an expected value of zero. The above shows that parameter β can be rationalized by a combination of an accidental correlation in changes between s_y and s_x, changes in x and a correlation between measurement

[20] For brevity, we drop all time subscripts in this section.

errors. Our chief concern is that a war might affect any one of these three relationships, and that the value of β might react due to a structural change in the economy. As noted already, the existing literature has focused more on estimating β than explaining it with micro foundations or providing causal evidence using micro data. It is therefore difficult to comment on how they change due to a war.

C.3. Google Topics

Google uses topics to group searches in different languages. A topic groups terms that share the same concept in any language. An example from Google: "if you search London, and choose the corresponding topic, your search includes results for topics like "Capital of the UK" and "Londres," which is "London" in Spanish". Guided by the literature, our model uses the following 30 topics: labor, unemployment, investment, loan, economy, interest rate, inflation, bankruptcy, export, mortgage loan, baggage, holiday, BMW, Mercedes, McDonald's, Calvin Klein, Gucci, emigration, washing machine, fridge, Sony, Apple, fitness, recruitment, travel, savings, computer security, unemployment benefits, fashion and MBA degree. Expanding the list of topics by adding the following, did not affect the results: birthday, job, recession, show business, vehicle, foreclosure, energy, health, car rental, tourist attraction, home, garden, supermarket, restaurant, beauty salon, film, audit, business, agriculture, chemical industry, manufacturing, real estate, import, aviation and tobacco.

C.4. Results of the Fixed Effects Model

The results of a model with region-fixed effects are almost identical to our favorite specification. In Figure 5, we present the results of a national time series (equivalent to Figure 3 in the main body) and there are no material differences at this level to our main results either in-sample or out-of-sample. In Figure 6, we present a scatter plot of monthly predictions in 2022 of the two models. It shows that the predictions are almost identical apart from for one region (Ternopil) where the FE model predicts a higher GDP.