



ON FORECASTING CONFLICT IN THE SUDAN

2009 - 2002

November 2015 | The Center on Conflict and Development at Texas A&M University

AUTHORS

David A. Bessler, Shahriar Kibriya, Junyi Chen and Edwin Price

ABSTRACT

There is considerable evidence from around the world of a fundamental link between food insecurity and conflict. Food prices in particular appear to be an important focus in much of the literature—a result consistent with historical assessments offered in non-technical literature on food prices and conflict. Hendrix and Brinkman (2013) review this correlation or degree of ‘fit’ evidence between observations on conflict events and poverty (or some metric of food scarcity). Our purpose in this note is to consider the forecasting merit of a particular fit relationship: wheat prices and conflict in the Sudan over post-2000 data.

The Center on Conflict and Development (ConDev) at Texas A&M University seeks to improve the effectiveness of development programs and policies for conflict-affected and fragile countries through multidisciplinary research, education and development extension. The Center uses science and technology to reduce armed conflict, sustain families and communities during conflict, and assist states to rapidly recover from conflict. condevcenter.org

On Forecasting Conflict in the Sudan: 2009–2012

DAVID A. BESSLER,^{1*} SHAHRIAR KIBRIYA,² JUNYI CHEN¹ AND EDWIN PRICE^{1,2}

¹ *Department of Agricultural Economics, Texas A&M University, College Station, Texas, USA*

² *Center on Conflict and Development, Texas A&M University, Texas, USA*

ABSTRACT

This paper considers univariate and multivariate models to forecast monthly conflict events in the Sudan over the out-of-sample period 2009–2012. The models used to generate these forecasts were based on a specification from a machine learning algorithm fit to 2000–2008 monthly data. The model that includes previous month's wheat price performs better than a similar model which does not include past wheat prices (the univariate model). Both models did not perform well in forecasting conflict in a neighborhood of the 2012 'Heglig crisis'. Copyright © 2015 John Wiley & Sons, Ltd.

KEY WORDS conflict; probability; calibration; encompassing

INTRODUCTION

There is considerable evidence from around the world of a fundamental link between food insecurity and conflict. Food prices in particular appear to be an important focus in much of the literature—a result consistent with historical assessments offered in non-technical literature on food prices and conflict.¹ Hendrix and Brinkman (2013) review this correlation or degree of 'fit' evidence between observations on conflict events and poverty (or some metric of food scarcity). Our purpose in this note is to consider the forecasting merit of a particular fit relationship: wheat prices and conflict in the Sudan over post-2000 data.

This paper is organized in three sections. In the next section we offer a brief review of literature on forecasting of conflict. We follow with estimated (fit) models, both univariate and multivariate, of conflict, fatalities from conflict and wheat prices on monthly observations from 2001 to 2008. These models are specified using a recently minted algorithm from the machine learning literature. In terms of 'fit' relationships we show that wheat prices in period T ($W(T)$) are not influenced by either conflict or fatalities numbers, but are best modeled as a first order autoregressive process ($W(T-1)$). Conflict numbers in period T ($C(T)$) are modeled as first-order autoregressions on themselves ($C(T-1)$) and ($W(T-1)$). Finally, fatalities from conflict are modeled as a function of current conflict numbers ($C(T)$) and two lag of fatalities from conflict ($F(T-1)$ and $F(T-2)$). These 'fit' results are generally in line with results reported in Chen *et al.* (2014). Based on the models 'fit' with 2001–2008 data we explore out-of-sample forecast performance of both univariate and multivariate forecasts of wheat price, conflict and fatalities from conflict over 2009–2012 monthly data. Both point (mean) forecasts and probability forecasts are studied.

This paper makes a substantive contribution in two areas: it is the first paper on probabilistic forecasts of conflict in the Sudan. By investigating the overall 'goodness' of such forecasts we are able to assess the likelihood that certain periods of high or low conflict numbers were attributable to grain (in particular wheat) prices. Second, our paper demonstrates where improvement in knowledge and understanding of the causes of conflict are needed. While we do offer evidence that wheat prices are a mover of conflict numbers, we also demonstrate time periods where wheat prices are clearly not the cause of conflict events.

What emerges from this paper is the need to develop models capable of offering credible forecasts of extreme events. In our case both models under consideration do a poor job of forecasting conflict numbers associated with the June 2012 'Heglig crisis'. While we are able to offer convincing evidence that the large jump in conflict numbers from May to June 2012 is not due to wheat price innovations or due to business as usual, we are not able to offer a model capable of offering credible 'prevision' on its occurrence. While over our entire 'out-of-sample' forecasting period our probabilities do pass a statistical test of credibility, they suffer from a form of overconfidence at particular critical periods.

*Correspondence to: David A. Bessler, Department of Agricultural Economics, Texas A&M University, College Station, TX 77840, USA.
E-mail: d-bessler@tamu.edu

¹See Carlyle (1837) and Lynn (2014) on bread prices and the French Revolution and the *Economist* (2012) on food prices and the Arab Spring.

LITERATURE

The literature on forecasting conflict is relatively recent. We do not provide a thorough review here but rely on that offered by Brandt *et al.* (2011). They review several recent forays into the area. Three threads characterizing these models are: expected utility–game theoretic models, logistic regression and Bayesian vector autoregressions. The expected utility–game theoretic models (Feder, 1995; Bueno de Mesquita, 2002) postulate preferences between belligerents and can solve for conflict probabilities; and, while they do allow for experts to help specify preferences and other model parameters, they do not produce real-time forecasts. A second thread follows the ‘State Failure Project’ and work done at the CIA. Papers by King and Zeng (2001) and Goldstone *et al.* (2010) use logistic regression on highly aggregated data. While such models can also produce a probability of conflict, the aggregation level of data used makes real-world application somewhat questionable (although it is not exactly clear what the CIA uses in its real-time efforts). A third strand of literature reviewed by Brandt *et al.* (2011) is the structural Bayesian vector autoregression work described in Brandt and Freeman (2006) and extended by Brandt *et al.* (2011). This work models the dynamic evolution of conflict relevant variables as a vector autoregression. Probabilistic forecasts are offered and evaluated. Expert opinion can be incorporated through an initial prior. If the data are measured with high frequency, and available in a timely manner, this model can produce near real-time forecasts. Further, if experts (human or artificial) can provide contemporaneous structure among innovations, reliable conditional (what-if type analysis) probability forecasts are capable of generation.

Granger and Newbold (1986, p. 265), and many others, suggest the comparison of a set of forecasts with another (alternative set) to offer evidence of actual performance. In Casillas-Olvera and Bessler (2006) accuracy was measured in terms of performance of the Bank of England’s forecasters relative to a set of ‘other forecasters’, who regularly issued subjective (expert opinion) forecasts on inflation rates and GDP growth. Here we are not aware of a set of alternative forecasters to serve as a benchmark from which to judge the forecasts forthcoming from our data-based models. Accordingly, below we compare our point (mean) and probability forecasts from a multivariate model of conflict (conditioning on past wheat prices) against a model that does not condition on previous wheat prices.

Chen *et al.* (2014) provide ‘fit’ evidence that grain prices in the Sudan Granger-cause conflict numbers in the Sudan based on monthly data. Using a model similar in its dynamic structure to that of Brandt *et al.* (2011) and Chen *et al.* (2014) show that wheat price explains a considerable proportion of the (within sample) forecast error variance of conflict numbers in the Sudan using monthly data observed over 2001–2012.

In undivided Sudan wheat was primarily consumed in the northern part (Abdelrahman, 1998) especially by the urban dwellers. Following the trend of other East African countries wheat has become the staple food of non-subsistence households in Sudan during the last decade (Mkumbwa, 2011; Mustafa *et al.*, 2013). Rapid urbanization in Sudan shifted consumer preference from sorghum to processed bread made from wheat. Although the production of wheat has increased marginally, Mustafa *et al.* (2013) report that the average consumption of wheat has increased to 1770,000 tons in the 2000s from 743,000 tons in the 1980s. The lack production of wheat in Sudan implies larger imports, higher price volatility and government intervention.

DATA AND MODELS

The data used in the current paper are monthly wheat prices, conflict numbers and fatalities from conflict numbers from the Sudan for the period 2001–2012. The nominal price data per ton of wheat reflect the monthly wholesale prices in Khartoum port. The source of the dataset is the Global Information and Early Warning System (GIEWS) Food Price Data and Analysis Tool, a real-time data-streaming source of the Food and Agricultural Organization (FAO) of the United Nations. While Chen *et al.* (2014) use data on both North and South Sudan after the 2010 separation of the two countries; we use only data on North Sudan, as Chen *et al.* (2014) find a break point in the combined data from both countries in 2011, in a neighborhood of the division of the countries. Plots are given in Figure 1.

The data for the number of conflict events are obtained from the Armed Conflict Location and Event Dataset (ACLED) (Raleigh *et al.*, 2010) over the same period. The ACLED dataset provides time-specific information on recorded and reported battles, killings, riots and recruitment activities of rebels, governments, militias, armed groups, protesters and civilians. Finally, Chen *et al.* (2014) use the procedure presented in Hsiao (1979) to specify a subset vector autoregression; we follow the causal search literature, using PC Algorithm (Spirtes *et al.*, 2000) for model specification.

The graphical pattern among current and lagged wheat prices, conflict numbers and fatalities from conflict numbers found with PC Algorithm² is given in Figure 2. Note first that there are several contemporaneous relations

²PC Algorithm is a machine learning algorithm for finding causal structure (specifying exogenous or endogenous status) among a set of causally sufficient variables. The latter is suggested by prior theory and the problem under study. Key to PC’s ability to assign causal direction is the finding of unshielded colliders (inverted forks, say $A \rightarrow B \leftarrow C$). Such colliders are uniquely identified if the $\rho(A, C) = 0$; while $\rho(A, B)$ and $\rho(B, C) \neq 0$. Here $\rho(i, j)$ represents the correlation between variables i and j , for $i, j = A, B$ and C ; see Spirtes *et al.* (2000) or Pearl (2000) for details.

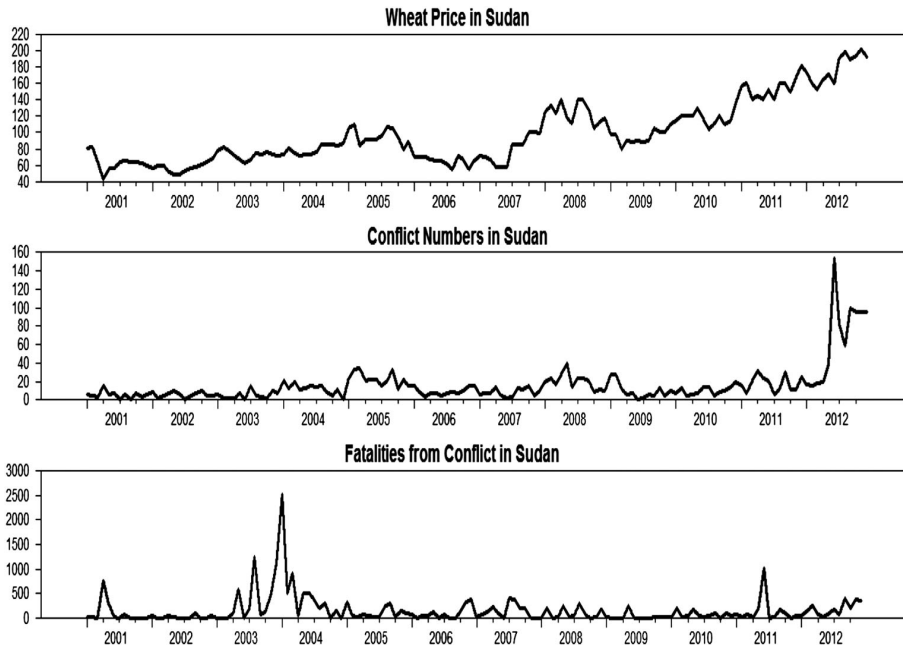


Figure 1. Time series plots of wheat price, conflict numbers, and fatalities from conflict in Sudan, 2001–2012, monthly data

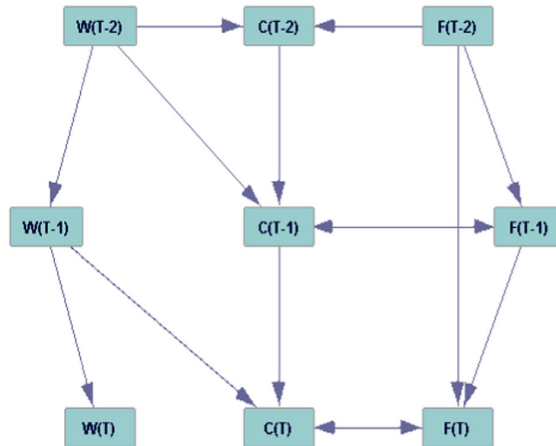


Figure 2. Graphical pattern found with PC Algorithm. Variable symbols are as follows: wheat price in period T is represented as $W(T)$, conflict in period T as $C(T)$ and fatalities from conflict in period T as $F(T)$. Lags of each of the above are indicated using $T-1$ and $T-2$ for one and two lags, respectively. Data used to fit the graph given here are monthly on wheat prices, conflict numbers and fatalities from conflict in the Sudan 2001–2008. The p -value used in PC Algorithm search was 10%. Details of PC Algorithm are given in Spirtes *et al.* (2000)

identified in the figure. $W(T-2)$ causes $C(T-2)$. $F(T-2)$, $F(T-1)$ and $F(T)$ and $C(T-2)$, $C(T-2)$ and $C(T)$ have contemporaneous relationships (the bidirectional arrow between conflict and fatalities indicates the presence of a latent variable that may move both conflict and fatalities in contemporaneous time). As these contemporaneous relationships cannot be exploited for forecasting purposes, we do not explore them further here. In terms of lagged (time-delayed) relationships we see wheat prices (W) are generated as a first-order (univariate) autoregression. Conflict (C) numbers are generated by one lag of wheat prices ($W(T-1)$) and one lag of conflict ($C(T-1)$) numbers. Fatalities from conflicts ($F(T)$) are generated by two lags of fatalities ($F(T-1)$, $F(T-2)$). The models fit³ over 2001–2008 monthly data are given as equations (1), (2) and (3) below:

³Models fit are with the heteroskedastic consistent ‘robust errors’ command for ordinary least squares found in RATS (Doan, 2008). Standard errors are given in parentheses below the parameter estimate. R^2 is the coefficient of determination and Q is the diagnostic Q -statistic associated with the null hypothesis of white noise errors (we reject the hypothesis with 23 degrees of freedom at the p -value indicated in each equation).

$$W(T) = 6.31(3.58) + 0.92(0.06)W(T-1); R^2 = 0.83; Q_{d.f.=23} = 30.4, p\text{-value} = 0.14 \quad (1)$$

$$C(T) = -5.65(2.40) + 0.17(0.04)W(T-1) + 0.30(0.11)C(T-1); R^2 = 0.47; Q_{d.f.=23} = 18.5, p\text{-value} = 0.73 \quad (2)$$

$$F(T) = 94.13(38.13) + 0.32(0.30)F(T-1) + 0.17(0.08)F(T-2); R^2 = 0.17; Q_{d.f.=23} = 16.8, p\text{-value} = 0.81 \quad (3)$$

Equation (1) indicates that wheat prices (W) are fit with a modestly high coefficient of determination (0.83), with no serious residual autocorrelation (as indicated by our failure to reject the null hypothesis of white noise residuals at usual (0.10 or 0.05) p -values). The degree of fit on the conflict equation (2) is much weaker ($R^2 = 0.47$) than that for wheat price. Interestingly, wheat prices lagged one period show a larger simple t -statistic ($t = 0.17/0.04$) than that associated with the estimated coefficient on conflict lagged one period in the second equation ($t = 0.30/0.11$). No strong evidence of autocorrelation amongst residuals is seen in equation (2) either. Finally, equation (3) has an even poorer fit than equation (2), as R^2 is just 0.17. It is the second lag of fatalities which shows the largest t -ratio ($t = 0.17/0.08$) in this third equation. Again, no evidence of autocorrelation in residuals is noted (p -value on Q is 0.81).

Point forecasts

Recursive forecasts of wheat price, conflict numbers and fatalities from conflict are generated from these equations (as recursively re-estimated) for the monthly data 2009–2012. Forecasting ‘goodness’ results for point forecasts (mean of the underlying probability distribution at each date) for one-step-, two-step- and 12-step-ahead horizons are given in Table I. These numbers are provided for both the vector autoregression (VAR) model summarized as equations (1), (2) and (3), as well as for the univariate forecasts (UNIV). As equations (1) and (3) are univariate, this comparison model is only different for equation (2), where lagged values of wheat price in the conflict equation are dropped. Accordingly, the results reported in Table I under the row labels *VAR* and *Univariate* are the same for *wheat price* and *fatalities*. The entries differ on *conflict* entries.

Table I. Forecast error metrics on one-step-ahead forecasts of wheat prices, conflict, and fatalities from conflict in the Sudan, 2009–2012, monthly data

Series/Model/Steps ahead	RMSE ^a	Theil U^b	Number of forecasts
Wheat			
<i>Univariate</i>			
1-step	11.89	1.03	48
2-steps	16.73	1.07	47
12-steps	45.80	1.49	37
<i>VAR</i>			
1-step	11.89	1.03	48
2-steps	16.73	1.07	47
12-steps	45.80	1.49	37
Conflict			
<i>Univariate</i>			
1-step	23.39	1.05	48
2-steps	29.51	1.08	47
12-steps	39.69	1.05	37
<i>VAR</i>			
1-step	21.55	0.97	48
2-steps	25.81	0.95	47
12-steps	35.92	0.95	37
Fatalities			
<i>Univariate</i>			
1-step	176.82	0.80	48
2-steps	179.30	0.76	47
12-steps	185.30	0.74	37
<i>VAR</i>			
1-step	176.82	0.80	48
2-steps	179.30	0.76	47
12-steps	185.30	0.74	37

^aRoot mean squared error (RMSE) is calculated as: $RMSE = (\sum_1^K (A(t) - F(t))^2 / K^{0.5})$, where $A(t)$ is the actual value, $F(t)$ is the forecast value from either the univariate or VAR model and K is the number of out-of-sample forecasts studied.

^bTheil U compares the MSE from the model (univariate or VAR) with that emanating from a random walk forecast (last period actual is the forecast of next period's actual). Theil $U < 1.0$ is preferred to Theil $U > 1.0$. Mean absolute per cent error (MAPE) is given at one step ahead as 0.07 (wheat price), 0.92 (conflict numbers) and 13.57 (fatalities). Numbers at higher levels are increasing and reflect the ordering $MAPE_{wheat} < MAPE_{conflict} < MAPE_{fatalities}$.

Table II. Encompassing tests on one-, two-, and three-step-ahead forecasts of conflict

Steps ahead	Null hypothesis	$\hat{\lambda}$	<i>p</i> -value	Decision
1	VAR encompasses UNIV	-0.64	0.11	Fail to reject
1	UNIV encompasses VAR	1.64	0.00	Reject
2	VAR encompasses UNIV	-0.80	0.02	Reject
2	UNIV encompasses VAR	1.80	0.00	Reject
12	VAR encompasses UNIV	-4.49	0.00	Reject
12	UNIV encompasses VAR	5.49	0.00	Reject

Note: The entries in the table refer to the estimated value of lambda ($\hat{\lambda}$) from an ordinary least squares regression, $e_{1t} = \lambda(e_{1t} - e_{2t}) + \varepsilon_t$, where e_{1t} refers to forecast errors from the encompassing model, e_{2t} refers to forecast errors from the encompassed model, both for out-of-sample data for period t , and ε_t is a white noise disturbance from the encompassing regression. Following Harvey and Newbold (2000), all regression models are estimated with White's 'robusterrors' procedure to account for possible heteroskedasticity in errors. We reject the null hypothesis (stated in the second column from the left for low *p*-values, i.e. values lower than 0.05. All estimation was carried out using RATS (Doan, 2008).

In terms of mean squared error (MSE) measures we see the usual relationship between forecast performance and horizon (steps ahead). As we look further out for all three variables our forecast performance deteriorates (as measured by higher MSE values). Note also that MSE measures for conflict forecasts associated with the VAR are lower than corresponding measures associated with the univariate model at all horizons (steps ahead). This latter result is consistent with Granger's notion of causality (Granger, 1980): wheat price causes conflict numbers. As a footnote to Table I, we report mean absolute percentage error at the one-step-ahead horizon. These metrics indicate that wheat prices are much easier to forecast (MAPE = 0.07), relative to conflict numbers (MAPE = 0.92) and fatalities (MAPE = 13.57)—a result consistent with the relative magnitudes of the R^2 values reported on each of equations (1), (2) and (3) above.

While the root mean squared error (RMSE) metrics presented in Table I indicate that forecasts of conflict numbers based on models that include past wheat prices in the model are 'better' than forecasts of conflict numbers that do not include past wheat prices (MSEs are lower for the former at all horizons), there may not be a significant difference between the measures compared. In Table II we present forecast encompassing results of the hypothesis that forecasts from the VAR encompass forecasts from the univariate model and vice versa. Encompassing tests address the question: Can we dispense with the encompassed model and focus decision making on the encompassing model? We offer results at all three horizons: Do forecasts from the VAR encompass forecasts from the univariate model? Do forecasts from the univariate model encompass forecasts from the VAR model? Each row of Table II offers a test of the encompassing hypothesis (see the note to the table for the specification of the test). We see that it is only at one step ahead that we do not reject the encompassing hypothesis: forecasts from the VAR encompass forecasts from the univariate model at one step ahead. At all other horizons the forecasts of each model do not encompass those of the other. This provides evidence that the 'wheat price causes conflict numbers' is a short-run phenomena (consistent with results reported in Chen *et al.* (2014) and in the graphical search results in Figure 2).

Continuing with the study of the relative merits of the two models (VAR and univariate) for forecasting conflict numbers we look at the d-separation (conditional independence) relation among forecasts from the univariate model, the VAR model and the actual realization. We do this for just the one-step horizon (given the outcomes of the encompassing test just reported). The d-separation⁴ results look for 'blockage of information flow' amongst variables. Figure 3 shows that forecasts of conflict numbers emanating from the univariate model are 'blocked' in their path to the actual realization of conflict numbers by forecasts from the VAR model—suggesting that one need not condition his/her decision making with respect to conflict numbers on the univariate model. The VAR forecasts capture all the relevant information (which is available from these two forecasting models) about next month's actual conflict numbers.

Probability forecasts

Of course, no point forecast can capture the extent of uncertainty of a future realization, whether such a forecast comes from a large-scale (many equations) model, a simple autoregression or from a human expert. Below we offer a comment on probability forecasts from the univariate autoregression and the VAR, whose point forecasts were studied above.

The literature on probability forecasts from models begins with Dawid (1984), although an earlier and certainly helpful literature on probability forecasting in psychology pre-dates Dawid's efforts by several years.⁵ Here we follow Kling and Bessler (1989) and bootstrap forecasts emanating from equations (1), (2) and (3) and the univariate form on equation (2). Uncertainty in errors on each equation and parameter uncertainty of each estimated parameter is accommodated by drawing on observed (historical) error terms in 1000 simulations at each out-of-sample data point. This gives us 1000 forecasts of each variable, wheat price, conflict numbers and fatalities, at each date, January 2009 to December 2012.

⁴D-separation (directional separation) is due to Judea Pearl (2000, p. 16) and is key to sorting out what needs to be conditioned on and what need not be conditioned on in decision-making environments. Bessler and Wang (2011) explore its use in forecast evaluation for agricultural prices and US GDP and inflation rates.

⁵See, for example, Lichtenstein *et al.* (1982, pp. 306–334).



Figure 3. Graphical pattern on forecasts of conflict and actual conflict numbers in the Sudan from monthly data 2009–2012. The forecasting models are the univariate (UAR) and multivariate (VAR) models found in Figure 2. Conflict has one lag of itself as its univariate (UAR) representation. Conflict has one lag of itself and one lag of wheat price as its multivariate (VAR) representation. The p -value used in the PC Algorithm search was 10%

Again, as in our point forecasting results, the model parameters are updated sequentially, by re-estimating at every month, as it moves through the out-of-sample data, January 2009 to December 2012.

Lichtenstein *et al.* (1982) suggests calibration as an indicator of ‘goodness’ of probability forecasts. So, for example, if our model (or our mind) issues a probability of $p = 0.25$ 100 times we should expect, after the fact, that 25 of these 100 outcomes should be realized (be ‘true’). Dawid (1984) builds on this calibration idea and the probability integral transform to offer a test of calibration through the associated cumulative density function (CDF). That is to say, if our issued probabilities are placed into their CDF form ($\int_{-\infty}^x P(x)dx$), where $P(x)$ is the probability distribution⁶ on conflict being x , generated from the 1000 draws simulation described above, then the realized fractile (fractile on event that actually occurs, after the fact), will be uniformly distributed if the issued distribution is well calibrated. A chi-squared test of good calibration is offered in Dawid; its application to our conflict study is presented below.

Calibration plots on conflict forecasts are presented in Figure 4. A calibration plot on well-calibrated forecasts would show the realized fractiles fall on a 45° line, in a plot of realized fractiles against the relative frequency (Bunn, 1984). Forecasts from both the univariate and VAR models of conflict have realized fractiles not on the 45° line, but generally above it. Visually, the univariate plots are somewhat closer to the line, relative to the VAR plots. A formal test of good calibration on both sets of forecasts is given in Table III. Here we do not reject the null hypothesis of good calibration for both the univariate and the VAR forecasts of conflict at a 5% level of significance. Similarly we do not reject the null of good calibration on the forecasts on wheat price and fatalities from conflict (Table III).

Given the decision reached in Table II, that our forecasts pass a chi-squared test of good calibration we investigate the issued probabilities on conflict for particularly interesting data realizations. First, we note that June 2009 was a particularly good month in that no conflicts occurred in the dataset. The previous months of April and May saw 6 and 7 conflict numbers, respectively, reported. The issued probabilities for June 2009 are given in histogram form in Figure 5. Both the VAR and the univariate model issue probabilities that contain the actual realization in their modal class ($0 \leq C_{\text{June 2009}} < 10$), where $C_{\text{June 2009}}$ is the conflict number in June of 2009.⁷ The VAR gave slightly preferred density forecasts, in the sense that its issued probability for this bin class was 0.595; the univariate model issued its probability on this class as 0.588.

A case where both models show less or no forecasting quality is found for the date June 2012. This date corresponds to the largest number of conflict numbers in our dataset (153 conflicts). Figure 6 shows the probability histograms for both

⁶Actually we use summation operations (not integrals) as our work, below, involves discrete probabilities, not continuous ones.

⁷Defining classes or bins for aggregating probabilities is somewhat arbitrary. There is an extensive literature that attempts to decide on the number of probability bins from historical data (see, for example, Scott, 1979; Wand, 1997). Here we argue that the proper number and size of each bin (the size does not have to be the same across all bins) is ‘best’ defined by subject matter experts (most likely in the field) who must deal with the consequences of conflict events. If we follow the statistics literature (and there is not universal agreement on selection of bin width) we find bins in the neighborhood of 25 conflict numbers (using the result derived by Wand (1997), $h = 3.49 (\sigma)/N^{1/3}$, where h is bin width, σ is the standard deviation of the data (prior to our forecasting interval) and N refers to the number of observations prior to our forecasting interval. Our subjective judgment is that bin width should be a bit narrower at low levels of conflict and perhaps higher at higher levels of conflict. Needless to say, this is an area for further research. We use the outcome bins in Figures 5–7 and apply the same bin widths, defined as follows: bin 1 = $P(0 \leq C_T < 10)$; bin 2 = $P(10 \leq C_T < 20)$; bin 3 = $P(20 \leq C_T < 30)$; bin 4 = $P(30 \leq C_T < 40)$; bin 5 = $P(40 \leq C_T < 80)$; and bin 6 = $P(80 \leq C_T < \infty)$; where C_T is conflict at time T .

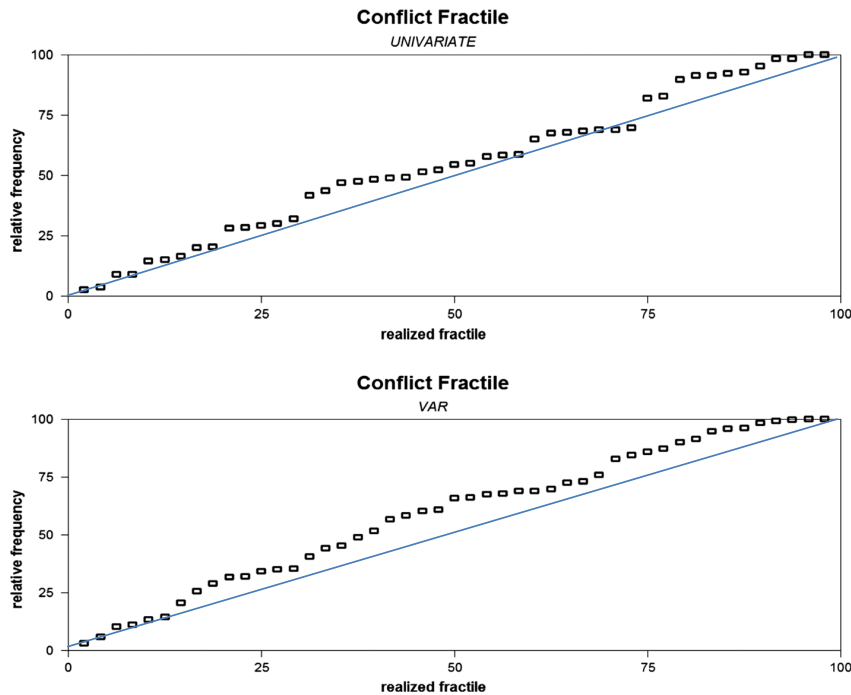


Figure 4. Realized fractiles and relative frequencies on conflict events from the univariate model and VAR model probability forecasts: Sudan, January 2009 to December 2012.

Table III. Chi-squared statistics on calibration of probability forecasts of wheat prices, conflict and fatalities from conflict in the Sudan, 2009–2012, monthly data

Series/Model	Calculated ^a χ^2	Critical $\chi_{(0.05,d.f.=9)}^2$ ^b	Decision with respect to the null hypothesis
<i>Wheat</i>			
Univariate ^b	11.58	16.9	Fail to reject
VAR ^c	11.58	16.9	Fail to reject
<i>Conflict</i>			
Univariate ^b	12.83	16.9	Fail to reject
Vector autoregression ^c	14.91	16.9	Fail to reject
<i>Fatalities</i>			
Univariate ^b	8.3	16.9	Fail to reject
Vector autoregression ^c	8.3	16.9	Fail to reject

^aIn the table, the null hypothesis in every case is that the issued probabilities are well calibrated. The χ^2 statistics are calculated as follows: $\chi^2 = \sum_{k=1}^{10} (m_i - 4.8)^2 / 4.8$; where m_i are the number of realized fractiles in category i , with categories defined as $c_1 = (0, 0.099)$, $c_2 = (0.100, 0.199)$, ..., $c_{10} = (0.900, 0.999)$.

^bUnivariate forecasts are generated from the fit model $Y_t = \beta_0 + \beta_1 Y_{t-1} + \dots + \beta_k Y_{t-k} + \varepsilon_t$, where Y_t is a scalar holding either observations at time t on conflict or fatalities, and $\beta_0, \beta_1, \dots, \beta_k$ are scalars, recursively updated at each data point; 1000 draws on the sampling distributions on $\beta_{0,i}, \beta_{1,i}, \dots, \beta_{k,i}$ and ε_{t+1} are used to form 1000 forecasts, $\hat{Y}_{t+1,i} = \beta_{0,i} + \beta_{1,i} Y_t + \dots + \beta_{k,i} Y_{t-k+1} + \varepsilon_{t+1,i}$; $i = 1, \dots, 1000$.

^cThe vector autoregressive forecasts are generated as: $Y_t = \beta_0 + \beta_1 Y_{t-1} + \dots + \beta_k Y_{t-k} + \varepsilon_{t+1}$; where Y_t is a vector holding wheat price, sorghum price, conflict and fatalities at time t , β_0 is a vector of constants, one for each series, and β_1, \dots, β_k are matrices holding estimated coefficients, relating Y_t to lags Y_{t-1}, \dots, Y_{t-k} . The estimated betas are recursively updated at each data point; 1000 draws on the sampling distributions on $\beta_{0,i}, \beta_{1,i}, \dots, \beta_{k,i}$ and ε_{t+1} are used to form 1000 forecasts, $\hat{Y}_{t+1,i} = \beta_{0,i} + \beta_{1,i} Y_t + \dots + \beta_{k,i} Y_{t-k+1} + \varepsilon_{t+1,i}$; $i = 1, \dots, 1000$.

models are centered on bin 3 = $P(20 \leq C_T < 30)$, with neither model issuing non-zero probabilities for the realized event (conflict for June 2012 = 153). This date corresponds to the ‘Heglig crisis’—an oil-related set of conflict events between Sudan and South Sudan. There is no (ex ante) hint of this extreme number in our data. Conflict numbers for the preceding 2 months (May and April) are reported as 38 and 19, respectively. Further, wheat price was within its recent historical range at \$170/ton for May 2012 and \$165/ton for April 2012, both down from the high of \$181/ton for December 2011. Clearly, both models did not offer credible forecasts of conflict for June 2012.

Another date which may be of interest is December 2011. This date corresponds to the highest wheat price observation in our post-fit, forecasting, period of study. Here wheat price (as indicated above) was \$181/ton. Figure 7 shows the associated histograms on conflict numbers for this month. Recall the VAR model has wheat price in its information set, while

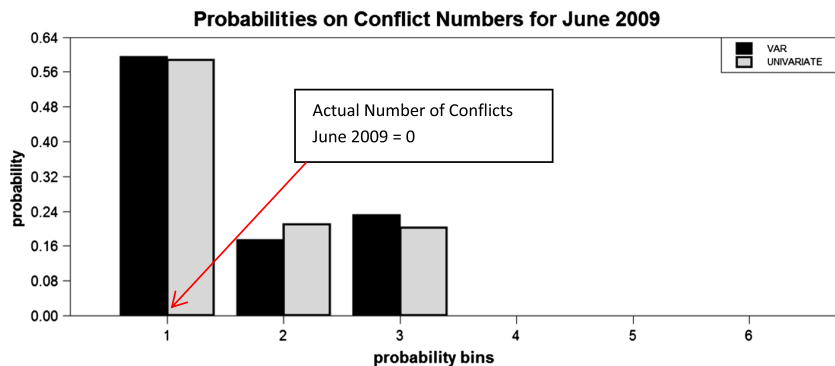


Figure 5. Probabilities issued on conflict numbers from univariate and VAR model for June 2009. The probability bins in Figure 5 are defined as follows: bin 1 = $P(0 \leq C_{\text{June 2009}} < 10)$; bin 2 = $P(10 \leq C_{\text{June 2009}} < 20)$; bin 3 = $P(20 \leq C_{\text{June 2009}} < 30)$; bin 4 = $P(30 \leq C_{\text{June 2009}} < 40)$; bin 5 = $P(40 \leq C_{\text{June 2009}} < 80)$; and bin 6 = $P(80 \leq C_{\text{June 2009}} < \infty)$

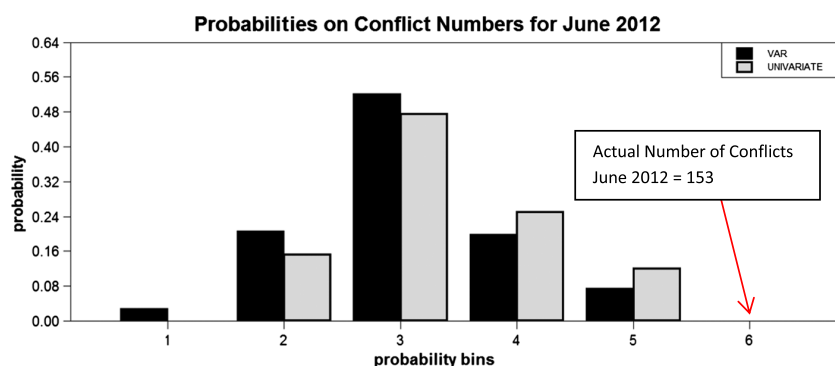


Figure 6. Probabilities issued on conflict numbers from univariate and VAR model for June 2012. The probability bins in Figure 6 are defined as follows: bin 1 = $P(0 \leq C_{\text{June 2012}} < 10)$; bin 2 = $P(10 \leq C_{\text{June 2012}} < 20)$; bin 3 = $P(20 \leq C_{\text{June 2012}} < 30)$; bin 4 = $P(30 \leq C_{\text{June 2012}} < 40)$; bin 5 = $P(40 \leq C_{\text{June 2012}} < 80)$; and bin 6 = $P(80 \leq C_{\text{June 2012}} < \infty)$.

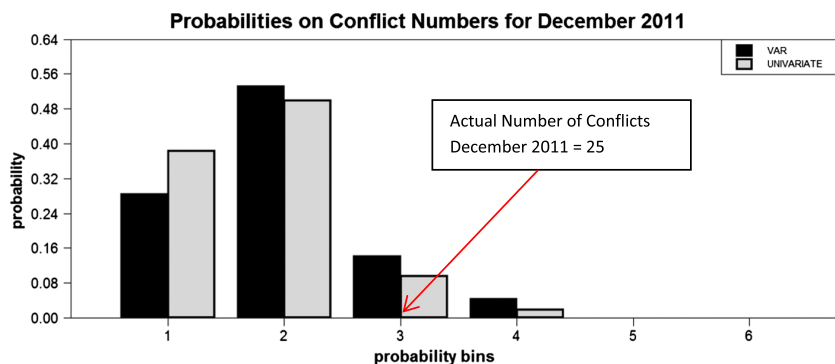


Figure 7. Probabilities Issued on conflict numbers from univariate and VAR model for December 2011. The probability bins in Figure 7 are defined as follows: bin 1 = $P(0 \leq C_{\text{December 2011}} < 10)$; bin 2 = $P(10 \leq C_{\text{December 2011}} < 20)$; bin 3 = $P(20 \leq C_{\text{December 2011}} < 30)$; bin 4 = $P(30 \leq C_{\text{December 2011}} < 40)$; bin 5 = $P(40 \leq C_{\text{December 2011}} < 80)$; and bin 6 = $P(80 \leq C_{\text{December 2011}} < \infty)$

the univariate model does not. The actual number of conflicts for December 2011 was 25. Conflict numbers for the preceding 2 months were 12 and 11, respectively. Both models issue their modal class as bin 2 = $P(10 \leq C_{\text{December 2011}} < 20)$; however, the VAR model places more density in this bin class relative to the univariate model ($0.141 > 0.096$). The latter places more density in bin 1 and less in bin 3 (the realized bin) relative to the VAR model.

DISCUSSION

Here we have explored the ability of two models in forecasting conflict numbers in the Sudan. An earlier paper found a ‘fit’ relation between wheat prices and conflict numbers in the Sudan. We use similar data and explore the out-of-

sample forecasting ability of similar models over 2009–2012 monthly data on wheat prices, conflict numbers and fatalities in North Sudan. Both point (mean) and probability forecasts are studied from two candidate models: a univariate model, which expresses conflict numbers at time t as a function of conflict numbers at time $t - 1$ and a vector autoregression model which expresses conflict numbers at time t as dependent on conflict numbers at $t - 1$ and wheat prices at time $t - 1$. Point forecasting results indicate that the VAR model performs better in terms of root mean squared forecast errors and forecast encompassing. Both models offer well-calibrated probability forecasts over our post-fit period. Clearly, knowledge of wheat prices in month t is helpful in forecasting conflict numbers for month $t + 1$ —a result that coheres well with ‘fit’ results found on earlier data via a machine learning algorithm.

While we do not reject good calibration for both models’ probability forecasts, neither performs well in their forecasts conflict for June 2012. This month corresponds to the ‘Heglig crisis’—an oil-related set of conflict events between North and South Sudan. There is no information in either the history of conflict numbers or wheat prices to hint of the large number of conflicts that occurred at this date. Both models issue probability forecasts of zero on conflict events in a neighborhood of the number actually observed. This clearly exposes the need to rely on more than formal models on historically relevant data in forecasting conflict (or in terms of forecasting human subject outcomes, in general). In our literature review we cite the need to bring experts into the forecasting exercise, as suggested by Brandt *et al.* (2011).

Some information variables are measured and reported regularly—market prices are one such variable. Also, as food is of particular importance for human existence, its price can offer us prevision of upcoming conflict. Other information is not reported regularly, as it occurs in a seemingly unsystematic pattern. The ‘Heglig crisis’, which erupted in extreme conflict in June of 2012, was clearly not an event caused by wheat prices. Most discussions cite oil as the root cause (see the UN report cited below). Here subjective judgment may well have improved our model’s forecasts. In May 2012 the Security Council Report from the United Nations discussed this crisis. This discussion indicates clearly the existence of evidence that an increase in conflict was eminent. The following quote of the UN report supports that such evidence was communicated to both parties of the potential (and finally realized) conflict: ‘A key issue is whether and how the Council can exert sufficient leverage on the parties to deter them from expanding their conflict, induce them to cease fighting, and convince them to return in good faith to the negotiating table. Since February, the Council has produced two press statements and two presidential statements regarding the situation in Sudan and South Sudan with what appears to be minimal impact on the calculations of the parties.’

Even if we can forecast increases in conflict, can we find some way to prevent or mitigate the possibility of such events coming to pass? Such efforts at mitigation would, in turn, require more elaborate forecasting models, as we would now have to offer our forecasts on the probability of conflict, given that a particular mitigation strategy has been undertaken.

ACKNOWLEDGEMENTS

This research was funded in part by the Howard G. Buffett Foundation and United States Agency for International Development. The views expressed in this article are solely those of the authors.

REFERENCES

- Abdelrahman AH-E. 1998. Trends in Sudanese Cereal Production, Consumption, and Trade. Center for Agricultural and Rural Development, Iowa State University: Ames, IA.
- Bessler DA, Wang Z. 2011. D-separation, forecasting, and economic science: a conjecture. *Theory and Decision* **73**: 295–314.
- Brandt PT, Freeman JR. 2006. Advances in Bayesian time series modeling and the study of politics: theory testing, forecasting, and policy analysis. *Political Analysis* **14**(1): 1–36.
- Brandt PT, Freeman JR, Schrodt PA. 2011. Real time, time series forecasting of inter-and intra-state political conflict. *Conflict Management and Peace Science* **28**: 41–64.
- Bueno de Mesquita B. 2002. Predicting Policies. Ohio State University Press: Columbus, OH.
- Bunn D. 1984. Applied Decision Analysis. McGraw-Hill: New York.
- Carlyle T. 1837. The French Revolution: A History (2002). Modern Library: New York.
- Casillas-Olvera G, Bessler DA. 2006. Probability forecasting and central bank accountability. *Journal of Policy Modeling* **28**: 223–234.
- Chen J, Kybriya S, Price E, Bessler DA. 2014. A causal exploration of conflict and commodity prices of undivided Sudan. Working paper, Conflict and Development Center, Texas A&M University.
- Dawid AP. 1984. Statistical theory: a prequential approach. *Journal of the Royal Statistical Society* **147**: 278–297.
- Doan T. 2008. RATS, version 7. Estima: Evanston, IL.
- Economist 2012. Food and the Arab Spring: let them eat baklava. Available: <http://www.economist.com/node/21550328> [13 August 2014].
- Feder S. 1995. Faction and Policon: new ways to analyze politics. In *Inside the CIA’s Private World: Declassified Articles from the Agency’s International Journal 1955–1992*, Bradford H (ed.). Yale University Press: New Haven, CT; 274–292.
- Goldstone JA, Bates RH, Epstein DL, Gurr TR, Lustik MB, Marshall MG, Ulfelder J, Woodward M. 2010. A global model of forecasting political instability. *American Journal of Political Science* **54**: 190–208.
- Granger CWJ. 1980. Testing for causality: a personal viewpoint. *Journal of Economic Dynamics and control* **2**: 329–352.

- Granger CWJ, Newbold P. 1986. *Forecasting Economic Time Series* (2nd edn). London: Academic Press.
- Harvey D, Newbold P. 2000. Tests for multiple forecast encompassing. *Journal of Applied Econometrics* **15**: 471–482.
- Hendrix C, Brinkman H. 2013. Food insecurity and conflict dynamics: causal linkages and complex feedbacks. *Stability* **2**(2): 1–18.
- Hsiao C. 1979. Autoregressive modeling of Canadian money and income data. *Journal of the American Statistical Association* **74**: 553–560.
- King G, Zeng L. 2001. Improving forecasts of state failure. *World Politics* **53**: 623–658.
- Kling J, Bessler DA. 1989. Calibration-based predictive distributions: an application of prequential analysis to interest rates, money, prices and output. *Journal of Business* **62**(4): 447–500.
- Lichtenstein S, Fischhoff B, Phillips L. 1982. Calibration of probabilities: the state of the art to 1980. In *Judgment under Uncertainty: Heuristics and Biases*, Kahneman D, Slovic P, Tversky A (eds). Cambridge University Press: Cambridge, UK; 306–334.
- Lynn MR. 2014. Riots and rye: bread and the French Revolution. Available: <http://www.ultimatehistoryproject.com/bread.html> [5 August 2014].
- Mkumbwa SS. 2011. Cereal food commodities in Eastern Africa: consumption–production gap trends and projections for 2020. MPRA paper 42113, 22 October 2012. Available: <http://mpra.ub.uni-muenchen.de/42113> [24 October 2015].
- Mustafa RH, Elagali MB, Abulgazim S. 2013. Trends in wheat production and consumption in Sudan. *International Journal of Research in Management, Economics and Commerce* **3**(4): 44–56.
- Pearl J. 2000. *Causality: Models, reasoning, and inference*. Cambridge University Press: Cambridge, UK.
- Raleigh C, Linke A, Hegre H, Karlsen J. 2010. Introducing ACLED: an armed conflict location and event dataset. *Journal of Peace Research* **47**(5): 1–10.
- Scott DW. 1979. On optimal and data-based histograms. *Biometrika* **66**: 605–610.
- Spirtes P, Glymour C, Scheines R. 2000. *Causality, prediction, and search*. MIT Press: Cambridge, MA.
- United Nations. 2012. Security Council Report of the UN, New York. Available: http://www.securitycouncilreport.org/monthly-forecast/2012-05/lookup_c_glKWLeMTIsG_b_8075191.php [13 August 2014].
- Wand MP. 1997. Data-based choice of histogram bin width. *American Statistician* **51**: 59–64.

Authors' biographies:

David A. Bessler is a Professor of Agricultural Economics at Texas A&M University. Over his career he has focused on research related to applications of time series methods and machine learning to price discovery in agricultural markets for products separated by space, time, and form.

Shahriar Kibriya is the research lead at the Center on Conflict and Development (ConDev) at Texas A&M University. He focuses on studying models of conflict, food security and social organization of developing nations. He is involved in the implementation of several research based pilot development projects in Central America and Africa.

Junyi Chen is a Ph.D. candidate in Agricultural Economics at Texas A&M University. She has focused on the field of applied econometrics including time series methods, machine learning, panel data, and semiparametric methods.

Edwin Price is the Howard G. Buffett Foundation endowed professor and the director of Center on Conflict and Development (ConDev) at Texas A&M University. He focuses on creating and utilizing the role of development in preventing and mitigating armed conflict.

Authors' addresses:

David A. Bessler, Department of Agricultural Economics, Texas A&M University, College Station, Texas, USA.

Shahriar Kibriya, Center on Conflict and Development, Texas A&M University, Texas, USA.

Junyi Chen, Department of Agricultural Economics, Texas A&M University, College Station, Texas, USA.

Edwin Price, Department of Agricultural Economics, Center on Conflict and Development, Texas A&M University, Texas, USA.