Conflict, social unrest and policy uncertainty measures are useful for macroeconomic forecasting *

Marina Diakonova
Banco de España
marina.diakonova@bde.es

Luis Molina
Banco de España
lmolina@bde.es

Hannes Mueller
IAE-CSIC and BSE
h.mueller.uni@gmail.com

Javier J. Pérez
Banco de España
javierperez@bde.es

Cristopher Rauh
University of Cambridge
cr542@cam.ac.uk

June 7, 2022

Abstract

It is widely accepted that episodes of social unrest, conflict, political tensions and policy uncertainty affect the economy. Nevertheless, the real-time dimension of such relationships is less studied, and it remains unclear how to incorporate them in a forecasting framework. This can be partly explained by a certain divide between the economic and political science contributions in this area, as well as the traditional lack of availability of high-frequency indicators measuring such phenomena. The latter constraint, though, is becoming less of a limiting factor through the production of text-based indicators. In this paper we assemble a dataset of such monthly measures of what we call "institutional instability", for three representative emerging market economies: Brazil, Colombia and Mexico. We then forecast quarterly GDP by adding these new variables to a standard macro-forecasting model in a mixed-frequency MIDAS framework. Our results strongly suggest that capturing institutional instability above a broad set of standard high-frequency indicators is useful when forecasting quarterly GDP. We also analyse relative strengths and weaknesses of the approach.

Keywords: forecasting; social unrest; social conflict; policy uncertainty; forecasting

*The opinions expressed in this paper are the responsibility of the authors and do not necessarily coincide with those of the Banco de España or the Eurosystem. We thank conference participants at the 41st International Symposium on Forecasting (June 2021), the 2021 Joint Statistical Meetings of the American Statistical Association (August 2021), and the IIF Workshop: Forecasting in a changing environment (December 2021), in particular Eva Senra. This paper is being written within the framework of the contract between the Banco de España and the Fundació d'Economia Analítica.

GDP; natural language processing; geopolitical risk.
JEL Classification: E37; D74; N16.

1 Introduction

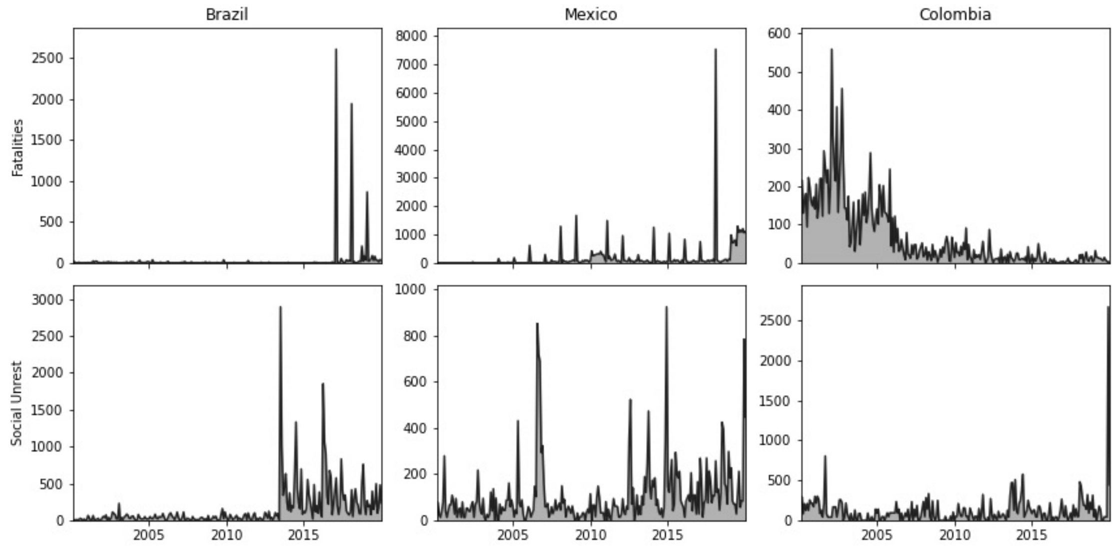
Social unrest, conflict, policy uncertainty, and, more generally, institutional instability, affect macroeconomic developments. The extant literature documents a negative relationship between such events and economic activity, in particular for emerging market economies (see, among others, Hadzi-Vaskov et al. (2021), Barrett et al. (2021), Saadi Sedik & Xu (2020)). Nonetheless, the economic forecasting literature has not elaborated on this substantial evidence and assessed, by unveiling real-time economic shocks, whether it also brings information that could help inform decision-taking by private and public agents. This might in part be related to the traditional scarcity of monthly and quarterly measurable indicators of such phenomena. Recent years, however, have brought about a host of typically text-based, databases which capture institutional instability for a large number of countries worldwide, like Mueller & Rauh (2022a), Barrett et al. (2021) and Caldara & Iacoviello (2022).

In this paper we incorporate variables of this sort in an otherwise standard, mixed-frequency time-series forecasting set up to assess their value for forecasters. We test the hypothesis that those variables are relevant focusing on three paradigmatic cases from Latin America: Brazil, Colombia, and Mexico. These are relatively large economies, with good macroeconomic data availability, while at the same time all have a history of social unrest and conflict, which makes them particularly suitable for the problem at hand.

Figure 1 shows examples of the kind of instability we are interested in. The first row shows the number of fatalities due to armed conflict according to the Uppsala Conflict Data Program (UCDP) whereas the lower panel shows the social unrest index from Barrett et al. (2021). In both timelines we see dramatic shifts in the three countries. Both Mexico and Brazil are destabilizing in our period of interest whereas Colombia is entering a peace process which is, however, accompanied by significant social unrest. The focus on these three economies allows us to provide a study on the contribution of different indicators, making the interpretation of results and messages transparent, something that might get somewhat blurred in studies covering large pools of countries.

We find that adding measures of institutional instability improves predictions. More specifically, for each of three analysed countries the forecast improves for at least 80% of the

Figure 1: Example of institutional instability measures such as social unrest and fatalities from political violence for Brazil, Mexico and Colombia.



Note: The top panel shows fatalities due to armed conflict events according to the UCPD GED dataset. The bottom panel shows data on social unrest from Barrett et al. (2020).

predicted periods by adding some combination of institutional instability proxies. For Mexico and Brazil we find a clear pattern in which the gain from adding institutional instability measures is largest at longer forecast horizons and are strongest in the first month of the quarter where traditional variables are less useful or less available.

Colombia features the strongest and most persistent gains from institutional instability but the pattern across forecast horizons and months is less clear here. We know from Figure 1 that Brazil and Mexico suffered dramatic escalations of instability in our sample range, whereas Colombia entered the sample with extremely high levels of violence and benefited from stabilization attempts. This might easily lead to gains for Colombia which are basically flat for longer horizons as expectations in the country did not undergo such dramatic shifts.

There is a growing literature on nowcasting and forecasting macroeconomic developments in emerging market economies. It tends to follow the usual approach taken for developed economies, along the lines of, e.g. Giannone et al. (2008). These papers use economic and financial indicators covering real economy activity and prices, including variables related to the housing market, the labor market, money and credit aggregates, and international developments (for studies covering the main Latin American economies see Leiva-Leon et al. (2020), Corona et al. (2020), López et al. (2021), and references therein). More recently Cepni et al. (2020) also account for global economic policy uncertainty and surprise indices based on a variety of different local and global datasets. The sort of international spillovers derived from the latter are particularly important for emerging markets, as also signaled by a related literature (see, among others, Carriere-Swallow & Cespedes (2013), or Gauvin & C. McLoughlin (2014)). We build on this literature, but expand it to include measures of social unrest, conflict, and institutional instability in general. By doing so, we put together two strands of the literature: that dealing with macroeconomic forecasting, as surveyed before, and another one that focuses on forecasting conflict, more from a political-economy point of view (see e.g. Mueller & Rauh (2022a)).

In our forecasting exercise we follow an additive approach, asking ourselves how much, if anything, the novel text-based variables add to forecasting models based on traditional macro-financial monthly indicators. The focus is on forecasting quarterly GDP for the three countries under study. The modeling framework is the time-series mixed-frequency MIDAS approach. Three types of text-based indicators are incorporated into the models: (i) conflict risk, updated following Mueller & Rauh (2022a), that have been developed by applying a combination of supervised and unsupervised ML techniques over a wealth of news sources; (ii) economic policy uncertainty (EPU) indicators, taken from Ghirelli et al. (2021), developed using supervised machine learning techniques, by means of computational text analysis applied to a wealth of Spanish newspaper sources; (iii) other measures of socio-political conflict Barrett et al. (2021), Caldara & Iacoviello (2022). In addition, we evaluate the usefulness of adding to the models the more standard political risk indicators, including those reflected in ratings of international agencies.

The rest of the paper is organized as follows. Section 2 deepens the discussion on why it makes sense to use "institutional instability" indicators to forecast macroeconomic developments. Section 3 describes the data used, while Section 4 outlines the econometric methodology used (MIDAS). Section 5 explains the empirical setup and Section 6 discusses the results of the forecast combination exercise. Finally, Section 7 draws the main conclusions and policy implications of our analysis.

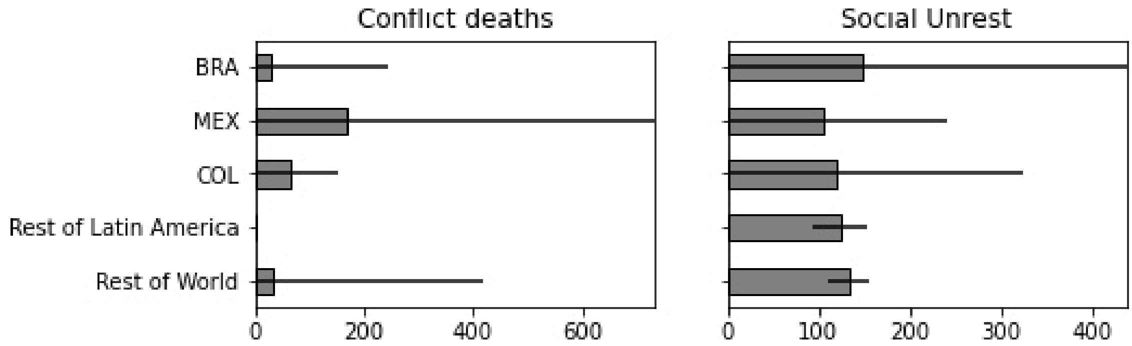
2 Institutional instability: Impact and measurement

The literature has identified several transmission mechanisms for how political turmoil and social unrest can affect the economy. Work on the economic effects of economic policy uncertainty suggests that political divisions could impede economic development by hindering investments and hiring (Baker et al., 2016; Bloom, 2009). Besley & Mueller (2018a) demonstrate that foreign investment shuns countries with weak executive constraints because economic volatility is larger in countries without strong checks and balances. They also find that sector-specific political factors such as political connections and bribery seem to play a role. Hassan et al. (2019) show how firms that are exposed to political risk retrench hiring and investment and actively lobby and donate to politicians to manage this risk. When social unrest leads to looting and destruction this will affect firms' investment and hiring decisions.

The security of property rights has long been identified as a central pillar of capitalism, and the corresponding insecurity has far-reaching consequences (Johnson et al., 2002; Besley & Ghatak, 2010). It is not only the direct loss due to predation that matters here, but also the effort spent on securing output and production that will distort the economy, Besley & Mueller (2018b). When political or social conflict escalates into armed conflict, the costs can increase dramatically (Collier, 1999; Abadie & Gardeazabal, 2003; Mueller & Tobias, 2016). Yet another important channel is expectations. In the theoretical discussion of Barro (2009) some of the political events that can be triggered by conflict could be regarded as rare disasters which affect asset prices. Asset prices can also be used empirically to show how expectations change with political violence. Zussman & Zussman (2006) and Willard et al. (1996) show, for example, that asset prices during conflict react to important conflict events like battles or ceasefire agreements. Besley & Mueller (2012) show that house prices seem to react to changes in expectations rather than violence itself. This implies that even once violence and turmoil are over it may happen that asset prices stay suppressed. Parts of the economy will then only recover once peace is regarded as stable.

Measuring instability is particularly relevant when a political situation might escalate or is escalating as during the Arab Spring or the wave of instability that dominated Latin America

Figure 2: Institutional instability in Brazil, Mexico and Colombia in the international perspective



Note: Figure shows the mean and standard deviation, in the time period of interest, of the estimate of fatalities due to armed conflict events according to the UCPD GED dataset (left) and of the social unrest index from Barrett et al. (2020) (right).

in the 2010s. This is clearly visible in the uncertainty measure provided by Baker et al. (2016) which shoots up when, for example, violence escalates, as this creates an extremely fluid situation. It is in these situations where we expect the biggest gains from quantitative measures of risk as this should allow for better forecasts.

The measurement of the different angles of "institutional instability", namely conflict (whether armed, or violence more broadly), social unrest, or policy uncertainty, has benefited dramatically from advancements in textual analysis, i.e. the use of machine learning algorithms to find topics or combination of relevant keywords in massive amounts of text related to the issues of interest (Mueller & Rauh, 2022a, 2018; Barrett et al., 2020; Ghirelli et al., 2021; Caldara & Iacoviello, 2022).

In this paper we focus on three Latin American economies, namely Brazil, Colombia, and Mexico. We consider that this is a relevant selection because: (i) there is already a reference literature on macroeconomic forecasting using traditional variables for these countries, given good macroeconomic data availability; (ii) the three economies present a history of social unrest, conflict and institutional instability.

Figure 2 illustrates the extent and variation in the number of fatalities during conflicts from UCDP and social unrest from Barrett et al. (2020) comparing the three selected countries to other countries in Latin America and the world. In the left panel we see that the countries in our sample are outliers in terms of the number of fatalities they have suffered when compared to the rest of Latin America. However, according to `conflictforecast.org`, armed conflict risks are currently spreading in several other countries in the region, and Mexico, Brazil and Colombia experienced large breaks in their timelines which makes their analysis particularly relevant in a situation of spreading risks. In addition, several emerging markets elsewhere, like Turkey or Pakistan, are suffering high levels of armed violence on their soil, and even developed countries in Europe or the United States have experienced waves of riots, racist, terrorist violence, and outright war, in their recent past.

In terms of social unrest our sample countries have sample means that are comparable to both the rest of Latin America and the world. Our analysis on this front is particularly relevant for other countries in Latin America which have a long history in economic and political turmoil (see e.g. works by Daron Acemoglu). Several countries in the region have recently been affected by outbreaks of social unrest and political crisis, often in a context of significant income inequality.[1] Given the high levels of inequality in our sample of countries, social unrests might continue to surface due to a lack of cohesion of the studied societies.

3 The data

We report the data used in the study in Table 1. The database starts in 2000. Beyond data availability, there are institutional reasons to select that year: Brazil had already adopted the current macroeconomic policy framework, after the currency crisis of January 1999 and the abatement of hyperinflation since mid 1994; Mexico was in transition to adopt the inflation targeting regime; and Colombia recovered from the September 1999 devaluation of the peso.

The series of interest is real GDP. The data sources are the National Institute for Geography and Statistics of Brazil (IBGE in Portuguese), the National Institute of Statistics and Geography of Mexico (INEGI in Spanish) and National Administrative Department of Statistics (DANE in Spanish). All series are seasonally adjusted by the respective statistical

[1]According to the Interamerican Development Bank (IDB) income inequality might be a key problem. According to the IDB (2020), the richest 10% of the population earns 22 times the income earned by the bottom 10% in the average country of Latin America and the Caribbean. The average Gini coefficient is 0.46. Both numbers are significantly higher than in the OECD or other comparable development countries.

offices.

For macro-financial data we use a set of widely-used indicators common to the three countries under study.[2] In particular: (i) "Hard indicators": industrial production index; retail sales index; unemployment rate; exports in volume; credit to private sector in real terms; fixed investment; and construction sector production (in the later two cases we use country-specific proxy variables, as detailed in Table 1); (ii) "Soft indicators": consumer confidence index; industrial or manufacturing confidence index ("Business Confidence Index"); (iii) Financial markets and political risk indicators: EMBI+ spread in basis points;[3] and the sovereign rating (an average of the ratings of the three major agencies, Standard and Poor's, Fitch and Moody's), linearized using a scale from 21 (AAA) to 12 (BBB-) and 0 (RD or D).

As regards social unrest, conflict and policy uncertainty, we rely on measures elaborated using textual analysis applied to newspaper sources. These are increasingly used to measure conflict events or other political risks and uncertainties. These news-based measures have the advantage of being available and updated at monthly frequencies or higher. One of the hallmarks of text-based indexes is the economic policy uncertainty (EPU) measure by Baker et al. (2016). The EPU index quantifies newspaper coverage of policy-related economic uncertainty through a combination dictionary which captures terms related to both economy, economic policy, and uncertainty. We use a novel adaption of this method developed by Ghirelli et al. (2019), Ghirelli et al. (2021), applied to Latin American economies for the period starting in January 1997 and using Spanish newspapers. The idea of the index is to count words in three lists. The *Economics*-terms or E-terms captures the extent to which economics is discussed. The *Policy*-terms or P-terms capture whether a text discussed something policy related. Finally, the *Uncertainty*-terms or U-terms captures whether the

[2]For some variables (retail sales, unemployment rate, credit, consumer confidence) it was necessary to interpolate a few data points at the beginning of the series, due to data limitations. To do so we resorted to either simple linear interpolation, or proxy variables to which the one of interest has a historical empirical relationship (for example the consumer confidence index with the retail sale index in Mexico for January 2000 to March 2001 and in Colombia for January 2000 to October 2001). Finally, all series have been sourced from Refinitiv, and those that originally were not available seasonally-adjusted from official sources have been adjusted using Tramo Seats via JDEMETRA+.

[3]The yield of a Brazilian or Mexican or Colombian synthetic external debt bond minus the equivalent yield of an US bond of the same maturity (in this case five years). This series are published daily by JP Morgan and have been sourced from Refinitiv.

writer considers a situation to be uncertainty. Only when at least one term from all three groups coincides in the text is it counted towards the EPU index.

The second measure we use is the Geopolitical Risk index elaborated by Caldara & Iacoviello (2022). This is a news-based index counting the occurrence of six groups of words related to geopolitical tensions in 11 leading international newspapers for 18 emerging market economies since 1985. The groups span explicit mentions of geopolitical risk and military tensions, nuclear tensions, war threats, terrorist threats, and actual adverse geopolitical events.

The third aspect of institutional instability we focus on comprises a set of risk measures from the webpage `conflictforecast.org` which follows the methodology of Mueller & Rauh (2022a,b). The page provides monthly out-of-sample forecasts for the outbreak of "armed conflict" and "any violence" three and twelve months in the future which we interpret here as a measure of broader political fragility. The forecast relies on variables that capture the conflict history of a country (monthly conflict event data updates from Uppsala conflict Data Program (UCDP)) and the news landscape through automated news summaries from a corpus of over five million articles. The methodology in Mueller & Rauh (2022a) uses a so-called *topic model* (or LDA) to summarize articles into topics first and then uses the share of each topic as a variable to predict the risk of a armed conflict outbreak in three and twelve months ahead. The method deviates from other methods as it is not based on a dictionary but on a supervised machine learning model which predicts the outbreak of armed conflict ahead of time and endogenously picks which the topics according to whether they predict risk - either through a positive or negative association. We use all eight models provided at the country level on the webpage. Apart from that, we include time-series of topic shares of three relevant topics which feature prominently in the construction of the conflict models.

The fourth measure we use is the new social unrest index developed by Barrett et al. (2020) at the IMF which is also based on counts of relevant media reports. The index uses a dictionary of words like "protest" or "riot" to produce individual monthly time series for 130 countries.

Table 1: Variables and related transformations

Variable	**Shorthand**	**Regressor Category**	**Notes**	**Sources**
GDP	GDP	Dependent variable	Real GDP seasonally adjusted (SA)	National Statistics Offices
Industrial production	Ind.Prod.	Traditional: hard	Industrial prod. index (SA)	National Statistics Offices
Retail sales	Ret.Sales	Traditional: hard	Retail sales index (SA)	National Statistics Offices
Credit	Credit	Traditional: hard	Nominal credit to private sector deflated by CPI	
Exports	Exports	Traditional: hard	Volume, SA	National Statistics Offices
Unemployment rate	Unempl.Rate	Traditional: hard	Rate, SA	National Statistics Offices
Production of construction	Prod.Constr.	Traditional: hard	Volume (SA) (Col.: building permits)	National Statistics Offices
Fixed investment	Fix.Inv.	Traditional: hard	Volume (SA) (Basil: prod. cap. goods; Col.: import cap. goods)	National Statistics Offices
Consumer confidence	Cons.Conf.	Traditional: soft	Level (SA)	
Business confidence index	Bus.Conf.Ind.	Traditional: soft	Level (SA)	Basil: F Getulio Vargas; Mex., Col.: OECD
Sovereign rating	Sov.Rat.	Political	Average SP, Moody's, Fitch	SP, Moody's and Fitch
Geopolitical risk index	GPR	Political	Level	Caldara-Iacovello
Emerging markets bond index	EMBI	Financial	Spread over US Treasury, bps	JP Morgan
Economic policy uncertainty	EPU	EPU	Level	Ghirelli et al.
Reported social unrest index	Soc.Unr.	Social Unrest	Level	RSUI IMF
Topic: politics	Top.Pol.	Conflict: media topics	topic1	Mueller and Rauh
Topic: economics	Top.Econ.	Conflict: media topics	topic6	Mueller and Rauh
Topic: conflict	Top.Conf.	Conflict: media topics	topic10	Mueller and Rauh
Armed conflict 12 months text	Arm.Conf.12.text	Conflict: indicators	Text model	Mueller and Rauh
Armed conflict 12 months best	Arm.Conf.12.best	Conflict: indicators	Best model	Mueller and Rauh
Armed conflict 3 months text	Arm.Conf.3.text	Conflict: indicators	Text model	Mueller and Rauh
Armed conflict 3 months best	Arm.Conf.3.best	Conflict: indicators	Best model	Mueller and Rauh
Any violence 12 months text	AnyViol.12.text	Conflict: indicators	Text model	Mueller and Rauh
Any violence 12 months best	AnyViol.12.best	Conflict: indicators	Best model	Mueller and Rauh
Any violence 3 months text	AnyViol.3.text	Conflict: indicators	Text model	Mueller and Rauh
Any violence 3 months best	AnyViol.3.best	Conflict: indicators	Best model	Mueller and Rauh

NOTE: Before analysis both GDP and the hard indicators in the "traditional" set variables are transformed using quarter-on-quarter and month-on-month difference of the logarithm of the baseline values. Month-on-month differencing is also applied to EMBI.

4 Methodology

Forecasting quarterly GDP using a combination of monthly and quarterly regressors requires a method capable of dealing with mixed frequencies. Specifically, we would like to forecast a low-frequency dependent variable from exogenous predictors at both equal or higher frequencies. However, in spite of the proliferation of models, there does not exist a consensus about which model should be used under the given circumstances

4.1 MIDAS

The MIDAS framework was proposed by Ghysels et al. (2004) in order to deal with the potential issue of proliferation of parameters. The general idea is to introduce a flexible parametric restriction weighing the time-lagged regressor values according to some *a priori* notion of their relative importance. This weighting can be accomplished with one of several common functional forms, typically polynomials on the lag order, or distributed lag polynomials. Such functions are useful as this way the entire relative weighting scheme becomes defined by a small number of hyperparameters. The additional constraint makes the regression non-linear on the lag orders. This, in turn, requires the extra-step of selecting the specific numerical optimisation algorithm, or fine-tuning the initial value of the hyperparameters to aid in convergence.

Consider the dependent low-frequency variable Y_t and high-frequency regressor (predictor) X_t. Note that since we are using the index t to stand for some "absolute" time, and hence capable of referencing both high and low frequency times, we are implicitly assuming that Y_t will not be defined for every t. Let L be the lag operator that, given some reference time t, obtains a vector of variable lags according to some specification, for instance with a varying lag starting point. This is often referred to as *frequency alignment.* The MIDAS model can then be written as

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i L^i Y_t + \gamma_0 \sum_{k=1}^{M} \Phi(k;\gamma) L_{HF}^k X_t + \epsilon_t,$$

where $\Phi(k;\gamma)$ is the weighting scheme, p is the order of the auto-regressive element, M is the number of lags (values) of the high-frequency predictor we aim to regress on, and the X_t and Y_t on the RHS are technically vectors of elements up to and including time t (the recent elements of which will therefore be frequency-aligned by the L operators). Note that since the auto-regressive coefficients β admit non-zero values, this form is also known as MIDAS-ADL (Clements & Galvão, 2008). Note also that the above form is naturally extendable to an arbitrary number of regressors.

There are several common weight specifications. One of the most widely used functional forms is the normalised exponential Almon polynomials (NEALMON), given by:

$$\Phi(k,\gamma) = \frac{\exp(\sum_{i=1}^{s}\gamma_i k^i)}{\sum_{j=1}^{M}\exp(\sum_{i=1}^{s}\gamma_i j^i)},$$

where s defines the polynomial order.

Setting all γ to zero is equivalent to weighting all the lagged values equally, and is sometimes referred to as time averaging (TA), since this implies the dependent variable is regressed on the mean of the lagged predictor values. Strictly speaking, this is a stand-alone method and does not require any restrictive weight specification setup. However, since it is still based on *frequency alignment*, and can be understood as a particular case of the MIDAS mathematical framework, we treat it as special cases of MIDAS. This is also arguably the simplest way of mapping higher-frequency variables into lower-frequency space. The implicit assumption in this approach is that the predictive power of each value of the regressor is the same, whether it is closer to the beginning or to the end of the low-frequency period. In our analysis we will use both NEALMON and TA methods.

We have also considered two additional variations of the MIDAS models, U-MIDAS and the Beta polynomials. U-MIDAS, or *unrestricted* MIDAS, refers to the special case when the non-linear restrictions on the lags, given by the weight specification function, are absent. Each lagged regressor value is treated separately, and the model turns into a standard linear regression, with coefficients that could be estimated using OLS. U-MIDAS can thus be prone to parameter proliferation if the number of lags or the number of regressors becomes too large. In our empirical specification U-MIDAS consistently underperformed compared to NEALMON and TA models, and so we only show U-MIDAS results in the appendix.

Models based on the Beta polynomials, on the other hand, have had consistent convergence problems, and the number of parameters for which such models would therefore be absent is an order of magnitude larger than the corresponding numbers for the NEALMON and TA methods. We therefore did not proceed with including the Beta-polynomial models in our final analysis.

Thus, all in all, the final results are based on four model specifications:

1. TA: Time averaging
2. TA-ADL: TA regression with an auto-regressive term
3. Nealmon: MIDAS regression with coefficients constrained by the normalised exponential Almon polynomials
4. Nealmon-ADL: Nealmon regression with an auto-regressive term.

For individual regressor results (available on request) we also consider a benchmark model, an AR(p) where the lag order p is defined by minimising the AIC in the initial training sample. The same lags are then used to define the autoregressive term in the ADL specifications used in the models that later form part of the forecast combinations. Note that in all the cases the information set timings are respected, and so the optimal lags are dependent on both the country and forecast month m.

Combining forecasts into a single model has been the approach preferred in literature for more than a decade, see Timmermann (2006). This shift to an a posteriori analysis gets round the curse of dimensionality, which would plague those models that, like MIDAS, admit multiple regressors (the number of variables in macroeconomic forecasting can typically reach hundreds). As such, it presents an alternative to dimensionality reduction techniques such as dynamic factor models or principal component analysis, though it does of course come with what is effectively an additional hyperparameter of the specific combination method. Nevertheless, if the ideal single regressor is not known, or, more likely, does not exist, it often results in more accurate forecasts as the errors of individual forecasts get mitigated. We therefore follow this strategy and follow the methodology of first, designing an optimal forecast for each individual regressor, and second, combining the forecasts using standard combination techniques.

Single predictor forecasts For each predictor we first use the initial window to estimate the optimal hyperparameters: the number of lags of the regressor, and the coefficients of the weight function in the case of normalised exponential Almon polynomials. The latter we take from a list of default parameters considered in Ghysels et al. (2016), Armesto et al. (2010), and own designs, which correspond to 21 weighting schemes, and represent a number of possible shapes (i.e. hypotheses about the relative importance of time lags). The maximal number of lags, starting from the ragged edge, for the monthly regressors is 12, and for the auto-regressive variable is four. These lags thus cover at most a year´s worth of past available data. After estimating the optimal hyperparameters, we carry out a recursive forecast, where the GDP growth is estimated based on the model fit within the new window. For our computations we use the **midasr** R package developed by Ghysels et al. (2016). We discard any model that has convergence issues, or that results in errors or warnings at any point during the training or forecasting stages.

Forecast combinations To create the final forecast we combine individual optimised forecasts computed above across specific groups of variables. Three weighting schemes are considered:

1. Equal weights (EW). Under this scheme each of the N forecasts carries a fixed equal weight of $1/N$. In spite of this scheme's simplicity, literature suggests that this leads to more accurate forecasts in terms of the mean square forecast error (Smith and Wallis 2009). In fact, the apparent contradiction between this simplicity and the performance is known as the "forecast combination puzzle" (see for instance Claeskens et al. (2016)).

2. Mean squared forecast error (MSFE). This scheme is part of the common set of schemes where the weights are estimated. In particular, the MSFE weighs high those forecasts that produce a lower mean square forecast error. Consider the T forecasts indexed by j, $f^j = (f_i^j)_{i=1}^N$, as well as the true values $y = (f_i)_{i=1}^N$. Then the weight of each

forecast j in the final combination forecast $w^j = \frac{v^j}{\sum_j v^j}$, where v^j is the inverse RMSFE of forecast j, $v^j = \left(\sum_i (f_i^j - y_i)^2/N\right)^{-1}$.

3. Discounted mean squared forecast error (DMSFE). The idea behind this approach is to give higher weights to those forecasts that performed well recently by modulating the components of the residual in an exponential manner. Thus each residual $r_i = (f_i^j - y_i)^2$ at time i appearing in v^j becomes $r_i \to r_i * \beta^{N-i}$, with $0 < \beta \leq 1$. Under $\beta = 1$ we recover the MSFE method. For the purpose of this exercise we use $\beta = 0.9$, a default value suggested within the module.

We also computed the forecast using Bayesian information criteria weights (BICW), but as the resulting trends varied greatly from the other weight schemes, and often corresponded to forecasts no different from the benchmark models, we do not consider this scheme appropriate for the task at hand.

4.2 Evaluation

We evaluate the forecast based on its predictive accuracy. Consider the forecast $f = (f_i)_{i=1}^N$, the benchmark forecast $b = (b_i)_{i=1}^N$, and the true values $y = (y_i)_{i=1}^N$. Let $e_i^f = y_i - f_i$ be an element of the residual of forecast f. Then, because we would like penalise the forecasts in a manner that weighs large errors more than smaller ones, we consider the squared loss function. The resulting root mean square forecast error is then $RMSFE = \sqrt{\frac{1}{N}\sum_i^N (e_i^f)^2}$.

We then follow the standard procedure of computing the forecast RMSFE relative to the RMSFE of the benchmark forecast b. For individual forecasts (mentioned in the appendix) the benchmark is AR(p), and for forecast combinations it is a combination of the more traditional variables, to be specified further in the text. The resulting value is less than unity if the forecast is going a better job at point predictions, and more than unity if it does not manage to beat the benchmark. As part of our analysis we have also calculated the Diebold-Mariano statistics, the Pesaran-Timmerman tests, and measured the accuracy of ex-post annual GDP forecasts. However, since they do form the focus of the results, we will only mention them in the footnotes/appendix where appropriate.

5 The forecasting exercise

5.1 Empirical approach

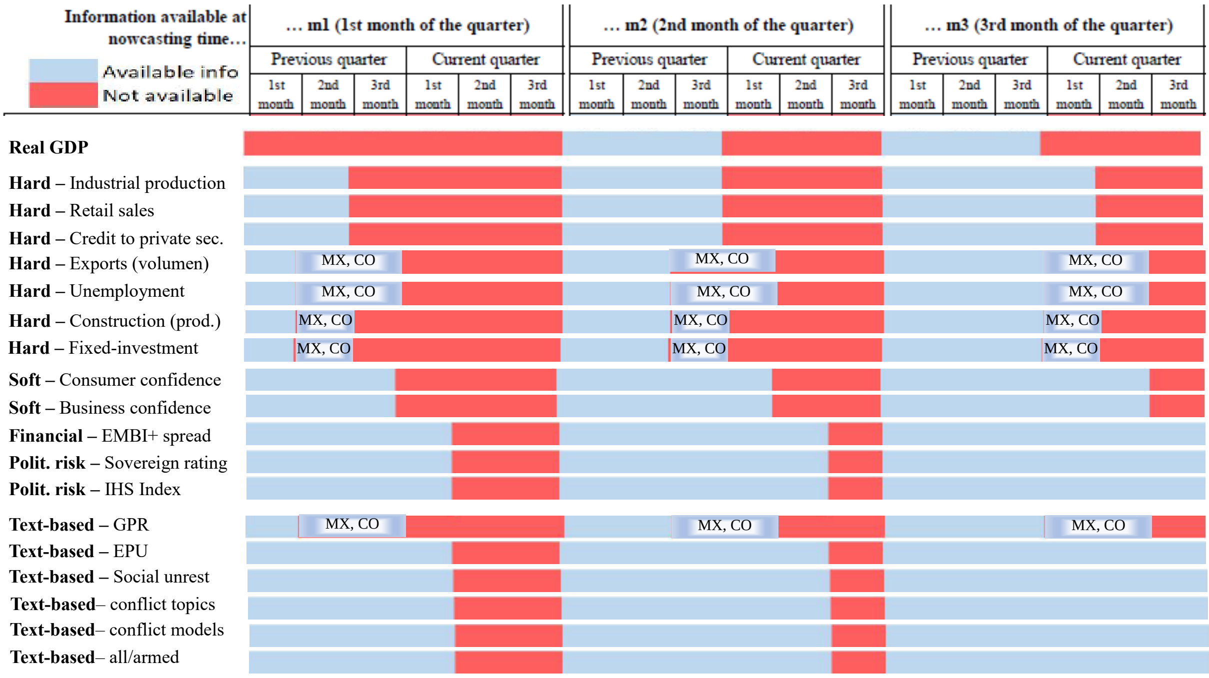
We use the MIDAS framework to test the predictive power of text-based indicators for forecasting quarterly GDP growth rates. Specifically, let time be indexed by month $m \in 1, .., 12$, quarter $q \in 1, .., 4$, and year y. Then every (m, q, y) is the forecast origin associated with some quarterly growth forecast at horizon h quarters ahead $\hat{Y}(m, q+h, y) = 100\ log(\hat{\text{GDP}}(q+h, y)) - 100\ log(\text{GDP}(q+h-1, y))$, with $h \in 0, 1, 2, 3$. This is our primary object of interest. Note the m dependency of the estimate, which reflects the fact that the information set grows as we get closer to end of the quarter. To evaluate quarterly forecasts we therefore compare $\hat{Y}$ with $Y(m, q+h, y) = 100\ log(\text{GDP}(q+h, y)) - 100\ log(\text{GDP}(q+h-1, y))$, for the pairs (m, h).

Our sample runs from Q1 of 2000 to Q4 of 2020. We subdivide our 20 year sample into two equal parts, and consider a recursive forecast with a training window of 10 years, the first pass from 2000 Q1 to 2009 Q4.

5.2 Information set available in (pseudo) real-time

A relevant question is to determine which information is available at the moment of running the forecasting models in real-time, that is, the publication lags of each of the above-mentioned indicators. This is typically referred to as the ragged-edge problem. Figure 3 summarizes this information. Given that the aim is to forecast quarterly GDP, we adopt the perspective of a forecaster that wishes to produce quarterly forecasts at each moment in time during the year. $m1$, $m2$ and $m3$ do refer, respectively, to the first, second and third month of each corresponding quarter within a calendar year. Over this structure, we recreate the (pseudo) real-time availability of each indicator based on average dissemination calendars by the respective sources over the past 20 years. The dark blocks correspond to no information being available for any country at that particular moment in time, while the light ones reflect that a particular data point is available, for a given country (labels BR-Brazil; MX-Mexico; CO-Colombia). Thus, for example, in month 1 the Exports figure for Brazil is known only

Figure 3: Information set available in (pseudo) real-time



The dark blocks correspond to no information being available for any country. The first row of the table corresponds to the month in which the forecast is computed; the second and third rows to the time for which the information is available. Thus, for example, in month 1 the Exports figure for Brazil is known only up and including until month 1 of the previous quarter, whereas for Mexico and Colombia the Exports are known until the end of the previous quarter, that is, for two more months. Note that the only quarterly variable in the Table is GDP.

up and including until month 1 of the previous quarter, whereas for Mexico and Colombia the Exports are known until the end of the previous quarter, that is, for two more months. The text-based variables are mostly available in real-time, as are financial and political risk variables.

5.3 On the use of different models

Throughout the paper we will show results for all the models outlined before, in some cases pooling them, under the assumption that an agnostic view of using several models might outweigh a strict selection of one particular model. This flows from the key objective of the paper: we are interested in evaluating the relative performance of adding certain groups of variables with respect to a benchmark group (typically the "Traditional" plus the financial ones), conditional on a given set of models.

That said, model forecasts may differ widely in terms of relative accuracy depending on the specific country, variable, and the month-horizon (m, h) pair. For the sake of transparency, in section A.2 in the appendix we analyse the differences in forecasting performance among different models on the basis of individual regressor forecasts. That section shows our reasons for not considering the unconstrained UMIDAS and UMIDAS-ADL models.

6 Results

Combining forecasts to optimise the predicting power of multiple regressors is a standard technique in both forecasting literature and in practice. In this section we use forecast combination to estimate the added value brought by institutional instability variables. Specifically, we look at the forecast gains made when institutional instability variables are added to the set of traditional and standard regressors.

While multiple categorisation schemes are possible, here we focus on splitting institutional instability variables into two broad categories, "Text Variables" and "Conflict Models". We also consider the entire additional set, that is, the union of the two.[4] The constituent variables are shown in Table 2. Note that while the number of variables in the two categories is different, that does not hinder our analysis since its aim is not to make a comparison between the value added by an average text variable and a conflict model, but rather to contrast these particular set of variables.

We first prune the two new variable combinations, TradStanCM and TradStanTV, by getting rid of variables that worsen their performance.[5] We select variables by checking the performance of the combination forecast when excluding the variable. The condition we impose is that the removal of a variable should lead to an improvement of the model for at

[4]Two alternative categorisation schemes with a large number of groups had also been considered, but the multiplicity of groupings, and the intra-country variance, did not improve the interpretability of results. They do, however, support the main conclusions. The forecasts and the related figures are available on request.

[5]Here is it is important to not simply discard the variables that were found to give the worst individual-regressor forecast results for each country, since a badly performing variable might, when taken as part of a combination, nevertheless steer the forecast in the correct direction (precisely what we see in the case of some Armed Conflict models for Brazil). Instead, we evaluate the improvement obtained by selectively removing individual variables from the forecast combinations, see Figure A21 in the appendix.

Table 2: Individual regressor categorisation scheme

Class	Category	Variable
Benchmark	Traditional (Trad)	Ind.Prod.
		Ret.Sales
		Credit
		Exports
		Unempl.Rate
		Prod.Constr.
		Fix.Inv.
		Cons.Conf.
		Bus.Conf.Ind.
	Standard (Stan)	EMBI
		Sov.Rat.
Institutional Instability Variables	Text Variables (TV)	GPR
		EPU
		Soc.Unr.
		Top.pol.
		Top.econ.
		Top.conf.
	Conflict Models (CM)	AnyViol.3.text
		AnyViol.12.text
		AnyViol.3.best
		AnyViol.12.best
		Arm.Conf.3.text
		Arm.Conf.12.text
		Arm.Conf.3.best
		Arm.Conf.12.best

Notes: Forecast combinations are labelled according to the shorthand of the constituent categories. For example, the combination "TradStanTV" includes Trad, Stan, and TV variables. The "Complete" set contains all the variables (Trad, Stan, TV, and CM).

Table 3: Variables removed from combination forecasts for each country.

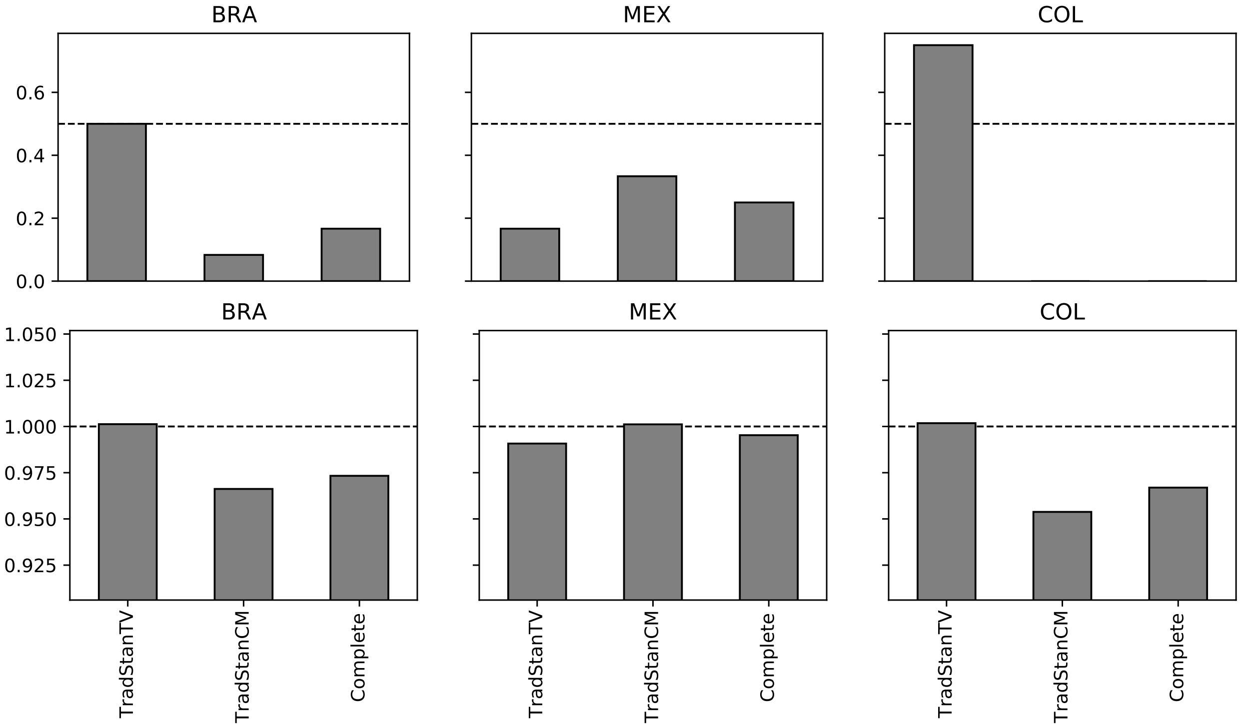
Country	Variables excluded
Brazil	Arm.Conf.3.best; Arm.Conf.12.best; Top.conf.
Mexico	AnyViol.3.text; Arm.Conf.12.text; Top.conf.
Colombia	AnyViol.3.text; AnyViol.12.text; EPU

least 90% of the (month, horizon) pairs. The variables excluded for each country are shown in Table 3. Their removal improves the relative RMSFE of TradStanCM, TradStanTV, and Complete with respect to the baseline of TradStan by up to a full percentage point.

We now evaluate how much forecasts with a combination of traditional and standard variables improve under the addition of these (optimised) variable categories. Figure 4 summarises the results. The top row shows the likelihood that adding variables to TradStan makes the forecast worse (the extensive margin) and the bottom row shows the RMSFE compared to TradStan. We see a clear overall tendency for an improvement of the forecast with all three variable combinations. The notable exception is the TradStanTV combination which leads to a relatively clear worsening for Colombia. However, the forecast of every country for at least 80% of the (month, horizon) pairs is improved by adding some combination of institutional instability variables. Brazil and Colombia benefit most from the Conflict Models, while Mexico benefits from the direct text-based variables. The improvements are relatively modest but it should be kept in mind that Figure 4 shows averages across forecast month and horizon, i.e. across very different informational sets and precision in the TradStan model.

Figure 5 shows how the relative forecast error of the optimal combination changes with both the initial month m (top) and horizon h (bottom) for all the four models and three combination methods. Two general trends immediately stand out: first, the overall improvement is largest for Colombia, and second, for Brazil and Mexico the added value of optimal institutional instability variables decreases with months and increases with horizon.

Figure 4: Effect of adding groups of variables to the TradStan benchmark combination on making the quarterly forecast less accurate (top row) and RMSFE (bottom row)

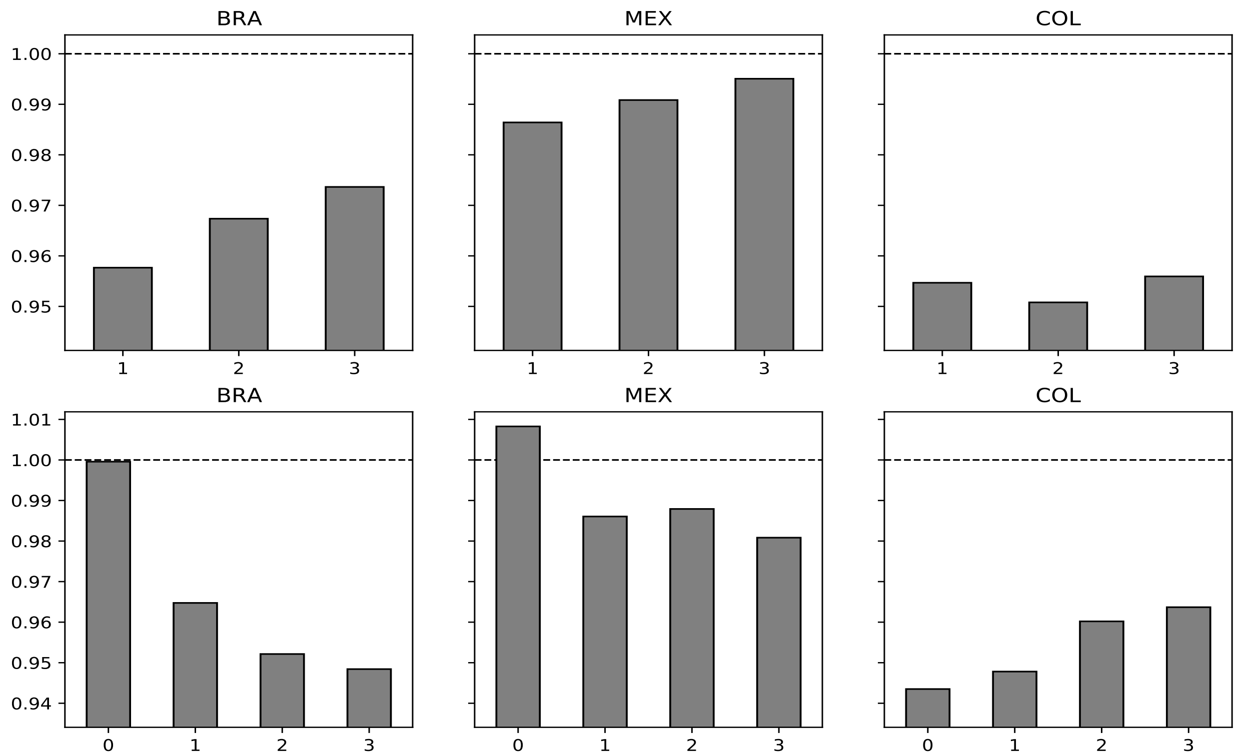


Top row: Probability that adding extra variables makes the quarterly forecast less accurate. The figures are (1 - fraction of (month, horizon) pairs where the average combination forecast gave a smaller RMSFE than the benchmark combination). Bottom row: RMSFE relative to the benchmark combination averaged first over the four models and three combination methods, and then over the (month, horizon) pairs. In both top and bottom row lower bars correspond to better combination performance compared to benchmark. The dashed horizontal line marks the probability of 0.5.

The explanation is intuitive: at higher m we are further into the quarter, more traditional variables become available, and so the benchmark becomes harder to beat. On the other hand, increasing the horizon increases the contribution of the instability variables. This could be because some of the instability measures predict future instability or because increasing risks affect investments today which affects outcomes in the future.

The difference between Colombia on the one hand and Mexico and Brazil on the other could be explained by the fact that Colombia has suffered a long-lasting, intense civil conflict for decades at the time when our sample begins. This makes keeping track of political instability particularly relevant for the country. At the same time, we do not expect large swings in expectations and economic behavior for Colombia as expectations in the peace process

Figure 5: Improvement of the quarterly combination forecast on the benchmark combination by month (top row) and horizon (bottom row)



RMSFE of the optimal combination relative to the TradStan benchmark combination averaged first over the four models and three combination methods, and then over either months (top row) or horizons (bottom row). For Brazil and Colombia, the optimal combination is TradstanCM and for Mexico it is TradStanTV, excluding country-specific worst performing variables.

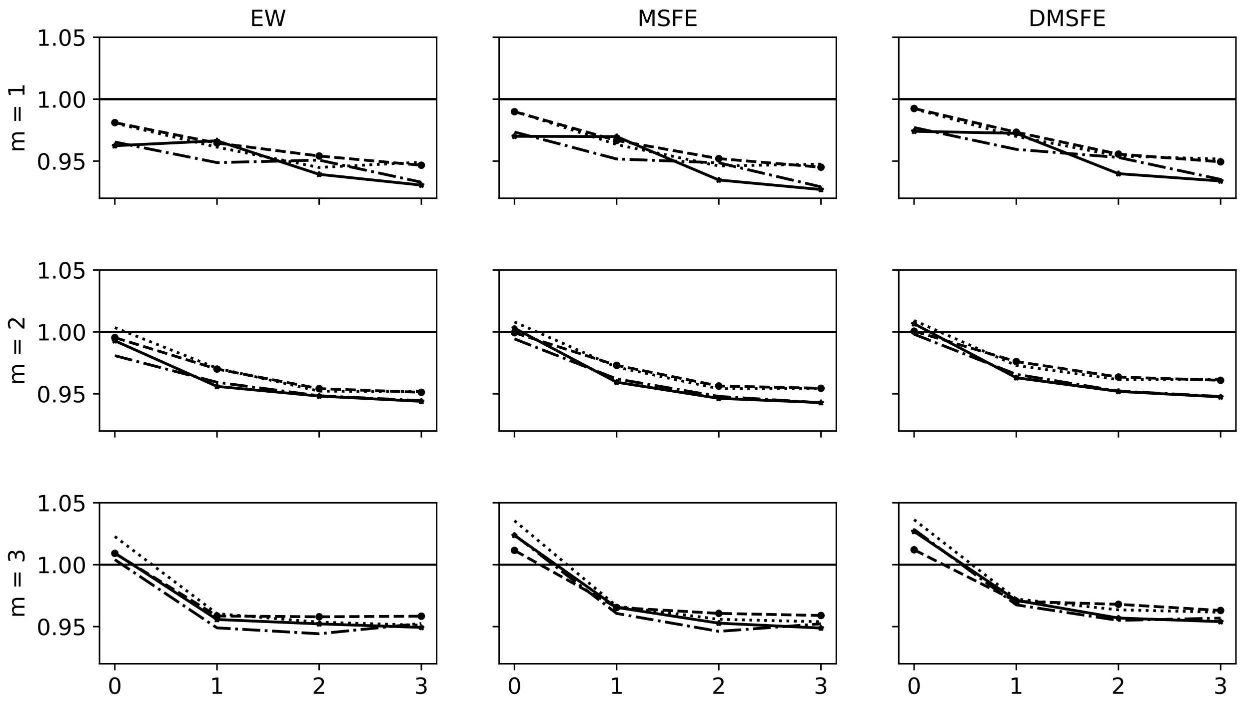
only change slowly. In Mexico and Brazil the opposite is true. Here, sudden escalations might be leading to particularly significant gains over longer forecast horizons.

This interpretation finds some support in Figures 6, 7 and 8. Here we show the data in a much more disaggregated way. Months of the quarter are displayed in rows, forecast horizons are displayed on the x-axis of each graph and the columns show different ways of measuring error. Finally, the different lines represent different forecast models. These figures show that independent of the country, these results are robust across the different models and combination methods.

From these figures it is clear that there is basically one case in which the additional variables do not improve the forecast: nowcasting in the third, and to some extent second, month of the quarter - but only when forecasting at low horizons. These are the months that contain the most of the Traditional variables, including the crucial previous quarter's GDP. Adding institutional instability variables to the traditional set for both Mexico and Brazil during those months can actually harm the nowcast if combined with low forecast horizons. Our interpretation is that political variables are most relevant for longer horizons or if the traditional variables are not yet available. However, for all other combinations institutional instability is a logical addition to forecast models.

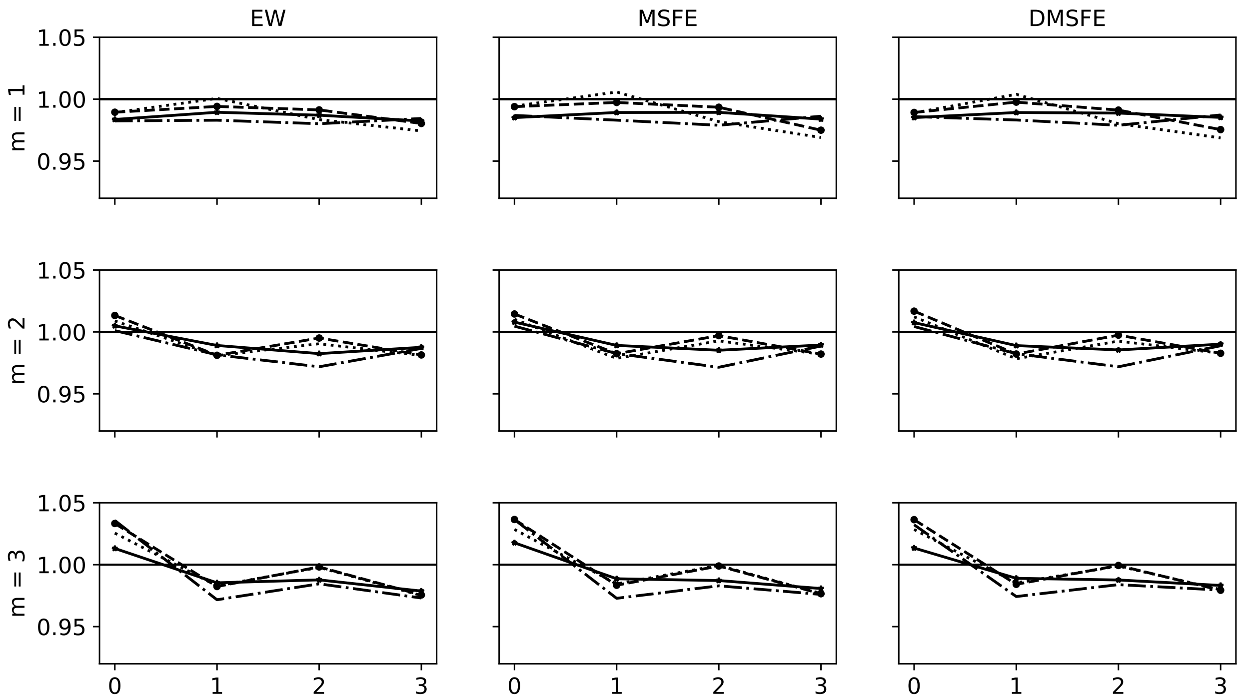
It is striking how similar the relative gains are for Mexico and Brazil when compared to Colombia. Our interpretation of this is, again, that the former suffered dramatic escalations of instability in our sample range, whereas Colombia entered the sample with extremely high levels of violence and benefited from stabilization attempts. The gains for Colombia are basically flat for longer horizons as expectations in the country did not undergo such dramatic shifts. It is more important to keep track of the day-to-day shifts than resolving large uncertainties at an escalation.

Figure 6: Improvement of quarterly combination forecast on the benchmark combination for different models and combination methods: Brazil



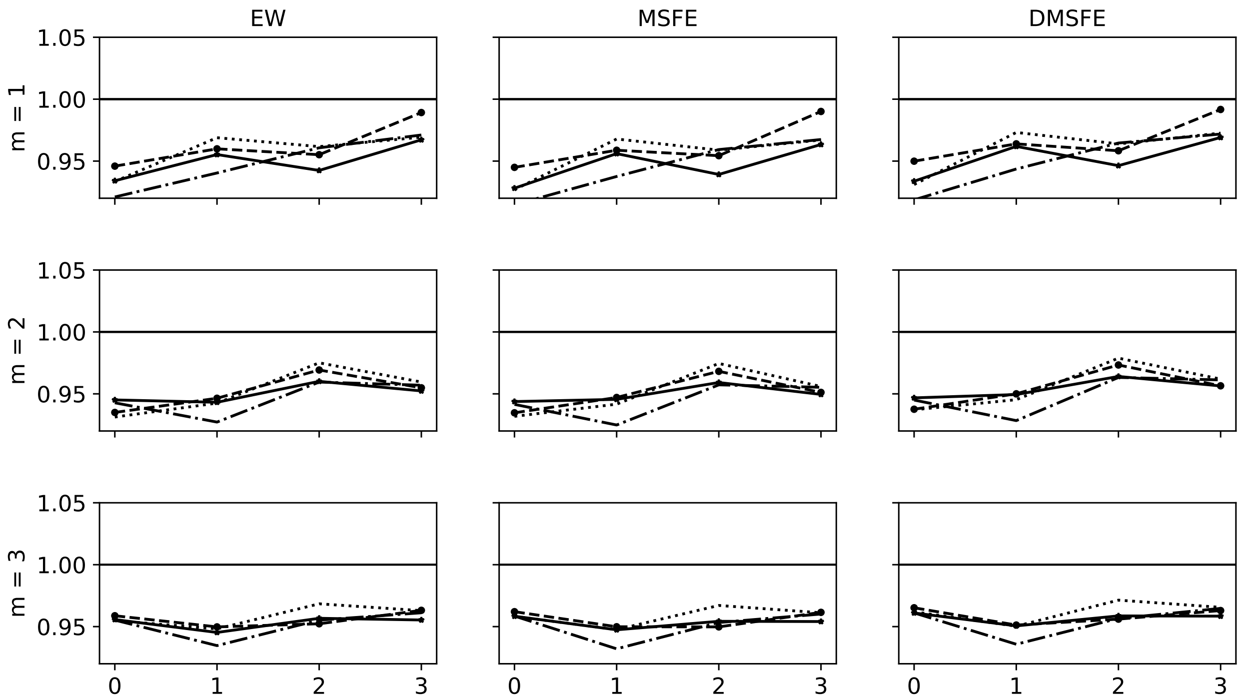
Relative RMSFE of the optimal combination forecast of Brazil (see caption to figure 5) compared to the combination of only the traditional and financial indicators (TradStan). Columns correspond to the different combination methods. Starting months are separated by rows, the abscissa shows the forecast horizons. Lines are individual regressor models: solid with marker: TA, dashed with marker: TA-ADL, dotdashed: NEALMON, dotted: NEALMON-ADL.

Figure 7: Improvement of quarterly combination forecast on the benchmark combination for different models and combination methods: Mexico



Relative RMSFE of the optimal combination forecast of Brazil (see caption to figure 5) compared to the combination of only the traditional and financial indicators (TradStan). Columns correspond to the different combination methods. Starting months are separated by rows, the abscissa shows the forecast horizons. Lines are individual regressor models: solid with marker: TA, dashed with marker: TA-ADL, dotdashed: NEALMON, dotted: NEALMON-ADL.

Figure 8: Improvement of quarterly combination forecast on the benchmark combination for different models and combination methods: Colombia



Relative RMSFE of the optimal combination forecast of Brazil (see caption to figure 5) compared to the combination of only the traditional and financial indicators (TradStan). Columns correspond to the different combination methods. Starting months are separated by rows, the abscissa shows the forecast horizons. Lines are individual regressor models: solid with marker: TA, dashed with marker: TA-ADL, dotdashed: NEALMON, dotted: NEALMON-ADL.

7 Concluding remarks

In this article we have shown that adding text-based measures of violent conflict risk and uncertainty can improve economic forecasts. Gains are largest in earlier months of a quarter and, in general, when time horizons are longer.

Our findings also suggest some country heterogeneity, i.e. gains depend on the political situation that a country is in. For Colombia we find the large and robust gains across very different horizons and months. Gains for Mexico and Brazil most clearly grow with longer horizons. For these two countries we find that a combination of short forecast horizons and a late forecast month, when more traditional variables become available, can actually harm the forecast.

This begs the question of whether stabilization of Colombia vs. the destabilization of Mexico and Brazil is the distinguishing feature that drives the different experiences in our forecast exercise. We would then expect similar gains in forecast accuracy to Mexico and Brazil for other economies which are suffering from sudden instability and gains like in Colombia for other economies with strong, ongoing but fluctuating instability. A natural avenue for enriching this analysis is, therefore, an extension of the dataset to more economies. This would allow us to see whether these speculations hold up in a larger sample with different political histories. Learning about this is important at a time when not only the the economic burden of the Covid-19 pandemic will find its political outlets and might destabilize more and more countries, but also in the wake of the unprecedented levels of geopolitical uncertainty accompanying the war in Ukraine.

On a final note, in this exercise we considered whether, on average, adding proxies of institutional instability such as conflict indicators improves the forecast of the GDP. But the value of these indicators need not be such that it is necessarily picked up on by regression-like models with its effective averaging over out-of-sample predictions. An open question is whether there are certain periods of time where the fluctuation in, for example, the likelihood of continued armed violence is key for determining variation in GDP. Characterising and detecting such instances would be instrumental to fully utilise the power of instability proxies for macroeconomic forecasting.

References

Abadie, A. & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American Economic Review*, 93(1), 113–132.

Armesto, M. T., Engemann, K. M., & Owyang, M. T. (2010). Forecasting with mixed frequencies. *Federal Reserve Bank of St. Louis Review*, 92(6), 521–36.

Baker, S., Bloom, N., & Davis, S. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4), 1593–1636.

Barrett, P., Appendino, M., Nguyen, K., & de Leon Miranda, J. (2020). *Measuring Social Unrest Using Media Reports*. Working paper, IMF.

Barrett, P., Bondar, M., Chen, S., Chivakul, M., & Igan, D. (2021). *Pricing Protest: The Response of Financial Markets to Social Unrest*. Working paper, IMF.

Barro, R. J. (2009). Rare disasters, asset prices, and welfare costs. *American Economic Review*, 99(1), 243–64.

Besley, T. & Ghatak, M. (2010). Property rights and economic development. *Handbook of Development Economics*, 5, 4525–4595.

Besley, T. & Mueller, H. (2012). Estimating the peace dividend: The impact of violence on house prices in northern ireland. *American Economic Review*, 102(2), 810–33.

Besley, T. & Mueller, H. (2018a). Institutions, volatility, and investment. *Journal of the European Economic Association*, 16(3), 604–649.

Besley, T. & Mueller, H. (2018b). Predation, protection, and productivity: A firm-level perspective. *American Economic Journal: Macroeconomics*, 10(2), 184–221.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, 77(3), 623–685.

Caldara, D. & Iacoviello, M. (2022). *Measuring Geopolitical Risk*. Technical report.

Carriere-Swallow, Y. & Cespedes, L. F. (2013). The impact of uncertainty shocks in emerging economie. *Journal of International Economics*, 90, 316–325.

Cepni, O., Güney, I. E., & Swanson, N. R. (2020). Forecasting and nowcasting emerging market GDP growth rates: The role of latent global economic policy uncertainty and macroeconomic data surprise factors. *Journal of Forecasting*, 39(18-36).

Claeskens, G., Magnus, J. R., Vasnev, A. L., & Wang, W. (2016). The forecast combination puzzle: A simple theoretical explanation. *International Journal of Forecasting*, 32(3), 754–762.

Clements, M. P. & Galvão, A. B. (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. *Journal of Business & Economic Statistics*, 26(4), 546–554.

Collier, P. (1999). On the economic consequences of civil war. *Oxford Economic Papers*, 51(1), 168–183.

Corona, F., González-Farías, G., & López-Pérez, J. (2020). A dynamic factor model for the mexican economy: are common trends useful when predicting economic activity? *Latin American Economic Review*, 26(7).

Gauvin, L. & C. McLoughlin, and D. Reinhardt, D. (2014). *Policy uncertainty spillovers to emerging markets: Evidence from capital flows*. Working paper, Bank of England.

Ghirelli, C., Pérez, J., & Urtasun, A. (2019). A new economic policy uncertainty index for Spain. *Economic Letters*, 182, 64–67.

Ghirelli, C., Pérez, J., & Urtasun, A. (2021). Economic policy uncertainty in Latin America: measurement using spanish newspapers and economic spillovers. *Latin American Journal of Central Banking*, 2(2), 100029.

Ghysels, E., Kvedaras, V., & Zemlys, V. (2016). Mixed frequency data sampling regression models: The R package midasr. *Journal of Statistical Software*, 72(4).

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2004). *The MIDAS Touch: Mixed Data Sampling Regression Models*. Working paper, CIRANO.

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55, 665–676.

Hadzi-Vaskov, M., Pienknagura, S., & Ricci, L. A. (2021). *The Macroeconomic Impact of Social Unrest*. Working paper, IMF.

Hassan, T. A., Hollander, S., van Lent, L., & Tahoun, A. (2019). Firm-level political risk: Measurement and effects get access arrow. *The Quarterly Journal of Economics*, 134(4), 2135–2202.

Johnson, S., McMillan, J., & Woodruff, C. (2002). Property rights and finance. *American Economic Review*, 92(5), 1335–1356.

Leiva-Leon, D., Pérez-Quirós, G., & Rots, E. (2020). *Real-time Weakness of the Global Economy: A First Assessment of the Coronavirus Crisis*. Working paper, Banco de Espana.

López, F., Yaselga, E., & Espinosa, F. (2021). A nowcasting model using dynamic factors to estimate real quarterly GDP in Ecuador. *Revista Cuestiones Económicas*, 31(1).

Mueller, H. & Rauh, C. (2018). Reading between the lines: Prediction of political violence using newspaper text. *American Political Science Review*, 112(2), 358–375.

Mueller, H. & Rauh, C. (2022a). The hard problem of prediction for conflict prevention. *Journal of the European Economic Association, forthcoming*.

Mueller, H. & Rauh, C. (2022b). *Using Past Violence and Current News to Predict Changes in Violence*. Technical report, Faculty of Economics, University of Cambridge.

Mueller, H. & Tobias, J. (2016). *The Cost of Violence: Estimating the Economic Impact of Conflict*. IGC Growth Brief 2211, International Growth Centre.

Saadi Sedik, T. & Xu, R. (2020). *A Vicious Cycle: How Pandemics Lead to Economic Despair and Social Unrest*. Working Paper 2020/216, IMF.

Timmermann, A. (2006). Chapter 4: Forecast Combinations. In G. Elliott, C. Granger, & A. Timmermann (Eds.), *Handbook of Economic Forecasting*, volume 1 (pp. 135–196). North-Holland.

Willard, K. L., Guinnane, T. W., & Rosen., H. S. (1996). Turning points in the civil war: Views from the Greenback Market. *American Economic Review*, 86(4), 1001–18.

Zussman, A. & Zussman, N. (2006). Assassinations: Evaluating the effectiveness of an israeli counterterrorism policy using stock market data. *Journal of Economic Perspectives*, 20(2), 193–206.

A Annex: data and additional results

A.1 Indicators

Real GDP as well as the raw variables for all the three countries are shown in figures A9 - A11.

A.2 Model selection

In this section we evaluate the four forecasting models used in the main article text, as well as the UMIDAS models, for their accuracy.

Figure A12 shows that forecast accuracy distributions associated with different models are in fact statistically distinct. Specifically, in the case of Brazil and Colombia, forecasts for UMIDAS and UMIDAS-ADL are statistically different from the other models but not from each other. More such qualitatively different model groupings can be distinguished in the case of Mexico. Given that the UMIDAS models represent unconstrained regression, an immediate implication is that the MIDAS approach, with its constraints on the coefficients, does lead to fundamentally different forecasts.

In addition, for each regressor we produce a model ranking based on average forecast accuracy relative to a benchmark. Figure A13 shows there is no clear relation between the type of variable and which model tends to be higher ranked. Nevertheless, the ranking distribution of the models themselves are highly distinct (see Figure A14). For all the three countries the shapes of the UMIDAS model ranking density are different not only from each other, but from the other four models as well. Furthermore, we see that TA and NEALMON are more similar to each other than to their auto-regressive counterparts. Finally, the fact that most curves are unimodal, with most values centered around a mean, imply that variables cannot be naturally clustered based on how well they perform in terms of model forecast accuracy (whether a multidimensional grouping exists is an open question). Figure A15 shows the mean ranking for the three countries. It is clear that while no single best model exists, the UMIDAS models tend to perform worse. The best forecasts for Brazil and Mexico are done with TA and NEALMON, while for Colombia these models are outperformed by their autoregressive counterparts. This analysis justifies dropping the MIDAS approach, as we see that constraining the coefficients does lead to better predictions.

A.3 Forecasts based on models with individual regressors

In this section we give a brief summary of quarterly GDP growth forecasts using using the 25 individual regressors. Our results suggest that it is hard to beat the autoregressive model, a fact that is already widely known in forecasting literature. It is nevertheless notable that the institutional instability variables consistently appear among best performing regressors across the three countries. This gives an indication of the potential forecast gains that could be made if these regressors are utilized properly. Such proper usage would in fact be vital since these variables can also be found on the other extreme of the evaluation spectrum where forecasts perform much worse than the benchmark models. This relative performance is country-specific. A case in point is *Social Unrest*, the fourth best predictor for Mexico, but second worst for Brazil and Colombia. Creating accurate forecasts would therefore depend on understanding the exact specification under which a variables is expected to make a positive contribution.

Accuracy We first test how well the regressors perform across the parameters by doing a ranking exercise. The performance variability is captured in figures A16 to A18. First, it is clear that, independent of the country, good performance is more likely to persist across the initial months of the current quarter than, for a given initial month, across the forecast horizons. Second, we note the dominance of the conflict model variables. Indeed, for any given month, these are the only regressors that consistently appear in the top five across the forecast horizons. This holds for all the countries in this analysis. They also form the majority of those regressors that stay in the top five across the initial months. Nevertheless, there is still substantial variability in terms of which specific variables perform well across all countries for any given month and horizon. Thus, only *AnyViol.3.best* ($h = 3$, $m = 1, 2$) and *Credit* ($h = 3$, $m = 3$) appear in the top five across all the countries. We therefore underline again the fact that while non-standard variables have the potential to add value over benchmark forecasts, some form of primary variable selection is required in order to make the most of such individual variable forecasting exercises, as there is no single regressor that works best for all the countries. Note, however, that this apparent inconsistency is not confined to the novel variables: the performance of standard variables such as Fixed

Investment or Retail Sales also varies greatly depending on the country. Finally, figure A19 formalises whether the difference between the forecast and the benchmark is statistically significant by showing the Diebold-Mariano test results (Diebold and Mariano, 1995).[6]

Directional accuracy A distinct but equally valuable aspect of a forecast is the extent to which it is able to predict the *direction* of GDP growth. Among the 25 individual regressors forecasting quarterly growth, on average 63% of the signs are predicted correctly for Brazil, 81% for Mexico, and 91% for Colombia (results tables available on request). The variation is, however, more related to the statistics of the dependent variable than it is to the quality of the forecasts: in the out-sample there are 64% positive values in Brazil, 87% in Mexico, and 92% in Colombia. Finally, we perform the Pesaran-Timmerman test in order to identify forecasts whose predictions of the direction of quarterly growth are statistically distinct from the baseline of throwing a biased coin, where the bias is defined by the true ratio of positive and negative values.[7] We further limit ourselves only to those forecasts that improve on the AR(p) benchmark. In the case of Mexico and Colombia the results are not encouraging, with very few regressors giving such statistically distinct predictions. This is in line with the results above that gains in directional accuracy (from whichever regressors) can be made mostly only for Brazil. To compare the contribution of different regressors we count, for Brazil, the fraction of such statistically improved forecasts. Figure A20 shows that institutional instability variables are twice as likely to produce better forecasts than traditional and standard variables. The same figure breaks down these numbers by horizon. First, we see that forecasting further afield leads to less accurate forecasts. Even

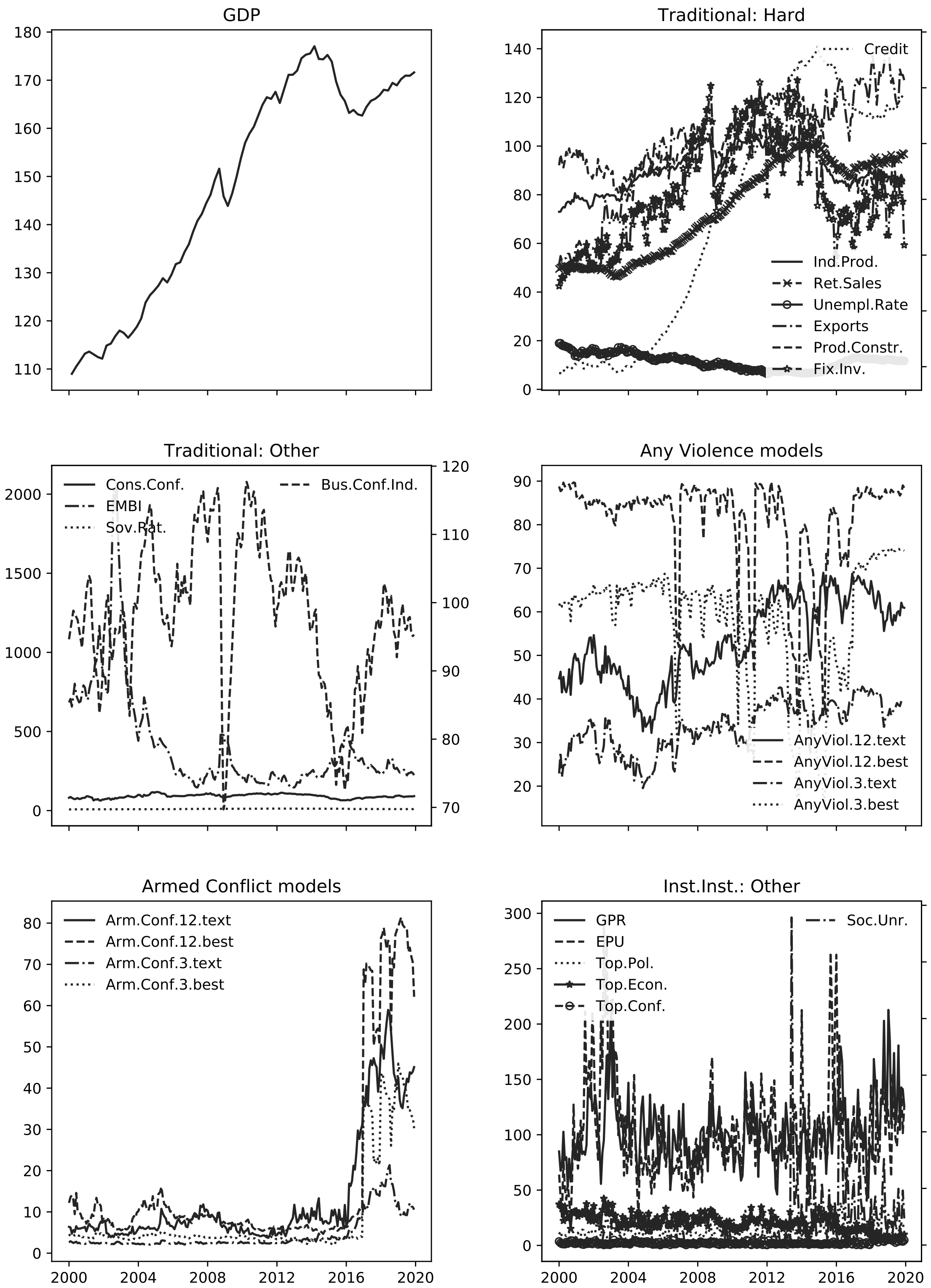
[6] First, define the loss differential $d_i = (e_i^f)^2 - (e_i^b)^2$, with its expected value $\mathbb{E}[d]$, as well as the respective autocorrelation function $a_k = \frac{1}{N}\sum_{i=k+1}^{N}(d_i - \bar{d})(d_{i-k} - \hat{d})$. Then the Diebold-Mariano statistic DM can be defined as $DM = \frac{\bar{d}}{\sqrt{(a_0 + 2\sum_{k=1}^{h-1} a_k)(1/N)}}$, where we consider $h = 1$ steps ahead. We furthermore modulate the D-M statistic with the Harvey adjustment (see Harvey et al., 2007) that balances the results for small datasets, so that the final formulae becomes $DM = DM\sqrt{(n + 1 - 2 * h + h(h-1)/N)/N}$. Through the exercise we consider $h = 1$ steps ahead. The D-M statistic is associated with the null hypothesis that there is no distinction between the forecast and the benchmark. Under this hypothesis, $DM \sim \mathbf{N}(0, 1)$.

[7] The P-T statistic (Pessaran and Timmerman, 1992) is associated with the null hypothesis that the forecast is statistically successful in predicting the signs. For a timeseries t, consider the fraction of positive values p_t, the respective quantity $q_t = \frac{p_t(1-p_t)}{N}$, and the fraction of correctly predicted value signs v. For a forecast f and the true values b let $p = p_f * p_b + (1 - p_f)(1 - p_b)$, $q = \frac{p(1-p)}{N}$, $w = (2p_y - 1)^2 q_b + (2p_b - 1)^2 q_f + 4q_f q_b$. Then the P-T statistic is $PT = \frac{(v-p)}{\sqrt{q-w}}$. Just like the D-M statistic, under the null hypothesis $PT \sim \mathbf{N}(0, 1)$.

so institutional instability variables outperform traditional ones for all horizons. Finally, the graph shows the difference between the variable groups growing with h, suggesting that institutional instability variables add more value when forecasting at longer timescales.

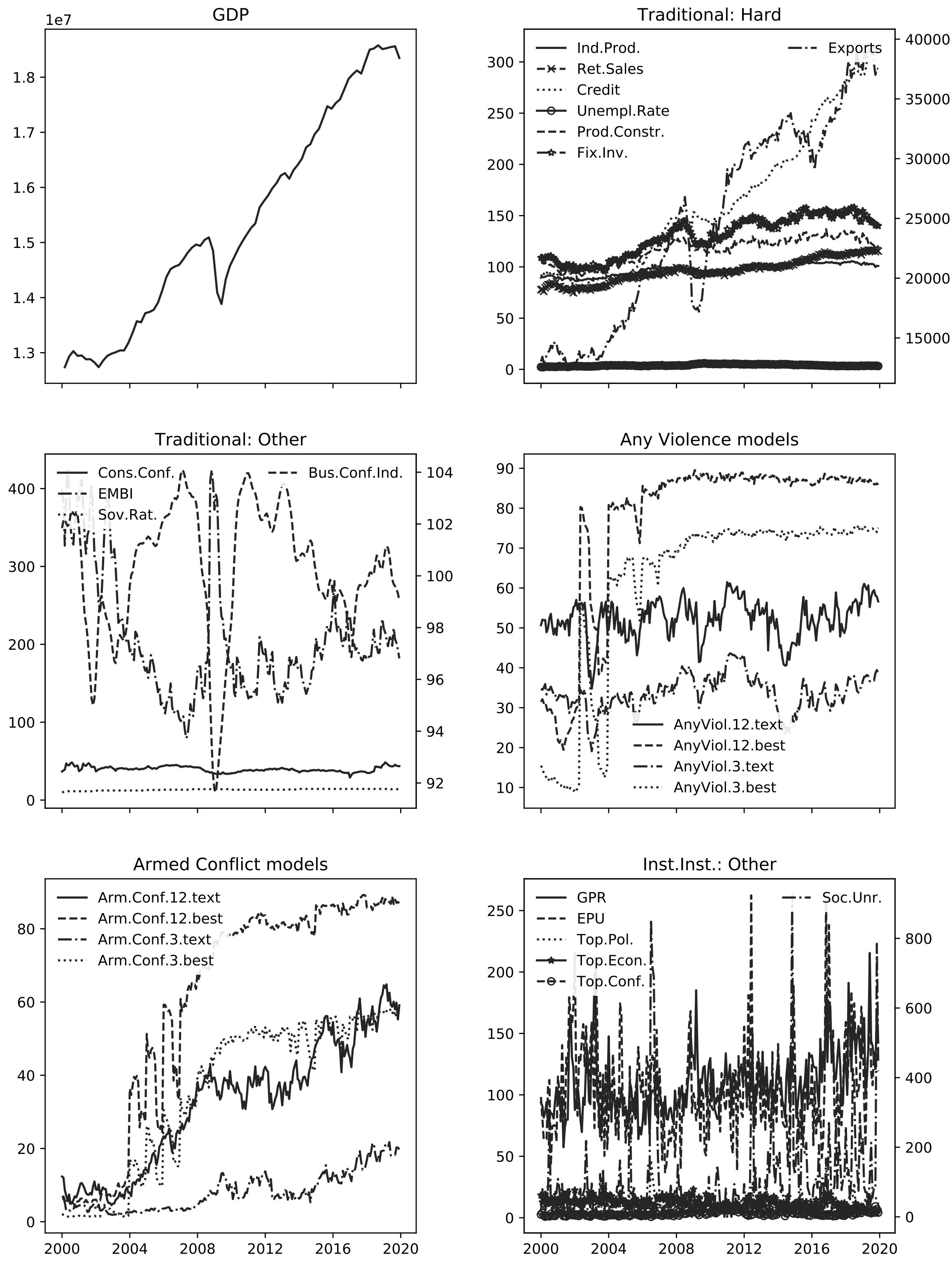
We conclude that in cases where the dependent variable growth show a statistically high chance of switching signs, institutional instability variables display a good potential of correctly predicting the said sign. However, the same caveat applies for directional accuracy as for the forecast accuracy above: while some of these novel regressors can improve the prediction, others can potentially mislead the forecaster with regards to sign change - and that to realise this potential more care needs to be taken in the preliminary analysis stage than that when working with standard variables.

Figure A9: Indicators for Brazil



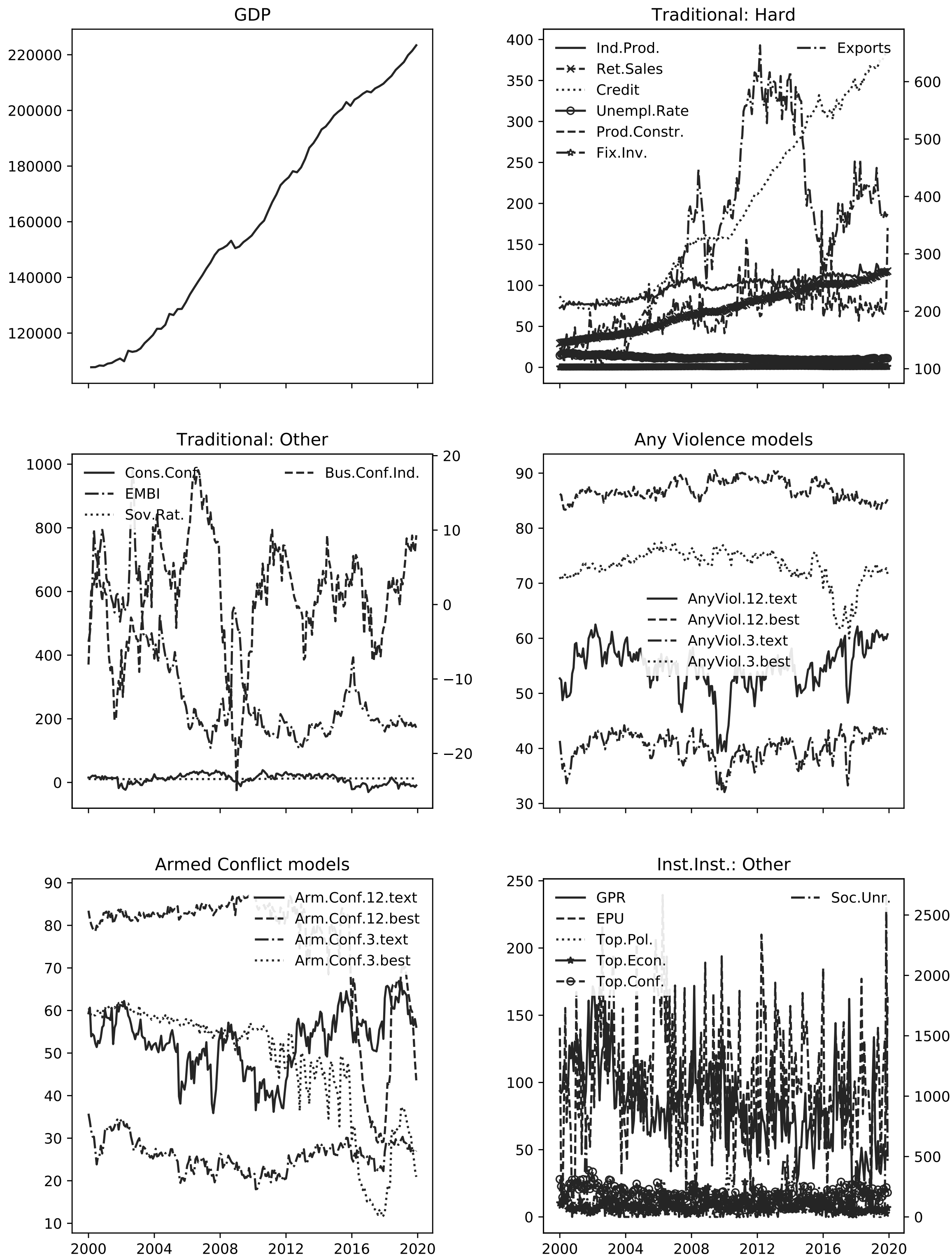
Variable trends are shown before any transformations. The twinned y-axis on the RHS corresponds to variables shown in the legend on the top right, if any.

Figure A10: Indicators for Mexico



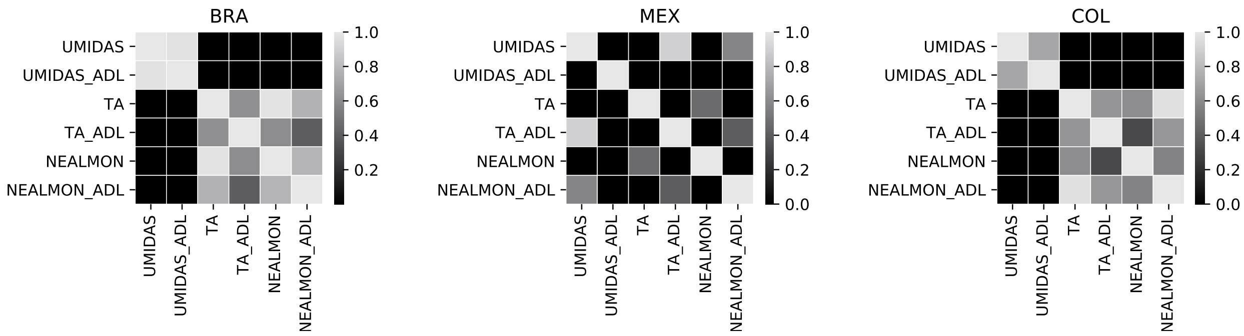
Variable trends are shown before any transformations. The twinned y-axis on the RHS corresponds to variables shown in the legend on the top right, if any.

Figure A11: Indicators for Colombia



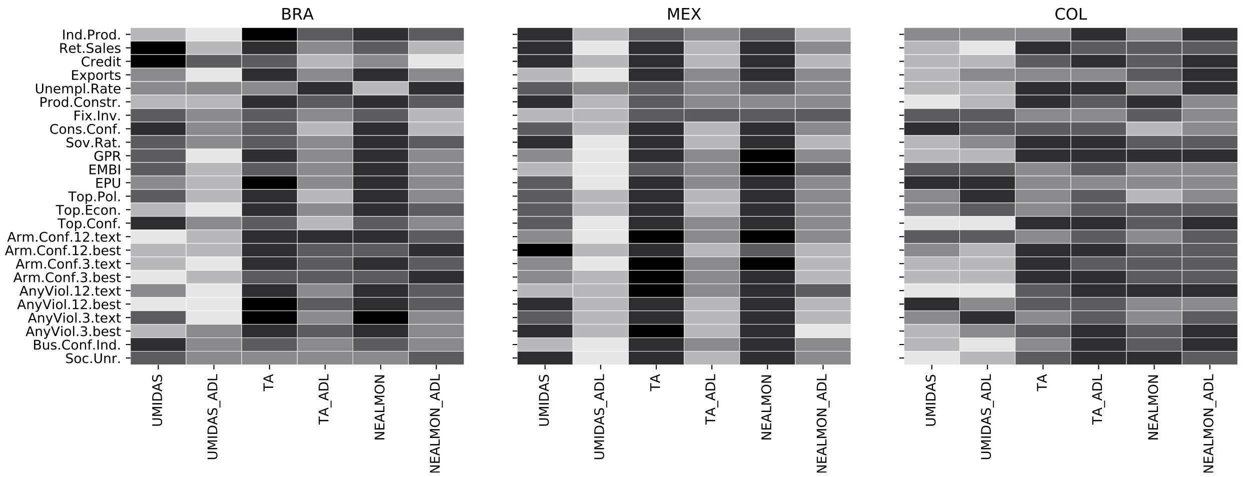
Variable trends are shown before any transformations. The twinned y-axis on the RHS corresponds to variables shown in the legend on the top right, if any.

Figure A12: Differences between the models for quarterly growth forecasts, based on individual regression variables.



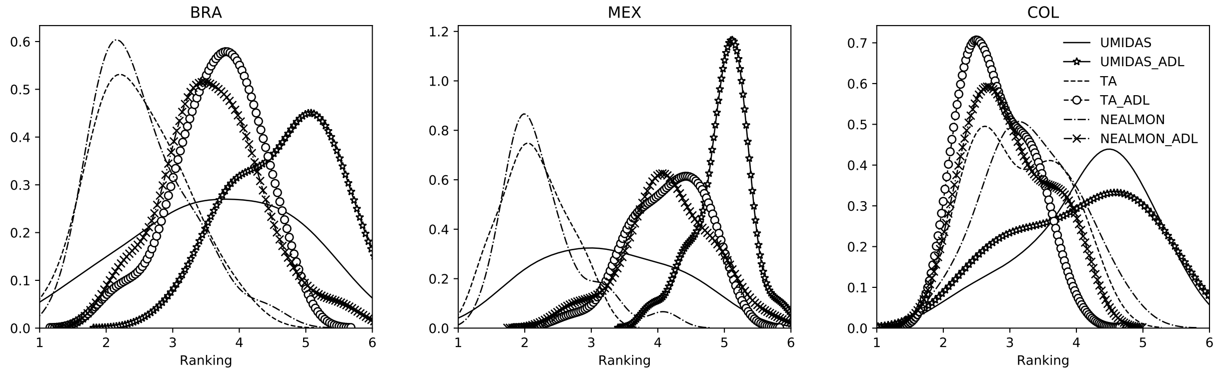
The p value of independent t-tests comparing forecast accuracy of the different models, where the list of values associated with each model is the relative RMSFE for all the regressors, the three starting months and the four forecast horizons.

Figure A13: Model ranking for each variable in the quarterly growth forecasts, based on individual regression variables.



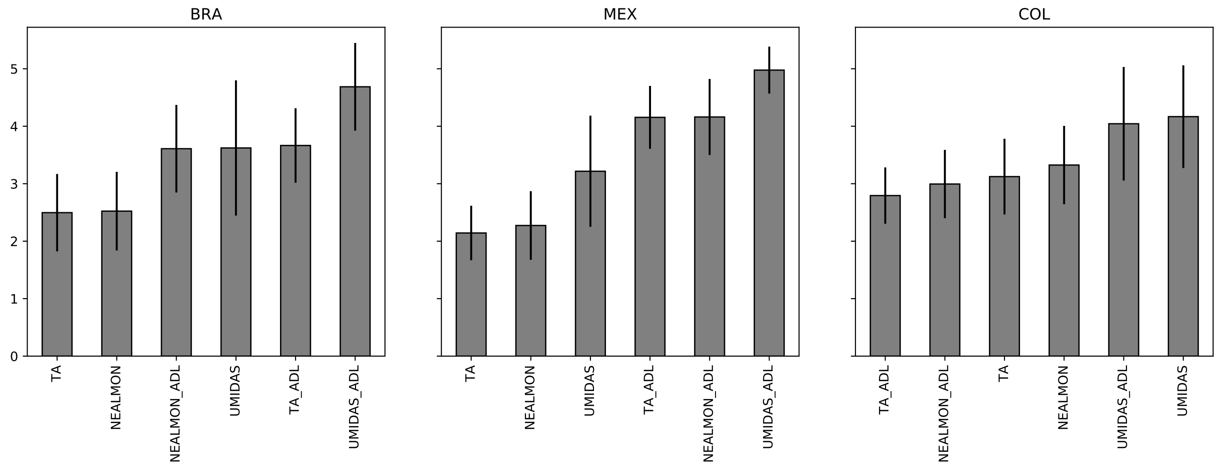
Each cell is the average ranking of the model for each variable, and correspond to the relative performance of a model with respect to the other models given a (m, h) pair, averaged over all the pairs. Relative performance obtained by computing the RMSFE of each model relative to the AR(p) benchmark and then getting the position (1 to 6) of the model in the list compiled in ascending order. Darker colours correspond to higher ranking (values closer to 1), and lighter colours to lower ranking (values closer to the 6, which is the number of models we are comparing).

Figure A14: Average ranking distribution quarterly growth forecasts based on individual regression variables.



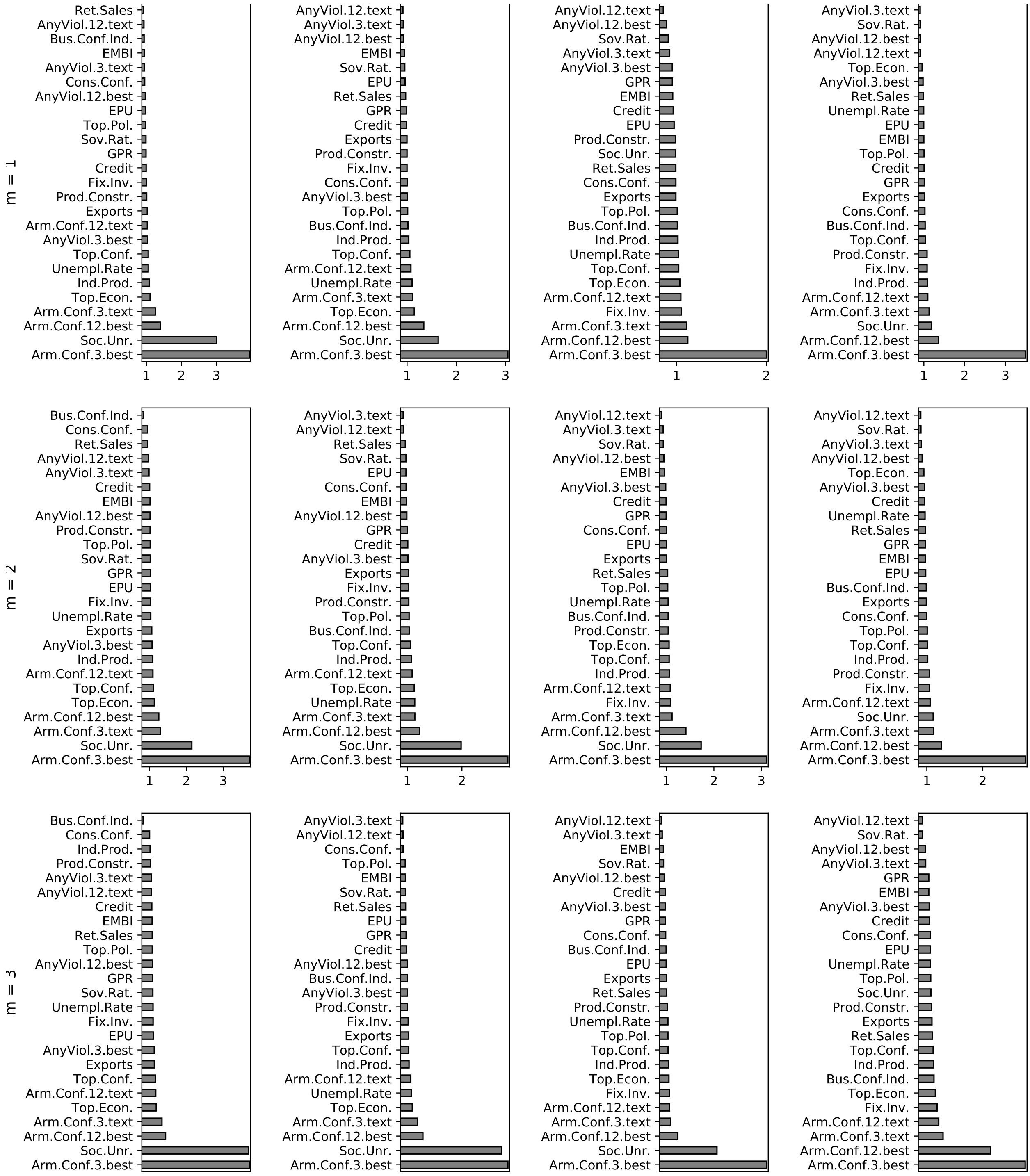
Estimated density functions of average ranking for each variable (the values appearing in figure A13). Lower abscissa values correspond to better model performance.

Figure A15: Illustrative model ranking for quarterly growth forecasts based on individual regression variables.



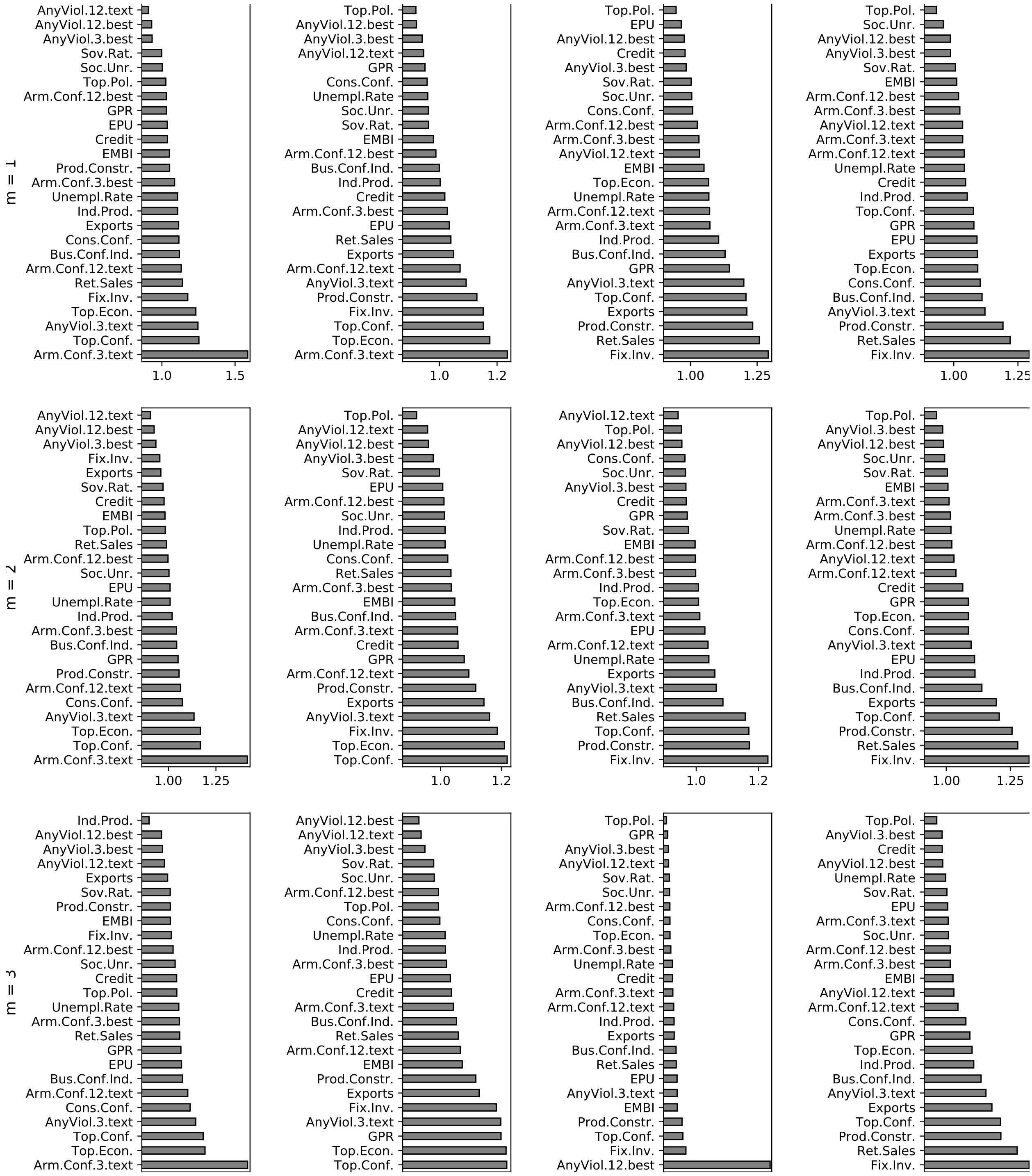
Average model ranking across all variables. The ranking for each variable is the same as in figure A13. Lower values correspond to better model performance.

Figure A16: Performance of individual regressors, quarterly GDP growth. Brazil.



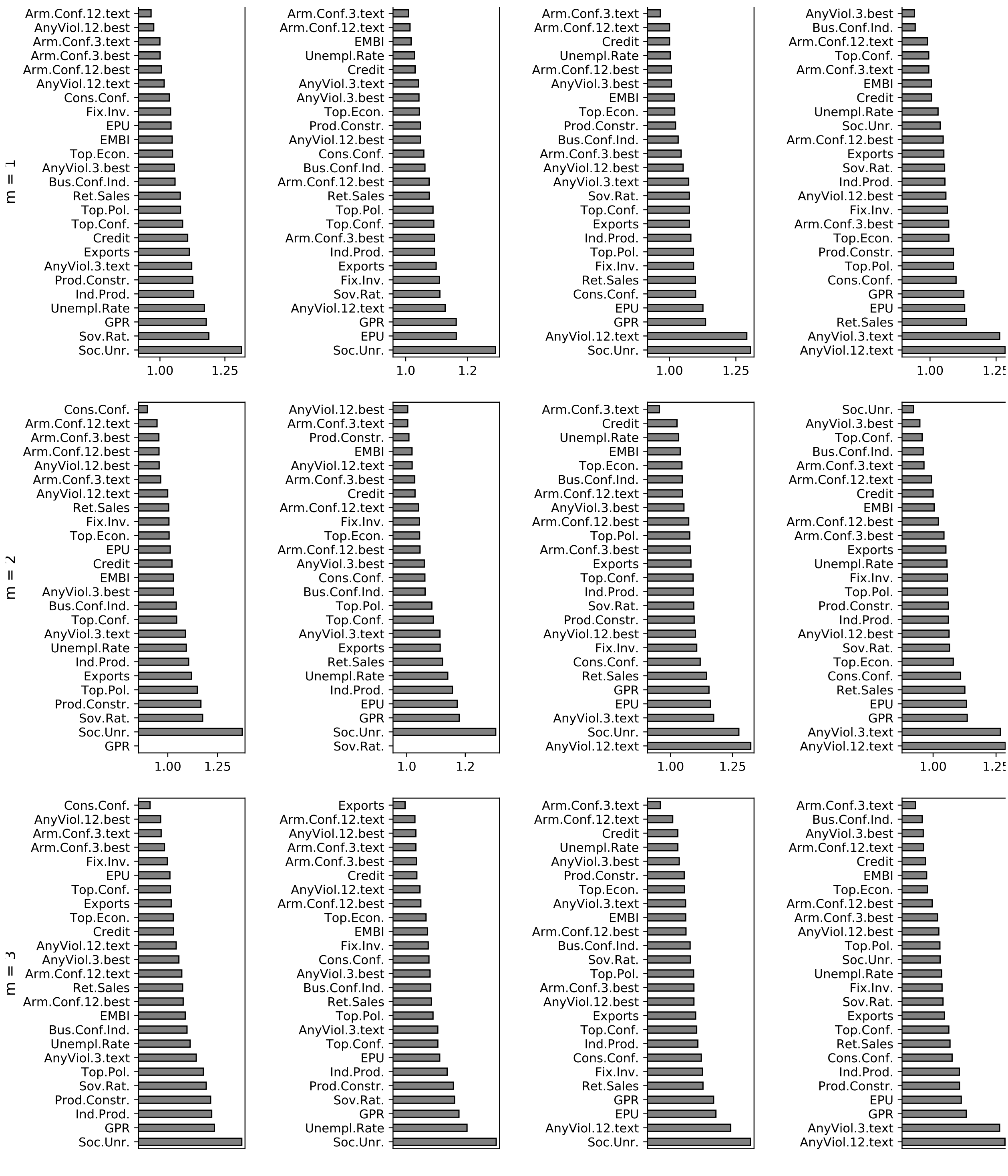
The relative RMSFE compared to the benchmark AR(p) for month m and horizon h for Brazil, averaged over the models.

Figure A17: Performance of individual regressors, quarterly GDP growth. Mexico.



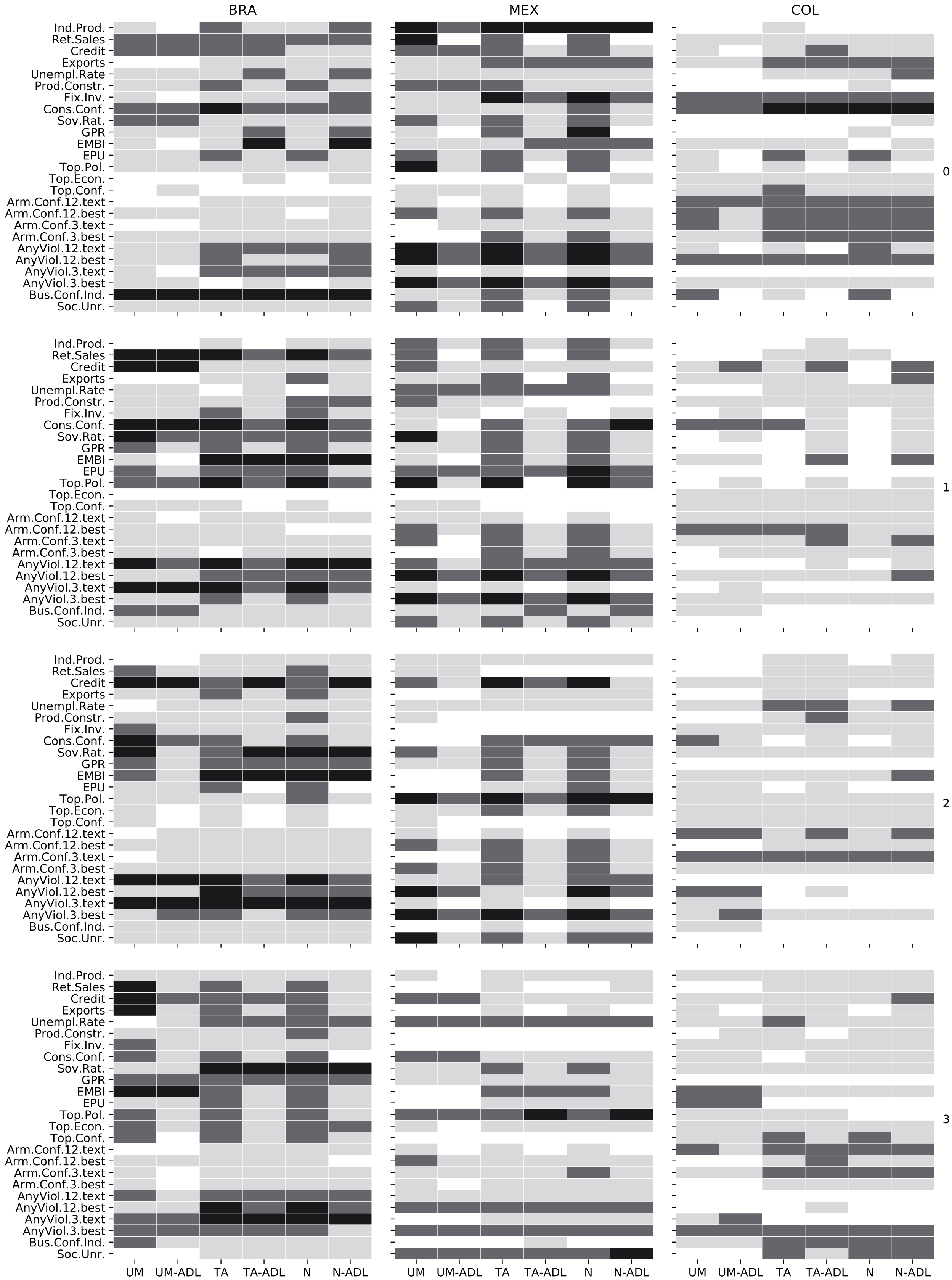
The relative RMSFE compared to the benchmark AR(p) for month m and horizon h for Mexico, averaged over the models.

Figure A18: Performance of individual regressors, quarterly GDP growth. Colombia.



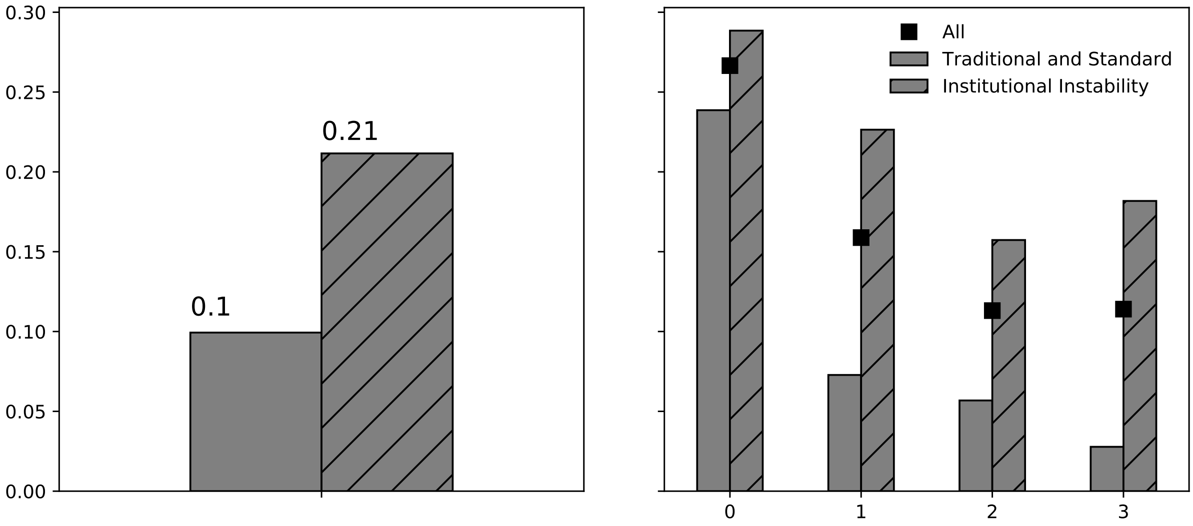
The relative RMSFE compared to the benchmark AR(p) for month m and horizon h for Colombia, averaged over the models.

Figure A19: Evaluation of individual regressor forecasts for quarterly growth



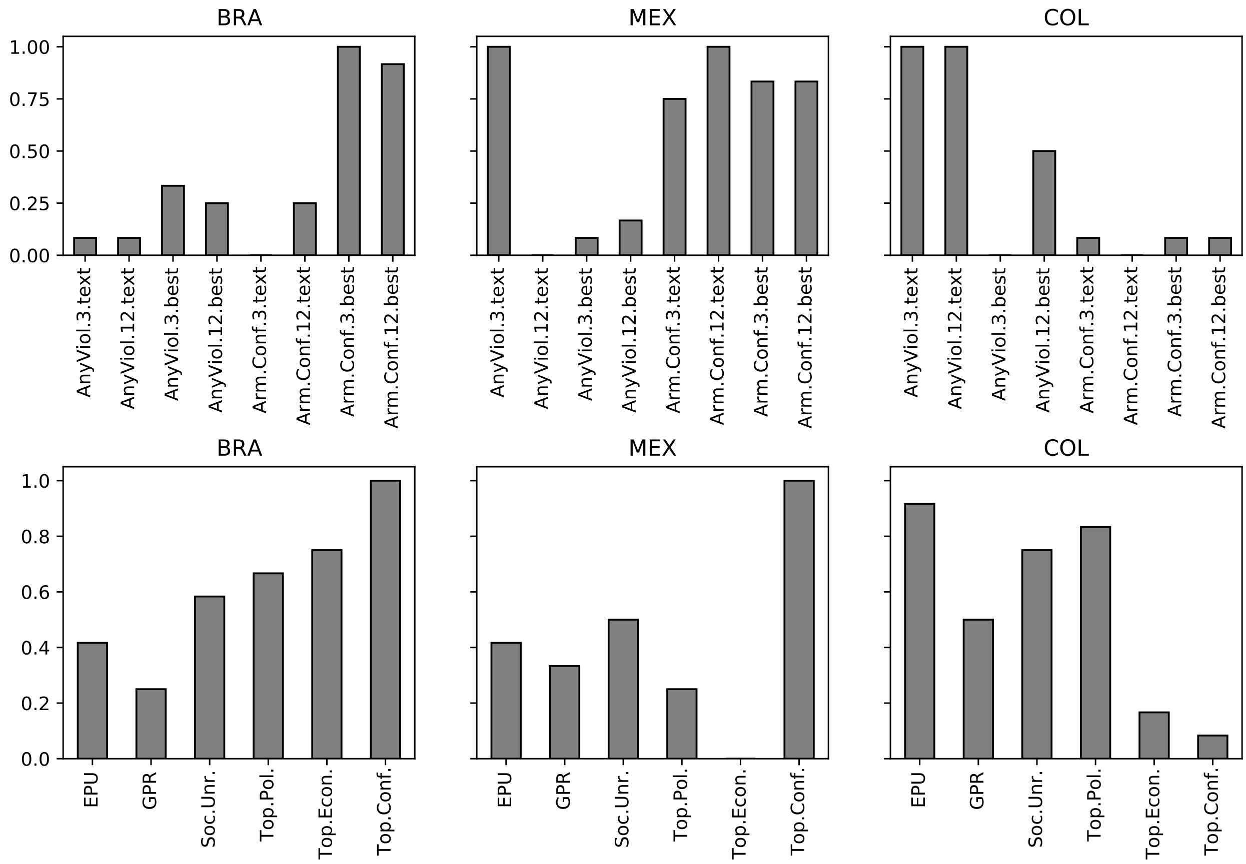
Each cell is associated with three forecasts for the three starting months, for that regressor (row in subplot), model (column in subplot), country (column in chart), and forecast horizon (row in chart). Black corresponds to the relative RMSFE compared to the benchmark AR(p) being less than 1, as well as being different to the benchmark with the D-M p-value of ≤ 0.1. Dark grey is similar but with $p > 0.1$. Light grey corresponds to a forecast that is worse than the benchmark, with relative error > 1, and $p > 0.1$, and white corresponds to a forecast that is statistically significantly worse than the benchmark, with the error > 1 and $p \leq 0.1$. To compute the relative error and statistical significance the three forecasts associated to each cell are averaged with equal weights.

Figure A20: Brazil: likelihood of good forecasts for directional quarterly growth



Averages over the respective variables (left), with a breakdown by horizon h (right). For each variable (and horizon) the computed value is the fraction of forecasts that are better than the benchmark, and are statistically distinct from a "coin toss" forecast with the Pesaran-Timmerman p-value of at most 0.1. In each case, the total number of forecasts is given by the number of initial months(at most 3)*number of models(at most 6)*number of horizons(at most 4, only for the left subplot). Traditional and standard variables number 11, institutional instability variables 14.

Figure A21: Improvement of quarterly forecast combination on elimination of individual variables



The fraction of (month, horizon) pairs where the average combination forecast excluding variable v gave a smaller relative RMSFE than when not excluding the variable. The variables v are shown on the x-axis. The top row considers TradStanCM combination forecasts, the bottom row TradStanTV forecasts. The RMSFE is taken relative to the RMSFE associated with the TradStan combination, which thus functions as a benchmark. The x-axis order is the same across each row. Only individual variable exclusion is considered. For each (month, horizon) the RMSFE is averaged over the four models and three combination methods..