

Analysis of market volatility via a dynamically purified option price process*

Chuong Luong, Nikolai Dokuchaev

Department of Mathematics & Statistics, Curtin University,
GPO Box U1987, Perth, 6845 Western Australia

Submitted: October 1st, 2014. Revised: November 4, 2014

Abstract

The paper studies methods of dynamic estimation of volatility for financial time series. We suggest to estimate the volatility as the implied volatility inferred from some artificial ‘dynamically purified’ price process that in theory allows to eliminate the impact of the stock price movements. The complete elimination would be possible if the option prices were available for continuous sets of strike prices and expiration times. In practice, we have to use only finite sets of available prices. We discuss the construction of this process from the available option prices using different methods. In order to overcome the incompleteness of the available option prices, we suggest several interpolation approaches, including the first order Taylor series extrapolation and quadratic interpolation. We examine the potential of the implied volatility derived from this proposed process for forecasting of the future volatility, in comparison with the traditional implied volatility process such as the volatility index VIX.

Keywords: econometrics, implied volatility, volatility index, approximation of missing data, dynamic forecasting, purified option prices.

MSC classification: 91G70

1 Introduction

In this paper, we study the statistics of option prices and implied volatilities of financial time series. Since the price of options depends on market prognosis for the future volatility, the option

*Accepted to Annals of Financial Economics. This is a pre-copy-editing, author-produced PDF of an article. The definitive publisher-authenticated version will be available online at <http://www.worldscientific.com/worldscinet/afe>

prices and the implied volatilities are used as important indicators of market sentiment about the future volatility. The implied volatilities are usually defined as the inverse of the Black-Scholes [1] pricing formula applied to the observed market prices; given fixed and known asset price, strike price, future interest rate, and time-to-maturity, the implied volatility is uniquely defined by the option price. In fact, the option prices and the implied volatilities are fluctuating along with the underlying assets prices; they have a stochastic “random walk” type pattern of movement, similar to the stock prices. Moreover, the implied volatility depends on the strike price and the expiration time. Therefore, one would find it difficult to decide which implied volatility to use among all possible versions of the implied volatility when describes the market expectations on the degrees of the future stock price deviations. For this purpose, the volatility indexes were created. For instance, the volatility index VXO on the Chicago Board of Options Exchange (CBOE) [26] and the AVX on Australian Securities Exchange (ASX) [15] used the Black-Scholes-Merton’s framework as the underlying model to construct the implied volatility for S&P 100 and S&P 200 respectively. The implied volatility of these indexes used options at different strike prices and maturity dates to approximate the at-the-money implied volatility. However, this approach has some limitations. It is restricted by the assumptions made in the Black-Scholes’ model and has an artificially induced upward bias [6]. With those limitations, Carr and Wu [7] introduced an alternative method for constructing the implied volatility using a model-free approach, which was then used by CBOE for constructing the volatility index VIX for S&P 500. The VIX index is constructed from the price of a portfolio including a number of out-the-money option prices that varies every day, depending on the number of options with non-zero bid price. While this implied volatility has a better known economic interpretation [7], the construction of the VIX process is complex and requires large samples of option data.

Motivated by these volatility processes, we suggest a modification in the approaches used for VXO, AVX, and VIX indexes with an aim to reduce the measurement errors and improve the computational robustness. We suggest to consider a ”dynamically purified” option price process such that impact of stock price movements is reduced. This helps to separate the impact of the stock price movements from the changes in the market forecast of the future volatility. In effect, the implied volatility calculated form this process could be more informative than traditionally calculated implied volatility, similarly to the popular volatility indexes such as VXO and VIX indexes.

In theory, “dynamically purified” option price process eliminates the impact of the stock price movements. However, this would be possible if the option prices were available for continuous sets of strike prices and expiration times. In practice, we have to use only finite sets of available prices. In order to restore this process from incomplete sets of available option prices, we suggest

to use a similar approach to the approach implemented in the calculation of the volatility index. Here, the implied volatilities from the missing option prices was replaced by linear combinations of implied volatilities using some observable options. However, instead of applying the linear interpolation on the implied volatility, we interpolate the missing options prices. In this paper, we discuss the use of both the first order Taylor series interpolation and quadratic interpolation. For this approach, the 'dynamically purified' option price process can be constructed using 18 observed option prices.

We study the statistical properties of the proposed process by using the S&P/ASX 200 Index Options data for the period from 1st January 2010 to 31st December 2012. For demonstration purposes, we consider only a special case of this process which represents the at-the-money implied volatility. It is found before that the classical volatility index VXO and VIX feature strongly negative correlations with the index return increments ([15], [26]). Our finding shows that the "dynamically purified" option price process has the same feature. In addition, the implied volatility of the purified option price process has a very strong positive correlation with the implied volatility index. This is a interesting observation given that VIX process is calculated using very different data and methods.

As a possible application of the proposed process, we consider the forecast of future volatility. We establish a number of regression models using the implied volatility index VIX and the implied volatility from the purified option prices as predictors of the future volatility. We find that for our selected dataset and constructed models, the forecasting ability of the new implied volatility is superior to that of the implied volatility index. Since calculation of the proposed process requires less option prices than calculation of the existing implied volatility index VIX, this process can be used as an alternative for VIX in some cases when there is no sufficient data to calculate VIX.

2 The model

We consider the diffusion model of a securities market consisting of a risk free bond or bank account with the price $B(t)$, $t \geq 0$, and a risky stock with the price $S(t)$, $t \geq 0$. The prices of the stocks evolve as

$$dS(t) = S(t) (a(t)dt + \sigma(t)dw(t)), \quad t > 0, \quad (2.1)$$

where $w(t)$ is a Wiener process, $a(t)$ is an appreciation rate, $\sigma(t)$ is a random volatility coefficient. The initial price $S(0) > 0$ is a given deterministic constant. The price of the bond evolves as

$$B(t) = \exp\left(\int_0^t r(s)ds\right)B(0), \quad (2.2)$$

where $r(t) \geq 0$ is a random process and $B(0)$ is given.

We assume that $w(\cdot)$ is a standard Wiener process on a given standard probability space $(\Omega, \mathcal{F}, \mathbf{P})$, where Ω is a set of elementary events, \mathcal{F} is a complete σ -algebra of events, and \mathbf{P} is a probability measure.

Let \mathcal{F}_t be a filtration generated by the currently observable data. We assume that the process $(S(t), \sigma(t))$ is \mathcal{F}_t -adapted and that \mathcal{F}_t does not depend on $\{w(t_2) - w(t_1)\}_{t_2 \geq t_1 \geq t}$. In particular, this means that the process $(S(t), \sigma(t))$ is currently observable and $\sigma(t)$ does not depend on $\{w(t_2) - w(t_1)\}_{t_2 \geq t_1 \geq t}$. We assume that \mathcal{F}_0 is the P -augmentation of the set $\{\emptyset, \Omega\}$, and that $a(t)$ does not depend on $\{w(t_2) - w(t_1)\}_{t_2 \geq t_1 \geq t}$. For simplicity, we assume that $a(t)$ is a bounded process.

The Black-Scholes price

Let $K > 0$ be given. We shall consider two types of options: vanilla call and vanilla put, with payoff function $f(S(T)) = F(S(T), K)$, where $F(S(T), K) = (S(T) - K)^+$ or $F(S(T), K) = (K - S(T))^+$, respectively. Here K is the strike price.

Let $T > 0$ be fixed. Let $H_{BS,c}(t, x, \sigma, r, K)$ and $H_{BS,p}(t, x, \sigma, r, K)$ denote Black-Scholes prices for the vanilla put and call options with the payoff functions $F(S(T), K)$ described above under the assumption that $S(t) = x$, $(\sigma(s), r(s)) = (\sigma, r)$ ($\forall s > t$), where $\sigma \in (0, +\infty)$ is non-random. The Black-Scholes formula for a call option can be rewritten as

$$\begin{aligned} H_{BS,c}(t, x, \sigma, r, K) &= x\Phi(d_+(t, x, \sigma, r, K)) - Ke^{-r(T-t)}\Phi(d_-(t, x, \sigma, r, K)), \\ H_{BS,p}(t, x, \sigma, r, K) &= H_{BS,c}(t, x, \sigma, r, K) - x + Ke^{-r(T-t)}, \end{aligned} \quad (2.3)$$

where

$$\Phi(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{s^2}{2}} ds,$$

and where

$$\begin{aligned} d_+(x, t, \sigma, r, K) &\triangleq \frac{\log(x/K) + (T-t)r}{\sigma\sqrt{T-t}} + \frac{\sigma\sqrt{(T-t)}}{2}, \\ d_-(x, t, \sigma, r, K) &\triangleq d_+(x, t, \sigma, r, K) - \sigma\sqrt{(T-t)}. \end{aligned} \quad (2.4)$$

Set

$$\tilde{S}(t) \triangleq S(t) \exp\left(-\int_0^t r(s)ds\right).$$

The risk neutral pricing

We assume that there exist a risk-neutral measure Q such that the process $\tilde{S}(t)$ is a martingale under Q , i.e., $\mathbf{E}_Q\{\tilde{S}(T) | \mathcal{F}_t\} = \tilde{S}(t)$, where \mathbf{E}_Q is the corresponding expectation.

The local risk minimization method, the mean variance hedging, and some other methods based on the risk-neutral valuation lead to the following pricing rule: given (a, σ, r) , the option price is

$$P_{RN}(t, \sigma(\cdot), r(\cdot)) \triangleq \mathbf{E}_Q\{e^{-\int_t^T r(s)ds} F(S(T)) | \mathcal{F}_t\}, \quad (2.5)$$

where Q is some risk neutral measure, and where \mathbf{E}_Q is the corresponding expectation. Usually, Q is uniquely defined by (a, σ, r) , and by the pricing method used.

For numerical simulation purposes, we assume that we have chosen one of these methods (for instance, local risk minimization method or mean variance hedging). Therefore, the risk neutral measure Q is uniquely defined by (a, σ, r) given the method of pricing.

For brevity, we shall denote by H_{BS} the corresponding Black-Scholes prices different options, i.e., $H_{BS} = H_{BS,c}$ or $H_{BS} = H_{BS,p}$, for vanilla call, vanilla put respectively. Let

$$v(t) \triangleq \frac{1}{T-t} \int_t^T \sigma(s)^2 ds, \quad \rho(t) \triangleq \frac{1}{T-t} \int_t^T r(s) ds.$$

The following lemma from [12] is a generalization for random $r(\cdot)$ of the lemma from Hull and White [19]:

Lemma 2.1 *Let $t \in [0, T)$ be fixed. Let $v(t)$ and $\rho(t)$ be \mathcal{F}_t -measurable. Then*

$$\mathbf{E}_Q\{e^{-\int_t^T r(s)ds} F(S(T)) | \mathcal{F}_t\} = H_{BS}(t, S(t), \sqrt{v(t)}, \rho(t), K).$$

Clearly, $\frac{1}{T-t} \int_t^T \sigma(s)^2 ds$ and $\frac{1}{T-t} \int_t^T r(s) ds$ are not \mathcal{F}_t -measurable in the general case of stochastic (r, σ) , and the assumptions of Lemma 2.1 are not satisfied.

Corollary 2.1 *Assume that $H_{BS} = H_{BS,c}$, or $H_{BS} = H_{BS,p}$, or $H_{BS} = H_{BS,s}$. Consider a market model with pricing rule (2.5). Let (σ, r) does not depend on w under Q . Then $P_{RN}(t) = \mathbf{E}_Q\{H_{BS}(t, S(t), \sqrt{v(t)}, \rho(t), K) | \mathcal{F}_t\}$, where (v, ρ) are defined in Lemma 2.1.*

3 “Purified” option price process

Let us consider dynamically adjusted parameters $T = T(t) = t + \tau$ and $K = K(t) = \kappa S(t)$, where $\kappa \in (0, +\infty)$ and $\tau > 0$ are some parameters, t is the current time. In this case, $F(S(T)) = F(S(T), K) = S(t)F(Y(t + \tau), \kappa)$, where

$$Y(T) = S(t + \tau)/S(t).$$

By rule (2.5), the option price given (a, σ, r) , is

$$\begin{aligned} P_{RN}(t, \sigma(\cdot), r(\cdot)) &\triangleq \mathbf{E}_Q\{e^{-\int_t^T r(s)ds} F(S(T), K) | \mathcal{F}_t\} \\ &= S(t)\mathbf{E}_Q\{e^{-\int_t^{t+\tau} r(s)ds} F(Y(t + \tau), \kappa) | \mathcal{F}_t\}, \end{aligned}$$

where Q is some risk neutral measure, and where \mathbf{E}_Q is the corresponding expectation.

Let

$$G(t) \triangleq \frac{P_{RN}(t, \sigma(\cdot), r(\cdot))}{S(t)}. \quad (3.1)$$

It follows that

$$G(t) = \mathbf{E}_Q\{e^{-\int_t^{t+\tau} r(s)ds} F(Y(t + \tau), \kappa) | \mathcal{F}_t\}. \quad (3.2)$$

Assume that $v(t)$ and $\rho(t)$ are \mathcal{F}_t -measurable in this case,

$$\begin{aligned} H_{BS,c}(t, 1, \sqrt{v(t)}, \rho(t), \kappa_i) &= G_C(t), \\ H_{BS,p}(t, 1, \sqrt{v(t)}, \rho(t), \kappa_i) &= G_P(t), \end{aligned} \quad (3.3)$$

for call and put options respectively.

The observations of option prices with dynamic adjusted strike price $K = \kappa S(t)$ with a fixed κ and a fixed period t can be useful for econometrics purposes even without calculation of the implied parameters. In particular, some features of the evolution law for implied parameters $(\sigma(t), r(t))$ can be restored directly from the observations of the processes $G(t)$. For instance, if $\rho(t)$ is a non-random process then the implied volatility $\sqrt{v(t)}$ can be calculated from (3.3) for call and put options. In addition, as the impact of the stock price movements is damped, one may expect that $G(t)$ is a relatively smooth process. Thus, the study of the process $G(t)$ will be of interest.

Up to the end of this paper, we will assume that $\rho(t)$ is non-random and known. This is an usual assumption since the risk-free rate is relatively stable.

3.1 Parametric approximation for absent option prices

In practice, option prices are available only for finite sets of possible option prices different strikes and time-to-maturity. Therefore, it is not possible to collect the prices $P_i(t)$ of the options at the exact strike prices $K_i = \kappa_i S(t)$ with fixed κ_i and t . In order to study the process $G(t)$ described above, we have to use the prices $\tilde{P}_i(t)$ of the corresponding options with the closest available strike prices $\tilde{K}_i(\tilde{t})$.

From this section onward, let \tilde{P}_C and \tilde{P}_P be the values of call and put options observed on the market.

Delta of the strike

The price change of the option price P with respect to K , when other factors remaining constant, is called the delta of the strike:

$$\Delta K = \frac{\delta P}{\delta K}.$$

From (2.3) and (2.4), the delta of the strike for a call and a put are:

$$\frac{\delta P_C}{\delta K} = -e^{-\tau \Delta t} \Phi(d_2) < 0,$$

and

$$\frac{\delta P_P}{\delta K} = e^{-\tau \Delta t} \Phi(-d_2) > 0,$$

respectively.

Theta of the option (time-decay)

Theta of an option is defined as the rate of change of its price P with respect to time t , while all other factors remaining constant:

$$\Theta = \frac{\delta P}{\delta t}.$$

From (2.3) and (2.4), it can be shown that

$$\theta_C = \frac{\delta P_C}{\delta t} = -\frac{S\sigma}{2\sqrt{\Delta t}} \Phi'(d_1) - \tau K e^{-\tau \Delta t} \Phi(d_2) < 0,$$

and

$$\theta_P = \frac{\delta P_P}{\delta t} = -\frac{S\sigma}{2\sqrt{\Delta t}} \Phi'(d_1) + \tau K e^{-\tau \Delta t} \Phi(-d_2) < 0.$$

Let's assume that one wishes to approximate the missing call option price P_C^* at strike price K^* with time-to-maturity Δt^* . The nearest available strike price is \tilde{K} with $\Delta \tilde{t}$ has a value of \tilde{P}_C . The first order approximation can be used such that:

$$P_C^* \approx \tilde{P}_C - m_1(\tilde{K} - K),$$

where $m_1 = \frac{\delta P_C}{\delta K}$, the delta of the strike for a call option. Thus, P_C^* can be approximated as followed:

$$P_C^*(\tilde{t}) \approx \tilde{P}_C - S e^{-\tau \Delta \tilde{t}} \Phi(d_2)(S - \tilde{K}). \quad (3.4)$$

We then need to adjust this approximated value $P_C^*(\tilde{t})$ in order to match $\Delta \tilde{t}$ with Δt^* . The equation (3.4) can be extended further by using the first-order-approximation on Δt where $m_2 = \theta = -\frac{\delta V}{\delta \Delta t}$, the time-decay of the option.

However, this approach can help to calculate $G(t)$ only for the case of small value $|K - \tilde{K}|$. In addition, one would have to obtain the implied parameters of the underlying option pricing model before estimating the missing option data with this method.

3.2 Non-parametric method and quadratic approximation of absent prices

For the construction of the implied volatility suggested in [26], the linear approximation of the implied volatility was used for determining the implied volatility of at-the-money options. We instead suggest using quadratic approximation for estimating the missing option price data before computing the implied volatility. As fitting the option price surface often leads to numerical difficulties [20], only some available option prices near the targeted κ_i will be used. Now let:

- (i) K^* be the strike price of the missing option;
- (ii) K_{j-1} be the strike price that is the second closest to and below; K^*
- (iii) K_j be the strike price that is the closest to and below K^* ;
- (iv) K_{j+1} be the strike price that is the closest to and just above K^* ;
- (v) K_{j+2} be the strike price that is the second closest to and above K^* ;
- (vi) T_1, T_2 and T_3 be the first-nearby, second-nearby and third-nearby expiration dates.

The selection criteria are as following:

▲ For $|K^* - K_j| < |K^* - K_{j+1}|$:

		1st Strike	2nd Strike	3rd Strike
Call Options	1st nearby	$\tilde{P}_{c,1}^{K_{j-1}}$	$\tilde{P}_{c,1}^{K_j}$	$\tilde{P}_{c,1}^{K_{j+1}}$
	2nd nearby	$\tilde{P}_{c,2}^{K_{j-1}}$	$\tilde{P}_{c,2}^{K_j}$	$\tilde{P}_{c,2}^{K_{j+1}}$
	3rd nearby	$\tilde{P}_{c,3}^{K_{j-1}}$	$\tilde{P}_{c,3}^{K_j}$	$\tilde{P}_{c,3}^{K_{j+1}}$
Put Options	1st nearby	$\tilde{P}_{p,1}^{K_j}$	$\tilde{P}_{p,1}^{K_{j+1}}$	$\tilde{P}_{p,1}^{K_{j+2}}$
	2nd nearby	$\tilde{P}_{p,2}^{K_j}$	$\tilde{P}_{p,2}^{K_{j+1}}$	$\tilde{P}_{p,2}^{K_{j+2}}$
	3rd nearby	$\tilde{P}_{p,3}^{K_j}$	$\tilde{P}_{p,3}^{K_{j+1}}$	$\tilde{P}_{p,3}^{K_{j+2}}$

▲ For $|K^* - K_j| > |K^* - K_{j+1}|$

		1st Strike	2nd Strike	3rd Strike
Call Options	1st nearby	$\tilde{P}_{c,1}^{K_j}$	$\tilde{P}_{c,1}^{K_{j+1}}$	$\tilde{P}_{c,1}^{K_{j+2}}$
	2nd nearby	$\tilde{P}_{c,2}^{K_j}$	$\tilde{P}_{c,2}^{K_{j+1}}$	$\tilde{P}_{c,2}^{K_{j+2}}$
	3rd nearby	$\tilde{P}_{c,3}^{K_j}$	$\tilde{P}_{c,3}^{K_{j+1}}$	$\tilde{P}_{c,3}^{K_{j+2}}$
Put Options	1st nearby	$\tilde{P}_{p,1}^{K_{j-1}}$	$\tilde{P}_{p,1}^{K_j}$	$\tilde{P}_{p,1}^{K_{j+1}}$
	2nd nearby	$\tilde{P}_{p,2}^{K_{j-1}}$	$\tilde{P}_{p,2}^{K_j}$	$\tilde{P}_{p,2}^{K_{j+1}}$
	3rd nearby	$\tilde{P}_{p,3}^{K_{j-1}}$	$\tilde{P}_{p,3}^{K_j}$	$\tilde{P}_{p,3}^{K_{j+1}}$

▲ For $|K^* - K_j| = |K^* - K_{j+1}|$

		1st Strike	2nd Strike	3rd Strike	4th Strike
Call/Put Options	1st nearby	$\tilde{P}_{c,1}^{K_{j-1}}$	$\tilde{P}_{c,1}^{K_j}$	$\tilde{P}_{c,1}^{K_{j+1}}$	$\tilde{P}_{c,1}^{K_{j+2}}$
	2nd nearby	$\tilde{P}_{c,2}^{K_{j-1}}$	$\tilde{P}_{c,2}^{K_j}$	$\tilde{P}_{c,2}^{K_{j+1}}$	$\tilde{P}_{c,2}^{K_{j+2}}$
	3rd nearby	$\tilde{P}_{c,3}^{K_{j-1}}$	$\tilde{P}_{c,3}^{K_j}$	$\tilde{P}_{c,3}^{K_{j+1}}$	$\tilde{P}_{c,3}^{K_{j+2}}$

Since the approximation uses limited data points, we suggest to apply the centring and scaling transformation of the data. This will in turns improve the numerical properties of the quadratic approximation. For example, in order to approximate the option price with strike price K^* by using n nearby available options, we find the coefficients of the quadratic equation in:

$$\hat{K}^* = \frac{K^* - \mu_K}{\sigma_K},$$

where

$$\mu_K = \frac{1}{n} \sum_{j=1}^{j=n} K_j, \quad \sigma_K = \sqrt{\frac{1}{n-1} \sum_{j=1}^{j=n} (K_j - \mu_K)^2}.$$

The Algorithm

- Approximate $\tilde{P}_{c,i}^{K^*}$ and $\tilde{P}_{p,i}^{K^*}$, where $i = T_1, T_2, T_3$: for each expiration date, apply the quadratic approximation on option prices at different strike prices to find the approximated option value at K^* ,

- Approximate $\tilde{P}_{c,t^*}^{K^*}$ and $\tilde{P}_{p,t^*}^{K^*}$: for each type of option, apply the quadratic approximation on option prices at different time-to-maturity to find the approximated option value with fixed Δt^* trading-day time horizon, by using the approximated option prices at \tilde{K}_{T_1} , \tilde{K}_{T_2} and \tilde{K}_{T_3} ,
- The G process for call-put options are constructed by the following equations:

$$G_C(t) = \frac{\tilde{P}_{c,t^*}^{K^*}}{S_{(t)}}, \quad G_P(t) = \frac{\tilde{P}_{p,t^*}^{K^*}}{S_{(t)}}.$$

For the rest of this paper, the G process is defined as the average of the dynamically purified call option and put option processes, with

$$G(t) = \frac{G_C(t) + G_P(t)}{2}. \quad (3.5)$$

It is noted that the approximations from our approach are found to fall within the optimal bounds of option prices when using the convex optimization approach suggested by Dimitris and Popescu [11].

3.3 The data

The S&P/ASX 200 index options are traded on the ASX with the underlying asset being the S&P/ASX 200 index. These option contracts was first listed on 31st March, 2001 and are European in exercise style, with quarterly expiry cycles: March, June, September and December. The exercise prices are set at intervals of 25 index points with new exercise prices automatically created as the underlying index oscillates. The S&P/ASX 200 index options are cash settled and the settlement amount is based on the opening prices of the stocks in the underlying index on the morning of the last trading date. Table 4.1 summarises the features of these index options.

The daily data for S&P/ASX 200 Index options and S&P/ASX 200 Index were obtained from SIRCA - the Securities Industry Research Centre of Asia-Pacific [24]. In this experiment, we used the last price for each trading day for both option prices and index level. We reported the statistical summary for the daily return and volatility of the index from January 2010 to December 2012 with 757 observations in Table 4.2. When computing the implied volatility from the G-process, we used the overnight interest rate and bank bill swap (BBSW) rates to interpolate the risk-free rate, with a fixed 22-trading-day time horizon. These rates are obtained from the Reserve Bank of Australia [25].

Table 3.1: S&P/ASX 200 Index Options Features

Underlying asset	ASX approved indexes (currently the ASX 200 Index)
Exercise style	European
Settlement	Cash settled based on the opening prices of the stocks in the underlying index on the morning of the last trading date.
Expiry day	The third Thursday of the month, unless otherwise specified by ASX.
Last trading day	Trading will cease at 12 noon on expiry Thursday. This means trading will continue after the settlement price has been determined.
Premium	Expressed in points
Strike price	Expressed in points
Index multiplier	A specified number of dollars per point e.g. AUD 10
Contract value	The exercise price of the option multiplied by the index multiplier

3.4 Some statistical properties of the purified option price process

We constructed the G-process for the selected period with a fixed time-horizon of 22 trading days ($\Delta t^* = \frac{22}{252}$) and $K^* = S$ ($\kappa = 1$) for the at-the-money options. The available option data were selected as discussed in Section 3.2. Table 3.2 provides a summary of statistics for the purified option price process and their logarithmic series. Figure 3.1 plots the S&P/ASX 200 level against this dynamically purified option price process G . It is noted that the purified option price process has a relatively small standard deviation.

Table 3.2: Summary statistics for at-the-money call/put options price process and their average.

Series	Mean	Std. Dev.	Skew.	Kurt.	Min.	Max.
$ATM_{G_C(t)}$	0.0158	0.0057	1.2610	2.4451	0.0061	0.0435
$ATM_{G_P(t)}$	0.0160	0.0055	1.3886	2.4133	0.0073	0.0422
$ATM_{G_C(t)}$	0.0159	0.0052	1.5093	2.6646	0.0083	0.0382
$\ln(ATM_{G_C(t)})$	-4.2038	0.3406	0.1209	0.0639	-5.0995	-3.135
$\ln(ATM_{G_P(t)})$	-4.1860	0.3113	0.4517	0.0713	-4.9199	-3.1653
$\ln(ATM_{G(t)})$	-4.1863	0.2946	0.6461	0.2068	-4.8036	-3.2649

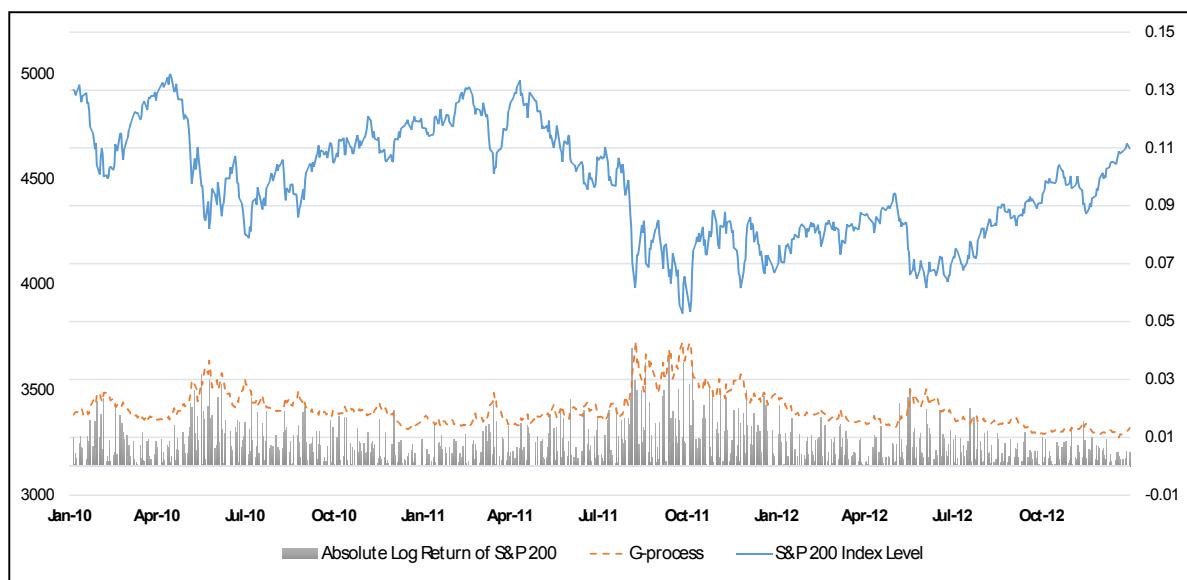


Figure 3.1: The daily S&P/ASX 200 index level with its absolute log return and the at-the-money purified option price process G

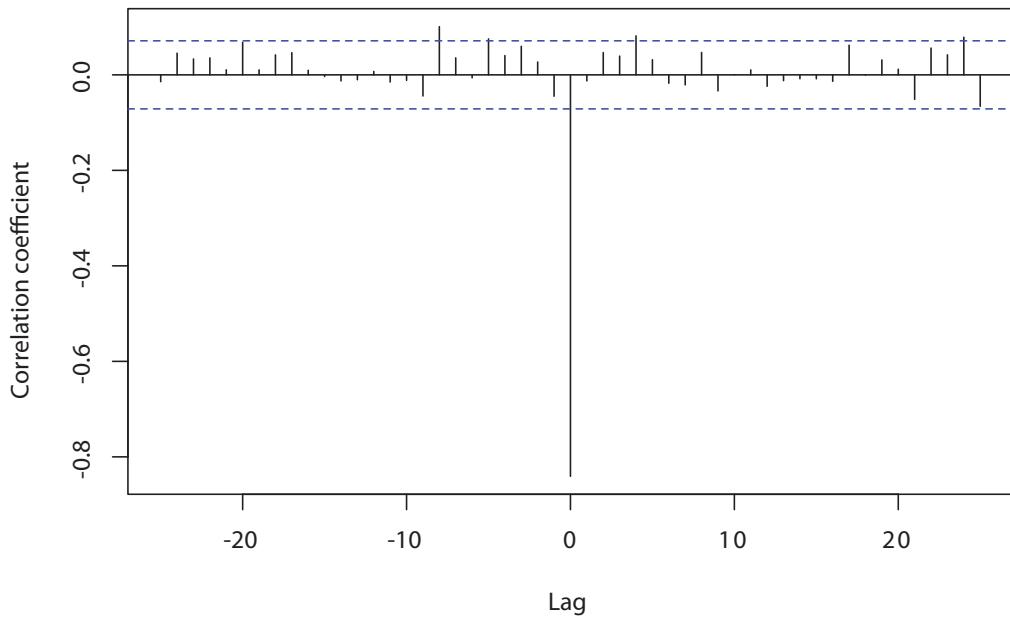


Figure 3.2: Cross-correlations between daily returns of the S&P 200 index level and daily changes in G .

To examine the properties of this new process against the index level, let's define:

- Increments for the dynamic log index level:

$$r_t = \Delta \ln S_t = \ln \frac{S_t}{S_{t-1}},$$

- Increments for the purified option price process:

$$q_t = \Delta G_t = G_t - G_{t-1}.$$

Figure 3.2 plots the cross-correlations between S&P 200 index returns at different leads and lags against daily changes in the G process, with the two dash-dotted lines denoting the 95% confidence band. We observed a strongly negative instantaneous correlation between q_t and r_t . In effect, it was found that $\text{corr}(q_t, r_t) = -0.8410$ for the whole period, while the correlation estimates at other leads and lags are smaller. Also, the statistical standard deviation of r_t and q_t were 27.76% and 4.35% respectively. A breakdown by years for these results is provided in below table.

Table 3.3: Summary statistics for daily log return of S&P 200 and the 22 trading-day G' process

		$\Delta \ln S_t$		$\Delta G'_t$		<i>Cross – correlation</i>
Period	No. Obs	Mean	SD(annual)	Mean	SD(annual)	$\text{corr}(r_t, q_t), lag = 0$
2010	253	-0.0001	15.80%	-0.0005	2.12%	-0.8167
2011	251	-0.0006	19.56%	0.001	3.42%	-0.8575
2012	253	0.0005	11.85%	-0.0025	1.50%	-0.7984
All	757	-0.0001	27.76%	-0.0005	4.35%	-0.8410

4 Forecasting the market volatility with the purify option price process

As discussed, option prices reflect the expectations of the future movements of the underlying assets. Therefore, the volatility implied from the options prices may contain useful information about the future stock market volatility. In this section, we look at the forecasting power of the implied volatility derived from the purified option prices against the traditional volatility index VIX, and their relationship with the future volatility.

We introduce the following processes:

- FV_t : the rolling ex-post (future) volatility measured at 22-trading-day windows, estimated by using the stock index prices at $s = t, t + 1, \dots, t + 22$.
- VIX_t : the non-parametric 22-trading-day volatility index S&P/ASX 200, using mid prices for S&P/ASX 200 put/call options.
- IV_t^G : the 22-trading-day till expiration implied volatility computed from the purified option prices, constructed by using the at-the-money options, average of $IV_t^{G_C}$ and $IV_t^{G_P}$.

For the future volatility, we estimate:

$$FV_t = \left(\frac{252}{\Delta t} \sum_{k=t}^{t+\Delta t} (\bar{R} - R(t_k))^2 \right)^{1/2}, \quad (4.1)$$

where

$$R(t_k) = \log S(t_k) - \log S(t_{k-1}), \quad \bar{R} = \frac{1}{\Delta t} \sum_{k=t}^{t+\Delta t} R(t_k), \quad \Delta t = 22.$$

Next, the volatility index VIX is derived from the near term and next term options on the S&P/ASX 200 using the out-of-money option. The overnight RBA rate, 1-month, 2-month and 3-month BBSW rates are used to interpolate the risk free rates at each maturity. The general formula to calculate this implied volatility is:

$$\sigma^2 = \frac{2}{T} \sum_i \frac{\Delta K_i}{K_i^2} e^{RT} O(K_i) - \frac{1}{T} \left(\frac{F}{K_0} - 1 \right)^2, \quad (4.2)$$

where: σ : implied volatility, T : time to expiration, F : forward index level, K_i : strike price of the i^{th} out-of-the-money option, ΔK_i : interval between strike prices, $K_0 = F$, R : risk-free interest rate, $O(K_i)$: strike mid-price of each option with strike K_i . More details about the construction of the S&P/ASX 200 VIX process can be found at [23]. Here we simply obtain the raw VIX data from SIRCA.

The Black-Scholes' model (2.4) is then used to derive the implied volatility from the at-the-money purified call/put option prices with fixed 22-trading-day time horizon. For the risk-free rate, we interpolate the RBA and BBSW rates similar to that was used for the volatility index as discussed above.

In Figure 4.1, we present a time series plot of the three volatilities over the next 22

trading days. It is observed that both IV_t^G and VIX_t could track FV_t 's movements, therefore can be used as predictors of the future volatility. We estimate the correlation between those volatility measures. We observe that $\text{corr}(FV_t, VIX_t) = 0.8511$ and $\text{corr}(FV_t, IV_t^G) = 0.8992$ respectively.

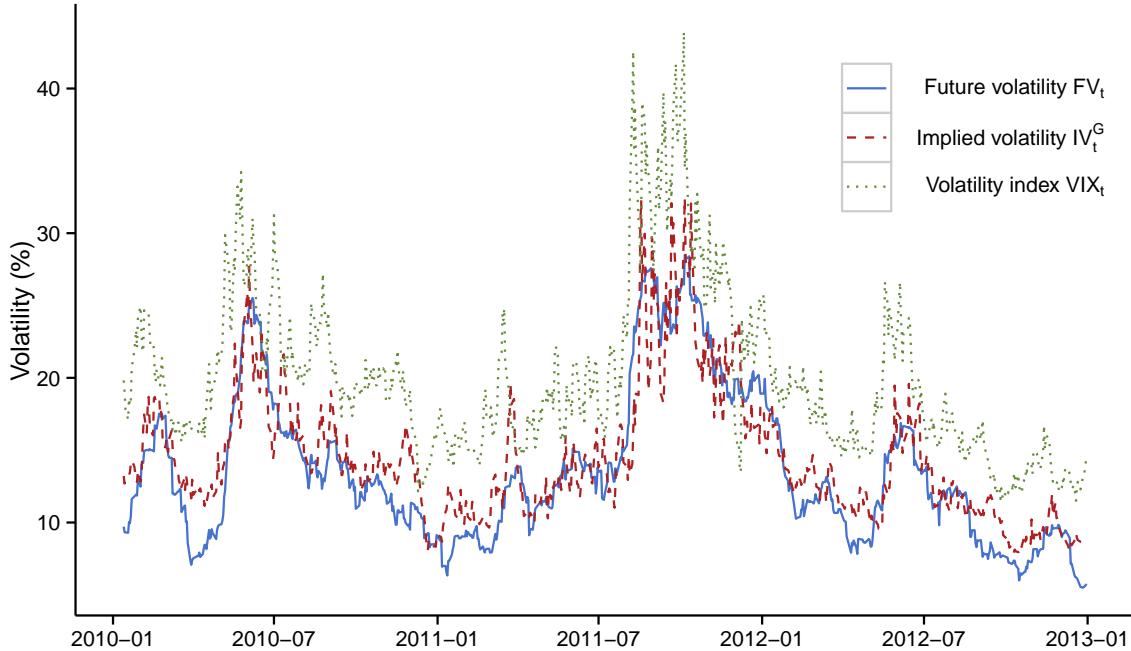


Figure 4.1: Future volatility, volatility index and implied volatility of the purified at-the-money option price process for S&P/ASX 200 from 01/01/2010 to 31/12/2012.

To examine the information content in the new implied volatility process IV_t^G and compare its ability in forecasting the future volatility against the implied volatility index, we consider the following multiple regressions:

Model (1)

$$FV_t = \eta + \alpha FV_{t-\Delta t} + \beta_0 VIX_t + \epsilon_t;$$

Model (2)

$$FV_t = \eta + \alpha FV_{t-\Delta t} + \beta_0 VIX_t + \beta_1 IV_t^G + \epsilon_t,$$

where $\epsilon_t \sim N(0, \sigma^2)$ are the residual errors of each model. It is noted that Model 1 is based on the conventional multiple regression model of the volatility with the inclusion of the implied volatility index [8] and $FV_{t-\Delta t}$ is the non-overlapped estimation of the future volatility. We extend Model 1 by adding the implied volatility derived from the purified option prices. If these predictors contain some information about the future volatility, the coefficients α and β_i

should be statistically significant.

We compute the residuals (the difference of the observed and the actual values) by:

$$\epsilon_t = \widehat{FV}_t - FV_t,$$

where \widehat{FV}_t is the predicted future volatility and FV_t is the observed future volatility. To compare the accuracy among the models, we compute the root-mean-square-error:

$$RMSE = \sqrt{MSE} = \left(\frac{1}{n} SSE \right)^{1/2} = \left(\frac{1}{n} \sum_{t=1}^n (\widehat{FV}_t - FV_t)^2 \right)^{1/2}. \quad (4.3)$$

for each model. Here, n is the number of observations in the dataset, i.e. 757 observations for our selected sample. The model with smaller RMSE would suggest that the predicted values on average are closer to the observed values, hence is a better model.

We also include the values from Akaike information criterion (AIC) test and Bayesian information criterion (BIC) test to measure the relative quality of the two models. Both criteria are capable of dealing with the trade-off between the goodness of fit and the complexity of the model as more variables are introduced. These criteria are based on a high log-likelihood value, but the penalty term of BIC ($k \ln n$) is potentially much more stringent than that of AIC ($2k$). These information criteria are estimated by:

$$AIC = n \ln MSE + 2k, \quad BIC = n \ln MSE + k \ln n,$$

where n is the number of observations, k is the number of estimated parameters and MSE is given by formula 4.3. The Durbin–Watson (DW) statistic is reported as a diagnostic check for the independence of errors in regression by detecting the presence of autocorrelation in the residuals, given by:

$$DW = \frac{\sum_{t=2}^n (\epsilon_t - \epsilon_{t-1})^2}{\sum_{t=1}^n \epsilon_t^2},$$

where $DW < 2$ suggests positive autocorrelation, $DW = 2$ for no autocorrelation and $DW > 2$ for negative autocorrelation. Since we have a lagged dependent in our model, Breusch–Godfrey [3] serial correlation Lagrange multiplier test' p-values are also reported to examine the presence of serial dependence that has not been included in a proposed model structure.

In our experiment, the data is split into two periods. The first subset includes data points from 01/01/2010 to 31/05/2011. This dataset is used as ‘in-sample’ data for determining the models’ parameters. The rest of the data from 01/06/2011 to 31/12/2012 is then used as ‘out-of-sample’ data for checking the efficiency of the models (1) and (2).

Usually, the coefficients of the regression models can be found by using the Ordinary Least Square (OLS) estimation method. However, previous study showed that the residuals computed from OLS can be highly autocorrelated for such models with Durbin-Watson test values are less than 1 [8]. This will raise the possibility of a spurious regression phenomenon [22] in our prediction. Therefore, the OLS estimation for the coefficients is inconsistent. The Feasible Generalized Least Squares (FGLS) estimation can be used an alternative consistent estimates of those models in the presence of autocorrelated errors. From our experiments, we observed that with OLS estimation, the coefficients of the predictive variables are statistically significant (see table 4.1). This suggests that the selected predictors are useful for predicting the future volatility. Moreover, from the model (2), we observe that the implied volatility from the purified option prices plays a more important role in forecasting the future volatility with its coefficient closer to 1, in comparison with the past volatility and the volatility index. However, when examine the Durbin-Watson statistical results, it is observed that the residuals from both models are autocorrelated (0.3619 and 0.3914 for model (1) and (2) respectively). Furthermore, the significant Breusch-Godfrey statistics for these regressions indicate the present of significant serial dependence in regression residuals (i.e. p-value is close to 0). Hence, the estimated coefficients from OLS for our model is inconsistent.

To overcome this, we performed the augmented Dickey–Fuller (ADF) [10] test for testing unit root and stationary of the residuals in model (1) and (2). As found, the insignificant ADF statistics in Table 4.1 do not indicate the presence of unit root in regression residuals. Therefore, we apply the Cochrane-Orcutt feasible generalised least square (CO-FGLS) estimation method [9] for these models (1) and (2) by modelling the first-order autoregressive on the error terms. The models (1) and (2) are adjusted as following:

Model (1')

$$\begin{aligned} FV_t &= 0.2393 - 0.3950 FV_{t-\Delta t} - 0.0468 VIX_t + \epsilon_t; \\ \epsilon_t &= 0.9951 \epsilon_{t-1} + e_t, \end{aligned}$$

Model (2')

$$\begin{aligned} FV_t &= 0.2376 - 0.3890 FV_{t-\Delta t} - 0.0468 VIX_t + 0.0569 IV_t^G + \epsilon_t; \\ \epsilon_t &= 0.9949 \epsilon_{t-1} + e_t, \end{aligned}$$

with e_t being the input noise. As a result, the Durbin-Watson's tests after using Cochrane-Orcutt's transformation are close to two, with the Breusch-Godfrey statistics for these regression indicate the absence of significant serial dependence in regression residuals. This confirms the relevance of Model (1) and (2).

Table 4.1: Regression results for in-sample data

<i>Dependent variable: FV_t</i>				
	OLS Estimation	Cochrane–Orcutt Estimation		
Coefficients	Model (1)	Model (2)	Model (1')	Model (2')
η	-0.0285	-0.0236	0.2393	0.23763
$FV_{t-\Delta t}$	0.1117	0.0980	-0.3950	-0.3890
VIX_t	0.8126	0.2807	-0.0084	-0.0468
IV_t^G		0.7060		0.0569
ρ			0.9951	0.9949
Durbin-Watson	0.3619	0.3914	1.9810	1.9834
Breusch-Godfrey	2.2×10^{-6}	2.2×10^{-6}	0.4867	0.6877
Dickey-Fuller	-7.0498 *	-7.647 **		
AIC	-2581.402	-2631.773	-3633.377	-3633.845
BIC	-2563.841	-2609.821	-3616.362	-3617.576
RMSE	0.0284	0.0263	0.0087	0.0072

Note: All coefficients are significant at p = 1%.

Critical values for augmented Dickey-Fuller test statistics:

* -2.58 at p = 1% ; ** -2.63 at p = 1%

In terms of accuracy of each model, the RMSE for the in-sample data from Model (1) is 0.0284 and 0.0263 from Model (2). With the out-sample data, RMSE are 0.0171 and 0.0168 for Models (1) and (2) respectively. The RMSE estimates are further improved via Cochrane-Orcutt estimation, with RMSE of model (2') be 0.0072 vs RMSE of model (1') be 0.0087. This suggests that the inclusion of the implied volatility from the purified option prices improves the accuracy of our forecast. This is also in agreement with the AIC and BIC tests from table 4.1.

In conclusion, for the selected dataset, we found that the implied volatility from the proposed process contains useful information about the movement of future volatility and can be used to improve the accuracy in forecasting future volatility.

5 Discussion

In this paper, we propose the use of a process G which represents the “dynamically purified” option price process where the impact of the stock price movement is reduced. The process is constructed by using observation of the market option prices. In our experiments, we constructed the process G for the stock index S&P 200 using the at-the-money options. We observed that there is a stable and strongly negative contemporaneous correlation between the increments of stock price return and the increments of G . In additions, we observed a strong correlation between the implied volatility computed from the at-the-money purified option price process and the non-parametric out-the-money implied volatility index VIX. This is an interesting feature since the VIX is calculated using very different data and methods. Similar to VIX, the implied volatility from the purified option prices can be used directly in volatility forecast. We found that the use of the implied volatility from the purified option prices can help improve the accuracy in predicting the future volatility in some experiments with a set of linear regression models. Besides, the new process G can be constructed using observations of just 18 option prices. This is significantly fewer prices than what VIX requires. Therefore, this process can be used to replace VIX in some cases when there is no sufficient data to calculate VIX or one interests in different ranges of strike prices.

Acknowledgment

This work was supported by ARC grant of Australia DP120100928. The authors extend their appreciations to the anonymous reviewer for the valuable suggestions that helped us to improve the paper.

References

- [1] Black, F. and M. Scholes. (1973): The valuation of options and corporate liabilities. *Journal of Political Economy*, 81, 637-654.
- [2] Brenner, B. and M.G. Subrahmanyam. (1988): A simple formula to compute the implied standard deviation. *Financial analysts journal*, 44, 80-83.
- [3] Breusch, T.S. (1979): Testing for Autocorrelation in Dynamic Linear Models. *Australian Economic Papers*, 17, 334–355.
- [4] Köksal, B. (2009): A comparison of conditional volatility estimators for the ISE National 100 Index Returns. *Journal of Economic and Social Research*, 11(2), 1-28.
- [5] Canina, L., and S. Figlewski. (1993): The information content of implied volatility. *Review of Financial Studies*, 6, 659-681.
- [6] Carr, P. and R. Lee. (2003): At-the-money implied as a robust approximation of the volatility swap rate. *Working paper*. New York University.
- [7] Carr, P. and L. Wu. (2006): A tale of two indexes. *Journal of Derivatives*, 13, 13–29.
- [8] Christensen, B.J., N.R. Prabhala. (1998): The relation between implied and realized volatility. *Journal of Financial Economics*, 50, 125-150.
- [9] Cochranea, D., G.H. Orcutta. (1949): Application of least squares regression to relationships containing auto-correlated error terms. *Journal of the American Statistical Association*, 44, 245, 32-61.
- [10] Dickey, D. A. and W. A. Fuller. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74, 427–431.
- [11] Dimitris, B. and I. Popescu. (2000): On the relation between option and stock prices: A convex optimization approach.
- [12] Dokuchaev, N. (2006): Two unconditionally implied parameters and volatility smiles and skews.
- [13] Dokuchaev, N. (2014): Volatility estimation from short time series of stock prices. In press in *Journal of Nonparametric Statistics* **26**(2), 373–384.
- [14] Duffe, D. (1992): Dynamic Asset Pricing Theory, 2nd ed. *Princeton, NJ: Princeton University Press*.

- [15] Frijns, B., Tallau C. and A.T. Rad. (2010): The information content of implied volatility: Evidence from Australia. *Journal of Futures Markets*, 30, 134-154.
- [16] Giot, P. (2005): Implied volatility indexes and daily value at risk. *The Journal of Derivatives*, 12, 54-64.
- [17] Harri, A., S.W. Brorsen. (2009): The overlapping data problem. *Quantitative and Qualitative Analysis in Social Sciences*, 3, 3, 78-115.
- [18] Heston, S.L. (1993): A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6, 327-343.
- [19] Hull, J. and A. White. (1987): The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42, 281-300.
- [20] Jiang, G.J. and Y.S. Tian. (2005) The model-free implied volatility and its information content. *Review of Financial Studies*, 18, 1305-1342.
- [21] Merton, R.C. (1973): Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4, 141-183.
- [22] Merton, R.C. (1973): Understanding spurious regressions in econometrics. *Journal of Econometric*, 33, 311-340.
- [23] S&P Dow Jones Indices. (2014): S&P/ASX 200 VIX Methodology. Retrieved from <http://au.spindices.com/>
- [24] Reserve Bank of Australia. (2013): *Interest Rates and Yields - Money Market - Daily - 1976 to 2013 - F1*. Retrieved from <http://www.rba.gov.au/statistics/tables/>
- [25] SIRCA. (2013): *Thomson Reuters Tick History*. Retrieved from <http://www.sirca.org.au/>
- [26] Whaley, R.E. (2000): The investor fear gauge. *The Journal of Portfolio Management*, 6, 12-17.