r b e

Variance Premium and Implied Volatility in a Low-Liquidity Option Market[*]

Eduardo Sanchez Astorino[†]
Fernando Chague[‡]
Bruno Giovannetti[§]
Marcos Eugênio da Silva[¶]

We propose an implied volatility index for Brazil that we name "IVol-BR". The index is based on daily market prices of options over Ibovespa—an option market with relatively low liquidity and few option strikes. Our methodology combines standard international methodology used in high-liquidity markets with adjustments that take into account the low liquidity in Brazilian option markets. We do a number of empirical tests to validate the IVol-BR. First, we show that the IVol-BR has significant predictive power over future volatility of equity returns not contained in traditional volatility forecasting variables. Second, we decompose the squared IVol-BR into (i) the expected variance of stock returns and (ii) the equity variance premium. This decomposition is of interest since the equity variance premium directly relates to the representative investor risk aversion. Finally, assuming Bollerslev, Tauchen, & Zhou (2009) functional form, we produce a time-varying risk aversion measure for the Brazilian investor. We empirically show that risk aversion is positively related to expected returns, as theory suggests.

Nós propomos um índice de volatilidade implícita para o mercado acionário do Brasil que chamamos de "IVol-BR". O índice é baseado nos preços diários das opções

sobre o Ibovespa—um mercado de opções com liquidez relativamente baixa e pou-cos preços de exercício. Nossa metodologia combina a metodologia internacional pa-drão usada em mercados de alta liquidez com ajustes que levam em conta a baixa liquidez do mercado brasileiro. Conduzimos uma variedade de testes empíricos a fim de validar o IVol-BR. Em primeiro lugar, demonstramos que o IVol-BR possui um poder de previsão significativo sobre a volatilidade futura do retorno do Ibovespa que não é encontrado em variáveis de previsão de volatilidade tradicionais. Em se-gundo lugar, decompomos o quadrado do IVol-BR em (i) a variância esperada do re-torno e (ii) o prêmio de variância. Essa decomposição é de interesse porque o prêmio de variância se relaciona diretamente com a aversão a risco do investidor represen-tativo. Finalmente, assumindo a forma funcional de Bollerslev, Tauchen, & Zhou (2009), produzimos uma medida de aversão a risco variante no tempo para o investi-dor brasileiro. Mostramos empiricamente que essa aversão a risco é positivamente correlacionada com os retornos esperados, como a teoria sugere.

1. INTRODUCTION

This is the first article to propose an implied volatility index for the stock market in Brazil.[1] We call our implied volatility index "IVol-BR". The methodology to compute the IVol-BR combines state-of-the-art international methodology used in the US with adjustments we propose that take into account the low liquidity in Brazilian option market. The average daily volume traded in Ibovespa Index options is US\$ 20 million[2] and, as consequence, few option strikes are traded. The methodology we propose can be applied to other low-liquidity markets. This is the first contribution of the paper.

The IVol-BR has good empirical properties. First, by regressing future realized volatility on the IVol-BR and a number of traditional volatility forecasting variables, we show that the IVol-BR does contain information about future volatility. Second, we decompose the squared IVol-BR into (i) the expected variance of stock returns and (ii) the equity variance premium (the difference between the squared IVol-BR and the expected variance). This decomposition is of interest since the equity variance premium directly relates to the representative investor's risk aversion. Then, we use such a decomposition to pin down a time-varying risk aversion measure for the Brazilian market. Finally, based on the results using a sample from 2011 to 2015, we show that both the risk aversion measure and the variance premium are good predictors of future stock returns in Brazil.[3] This is the second contribution of this paper.

An implied volatility index is useful for both researchers and practitioners. It is commonly referred to as the "fear gauge" of financial markets. The best known example is the VIX, first introduced by the Chicago Board of Options Exchange (CBOE) in 1993.[4] The squared implied volatility of a stock market reflects the dynamics of two very important variables. The first relates to the level, or quantity, of risk that the representative investor faces: the expected future variance of the market portfolio. The second relates to the price of such risk, the risk aversion of this investor.

---

[1] The Chicago Board of Options Exchange (CBOE) produces a volatility index called VXEWZ, which is constructed from options over a *dollar-denominated* index of the Brazilian stock market (called EWZ). As such, the VXEWZ contains both the implied volatility of equities and the implied volatility of the FX market (R\$/US\$). The index we propose, in turn, is a clean measure of the implied volatility of the Brazilian stock market for local investors—not polluted by the exchange rate volatility. Other related works are Kapotas, Schirmer, & Manteiga (2004), that studied implied volatility of options over Telemar stocks, and Dario (2006) and Brostowicz Jr. & Laurini (2010), that studied volatility indices for the Brazilian FX market.

[2] A small fraction of the daily volume of options over S&P 500 traded in the US (about 1.5%).

[3] Ornelas (2016) also finds that the currency variance premium has good predictive properties over many US dollar exchange rates, including the Brazilian Real.

[4] See CBOE (2009).

The economic intuition for this is the following. Since options' payoffs are asymmetric, the value of any option is increasing in the expected variance of the underlying asset. Because of that, options are often used as a protection against changes in variance. Since the typical risk-averse investor dislikes variance, options are traded with a premium because of their insurance value. As a consequence, the squared implied volatility (which is directly computed from option prices) also has a premium with respect to the (empirical) expected variance: the first should always be higher than the second. This is the so-called "variance premium". The more the investor dislikes variance, the more she is willing to pay for the insurance that options provide. Therefore, the higher the risk aversion, the higher the variance premium (see, for instance, Bollerslev et al., 2009 and Bollerslev, Gibson, & Zhou, 2011).

We decompose the squared IVol-BR (hereinafter "IVar") into (i) the expected variance conditional on the information set at time t and (ii) the variance premium at time t. We estimate component (i) by searching for the best forecasting model for future variance, based on Bekaert & Hoerova (2014). Following the recent literature on variance forecasting (Chen & Ghysels, 2011; Corsi, 2009; and Corsi & Renò, 2012), we use high-frequency data for this task. Then, we compute the variance premium, i.e. component (ii), as the difference between the implied variance and the estimated expected variance. Finally, we use a closed-form equation for the variance premium, based on Bollerslev et al. (2009), which is an increasing function of the risk aversion coefficient of the representative investor, to pin down a time-varying risk aversion measure of the representative investor in the Brazilian market.

Our paper relates to Bekaert & Hoerova (2014) that shows the variance premium calculated with an econometric model for expected variance has a higher predicting power over future returns than the variance premium calculated under the assumption that the expected variance follows a random walk process, as in Bollerslev et al. (2009). In particular, we confirm Bekaert & Hoerova (2014) finding that the predicting properties of the variance premium can be significant as early as a month in advance rather than only at the quarterly frequency found by Bollerslev et al. (2009). However, we go beyond Bekaert & Hoerova (2014) and directly relate the risk aversion series obtained from the variance premium with future returns. We show the risk-aversion series is a strong predictor of future returns with a slightly superior fit than the variance premium.

The US evidence on the predicting properties of the variance premium, first shown by Bollerslev et al. (2009), has recently been extended to international developed markets. Bollerslev, Marrone, Xu, & Zhou (2014) finds that the variance premium is a strong predictor of stock returns in France, Germany, Switzerland, The Netherlands, Belgium, the UK and, marginally, in Japan. Our results confirm Bollerslev et al. (2014) findings for the Brazilian market. To the best of our knowledge, our study is the first to show this for an emerging economy.

The paper is divided as follows. Section 2 presents the methodology to compute the IVol-BR. Section 3 decomposes the squared IVol-BR into expected variance and variance premium, computes a time-varying risk-aversion measure of the representative investor in the Brazilian market, and documents the predictive power of both the variance premium and the risk-aversion measure over future stock returns. Section 4 concludes.

2. IMPLIED VOLATILITY INDEX FOR THE BRAZILIAN STOCK MARKET

The methodology we use to compute an index that reflects the implied volatility in the Brazilian stock market (called IVol-BR) combines the one used in the calculation of the "new" VIX (described in Carr & Wu, 2009) with some adjustments we propose, which take into account local aspects of the Brazilian stock market—mainly, a relatively low liquidity in the options over the Brazilian main stock index (Ibovespa) and, consequentially, a low number of option strikes.

Options over Ibovespa expire on even months: February, April, etc. Because of that, we compute the IVol-BR in order to reflect the implied volatility with a 2-month ahead horizon. It is calculated as

a weighted average of two volatility vertices: the "near-term" and "next-term" implied volatilities of options over the Ibovespa spot. On a given day t, the near-term refers to the closest expiration date of the options over Ibovespa, while the next-term refers to the expiration date immediately following the near-term. For instance, on any day in January 2015, the near-term refers to the options that expire in February 2015, while the next-term refers to the options that expires in April 2015.[5] On Table 1 we show the quarterly daily averages of the number and financial volume of options contracts traded.

The formula for the square of the near and next-term implied volatilities combines the one used in the calculation of the "new" VIX (described in Carr & Wu, 2009) with a new adjustment factor needed to deal with the low liquidity in Brazil:

$$\sigma_k^2(t) = \frac{2}{T_k - t} \sum_i \frac{\Delta K_i}{K_i^2} e^{r_t(T_k - t)} O_t(K_i) - \frac{j}{T_k - t} \left[\frac{F(t, T_k)}{K_0} - 1\right]^2, \tag{1}$$

where

- $k = 1$ if the formula uses the near-term options and $k = 2$ is the formula uses the next-term options—that is, $\sigma_1^2(t)$ and $\sigma_2^2(t)$ are, respectively, the squared near—and the next-term implied volatilities on day t;[6]

- $F(t, T_k)$ is the settlement price on day t of the Ibovespa futures contract which expires on day T_k (T_1 is the near-term expiration date and T_2 is the next-term expiration date);

- K_0 is the option strike value which is closest to $F(t, T_k)$;

- K_i is the strike of the i-th out-of-the-money option: a call if $K_i > K_0$, a put if $K_i < K_0$ and both if $K_i = K_0$.

- $\Delta K_i = \frac{1}{2}(K_{i+1} - K_{i-1})$;

- r_t is risk-free rate from day t to day T_k, obtained from the daily settlement price of the futures interbank rate (DI);

- $O_t(K_i)$ is the market price on day t of option with strike K_i;

- j is a new adjustment factor that can take the values 0, 1 or 2—in the methodology described in Carr & Wu (2009) j is always equal to 1 (we explain this adjustment below).

After calculating both the near- and next-term implied volatilities using equation (1), we then aggregate these into a weighted average that corresponds to the 2-month (42 business days) implied volatility, as follows:

$$IVol_t = 100 \times \sqrt{\left[(T_1 - t)\sigma_1^2\left(\frac{N_{T_2} - N_{42}}{N_{T_2} - N_{T_1}}\right) + (T_2 - t)\sigma_2^2\left(\frac{N_{42} - N_{T_1}}{N_{T_2} - N_{T_1}}\right)\right] \times \frac{N_{252}}{N_{42}}}, \tag{2}$$

where $IVol_t$ is the IVol-BR in percentage points and annualized at time t; N_{T_1} is the number of minutes from 5 pm of day t until 5 pm of the near-term expiration date T_1; N_{T_2} is the number of minutes from 5 pm of day t until 5 pm of the next-term expiration date T_2; N_{42} is the number of minutes in 42 business days (42×1440); and N_{252} is the number of minutes in 252 business days (252×1440).[7]

---

[5] All options expire on the Wednesday closest to the 15th day of the expiration month.

[6] The rollover of maturities occurs when the near-term options expire. We tested rolling-over 2, 3, 4, and 5 days prior to the near-term expiration to avoid microstructure effects, but the results do not change.

[7] On days when the weight of the second term of equation (2) is negative, we do not use the next-term volatility, i.e., the IVol-Br index equals the near-term volatility.

The market of options over Ibovespa presents a low liquidity. The average daily volume traded in this market is about US\$ 20 million (about 1.5% of the daily volume of options over S&P 500 traded in the US). As a consequence, we have few strikes to work with. On average, 10 different strikes for the near-term and 10 different strikes for the next-term, considering both calls and puts. Table 1 reports the quarterly daily averages of the number of strikes that we use.[8]

Table 1. NUMBER OF OPTION STRIKES USED IN THE IVOL-BR

The table shows the quarterly daily averages of the number of strikes that were used in the construction of the IVol-BR.

Period	Call Strikes	Near Term Put Strikes	Total Strikes	Call Strikes	Near Term Put Strikes	Total Strikes
2011Q3	3	5	8	2	4	6
2011Q4	5	5	10	4	4	9
2012Q1	4	6	10	4	5	9
2012Q2	5	5	10	4	4	9
2012Q3	5	6	11	4	5	9
2012Q4	5	6	11	5	4	10
2013Q1	5	6	11	5	5	10
2013Q2	6	6	12	5	5	10
2013Q3	5	6	11	4	4	9
2013Q4	3	3	7	4	3	7
2014Q1	4	5	9	3	4	7
2014Q2	4	5	8	3	4	7
2014Q3	4	5	9	5	7	12
2014Q4	4	5	9	4	5	9

The methodology presented above departs from the standard one described in Carr & Wu (2009) for three reasons which are related to the relatively low liquidity and low number of strikes traded in the options market in Brazil:

1) We introduce the adjustment factor j in equation (1) to account for the following: (i) there are days when only a call or a put at K_0 is traded—Carr & Wu (2006) have always both a call and a put—; moreover, (ii) we have to define K_0 as the option strike value which is closest to $F(t, T_k)$ and, because of that, we may have either $K_0 > F(t, T_k)$ or $K_0 < F(t, T_k)$—Carr & Wu (2006) define K_0 as the option strike value immediately below $F(t, T_k)$. Depending on the situation we face regarding (i) and (ii), the value of j is set to 0, 1 2, as follows (an explanation about this can be found in the Appendix):

Table 2. POSSIBLE VALUES OF j

	$K_0 < F$	$K_0 > F$	$K + 0 = F$
$\exists$ call, $\exists$ put	1	1	1
$\exists$ call, $\nexists$ put	2	0	0
$\nexists$ call, $\exists$ put	0	2	0

---

[8]The fraction of options that were excluded because they were in-the-money are the following: 52.42% in-the-money near-term calls, 46.53% near-term puts, 60.34% next-term calls, and 52% next-term puts.

2) We widen the time frame of options prices to the interval [3 pm, 6 pm]. For each strike, we use the last deal in this interval to synchronize the option price with the settlement price of the Ibovespa futures;

3) We only calculate $\sigma_1^2(t)$ and $\sigma_2^2(t)$ if, for each vertex, there are at least 2 trades involving OTM call options at different strikes and 2 trades involving OTM put options also at different strikes—this is done in order to avoid errors associated with lack of liquidity in the options market. If on a given day only one volatility vertex can be calculated, we suppose that the volatility surface is flat and the IVol-Br is set equal to the computed vertex. If both near- and next-term volatilities cannot be calculated, we report the index for that day as missing.

The volatility index calculated according to equations (1) and (2) could be biased because it considers only traded options at a finite and often small number of strikes. To assess the possible loss in the accuracy of the integral calculated with a small number of points, we refine the grid of options via a linear interpolation using 2,000 points of the volatility smile that can be obtained from the traded options (based on the procedure suggested by Carr & Wu, 2009).[9] The results did not change.

The IVol-BR series, computed according to the methodology described above, is available for download at the webpage of the Brazilian Center for Research in Financial Economics of the University of Sao Paulo (NEFIN).[10] Figure 1 plots the IVol-Br for the period between August 2011 and February 2015, comprising 804 daily observations.

Figure 1. IMPLIED VOLATILITY IN BRAZIL – THE IVOL-BR
This figure shows the daily time-series of the IVol-Br in percentage points and annualized.

---
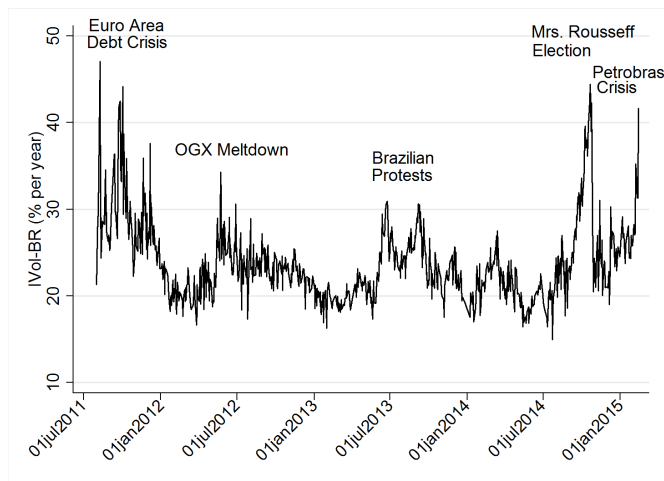
[9]The "coarse" volatility smile for both near and next-term is retrieved from the options market values and the Black-76 formula. We then refine the grid of strike prices K_i using the implied volatilities and implied deltas of the options with the formula:

$$K_i = F(t, T) \exp\left[-w\sigma_i \sqrt{T-t} N^{-1}\left(|\Delta_i| + \tfrac{1}{2}\sigma_i^2(T-t)\right)\right],$$

where $w = 1$ for calls and $w = -1$ for puts; $N^{-1}(\cdot)$ is the inverse of the standard normal cumulative density function. To simplify the process of retrieving K_i, we transform all traded options in calls (via put-call parity) and create a smile in the $(\Delta_{\text{call}}, \sigma)$ space. We then generate 2,000 points by linearly interpolating this smile considering two intervals: (i) the interval $\left[\Delta_{\max}; \Delta_{\min}\right]$ of deltas of the traded options; (ii) the interval $[99; 1]$ of deltas.

[10]http://www.nefin.com.br

An implied volatility index should reflect the dynamics of (i) the level, or quantity, of risk that investors face—the expected future volatility—and (ii) the price of such risk—the risk-aversion of investors. Given that, the IVol-BR should be higher in periods of distress. As expected, as Figure 1 shows, the series spikes around events that caused financial distress in Brazil, such as the Euro Area debt crisis (2011), the meltdown of oil company OGX (2012), the Brazilian protests of 2013, the second election of Mrs. Rousseff (2014) and the corruption and financial crisis in Petrobras (2015).

It is also interesting to compare the IVol-BR with the VXEWZ, the CBOE's index that tracks the implied volatility of a dollar-denominated index (EWZ) of the Brazilian stock market. Figure 2 shows the evolution of both series.

As Figure 2 presents, the VXEWZ is often higher than IVol-BR. This happens because the VXEWZ, which is constructed using options over the EWZ index (that tracks the level in dollars of the Brazilian stock market), embeds directly the exchange rate volatility. In turn, the IVol-BR is constructed using options over the Ibovespa itself and, hence, reflects only the stock market volatility. Thus the IVol-BR is better suited to describe the implied volatility of the Brazilian stock market for local investors or foreign investors that have hedged away the currency risk. During the period depicted in Figure 2 there were important changes in the exchange rate volatility that directly impacted the VXEWZ but not the IVol-BR.

Figure 2. COMPARING IVOL-BR AND VXEWZ
This figure shows the daily time-series of the IVol-Br and the VXEWZ. Both series are in percentage points and annualized. VXEWZ is the implied volatility index on the Brazilian stocks ETF EWZ and is calculated by CBOE.
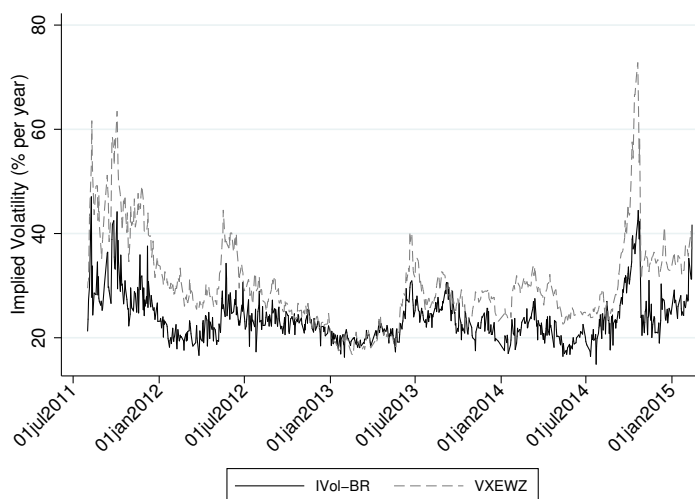
2.1. Comparison of IVol-BR with ATM volatility

In this section we compare the IVol-BR to the Black-Scholes implied volatility calculated using only at-the-money options; we call this volatility index "IVol-ATM." To compute it, we first calculate the Black-Scholes put and call implied ATM-volatilities from the two strikes immediately below and above the current futures' settlement price. Then, we linearly interpolate the put and call implied volatilities to obtain the forward implied volatility. This procedure is done to both near- and next-term options. Finally, we standardize the near- and next-term forward implied volatilities to have a basis period of two months (42 business days) to obtain the Ivol-ATM. Figure 3 shows the daily time-series of the Ivol-BR and the Ivol-ATM from 2011 to 2015.

As can be seen in Figure 3, both Ivol-BR and Ivol-ATM series are highly correlated (0.92). This is not surprising, since a similar pattern is also observed between the VIX and the VXO, the Black-Scholes implied volatility version of the VIX that also uses only only at-the-money options on the S&P 500 index. Indeed, both VIX and VXO have a correlation of 0.99 during the same sample period.[11]

In Carr & Wu (2006) we can find a discussion on the reasons which led to the adoption of the VIX. Firstly, the VIX is model-free and the VXO depends on the Black-Scholes framework. Secondly, although both VIX and VXO provide good approximations of the volatility swap rate, the VIX is much easier to replicate using option contracts than the VXO. Indeed, despite its popularity as a general volatility reference index, no derivative products have been launched on the VXO index. In contrast, less than a year after its introduction, the VIX was already serving as basis for new futures contracts.

Figure 3. COMPARING IVOL-BR AND VXEWZ
This figure shows the daily time-series of the IVol-Br and the IVol-ATM. Both series are in percentage points and annualized. IVol-ATM is the Black-Scholes implied volatility calculated using only at-the-money options.
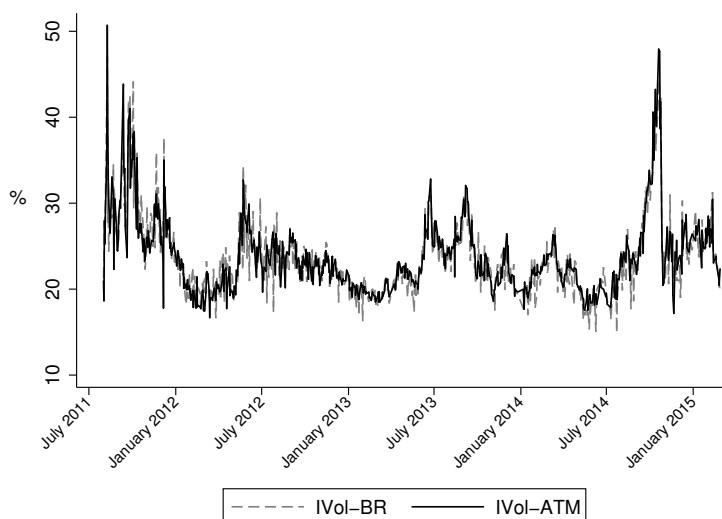
3. EMPIRICAL ANALYSIS USING THE IVOL-BR

In this section we use the squared IVol-BR series, which we call IVar, in some interesting empirical exercises. We first decompose the IVar into (i) the actual expected variance of stock returns and (ii) the variance premium.[12] Then, from the variance premium, we produce a time-varying risk-aversion measure for the Brazilian investor. Finally, we show empirically that higher risk-aversion is accompanied with higher expected returns, as asset pricing theory suggests. The reason for working with the IVar, instead of the IVol-BR, is that theoretical models usually produce closed-form equations that relate risk aversion to the variance premium and not the volatility premium.

---

[11] Prior to September 2003 the Chicago Board of Options Exchange (CBOE) calculated a volatility index that used only ATM options. However, even after introducing the new VIX, the CBOE continued to calculate the older, ATM index, under a new name (VXO). Both data sets can be downloaded from the CBOE website.

[12] It can be shown that the implied variance approximates the expected variance under the risk neutral measure (see for instance Carr & Wu (2006)). Given that, what we call "variance premium" is the difference between the expected variance under the risk neutral measure and the expected variance under the empirical, or historical, measure.

In section 3.1 we decompose the implied variance, calculated as the IVol-BR squared, into (i) the expected variance of stock returns and (ii) the equity variance premium. To do this, we first estimate a model that represents the conditional expectation of investors of future variance. Then, by calculating the difference between the implied variance and the estimated expected variance, we arrive at a daily measure of the variance premium. In section 3.2, from the volatility premium, we produce a time-varying risk aversion measure for the Brazilian investor from the variance premium. In section 3.3 we show that the variance premium and the risk aversion measure are able to predict future stock returns as theory suggests: when variance premium (risk-aversion) is higher, expected returns are higher.

3.1. Decomposing implied variance into expected variance and variance premium

To decompose implied variance into expected variance and a premium, we first search for the model that best forecasts variance. Because the implied variance, calculated by squaring the IVol-BR, reflects markets expectations for the two-months ahead, the measure of expected variance of interest is also over the same two-month period.

The variance of returns is a latent, unobservable variable. Fortunately, we can obtain a good estimator of the variance of returns from high frequency data and use the estimated time-series, the so-called realized variance, as the dependent variable of our forecasting model. Formally, the realized variance over a two-month period at day t is calculated by summing squared 5-minute returns over the last 42 trading days:

$$RV_t^{(42)} = \frac{252}{42} \times \sum_{i=1}^{\lfloor 42/\Delta \rceil} r_i^2,$$

where $\Delta = 5/425$ is the 5-minute fraction of a full trading day with 7 hours including the opening observation; $\lfloor \cdot \rceil$ is the operator that approximates to the closest integer; and $r_i = 100 \times \left[\ln\left(Ibov_i\right) - \ln\left(Ibov_{i-i}\right) \right]$ is a 5-minute log-return in percentage form on the Ibovespa index, except when i refers to the first price of the day, in which case r_i corresponds to the opening/close log-return.

Following the recent literature on variance forecasting (Chen & Ghysels, 2011, Corsi, 2009, Corsi & Renò, 2012, and Bekaert & Hoerova, 2014), we construct several explanatory variables (predictors) from a 5-minute returns data set.[13] First, we include in the set of explanatory variables lags of realized variance at heterogeneous frequencies to account for the clustering feature of stock returns variance. In the spirit of Corsi's HAR model (Corsi, 2009), lags of bimonthly, monthly, weekly and daily realized variances are included: $RV_t^{(42)}$, $RV_t^{(21)}$, $RV_t^{(5)}$ and $RV_t^{(1)}$. Formally:

$$RV_t^{(k)} = \frac{252}{k} \times \sum_{i=1}^{\lfloor k/\Delta \rceil} r_i^2$$

for each $k = 42, 21, 5, 1$.

One important feature of variance is the asymmetric response to positive and negative returns, commonly referred to as leverage effect. To take this into account, Corsi & Renò (2012) suggests including lags of the following "leverage" explanatory variables:

$$Lev_t^{(k)} = -\frac{42}{k} \sum_{i=1}^{\lfloor k/\Delta \rceil} \min\{r_i, 0\},$$

with $k = 42, 21, 5, 1$. For a convenient interpretation of the estimated parameters, we take the absolute value of the cumulative negative returns.

---

[13]We thank BM&FBovespa for providing the intraday data set.

Andersen, Bollerslev, & Diebold (2007) show that jumps help in predicting variance. Following the theory laid out by Barndorff-Nielsen & Shephard (2004), realized variance can be decomposed into its continuous and jump components with the usage of the so-called bipower variation (BPV). As shown by these authors, under mild conditions, the BPV is robust to jumps in prices while the realized variance is not. This insight allows one to identify jumps indirectly by simply calculating the difference:

$$J_t = \max\{RV_t - BPV_t, 0\},$$

where $BPV_t = \frac{252}{42} \times \sum_{i=1}^{\lfloor 42/\Delta \rceil} |r_i| |r_{i-1}|$. The maximum operator is included to account for the situation when there are no jumps and the BPV is eventually higher than the realized variance. The continuous component of the realized variance is defined as follows:

$$C_t = RV_t - J_t.$$

Lagged variables of the continuous and jump components at other time frequencies are also included. Using the same notation as before, the following eight variables are added $C_t^{(k)}$, $J_t^{(k)}$ with $k = 42$, 21, 5, 1.

Finally, we follow Bekaert & Hoerova (2014) and include the lagged implied variance as explanatory variable. Importantly, as will be shown, this variable contains information about future realized variance that is not contained in lagged realized variance and other measures based on observed stock returns.

To find the best forecasting model, we apply the General-to-Specific (GETS) selection method proposed by David Hendry (see for instance Hendry, Castle, & Shephard, 2009). The starting model, also called GUM or General Unrestricted Model, includes all the variables described above plus a constant:

$$RV_{t+42} = c + \gamma_1 IVar_t + \gamma_2 C_t^{(42)} + \gamma_3 C_t^{(21)} + \gamma_4 C_t^{(5)} + \gamma_5 C_t^{(1)} + \gamma_6 J_t^{(42)} + \gamma_7 J_t^{(21)} + \gamma_8 J_t^{(5)} + \gamma_9 J_t^{(1)}$$
$$+ \gamma_{10} Lev_t^{(42)} + \gamma_{11} Lev_t^{(21)} + \gamma_{12} Lev_t^{(5)} + \gamma_{13} Lev_t^{(1)} + \epsilon_t.$$

To avoid multicolinearity, the lagged realized variance measures were excluded from the initial set of explanatory variables since by construction they are approximately equal to $RV_t^{(k)} \approx C_t^{(k)} + J_t^{(k)}$. However, in a robustness exercise below, we include these variables in other forecasting models.

Following an iterative process, the method searches for variables that improve the fit of the model but penalizes for variables with statistically insignificant parameters. The regressions are based on daily observations. Table 3 shows the estimates of the final model—the GETS model. Eight variables plus a constant remain in the GETS model: $IVar_t$, $C_t^{(42)}$, $C_t^{(5)}$, $J_t^{(21)}$, $J_t^{(5)}$, $Lev_t^{(42)}$, $Lev_t^{(21)}$ and $Lev_t^{(5)}$.

Importantly, the coefficient on the implied variance is positive (0.152) and highly significant. This indicates that, as expected, IVar does contain relevant information about future variance, even after controlling for traditional variance forecasting variables.

From the GETS model, we calculate a time-series of expected variance. We name the difference between implied variance and this time-series of expected variance as the variance premium:

$$VariancePremium_t = IVar_t - \widehat{\sigma}_t^2, \tag{3}$$

where $\widehat{\sigma}_t^2 = \mathbb{E}_t\left[RV_{t+42}\right] = \mathbb{E}_t\left[\sigma_{t,t+42}^2\right]$ is the GETS model expected variance computed using information up until day t; the subscript $t + 42$ emphasizes the fact that it is the expected variance over the same horizon as the implied variance, $IVar_t$. Figure 4 shows both series and Figure 5 shows the variance premium. We observe that the premium varies considerably. The 3-month moving average shown in Figure 5 suggests that the average premium varies and remains high for several months.

Figure 4. IMPLIED VARIANCE AND EXPECTED VARIANCE

This figure shows the weekly time-series of the implied variance—the squared of the IVol-Br—and the estimated expected variance. The model for expected volatility is the GETS model shown on Table 3. Both series are in percentage points and annualized.
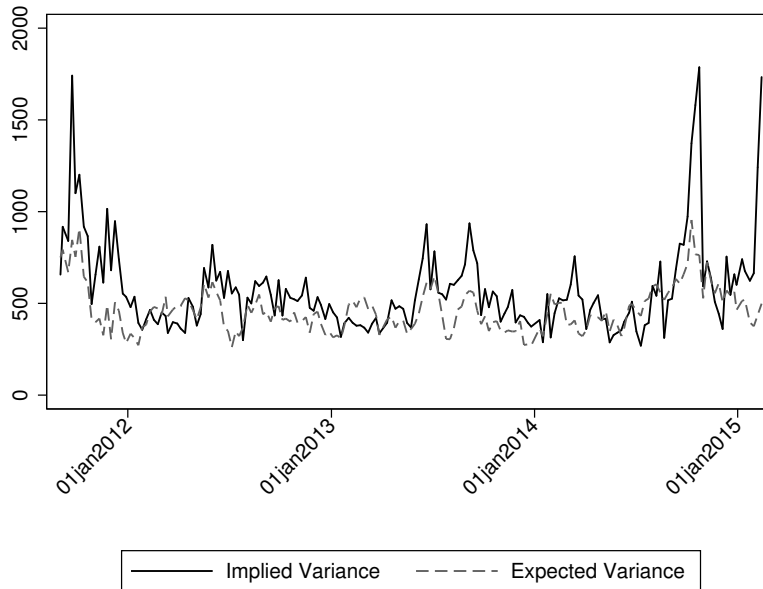
Figure 5. THE VARIANCE PREMIUM

This figure shows the weekly time-series of the variance premium calculated by the difference of the implied variance and expected variance as predicted by the GETS model shown on Table 3, and its three month moving average.
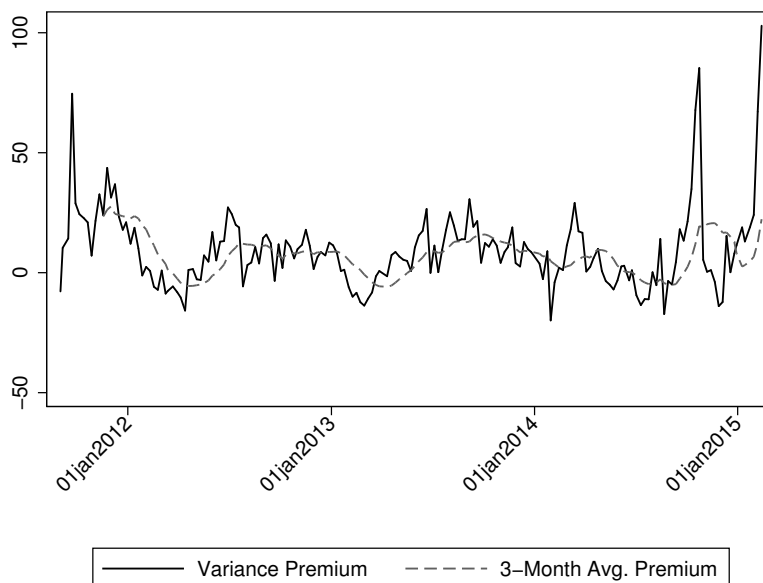
Table 3. GENERAL-TO-SPECIFIC BEST MODEL

The table shows the estimates of the best variance forecasting model following the General-to-Specific selection method. The starting model, also called GUM or General Unrestricted Model, comprises of all independent variables. The standard errors reported in parenthesis are robust to heteroskedasticity. Regressions are based on daily observations. The corresponding p-values are denoted by $*$ if $p < 0.10$, $**$ if $p < 0.05$ and $***$ if $p < 0.01$.

$IVar_t$	0.152^{***}
	(0.040)
$C_t^{(42)}$	3.215^{***}
	(0.267)
$C_t^{(5)}$	-0.656^{***}
	(0.123)
$J_t^{(21)}$	-0.540^{***}
	(0.127)
$J_t^{(5)}$	0.307^{***}
	(0.082)
$Lev_t^{(42)}$	-2.237^{***}
	(0.155)
$Lev_t^{(21)}$	0.722^{***}
	(0.084)
$Lev_t^{(5)}$	0.367^{***}
	(0.066)
Constant	950.186^{***}
	(73.048)
Number of Obs.	741
R^2	0.400
Adjusted R^2	0.393
RMQE	176.345
BIC	9818.8

3.2. The variance risk premium and the risk aversion coefficient

An implied variance index reflects the dynamics of two very important variables. The first relates to the level, or quantity, of risk that investors face: the expected future variance of the market portfolio, estimated above. The second relates to the price of such risk: the risk aversion of the representative investor.

Since options' payoffs are asymmetric, the value of any option (call or put) is increasing in the expected variance of the underlying asset. Because of that, options are often used as a protection against changes in expected variance. Since the typical risk-averse investor dislikes variance, options are traded with a premium because of such an insurance value. As a direct consequence, the implied variance (IVar, the IVol-BR squared), which is computed directly from options prices, also has a premium with respect to the expected variance. That is, the more risk-averse the investor is, the more she is willing to pay for the insurance that options provide, i.e., the higher the variance premium.

In order to make this connection between risk aversion and variance premium more precise, we need to impose some economic structure. To do this, we use Bollerslev et al. (2009) economic model, which is an extension of the long-run risk model of Bansal & Yaron (2004). We assume that the following

closed-form equation for the variance premium holds for each t:[14]

$$VariancePremium_t = \frac{\psi^{-1} - \gamma_t}{1 - \psi^{-1}} \times \frac{\kappa_1 (1 - \gamma_t)^2}{2\left(\frac{1 - \gamma_t}{1 - \psi^{-1}}\right)(1 - \kappa_1 \rho_\sigma)} q, \tag{4}$$

where ψ is the coefficient of elasticity of intertemporal substitution, γ_t is the time-varying risk aversion coefficient, q the volatility of the volatility, and ρ_σ is the auto-regressive parameter in the volatility of consumption.

Using the estimated weekly series for the variance premium computed above and usual parameter calibration,[15] we pin down a time-series for the time varying risk aversion coefficient of the representative investor in Brazil.[16] The resulting series is plotted in Figure 6. The smallest value for γ_t is 1 on August 22, 2014; and the highest value is 57 on February 13, 2015. The average risk aversion level is 26. Such values are consistent with the results in Zhou (2009)—an average risk aversion higher than 10 is needed to match the empirical moments of the variance premium (see his Table 8).

Figure 6. RISK AVERSION
This figure show a time-series for the risk aversion index in Brazil. It is computed by combining the weekly series for the variance premium with Bollerslev et al. (2009) functional form for the variance premium, as explained in section 3.3.
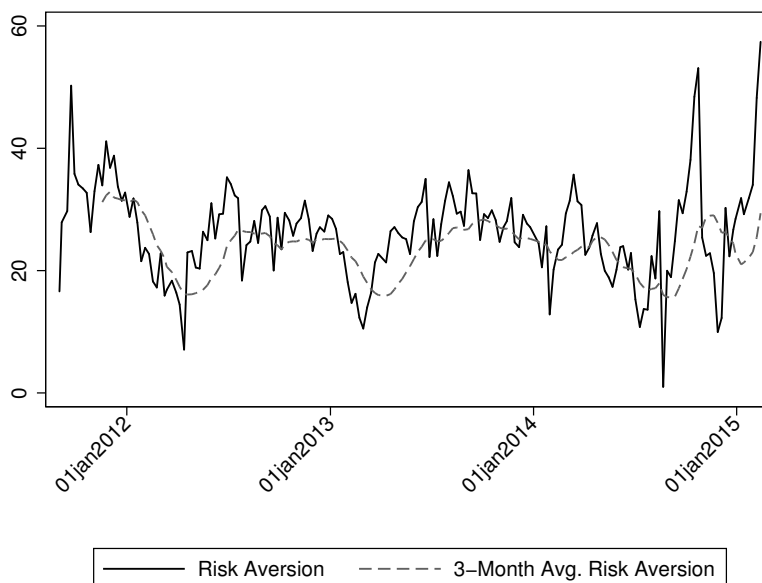
Using different methodology and data sets, other papers find (unconditional) lower estimates of risk aversion for the Brazilian representative investor. For instance, Fajardo, Ornelas, & Farias (2012) using data on currency options for the Brazilian Real from 1999 to 2011, estimate the implied risk-neutral density and, by comparing it to the objective density, find a coefficient of relative risk aversion of around 2.7. Similarly, Issler & Piqueira (2000) using data on aggregate consumption from 1975 to 1994,

---

[14]We use their simpler equation, where they assume a constant volatility of volatility (the process q is constant at all t).

[15]We set $\psi = 1.5$, $q = 10^{-6}$, $\kappa_1 = 0.9$ and $\rho_\sigma = 0.978$ following the calibration in Bansal & Yaron (2004) and Bollerslev et al. (2009).

[16]Equation (4) is quadratic on the risk-aversion coefficient γ_t. In order to avoid complex roots, we shift the variance premium upward so that the minimum variance premium corresponds to the minimum value of $\gamma_t = 1$.

document a coefficient of relative risk aversion lower than five. In the next section we assess to which extent the lower risk aversion found in other studies can be attributed to differences in calibration.

### 3.2.1. Sensitivity of risk-aversion to calibration parameters

A number of papers estimate the elasticity of intertemporal substitution for the Brazilian representative investor. Issler & Piqueira (2000) find a relatively low ψ of 0.29 using annual consumption data from 1975 to 1994—less than half of the ψ they find for the US (0.72) for the same period and using the same methodology. A low value of ψ for the Brazilian economy is also reported by Havranek, Horvath, Irsova, & Rusnak (2015); the authors find that households in developing countries and countries with low stock market participation, such as Brazil, substitute a lower fraction of consumption intertemporally in response to changes in expected asset returns, with implied values for ψ often below one. On the other hand, using a quarterly data set of 1975 to 2000, Araújo (2005) finds a much higher range of possible ψ's, with the minimum value of ψ being equal to 2.5. Given that there is no consensus on which is the ψ that best characterize the behavior of the representative investor in Brazil, we assess the sensitivity of our results to the choice of ψ by comparing the estimates of γ_t for three different values of ψ: 0.50, 1.50, and 3.00.

Another parameter that was used in the calibration that can vary across countries is κ_1. The parameter κ_1 is equal to the ratio $\overline{PD}/(1+\overline{PD})$, where $\overline{PD}$ is a long-run average of the price-dividend ratio. The number used in our calibration was 0.9 and is the same one used Bollerslev et al. (2009). According to its formula, the value of $\kappa_1 = 0.9$ corresponds to a price-dividend ratio of around 10. This number, however, can vary depending on the data set considered. Indeed, the average price-dividend ratio computed for Brazil in the period 2001–2015 is around 40. Thus, we also assess the sensitivity of our results to the choice of κ_1 by comparing the estimates of γ_t for another two different values of κ_1: 0.98 and 0.95. These two values correspond, respectively, to price-dividend ratios of 40 and 20.

Figure 7 shows the image of the function γ_t —the inverse of the function in equation (4)—for plausible values of the variance premium and for nine pairs of (ψ, κ_1), each one a possible combination

Figure 7. RISK AVERSION AS A FUNCTION OF ψ AND κ_1
This figure shows the image of the γ_t function for reasonable values of the variance premium and for different choices of ψ and κ_1.
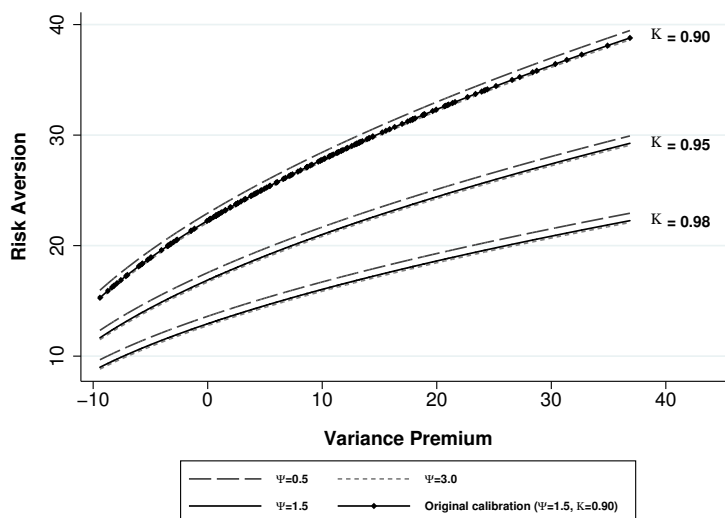
of $\psi = (0.5, 1.5, 3.0)$ with $\kappa_1 = (0.90, 0.95, 0.98)$. The estimated values of γ_t differ depending on the chosen parametrization, particularly across κ_1. At a variance premium of five, γ_t can be 13, 16, or 23 if κ_1 is, respectively, 0.98, 0.95, or 0.90. On the other hand, the values of γ_t do not vary much across ψ; for any given κ_1 and value of variance premium, the range of possible values for γ_t is less than one. Overall, at any pair of (ψ, κ_1) considered the estimates of γ_t are higher than 10.

3.3. Predicting future returns

If the variance premium positively comoves with investors risk-aversion, it should predict future market returns: when risk aversion is high, prices are low; consequentially, future returns (after risk aversion reverts to its mean) should be high. Moreover, the risk aversion measure itself, computed in section 3.2, should also predict future returns. In this Section we test these predictions by regressing future market returns on both the variance premium and the risk-aversion measure.

Table 4 shows the results of our main regression. The dependent variable is the return on the market portfolio 4 weeks ahead. To limit the overlapping of time-series, we reduce the frequency of our data set from daily to weekly by keeping only the last observation of the week. Additionally, to account for the remaining serial correlation in the error term, the standard errors are computed using Newey-West estimator. Columns (1) and (2) show that implied variance $IVar_t$ and expected variance $\widehat{\sigma}_t^2$ alone are not very good predictors of future returns. On the other hand, Column (3) shows that the variance premium, resulting from a combination of both variables, $IVar_t - \widehat{\sigma}_t^2$, strongly predicts future returns at the 4-week horizon. The estimated coefficient is positive, 0.089, and significant at the 1% confidence level. Column (4) shows that the risk aversion measure also predicts future returns at the 4-week horizon. The estimated coefficient is positive, 0.180, and significant at the 1% confidence level.

The predictive power of the variance premium and the risk aversion measure remains after we include in the regression the divided yield $\log(D_t/P_t)$, another common predicting variable. Again, columns (5) and (6) show that implied variance and expected variance alone are poor predictors of returns. On the other hand, both the variance premium and the risk aversion measure do predict future returns. Column (7) shows a positive coefficient for the variance premium, 0.066, significant at the 5% confidence level. Column (8) shows a positive coefficient for the risk-aversion measure, 0.135, also significant at the 5% confidence level.

In columns (1) through (8) of tables 5 and 6, the regressions are the same as the one in Column (7) and (8), respectively, of Table 4, except for the horizon of future returns. As the significance and values of the estimates indicate, the variance premium predictability is stronger at the 4-week horizon—columns (7) and (8).

A concern is that the standard errors in the first eight Columns in tables 5 and 6 may be biased due to the presence of a persistent explanatory variable such as the log dividend yield (see for instance Stambaugh, 1999) combined with a persistent dependent variable (overlapping returns). To address this concern, columns (9) and (10) in both tables show the same regressions of columns (7) and (8) but based on *non-overlapping* 4-week returns. As we can see, the coefficients on the variance premium and risk-aversion remain positive and significant.

Another concern may be that the actual expected variance by market participants cannot be observed. Hence, our measure of expected variance depends on the model chosen by the econometrician. To address this concern, we also assess to which extent our results depend on the chosen variance model.

Tables 7 and 8 show the estimates of several models. Table 7 brings the estimates of Corsi's HAR model (Corsi, 2009) in Column (1), with the addition of a 42-day realized variance lag in accordance with the frequency of the dependent variable. In columns (2), (3) and (4) we include the lagged implied variance, $IVar_t$, that was shown to contain important predictive information. Columns (3) and (4) include leverage variables to account for the asymmetric response of variance to past negative returns.

Table 4. PREDICTABILITY REGRESSIONS

The table shows the estimates of predictability regressions. The dependent variable is the return on the market portfolio 4-weeks ahead. The explanatory variables are: i) $\widehat{\sigma}_t^2$, the expected variance on the next 8 weeks estimated by best model following the General-to-Specific selection method; ii) $IVar_t$, the expected implied variance on the next 8 weeks estimated from prices of options contracts at time t; iii) $IVar_t - \widehat{\sigma}_t^2$, the variance premium; iv) γ_t is the risk aversion computed using the functional form in Bollerslev et al. (2009) and the variance premium; and v) $\log(D_t/P_t)$, the log dividend yield. Regressions are based on weekly observations. To account for error correlation, the standard errors are computed using Newey-West lags. The standard errors are reported in parenthesis. The corresponding p-values are denoted by * if $p < 0.10$, ** if $p < 0.05$ and *** if $p < 0.01$.

| | 4 weeks | | | | | | | |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\widehat{\sigma}_t^2$	0.000				-0.002			
	(0.005)				(0.004)			
$IVar_t$		0.004*				0.002		
		(0.002)				(0.002)		
$IVar_t - \widehat{\sigma}_t^2$			0.089***				0.066**	
			(0.032)				(0.031)	
γ_t				0.180***				0.135**
				(0.057)				(0.059)
$\log(D/P)$					14.928***	12.684***	11.449***	11.303***
					(5.482)	(6.169)	(6.157)	(6.132)
Constant	0.589	-1.854	-0.314	-4.228***	47.919***	38.601*	35.594*	32.204
	(2.349)	(1.345)	(0.579)	(1.538)	(17.131)	(19.719)	(19.315)	(19.956)
Number of Obs.	175	175	175	175	175	175	175	175
R^2	0.000	0.044	0.082	0.086	0.096	0.108	0.135	0.137
Adjusted R^2	-0.001	0.038	0.077	0.081	0.086	0.098	0.125	0.127

Table 5. PREDICTABILITY REGRESSIONS AT DIFFERENT HORIZONS WITH VARIANCE PREMIUM

The table shows regressions having future market returns 1-week ahead, 2-, 3- and 4-weeks ahead as the dependent variable. The explanatory variables are: i) $IVar_t - \widehat{\sigma}^2_t$ is the *ex-ante* volatility premium; and ii) $\log(D_t/P_t)$ is the log dividend yield. Regressions in columns (1) through (8) are based on weekly observations. Regressions in columns (9) and (10) are non-overlapping on the dependent variable and are based on monthly observations. To account for error correlation, standard errors in columns (3) through (8) are computed using Newey-West lags. The standard errors reported in parenthesis. The corresponding *p*-values are denoted by * if $p < 0.10$, ** if $p < 0.05$ and *** if $p < 0.01$.

	1 week		2 weeks		3 weeks		4 weeks		3 weeks N-O	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$IVar_t - \widehat{\sigma}^2_t$	0.028*	0.023	0.038*	0.026	0.059*	0.043	0.089**	0.066*	0.136**	0.112*
	(0.017)	(0.018)	(0.021)	(0.020)	(0.030)	(0.028)	(0.032)	(0.031)	(0.050)	(0.063)
$\log(D/P)$		3.067		6.102*		7.939		11.449*		6.032
		(2.399)		(3.571)		(5.385)		(6.157)		(9.099)
Constant	−0.129	9.487	−0.095	19.042*	−0.191	24.707	−0.314	35.594*	−0.431	18.508
	(0.208)	(7.559)	(0.338)	(11.188)	(0.462)	(16.933)	(0.579)	(19.315)	(0.730)	(28.585)
Number of Obs.	178	178	177	177	176	176	175	175	41	41
R^2	0.033	0.047	0.028	0.056	0.047	0.079	0.082	0.135	0.135	0.149
Adjusted R^2	0.028	0.036	0.023	0.045	0.041	0.067	0.077	0.125	0.113	0.104

Table 6. PREDICTABILITY REGRESSIONS AT DIFFERENT HORIZONS WITH RISK AVERSION

The table shows regressions having future market returns 1-week ahead, 2-, 3- and 4-weeks ahead as the dependent variable. The explanatory variables are: i) γ_t is the risk aversion computed using the functional form in Bollerslev et al. (2009); and ii) $\log(D_t/P_t)$ is the log dividend yield. Regressions in columns (1) through (8) are based on weekly observations. Regressions in columns (9) and (10) are non-overlapping on the dependent variable and are based on monthly observations. To account for error correlation, standard errors in columns (3) through (8) are computed using Newey-West lags. The standard errors reported in parenthesis. The corresponding p-values are denoted by * if $p < 0.10$, ** if $p < 0.05$ and *** if $p < 0.01$.

	1 week		2 weeks		3 weeks		4 weeks		3 weeks N-O	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
γ_t	0.052	0.039	0.078*	0.054	0.128**	0.097*	0.180***	0.135**	0.238***	0.198*
	(0.033)	(0.035)	(0.043)	(0.043)	(0.051)	(0.052)	(0.057)	(0.059)	(0.087)	(0.106)
$\log(D/P)$		3.185		6.018*		7.601		11.303*		6.631
		(2.336)		(3.587)		(5.332)		(6.132)		(8.894)
Constant	-1.233	9.026	-1.804	17.593	-3.010**	21.491	-4.228***	32.204	-5.534**	16.142
	(0.875)	(7.711)	(1.164)	(11.540)	(1.340)	(17.328)	(1.538)	(19.956)	(2.271)	(29.341)
Number of Obs.	178	178	177	177	176	176	175	175	41	41
R^2	0.027	0.042	0.030	0.057	0.056	0.086	0.086	0.137	0.135	0.153
Adjusted R^2	0.022	0.031	0.025	0.046	0.051	0.076	0.081	0.127	0.113	0.109

Table 7. Robustness: Different variance models

The table shows the estimates of different models of expected variance. The dependent variable is the realized variance over the following 42 days, calculated from 5-minute returns on the Ibovespa portfolio. The explanatory variables are: i) $IVar_t$ is the expected implied variance on the next 8 weeks estimated from prices of options contracts at time $t-1$; ii) $RV_{t-1}^{(k)}$ is the realized volatility on the following k days at time $t-1$, where $k = 42$, 21, 5, 1; computed iii) $Lev_{t-1}^{(k)}$ is the cumulative negative 5-minute returns continuous component of the realized variance on the following k days at time $t-1$, where $k = 42$, 21, 5, 1. The standard errors reported in parenthesis. The corresponding p-values are denoted by $*$ if $p < 0.10$, $**$ if $p < 0.05$ and $***$ if $p < 0.01$.

	M1	M2	M3	M4
$RV_t^{(42)}$	−0.113 $*$	−0.166 $***$		1.273 $***$
$RV_t^{(21)}$	0.256 $***$	0.247 $***$		−0.532 $***$
$RV_t^{(5)}$	0.139 $***$	0.069 $*$		−0.009
$RV_t^{(1)}$	0.021	0.011		0.000
$IVar_t$		0.214 $***$	0.218 $***$	0.194 $***$
$Lev_t^{(42)}$			−0.513 $***$	−1.837 $**$
$Lev_t^{(21)}$			0.484 $***$	1.031 $***$
$Lev_t^{(5)}$			0.091 $***$	0.101
$Lev_t^{(1)}$			0.029	0.031
Constant	324.831 $***$	271.370 $***$	251.331 $***$	691.146 $***$
Number of Obs.	762	741	741	741
R^2	0.209	0.228	0.277	0.369
Adjusted R^2	0.205	0.223	0.272	0.361
RMQE	201.504	199.543	193.193	180.900
BIC	10,276.7	9,985.2	9,937.2	9,862.2

Table 8. ROBUSTNESS: DIFFERENT VARIANCE MODELS (CONT.)

The table shows the estimates of different models of expected realized variance. The dependent variable is the realized variance over the following 8-weeks, calculated from 5-minute returns on the Ibovespa portfolio. The explanatory variables are: i) $IVar_{t-1}$ is the expected implied variance on the next 8 weeks estimated from prices of options contracts at time $t-1$; ii) $C_{t-1}^{(k)}$ is the continuous component of the realized variance during the following k days at time $t-1$, where $k = 42, 21, 5, 1$; iii) $J_{t-1}^{(k)}$ is the jump component of the realized variance during the following k days at time $t-1$, where $k = 42, 21, 5, 1$; and iv) $Lev_{t-1}^{(k)}$ is the absolute of the sum 5-minute negative returns during the following k days at time $t-1$, where $k = 42, 21, 5, 1$. The standard errors reported in parenthesis. The regressions are based on daily observations. The corresponding p-values are denoted by $*$ if $p < 0.10$, $**$ if $p < 0.05$ and $***$ if $p < 0.01$.

	M5	M6	M7	M8
$IVar_{t-1}$	0.147***	0.142***		0.286***
$C_{t-1}^{(42)}$	3.282***	3.679***		−1.675***
$C_{t-1}^{(21)}$	−0.514	−1.171***		1.545***
$C_{t-1}^{(5)}$	−0.461***	−0.225*		0.010
$C_{t-1}^{(1)}$	−0.057	−0.025		−0.015
$J_{t-1}^{(42)}$	0.294		1.583***	1.308***
$J_{t-1}^{(21)}$	−0.540**		−0.768***	−1.126***
$J_{t-1}^{(5)}$	0.247***		0.052	0.163
$J_{t-1}^{(1)}$	0.021		0.007	0.026
$Lev_{t-1}^{(42)}$	−2.433***	−2.535***	−1.289***	
$Lev_{t-1}^{(21)}$	1.023***	1.178***	0.802***	
$Lev_{t-1}^{(5)}$	0.263***	0.211***	0.077	
$Lev_{t-1}^{(1)}$	0.043	0.043	0.021	
$IVol_{t-1}$			13.969***	
Constant	930.537***	935.469***	337.173***	256.965***
Number of Obs.	741	741	741	741
R^2	0.406	0.397	0.349	0.269
Adjusted R^2	0.396	0.390	0.341	0.260
RMQE	175.994	176.881	183.726	194.766
BIC	9,843.8	9,828.9	9,885.2	9,971.7

In Table 8 we separate the realized variance into its continuous and jump components and use these variables instead. Column (1) shows the estimates of the GUM model, the starting model in the General-to-Specific selection method adopted in section 3.1. The GUM regression includes all the variables initially selected as candidate variables to forecast variance. Columns (2) through (4) are variants of this more general model.

As we can conclude by comparing the statistical properties of each regression in tables 3, 7 and 8, the GETS model has the lowest information criterion, BIC, as the selection method strongly penalizes the inclusion of variables and favors a more parsimonious model. Models M4, M5 and M6 have comparable R^2 to the GETS models, explaining more than 35% of the variation of the dependent variable, but with the inclusion of extra regressors.

We now assess how sensitive is our predictive regression to the selection of the variance model. For each one of the regression models shown in tables 7 and 8 we calculate a volatility premium as in equation (3). The results of the predictability regressions at the 4-week return horizon are shown in Table 9. In Column (1) we use a simple model to predict future variance and set $\widehat{\sigma}_t^2 = \sigma_{t-1}^2$ following the definition of Bollerslev et al. (2009). Column (2) replicates our main regression that uses the GETS model to predict variance. Columns (3) through (10) show the predictability regressions for each of the 8 models presented in tables 7 and 8. As can be seen, the results are largely robust to the selection of the variance model.

4. CONCLUSION

This is the first article to propose an implied volatility index for the Brazilian stock market based on option and futures prices traded locally. The methodology we propose has to deal with the relatively low liquidity of contracts used. This is a first contribution of this paper.

We use our implied volatility index to calculate the so-called variance premium for Brazil. Assuming Bollerslev et al. (2009) economic structure, we also pin down a time-varying risk aversion measure of the representative investor in the Brazilian market. In line with international evidence, we show the variance premium strongly predicts future stock returns. Interestingly, we also find that our measure of risk aversion is a strong predictor of future returns with a slightly superior fit than the variance premium. To the best of our knowledge, this is the first analysis of this kind for an emerging market. This is the second contribution of this paper.

Further extensions of this work include applying our methodology to construct implied volatility indices for other markets with low liquidity. With respect to the risk aversion measure, different economic models and parameter calibration can be tested.

Table 9. Robustness: Predictability regression

The table shows the estimates of predictability regressions having the 4-week ahead return as the dependent variable and using various measures of variance premium as regressors. Each measure of variance premium is calculated with a different model for expected variance: $\widehat{\sigma}^2_t = \sigma^2_{t-42}$, and was proposed by Bollerslev et al. (2009). Column (1) uses a simple model for expected variance. Columns (2) replicates our main regression that uses the GETS model to forecast variance. Columns (3) through (10) varies the expected variance model from M1 to M10. $\log(D_t/P_t)$ is the log dividend yield. Regressions are based on weekly observations. To account for error correlation, the standard errors are computed using Newey-West lags. The standard errors are reported in parenthesis. The corresponding p-values are denoted by * if $p < 0.10$, ** if $p < 0.05$ and *** if $p < 0.01$.

		GETS									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Constant	50.656***	35.594*	37.988**	38.108**	37.454*	33.082*	34.887*	34.788*	33.337*	37.230*
	(17.423)	(19.315)	(18.852)	(19.122)	(19.386)	(19.675)	(19.394)	(19.569)	(19.601)	(19.185)
$\log(D/P)$	16.234***	11.449*	12.188**	12.235**	12.007*	10.675*	11.233*	11.201*	10.751*	11.964*
	(5.583)	(6.157)	(6.021)	(6.098)	(6.179)	(6.268)	(6.181)	(6.235)	(6.245)	(6.114)
$IVar_t - \sigma^2_{t-1}$	0.005***									
	(0.002)									
$IVar_t - \widehat{\sigma}^2_t$		0.007**								
		(0.003)								
$IVar_t - \widehat{\sigma}^2_t$ (M1)			0.005**							
			(0.002)							
$IVar_t - \widehat{\sigma}^2_t$ (M2)				0.005*						
				(0.003)						
$IVar_t - \widehat{\sigma}^2_t$ (M3)					0.004*					
					(0.002)					
$IVar_t - \widehat{\sigma}^2_t$ (M4)						0.006**				
						(0.003)				
$IVar_t - \widehat{\sigma}^2_t$ (M5)							0.006**			
							(0.003)			
$IVar_t - \widehat{\sigma}^2_t$ (M6)								0.006**		
								(0.003)		
$IVar_t - \widehat{\sigma}^2_t$ (M7)									0.006**	
									(0.003)	
$IVar_t - \widehat{\sigma}^2_t$ (M8)										0.005**
										(0.003)
Number of Obs.	175	175	175	175	175	175	175	175	175	175
R^2	0.015	0.135	0.127	0.119	0.114	0.147	0.140	0.138	0.143	0.126
Adjusted R^2	0.140	0.125	0.116	0.109	0.104	0.137	0.130	0.128	0.133	0.116

REFERENCES

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *Review of Economics and Statistics*, *89*(4), 701–720. doi: 10.1162/rest.89.4.701

Araújo, E. (2005). Avaliando três especificações para o fator de desconto estocástico através da fronteira de volatilidade de Hansen e Jagannathan: Um estudo empírico para o Brasil. *Pesquisa e Planejamento Econômico*, *35*(1). Retrieved from http://repositorio.ipea.gov.br/handle/11058/4395

Bansal, R., & Yaron, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance*, *59*(4), 1481–1509. doi: 10.1111/j.1540-6261.2004.00670.x

Barndorff-Nielsen, O. E., & Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, *2*(1), 1–37. doi: 10.1093/jjfinec/nbh001

Bekaert, G., & Hoerova, M. (2014). The VIX, the variance premium and stock market volatility. *Journal of Econometrics*, *183*(2), 181–192. doi: 10.1016/j.jeconom.2014.05.008

Bollerslev, T., Gibson, M., & Zhou, H. (2011). Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Journal of Econometrics*, *160*(1), 235–245. doi: 10.1016/j.jeconom.2010.03.033

Bollerslev, T., Marrone, J., Xu, L., & Zhou, H. (2014). Stock return predictability and variance risk premia: Statistical inference and international evidence. *Journal of Financial and Quantitative Analysis*, *49*(3), 1–50. doi: 10.1017/S0022109014000453

Bollerslev, T., Tauchen, G., & Zhou, H. (2009). Expected stock returns and variance risk premia. *Review of Financial Studies*, *22*(11), 4463–4492. doi: 10.1093/rfs/hhp008

Brostowicz Jr., R. J., & Laurini, M. P. (2010). Swaps de variância na BM&F: Apreçamento e viabilidade de hedge [Variance swaps in BM&F: Pricing and viability of hedge]. *Brazilian Review of Finance*, *8*(2), 197–228.

Carr, P., & Wu, L. (2006). A tale of two indices. *The Journal of Derivatives*, *13*(3), 13–29. doi: 10.3905/jod.2006.616865

Carr, P., & Wu, L. (2009). Variance risk premiums. *Review of Financial Studies*, *22*(3), 1311–1341. doi: 10.1093/rfs/hhn038

CBOE–Chicago Board Options Exchange. (2009). *CBOE Volatility Index – VIX: The powerful and flexible trading and risk management tool from the Chicago Board Options Exchange* [White Paper].

Chen, X., & Ghysels, E. (2011). News—good or bad—and its impact on volatility predictions over multiple horizons. *Review of Financial Studies*, *24*(1), 46–81. doi: 10.1093/rfs/hhq071

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196. doi: 10.1093/jjfinec/nbp001

Corsi, F., & Renò, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics*, *30*(3), 368–380. doi: 10.1080/07350015.2012.663261

Dario, A. D. G. (2006). Apreçamento de ativos referenciados em volatilidade [Pricing volatility referenced assets]. *Brazilian Review of Finance*, *4*(2), 203–228. Retrieved from http://bibliotecadigital.fgv.br/ojs/index.php/rbfin/article/view/1162/363

Fajardo, J., Ornelas, J. R. H., & Farias, A. R. d. (2012). Estimating risk aversion, risk-neutral and real-world densities using Brazilian Real currency options. *Economia Aplicada*, *16*(4), 567–577.

Havranek, T., Horvath, R., Irsova, Z., & Rusnak, M. (2015). Cross-country heterogeneity in intertemporal substitution. *Journal of International Economics*, *96*(1), 100–118. doi: 10.1016/j.jinteco.2015.01.012

Hendry, D. F., Castle, J., & Shephard, N. (2009). *The methodology and practice of econometrics: A festschrift in honour of David F. Hendry*. Oxford University Press.

Issler, J. V., & Piqueira, N. S. (2000). Estimating relative risk aversion, the discount rate, and the intertemporal elasticity of substitution in consumption for Brazil using three types of utility function. *Brazilian Review of Econometrics*, *20*(2), 201–239.

Kapotas, J. C., Schirmer, P. P., & Manteiga, S. M. (2004). Apreçamento de contratos de volatilidade a termo no mercado brasileiro [Forward volatility contract pricing in the Brazilian market]. *Brazilian Review of Finance*, *2*(1), 1–21. Retrieved from http://bibliotecadigital.fgv.br/ojs/index.php/rbfin/article/view/1133

Ornelas, J. R. H. (2016, June). *Expected currency returns and volatility risk premia*. doi: 10.2139/ssrn.2809141

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics*, *54*(3), 375–421. doi: 10.1016/S0304-405X(99)00041-0

Zhou, H. (2009, May). *Variance risk premia, asset predictability puzzles, and macroeconomic uncertainty*. Retrieved from http://papers.ssrn.com/abstract=1400049

APPENDIX.

The j adjustment

In this section we demonstrate how to obtain the adjustment term j. In the following derivations we refer to an out-of-the-money option as OTM, and to an in-the-money option as ITM.

Under the risk neutral measure, it can be shown that the variance is approximated by a portfolio of OTM calls and puts. However, in practice, the portfolio used is

$$\sigma^2(t) \approx \frac{2}{T-t} \sum_i \frac{\Delta K_i}{K_i^2} e^{r_t(T-t)} O_t(K_i), \tag{A-1}$$

where

- K_i is the strike of the i-th out-of-the-money option: a call if $K_i > K_0$, a put if $K_i < K_0$ and both if $K_i = K_0$;

- K_0 is the strike closest to the futures price F;

- $\Delta K_i = \frac{1}{2}(K_{i+1} - K_{i-1})$;

- r_t is risk-free rate from day t to day T, obtained from the daily settlement price of the futures interbank rate (DI);

- $O_t(K_i)$ is the market price on day t of option with strike K_i.

Since we don't necessarily have a call and a put at K_0, an adjustment in the formula above is needed. The following 6 cases can arise:

Case 1: If $K_0 \leq F$ and we have data on call and put prices at K_0.

This is the standard case set by Carr & Wu (2006). It follows from the Put-Call parity that:

$$O(K_0) = \frac{P(K_0) + C(K_0)}{2} = \frac{P(K_0) + \left(P(K_0) + (F - K_0)e^{-r(T-t)}\right)}{2}.$$

Therefore, substituting for the $O(K_0)$ term in Equation (A-1), we obtain

$$\frac{2}{T-t}\frac{\Delta K_0}{K_0^2}e^{r_t(T-t)}O(K_0) = 2\frac{\Delta K_0}{K_0^2}\frac{P(K_0)}{T-t}e^{r_t(T-t)} + \frac{1}{T-t}\frac{\Delta K_0}{K_0^2}(F-K_0)$$

$$= 2\frac{\Delta K_0}{K_0^2}\frac{P(K_0)}{T-t}e^{r_t(T-t)} + \frac{1}{T-t}\left(\frac{F}{K_0}-1\right)^2,$$

where, the last equality, follows from the assumption that $\Delta K_0 = F - K_0$.

Substituting back in Equation (A-1) we obtain that the last term below is zero:

$$\sigma^2(t) = \frac{2}{T-t}\sum_i \frac{\Delta K}{K_i^2}e^{r(T-t)}O_t(K_i) + \underbrace{\frac{1}{T-t}\frac{\Delta K_0}{K_0^2}(F-K_0) - \frac{1}{T-t}\left(\frac{F}{K_0}-1\right)^2}_{=0},$$

where at $i=0$ we have $O(K_0) = P(K_0)$, that is, all options are OTM.

Equivalently, we can write the above equation as

$$\sigma^2(t) = \frac{2}{T-t}\sum_i \frac{\Delta K}{K_i^2}e^{r(T-t)}O_t(K_i) - \frac{1}{T-t}\left(\frac{F}{K_0}-1\right)^2,$$

where $O(K_0) = \frac{P(K_0)+C(K_0)}{2}$ and $C(K_0)$ is ITM.

In Brazil, there are days when only a call or a put at K_0 is traded. Besides, we have to define K_0 as the option strike value which is closest to $F(t,T_k)$ and, because of that, we may have either $K_0 > F(t,T_k)$ or $K_0 < F(t,T_k)$. Given that, we have to create the following 5 additional cases.

Case 2: If $F < K_0$ and we have data on call and put prices at K_0.

In this case, $P(K_0)$ is ITM and, by the Put-Call parity, we obtain analogously:

$$\sigma^2(t) = \frac{2}{T-t}\sum_i \frac{\Delta K}{K_i^2}e^{r(T-t)}O_t(K_0) + \frac{1}{T-t}\frac{\Delta K_0}{K_0^2}(F-K_0) - \frac{1}{T-t}\left(\frac{F}{K_0}-1\right)^2,$$

where $O(K_0) = C(K_0)$, that is, all options are OTM.

Equivalently,

$$\sigma^2(t) = \frac{2}{T-t}\sum_i \frac{\Delta K}{K_i^2}e^{r(T-t)}O_t(K_i) - \frac{1}{T-t}\left(\frac{F}{K_0}-1\right)^2,$$

where $O(K_0) = \frac{P(K_0)+C(K_0)}{2}$ and $P(K_0)$ is ITM.

Case 3: If $K_0 \le F$, we have data on put prices and don't have data on call prices at K_0.

In this case, all options are OTM and no adjustment is needed. That is, we set $j=0$ in the formula:

$$\sigma^2(t) = \frac{2}{T-t}\sum_i \frac{\Delta K_i}{K_i^2}e^{r_t(T-t)}O_t(K_i) - \frac{j}{T-t}\left(\frac{F}{K_0}-1\right)^2,$$

where $O(K_0) = P(K_0)$.

Case 4: If $K_0 > F$, we have data on call prices and don't have data on put prices at K_0.

In this case, all options are OTM and no adjustment is needed. That is, we set $j=0$ in the formula:

$$\sigma^2(t) = \frac{2}{T-t}\sum_i \frac{\Delta K_i}{K_i^2}e^{r(T-t)}O(K_i) - \frac{j}{T-t}\left(\frac{F}{K_0}-1\right)^2,$$

where $O(K_0) = C(K_0)$.

Case 5: If $K_0 \leq F$, we have data on call prices and don't have data on put prices at K_0.

In this case, $C(K_0)$ is ITM and should be transformed into a OTM $P(K_0)$ by the Put-Call parity. Using the result $O(K_0) = C(K_0) = P(K_0) + (F - K_0)e^{-r(T-t)}$, and substituting for the $O(K_0)$ term in equation (A-1) we obtain

$$\frac{2}{T-t}\frac{\Delta K_0}{K_0^2}e^{r_t(T-t)}O(K_0) = \frac{2}{T-t}\frac{\Delta K_0}{K_0^2}P(K_0)e^{r_t(T-t)} + \frac{2}{T-t}\frac{\Delta K_0}{K_0^2}(F - K_0).$$

Following the same steps of Case 1, we obtain

$$\sigma^2(t) = \frac{2}{T-t}\sum_i \frac{\Delta K}{K_i^2}e^{r(T-t)}O_t(K_i) + \underbrace{\frac{2}{T-t}\frac{\Delta K_0}{K_0^2}(F - K_0) - \frac{2}{T-t}\left(\frac{F}{K_0} - 1\right)^2}_{=0},$$

where $Q(K_0) = P(K_0)$, that is, all options are OTM.

Equivalently,

$$\sigma^2(t) = \frac{2}{T-t}\sum_i \frac{\Delta K_i}{K_i^2}e^{r(T-t)}O_t(K_i) - \frac{j}{T-t}\left(\frac{F}{K_0} - 1\right)^2,$$

where now we have $j = 2$, $O(K_0) = C(K_0)$, and $C(K_0)$ is ITM.

Case 6: If $K_0 > F$, we have data on put prices and don't have data on call prices at K_0.
This can be solved similarly as Case 5 with $j = 2$ and $P(K_0)$ ITM.