WP/10/181

IMF Working Paper

A New Framework To Estimate the Risk-Neutral Probability Density Functions Embedded in Options Prices

Kevin C. Cheng

INTERNATIONAL MONETARY FUND

 WP/10/181

IMF Working Paper

Research Department

A New Framework to Estimate the Risk-Neutral Probability Density Functions Embedded in Options Prices

Prepared by Kevin C. Cheng[1]

Authorized for distribution by Thomas Helbling

August 2010

Abstract

Building on the widely-used double-lognormal approach by Bahra (1997), this paper presents a multi-lognormal approach with restrictions to extract risk-neutral probability density functions (RNPs) for various asset classes. The contributions are twofold: first, on the technical side, the paper proposes useful transformation/restrictions to Bahra's original formulation for achieving economically sensible outcomes. In addition, the paper compares the statistical properties of the estimated RNPs among major asset classes, including commodities, the S&P 500, the dollar/euro exchange rate, and the US 10-year Treasury Note. Finally, a Monte Carlo study suggests that the multi-lognormal approach outperforms the double-lognormal approach.

JEL Classification Numbers: C13, G13, G17
Keywords: Implied risk-neutral density functions, option pricing, market expectations.
Author's E-Mail Address: kcheng@imf.org

[1] This paper has benefited from discussions with Thomas Helbling and Shaun Roache. The author is very grateful to Ying He for her Matlab assistance and Marina Rousset for her research assistance.

Contents Page

Tables

Figures

I. Introduction

Since asset prices reflect discounted present values of expected future cash flows, they contain useful information on market expectations. Thus, information embedded in asset prices has long been used to analyze economic and financial prospects. One popular practice in this area is to use option prices to derive the risk-neutral probability density function for the expected price of the underlying security in the future. The logic of this practice is simple: given that an option's payoff is a function of the future developments of the underlying asset, the option premium paid by the investor for a certain exercise price reflects her view of the probability distribution of the expected underlying security prices.

Since the early 1990s, numerous methodologies in this area—from Shimko's (1993) interpolation of implied volatility to Bahra's (1997) double lognormal to Ait-Sahalia and Duarte's (2003) nonparametric kernel smoothing procedure to Figleski's (2008) generalized extreme value distribution tail-completion technique—have been developed. These techniques have been applied to options on different asset classes—from individual stocks to equity futures, interest rates futures, and currency futures.

With the notable exception of gold and crude oil, however, most of these studies have not been applied to commodities—an alternative asset class that has seen rapid growth over the past few years. This neglect largely reflects data hurdles in commodity markets in the implementation of these techniques: Specifically, most methodologies require a dense set of observations of option/strike prices. However, commodities futures options are usually not very liquid and the number of available option contracts is low.

This paper attempts to fill this gap. Specifically, it proposes to use a multi-lognormal parametric estimation framework—an extension and modification of the double-lognormal method formulated by Bahra (1997). The advantage of this method is that it does not require a large number of observations for options/strike prices. Furthermore, apart from extending Bahra's double-lognormal to a more generalized multi-lognormal framework, the paper also addresses certain known technical shortcomings associated with the double-lognormal approach by proposing some generic transformation/restrictions to Bahra's original framework. In addition, the paper compares and contrasts the statistical properties of the probability density functions of commodities vis-à-vis other asset classes such as the S&P 500 index, the dollar/euro exchange rate, and the 10-year US Treasury Note. Finally, the paper presents a Monte-Carlo study to compare the properties of various lognormal methods

Major findings/proposals of the paper include:

- On the technical side, the paper suggests that the multi-lognormal approach would yield more stable results and become more manageable if the optimization procedure is formulated in terms of the expected asset return μ and return volatility σ rather than the lognormal parameters α and β as proposed by Bahra (1997). In addition, restrictions—based on the researcher's assessment—should be imposed on both μ and σ to anchor the numerical procedure. Moreover, a Monte-Carlo simulation suggests that the multi-lognormal approach outperforms that of the more common double-lognormal approach.

- In terms of empirical implications, like the S&P 500 index, commodities—except gold—are found to have a noticeably higher positive skewness and kurtosis (fatter tails) than the dollar/euro exchange rate and Treasury bond futures.

The rest of the paper proceeds as follows: Section II discusses the theoretical background and presents an overview of existing methodologies; Section III discusses the multi-lognormal approach with transformation/ restrictions; Section IV applies the procedure to five commodities and other assets and compares/contrasts the results; Section V presents the Monte-Carlo simulation; and finally, Section VI concludes.

II. Theoretical Background and Existing Methodologies

A. Theoretical Background

Every financial asset with payoff Z_τ at time τ can be priced by the following Euler equation at time zero:[2]

$$P_0 = E[Z_\tau M_\tau] = \int_{-\infty}^{\infty} Z_\tau(\theta) \frac{e^{-\rho\tau} U'(C_\tau(\theta))}{U'(C_0)} f(\theta) d\theta = \int_{-\infty}^{\infty} Z_\tau(\theta) M_\tau(\theta) f(\theta) d\theta, \tag{1}$$

where $f(\theta)$ is the *objective* probability density function at time zero for some random outcomes θ to be realized at time τ; ρ is the consumer's subjective discount rate; and $M_\tau(\theta) \equiv \frac{e^{-\rho\tau} U'(C_\tau(\theta))}{U'(C_0)}$ is the intertemporal rate of marginal substitution of consumption—often referred to as the stochastic discount factor or the pricing kernel in the finance literature.

Since an investor's preferences are not directly observable, equation (1) is often rewritten in terms of risk-neutral probability distribution given as follows:

$$P_0 = e^{-r\tau} \int_{-\infty}^{\infty} Z_\tau(\theta) f^N(\theta) d\theta \equiv e^{-r\tau} E^N[Z_\tau], \tag{2}$$

where $E^N[\bullet]$ is the risk-neutral expectation at time zero; *r* is the risk-free interest rate during the horizon τ; $f^N(\theta)$ can be interpreted as the *risk-neutral* probability (RNP) distribution:[3]

---

[2] Equation 1 is the Euler equation derived from dynamic utility maximization problem. See, for example, Cochrane (2001) for a detailed discussion of the consumption-based asset pricing model.

[3] The derivation utilizes the fact that $e^{r\tau} = \frac{1}{E(M)}$, since $e^{-r\tau} = E(M \cdot 1)$ because the price at time zero of a risk-free bond that will pay \$1 at time τ is $e^{-r\tau}$.

$$f^N(\theta) \equiv \frac{M_\tau(\theta) f(\theta)}{\int_{-\infty}^{\infty} M_\tau(\theta) f(\theta) d\theta} \equiv \frac{M_\tau(\theta) f(\theta)}{E(M)} = e^{r\tau} M_\tau(\theta) f(\theta) \tag{3}$$

Equation (2) suggests that the price of an asset equals the present value (discounted) of its expected payoff under the risk-neutral probability distribution.

Using equation (2), Cox and Ross (1976) showed that a European-style call option that entitles an owner the right to purchase the underlying asset at strike price X at time τ can be priced as follows:

$$C(X,\tau) = \max(S_\tau - X, 0) = e^{-r\tau} \int_X^{\infty} (S_\tau - X) f^N(S_\tau) dS_\tau, \tag{4}$$

where S_τ is the terminal price of the underlying asset at time τ.

Furthermore, as shown by Breeden and Litzenberger (1978), the risk-neutral probability can be recovered by the second derivative of *C(X)*:[4]

$$C''(X) = e^{-r\tau} f^N(X) \tag{5}$$

Thus, given a set of cross-sectional data on option prices $\left(C_1, C_2, C_3, ..., C_K\right)$ and their corresponding strike prices $\left(X_1, X_2, X_3, ..., X_K\right)$, one can use (4) or (5) to extract $f_t^N(S_\tau)$—the risk neutral probability distribution at time zero for the underlying asset price at time τ through various methods to be discussed below.

B. Existing Estimation Methods

Existing frameworks to extract the risk-neutral probability density can be classified into three main approaches:

- specifying a parameterized stochastic process for the underlying asset price S_τ;
- exploiting equation (4) by Cox and Ross; and
- exploiting of equation (5) by Breeden and Litzenberger.

---

[4] Specifically, differentiating equation (4) using the Leibniz rule, we get:

$C'(X) = -e^{-r\tau}(X - X) f^N(X) + e^{-r\tau} \int_x^{\infty} -1 * f^N(S_\tau) dS_\tau = -e^{-r\tau} \int_x^{\infty} f^N(S_\tau) dS_\tau$. Differentiating this again yields equation (5).

The first approach has been used by Bates (1991) and Malz (1996). The approach assumes a stochastic process for the underlying asset prices—such as a jump-diffusion process or a geometric Brownian motion—which determines the RNP.[5] This approach, however, is less popular than the other two because it is relatively inflexible, as the assumption about the stochastic process imposes strong restrictions on the shape of the RNP of the underlying asset.

In the second approach, a functional form for $f^N(S_\tau)$ is assumed. A form commonly used in practice is Bahra's a double-lognormal approach. Specifically, as discussed in Bahra (1997), the double-lognormal approach is given by:

$$f^N(S_\tau) = \theta L(\alpha_1, \beta_1) + (1-\theta) L(\alpha_2, \beta_2),$$

where $\theta, \alpha_1, \alpha_2, \beta_1, \beta_2$ are parameters to be estimated.[6] An advantage of this approach is that it is relatively flexible and could capture various 'non-Black-Scholes" properties such as a very fat tail or a high degree of skewness.[7]

With the double-lognormal assumption, using (4), we can compute the fitted call and put prices as follows:[8]

$$\hat{C}_j = e^{-r\tau} \int_{X_j}^{\infty} (S_\tau - X_j)[\theta L(\alpha_1, \beta_1) + (1-\theta) L(\alpha_2, \beta_2)] dS_\tau$$

$$\hat{P}_j = e^{-r\tau} \int_0^{X_j} (X_j - S_\tau)[\theta L(\alpha_1, \beta_1) + (1-\theta) L(\alpha_2, \beta_2)] dS_\tau$$

In addition, if markets are efficient, the futures price of the underlying asset to be delivered at time τ (as of time zero) should be equal to the expected value of the underlying asset under the risk-neutral probability density:[9]

$$F_\tau = E^N(S_\tau) = \theta e^{\alpha_1 + \frac{1}{2}\beta_1^2} + (1-\theta) e^{\alpha_2 + \frac{1}{2}\beta_2^2} \tag{6}$$

---

[5] For example, an assumption of a geometric Brownian motion for the underlying asset price would imply that a lognormal distribution for the RNP. This is the famous Black-Scholes model (1973).

[6] The idea of lognormal mixtures was introduced by Melick and Thomas (1997) who used a mixture of three lognormal functions to estimate the RNP for the oil market. However, the formulation of Melick and Thomas is rather complicated and the framework of this paper is built on the formulation of Bahra, which is rather different from that of his predecessors.

[7] As pointed out Melick and Thomas, a specific stochastic process would imply a particular RNP; however, a given RNP is consistent with many different stochastic process.

[8] A put contract entitles an owner the right to sell the underlying at the strike price. Thus the pricing formula is the reverse of equation 4.

[9] The mean of a lognormal distribution with parameters α and β is $e^{\alpha + \frac{1}{2}\beta^2}$.

Then, given observations of K actual call prices $\left(\widetilde{C}_1, \widetilde{C}_2, \widetilde{C}_3 ..., \widetilde{C}_K\right)$ and L put prices $\left(\widetilde{P}_1, \widetilde{P}_2, \widetilde{P}_3, ..., \widetilde{P}_L,\right)$, the parameters can be estimated by minimizing the following objective function:

$$\min_{\alpha_1,\alpha_2,\beta_1,\beta_2,\theta} \sum_{j=1}^{K}[\widetilde{C}_j - \hat{C}_j]^2 + \sum_{j=1}^{L}[\widetilde{P}_j - \hat{P}_j]^2 + [\theta e^{\alpha_1 + \frac{1}{2}\beta_1^2} + (1-\theta) e^{\alpha_2 + \frac{1}{2}\beta_2^2} - F_\tau]^2 \ . \tag{7}$$

The advantage of Bahra's double-lognormal approach is that it only requires estimation of five parameters and therefore is not as data-demanding as other methods. It is more appropriate for less liquid options markets, such as those for commodity futures. A drawback of this approach is its instability in the case of low volatility and high skewness (Cooper, 1999).

A third approach is to exploit equation (4) from Breeden and Litzenberger by calculating the second derivative of $C(X)$ numerically. Since markets usually only offer a limited number of options with strike prices near the spot price of the underlying asset (i.e. options that are "near the money"), the actual observations are typically extended by interpolation between observed prices and extrapolation outside the range to model the tail. In addition, to make sure that $C(X)$ is indeed twice-differentiable, observations are typically smoothened to ensure enough curvature.[10]

A main advantage is that these procedures require no assumption on the stochastic process of the underlying asset or on the functional form of RNP. A main disadvantage, however, is that they can be quite data-demanding and unstable.

III. The Multi-Lognormal Approach with Restrictions

A. The Framework

In many options markets, only a limited number of discrete strike prices are traded, including in commodity futures markets. Consequently, the third approach that requires $C(X)$ to be twice-differentiable as described in the previous section may not be a workable solution for extracting RND for these assets.

Against this background, this paper opts for a parametric procedure that involves a mixture of lognormal as formulated by Bahra (1997). However, while Bahra's formulation is flexible, simple, and parsimonious, it is also known to have undesirable properties. In particular, one drawback, as discussed in Cooper (1999) is that it can generate spikes when one of the estimated lognormals has a very small standard deviation. Indeed, since the optimization problem described in equation (7) involves complex non-linear optimization, potential multiple solutions or local optima could arise. Therefore, imposing restrictions on the parameters to be

[10] Most common techniques have been described by Shimko (1993), Ait-Sahalia and Duarte (2003), and Figlewski (2008).

estimated and picking a sensible initial condition for the numerical optimization procedure would greatly facilitate the process and help ensure that the final results would have desirable properties.

Against this background, the remainder of the section will discuss appropriate transformation and propose useful restrictions to Bahra's framework. In addition, to make the procedure more general so that it can capture a wider possible range of stochastic processes, the discussion will be coined in terms of a generalized multi-lognormal approach with n mixtures.[11]

First, Bahra's original formulation problem should be transformed: instead of optimizing by choice of lognormal parameters, i.e., (α_i, β_i), as specified by equation (7), the problem should be solved by choosing (μ_i, σ_i), which are given by:[12]

$$\alpha_i = \ln S_0 + \left(\mu_i - \frac{\sigma_i^2}{2}\right)\tau, \beta_i = \sigma_i \sqrt{\tau} \qquad (8)$$

Since the pairs (α_i, β_i) and (μ_i, σ_i) have a one-to-one relation, from a purely mathematical perspective, the change of variable should not alter the optimization problem. In a practical sense, however, the transformation could facilitate the calibration of appropriate parameter restrictions and initial conditions (to be discussed below) because both μ_i and σ_i have an "intuitive" interpretation while the lognormal parameters α_i and β_i do not.

Specifically, in the case of a single-lognormal distribution, μ and σ can be interpreted as the mean and volatility (measured by the standard deviation) of the asset return.[13] Strictly speaking, the precise mathematical relation of (μ_i, σ_i) with respect to the expected return and return volatility is unknown in a multi-lognormal case. Nonetheless, knowledge of this interpretation in the single lognormal case can help the researcher to "anchor" permissible range for the μ_i and σ_i and calibrate a sensible initial condition for the numerical procedure.

Another modification is to impose equation (6)—the relation that futures prices should equal expected prices—as a constraint rather than in the objective function as in Bahra (as specified in equation (7)). The advantage is that this will ensure that the relation will hold more precisely because putting the condition in the objective function will entail a tradeoff vis-à-vis the first two parts of the objective function.

---

[11] Although a multi-lognormal will increase the number of parameters to be estimated, it is still far less data-demanding than other approaches.

[12] To avoid notational confusion, the following conventions are used throughout the paper: i denotes the index across the mixtures of lognormal; j denotes the index across the observations of options/strike prices; n denotes the total number of mixtures used; while K and L denotes the number of available call and put contracts, respectively.

[13] See Chapters 12-13 in Hull (2005) for a lucid explanation. Also see Black and Scholes (1973) for further details.

Thus, substituting (8) into this constraint as given by $F_\tau = \sum_{i=1}^{n} \theta_i e^{\alpha_i + \frac{1}{2}\beta_i^2}$, the constraint can then be simplified into:

$$F_\tau = S_0 \sum_{i=1}^{n} \theta_i e^{\mu_i \tau} \tag{9a}$$

Equation (9a) has an interesting economic interpretation in the case of a single lognormal (i.e. n=1) , because equation (9a) can then be simplified into

$$F_\tau = S_0 e^{\mu\tau} . \tag{9b}$$

Comparing (9b) to the well-known spot-forward relation:[14]

$$F_\tau = S_0 e^{(r-\gamma)\tau} ,$$

where r and γ are the risk-free rate and the dividend/convenience yield, respectively, implies that the expected return of an asset equals the risk-free rate minus the dividend/convenience yield.

To recap, putting all pieces together, the transformed multi-lognormal approach with n mixtures is to choose of a set of $(\mu_1, \mu_2, ..., \mu_n)$, $(\sigma_1, \sigma_2, ..., \sigma_n)$, and $(\theta_1, \theta_2, ..., \theta_n)$ to solve the following constrained non-linear program:[15]

$$\min \sum_{j=1}^{K} [\tilde{C}_j - \hat{C}_j]^2 + \sum_{j=1}^{L} [\tilde{P}_j - \hat{P}_j]^2$$
$$\textit{subject to } F_\tau = S_0 \sum_{i=1}^{n} \theta_i e^{\mu_i \tau} \text{ and } \sum_{i=1}^{n} \theta_i = 1, \theta_i \geq 0 \forall i \quad , \tag{10a}$$

where $(\tilde{C}_1, \tilde{C}_2, ..., \tilde{C}_K)$ are K observed actual call prices; $(\tilde{P}_1, \tilde{P}_2, ..., \tilde{P}_L)$ are L observed actual put prices; $(\hat{C}_1, \hat{C}_2, ..., \hat{C}_K)$ and $(\hat{P}_1, \hat{P}_2, ..., \hat{P}_L)$ are calculated call and put prices, based the following closed forms:[16]

$$\hat{C}(X) = e^{-r\tau} \sum_{i=1}^{n} \theta_i \left\{ e^{\mu_i \tau} N(d_{1,i}) - XN(d_{2,i}) \right\} \text{ and}$$

[14] This relation is derived from the implication of no arbitrage.

[15] In *Matlab* version 8, this problem can be solved by the command ***fmincon***, a procedure for tackling complex constrained non-linear minimization problems.

[16] The derivations of these closed-form solutions are similar to those of the Black-Scholes model, which is essentially a single-lognormal model. Bahra (1997) also has similar closed forms for his double-lognormal, but they are in terms of α and β .

$$\hat{P}(X) = e^{-r\tau}\sum_{i=1}^{n}\theta_i\left\{-e^{\mu_i\tau}N(-d_{1,i}) - XN(-d_{2,i})\right\} \tag{10b}$$

where $N(\bullet)$ is the normal cumulative distribution function and

$$d_{1,i} = \frac{1}{\sigma_i\sqrt{\tau}}\ln\left(\frac{S_0}{X}\right) + \sqrt{\tau}\left[\frac{\mu_i}{\sigma_i} + \frac{1}{2}\sigma_i\right] \text{ and } d_{2,i} = d_{1,i} - \sigma_i\sqrt{\tau}\ .$$

B. Useful Restrictions and Initial Condition

As discussed above, given the complex nonlinear structure of (10), multiple solutions and local optima may exist. Therefore, some parameter restrictions and initial conditions can facilitate the numerical procedure and help ensure an economically sensible outcome.

First, since μ_i and σ_i are related to the expected return of the underlying asset and its volatility (standard deviation), respectively, it would be reasonable to restrict μ_i to be within an interval determined by multiples of standard deviations around the historical expected return:

$$\bar{\mu} - \lambda\bar{\sigma} \leq \mu_i \leq \bar{\mu} + \lambda\bar{\sigma}\ . \tag{11a}$$

where $\bar{\mu}$ is some historical value or another value the researcher deems appropriate. A reasonable value for λ would be two, since that would cover a 95-percent confidence interval, if the distribution of the asset return is close to a normal distribution.

Likewise, σ_i should be restricted in a similar fashion. However, since σ_i should always be positive, the restriction could thus take the form:

$$\frac{1}{\xi}\bar{\sigma} \leq \sigma_i \leq \xi\bar{\sigma}\text{, where } \xi > 1, \tag{11b}$$

where $\bar{\sigma}$ is some historical value of appropriate value in the researcher's judgment. The value of ξ should depend on the expected "volatility of volatility" of the underlying asset return.

A delicate balance needs to be struck between imposing constraints that are too tight and constraints that are too loose. On the one hand, if the constraints are too tight—a too small λ in (11a) and/or a too small ξ in (11b)—the flexibility of the optimization procedure could be compromised, thereby hampering the data from "speaking for themselves". On the other hand, if the permissible intervals in (11a) and (11b) are too wide, the procedure can yield implausible or unwieldy results with undesirable properties such as spikes as discussed in Copper (1999).

Finally, regarding the initial condition, a natural choice would be to start the procedure with an equally-weighted mixture, with $\bar{\mu}$ and $\bar{\sigma}$ being the initial values for the parameters.

IV. Applications

A. The Setup

This section applies the multi-lognormal approach with four lognormal mixtures to a variety of asset classes, including five commodities—WTI crude oil, gold, copper, corn, and wheat—together with the Continuous Commodity Index (CCI)—a commodity index of 17 component commodities—as well as the S&P 500 index, the Dow Jones Index, the dollar/euro exchange rate, and the US 10-year Treasury Bond. The underlying assets for these options contracts are all futures contracts, as specified in Table 1:

Table 1. Futures Contracts Specification

Contract	Exchange 1/	Description	Bloomberg Ticker
WTI	NYMEX	1,000 barrels	CLA Comdty
CCI 2/	NYF-ICE	500 USD x Index	CIA Index
Gold	COMEX	100 troy ounces	GCA Comdty
Copper	COMEX	25,000 pounds	HGA Comdty
Wheat	CBOT	5,000 bushels	W A comdty
Corn	CBOT	5,000 bushels	C A comdty
S&P 500	CME	250 USD x Index	SPA Comdty
Dow Jones	CBOT	10 USD x Index	DJI Index
USD/Euro	CME	Exchange rate; 125,000 Euro	ECA Curncy
Treasury Note	CBOT	U.S. Treasury 10-year bond; 100,000 USD	TYA Comdty

Sources: COMEX division of NYMEX; NYMEX; CBOT; and CME

1/ COMEX is a division of NYMEX, the New York Mercantile Exchange. CBOT is an abbreviation of the Chicago Board of Trade. CME is an abbreviation of the Chicago Mercantile Exchange. NYF-ICE stands for the Intercontinental Exchange, New York.

2/ Continuous Commodity Index, average of 17 commodity futures contracts, 1995 revision of the Commodity Research Bureau Index.

Data on settlement options/strike prices were collected on March 24-25, 2010 from Bloomberg.[17] A caveat is in order here: it is possible that the settlement data may not truly reflect market expectation across all strike prices because of low trading activity for certain options contracts that are deeply out of money.[18]

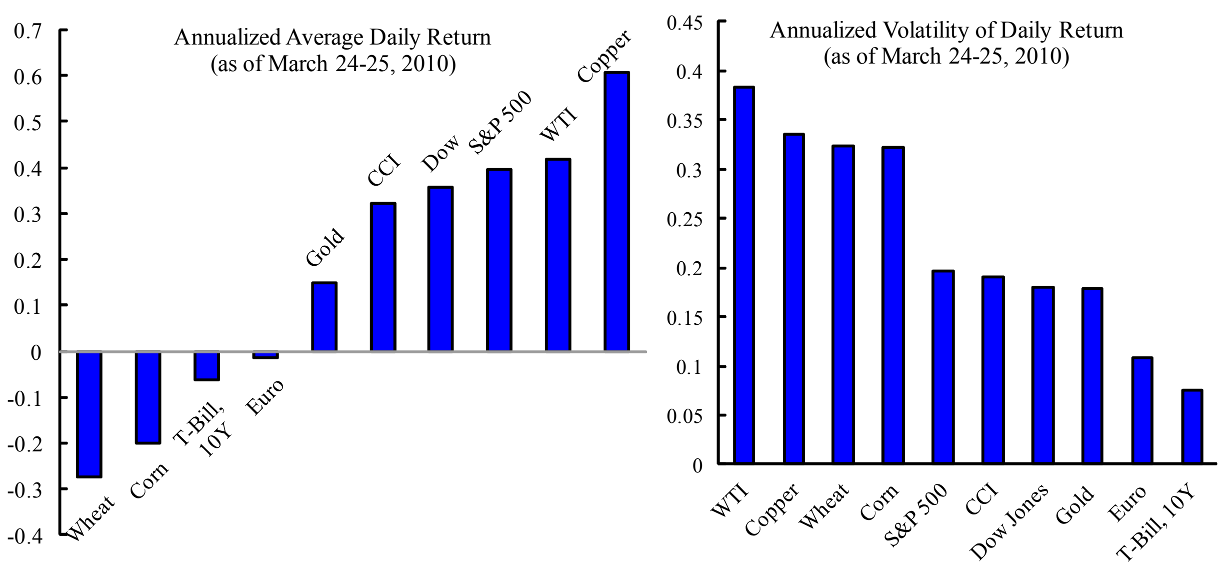
[17] Another set of data around end-April and early May. The estimated probability density functions will then be compared to those estimated earlier.

[18] A potential remedy for the situation (to be implemented in a future draft of the paper) is to augment the data with the bid-ask quotes of the market makers (brokers) because these quotes tend to incorporate up-to-date information even though an actual market transaction has not taken place. In addition, this would also be beneficial to the numerical procedure as it increases the number of observations.

The historical expected asset return, $\overline{\mu}$, is approximated by the annualized average daily return of the asset prices during March 25, 2009-March 24, 2010; and the historical return volatility, $\overline{\sigma}$, is approximated by the annualized volatility (standard deviation) shown in Figure 1.[19] Return is calculated by the daily changes in logarithm of prices. The risk-free interest rate, r, is given by the Treasury bill/bond rate with a maturity similar to the horizon between March 24, 2010 and the expiration date of the option.[20]

To ensure comparability and consistency, all assets are subject to the same sets of generic parameter restrictions. Specifically, μ_i is restricted to be within plus or minus two historical standard deviations ($\overline{\sigma}$) from the historical mean return ($\overline{\mu}$): i.e. $\overline{\mu} - 2\overline{\sigma} \le \mu_i \le \overline{\mu} + 2\overline{\sigma}$; while σ_i is restricted to be within the range between $1/3\,\overline{\sigma}$ and $3\,\overline{\sigma}$.

Figure 1. Annualized Average Daily Returns and Return Volatilities



Sources: Bloomberg and the author's calculations

B. Results

The general story presented in the Table 2 and Figures 2a and 2b is: as of end-March, 2010,

- With the exception of gold, for all key commodities prices, such as crude oil, corn, and wheat, prices are not expected to recoup their 2008 losses by mid-2010 or end-2010,

[19] Return is annualized by multiplying by 260, which is the approximate number of trading days within one year; volatility is annualized by multiplying by the square root of 260.

[20] For horizon less than three months, the 3-month Treasury Bill rate is used. For horizon higher than 3 months, a weighted average of interest rates is used. For example, the 5 month rate is approximated by two third of the 6-month rate and one third of the 3-month rate.

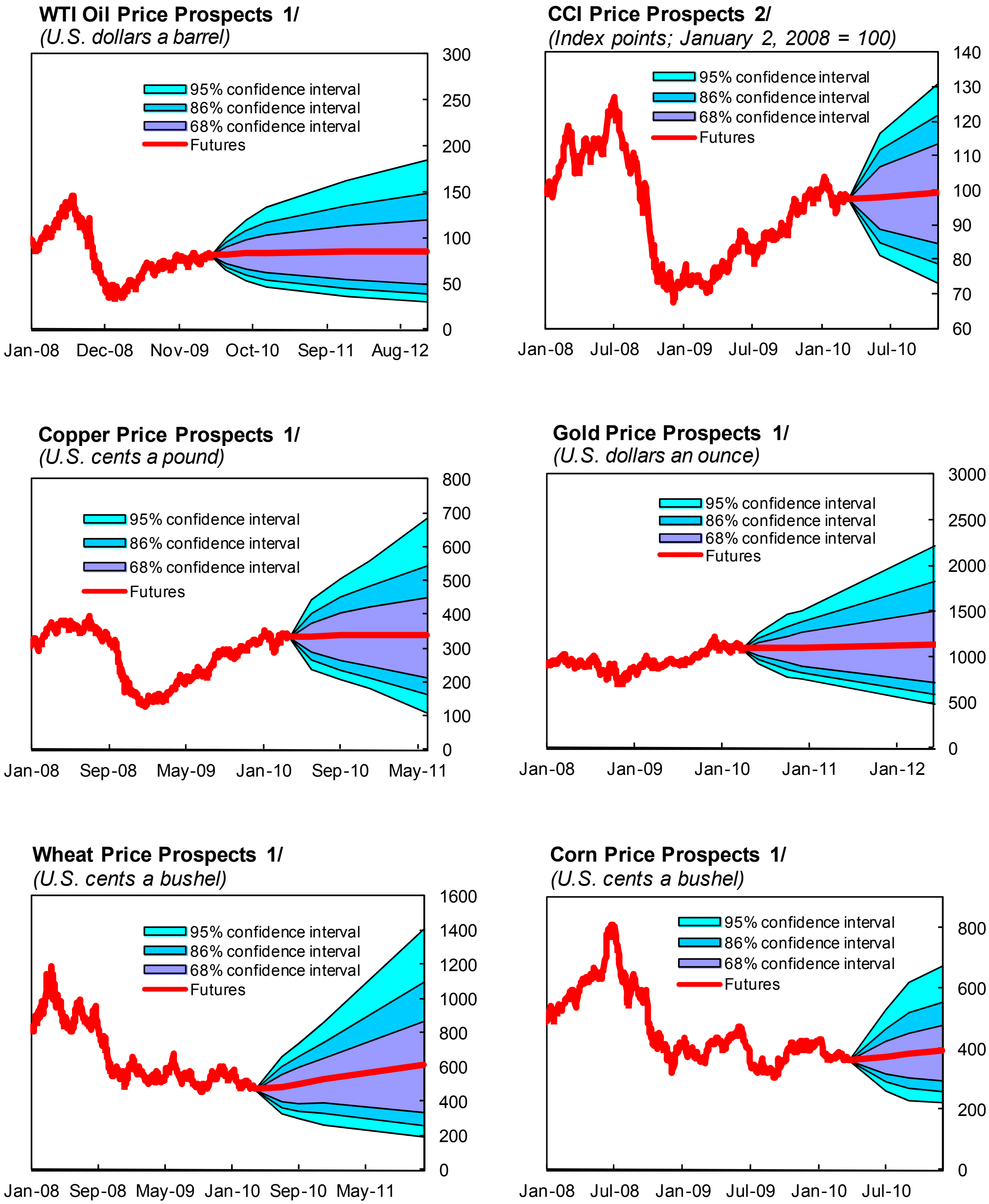
Table 2. Outlook for Major Commodity and Financial Prices as of March 24-25, 2010
(Probability and yield in percent; prices in U.S. dollars)

	Continuous Commodity Index (January 2, 2008=100) 1/		WTI Crude Oil		Gold	
	Jun-10	Nov-10	Jun-10	Dec-10	Jun-10	Dec-10
Futures Prices	98	99	81	83	1094	1097
Prob(higher than 2007 mean)	91	81	85	67	100	99
Prob(higher than 2008 peak)	0	4	0	1	88	65
Prob(higher than 2008 mean)	36	42	3	20	100	89
Prob(higher than 2009 Q1-Q2 average)	99	94	100	95	99	82
	Copper		Corn		Wheat	
	Jun-10	Dec-10	Jul-10	Dec-10	Jul-10	Dec-10
Futures Prices (in U.S. cents)	334	338	376	394	483	527
Prob(higher than 2007 mean)	59	52	47	51	4	19
Prob(higher than 2008 peak)	6	20	0	1	0	0
Prob(higher than 2008 mean)	68	57	3	10	0	5
Prob(higher than 2009 Q1-Q2 average)	100	97	34	43	17	37

	S&P 500		Dow Jones		USD/Euro Exchange 5/	
	Jun-10	Dec-10	Apr-10	Jun-10	Jun-10	Dec-10
Futures Prices	1170	1155	10778	10790	1.330	1.330
Prob(higher than pre-crisis level) 2/	0	5	0	0	33	40
Prob(higher than pre-Bear-Stearn level) 3/	4	20	0	2	0	8
Prob(higher than pre-Lehman level) 4/	11	26	2	13	3	15
Prob(higher than 2009 Q1-Q2 average)	100	97	100	100	46	47
	Treasury Bond Price		Treasury Yield (derived) 6/			
	Jun-10	Dec-10	Jun-10	Dec-10		
Futures Prices	116	113				
Prob(higher than pre-crisis level) 2/	100	91	0	10		
Prob(higher than pre-Bear-Stearn level) 3/	16	16	84	84		
Prob(higher than pre-Lehman level) 4/	39	27	61	73		
Prob(higher than 2009 Q1-Q2 average)	3	6	97	94		

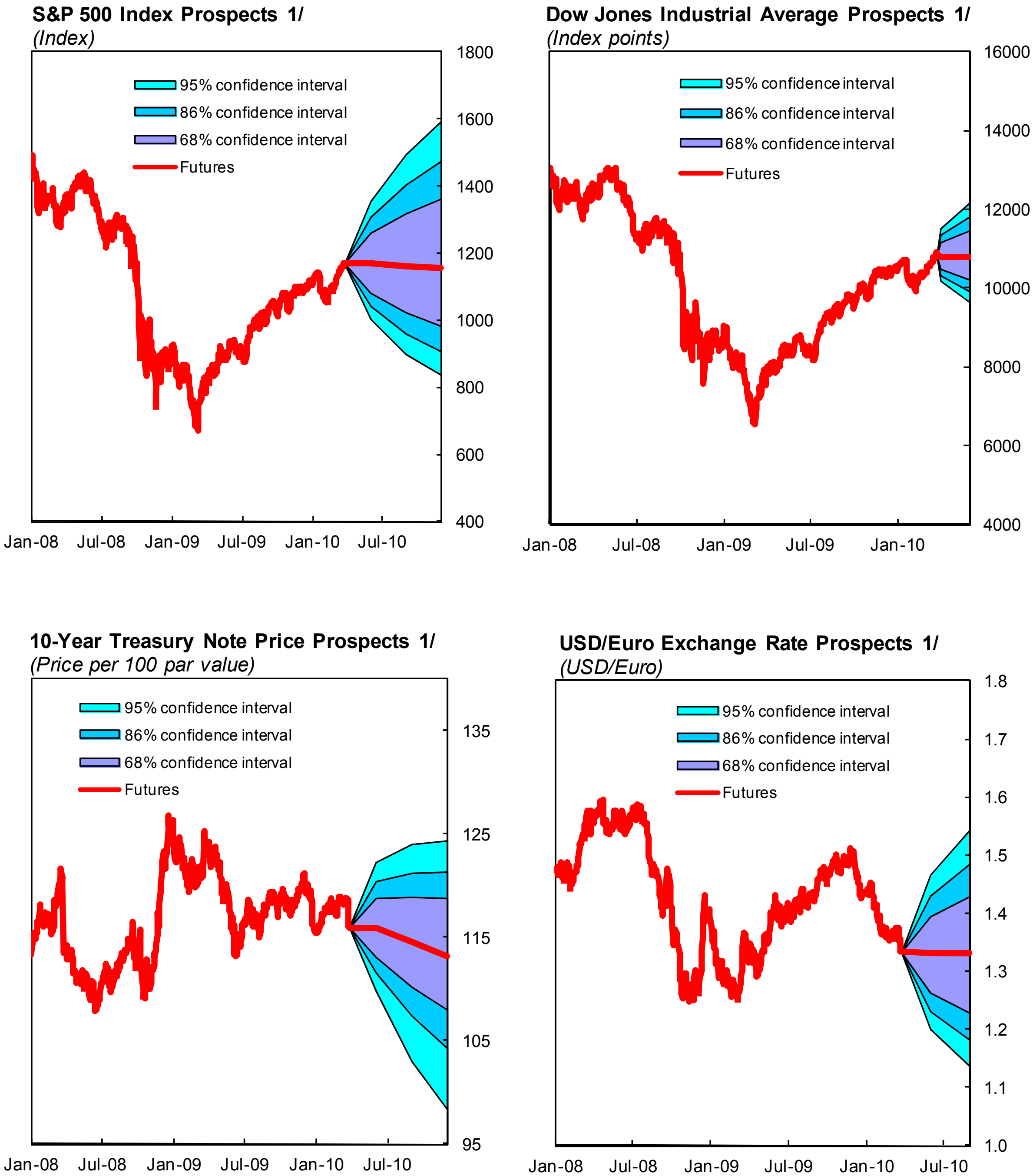
Sources: IMF staff calculations.
1/ 1995 Revision of the Commodity Research Bureau Index; average of 17 commodity futures prices; traded at NYBOT.
2/ defined as end-June 2007.
3/ defined as end-February 2008.
4/ defined as end-August 2008.
5/ An increase implies an appreciation in Euro.
6/ "Yield" refers to the yield to maturity.

Figure 2a. Fan Charts for Selected Commodities (as of March 24-25, 2010)



Sources: Bloomberg; and IMF staff calculations.
1/ Derived from prices of futures options on March 24-25, 2010.
2/ 1995 Revision of the Commodity Research Bureau Index; average of 17 commodity futures prices; traded at NYBOT; Derived from prices of futures options on March 10, 2010.

Figure 2b. Fan Charts for Selected Financial Instruments (as of March 24-25, 2010)



Sources: Bloomberg; and IMF staff calculations.
1/ Derived from prices of futures options on March 24-25, 2010.

although there is a 2-in-3 chance that copper prices could attain its 2008 average price by mid-2010. Except for corn and wheat prices, commodity prices, however, are very likely to be higher in 2010 than their levels during the first six months of 2009. By the end of 2010, there is a 1 in 5 chance that crude oil price could attain the average level of 2008, but it is still very unlikely that crude oil could attain its historical high of over $147 per barrel by end-2010. Also, the price of gold, which has been largely immune to the financial crisis, is expected to stay high.

- For non-commodities asset prices, the main story is that there is virtually no chance that either the Dow Jones Industrial Average or the S&P 500 would rebound to their pre-crisis levels (defined as end-June 2007) by the end of 2010. Relative to the Lehman collapse, there was a pretty small chance—around 11-13 percent— that these two indices would recoup their losses since the Lehman Brother collapse by June 2010. However, these two indices are almost certainly to exceed their 2009 Q1-Q2 levels by the end-2010.

- For the dollar/euro exchange rate, it is unlikely that the euro would be stronger against the US dollar by the end of 2010, compared with its pre-Lehman level.

- In addition, the 10-year US Treasury yield is very likely to be higher by the end of 2010 than its average level during the first six months of 2009.

The main statistical properties of the estimated distribution functions for the three-month-ahead (or closet) and for the eight- or nine-month-ahead (or closest) contracts as of mid-September 2009 and end-March 2010 are shown in Tables 3a and 3b. Figures 3 plot the estimated risk-neutral probability density functions (PDFs) for these contracts estimated during these two periods, which provide us not only a sense of the direction of expected price changes, but also with shifts in the perception of risks.

Main findings include:

- For most commodities, volatilities—as measured by the standard deviation—decreased or stayed roughly the same for all selected contracts during September 2009-March 2010, with the notable exception of corn and wheat. Similar conclusions can also be drawn by gauging the fatness and tallness in the charts of the risk-neutral PDFs probability density functions, with "taller" and "thinner" PDFs associated with a lower volatility and "fatter" and "shorter" PDFs associated with a higher volatility.

- Skewness for most assets has declined during September 2009-March 2010—except copper and corn.

- Most commodity prices have increased, with the distribution functions shifting to the right, except for the wheat price distribution, which has barely moved.

- Turning to non-commodities, the distributions for the S&P 500 and Dow Jones equity price indices have moved to the right while the dollar/euro exchange rate has moved to the left, reflecting the depreciation of euro during the period. The distribution for the

Treasury Note price has also moved to the left, reflecting expectations of higher long-term yield.

- Distributions for commodities and equity (represented by S&P 500) appear to be more positively skewed than those for the dollar/euro exchange rate and the Treasury bond price, which appear rather symmetric. This pattern is shown not only by a higher skewness measure, but also a lower median-to-mean ratios for commodities (except gold) . Among commodities, gold appears to be the least skewed.

- "Fat-tails" as indicated by kurtosis values also appear to be prevalent for most commodities (except gold) and equity, whose kurtosis values are significantly higher than three—the kurtosis value for a normal distribution. For the dollar/euro exchange rate and the Treasury bond price, the excess kurtosis is much closer to zero. For gold, the excess kurtosis is significantly smaller than those of other commodities.

Table 3a. Statisitical Properties for Three-Month Contracts or Closest (in prices)

	WTI	CCI	Gold	Copper	Corn	Wheat	S&P 500	Dow Jones	Dollar/Euro	Treasury Bond
As of mid-September 2009										
Spot	68.83	426.00	998.60	280.90	317.75	461.00	1051.70	9713.79	1.47	117.03
Futures	69.88	432.50	998.60	279.75	317.75	461.00	1051.70	9663.00	1.47	117.03
Expected Values	69.88	432.50	998.60	279.75	317.75	461.00	1051.56	9663.00	1.47	117.03
Mean 1/	69.88	432.50	998.60	279.75	317.75	461.00	1051.56	9663.00	1.47	117.03
Median	68.73	430.09	993.71	275.87	314.60	456.67	1039.18	9573.90	1.47	116.98
Median/Mean	0.98	0.99	1.00	0.99	0.99	0.99	0.99	0.99	1.00	1.00
Mode	67.15	425.62	983.22	269.55	308.57	451.59	1014.47	9394.17	1.46	116.81
Mode/Mean	0.96	0.98	0.98	0.96	0.97	0.98	0.96	0.97	0.99	1.00
Standard Deviation	14.50	44.87	101.41	49.08	45.43	72.58	164.05	1326.81	0.10	4.19
Dispersion 2/	0.21	0.10	0.10	0.18	0.14	0.16	0.16	0.14	0.07	0.04
Interquartile Range 3/	9.26	31.10	69.91	32.87	31.46	41.72	114.12	922.90	0.07	2.75
Skewness 4/	0.73	0.31	0.31	0.58	0.45	0.86	0.47	0.41	0.21	0.12
Kurtosis 4/	4.86	3.17	3.17	4.07	3.53	7.12	3.40	3.31	3.08	3.40
Excess Kurtosis 5/	1.86	0.17	0.17	1.07	0.53	4.12	0.40	0.31	0.08	0.40
As of end-March 2010										
	WTI	CCI	Gold	Copper	Corn	Wheat	S&P 500	Dow Jones	Dollar/Euro	Treasury Bond
Spot	81.00	477.50	1092.50	333.15	365.00	470.75	1169.70	10845.37	1.33	115.95
Futures	81.46	476.50	1093.90	334.10	376.00	483.25	1169.70	10790.00	1.33	115.95
Expected Values	81.46	477.33	1093.38	334.10	374.23	477.64	1139.98	10790.00	1.33	115.95
Mean 1/	81.46	477.33	1093.38	334.10	373.72	474.20	1139.81	10790.00	1.33	115.95
Median	80.97	475.16	1090.42	331.52	368.69	467.03	1131.80	10770.53	1.33	115.90
Median/Mean	0.99	1.00	1.00	0.99	0.99	0.98	0.99	1.00	1.00	1.00
Mode	79.95	471.10	1084.96	328.69	358.83	455.03	1126.53	10732.57	1.33	115.85
Mode/Mean	0.98	0.99	0.99	0.98	0.96	0.96	0.99	0.99	1.00	1.00
Standard Deviation	9.05	44.45	79.45	49.73	61.84	79.36	147.27	632.55	0.06	3.12
Dispersion 2/	0.11	0.09	0.07	0.15	0.17	0.17	0.13	0.06	0.05	0.03
Interquartile Range 3/	6.28	30.80	54.63	28.49	43.07	54.61	91.25	433.86	0.04	1.91
Skewness 4/	0.33	0.28	0.22	0.68	0.50	0.51	0.45	0.18	0.15	0.12
Kurtosis 4/	3.20	3.14	3.08	5.72	3.45	3.46	3.85	3.06	3.04	4.65
Excess Kurtosis 5/	0.20	0.14	0.08	2.72	0.45	0.46	0.85	0.06	0.04	1.65

Sources: Author's calculations

1/ The mean may be slightly different from the futures prices because of the discretization of the sample space.

2/ Dispersion is measured by the coefficient of variation given by the standard deviation divided by the mean.

3/ Interquartile range is calculated by the difference between the first and third quartile.

4/ Since both skewness and kurtosis have been standardized by the standard deviation, they can be compared across commoditites.

5/ Excess kurtosis is the kurtosis minus 3, since the kurtosis of a normal distribution is 3.

Table 3b. Statisitical Properties for Eight-Month Contracts or Closest (in prices)

	WTI	CCI	Gold	Copper	Corn	Wheat	S&P 500	Dow Jones	Dollar/Euro	Treasury Bond
As of mid-September 2009										
Spot	68.83	426.00	998.60	280.90	317.75	461.00	1051.70	9713.79	1.47	117.03
Futures	72.62	438.50	1000.90	281.15	331.25	493.75	1037.40	9551.00	1.47	115.19
Expected Values	72.62	431.30	1000.61	281.15	331.33	493.75	1037.40	9551.00	1.47	115.19
Mean 1/	72.62	431.29	1000.61	280.77	331.33	493.74	1037.40	9551.00	1.47	115.19
Median	68.77	427.11	982.72	264.77	323.82	481.56	1000.79	9374.74	1.46	115.11
Median/Mean	0.95	0.99	0.98	0.94	0.98	0.98	0.96	0.98	1.00	1.00
Mode	62.40	419.23	948.27	240.47	309.20	461.97	930.85	9029.87	1.45	114.88
Mode/Mean	0.86	0.97	0.95	0.86	0.93	0.94	0.90	0.95	0.99	1.00
Standard Deviation	25.87	59.87	191.32	107.63	71.87	116.94	283.74	1863.37	0.14	8.29
Dispersion 2/	0.36	0.14	0.19	0.38	0.22	0.24	0.27	0.20	0.10	0.07
Interquartile Range 3/	17.45	41.75	134.33	67.70	50.37	74.23	198.79	1306.64	0.10	5.15
Skewness 4/	1.15	0.42	0.58	1.63	0.66	1.27	0.84	0.59	0.29	0.00
Kurtosis 4/	5.81	3.31	3.61	10.03	3.79	9.24	4.28	3.63	3.15	4.93
Excess Kurtosis 5/	2.81	0.31	0.61	7.03	0.79	6.24	1.28	0.63	0.15	1.93
As of end-March 2010										
	WTI	CCI	Gold	Copper	Corn	Wheat	S&P 500	Dow Jones	Dollar/Euro	Treasury Bond
Spot	81.00	477.50	1092.50	333.15	365.00	470.75	1169.70	10845.37	1.33	115.95
Futures	83.16	482.50	1096.70	338.30	394.00	527.25	1155.40	10790.00	1.33	113.22
Expected Values	83.16	482.50	1096.70	338.30	389.63	483.80	1096.74	10790.00	1.33	113.22
Mean 1/	83.16	482.50	1096.70	338.30	389.29	480.06	1077.45	10790.00	1.33	113.22
Median	80.93	477.31	1082.23	327.19	375.26	462.79	1056.34	10770.53	1.32	113.29
Median/Mean	0.97	0.99	0.99	0.97	0.96	0.96	0.98	1.00	1.00	1.00
Mode	77.81	467.04	1071.30	311.69	348.79	433.60	1048.16	10732.57	1.31	113.35
Mode/Mean	0.94	0.97	0.98	0.92	0.90	0.90	0.97	0.99	0.99	1.00
Standard Deviation	22.72	72.12	199.45	96.61	107.33	127.16	225.55	632.55	0.13	6.24
Dispersion 2/	0.27	0.15	0.18	0.29	0.28	0.26	0.21	0.06	0.10	0.06
Interquartile Range 3/	13.89	50.14	115.82	62.14	75.02	87.80	124.00	433.86	0.09	3.65
Skewness 4/	1.20	0.45	0.75	1.06	0.85	0.81	0.87	0.18	0.29	-0.05
Kurtosis 4/	8.10	3.37	5.01	6.06	4.31	4.20	5.51	3.06	3.15	5.01
Excess Kurtosis 5/	5.10	0.37	2.01	3.06	1.31	1.20	2.51	0.06	0.15	2.01

Sources: Author's calculations

1/ The mean may be slightly different from the futures prices because of the discretization of the sample space.

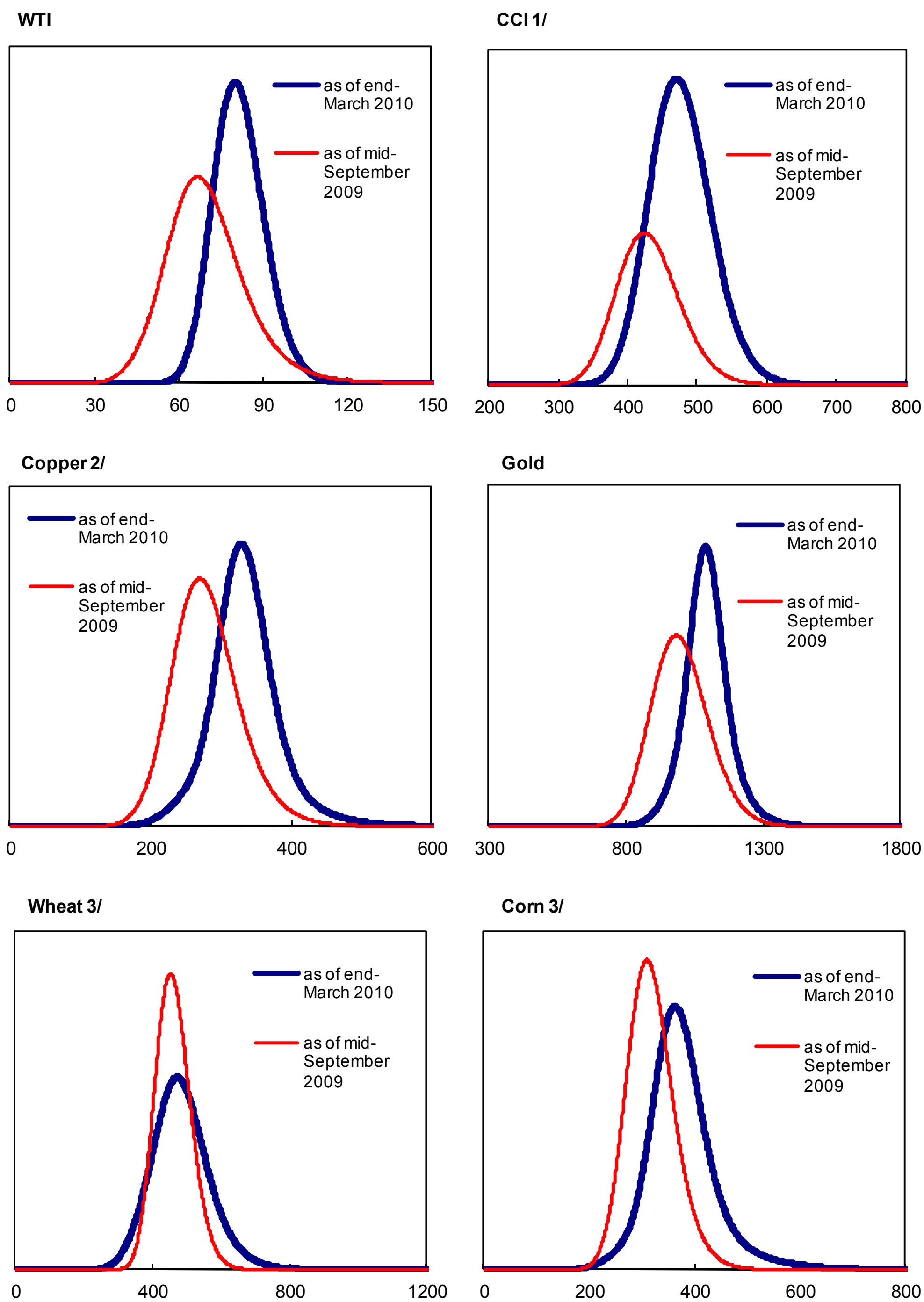
2/ Dispersion is measured by the coefficient of variation given by the standard deviation divided by the mean.

3/ Interquartile range is calculated by the difference between the first and third quartile.

4/ Since both skewness and kurtosis have been standardized by the standard deviation, they can be compared across commoditites.

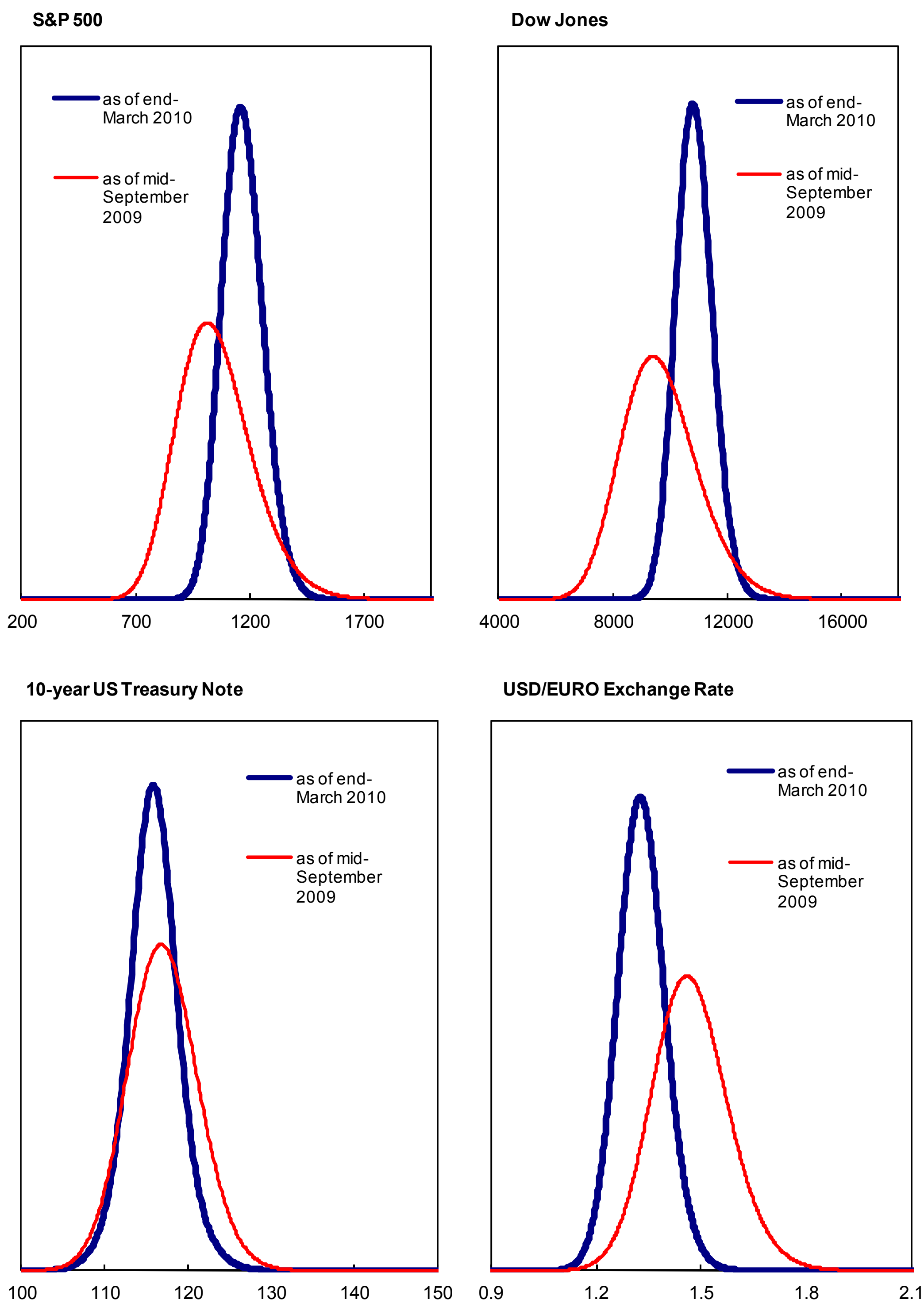
5/ Excess kurtosis is the kurtosis minus 3, since the kurtosis of a normal distribution is 3.

Figure 3a. Probability Density Functions for 3-month ahead (or closest) contracts as of mid-September 2009 and end-March 2010



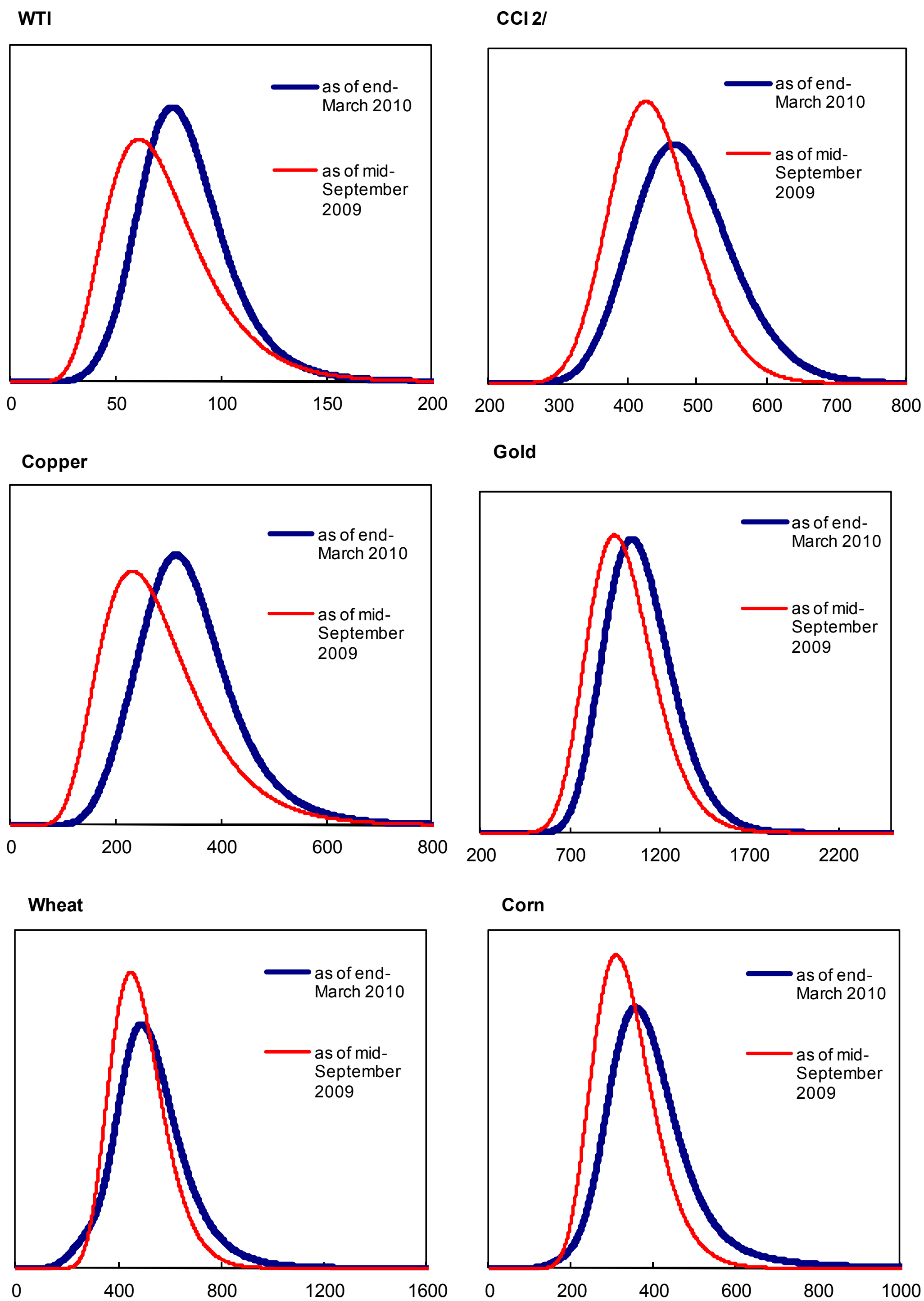
Source: Bloomberg and IMF staff calculations.
1/ Continuous Commodity Index: 1995 Revision of the Commodity Research Bureau Index; average of 17 commodity futures prices; traded at NYBOT.
2/ For copper, December 2009 (3-months forward) contract was not available in September 2009, so it was substituted with November 2009 (2-months forward) contract.
3/ For wheat and corn, June 2010 (3-months forward) contracts were not available in March 2010, so they were substituted with July 2010 (4-months forward) contracts.

Figure 3b. Probability Density Functions for 3-month ahead (or closest) contracts as of mid-September 2009 and end-March 2010



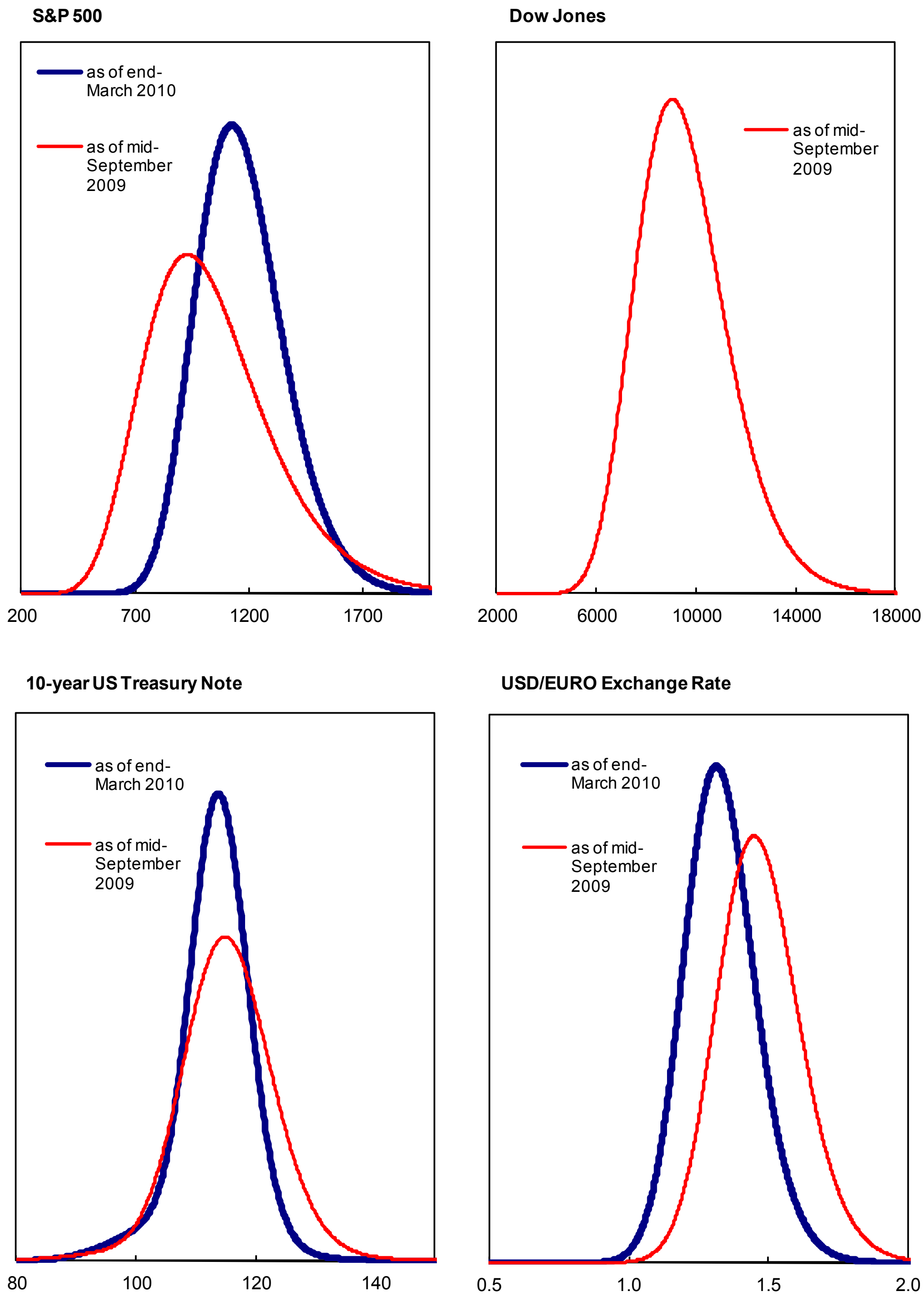
Source: Bloomberg and IMF staff calculations.

Figure 3c. Probability Density Functions for 9-month (or closest) ahead contracts as of mid-September 2009 and end-March 2010 1/



Source: Bloomberg and IMF staff calculations.
1/ For September 2009, 7-month contracts were used for CCI and gold, while 8-month contracts were used for the rest.
2/Continuous Commodity Index: 1995 Revision of the Commodity Research Bureau Index; average of 17 commodity futures prices; traded at NYBOT.

Figure 3d. Probability Density Functions for 9-month (or closest) ahead contracts as of mid-September 2009 and end-March 2010 1/



Source: Bloomberg and IMF staff calculations.
1/ Due to data availability, time to maturity might differ. For September 2009, June 2010 (9-months forward) contracts were used for S&P 500 and 10-year Treasury Note. For the Dow Jones and USD/EURO exchange rate, 6-month contracts were used. For March 2010, nine-month contracts were used except Dow Jones, where

C. Caveats

The results above should be interpreted with two caveats in mind: first, the probability distribution derived is the risk-neutral probability distribution, not the objective probability distribution of future events. In fact, if investors are risk-averse, the estimated risk-neutral probability would exaggerate the likelihood of an undesirable outcome. To see why, recall from equation (3) that the risk-neutral probability is the objective probability multiplied by the intertemporal marginal rate of substitution of consumption and the discount factor: i.e.

$$f^N(\theta) = e^{r\tau} M_\tau(\theta) f(\theta) = e^{r\tau} \frac{e^{-\rho\tau} U'(C_\tau(\theta))}{U'(C_0)} f(\theta).$$

If investors are risk-averse, their utility functions are concave. Since an undesirable outcome is associated with a lower consumption, the marginal utility of a bad state is higher because of the concavity of the utility function, thereby overstating the risk-neutral probability. Formally:

$$U'(C_\tau^{bad}) > U'(C_\tau^{good}) \text{ because } C_\tau^{bad} < C_\tau^{good} \text{ and } U''(\bullet) < 0.$$

Intuitively, a risk-averse investor is willing to pay a higher premium to insure against an unlikely but disastrous outcome than would a risk-neutral investor. For example, a risk-averse investor would be willing to pay a higher premium to purchase a put option to safeguard against an outcome of sharp stock price decline than if she were risk-neutral. If we estimate the probability of such an outcome based on the *actual observed* premium paid by this risk-averse investor—under the assumption that the investor is risk-neutral—the estimated probability would be higher than the *objective* probability.

To gauge the magnitude of this bias, the utility function is assumed to be the standard constant relative risk aversion (CRRA):

$$U(C) = \frac{C^{1-\gamma}}{1-\gamma}.$$

Then the ratio of the risk-neutral probability to the risk-averse probability can be expressed as:

$$\frac{f^N(\theta)}{f(\theta)} = e^{(r-\rho)\tau} \left[\frac{C_0}{C_\tau(\theta)} \right]^\gamma$$

For simplicity, let's assume that $\rho = r$. Then the ratio will be equal to the intertemporal rate of substitution of consumption, which in turn equals to the ratio of current consumption (which is known at time zero) to the future unknown consumption (which depends on the random variable, namely the asset price). $C_\tau(\theta)$ is then estimated by a simple reduced-reduced form.[21]

The result is presented in Table 4 for various risk aversion coefficient. The bias is very huge for some very "averse" outcomes (such as a S&P index below 200). However, for the likely outcome range, the bias is relatively modest.

The second caveat is the multi-lognormal approach—as well as all other approaches described in the previous section—are designed for European-styled options. Strictly speaking, to apply the techniques for American-style options—a family to which most of the liquid options belong—some adjustment would needs to be made for the early-exercise premium.[22]

In practice, however, it is rarely optimal for an American-styled owner to exercise the option before expiration, because the time value of an option is usually greater than the benefits for the early exercise. Therefore, many practitioners would just simply use European-styled approaches (such as the Black-Scholes model) to price an American-styled model.

A well-known complication arises when the underlying assets pay dividends or entail "convenience yield" in the case of commodities. In this situation, if the dividend payout or convenience yield before the expiration date is larger than the time value of the option, the option holder may have an incentive to exercise a call option in order to capture the ludicrous dividend payout or convenience yield. Fortunately, however, the underlying assets for this paper, are all futures contracts on dividend-paying stocks (S&P500 futures) or futures contracts on commodities. In other word, early exercise will entitle the investor to the futures contract, not the stocks that pay dividends or the physical commodities that give "convenience" for consumption or production.

---

[21] Ideally, such a relation should be estimated structurally. But this is not the focus of this paper. The aim here is merely to illustrate the relation between the risk-neutral and risk-averse probabilities.

[22] Typically, these methods are rather complex. One method is to obtain an implied volatility for the American-styled option using a binomial pricing model; then use the calculated implied volatility to calculate the price of an equivalent European-styled option with a desired maturity date, and then proceed with the multi-lognormal approach. Alternatively, as in Melick and Thomas (1994), some bounds can be derived to allow for the possibility of early exercise.

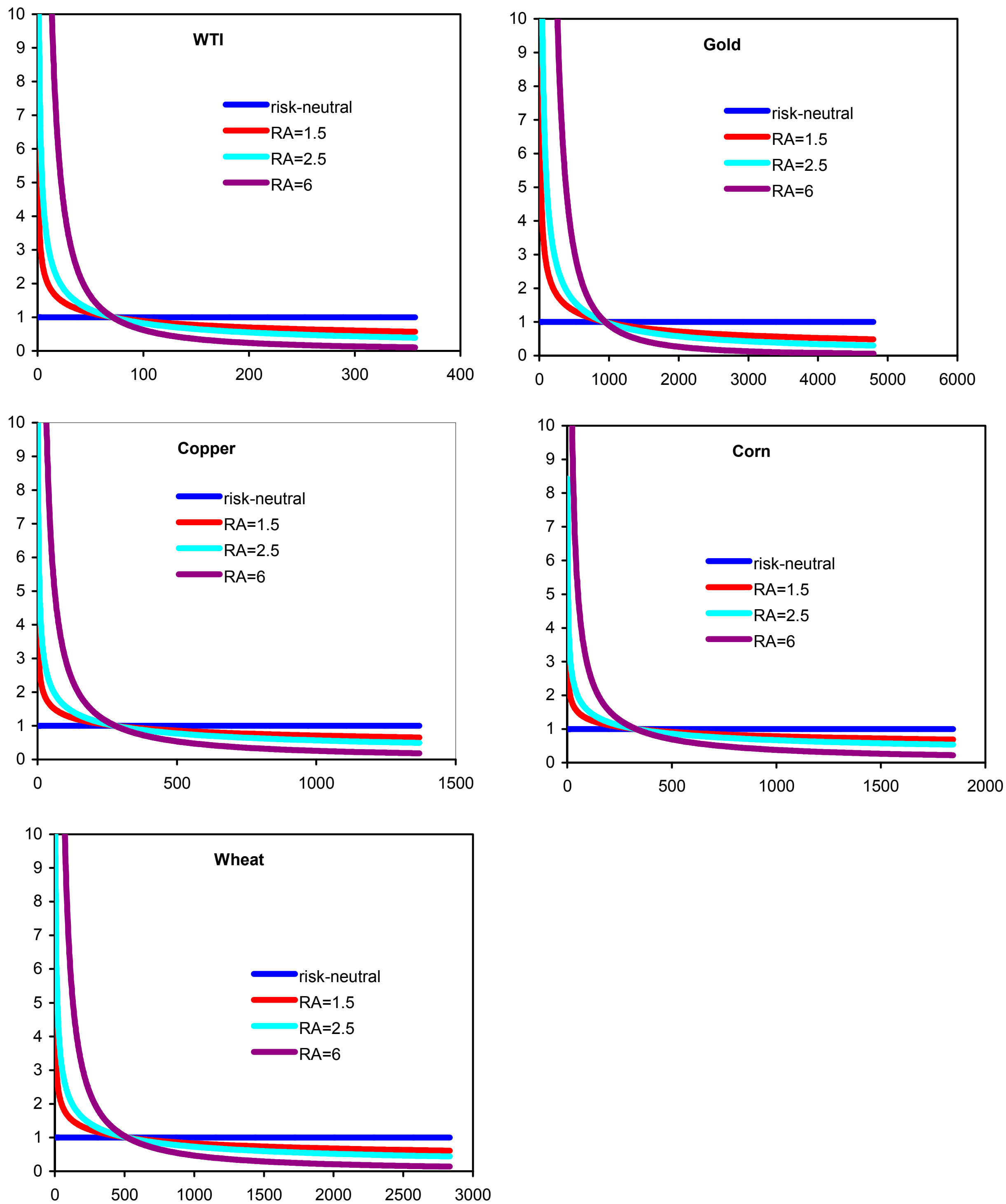
Table 4. Sum of Squared Errors for the Monte Carlo Study with 10,000 simulations 1/

	Sum of Squared Errors of PDF			Sum of Squared Errors of Estimated Options Prices		
	Double	Triple	Quadruple	Double	Triple	Quadruple
	When the true model is four lognormal					
Mean	0.0109	0.0042	0.0034	4.6781	0.3537	0.0709
Median	0.0052	0.0015	0.0012	0.0108	0.0000	0.0001
Maximum	0.2670	0.1213	0.1619	316.5097	250.5623	230.7967
Minimum	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
First Quartile	0.0018	0.0004	0.0003	0.0018	0.0000	0.0000
Third Quartile	0.0124	0.0048	0.0037	0.0355	0.0003	0.0004
Average between 1st-3rd quartile	0.0058	0.0019	0.0014	0.0132	0.0001	0.0001
	When the true model is three lognormal					
Mean	0.0103	0.0039	0.0030	1.6522	0.1950	0.0408
Median	0.0023	0.0007	0.0007	0.0006	0.0000	0.0000
Maximum	0.2765	0.3570	0.1286	550.9031	196.1020	95.9168
Minimum	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
First Quartile	0.0004	0.0002	0.0002	0.0001	0.0000	0.0000
Third Quartile	0.0102	0.0032	0.0027	0.0055	0.0001	0.0001
Average between 1st-3rd quartile	0.0032	0.0010	0.0009	0.0013	0.0000	0.0000
	When the true model is two lognormal					
Mean	0.0039	0.0016	0.0017	2.3540	0.2323	0.1883
Median	0.0000	0.0001	0.0001	0.0000	0.0000	0.0000
Maximum	0.3579	0.1993	0.2207	422.9787	279.6514	1304.9594
Minimum	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
First Quartile	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Third Quartile	0.0003	0.0005	0.0008	0.0001	0.0000	0.0001
Average between 1st-3rd quartile	0.0000	0.0001	0.0002	0.0000	0.0000	0.0000

Sources: The author's calculations

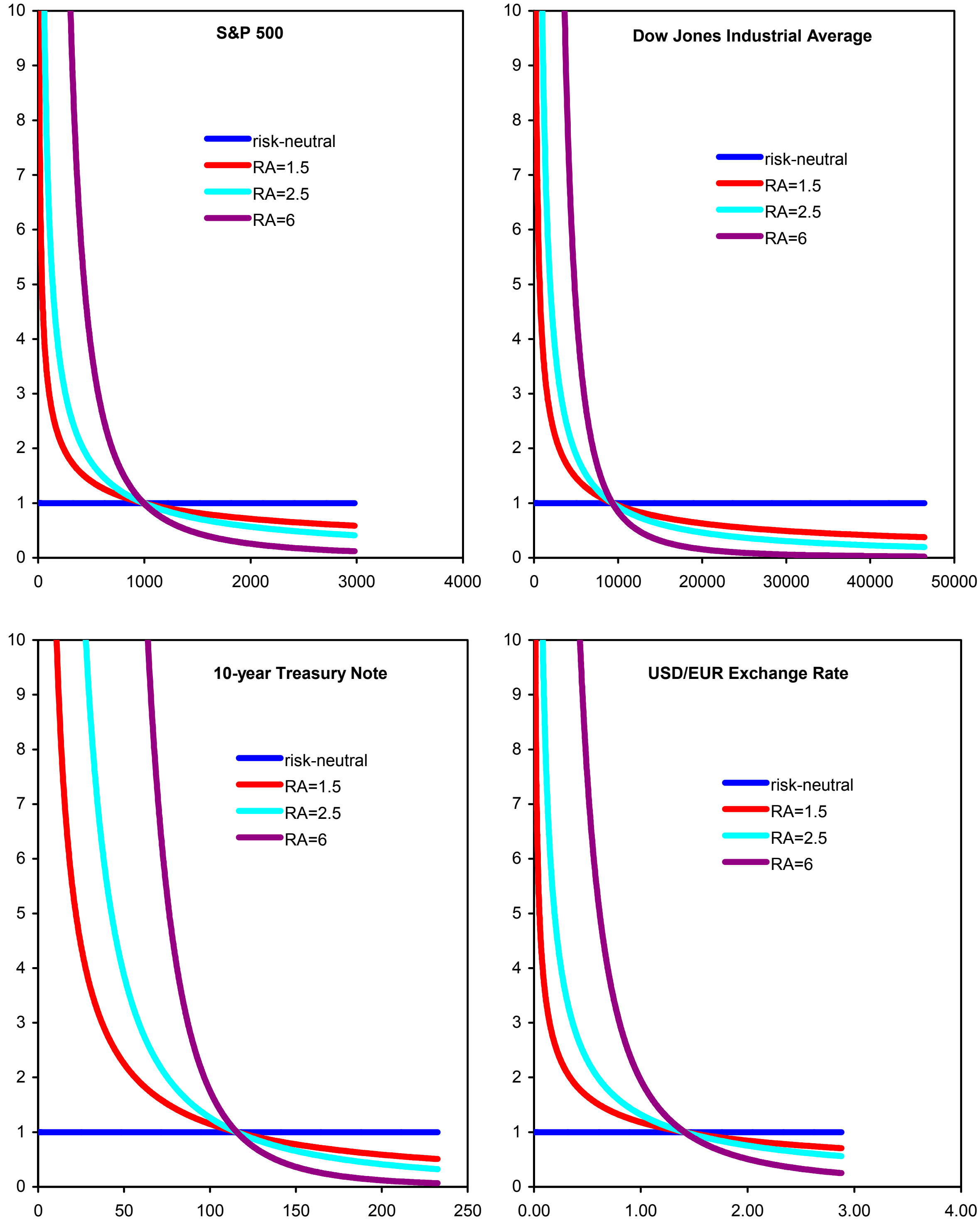
1/ For each simulation under the Monte Carlo, a sum of squared errors is calculated for each estimating technique. The errors are defined as the difference between the true distribution and the calculated distribution or between the actual options prices and the estimated options prices . The above statistics summarizes the outcomes of the 10,000 simulations.

Figure 4a. Commodities: Ratio of Risk-Neutral Probability to Risk-Averse Probability



Source: Bloomberg, L.P.; and IMF staff calculations

Figure 4b. Financial Securities: Ratio of Risk-Neutral Probability to Risk-Averse Probability



Source: Bloomberg, L.P.; and IMF staff calculations

V. A MONTE-CARLO SIMULATION

This section evaluates the procedure discussed in the previous section. One approach to test these techniques is to examine how accurately previously estimated distributions have predicted actual outcomes in the past. Since such an approach would require a large amount of time-series data on options/strikes prices, the data collection and management process could become a daunting task. In addition, since such as test would require ex-post actual data outturn, such an exercise is a joint test of how accurately the technique has estimated market expectation in addition to whether or not market expectations have been right in the first place.

Another approach is to assume the *true* RND of the underlying asset price and then simulate the artificial options price data. Next, the multi-lognormal procedure is used to recover the RNP distribution. The method can then be evaluated by gauging the "goodness of fit" between the *true* and *estimated* distributions.

Given that a wide range of papers have already evaluated the performance of the double-lognormal vis-à-vis other classes of methods,[23] this section focuses on the relative performance among the multi-lognormal class. Specifically, the performance of the quadruple-lognormal is compared with those of the triple- and the double-lognormal by means of a Monte Carlo.

For simplicity, the Monte Carlo simulation assumes that the *true* RND of the underlying asset prices at time τ (as of time zero) is a mixture of lognormal distributions of various orders:

$$f^{true}(S_\tau) = \sum_{i=1}^{I} \theta_i L(\alpha_i, \beta_i), \tag{12}$$

where $\alpha_i = \ln S_0 + (\mu_i - \frac{1}{2}\sigma_i^2)\tau$ and $\beta_i = \sigma_i\sqrt{\tau}$.

Three cases are considered: in the first case, the true RND is assumed to be a mixture of four lognormal; in the second case, the true RND is assumed to be a mixture of three lognormal; and in the third case, the true RND is assumed to be a mixture of four lognormal.
For each case, $\mu_i = \overline{\mu} + \varepsilon_i\overline{\sigma}$ and $\sigma_i = \xi_i\overline{\sigma}$. To ensure that the true RNP is true mixture of various lognormal, each of the ε_i and ξ_i is drawn randomly from uniform distributions on different intervals. For example, in the case where the true RNP is assumed to be four lognormals, ε_1 is drawn from the uniform distribution on the real interval *[-2, -1]*; ε_2 from the interval *[-1,0]*; ε_3 from the interval *[0,1]*; and ε_4 from the interval *[1,2]*; similarly, ξ_1 is drawn randomly from the interval *[1/3, 2/3]*; ξ_2 from the interval *[2/3,4/3]*; ξ_3 from the interval *[4/3,2]*; and ξ_4 from the interval *[2,3]*. Similar methods are applied for the other two cases.

[23] For example see Cooper (1999), and Syrdal (2002).

Te spot price, S_o, is randomly drawn from the uniform distribution on the real interval [65,80]. The futures price F_τ is given by the expected value of (9), which is equal to $S_0 \sum_{i=1}^{5} \theta_i e^{\mu_i \tau}$.

Thirty call contracts and thirty put contracts are generated, with prices given by equations (10') . For call contracts, the domain of the available strike prices is assumed to be $[0.8F_\tau, 1.5F_\tau]$, with the domain for available strike prices for put contracts being $[0.3F_\tau, 1.1F_\tau]$. Finally, the rest of the parameters are assumed to take the following values:[24] r=0.0040 (i.e. 0.40 percent); $\tau = 0.30; \overline{\mu} = -0.5; \overline{\sigma} = 0.8$.

Given the simulated call/put prices and their corresponding strike prices, the multi-lognormal technique discussed in the previous section is used to recover the RND. Then all three variations of the procedure with different numbers of mixtures—double-lognormal, triple-lognormal, and quadruple-lognormal—are implemented for each case.

The goodness of fit is measured by two measures of the sum of squared errors (SSE). First, a SSE related to the estimated and true RNP are calculated as follows:

$$SSE = \sum_x \left[f^{true}(x) - f^{estimated}(x) \right]^2$$

Similarly, another SSE related the difference between the actual observed options prices and the estimated prices—which is very similar to the objective function of the optimization problem in (10)—is given by:

$$\sum_{j=1}^{30} [\widetilde{C}_j - \hat{C}_j]^2 + \sum_{j=1}^{30} [\widetilde{P}_j - \hat{P}_j]^2$$

This procedure is repeated 10,000 times for the three variations of the multi-lognormal technique.[25] Then for each approach, the SSE is ranked from the lowest to the highest and the statistics are summarized in Table 5.

Overall, the quadruple-lognormal appears to outperform the triple-lognormal, which in turn outperforms the double-lognormal in most cases. Specifically, the quadruple the other two methods in terms of producing a smaller mean, median, first-quartile, and third-quartile SSEs

[24] These numbers are the actual values for the WTI September 2009 contracts as of April 2, 2009.

[25] Given the complex numerical optimization process for each procedure, this Monte Carlo study with 10,000 simulations took Matlab over 2 days to run on a high-performance computer on the IMF server. Before this "super" round, two *trial* rounds with only 1,000 simulations were run and the results were consistent with the main findings here.

than does others double-lognormal in all cases. The fact that the quadruple outperforms the double-lognormal when the true RNP is assumed to be a double-lognormal may seem puzzling. One plausible explanation may be that since the numerical procedure is based on a Newton-Method-type optimization procedure, the solution may not necessarily be the global optimum. Since the quadruple-lognormal increases the degree of freedom, it also produces a better solution.

VI. Conclusion and Further Studies

Building on the double-lognormal approach by Bahra (1997), this paper develops a multi-lognormal technique with transformation/restrictions to extract RNPs for a variety of assets. In general, the paper suggests that restrictions should be imposed to ensure economically-sensible results. On the empirical side, the paper finds that probability distributions for commodities except gold and S&P 500 are more skewed and have fatter tails than are for the dollar/euro exchange rate and the 10-year Treasury Note price.

References

Ait-Sahalia Yacine and Jefferson Duarte, 2003, "Nonparametric Option Pricing under Shape Restrictions," *Journal of Econometrics*, Vol. 116 (September-October), pp. 9–47.

Bahra, B., 1997, "Implied Risk-Neutral Probability Density Functions from Options Prices: Theory and Application," Working Paper No. 66, Bank of England.

Bates, David, 1991, "The Crash of 87: Was it Expected? The Evidence from Options Markets," *Journal of Finance,* Vol. 46 (July), pp. 1009-44.

Black, Fischer and Myron Scholes, 1973, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, Vol. 81 (May-June), pp. 637–54 (Chicago: The University of Chicago Press).

Breeden, Douglas and Robert Litzenberger, 1978, "Prices of State-Contingent Claims Implicit in Option Prices," *Journal of Business*, Vol. 51 (October), pp. 621–51 (Chicago: The University of Chicago Press).

Cochrane, John, 2001, *Asset Pricing*, Princeton University Press.

Cooper, Neil, 1999, "Testing Techniques for Estimating Implied RNDS From the Prices of European-Style Options," BIS Workshop, Basel, Switzerland.

Cox, John, C. and Stephen A. Ross, 1976, "The Valuation of Options for Alternative Stochastic Processes," *Journal of Financial Economics,* Vol. 3, No. 1-2, pp. 145–66.

Figlewski, Stephen, 2008, "Estimating the Implied Risk Neutral Density for the U.S. Market Portfolio," Working Paper.

Hull, John, 2005, *Options, Futures, and other Derivatives*, sixth edition, Prentice Hall.

Malz, Allan M., 1996, "Using Option Prices to Estimate Realignment Probabilities in the European Monetary System: the case of Sterling-Mark," *Journal of International Money and Finance*, Vol. 15 (October), pp. 717–48.

Melick, William and Charles Thomas, 1997, "Recovering an Asset's Implied PDF from Option Prices: An Application to Crude Oil During the Gulf Crisis," *Journal of Financial and Quantitative Analysis*, Vol. 32 (March), pp. 91–115.

Shimko, David, 1993, "*Bounds of Probability*," RISK, Vol.6 (April), pp. 33–37.
.
Syrdal, Stig Arild, 2002, "A Study of Implied Risk-Neutral Density Functions in the Norwegian Option Market," Norges Bank Working Paper, ANO 2002/13.