RESEARCH PAPER No. 6

ASSESSING VOLATILITY PATTERNS IN FOOD CROPS

MARCH 2017

AMIS

This study was prepared for the Agricultural Market Information System of the G20 (AMIS) by Friederike Greb and Adam Prakash. It has been endorsed by the Steering Committee of the AMIS Secretariat, which represents all member organizations of AMIS. The designations employed and the presentation of materials in this information product do not imply the expression of any opinion whatsoever on the part of AMIS, its Secretariat or participating countries concerning the legal or development status of any country, territory, city or area of its authorities, or concerning the delamination of its frontiers or boundaries.

Contents

Introduction

Prices for food crops are naturally volatile – their supply depends on unpredictable factors like the weather. While volatility is not problematic *per se*, uncertain and excessive price movements present a threat. Abnormally high agricultural price volatility can have severe impacts on governments, which have to finance imports of foodstuffs and also rely on export earnings from commodities. Large effects are also felt by farmers and consumers, especially in developing countries, where about two billion people live off small farms and spend large shares of their income on food items (IFAD, 2016). Especially in the absence of coping mechanisms such as storage, savings or access to credit and insurance, the repercussions of extreme price instability can be devastating. Vulnerable households are left with little scope to mitigate unusually high prices other than by lessening the intake of nutritious food, dropping out of school, lowering access to healthcare or distress sales of land and livestock. These responses can result in poverty traps and, accordingly, have long-term consequences. Producers are affected too. As sellers of commodities, high volatility brings with it considerable downside price risk, which affects planting decisions and undermines agricultural investment where it is needed most.

We have witnessed episodes of low volatility alternating with phases of high volatility (Greb and Prakash, 2015). However, it is not well understood what brings about these changes and which variables could forewarn increased volatility levels in advance. It becomes, therefore, crucial to anticipate an increase in price volatility, to prepare for such periods, or even to design policies to prevent them. Theoretically, mechanisms that influence the frequency and amplitude of price movements are firmly established in the literature. An example is the role of stocks in buffering supply shocks (Wright, 2011). The demand curve for wheat incorporates two different demands, the demand for immediate consumption and the demand for stocks to be accumulated for deferred consumption. This additional demand prevents price slumps in case of abundant supplies and at the same time allows to smooth prices in case of a shortfall – unless stocks are too low. Yet, in order to derive concrete policy implications, it is essential to acknowledge that not only low inventories can trigger an increase in the amplitude of price spikes, but pinpoint the critical inventory level that needs to be maintained to prevent such spikes from happening.

This study aims at identifying such thresholds associated with transitions between different levels of price volatility. Being statistically driven, the study surveys a host of factors related to price dynamics of food crops, and tries to isolate relevant determinants of volatility transition and to discover their change points. In terms of exact methodology, our analysis rests on a statistical estimation and variable selection algorithm termed component-wise gradient boosting. We adopt this approach because it enables us to flexibly model the influence of different variables, including abrupt transitions. In addition, it allows one to incorporate and select from a multitude of potentially relevant variables, even collinear ones. While policy recommendations naturally follow from the results, there are also findings that demand further investigation.

The study is structured as follows. We explain our modelling and estimation strategy in the next section. In the third section, we describe the data employed and covariates used. Following a presentation and discussion of the estimation results in the fourth section, the fifth section concludes the paper. We defer complete estimation results to an appendix.

Methods

We closely follow the methodology suggested in Mittnik *et al.* (2015). They analyse the volatility of the S&P 500 index employing componentwise gradient boosting to an exponential ARCH model including exogenous variables. This setup allows to not only identify relevant external factors associated with stock market volatility and quantify their impact, but also formulate a model with higher predictive performance than the benchmark models, which Mittnik *et al.* (2015) argue are the GARCH(1,1) model (Hansen and Lunde, 2005) and, in our specific context, the exponential GARCH model.

Modelling framework

While the GARCH model is arguably the most prolific way to represent conditional variances, we adopt Nelson's (1991) exponential ARCH model, which overcomes certain drawbacks of the GARCH framework (such as ruling out a negative correlation between current returns and future volatility or imposing restrictions on the parameters that might unduly limit the dynamics of the conditional variances). More precisely, for prices p_t we model logarithmic returns $r_t = \log(p_t/p_{t-1})$ as

$$
\begin{aligned}
r_t &= \exp(\eta_t/2)\,\varepsilon_t \\
\eta_t &= \beta_0 + \beta_1 \cdot t + f_{year}(y_t) + f_{month}(m_t) + \\
&\quad \sum_{i=1}^{3} f_i(r_{t-i}) + \sum_{j=1}^{J}\sum_{k=0}^{K_j} f_{j,k}(x_{j,t-k}) \qquad (1)
\end{aligned}
$$

where $\varepsilon_t \sim \mathcal{N}(0,1)$ are independent normal errors and $t = 1, \ldots, T$. The variable y_t indicates the year of the t-th observation, the variable m_t the month, and $x_{j,t}$, $j = 1, \ldots, J$, denote additional explanatory variables with a potential impact on volatility. The number of lags K_j to be included is variable-specific and varies between zero and three. We define functions $f_\cdot : \mathbb{R} \mapsto \mathbb{R}$ as so-called "trees"

$$f_\cdot(z) = \sum_{s=1}^{S} \gamma_s I_{A_s}(z) \quad \text{where} \quad I_{A_s}(z) = \begin{cases} 1 & \text{if } z \in A_s \\ 0 & \text{if } z \notin A_s \end{cases}$$

for disjoint intervals $A_s \subset \mathbb{R}$ $s = 1 \ldots S$, partitioning $\mathbb{R}$. This functional form is very flexible and allows to capture nonlinearities such as abrupt transitions in the dependencies.

For $S = 2$, $A_1 = \{z|z \le c\}$ and $A_2 = \{z|z > c\}$ $c \in \mathbb{R}$, we get a function

$$f(z) = \begin{cases} \gamma_1 & \text{if } z \le c \\ \gamma_2 & \text{if } z > c, \end{cases}$$

that is, our modelling framework is versatile enough to, for example, include a stocks-to-disappearance ratio z triggering an increase in price volatility by $\exp(0.5 \cdot (\gamma_1 - \gamma_2))$ when falling below a critical threshold c.

Boosting (estimation and variable selection)

There are distinct advantages to using componentwise gradient boosting to estimate model (1). This estimation method allows to include a very large number of covariates, even collinear ones, suspected to be related to the outcome. The fitting process incorporates a mechanism for model choice, which selects variables that are relevant. In addition, while allowing for considerable flexibility in the shape of the dependencies between response and explanatory variables, it still yields interpretable results.

Boosting aims to estimate a function η relating explanatory variables $Z_1, \ldots, Z_p$ to a response variable Y by minimizing the expected loss $\mathbb{E}\left[\ell(Y, \eta(Z_1, \ldots, Z_p))\right]$ for a loss function ℓ. It is a stepwise gradient descent algorithm approaching the minimum of the observed mean loss, $\frac{1}{T}\sum_{t=1}^{T} \ell(y_t, \eta(z_{1,t}, \ldots, z_{p,t}))$, until reaching a stopping iteration m_{stop}.

Buehlmann and Hothorn (2007) detail the boosting algorithm after specifying a base learning procedure – i.e. a way to generate an estimate $\hat{g}$ of a function g relating a real-valued response Y_t to p-dimensional predictor variables Z_t for $1, \ldots, T$ – as follows,

(i) Initialize $\hat{\eta}^0$ with an offset value and set $m = 0$.

(ii) Increase m by 1, compute the negative gradient $-\frac{\partial}{\partial \eta}\ell(y_t, \eta)$ and evaluate it at $\hat{\eta}^{m-1}(z_t)$ for $t = 1, \ldots, T$,

$$u_t^m = -\frac{\partial}{\partial \eta}\ell(y_t, \eta)|_{\eta = \hat{\eta}^{m-1}(z_t)}$$

The negative gradient can be thought of as a generalized residual; for the squared error loss function $\ell(y_t, \eta) = 1/2\ (\eta - y_t)^2$ we get $u_t^m = y_t - \eta$, the residual.

(iii) Fit the so-called pseudo-residuals $u_1^m, \ldots, u_T^m$ to $z_1, \ldots, z_T$ by the base learning procedure

$$(z_t, u_t^m)_{t=1}^T \xrightarrow{\text{base procedure}} \hat{g}^m$$

(iv) Update

$$\hat{\eta}^m = \hat{\eta}^{m-1} + \nu \cdot \hat{g}^m$$

for a small step-length $\nu \in (0, 1]$

(v) Iterate steps (2) to (4) until $m = m_{stop}$ is reached for some fixed m_{stop} .

Componentwise gradient boosting refers to a modification of this algorithm. In the third step "base-learners" or regression estimators based on a fixed subset of the predictor variables, are fitted individually to the negative gradients $u_1^m, \ldots, u_T^m$ and only the fit of the base-learner correlating most with the gradients is included in step (iv) as $\hat{g}^m$. More precisely, this means changing step (iii) to

(iii) For each of the Q base learners, fit g_q to the negative gradient vector $u_1^m, \ldots, u_T^m$

$$(z_t, u_t^m)_{t=1}^T \xrightarrow{\text{q-th base learner}} \hat{g}_q^m$$

Select the base-learner that yields the best fit according to the sum of squared residuals

$$q^* = \underset{1 \le q < Q}{\operatorname{argmin}} \sum_{t=1}^{T} \left(u_t^m - \hat{g}_q^m(z_t)\right)^2$$

and set $\hat{g}^m = \hat{g}_{q^*}^m$. (To simplify notation we write $\hat{g}_q^m(z_t)$ although the q-th base-learner does not necessarily depend on all p variables $Z_1, \ldots, Z_p$ but only a subset.)

The stopping iteration m_{stop} is a crucial tuning parameter of the algorithm. We use cross-validation based on 250 bootstrap samples of size T to choose the stopping iteration m_{stop}, however, other approaches such as AIC-based criteria are also common. Clearly, only one base-learner enters the estimate $\hat{\eta}$ (and, hence, the model) at each iteration. Covariates that have not been selected as the best-fitting component at any step $m \le m_{stop}$ become redundant and drop out of the model.

In case of our model (1) there are $p = 6 + J + \sum_{j=1}^{J} K_j$ base-learners. We specify all but one – an ordinary least squares base-learner for a linear time trend $\beta_0 + \beta_1 \cdot t$ – as regression trees with two nodes and the estimator proposed by Hothorn *et al.* (2006). The effects of year, month, lagged returns and additional exogenous variables and their lags enter in this way. To complete the description of our estimation procedure, we define ℓ as the negative log-likelihood loss function according to (1).

It might trouble some readers that our manuscript does not present any significance tests. This is inherent in the modelling approach we take. In his paper on the "two cultures" of statistical modelling Breiman (2001) distinguishes two different goals of data analysis – information and prediction – leading to explanatory and predictive modelling, respectively. In the case of explanatory modelling, theory determines the specific model relating independent variables $Z_1, \dots, Z_p$ to the dependent variable Y. Hypotheses about this theoretical construction are then tested. However, when data feature complex patterns and relationships, the attempt to formulate a model based on theory does not seem a natural or promising approach. Instead, an algorithmic solution associating the variables according to a criterion of predictive accuracy can be used – predictive modelling. Shmueli (2010) points out several dimensions of the differences between these two ways to analyse data. Namely, the first assumes a causal relation between $Z_1, \dots, Z_p$ and Y while it is mere association in the second case; in the first case model-building is theory-driven, while it is data-driven in the second; we take a backward-looking perspective in the first, a forward-looking one in the second case. In particular, model validation differs. Explanatory models are assessed testing the goodness of fit and examining residuals. Predictive performance, on the contrary, is the criterion to validate predictive models. Given that commodity price volatility is determined by a complex interplay of numerous factors, it is predictive power, not statistical significance that guides our analysis. Whereas it might still be interesting to formulate and test hypotheses about the model, to the best of our knowledge, inference within a boosting framework is ongoing research and significance tests are not yet available.

Data and covariates

Price data

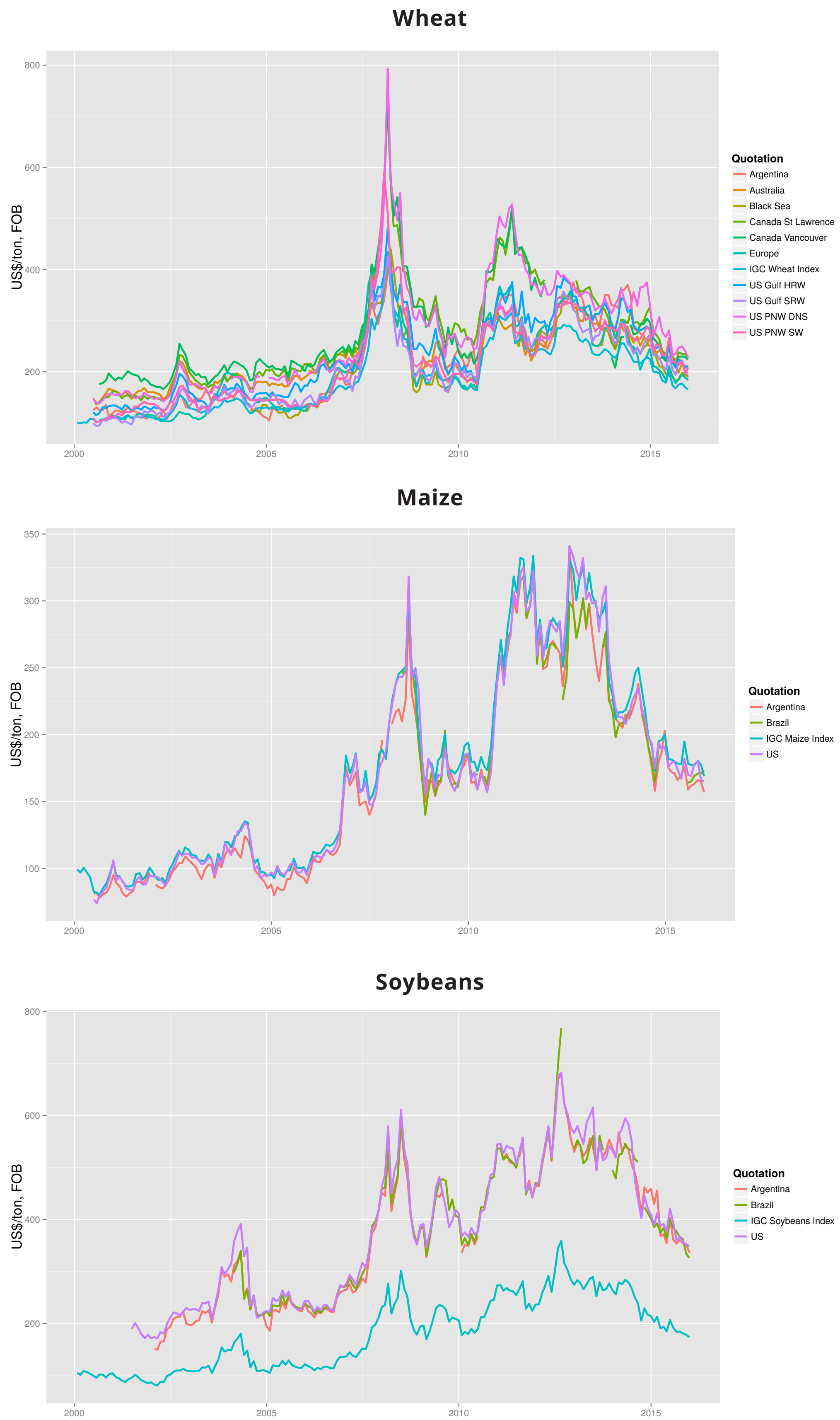
We examine price volatility for three of the four AMIS commodities – wheat, maize, and soybeans. Because of the idiosyncrasies of the rice market, namely its absence on global commodity exchanges and the importance of domestic policies in price formation which cannot be readily quantified, we omit this commodity. We study price data obtained from the International Grains Council (IGC), namely their Grain and Oilseed Index' (GOI) sub-indexes for wheat, maize and soybeans as well as the quotations included in these indexes. The GOI's sub-index for wheat is an unweighted average of ten export quotations – Argentinian Trigo Pan (Up River), Australian ASW (Port Adelaide), Black Sea milling wheat, Canadian No. 1 CWRS, 13.5% (St. Lawrence), Canadian No. 1 CWRS, 13.5% (Vancouver), European/French standard grade wheat (Rouen), US No. 2 HRW (Gulf), US No. 2 SRW (Gulf), US No. 2 DNS, 14% (PNW), and US No. 2 SW (PNW). Quotations included in the GOI's sub-index for maize are from Argentina (Rosario, Up River), Black Sea, Brazil (Paranagua), as well as US No.3 Yellow (Gulf). Again, all quotations enter the average with identical weight. As we do not have prices for the Black Sea before August 2010, we do not analyze this quotation separately. Similarly, prices from Argentina (Rosario, Up River), Brazil (Paranagua), and US No.2 Yellow (Gulf) enter the GOI's sub-index for soybeans (again, an unweighted average). All quotations are export prices in USD per ton, FOB. In Figure 1 we display indexes together with the individual quotations by commodity.

Our analysis focuses on monthly data since we try to quantify the impact of variables that change at a monthly, but not necessarily daily frequency. For example, the most influential update of the stocks-to-disappearance ratio (see following section), one of the variables we look at, becomes available with the release of the World Agricultural Supply and Demand Estimates Report (WASDE) by the United States Department of Agriculture (USDA). This happens once a month, generally between the 8th and the 12th of each month. Our data series start in January 2000 and run until December 2015. While we study monthly observations, our underlying IGC data are weekly or even daily. To assure robustness of our findings in view of different possible ways to aggregate these, we decide to simultaneously analyze three sets of monthly price observations per commodity. For each, we create a series of the last week's (or last day of the last week's) price, the third week's (or last day of the third week's) price and a monthly average price.

Covariates

As suggested in Algieri's (2014) investigation of the main drivers of the international wheat price, we consider four broad classes of variables – market fundamentals, macro-economic, financial and weather variables. To avoid ambiguities and facilitate recognizing the variables in Table 2 to Table 7 (in the appendix), we include short names identifying the variables in bold type and parenthesis in the text.

Figure 1: GOI sub-indexes and individual quotations for wheat, maize, and soybeans



Market fundamentals

Inventories

Inventories are crucial to smooth consumption. Thus, they can also help to even out prices. A bumper crop triggers less of a fall in prices when part of it is stored. Releasing stocks can absorb the supply shock of a bad harvest and prevent a spike in prices. Naturally, this is only possible if stocks are sufficient to cover the loss. When stocks are low commodity prices are more prone to big movements. The relationship between stocks and commodity prices has been explored in detail in Deaton and Laroque (1992), Bobenrieth *et al*. (2013), Wright (2011), and Pindyck (2001) to name just a few. We include four variables related to inventories in our analysis – the world's stocks-to-use ratio (**stocks2useW**), the stocks-to-use ratio for the United States (**stocks2useUS**), and the stocks-to-disappearance ratio for major exporters with and without the United States (**stocks2disMExUS** and **stocks2disMEx**, respectively). The stocks-to-use ratio is defined as ending stocks over utilization, the stocks-to-disappearance ratio as ending stocks over domestic utilization together with exports.

Yield

Balcombe (2011) proposes yield as a potential determinant of volatility (**yield**). He conjectures that the impact of extraordinarily high and low yields on volatility might differ. However, he points out that even if a dramatic fall in prices might result in an increase in volatility, it might also result in a decrease due to replenishment of stocks.

Market thinness

A thin market is characterized by a low number of buyers and sellers, hence, few transactions, which makes it more susceptible to price fluctuations. Relating the number of transactions to the quantities traded, we follow Algieri (2014) in measuring thinness as the ratio of exports to global consumption (**thinness1**) or, alternatively, exports to global production (**thinness2**).

Market concentration

Another trade-related characteristic of global commodity markets is the degree of export concentration (**Herfindahl**). If only two or three important players account for nearly all exports of a commodity, domestic weather shocks or changes in their policies can have a strong impact on prices. The more diverse are the exporters and the smaller their respective shares, the better is the insulation of the market against idiosyncratic shocks. We use the Herfindahl index to measure export concentration (Balcombe, 2011). It is defined as

$$H = \sum_{n=1}^{N} s_n^2 \quad \text{where} \quad s_n = x_n \Big/ \sum_{i=1}^{N} x_i$$

and x_i denotes the quantity exported by the i-th country. The closer the index is to zero, the less concentrated the market. A single player, in contrast, would result in a Herfindahl index of one.

For all of the above variables we use the USDA's WASDE as our data source. An intrinsic difficulty when compiling data series from commodity balance sheets is to determine the relevant marketing year for countries' crops. The WASDE shows harmonized estimates across a given marketing year (MY) and forecasts for the next. To give an example, the report released in January 2000 shows an estimate for the ending stocks of 1998/99 MY and a forecast for the 1999/2000 MY. Among the two of these, the level of the 1999/2000 MY ending stocks has a stronger impact on the market in January 2000, and hence will constitute the value of our time series for January 2000. In May of any given year, however, WASDE shifts marketing years. That is, while the April 2000 report gives estimates for the 1998/99 MY and forecasts for the 1999/2000 MY, the May 2000 report gives estimates for the 1999/2000 MY and forecasts for the 2000/2001 MY. As it is unlikely that the 1999/2000 MY ending stocks suddenly become irrelevant for the market, we decide to take a weighted average of the 1999/2000 MY's ending stocks estimate and the 2000/2001 MY's ending stocks forecast to generate a single entry for the time series in May 2000. We create all variables based on data from commodity sheets as weighted averages (with the new marketing year linearly gaining influence while the old one looses it) for two different marketing years for the months of May, June, and July; and we take forecasts for the current marketing year for the remaining months.

Macroeconomic variables

Interest rates

Part of the cost of carry of a commodity inventory is the opportunity cost of forgone interest. A lower interest rate, thus, reduces the cost of storing a commodity, resulting in upward pressure on its price. An increase in interest rates, on the contrary, triggers the reverse price movement. Frankel (2008) provides a detailed investigation of the interconnections between monetary policy and commodities. Another consequence of a change in interest rates can be increased (in case of a lower interest rate) or decreased (in case of a higher interest rate) flow of money into commodity futures or options, perceived as financial assets by non-commercial traders. We include the Effective Federal Funds Rate (**FEDFUNDS**) as well as the 6-Month Treasury Bill (**TB6MS**) as variables in our model.

Oil price
The oil price affects the price for agricultural commodities through different channels. Input costs depend on energy costs via processing and fertilizer costs. Transportation and distribution costs also vary with the oil price. In addition, prices of oil and agricultural commodities are linked to each other via biofuel. It might be important to not only focus on the price of oil but also on its volatility. Here we make two volatility estimates part of our analysis, realized volatility and conditional volatility. We compute realized volatility from daily data as the sample standard deviation of logarithmic returns in a month. Conditional volatility estimates are based on a GARCH model. Level, realized and conditional volatility for two crude oil benchmark prices, the Western Texas Intermediate price in Cushing, Oklahoma, (**WTI**) and the European Brent price (**BRENT**), enter our study.

Stock market volatility
As a general measure of economic risk and uncertainty we include the Chicago Board Options Exchange's Volatility Index VIX (**VIX**). The index measures market expectation of near term volatility conveyed by stock index option prices. It is based on the S&P 500 index' expected volatility, computed as the mean of weighted prices of put and call options on the S&P 500 index over a wide range of strike prices. Details on its calculation are gathered in CBOE (2015). The VIX, frequently cited in the media, is sometimes referred to as the "fear gauge" (although CBOE (2009) claims this to be a misnomer).

Our data source for interest rates, oil prices as well as VIX is the Board of Governors of the US Federal Reserve System. We retrieve them from FRED, Federal Reserve Bank of St. Louis (https://research.stlouisfed.org/fred2/).

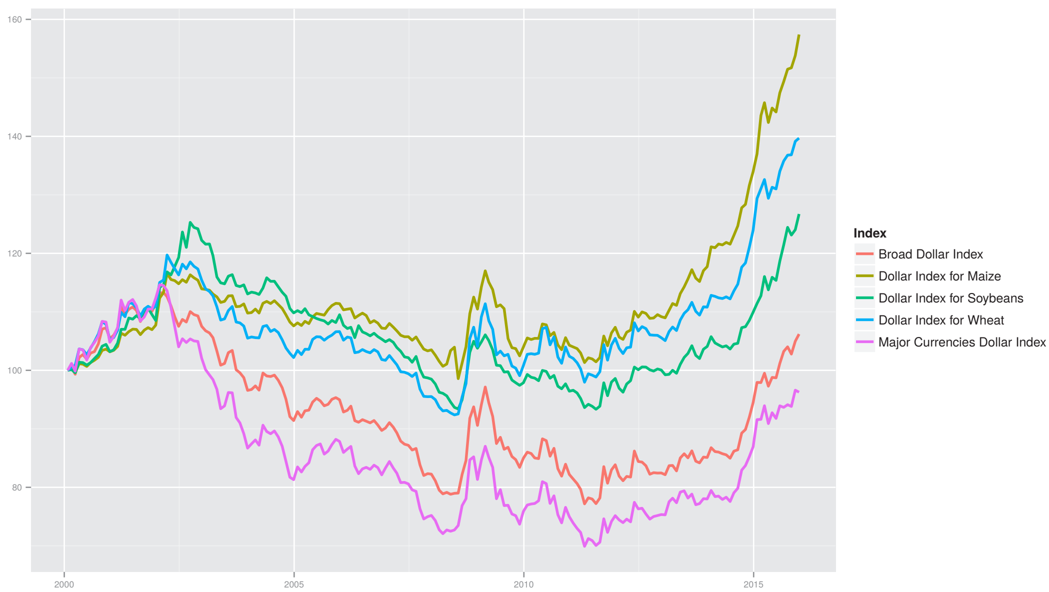
Foreign exchange
International commodity prices have typically had an inverse relationship with the value of the USD. When the USD strengthens against other major currencies, commodity prices tend to fall. On the contrary, when the value of the USD weakens against other major currencies, the prices of commodities increase. The relationship is chiefly a result of commodities being priced in USD and of international buyers being required to purchase them with USD. When the value of the USD rises (falls), buyers have less (more) purchasing power and so demand usually weakens (strengthens). Hence, changes in exchange rates reallocate purchasing power and price incentives for buyers and sellers. We construct indexes (**FX**) of the foreign exchange value of the USD by commodity along the lines of the Trade-Weighted U.S. Dollar Index published by the US Federal Reserve System (Loretan, 2005), a weighted geometric mean of bilateral exchange rates,

$$I_t = I_{t-1} \times \prod_{j=1}^{J} \left(e_{j,t} \middle/ e_{j,t-1} \right)^{w_{j,t}}$$

I_t is the value of the index at time t, $e_{j,t}$ and $w_{j,t}$ are USD exchange rate and the weight of the j-th currency at time t, respectively. For each commodity, we determine the set of relevant currencies as those of countries that together account for 90% of the exports and imports (in any year between 2000 and 2015). Weights are updated yearly and based on trade shares,

$$w_{j,t} = 0.5 \cdot s_{j,t}^{IMPORT} + 0.5 \cdot s_{j,t}^{EXPORT} ,$$

Figure 2: Dollar indexes by commodity, major currencies and broad dollar index



where $s_{j,t}^{IMPORT}$ ($s_{j,t}^{EXPORT}$) is the import (export) share of the j-th country at time t. The index is set to equal 100 in January 2000. We obtain trade data from FAOSTAT and exchange rates from OANDA (http://www.oanda.com/currency/historical-rates/). Indexes are shown in Figure 2 along with the Major Currencies and Broad Dollar Index (standardized to equal 100 in January 2000 as well).

Variables related to financialization

There has been extensive discussion in the aftermath of the 2007/08 price spikes in agricultural commodity markets about the effect of financialization on price levels and volatility. Financialization refers to "the process of alignment of commodities returns with pure financial assets ('pooling effect'), so increasing co-movements among asset classes that have been historically seen as following opposite causal pattern" (Valiante, 2013, page 52). There is no consensus in the literature regarding the role of financialization, whether it led to price spikes and increased volatility. Will *et al.* (2012) survey this literature; Lagi *et al*. (2015) and Etienne *et al*. (2015) provide more recent reviews.

Studies on the topic of financialization of agricultural commodity markets rely on different variables. Our analysis, thus, contains not just one but a number of them. A common indicator (Valiante, 2013; Robles *et al.*, 2009; Algieri, 2014) is the number of trades in futures contracts (**Volume**). We focus on volumes at the Chicago Board of Trade (CBoT) as a leading exchange. According to Robles *et al*. (2009) contracts traded typically expire within the next 24 months. As the CBoT's wheat, maize, and soybeans future contracts expire five times a year, we add up volumes traded for ten (or nine, depending on data availability) future continuations. Data source is Thomson Reuters.

The U.S. Commodity Futures Trading Commission (CFTC) publishes a weekly report on commitments of traders showing open interest broken down into reportable – commercial and non-commercial – and non-reportable positions. A trader's position becomes reportable once it exceeds levels set by the CFTC. A trader is further identified as commercial if futures contracts held serve hedging purposes, while non-commercial traders typically do not hold a position in the underlying commodity (CFTC, 2016). We use the CFTC's Commitment of Traders reports to create a set of variables. One of these is open interest, the total of all futures contracts entered into and not yet offset by a transaction, delivery, or exercise (**OpenInterest**).

If we assume that those traders who view agricultural commodity futures and options as just another asset class and without interest in the actual underlying commodity show a trading behavior distinct from commercial players, it can be instructive to examine the ratio of volume to open interest (**Vol2OpenInterest**). We suspect that these speculative traders quickly get into and out of the market whenever opportunities arise, hence, augmenting the number of trades, i.e. volume, but not necessarily increasing open interest at the same time.

We further use the CFTC's separation of commercial from non-commercial traders to capture the relative amount of actors in pursuit of financial profits instead of hedging existing risk. We include the ratio of both long and short non-commercial to total positions in CBoT futures contracts in our study (**S2TotalL** and **S2TotalS**, respectively).

Valiante (2013, Figure 36) shows evidence for differences in net positions in agricultural commodity markets. While commercial traders tend to be net short on average, it is their financial counterparts that are net long. We include net non-commercial long positions as another measure of speculative activity frequently used in the literature (**netL**) (Micu, 2005; Domanski and Heath, 2007).

Working's speculative index goes a step further, distinguishing speculative positions necessary to absorb residual positions of commercial actors from those going beyond this position (**Working**) (McPhail *et al.*, 2012; Bastianin *et al.*, 2012; Algieri, 2014). To this end, the index is defined as

$$W = \begin{cases} 1 + S_{nc}/(S_c + L_c), & \text{if } S_c \geq L_c \\ 1 + L_{nc}/(S_c + L_c), & \text{if } L_c > S_c \end{cases}$$

where $S_{(n)c}$ are (non) commercial short and $L_{(n)c}$ (non) commercial long positions. Because the aggregate of all short open interest has to equal that of all long open interest, $S_{nc} + S_c = L_{nc} + L_c$, it is clear that whenever $S_c \neq L_c$ some non-commercial position $S_{nc} > 0$ or $L_{nc} > 0$ is needed to equilibrate excess L_c or S_c, respectively. Hence, in case of $S_c > L_c$, L_{nc} has to be greater than zero for long positions to offset short positions. However, if, in addition, $S_{nc} > 0$ this indicates that speculation positions exceed what is necessary to balance hedges by commercial market participants.

Weather variables

Weather determines conditions for crop growth and, ultimately, crop outcomes. Thus, weather variability can cause fluctuations in supply of agricultural commodities and their prices. We include FAO's Agricultural Stress Index as a broad measure of vegetation health (**ASI**). This index is based on the integration of the so-called Vegetation Health Index in two dimensions that are crucial for the assessment of a drought event in agriculture, time and space (Rojas, 2015). We use its annual summary, which reports the percentage of arable land with a Vegetation Health Index below a critical threshold. FAO is the source of these data.

A different group of indicators quantifying weather conditions relevant for crop growth is based on measurements related to El Niño and La Niña. These are irregularly occurring changes in wind and rainfall patterns, which affect agriculture. They are the result of a chain of events prompted by unusually high (El Niño) or low (La Niña) sea surface temperatures in a specific area in the Pacific Ocean. Hence, a natural indicator to consider is sea surface temperature anomalies in this particular region, the Niño 3.4 region (**NINO34**) (Ubilava, 2014; Algieri, 2014). An El Niño (La Niña) event is characterized by the Oceanic Niño Index, the running 3-months average sea surface temperature anomaly for the Niño 3.4 region, being above 0.5°C (below −0.5°C) in five consecutive instances (Rojas *et al*., 2014). For this reason, we also consider the Oceanic Niño Index in our analysis (**ONI**).

Another variable indicating episodes of El Niño or La Niña is the Southern Oscillation Index (**SOI**). It measures large-scale fluctuations in air pressure between Tahiti and Darwin, Australia, typical for El Niño and La Niña episodes. Unusually low air pressure at Tahiti combined with abnormally high air pressure at Darwin results in negative values for the Southern Oscillation Index. These coincide with changes in the water temperature that characterize El Niño. The reverse – positive values for the Southern Oscillation Index – holds for episodes of La Niña.

We collect El Niño related data from the National Oceanic and Atmospheric Administration's Climate Prediction Center (http://www.cpc.ncep.noaa.gov/).

Results and discussion

Interpretation of estimates

Before we enter a discussion of the estimates, some comments might be helpful to correctly interpret the figures and tables. Looking at formula (1), we see that volatility, i.e. the standard deviation of the logarithmic price returns $\exp(\eta_t/2)$,is modeled as a product of different factors,

$$\begin{aligned}&\exp(\eta_t/2)\\ &= \exp\left\{\frac{1}{2}\left[\beta_0 + \beta_1 \cdot t + f_{year}(y_t) + f_{month}(m_t) + \sum_{i=1}^{3} f_i(r_{t-i}) + \sum_{j=1}^{J}\sum_{k=0}^{K_j} f_{j,k}(x_{j,t-k})\right]\right\}\\ &= \exp(\beta_0/2)\cdot\exp(\beta_1\, t/2)\cdot\exp\left[f_{year}(y_t)/2\right]\cdot\exp\left[f_{month}(m_t)/2\right]\cdot\\ &\prod_{i=1}^{3}\exp\left[f_i(r_{t-i})/2\right]\cdot\prod_{j=1}^{J}\prod_{k=0}^{K_j}\exp\left[f_{j,k}(x_{j,t-k})/2\right]\end{aligned}$$

Each of the factors involving a tree $f_{\cdot}(z) = \sum_{s=1}^{S}\gamma_s I_{A_s}(z)$ takes S different values, $\exp(\gamma_1/2), \dots, \exp(\gamma_S/2)$. We divide these by the value the function $f_{\cdot}(z)$ takes for an arbitrarily small z, e.g. $\exp(\gamma_1/2)$ in case $A_1 = (-\infty, c\,]$, and augment the product $\exp(\eta_t/2)$ by the same factor. This leaves us with trees $\tilde{f}_{\cdot}(z)$ that take the value one for $z \to -\infty$–that is, $\tilde{\gamma}_1 = 1$– still assuming $A_1 = (-\infty, c\,]$ which facilitates assessment of the impact of passing a critical threshold and its comparison across variables and between different price series.

For illustration we return to the very simple tree introduced in section 2,

$$f(z_t) = \begin{cases}\gamma_1 & \text{if } z_t \le c\\ \gamma_2 & \text{if } z_t > c,\end{cases}$$

where $z_t \in V_t = \{y_t, m_t, r_{t-1}, r_{t-2}, r_{t-3}, x_{1,t}, \dots, x_{J,t-K_J}\}$,and re-write

$$\begin{aligned}&\exp(\eta_t/2) = \exp\left[\eta_t/2 - f(z_t)/2\right]\cdot\exp\left[f(z_t)/2\right]\\ &= \exp\left[\eta_t/2 - f(z_t)/2\right]\cdot\exp(\gamma_1/2)\cdot\begin{cases}1 & \text{if } z_t \le c\\ \exp\left[(\gamma_2-\gamma_1)/2\right] & \text{if } z_t > c.\end{cases}\end{aligned}$$

Given that the term $\exp\left[\eta_t/2 - f(z_t)/2\right]\cdot\exp(\gamma_1/2)$ does not depend on the variable z_t, we see that holding all variables $v_t \in V_t$ $v_t \neq z_t$, constant, the effect of the variable z_t passing the critical threshold c is a change in volatility by the factor $\exp\left[(\gamma_2-\gamma_1)/2\right]$ relative to its level for arbitrarily small z_t.

As an example of how to interpret Figure 3 to Figure 5 (as well as the figures in the appendix), we focus on the upper left panel of Figure 3. Here we observe a decrease in volatility of the price of Canadian wheat at the port of St Lawrence by nearly 30 percent (i.e. multiplication by a factor of 0.72 instead of 1 or a jump down by 0.28) when the stocks-to-use ratio in the United States is above a critical value of 37 percent. The volatility of the International Grain Council's wheat index also decreases when the US stocks-to-use ratio grows larger than 37 percent, however, only by 20 percent.

Interrelation between estimates

Before discussing details, it might be helpful to comment on how we relate estimation results for different price series but the same commodity to each other. While model selection and estimated trees should be somewhat similar among prices for the same commodity, we do not expect them to be identical. Indeed, we find that, for example, there is a 39 percent threshold for the United States' stocks-to-use ratio for US No.2 Soft Red Winter wheat, below which the level of volatility jumps up. However, for most

Figure 3: Selected estimates for wheat index and prices (monthly averages)

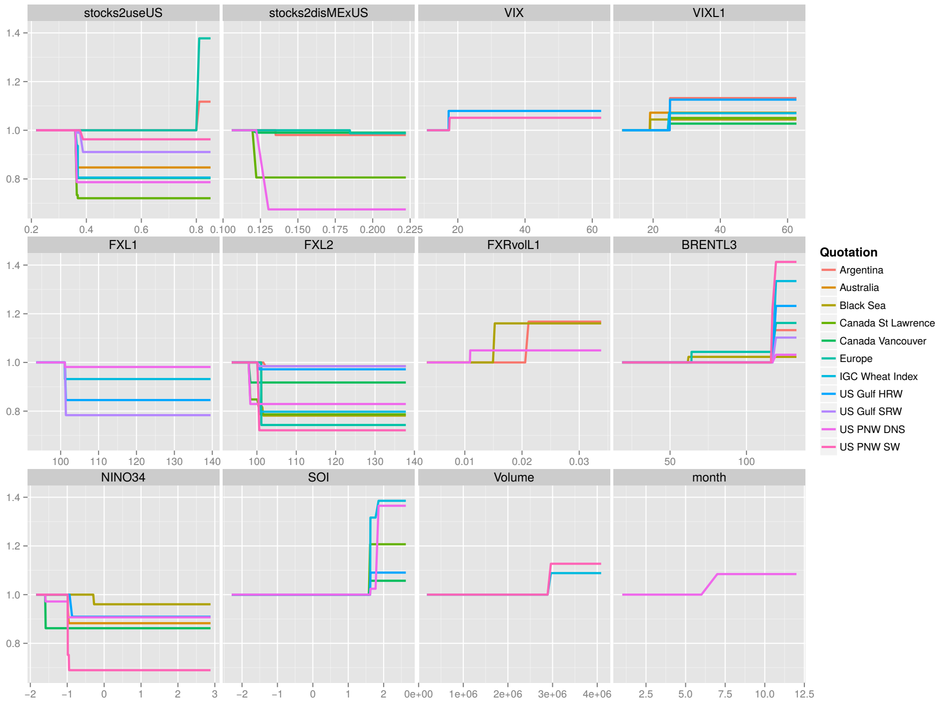
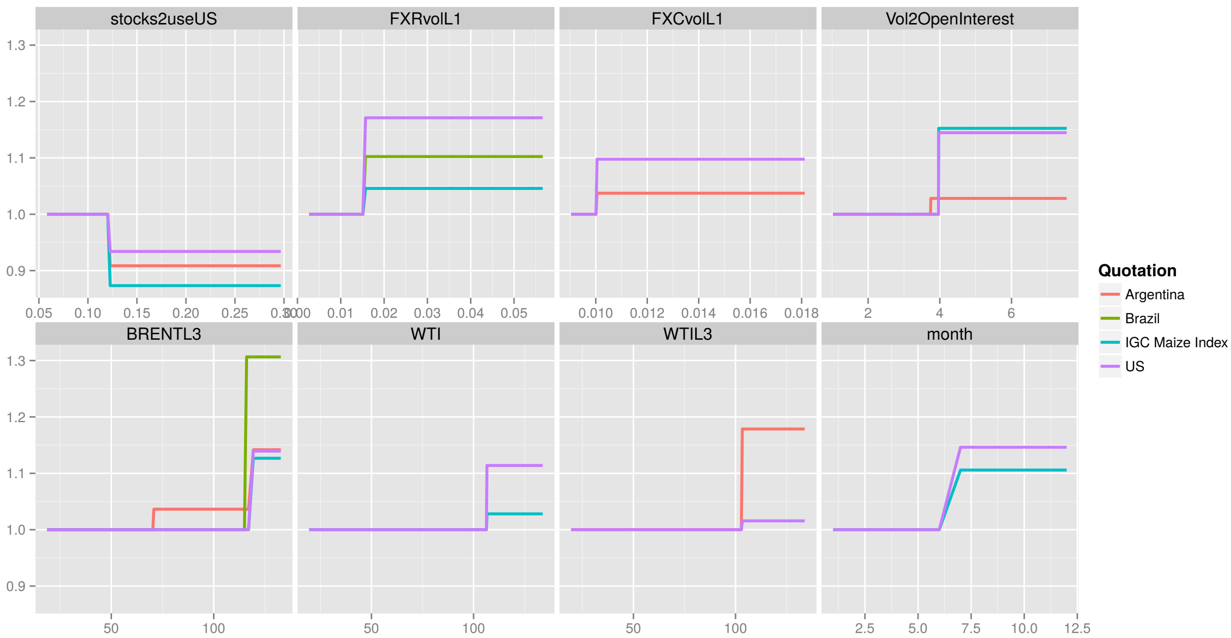


Figure 4: Selected estimates for maize index and prices (monthly averages)



other quotations examined (within as well as outside the US) the threshold is 37 percent (upper-left panel in Figure 3). Hence, we face the question of how to combine results into a coherent whole and distil a clear set of indicators per commodity market. It would be ingenious to suggest individual critical US stocks-to-use thresholds for wheat by quotation – would the policy recommendation then aim to maintain US inventories above 37 or 39 percent relative to use?

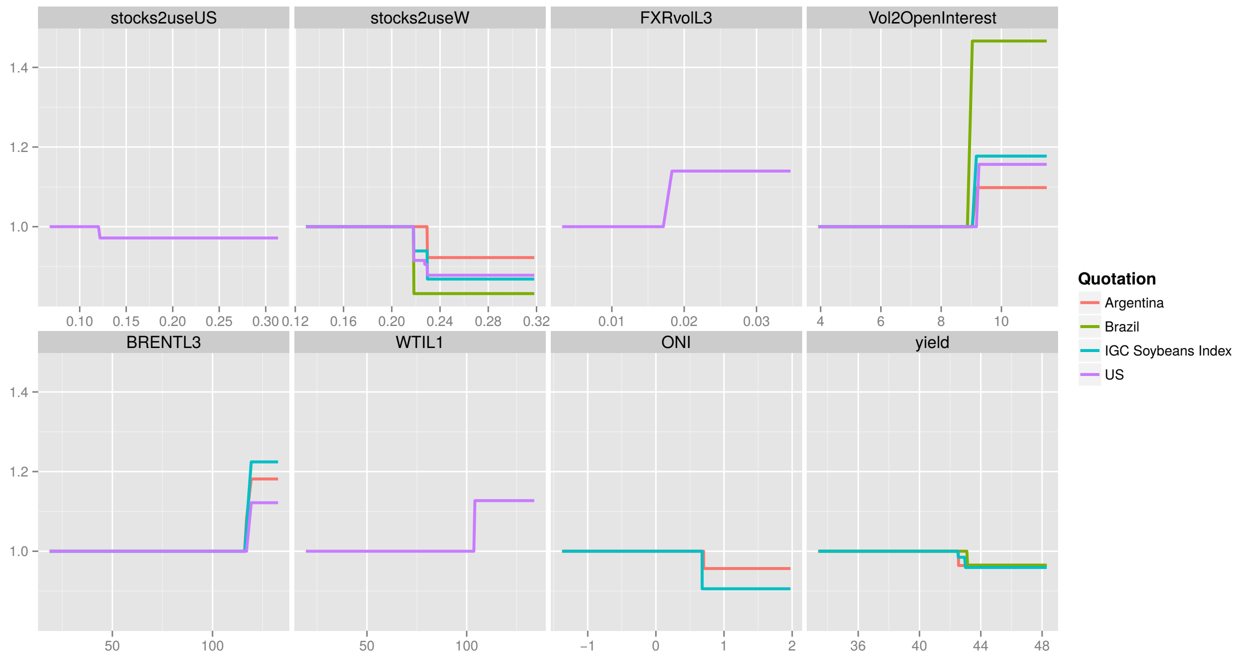
Naturally, it would be easier to interpret estimates had we only examined one benchmark price or index per commodity. However, a complex interplay of factors determines price volatility. It is plausible for certain effects to only play out under particular circumstances, which could be given in one location but not in another. For example, a country could insulate itself from volatility due to low global inventories by changing its trade policies. These opposite effects on volatility (an increase due to low stocks versus a decrease due to policy response) offset each other. It becomes impossible to detect and distinguish the two mechanisms when looking at the respective price time series. Still, it is unlikely for two effects to occur simultaneously in all price series examined. Thus, by considering more than one series, we have higher chances of discovering dynamics and get a more complete picture.

Moreover, when there is trade prices in different locations influence each other. They are connected via the Law of One Price (Fackler and Goodwin, 2001). Price movements transmit in space and volatility can spill over. Consequently, if we find an indicator related to an increase in volatility for one series, this might be relevant not only for this particular series, but for others as well, even if not directly reflected so in the estimation results.

Finding similar critical thresholds prompting a change in the level of volatility in different series gives confidence that we do not observe statistical artefacts, but that outcomes are robust. Yet, differences should not be unsettling. They could indicate that an effect has been cancelled by another one; that it has been concealed when aggregating (in case of the price index); that it may not be crucial for every series directly, but possibly indirectly through spillover of price movements between locations.

When interpreting the estimates, it is further helpful to keep in mind that not all of them are equally reliable. Some series feature long stretches of missing data, which can reduce the quality of estimates. For wheat this is the case for quotations in the Black Sea, but the Australian and Canadian (Vancouver) series are not complete either. For maize, some Brazilian data are missing. We

Figure 5: Selected estimates for soybean index and prices (monthly averages)



do not analyze Black Sea maize prices because of data availability. The least reliable soybean estimates are those for Argentina and Brazil because of gaps in the data series.

Results by covariate

To facilitate the presentation of results, unless we state otherwise and in particular in the figures we refer to estimates for monthly price averages. Still, we do not merely estimate the remaining two sets of price data (prices for the last and third week of each month) to check robustness of our model, but also to gain additional insights and discover patterns that might have been hidden by smoothing the time series via averaging and potentially reducing the signal-to-noise ratio. What we call our first, second, and third set of estimates correspond to estimates based on the last week of each month's price, the third week's price and the monthly average, respectively. Complete results showing estimates for all selected covariates and the three different ways to aggregate price data are gathered in the appendix.

Clearly, the boosting algorithm repeatedly selects certain variables as relevant into the model. This becomes visible when glancing over Table 2 to Table 7 and focusing on blank spaces versus those filled with numbers. These variables are related to inventories, foreign exchange rates, the VIX, financialization, oil prices, weather and time of the year. We display selected estimates in Figure 3 to Figure 5, and we discuss them in the following section and broadly summarize findings in Table 1. This table is an attempt, for each commodity, to combine findings from different quotations and sets of price data into a group of indicators and corresponding critical values. Admittedly, this is a somewhat subjective exercise, which does not follow strictly defined criteria, but takes into account differential reliability of estimates and the like.

Among the stocks-to-use (stocks-to-disappearance) class of variables we find the global stocks-to-use ratio as well as that for the United States to be relevant for all three commodities and, in addition, for wheat and maize the stocks-to-disappearance ratio for major exporters. For wheat, global inventories below 18 percent relative to use prompt higher volatility. This increase can amount to more than 30 percent (as is the case for Canadian wheat at the port of St Lawrence, second set of estimates). The same holds when the US stocks-to-use ratio falls below 37 percent. The boosting algorithm selects this variable and threshold even more consistently, for around half of the quotations in each of the three data sets. For the majority of quotations, falling below this threshold is associated with an increase in volatility between five and 20 percent. Major exporter's stocks-to-disappearance ratio also features as one of the variables prompting elevated wheat price volatility. The critical change point is at around 13 percent. For maize, these three ratios also appear to be decisive. Estimated thresholds do not differ much between quotations – critical ratios are around 17 percent for the world's stocks-to-use ratio; 12 percent for the US stocks-to-use ratio; and 12 percent for major exporters stocks-to-disappearance ratio. A ratio below these thresholds is associated with volatility shooting up by up to 20 percent, but nine percent on average. Compared with the other two commodities, thetransition between different levels of volatility looks less abrupt for soybeans, especially when examining estimates for the first data set. Several smaller steps are estimated instead of a single step of larger size. The largest, and thus critical, estimated steps coincide to be around 12 percent for the US and 22 percent for the world stocks-to-use ratio across quotations. The estimated impact of falling below these ratios is an increase in price volatility of up to 25 percent and, similar to maize, it is around ten percent on average. The wide interval on which the thresholds rest, by commodity and by exporter, might be explained by the importance of trade to production, in that a higher share of exported production yields higher thresholds and also when the particular exporting country is dominant in the global marketplace.

The foreign exchange value of the USD is also crucial for wheat price volatility. Volatility soars when the FX index falls below 101. This does not necessarily happen immediately – the impact is most pronounced after a lag of two months. For example, for the IGC wheat index, the impact of an FX index smaller than 101 is estimated as an 19 to 35 percent increase in volatility for the different quotations. In addition to the level of the FX index, rising FX index volatility is passed on to commodity price volatility. For wheat conditional volatility of the FX index above 1.3 percent and realized FX index volatility above 1.5 percent appear to be critical. While the value of the USD does not seem to influence maize or soybean price volatility, we observe an increase in maize price fluctuations when the realized volatility of the FX index raises above 1.3 percent.

Our estimation results suggest the VIX to be associated with changes in volatility levels, at least for wheat. Both the VIX and its first lag are selected for various quotations and data sets. Lee and Han's (2015) recent paper corroborates this link between perceived overall stock market risk and uncertainty and agricultural price volatility. Lee and Han (2015) propose and extract a common stochastic volatility factor for energy and commodity prices (namely, for oil, wheat, and maize prices), which they find to be strongly correlated with the VIX. Whereas we find some evidence for a VIX-related

Table 1: Critical thresholds by commodity

	Wheat	Maize	Soybeans
Global stocks-to-use ratio (percent)	18	17	22
US stocks-to-use ratio (percent)	37	12	12
Major exporters stocks-to-disappearance ratio (percent)	13	12	
Foreign Exchange Index	101		
FX Index conditional volatility (percent)	1.3		
FX Index realized volatility (percent)	1.5	1.3	
VIX	23		
Brent oil price (USD per Barrel), third lag	118	105	119
WTI oil price (USD per Barrel)		105	104
Volume (thousand)	2700		
Open Interest (thousand)	485		
Volume-to-open-interest ratio		4.3	8.6
Non-commercial long positions (thousand)	40		
Sea surface temperature anomalies in Niño 3.4 region	-1.0		
Oceanic Niño Index			0.6
Month		June/July	

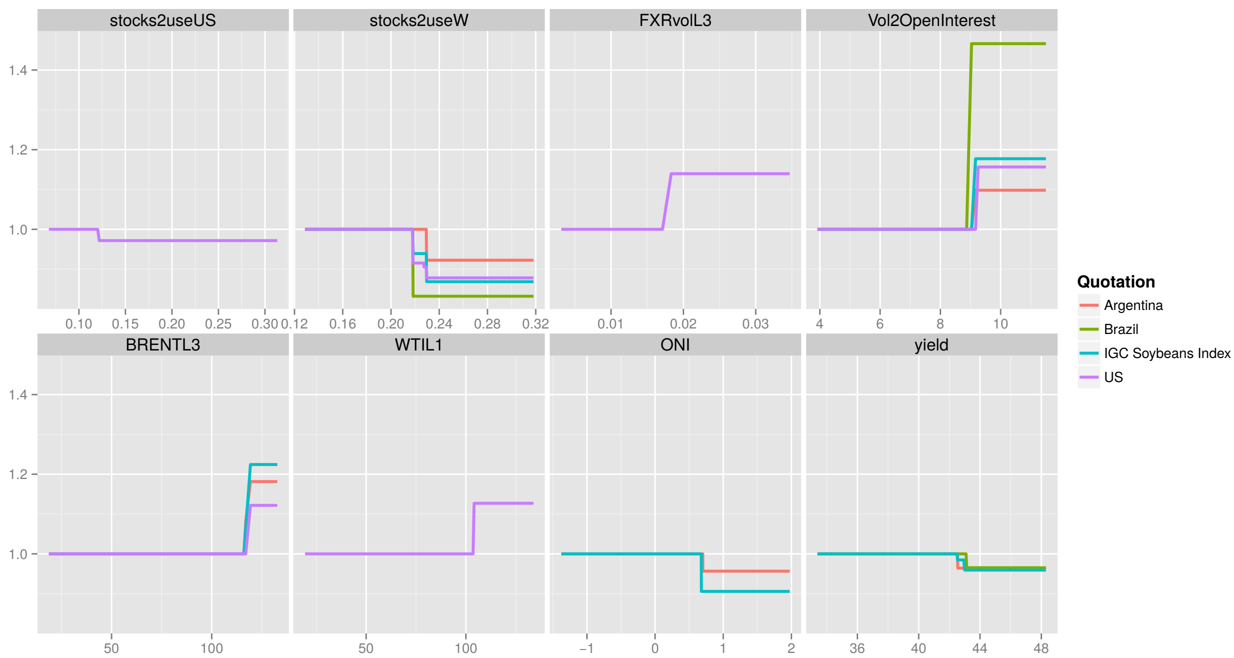
change point for maize in the first data set, it is weak in view of inconsistent and counter-intuitive estimates for the other two data sets. Similarly, the VIX does not seem to be associated with patterns in soybean price volatility.

Oil prices come out relevant for wheat, maize, and soybeans, although more so for the latter two commodities. For Brent oil, critical points are located at a price of 118 USD per barrel for wheat, 105 USD per barrel for maize, and 119 USD per barrel for soybeans, at a three-months lag. The estimated impact of a rise in oil price beyond this limit is explosive in some cases – for US PNW No.2 SW wheat it amounts to a rise in volatility by more than 60 percent. While this value might be overestimated, the Brent oil price features as an important determinant of wheat price volatility for all three sets of pricedata. We do not find the WTI oil price among the covariates relevant for wheat price volatility, but it appears to be associated with price volatility for the remaining two food crops. Critical thresholds are estimated at 105 USD per barrel for maize, and 104 USD per barrel for soybeans.

Variables indicating the degree of financialization turn out to be relevant for all three commodities. Looking at volumes for wheat, a critical threshold appears to be between 2300000 and 2900000 trades per month. Regarding open interest, surpassing 465000 to 495000 contracts entered into but yet to be offset corresponds to a rise in wheat price volatility. Instead of either volume or open interest separately, a growing ratio of the one to the other is correlated with an increase in volatility for maize and soybeans. While the change in volatility associated with the volume-to-open-interest ratio can be gradual – especially for soybeans – the transition between volatility levels seems to take place around a quotient of four for maize and eight for soybeans. As pointed out before, our analysis does not allow to distinguish which way the causality runs. The entry of large numbers of non-commercial market participants might cause price fluctuations to amplify. But what if it is increased price volatility itself that attracts speculators looking for opportunities to take risks to get positions in the market? Causation might even work both ways, reinforcing itself in a spiral of higher volatility gaining the attention of non- commercial traders and the increase in volume in turn resulting in yet higher volatility. While this uncertainty might not permit to draw certain policy implications, we can use the association between volatility levels and certain measures of financialization as an early warning indicator regardless of the direction of causality. With the expectation of high volatility attracting non-commercial market participants, volume and related variables can be understood as a forward-looking measure of volatility, similar to implied volatility, regardless of their own causal impact.

The boosting algorithm selects variables related to El Niño and La Niña, mainly for wheat and soybeans. It is very hard to make a general statement on the influence

Figure 6: Estimated impact of lagged (by one month) returns on wheat price volatility (monthly averages)



of these phenomena on agriculture. Effects vary between regions and further depend on the exact moment in the growing phase when changed weather conditions hit (Baffes, 2015). As a very rough rule, in the northern hemisphere El Niño means more moisture, hence, higher yields. Assuming a positive association between levels of prices and their volatility (Prakash, 2011), this implies lower volatility than usual during phases of El Niño. Indeed, our estimates show a rise in wheat price volatility associated with episodes of moderate or strong La Niña. Weather-related variables seem to be less important for maize and soybeans than for wheat. This might be linked to the fact that the former are considered irrigated crops in the main. Still, Brazilian maize and US soybean price volatility goes down whenever the Oceanic Niño Index passes a 0.2 and 0.6 threshold, respectively, which is a signal of El Niño. In case of maize in Brazil, at least for the southern region of Rio Grande do Sul, this is in line with Berlato and Fontana's (2001) observation that while La Niña causes harvest losses, non-irrigated spring-summer crops benefit from El Niño. Instead of trying to trace weather effects on volatility back to its differential impacts on yields, one can alternatively make sense of the estimation results by recalling that during phases of El Niño it is fairly clear what weather to expect. Lower uncertainty concerning yields in turn is reflected in reduced fluctuations in prices.

In addition to the factors discussed so far, we observe a seasonal effect. An upswing in volatility is estimated for the second half of the year. It is particularly pronounced for maize. This could be related to USDA's switch in projections to a new marketing year (detailed in section 3) in its WASDE.

Additional remarks

It is interesting to observe that the estimated effect of lagged negative returns on volatility differs from that of positive returns (Figure 6). Part of the rationale for Nelson (1991) to introduce the exponential ARCH model was to allow for exactly this differential effect; in the conventional ARCH model, only the magnitude but not the sign of lagged residuals drives volatility. Contrary to the evidence of a negative association between stock returns and volatility drawn on by Nelson (1991), of volatility rising with bad and falling with good news (measured as excess returns lower or higher than expected), we find volatility to increase when returns are unusually high, i.e. prices are increasing.

Reflecting on our estimation results, it becomes clear that among the variables the boosting algorithm chooses as model constituents pertinent to volatility, most are factors not leading to immediate policy implications. For model components such as the stocks-to-use ratio we can derive the recommendation to design policies such as to keep global wheat inventories above a threshold of 18 percent relative to use. However, it is not feasible to shape variables such as the Oceanic Niño Index, VIX, oil price or foreign exchange rate through policies. (Of course, it is possibly to try and control some of these variables to a certain extend. However, the consequences of influencing e.g. exchange rates would reach so far beyond agricultural commodity price volatility that it would never occur to anyone to do so with this aim in mind. Thus, it is not realizable in practice.) Whereas it is crucial to have just one set of indicators per commodity to guide policy-making, even for variables impossible to impact such a compilation is helpful to get a sense of potential future volatility levels. Still, when the focus shifts from what influences volatility to anticipating volatility ahead, a question arises on whether to combine estimates for different quotations into a single set of variables and critical thresholds (as was done in the previous sub-section) would constitute the best possible approach. Another question arises on whether to use the estimated $\exp(\hat{\eta}_t/2)$ for each quotation instead – leaving out terms involving lagged returns while using expected future values for the remaining ex-planatory variables – and then generate projected volatility levels for different scenarios? While this strategy has the advantage of readily providing a range of volatility projections per price quote, we believe that a more complete picture evolves when considering estimation results jointly for reasons detailed above (namely, transmission of price shocks across locations and complex mechanisms potentially hidden for some of the price series).

Notably, one important variable that would have immediate policy implications is missing in our study – trade policies. These can have a substantial impact on price movements. A price hike following a shortage of supply on the world market can be exacerbated if a major exporter decides to put in place an export ban in order to secure sufficient domestic supplies. The literature provides considerable evidence that insulating trade policies amplify volatility (Martin and Anderson, 2012; Ivanic and Martin, 2014). Martin and Anderson (2012) attribute almost 30 percent of the observed change in the international price of wheat between 2005 and 2008 to modifications in border protection rates. Unfortunately, it is an intricate task to assign numerical values to these policies. First, there is not just one type of trade-related policy measure but various, whose relative importance is unclear. Second, their significance has to be seen in the context of the importance of the implementing country for trade in the respective commodity. Third, the same measure should have very different domestic and foreign effects. For example, an export tax on wheat implemented by Argentina should have a different effect on Argentinian wheat prices than one applied by Russia. A different, and probably even more serious concern is that in case of an export ban, export prices are typically not available for the duration of the policy. Russia prohibited wheat exports between August 2010 and June 2011; we do not have wheat prices for the Black Sea between October 2010 and May 2011. How can we possibly identify the impact of a variable during a specific period when it is for this exact period that data are missing?

Conclusions

Our statistical assessment of volatility patterns in food crops has drawn attention towards a selection of factors and associated critical thresholds that appear key when trying to anticipate future price variability. The main findings of this paper – critical thresholds above which price volatility rises or falls abruptly – are condensed in Table 1.

Clearly though, we have to understand the relative importance of variables in the context of our modelling framework. There is, of course, a rationale to believe that the model set-up is appropriate and is probably among the best suitable approaches to address our question under given constraints, such as data availability. Nevertheless, while we keep its structure very flexible, our model might still oversimplify an immensely complex interplay of different factors that drives volatility. Variables that are either missing, or have been quantified inadequately or observed at a wrong time scale or frequency, cannot be detected by the model. For example, we do not include a measure for restrictive trade policies. Consequently, although well established, it is impossible to reflect the impact of policy on commodity price volatility within our modelling framework.

In addition, it would be tempting to infer causation from the results, that is, to interpret variables selected by the boosting algorithm as relevant for price volatility as drivers of volatility. While the channels through which certain variables affect price fluctuations are particularly well understood, for other variables, such as those related to financialization, there is less congruence. Finding correlation between an increase in price volatility and the number of non-commercial players entering the market, one might conclude that activities of these market participants drive price formation that result in higher volatility. However, correlation does not imply causation. Correlation could indicate causation, or it could be mere coincidence or even the causal chain might even run the other way. While comprehension of causal mechanisms is, of course, ultimately desirable, for the purpose of anticipating changes in volatility mere correlation is also useful.

Still, awareness of the limitations of the study certainly does not refrain from the recommendation to be diligent when certain variables approach specific thresholds. It is insightful that some of these variables turn out relevant when lagged. This might be hinting at causality instead of mere association. More importantly, it allows us to have an indication of possibly elevated future volatility in advance.

Naturally, our analysis has also generated further research questions and possible refinements to the model set-up. This includes a review of the frequency that data are modelled. Would an analysis based on a daily frequency be able to provide additional insights? Would a higher frequency mirror developments in market more appropriately? Do results change under temporal aggregation? Further, although it might be difficult to formulate, a variable quantifying trade policy measures would be a valuable addition to the model. Finally, understanding if threshold levels have evolved over time would be important to address, that is, if they are subject to trend. For example, is it that volatility is triggered with a lower stock-to-use ratio than was the case in the past? Has the influence of oil prices grown with the emergence of biofuels? Are exchange rates more influential with expanding trade? While beyond the scope of this paper, these questions warrant further enquiry.

References

Algieri, B. (2014). A roller coaster ride: An empirical investigation of the main drivers of the international wheat price. *Agricultural Economics*, 45(4):459–475.

Baffes, J. (2015). Understanding El Niño: What does it mean for commodity markets? In *Commodity Market Outlook*. The World Bank, Washington, DC.

Balcombe, K. (2011). The nature and determinants of volatility in agricultural prices: An empirical study. In Prakash, A., editor, *Safeguarding food security in volatile global markets*. Food and Agriculture Organization of the United Nations, Rome, Italy.

Bastianin, A., Manera, M., Nicolini, M., and Vignati, I. (2012). Speculation, returns, volume and volatility in commodities futures markets. *Review of Environment, Energy and Economics*.

Berlato, M. A. and Fontana, D. C. (2001). Impacts of El Niño and La Niña on agricultural production in southern Brazil and the use of climate forecasts in agriculture. In Cunha, G. R., Haas, J. C., and Berlato, M. A., editors, *Applications of climate forecasting for better decisionmaking processes in agriculture*. Embrapa Trigo, Passo Fundo, RS, Brazil.

Bobenrieth, E., Wright, B., and Zeng, D. (2013). Stocks-to-use ratios and prices as indicators of vulnerability to spikes in global cereal markets. *Agricultural Economics*, 44(1):43–52.

Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, 16(3):199– 231.

Buehlmann, P. and Hothorn, T. (2007). Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science*, 22(4):477–505.

CBOE (2009). VIX – fact & fiction. Research note, Chicago Board Options Exchange.

CBOE (2015). The CBOE volatility index – VIX. White paper, Chicago Board Options Exchange.

CFTC (2016). Explanatory notes to the Commitment of Traders Report. http://www.cftc.gov/MarketReports/CommitmentsofTraders/ExplanatoryNotes.

Deaton, A. and Laroque, G. (1992). On the behaviour of commodity prices. *The Review of Economic Studies*, 59(1):1–23.

Domanski, D. and Heath, A. (2007). Financial investors and commodity markets. In *BIS Quarterly Review*. Bank for International Settlements, Basel, Switzerland.

Etienne, X. L., Irwin, S. H., and Garcia, P. (2015). New evidence that index traders did not drive large bubbles in grain futures markets. Discussion paper.

Fackler, P. L. and Goodwin, B. K. (2001). Spatial price analysis. In Gardner, B. and Rausser, G., editors, *Handbook of Agricultural Economics*, volume 1B, pages 971–1024. Elsevier.

Frankel, J. A. (2008). The effect of monetary policy on real commodity prices. In Campbell, J. Y., editor, *Asset Prices and Monetary Policy*, pages 291–333. University of Chicago Press, Chicago.

Greb, F. and Prakash, A. (2015). Has price volatility changed? In *Food Outlook*. Food and Agriculture Organization of the United Nations, Rome, Italy.

Hansen, P. R. and Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH (1, 1)? *Journal of Applied Econometrics*, 20(7):873–889.

Hothorn, T., Hornik, K., and Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, 15(3):651–674.

IFAD (2016). Food prices: smallholder farmers can be part of the solution. https://www.ifad.org/what/operating model/tags/1963528.

Ivanic, M. and Martin, W. (2014). Implications of domestic price insulation for global food price behavior. *Journal of International Money and Finance*, 42:272–288.

Lagi, M., Bar-Yam, Y., Bertrand, K. Z., and Bar-Yam, Y. (2015). Accurate market price formation model with both supply-demand and trend-following for global food prices providing policy recommendations. *Proceedings of the National Academy of Sciences*, 112(45):E6119–E6128.

Lee, E. and Han, D. B. (2015). A common factor of stochastic volatilities between oil and commodity prices. Discussion paper.

Loretan, M. (2005). Indexes of the foriegn exchange value of the dollar. *Federal Reserve Bulletin*, 91:1–8.

Martin, W. and Anderson, K. (2012). Export restrictions and price insulation during commodity price booms. *American Journal of Agricultural Economics*, 94(2):422–427.

McPhail, L. L., Du, X., and Muhammad, A. (2012). Disentangling corn price volatility: The role of global demand, speculation, and energy. *Journal of Agricultural and Applied Economics*, 44(03):401–410.

Micu, M. (2005). Declining risk premia in the crude oil futures market. In *BIS Quarterly Review*. Bank for International Settlements, Basel, Switzerland.

Mittnik, S., Robinzonov, N., and Spindler, M. (2015). Stock market volatility: Identifying major drivers and the nature of their impact. *Journal of Banking & Finance*, 58:1–14.

Nelson, D. (1991). Conditional heteroskedasticity in asset returns: a new approach.*Econometrica*, 59(2):347–370.

Pindyck, R. S. (2001). The dynamics of commodity spot and futures markets: A primer. *Energy Journal*, 22(3).

Prakash, A. (2011). Why volatility matters. In Prakash, A., editor, *Safeguarding food security in volatile global markets*. Food and Agriculture Organization of the United Nations, Rome, Italy.

Robles, M., Torero, M., and Von Braun, J. (2009). *When speculation matters*. International Food Policy Research Institute, Washington, DC.

Rojas, O. (2015). *Protocol for Country-Level ASIS calibration and national adaptation*. Food and Agriculture Organization of the United Nations, Panama City, Panama.

Rojas, O., Li, Y., and Cumani, R. (2014). *Understanding the Drought Impact of El Niño on the Global Agricultural Areas: An assessment using FAO's Agricultural Stress Index (ASI)*. Food and Agriculture Organization of the United Nations, Rome, Italy.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3):289–310.

Ubilava, D. (2014). The ENSO effect on world wheat market dynamics: Smooth transitions in asymmetric price transmission. In *2014 Annual Meeting, July 27-29, 2014, Minneapolis, Minnesota*. Agricultural and Applied Economics Association.

Valiante, D. (2013). Commodities price formation: Financialisation and beyond. CEPS-ECMI Task Force Report, Centre for European Policy Studies, Brussels.

Will, M. G., Prehn, S., Pies, I., and Glauben, T. (2012). Is financial speculation with agricultural commodities harmful or helpful? A literature review of current empirical research. Discussion Paper 2012-27, Chair of Economic Ethics, Martin-Luther-University Halle-Wittenberg.

Wright, B. D. (2011). The economics of grain price volatility. *Applied Economic Perspectives and Policy*, 33(1):32–58.

Appendix

While Figure 3 to 5 and Table 1 display selected results only, Figure 7 to Figure 15 together with Table 2 to Table 7 show the complete set of estimates. Empty cells indicate that the algorithm has not selected this particular variable. Entries show critical thresholds. Clearly, a tree is not limited to one threshold, but can have several (see, for example, the case of the volume-to-open-interest ratio for soybeans in the second panel, third row of Figure 14). For ease of presentation we only report one threshold per variable in the tables, the one making most difference in terms of changes in volatility, e.g. for the IGC soybean index' volume-to-open-interest ratio we list 9.2 as a ratio above this value prompt an rise in volatility by almost 20 percent, about twice as much as the other thresholds estimated.

Table 2: Estimated critical thresholds for wheat index and quotations (last week's value of each month). **L1**, **L2** or **L3** added to the variable name indicates the first, second or third lag. **Rvol** and **Cvol** specify realized and conditional volatility of the variables

	Argentina Up River Trigo Pan	Australia Adelaide ASW	Black Sea Milling Wheat	Canada No. 1 CWRS 13.5%		EU Rouen St Grade	US Gulf No. 2		US PNW No. 2		IGC
				St Lawrence	Vancouver		HRW	SRW	DNS	SW	
year					2007						
month		6									
thinness1											
thinness2											
Herfindahl											
stocks2useW		0.28		0.18							
stocks2useUS	0.80			0.36	0.37			0.42	0.36	0.42	0.37
stocks2disMEx											
stocks2disMExUS				0.12	0.13				0.12	0.13	
yield											
FEDFUNDS											
FEDFUNDSL1											
FEDFUNDSL2											
FEDFUNDSL3											
TB6MS											
TB6MSL1											
TB6MSL2											
TB6MSL3				0.1							
BRENT					60						
BRENTL1											
BRENTL2											
BRENTL3	117									117	
BRENTCvol		0.11									
BRENTRvol											
WTI					66						
WTIL1	106										
WTIL2											
WTIL3											
WTICvol						0.08					
WTIRvol											
FX									97		
FXL1			101				101				101
FXL2		101	97		100	101	101	101	97	101	101
FXL3	105						105	104		117	
FXCvol	0.015	0.015		0.015	0.015				0.015	0.015	
FXCvolL1	0.018	0.018	0.018			0.018	0.018	0.017		0.018	0.018
FXCvolL2	0.019		0.015		0.015	0.015		0.015			0.015
FXCvolL3					0.015						
FXRvol											
FXRvolL1	0.020					0.011	0.011	0.010	0.010		0.011
FXRvolL2	0.017										
FXRvolL3	0.009										
VIX	23	23	23								
VIXL1	25		27	25	23		25				
VIXL2											
VIXL3											
Volume (thousand)				2682	2670	1364	2343			2753	2343
OpenInterest (thousand)		490	467			490	490				490
Vol2OpenInterest											
S2TotalL							0.53		0.55	0.54	
S2TotalS											
netL (thousand)				35	35	79			35		
Working											
ASI											
ONI											
NINO34		-1.2	-0.2	-1.1	-1.1		-1.2		-1.1	-1.1	-1.2
SOI	1.3		1.0								

Table 3: Estimated critical thresholds for maize and soybean index and quotations (last week's value of each month). **L1**, **L2** or **L3** added to the variable name indicates the first, second or third lag. **Rvol** and **Cvol** specify realized and conditional volatility of the variables

	Maize				Soybeans			
	Argentina Up River	Brazil Paranagua	US Gulf No. 3 Yellow	IGC	Argentina Up River	Brazil Paranagua	US Gulf No. 2 Yellow	IGC
year								
month			6					
thinness1								
thinness2								
Herfindahl								
stocks2useW			0.17	0.17	0.22	0.22		
stocks2useUS	0.12		0.12	0.12		0.16	0.12	0.08
stocks2disMEx								
stocks2disMExUS		0.12	0.12	0.12				
yield						43		
FEDFUNDS								
FEDFUNDSL1								
FEDFUNDSL2								
FEDFUNDSL3								
TB6MS								
TB6MSL1								
TB6MSL2								
TB6MSL3		0.4						
BRENT								
BRENTL1								
BRENTL2			116			43		
BRENTL3	72			72				
BRENTCvol								
BRENTRvol								
WTI								
WTIL1								
WTIL2			109					
WTIL3	105	105	105	105				
WTICvol				0.08				
WTIRvol					0.08			
FX								
FXL1								118
FXL2							118	
FXL3	104		104	104				
FXCvol	0.012						0.001	
FXCvolL1								
FXCvolL2	0.011	0.012	0.011	0.011			0.004	0.003
FXCvolL3								
FXRvol					0.014	0.015		0.016
FXRvolL1		0.009						
FXRvolL2	0.013		0.013	0.013				
FXRvolL3								
VIX		29	41					
VIXL1								
VIXL2					15	15		
VIXL3								
Volume (thousand)								
OpenInterest (thousand)					742			
Vol2OpenInterest					7.4	7.4	8.9	7.4
S2TotalL								
S2TotalS	0.37		0.30	0.37		0.31		
netL (thousand)								
Working								
ASI					10			
ONI		0.2					0.4	0.6
NINO34					0.7			
SOI								

Table 4: Estimated critical thresholds for wheat index and quotations (third week's value of each month). **L1**, **L2** or **L3** added to the variable name indicates the first, second or third lag. **Rvol** and **Cvol** specify realized and conditional volatility of the variables

	Argentina Up River Trigo Pan	Australia Adelaide ASW	Black Sea Milling Wheat	Canada No. 1 CWRS 13.5%		EU Rouen St Grade	US Gulf No. 2		US PNW No. 2		IGC
				St Lawrence	Vancouver		HRW	SRW	DNS	SW	
year											
month									7		
thinness1					0.178						
thinness2	0.171						0.170				0.170
Herfindahl											
stocks2useW				0.18	0.18				0.18		
stocks2useUS	0.80	0.36		0.36				0.37	0.36	0.37	0.37
stocks2disMEx				0.12	0.12	0.10				0.12	
stocks2disMExUS					0.13				0.12		
yield											
FEDFUNDS											
FEDFUNDSL1											
FEDFUNDSL2											
FEDFUNDSL3											
TB6MS					2.0						
TB6MSL1											
TB6MSL2											
TB6MSL3											
BRENT					70						
BRENTL1			65								
BRENTL2						66		65			
BRENTL3		116	117			118	118	118		118	118
BRENTCvol											
BRENTRvol			0.08								
WTI											
WTIL1							64				
WTIL2											
WTIL3	104		108			108					
WTICvol											
WTIRvol											
FX								102	97		
FXL1							102				
FXL2	101		101	99		101			99	101	101
FXL3			101					107			
FXCvol											
FXCvolL1		0.008	0.013				0.007	0.007	0.006	0.007	0.007
FXCvolL2								0.003			
FXCvolL3				0.012		0.003	0.003				
FXRvol	0.011										
FXRvolL1	0.017		0.017								
FXRvolL2						0.016					
FXRvolL3		0.017									
VIX								16			
VIXL1	24		20	24	24						
VIXL2											
VIXL3											
Volume (thousand)		1919			2769						2962
OpenInterest (thousand)	495										
Vol2OpenInterest											
S2TotalL				0.53							
S2TotalS	0.36	0.41	0.43	0.24			0.41	0.41			0.41
netL (thousand)				36	34				36		
Working											
ASI											
ONI			-0.0								
NINO34	-1.0	-0.9			-1.6		-1.0		-1.0	-1.0	-1.0
SOI			-0.1	2.2		1.8					1.8

Table 5: Estimated critical thresholds for maize and soybean index and quotations (third week's value of each month). **L1**, **L2** or **L3** added to the variable name indicates the first, second or third lag. **Rvol** and **Cvol** specify realized and conditional volatility of the variables

	Maize				Soybeans			
	Argentina Up River	Brazil Paranagua	US Gulf No. 3 Yellow	IGC	Argentina Up River	Brazil Paranagua	US Gulf No. 2 Yellow	IGC
year								
month			7	6				
thinness1								
thinness2	0.117				0.360			
Herfindahl								
stocks2useW	0.16		0.16	0.16	0.23	0.22		0.23
stocks2useUS							0.12	
stocks2disMEx								
stocks2disMExUS	0.11			0.11				
yield						41	43	
FEDFUNDS								
FEDFUNDSL1		0.2			0.2			
FEDFUNDSL2								
FEDFUNDSL3								
TB6MS								
TB6MSL1								
TB6MSL2								
TB6MSL3								
BRENT								
BRENTL1	56							
BRENTL2								
BRENTL3		117			118	118	118	118
BRENTCvol					0.09		0.09	0.09
BRENTRvol					0.15			
WTI	108		108	108				
WTIL1			57	57		39	108	
WTIL2								
WTIL3	103		103	103		103	103	
WTICvol								
WTIRvol								
FX								
FXL1								
FXL2								
FXL3								
FXCvol								
FXCvolL1								
FXCvolL2								
FXCvolL3								
FXRvol								
FXRvolL1		0.011						
FXRvolL2	0.004	0.014	0.012					
FXRvolL3				0.004			0.017	
VIX	11			11				
VIXL1								
VIXL2								
VIXL3								
Volume (thousand)			1654					
OpenInterest (thousand)								
Vol2OpenInterest		5.8	4.6	3.7	8.6	8.6	9.2	9.2
S2TotalL								
S2TotalS	0.33			0.33				
netL (thousand)								
Working								
ASI								
ONI		0.2						
NINO34								
SOI								-0.0

Table 6: Estimated critical thresholds for wheat index and quotations (monthly average). **L1**, **L2** or **L3** added to the variable name indicates the first, second or third lag. **Rvol** and **Cvol** specify realized and conditional volatility of the variables

	Argentina	Australia	Black Sea	Canada		EU	US		US		IGC
	Up River	Adelaide		St Lawrence	Vancouver	Rouen	Gulf		PNW		
	Trigo Pan	ASW	Milling	No. 1		St Grade	No. 2		No. 2		
			Wheat	CWRS 13.5%			HRW	SRW	DNS	SW	
year											
month									7		
thinness1											
thinness2											
Herfindahl											
stocks2useW		0.27									
stocks2useUS	0.81	0.37		0.37	0.37	0.81		0.39	0.37	0.39	0.37
stocks2disMEx					0.11						
stocks2disMExUS	0.14			0.12	0.12	0.18			0.13		
yield											
FEDFUNDS											
FEDFUNDSL1											
FEDFUNDSL2											
FEDFUNDSL3											
TB6MS					2.1						
TB6MSL1											
TB6MSL2											
TB6MSL3											
BRENT					64						
BRENTL1							62				
BRENTL2											
BRENTL3	117		62			119	119	119	119	117	119
BRENTCvol											
BRENTRvol											
WTI											
WTIL1	104										
WTIL2											
WTIL3	102		107	106							
WTICvol											
WTIRvol											
FX											
FXL1							101	101	101		101
FXL2	102		101	98	98	101	101	101	98	101	101
FXL3	104		97								
FXCvol	0.010				0.013	0.014		0.013			
FXCvolL1	0.014	0.013					0.014	0.010		0.014	
FXCvolL2											
FXCvolL3		0.011						0.011			
FXRvol											
FXRvolL1	0.021		0.015						0.011		
FXRvolL2			0.021			0.015					
FXRvolL3		0.014									
VIX							17			18	
VIXL1	25	19	19	25	25		25				25
VIXL2											
VIXL3											
Volume (thousand)										2962	2962
OpenInterest (thousand)											
Vol2OpenInterest											
S2TotalL				0.53							
S2TotalS				0.25			0.41	0.41			
netL (thousand)				36	36						
Working											
ASI											
ONI			0.2			0.3					
NINO34		-1.0	-0.3		-1.6		-0.9		-1.0	-1.0	
SOI				1.6	1.6		1.6		1.9		1.6

Table 7: Estimated critical thresholds for maize and soybean index and quotations (monthly average). **L1**, **L2** or **L3** added to the variable name indicates the first, second or third lag. **Rvol** and **Cvol** specify realized and conditional volatility of the variables

	Maize				Soybeans			
	Argentina Up River	Brazil Paranagua	US Gulf No. 3 Yellow	IGC	Argentina Up River	Brazil Paranagua	US Gulf No. 2 Yellow	IGC
year								
month			7	7			7	
thinness1								
thinness2					0.333			
Herfindahl								0.294
stocks2useW					0.23	0.22	0.22	0.23
stocks2useUS	0.12		0.12	0.12			0.12	
stocks2disMEx								
stocks2disMExUS								
yield					43	43		43
FEDFUNDS								
FEDFUNDSL1		0.2			0.1			0.1
FEDFUNDSL2								
FEDFUNDSL3								
TB6MS								
TB6MSL1								
TB6MSL2								
TB6MSL3		0.3						
BRENT								
BRENTL1								
BRENTL2								
BRENTL3	119	116	119	119	119		119	119
BRENTCvol					0.08			
BRENTRvol					0.11			
WTI			107	107				
WTIL1			104	104			104	
WTIL2	107			107	107			107
WTIL3	103		103			103		
WTICvol								
WTIRvol		0.06			0.09			
FX								
FXL1								
FXL2								
FXL3								
FXCvol			0.010					
FXCvolL1	0.010		0.010					
FXCvolL2	0.010							
FXCvolL3	0.009					0.009		
FXRvol								
FXRvolL1		0.016	0.016	0.016				
FXRvolL2								
FXRvolL3		0.012					0.018	
VIX	12			12		18		
VIXL1								
VIXL2								
VIXL3								
Volume (thousand)			1654	1584				
OpenInterest (thousand)		1184						
Vol2OpenInterest	3.7		4.0	4.0	9.2	9.0	9.3	9.2
S2TotalL								
S2TotalS				0.32				
netL (thousand)								
Working								
ASI								
ONI					0.7			0.7
NINO34								
SOI				1.9				

Figure 7: Estimates for wheat index and prices (last week's value of each month)

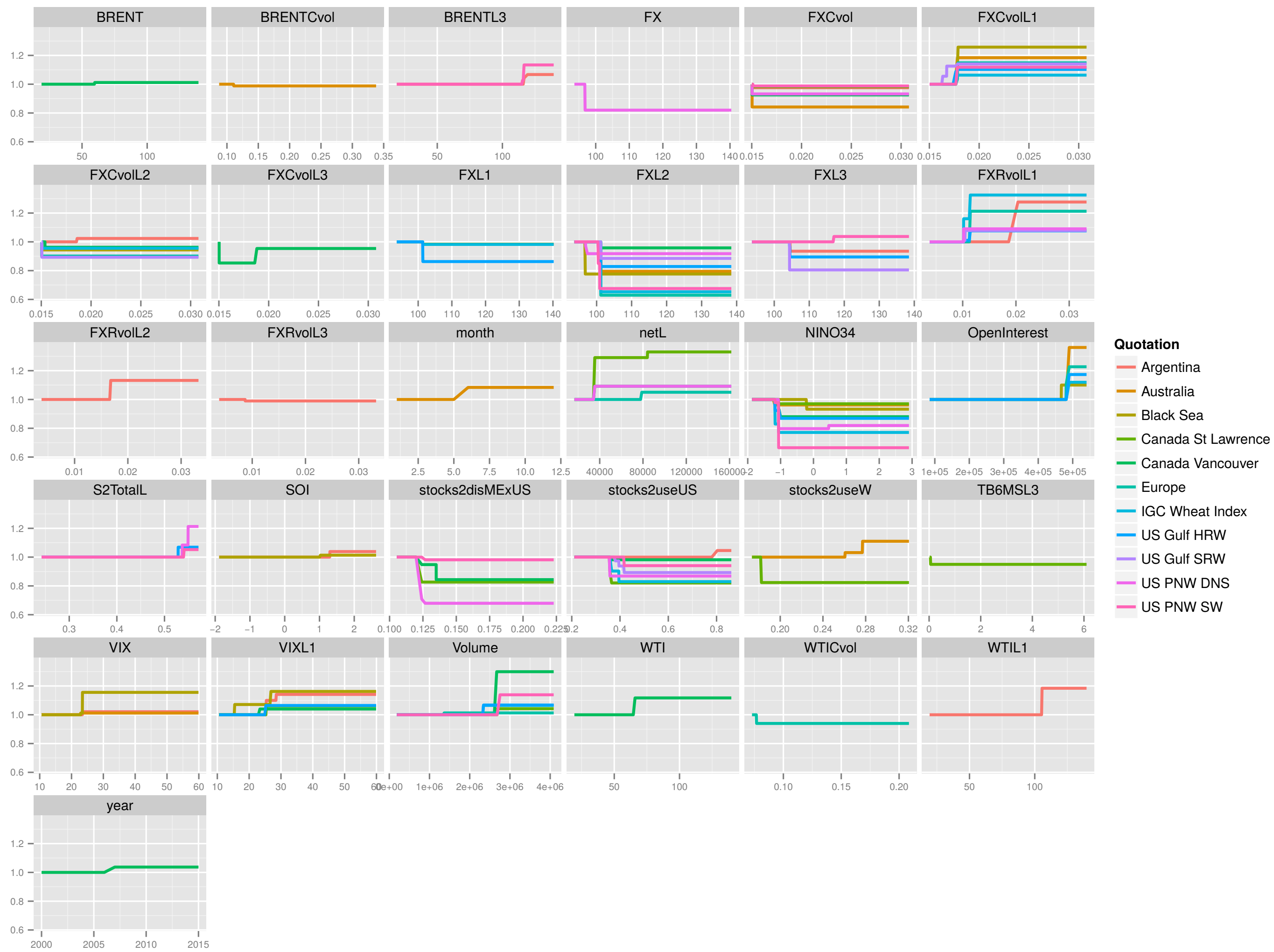


Figure 8: Estimates for wheat index and prices (third week's value of each month)

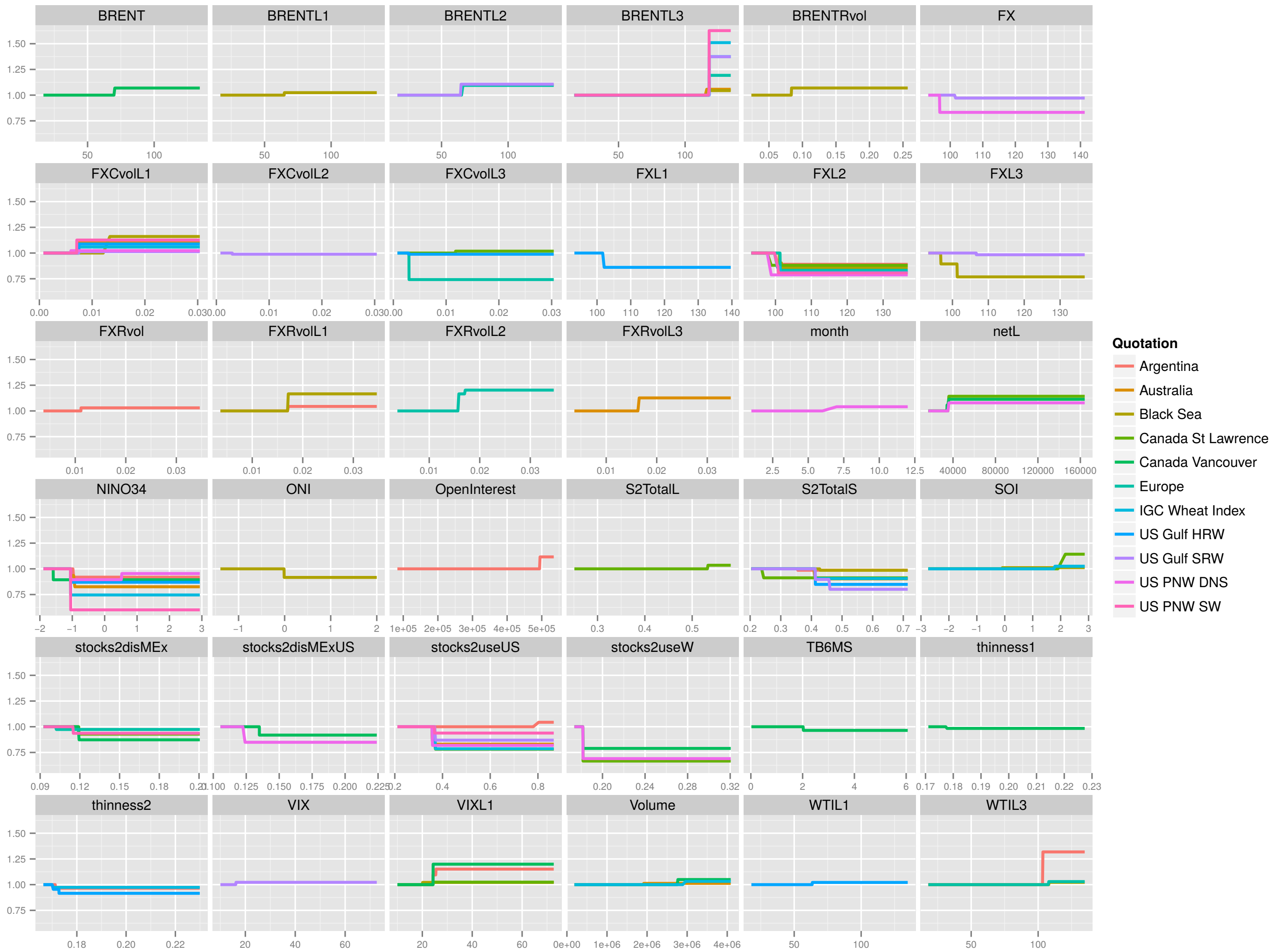


Figure 9: Estimates for wheat index and prices (monthly average)

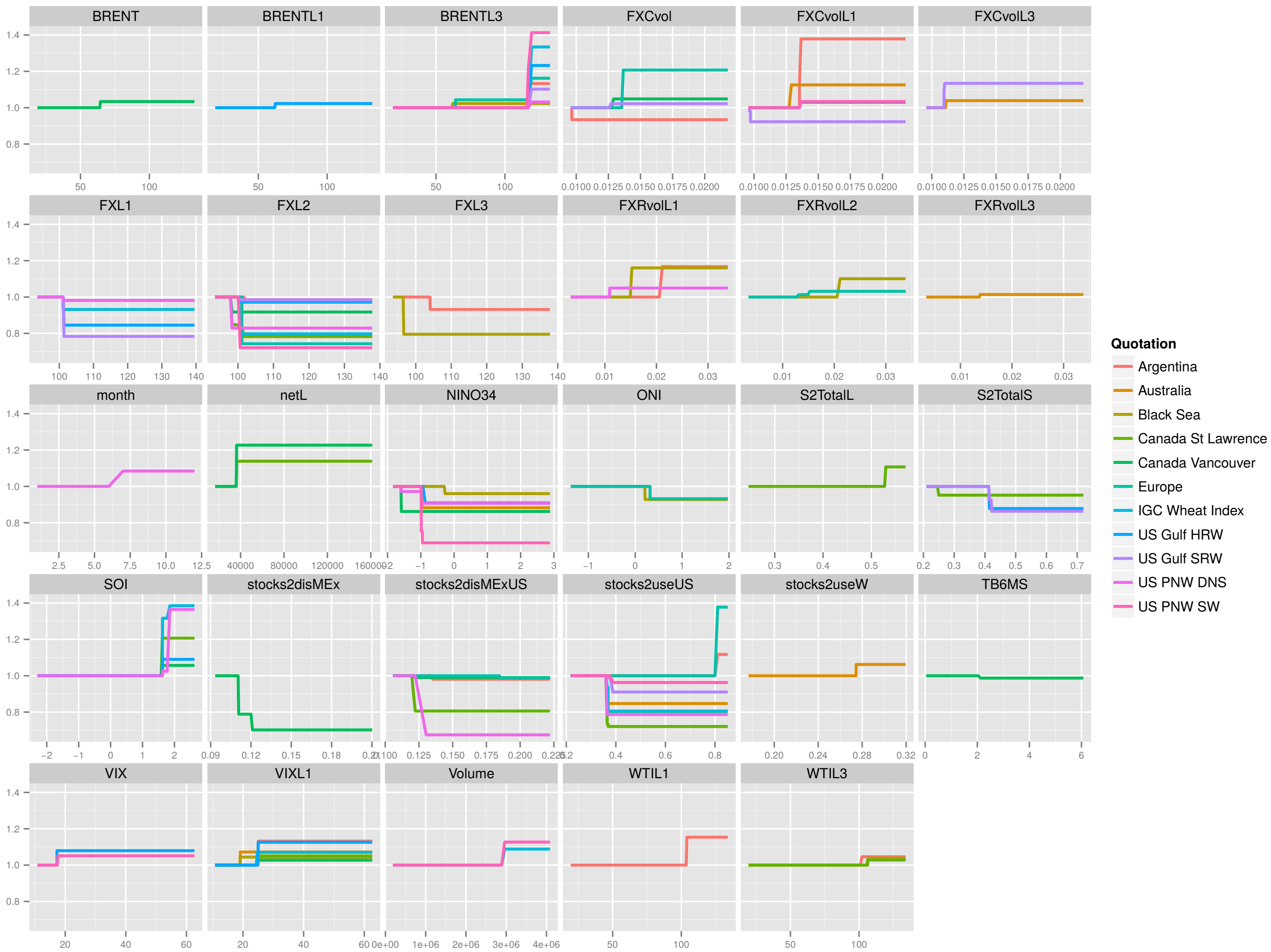


Figure 10: Estimates for maize index and prices (last week's value of each month)

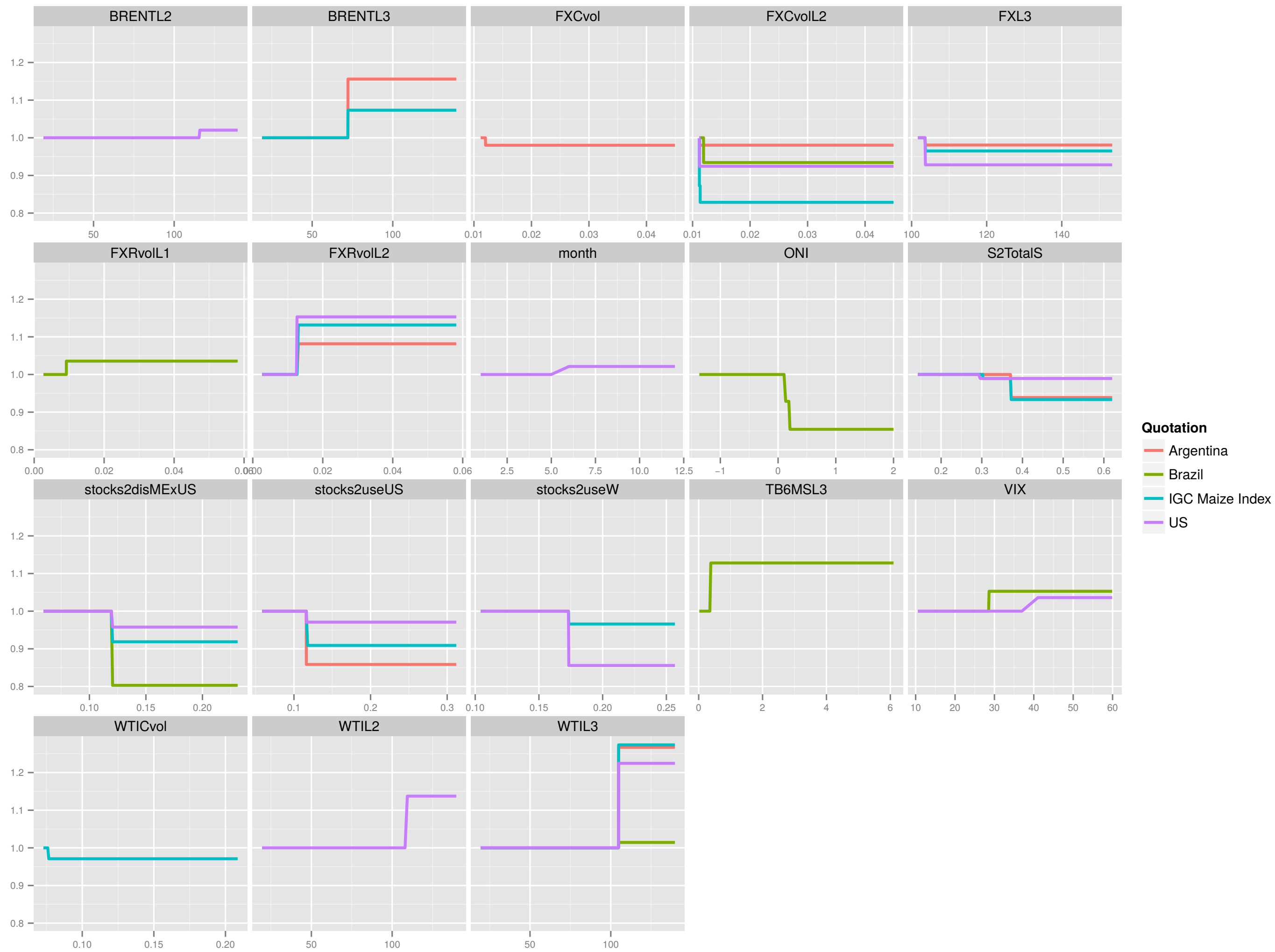


Figure 11: Estimates for maize index and prices (third week's value of each month)

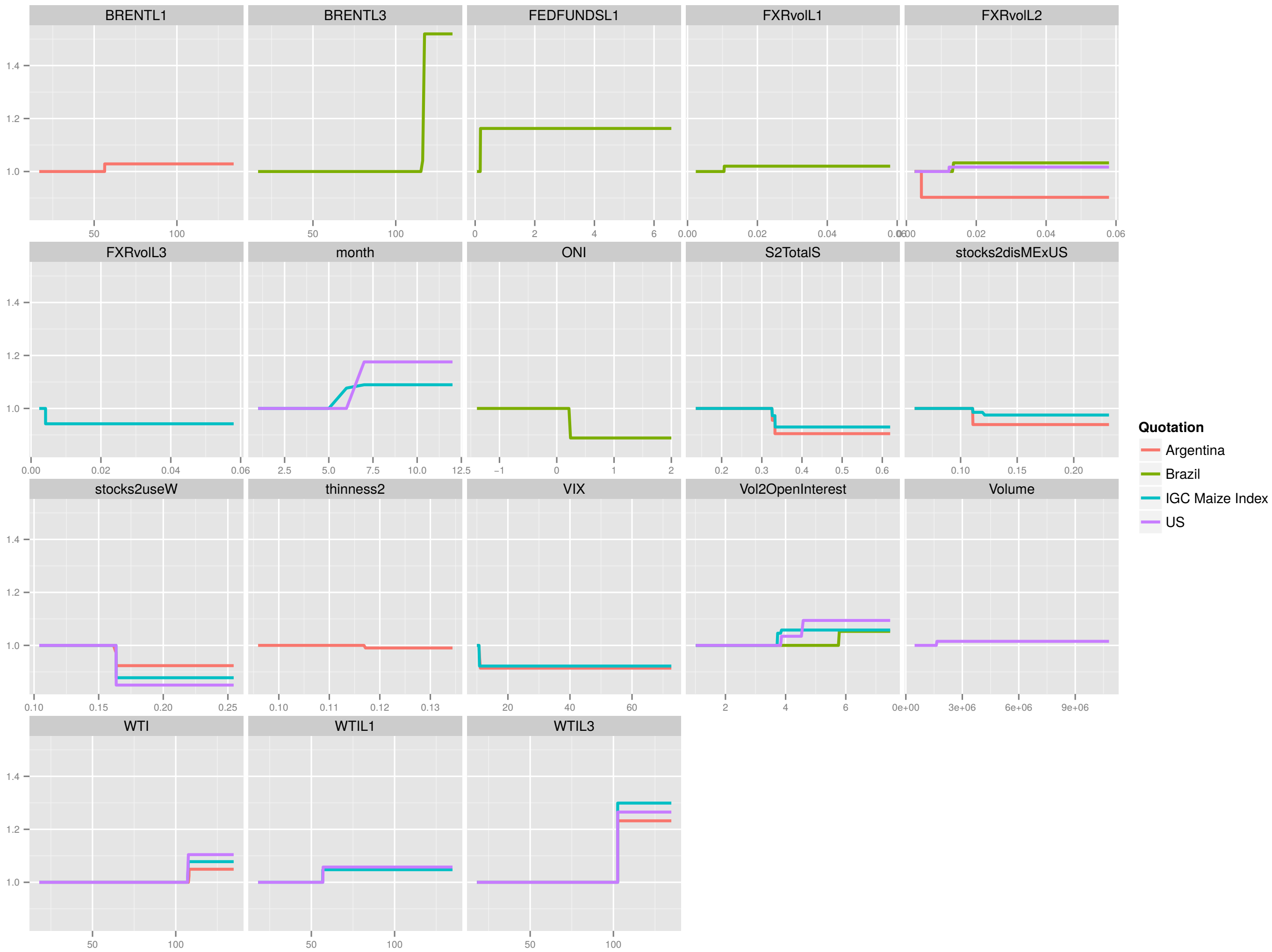


Figure 12: Estimates for maize index and prices (monthly average)

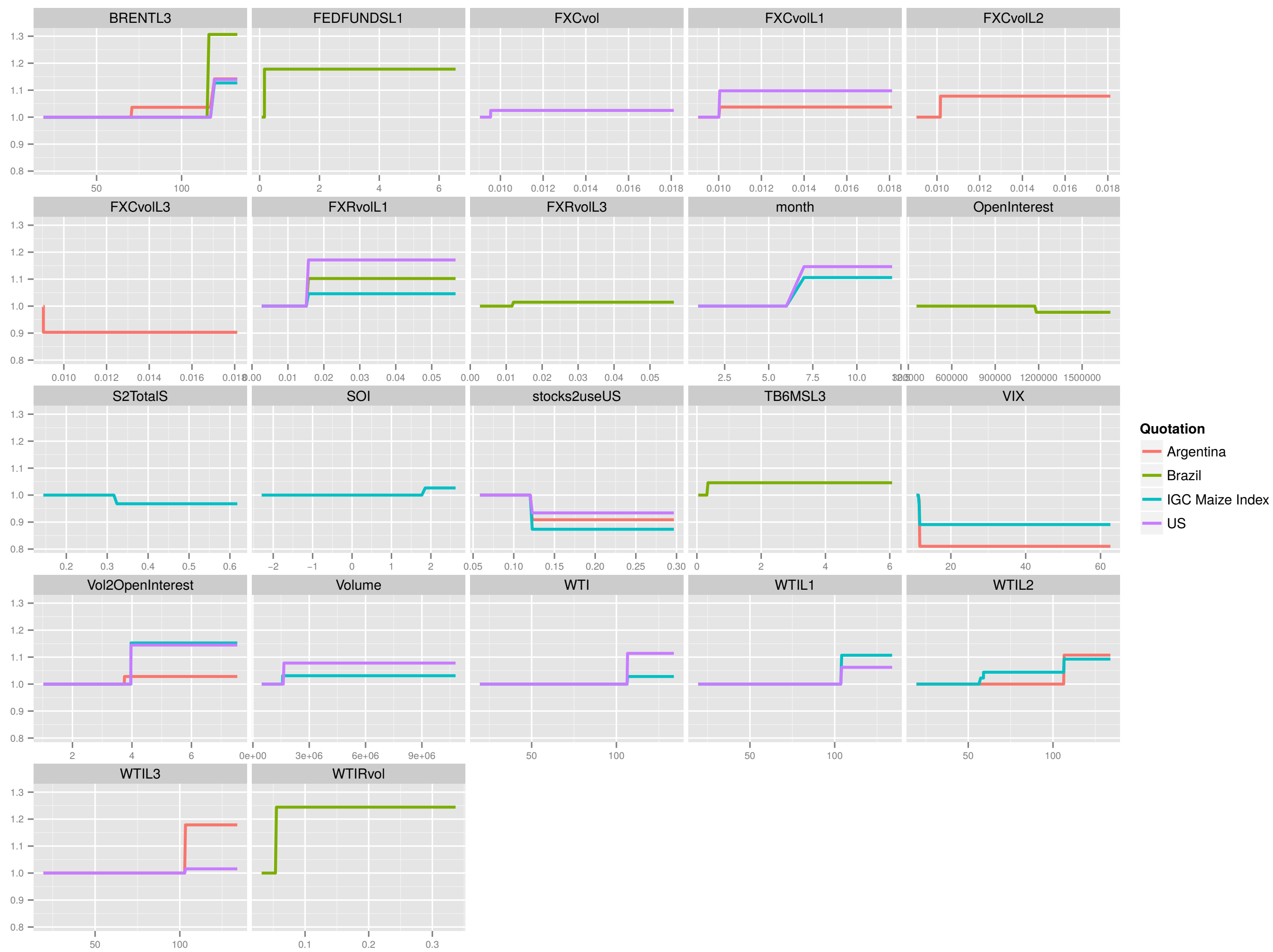


Figure 13: Estimates for soybean index and prices (last week's value of each month)

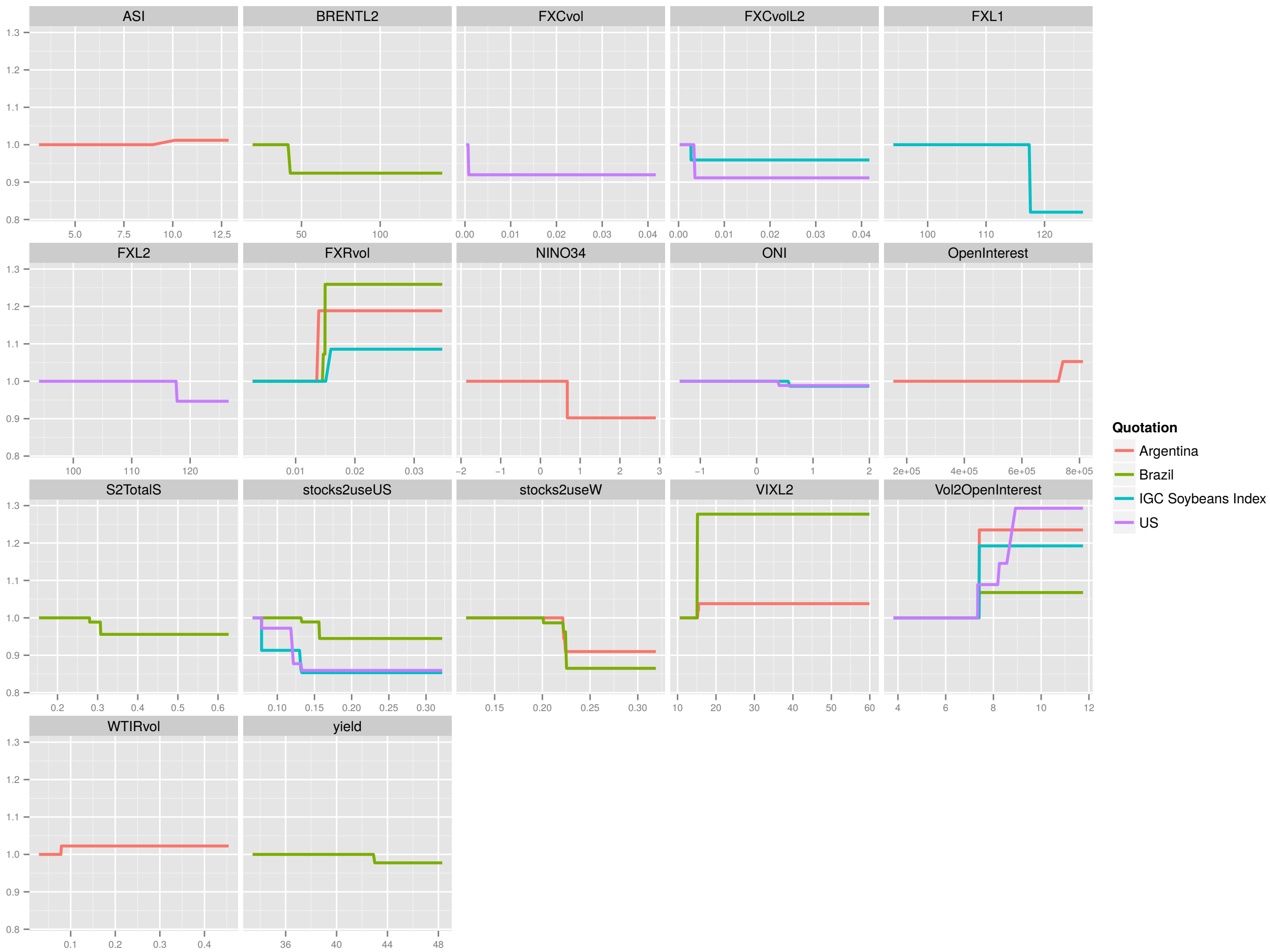


Figure 14: Estimates for soybean index and prices (third week's value of each month)

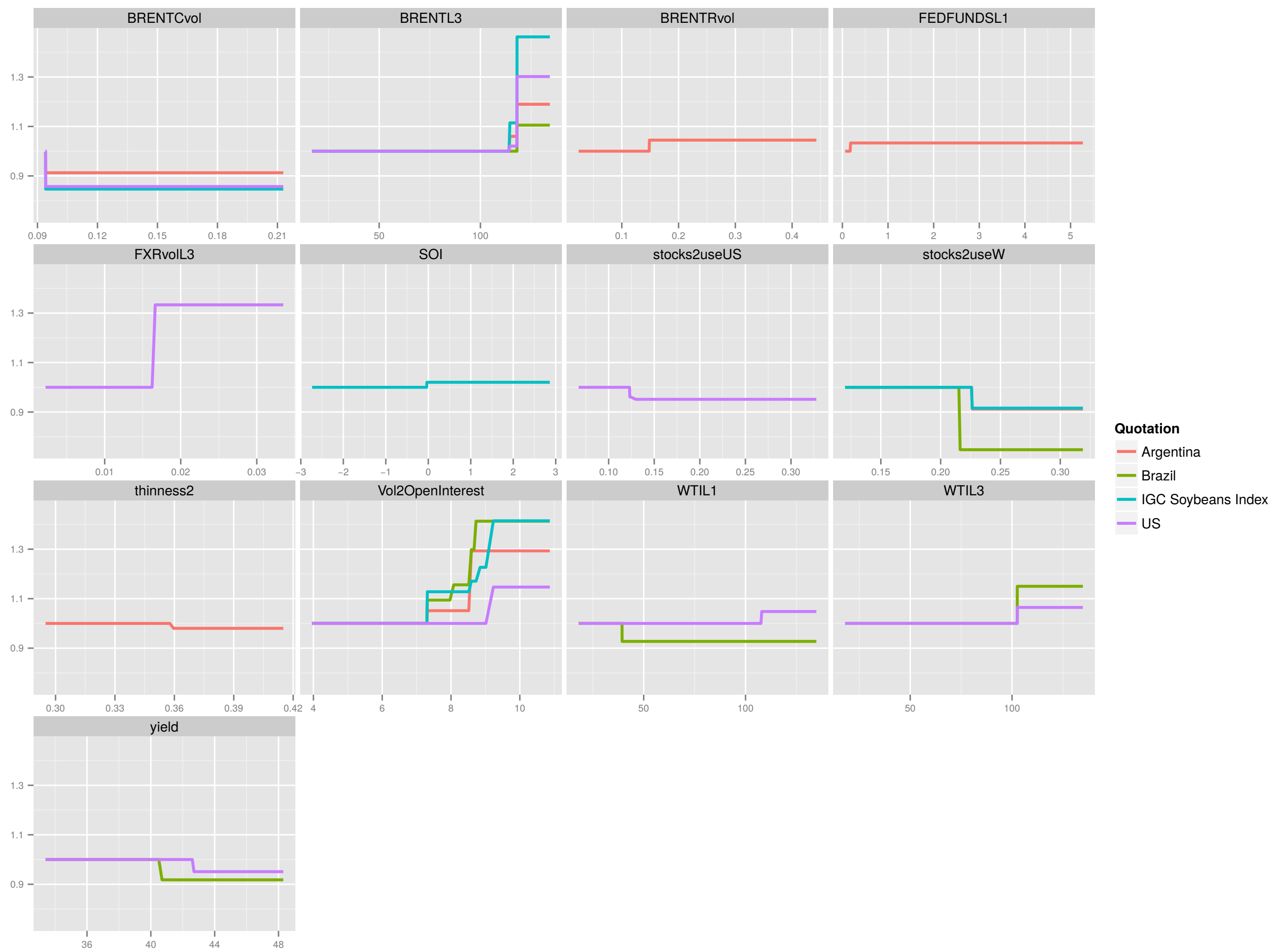
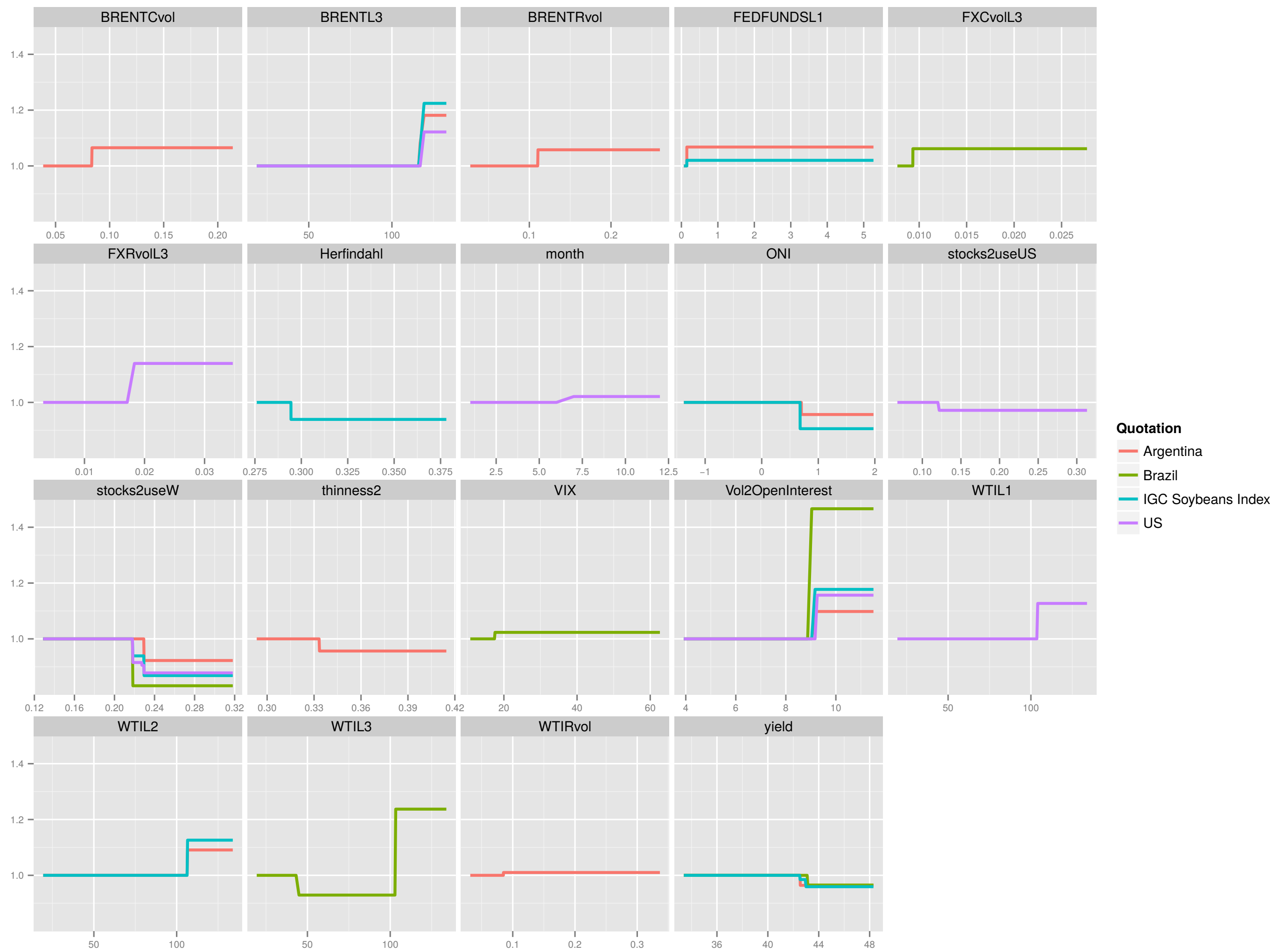


Figure 15: Estimates for soybean index and prices (monthly average)



©/Morguefile