ABSTRACT

Name: Anthony J. F. Biagioli

Department: Economics

Title: Cross-Hedging with Futures and Options: Bivariate Lognormal and Other Distributions

Major: Economics

Degree: Doctor of Philosophy

Approved by:

Date:

Dissertation Director

04/03/07

NORTHERN ILLINOIS UNIVERSITY

ABSTRACT

Chang and Wong investigated the optimal hedging strategy for a multinational firm which has future cash flows in a foreign currency but is unable to directly hedge the exchange rate risk. The firm then uses a third currency to partially hedge the risk. This paper generalizes the paper of Chang and Wong by showing that some of the assumptions about the distributions of the stochastic process generating the exchange rates are more restrictive than necessary, i.e., that the same results hold under weaker assumptions. It then does specific calculations for the case of bivariate lognormal distributions and compares the results to those of Chang and Wong. Using the bivariate lognormal model with a term for inflation gives the best performance under a real-life data set.

NORTHERN ILLINOIS UNIVERSITY

CROSS-HEDGING WITH FUTURES AND OPTIONS:
BIVARIATE LOGNORMAL AND
OTHER DISTRIBUTIONS

A DISSERTATION SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

DOCTOR OF PHILOSOPHY

DEPARTMENT OF ECONOMICS

BY

ANTHONY J. F. BIAGIOLI

DEKALB, ILLINOIS

MAY 2007

UMI Number: 3272143

INFORMATION TO USERS

Certification: In accordance with departmental and Graduate School policies, this dissertation is accepted in partial fulfillment of degree requirements.

Nula Jan

Dissertation Director

03/28/07

Date

ACKNOWLEDGEMENTS

The author wishes to express his gratitude to Professors Neelam Jain, Susan Porter-Hudak, Evan Anderson, and Ardeshir Dalal for their guidance and encouragement in my graduate studies and in the production of this thesis.

I would also like to express my appreciation to my wife, without whose support I would neither have begun work on my degree in economics, nor would I have been able to carry it through.

DEDICATION

To Cathy, Natalie, Brandon, Vincent, and Meryl,

in gratitude for their patience with me while I worked through this

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This paper investigates some extensions of the paper by Chang and Wong (2003), hereafter referred to as CW. The situation investigated by CW is to find the optimal hedging strategy for a firm based on the following scenario. The firm is multinational and has a future cash flow in a foreign currency, but the firm is unable to directly hedge the exchange rate risk of that currency due to incomplete markets. The firm then attempts to partially hedge the risk using derivatives for the currency of a third country. In the industrialized countries of the world there are mature markets for currency futures and derivatives. For less developed countries such markets may be either absent or immature. Often the contracts which can be purchased are more akin to insurance contracts on the future cash flows, with large spreads to cover the risk assumed. The optimal cross-hedging strategy is investigated in both a one period model and in a multi-period, dynamic model.

CW derives the optimal hedge for the firm using futures and options, but their principal results are obtained under several assumptions. One assumption is that the firm maximizes expected utility with a quadratic utility function. Another is that the distributions for the exchange rates are bounded and symmetric and that when one exchange rate is regressed on the other the regression error is independent. A third assumption is that the volatilities of the exchange rates are constant over time. These assumptions are restrictive, but this is justified by the difficulty of getting concrete results without them.

The cash flows resulting from futures contracts are a linear function of the underlying exchange rate, while those resulting from options are nonlinear, as are those resulting from other derivatives. One of the frustrations of building models for hedging strategies is that most of the models which are simple enough to analyze in detail give the result that any use of non-linear instruments is not optimal. For instance, in the Black-Scholes model (Black and Scholes, 1973), all investors are risk-neutral, and this leads to the conclusion that there is no reason to use derivatives at all. The use of derivatives, including options, is typical of real-life behavior, since derivatives markets are very large and options are a considerable portion of this market. Bodnar et al. (1995, 1996, 1998) attempted to survey U.S. firms on their use of financial derivatives, and this was followed up by several related surveys and comparative studies by Bodnar and Gebhardt (1999), Bodnar, de Jong, and MacRae (2003), Børsum and Ødegaard (2005), and Jakoniuk et al. (2000). These surveys show that many firms, especially the smaller ones, use only futures for hedging purposes, but a significant proportion of the larger companies use options as well. In using these survey results, one should keep in mind that they are limited by self-selection, since, as is typical of surveys, the response rates were around 20%–25%. A detailed series of surveys of derivatives usage, specific to the gold mining industry, was made by Ted Reeve (1991, 1993, 1994), and these were used by Tufano (1996) to analyze the motivations behind hedging behavior. The gold mining industry has a tradition of publicly declaring such details of its operations and so is amenable to the construction of such a data set.

The fact that CW obtained an optimal hedge which used nonlinear derivatives, namely options, is therefore noteworthy in itself. Our results allow this analysis to be extended to models for the exchange rate process that more closely approach the real-world processes.

This paper shows that the assumptions about the distributions of the exchange rates are unnecessary. The assumption that the exchange rate processes are symmetric and bounded, with an independent error for the regression, and the assumption that the volatility between periods is nonrandom and independent over time can be omitted without any change in the conclusions. The same results that CW obtained are shown to still be true when these assumptions are relaxed, both in the one period model and in the dynamic model. This allows for consideration of more realistic assumptions about the stochastic process governing the exchange rates. We then apply these results to the case of a bivariate lognormal distribution. We obtain explicit formulas for the optimal hedge, and we compare the performance of this hedging strategy for the same three countries used by CW: Japan, Taiwan, and the U.S., over the period 1997 through 2001. The hedge based on the assumption of a bivariate lognormal distribution does do better at minimizing the variance of the cash flows, and a model including a trend term for the exchange rate improves the hedging result a bit more.

One assumption from CW is still preserved, the assumption that the firm's utility is quadratic. A satisfactory theory must deal with something more general than the quadratic utility function, but our analysis depends in an essential way on this assumption.

The outline of this thesis is as follows. In Chapter 2 we develop the basic model for the case of one period, and in Chapter 3 we develop the basic dynamic model, which involves periods $1, \ldots, T$. Then in Chapter 4 we implement the details of the one period model for the case of bivariate lognormal distributions, and in Chapter 5 we implement the details of the dynamic model for bivariate lognormal distributions. Chapter 6 discusses an adaptation of Newton's method for a system of equations, which is necessary for finding the maximum likelihood estimators (MLEs) in variations of the model. Chapter 7 then shows how to calculate these MLEs. In Chapter 8 we calculate the estimators and compare the performance of the variations on the model for the bivariate normal distributions and the bivariate lognormal distributions. Chapter 9 discusses the results of simulations which were run to assess the consistency of the underlying model and its parameters. Finally, Chapter 10 discusses the overall conclusions and considers some possible directions for future research.

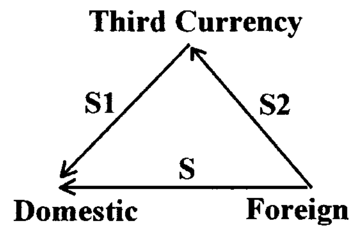
CHAPTER 2

THE ONE PERIOD MODEL

In this chapter we discuss the general outlines of our model for cross-hedging with currency futures and options and develop the consequences of this model as far as possible without knowing the specifics of the underlying distributions.

For the one period model, the situation is this. A firm is expecting a future cash flow of X, denominated in a foreign currency, which is known with certainty. The exchange rate to the domestic currency at the end of the period is a random variable S, so that the cash flow converted to domestic currency will be SX. (I shall not use tildes to distinguish random variables from their realizations, as done in CW, but will rely on context or a specific comment for this.) The firm wishes to hedge its risk, but there is no market, or an inadequate one, for exchange rate derivatives in the foreign currency. Instead, the firm will hedge the cash flow with derivatives of a third currency, one which does have a mature market.

Let S_1 be the random variable for the exchange rate from the third currency into the domestic currency, and let S_2 be the random variable for the exchange rate of the foreign currency into the third currency, so that $S = S_1 S_2$. We assume that the firm's internal assessment of future exchange rates is objectively accurate, so all expectations are taken with the same distributions.



To hedge the cash flow the firm will sell H futures contracts and Z put options. We allow that the firm might decide to buy these derivatives, instead of selling them, by permitting H and Z to be negative.

We are assuming transaction costs to be zero, so the put-call parity relation shows that allowing the purchase or sale of call options would be redundant. The price F of a futures contract is assumed to be revenue neutral, so that it is equal to the expected value $F = \bar{S}_1 = E[S_1]$. The put options, for simplicity, are assumed only to be available at a single strike price, which we take to be $\bar{S}_1$. Since this is a one period model, there's no sense in asking whether the options are American or European. To simplify the analysis, we assume the interest rate to be zero. Then for the futures contract, it receives $\bar{S}_1 H$ at the beginning of the period and pays out $S_1 H$ at the end of the period, for a net cash flow of $(\bar{S}_1 - S_1)H$. If the interest rate were a deterministic amount $\delta \neq 0$, then the price $F = E[\frac{1}{1+\delta} S_1] = \frac{1}{1+\delta} E[S_1]$ would reflect this, and the cash flow at the end of the period would be the same. If δ were random, then a considerable increase in complexity would result, and we are avoiding this. We let m represent the random variable for the cash outflow at the end of the period from selling one put option. Thus

$$m = \max(0, \bar{S}_1 - S_1) = \begin{cases} 0, & \text{if } S_1 \geq \bar{S}_1 \\ \bar{S}_1 - S_1, & \text{if } S_1 \leq \bar{S}_1. \end{cases} \tag{2-1}$$

We also assume the price P of a put option is such that the cash flow at the end of the period is revenue neutral, $P = E[m]$. For the put option, the firm's net cash flow for the period is $(P-m)Z$. The cash flow for the period, including the results of the hedge, is now

$$\Pi = \Pi(H,Z) = S_1 S_2 X + (\bar{S}_1 - S_1)H + (P-m)Z. \tag{2-2}$$

We assume that the firm maximizes expected utility for a utility function $U(\Pi)$, so its objective is to maximize the function

$$\mathcal{O}(H,Z) = E\big[U(\Pi(H,Z))\big]. \tag{2-3}$$

The minimum assumptions necessary for this to make sense are that U is increasing and strictly concave. With only these assumptions, we can only conclude that U is continuous, U has a left derivative $U'(\Pi^-)$ and a right derivative $U'(\Pi^+)$ at every point; that these are decreasing, in the sense that $\Pi_1 < \Pi_2$ implies

$$U'(\Pi_1^-) \geq U'(\Pi_1^+) > U'(\Pi_2^-) \geq U'(\Pi_2^+) > 0; \tag{2-4}$$

and that, if it exists, $U''(\Pi) \leq 0$. We note here that even if $U''(\Pi)$ exists everywhere, strict concavity does not imply $U''(\Pi) > 0$ for every Π; it only implies that every interval contains one point at which the inequality is strict. In order to apply standard calculus techniques to the maximization problem, it's necessary that $\mathcal{O}(H,Z)$ have continuous first and second derivatives at every point. Therefore, we first assume that $U'(\Pi)$ and $U''(\Pi)$ exist for every Π.

Some further assumptions are necessary, since rather than the derivatives of $U(\Pi)$, we want the derivatives of $\mathcal{O}(H,Z)$. The ensuing analysis depends on being able to take each of the operators $\frac{\partial}{\partial H}$, $\frac{\partial}{\partial Z}$, $\frac{\partial^2}{\partial H^2}$, $\frac{\partial^2}{\partial H \partial Z}$, and $\frac{\partial^2}{\partial Z^2}$ inside the expectations operator, as, for example, in

$$\frac{\partial}{\partial H}\Big(E[U(\Pi)]\Big) = E\Big[\frac{\partial}{\partial H}U(\Pi)\Big]. \tag{2-5}$$

For our purposes, the simplest assumptions which permit this are, first of all, that U is a thrice-continuously differentiable function and, second, that all of the moments

$$\begin{aligned} &E[U(\Pi)], \quad E[(\bar{S}_1 - S_1)U'(\Pi)], \quad E[(P-m)U'(\Pi)], \\ &E[(\bar{S}_1 - S_1)^k (P-m)^{2-k} U''(\Pi)], \quad \text{where } k = 2,1,0, \quad \text{and} \\ &E[(\bar{S}_1 - S_1)^k (P-m)^{3-k} U'''(\Pi)], \quad \text{where } k = 3,2,1,0, \end{aligned} \tag{2-6}$$

are finite. Then the generalized mean value theorem says that for any $\Delta\Pi$, there is a $\Delta_3\Pi$ which lies in the range $0 < \Delta_3\Pi < \Delta\Pi$ or $\Delta\Pi < \Delta_3\Pi < 0$, whichever is relevant, so that

$$U(\Pi + \Delta\Pi) = U(\Pi) + \Delta\Pi \cdot U'(\Pi) + \tfrac{1}{2}(\Delta\Pi)^2\, U''(\Pi) + \tfrac{1}{6}(\Delta_3\Pi)^3\, U'''(\Pi). \tag{2-7}$$

Then, for instance, if $\Delta\Pi = \Pi(H + \Delta H, Z) - \Pi(H, Z) = (\bar{S}_1 - S_1)\Delta H$, we have

$$\begin{aligned} \frac{\partial}{\partial H}\Big(E[U(\Pi)]\Big) &= \lim_{\Delta H \to 0} E\Big[\frac{U(\Pi + \Delta\Pi) - U(\Pi)}{\Delta H}\Big] \\ &= \lim_{\Delta H \to 0} \Big\{ E[(\bar{S}_1 - S_1)U'(\Pi)] + \tfrac{1}{2}\Delta H \cdot E\Big[(\bar{S}_1 - S_1)^2\, U''(\Pi)\Big] \\ &\qquad\qquad + \tfrac{1}{6}(\Delta_3 H)^2 E\Big[(\bar{S}_1 - S_1)^3 U'''(\Pi)\Big]\Big\} \\ &= E\Big[(\bar{S}_1 - S_1)U'(\Pi)\Big]. \end{aligned} \tag{2-8}$$

Similar computations show that all of the other derivative operators can be taken inside the expectations operator. We proceed under the assumption that the moments in (2-6) are finite, so that this change of order is justified.

The first-order conditions for the maximum are then

$$\begin{cases} E[U'(\Pi)(\bar{S}_1 - S_1)] = 0, \\ E[U'(\Pi)(P - m)] = 0, \end{cases} \tag{2-9}$$

and the second-order condition, since there are no constraints, is that the Hessian $\mathcal{H}$ be negative semi-definite. For any $h = [h_1, h_2]'$, we calculate

$$h'\mathcal{H}h = E\left[U''(\Pi)\left\{h_1(\bar{S}_1 - S_1) + h_2(P - m)\right\}^2\right] \leq 0, \tag{2-10}$$

so we see that $\mathcal{H}$ is negative semi-definite, and if we make the very mild assumption that

$$\text{for all } h \neq [0,0]', \qquad \text{Prob}\left[U''(\Pi)\left\{h_1(\bar{S}_1 - S_1) + h_2(P - m)\right\}^2 \neq 0\right] > 0, \tag{2-11}$$

we can say in addition that $\mathcal{H}$ is negative definite, so that the solution to the first-order equations, if it exists, is unique and is the global maximum.

To see how mild this condition really is, since $m = \max(\bar{S}_1 - S_1, 0)$, we can say that for any fixed h_1 and h_2, the quantity $\left\{h_1(\bar{S}_1 - S_1) + h_2(P - m)\right\}$, as a function of S_1, consists of two straight-line portions, one to either side of $S_1 = \bar{S}_1$. Consequently, we can have $\left\{h_1(\bar{S}_1 - S_1) + h_2(P - m)\right\} = 0$ at two values of S_1 at the most. Then at all other values of S_1, we can have $U''(\Pi)\left\{h_1(\bar{S}_1 - S_1) + h_2(P - m)\right\}^2 = 0$ only when $U''(\Pi) = 0$, and the set $U''(\Pi) \neq 0$, when these two values are excluded, must have probability zero. It is possible to have strictly convex functions U with many points at which $U''(\Pi) = 0$, but for most of the models used in the literature, we have $U''(\Pi) < 0$ at all Π, and for these models the condition (2-11) can fail only when S_1 is concentrated at two points. In particular, this is the case under the assumption of quadratic utility, which we now consider.

The above holds for any $U(\Pi)$, but for the rest of this section we assume that

$$U(\Pi) = \Pi - b\Pi^2 \tag{2-12}$$

for some $b > 0$. Every quadratic utility function is equivalent to one with this form. Its marginal utility $U'(\Pi) = 1 - 2b\Pi$ is linear, and $U'' = -2b < 0$ is a constant. It is defined only for $\Pi \leq 1/2b$.

Quadratic utility is always disconcerting to deal with because there is an amount of money, $1/2b$, beyond which utility is undefined. If S_1 and S_2 have a distribution which allows arbitrarily large values, then Π will sometimes be assuming meaningless values with positive probability. We ultimately have to rationalize this by assuming that the true utility function is defined for all values of Π and is equal to the quadratic utility on a region including the global maximum. This throws off the accuracy of the first-order conditions (2-9), since the expectations involved in them are affected by all values of $U(\Pi)$, not just the values near the optimum. To remedy this we have to assume, in addition, that the probability of events outside the region where U is quadratic is so small that the solution of the first-order conditions under the assumption of quadratic utility is a good approximation to the true maximum. We shall proceed under these assumptions.

Under quadratic utility, since $U(\Pi)$ is quadratic and Π is linear, in order for the moments in (2-6) to be finite, it is equivalent that the moments of S_1 and S_2 of orders 0, 1, and 2 be finite.

Since marginal utility is now linear, the first-order conditions simplify to

$$\begin{cases} E[\Pi(\bar{S}_1 - S_1)] = -\mathrm{Cov}[\Pi, S_1] = 0 \\ E[\Pi(P - m)] \;\; = -\mathrm{Cov}[\Pi, m] = 0. \end{cases} \tag{2-13}$$

Before proceeding, we want to note that these same equations produce the optimizing values of H^* and Z^* for the expected value of any quadratic function of Π. In particular, the variance of Π is the expectation of a quadratic function of Π, and so it is minimized for exactly the same values of H^* and Z^* which maximize $E[U(\Pi)]$. This is useful in Chapter 5, where we calculate the variances of the cash flows, since this result shows that, under quadratic utility, minimizing the variance is equivalent to maximizing the expected utility.

Applying the definition (2-2) of Π we can rewrite (2-13) as

$$\begin{cases} A_{11}H + A_{12}Z = B_1 X \\ A_{21}H + A_{22}Z = B_2 X, \end{cases} \tag{2-14}$$

where

$$\begin{aligned} A_{11} &= \mathrm{Var}[S_1] \\ A_{12} = A_{21} &= \mathrm{Cov}[S_1, m] \\ A_{22} &= \mathrm{Var}[m] \\ B_1 &= \mathrm{Cov}[S_1 S_2, S_1] \\ B_2 &= \mathrm{Cov}[S_1 S_2, m]. \end{aligned} \tag{2-15}$$

The determinant of this system of equations is thus

$$\Delta = \mathrm{Var}[S_1]\mathrm{Var}[m] - \mathrm{Cov}[S_1, m]^2 \quad = \mathrm{Var}[S_1]\mathrm{Var}[m](1 - R^2), \tag{2-16}$$

where R is the correlation coefficient between S_1 and m. When $\Delta \neq 0$ the solution to the system of equations will exist and be unique, so we examine the case $\Delta = 0$. This is equivalent to having an equation $p(S_1 - \bar{S}_1) + q(m - P) = 0$ being true on a set of probability 1, for some p, q, not both zero. (This is the condition for having equality in the Cauchy-Schwarz inequality $\mathrm{Cov}[S_1, m]^2 \leq \mathrm{Var}[S_1]\mathrm{Var}[m]$.)

The case $q = 0$ is equivalent to the case that S_1 is a point mass, so we now assume that S_1 is not a point mass.

Suppose S_1 is concentrated at two points $S_1^- < S_1^+$ and that $\text{Prob}[S_1 = S_1^+] = p$. Then $\bar{S}_1 = pS_1^+ + (1-p)S_1^-$, so that $m = 0$ at $S_1 = S_1^+$ and $m = p(S_1^+ - S_1^-)$ at $S_1 = S_1^-$. Consequently $P = p(1-p)(S_1^+ - S_1^-)$, and we see that $p(S_1 - \bar{S}_1) + (m - P) = 0$ is true at both $S_1 = S_1^-$ and at $S_1 = S_1^+$. It follows that $\Delta = 0$ when S_1 is concentrated at two points.

Conversely, suppose $\Delta = 0$ and S_1 is not a point mass. Take $p \neq 0$ such that $p(S_1 - \bar{S}_1) + (m - P) = 0$ on a set of measure 1. There must be at least one value $S_1^+ \geq \bar{S}_1$ for which $S_1 = S_1^+$ occurs with positive probability (or probability density). But for this value of S_1 we have $m = 0$, and so we can calculate $S_1^+ = \bar{S}_1 + \frac{1}{p}P$. This is then the only value greater than or equal to $\bar{S}_1$ for which the probability is positive. Similarly, the only value less than or equal to $\bar{S}_1$ for which the probability is positive is $S_1^- = \bar{S}_1 - \frac{1}{1-p}P$. From $S_1^- < S_1 < S_1^+$ we obtain $0 < p < 1$, and from $E[m] = P$ and $E[S_1] = \bar{S}_1$ we find that $\text{Prob}[S_1 = S_1^-] = 1 - p$ and $\text{Prob}[S_1 = S_1^+] = p$. Hence S_1 is concentrated at two points.

This proves that $\Delta = 0$ if and only if S_1 is concentrated at one or two points. In these cases the equations will not have a unique solution but will have two degrees or one degree of freedom, respectively. However, these cases are not interesting for our hedging problem. If S_1 is concentrated at one point, then there is no risk to hedge. If S_1 is concentrated at two points, then, as mentioned above, we have $(m - P) = -p(S_1 - \bar{S}_1)$ a.e., so the two instruments are equivalent; you can hedge whatever risk there is equally well with either derivative. Since these cases have no interest to us, we shall omit them from all future discussion. We have proved the following.

Proposition 1. *Under quadratic utility, so long as S_1 is not concentrated at one or two points, the hedging problem has a unique global maximum.*

We want to make a comparison with the results in CW. By Cramer's rule the solution to the hedging problem is

$$\begin{cases} H^* = \quad (A_{22}B_1 - A_{12}B_2)/\Delta \\ \\ Z^* = (-A_{21}B_2 + A_{11}B_2)/\Delta. \end{cases} \tag{2-17}$$

The solution in proposition 2 of CW has a distinctly different appearance. Specifically, if we let H^0 be the optimal number of futures contracts to sell if options are not used, then combining (CW-14) and (CW-27), their results say

$$H^* = H^0 + \tfrac{1}{2}Z^*. \tag{2-18}$$

This can be reconciled as follows. To solve for H^0, we ignore the second equation in (2-14) and solve the first equation with Z set equal to 0. Doing so, we see that

$$H^0 = \frac{B_1}{A_{11}}X. \tag{2-19}$$

Then from (2-17), with no restrictions on Z, we get

$$H^* = H^0 + \beta Z^*, \tag{2-20}$$

where

$$\beta = \frac{-A_{12}}{A_{11}} = -\frac{\mathrm{Cov}[S_1, m]}{\mathrm{Var}[S_1]}. \tag{2-21}$$

The covariance of S_1 and m is

$$\mathrm{Cov}[S_1, m] = E[(S_1 - \bar{S}_1)m] = -\int_{-\infty}^{\bar{S}_1} (\bar{S}_1 - S_1)^2 f(S_1)\, dS_1, \tag{2-22}$$

which shows that $\beta > 0$ is always true. CW assumed symmetry, and from this follows

$$\mathrm{Cov}[S_1, m] = -\tfrac{1}{2}\int_{-\infty}^{\infty} (\bar{S}_1 - S_1)^2 f(S_1)\, dS_1 = -\tfrac{1}{2}\mathrm{Var}[S_1], \tag{2-23}$$

so that $\beta = \frac{1}{2}$ and we have (2-18), which shows that we can obtain the same format for our answer as CW did.

This concludes the development of the basic model, insofar as we can know it without having details of the underlying distributions. In the next chapter we do the same for the dynamic model, with a discrete number of periods $1, \ldots, T$.

CHAPTER 3

THE DYNAMIC MODEL

We now turn to the dynamic model, reverting for the moment to the situation in which no assumptions are made about the distribution of S_1 and S_2. For the next few paragraphs, until equation (3-8), we shall suspend the assumption that utility is quadratic and discuss, for an arbitrary utility function, the time dependence of knowledge of the optimal hedge.

We suppose the firm is facing a series of cash flows $X_1, \ldots, X_T$ at the end of each of the next T periods, all of the periods being the same length. Time is now measured from $t = 0$ to $t = T$, in integer increments, and period t runs from time $t-1$ to time t. Instead of making one hedge, the firm will make a separate hedge for each period, depending on the results of the previous periods and its expectations for the future. For period t, the firm will at time $t-1$ sell H_t futures contracts and Z_t options, which mature at time t. At time t these will be cashed out, and if $t < T$, a new hedge of H_{t+1} futures contracts and Z_{t+1} options will be established.

The exchange rates at time t are $S_{1,t}$ and $S_{2,t}$, and these form a stochastic process. That is, there is some collection ω_t of random variables for period t, that determines $S_{i,t}$ and whose distributions may depend on the realizations ω_τ for $\tau < t$. Let ω_0 be a set of initial values and parameters, and let $\Omega_t = \omega_0 \cup \ldots \cup \omega_t$. In order to deal with expectations for random variables over different lengths of time, the notation $E_{t_1..t_2}[F]$ for $t_1 \leq t_2$ will be used to denote the expectation of F that is expected at time t_2, as seen from time $t_1 - 1$. That is, $\omega_0, \ldots, \omega_{t_1-1}$ have been realized, and the expectation is being taken over the set of variables $\omega_{t_1} \cup \ldots \cup \omega_{t_2}$, conditional on these realizations. When there is only one period t involved, this will be shortened to $E_t[F] = E_{t..t}[F]$.

The net cash flow at the end of period t is

$$\Pi_t = S_{1,t}S_{2,t}X_t + (\bar{S}_{1,t} - S_{1,t})H_t + (P_t - m_t)Z_t, \tag{3-1}$$

and the accumulated cash flows over the first t periods is

$$W_t = \sum_{\tau=1}^{t} \Pi_\tau. \tag{3-2}$$

The company's aim is to maximize its expected utility of W_T, but it does not have to choose H_t and Z_t until time $t - 1$, so this is a nested series of optimization problems. At time $t - 1$ the firm knows the realizations of Ω_{t-1} and it also knows its past decisions $H_1, Z_1, \ldots, H_{t-1}, Z_{t-1}$, and can only plan on the future choices for $H_{t+1}, Z_{t+1}, \ldots, H_T, Z_T$. The optimal solutions H_t^* and Z_t^* will be functions of some or all of these quantities. Note that when we are considering the period t optimization, the quantities H_τ or Z_τ for $\tau < t$ are presumed to have been chosen earlier, and so they are constants, but they can assume any feasible values. Also, the variables of Ω_{t-1} are realized and are also treated as constant. The solution for the hedging problem is done one period at a time, working backwards from time T.

The optimization problem for period T, the last period, is to take the values of Ω_{T-1} and $H_1, Z_1, \ldots, H_{T-1}, Z_{T-1}$ as given and to choose H_T^* and Z_T^* so as to maximize

$$\mathcal{O}_T(H_T, Z_T) = E_T[U(W_T)]. \tag{3-3}$$

Here $H_1, Z_1, \ldots, H_{T-1}, Z_{T-1}$ are not assumed to necessarily be optimal, but simply whatever choices the firm has made in the past, and all of the variables in Ω_{T-1} are realized. We denote the solution to this problem by H_T^* and Z_T^*. These will be functions of $\Omega_{T-1}, H_1, Z_1, \ldots, H_{T-1}$, and Z_{T-1}. Knowledge of H_T^* will be different at different points in time, so we denote by $H_T^{*(t)}$ the value of H_T^* as it is known at time t. Accordingly, this solution is denoted $H_T^{*(T-1)}$ and $Z_T^{*(T-1)}$. Similarly $H_\tau^{*(t)}$ and $Z_\tau^{*(t)}$ for any $t < \tau$ will denote H_τ^* and Z_τ^*, the solutions to the optimization problem for period τ, as they are seen from time t. Once these are known, let

$$\Pi_\tau^{*(t)} = S_{1,\tau}S_{2,\tau}X_\tau + (\bar{S}_{1,\tau} - S_{1,\tau})H_\tau^{*(t)} + (P_\tau - m_\tau)Z_\tau^{*(t)} \tag{3-4}$$

be the optimized cash flow for period τ, as seen from time t, and let

$$W_{t,T}^* = \sum_{\tau=1}^{t} \Pi_\tau + \sum_{\tau=t+1}^{T} \Pi_\tau^{*(t)}. \tag{3-5}$$

All of the variables $S_{i,\tau}$ might also be noted as depending on the time t, since their distributions may be conditional on the variables Ω_t, which are realized at time t, and also on the variables $\omega_{t+1} \cup \ldots \cup \omega_{\tau-1}$, which are still random variables. However, we will leave this dependence on time as implicit.

Suppose that at some time $t \leq T-1$ all of the future optimal choices $H_\tau^{*(t)}$ for periods $\tau > t$ are known. We consider the firm's optimization problem for period t, which is done at time $t-1$. It needs to choose $H_t^* = H_t^{*(t-1)}$ and $Z_t^* = Z_t^{*(t-1)}$ so as to maximize

$$\mathcal{O}_t(H_t, Z_t) = E_{t..T}[U(W_{t,T}^*)]. \tag{3-6}$$

As an induction assumption, we assume $H_\tau^{*(t)} = H_\tau^{*(t)}(\Omega_{\tau-1}, H_1, Z_1, \ldots, H_t, Z_t)$ for all $\tau > t$. This assumption says that the solution for H_τ^*, as known at time t, depends on all of the variables in $\Omega_{\tau-1}$, some of which are realized and some of which are still random variables, but only on the past choices for $H_1, Z_1, \ldots, H_t, Z_t$ that were made up to time $t-1$. They are not dependent on $H_{t+1}, Z_{t+1}, \ldots, H_\tau, Z_\tau$. This is true for $t = T-1$, since there are no times $\tau > t$, which starts the induction. Then $W_{t,T}^*$ is a function of Ω_T and $H_1, Z_1, \ldots, H_t, Z_t$, and the expectations operator in (3-6) removes all of the variables in $\omega_t \cup \ldots \cup \omega_T$. This means the solution to the period t optimization problem is a function $H_t^* = H_t^*(\Omega_{t-1}, H_1, Z_1, \ldots, H_{t-1}, Z_{t-1})$ and a similar equation for Z_t^*. We then use this to replace each occurrence of H_t and Z_t in $H_\tau^{*(t)}$ with H_t^* and Z_t^*, respectively:

$$\begin{aligned} H_\tau^{*(t-1)} &= H_\tau^{*(t)}(\Omega_{\tau-1}, H_1, Z_1, \ldots, H_{t-1}, Z_{t-1}, H_t^*, Z_t^*) \\ &= H_\tau^{*(t-1)}(\Omega_{\tau-1}, H_1, Z_1, \ldots, H_{t-1}, Z_{t-1}), \end{aligned} \tag{3-7}$$

and a similar equation for $Z_\tau^{*(t)}$. The function $H_\tau^{*(t-1)}$ is obtained from $H_\tau^{*(t)}$ by replacing H_t and Z_t by H_t^* and Z_t^*. By induction, we now conclude that for all $1 \leq t < \tau \leq T$, it is true that $H_\tau^{*(t)} = H_\tau^{*(t)}(\Omega_{\tau-1}, H_1, Z_1, \ldots, H_t, Z_t)$. The set of random variables $\Omega_{\tau-1}$ is determined by the later time τ and the set of decision variables $H_1, Z_1, \ldots, H_{t-1}, Z_{t-1}$ is determined by the earlier time t.

This has been done for any form of the firm's utility function, and this result will be true, with a similar proof, for any optimization problem which consists of a similar nested series of optimizations, solvable by backward induction.

We now reinstate the assumption that the utility function is quadratic, and we want to show that in this case $H^*_\tau = H^*_\tau(\Omega_{\tau-1})$ and $Z^*_\tau = Z^*_\tau(\Omega_{\tau-1})$ also have no dependence on the past hedging choices. For this part we leave the perspective time t implicit and simply write H^*_τ and Z^*_τ.

The first-order conditions for the optimization problem are

$$\begin{cases} E_{t..T}\left[U'(W^*_{t,T})\dfrac{\partial W^*_{t,T}}{\partial H_t}\right] = 0 \\ E_{t..T}\left[U'(W^*_{t,T})\dfrac{\partial W^*_{t,T}}{\partial Z_t}\right] = 0. \end{cases} \tag{3-8}$$

We now prove that the first-order conditions (3-8) can be reduced to the following. This is highly dependent on the assumption of quadratic utility.

$$\begin{cases} \mathrm{Var}_t[S_{1,t}] \cdot H_t + \mathrm{Cov}_t[S_{1,t}, m_t] \cdot Z_t = \sum_{\tau=t}^{T} \mathrm{Cov}_{t..\tau}[S_{1,\tau}S_{2,\tau}, S_{1,t}] \cdot X_\tau \\ \mathrm{Cov}_t[S_{1,t}, m_t] \cdot H_t + \mathrm{Var}_t[m_t] \cdot Z_t = \sum_{\tau=t}^{T} \mathrm{Cov}_{t..\tau}[S_{1,\tau}S_{2,\tau}, m_t] \cdot X_\tau. \end{cases} \tag{3-9}$$

Since there is no appearance of $H_1, Z_1 \ldots, H_{t-1}, Z_{t-1}$ in these equations, the solutions H^*_t and Z^*_t are independent of them. The variance and covariance quantities on the left are dependent only on Ω_{t-1}, and the operators $\mathrm{Cov}_{t..\tau}[\ \ ]$ on the right remove all of the variables in $\omega_t \cup \ldots \cup \omega_T$, so the lefthand side is also dependent only on Ω_{t-1}, and this will justify writing $H^*_t = H^*_t(\Omega_{t-1})$ and $Z^*_t = Z^*_t(\Omega_{t-1})$.

Proof of Reduction of the First-Order Conditions

Suppose this is true for all $\tau > t$. For the case $t = T$, this supposition is vacuously true and the argument that follows will also serve to begin the induction.

By the induction assumption,

$$\frac{\partial \Pi_\tau^*}{\partial H_t} = \frac{\partial \Pi_\tau^*}{\partial Z_t} = 0 \tag{3-10}$$

for all $\tau > t$, and the values of $H_1, Z_1, \ldots, H_{t-1}, Z_{t-1}$ are known and constant, so we also conclude that

$$\frac{\partial \Pi_\tau}{\partial H_t} = \frac{\partial \Pi_\tau}{\partial Z_t} = 0 \tag{3-11}$$

for all $\tau < t$. From these and the definitions (3-4) and (3-5), we conclude that $\partial W_{t,T}^*/\partial H_t = (\bar{S}_{1,t} - S_{1,t})$ and $\partial W_{t,T}^*/\partial Z_t = (P_t - m_t)$. This allows us to simplify (3-8) to

$$\begin{cases} E_{t..T}\left[U'(W_{t,T}^*)(\bar{S}_{1,t} - S_{1,t})\right] = 0 \\ \\ E_{t..T}\left[U'(W_{t,T}^*)(P_t - m_t)\right] = 0. \end{cases} \tag{3-12}$$

We remark here that, because of the induction assumption, we can now conclude, in exactly the same way as in the one period case, that the Hessian is negative definite, so the solution, if it exists, is unique and is a global maximum.

Since we are assuming quadratic utility, we can use the fact that $U'(W)$ is linear and that $E_{t..T}[\bar{S}_{1,t} - S_{1,t}] = E_{t..T}[P_t - m_t] = 0$ to rewrite (3-12) as

$$\begin{cases} 0 = E_{t..T}\left[W_{t,T}^* \cdot \left(\bar{S}_{1,t} - S_{1,t}\right)\right] \\ \\ 0 = E_{t..T}\left[W_{t,T}^* \cdot (P_t - m_t)\right]. \end{cases} \tag{3-13}$$

We can use (3-5) to expand the quantity $W^*_{t,T}$ and thereby obtain

$$\begin{cases} 0 = \sum_{\tau=1}^{t-1} E_{t..T}\left[\Pi_\tau\left(\bar{S}_{1,t} - S_{1,t}\right)\right] + E_{t..T}\left[\Pi_t\left(\bar{S}_{1,t} - S_{1,t}\right)\right] \\ \qquad\qquad + \sum_{\tau=t+1}^{T} E_{t..T}\left[\Pi^*_\tau\left(\bar{S}_{1,t} - S_{1,t}\right)\right] \\ 0 = \sum_{\tau=1}^{t-1} E_{t..T}\left[\Pi_\tau\left(P_t - m_t\right)\right] + E_{t..T}\left[\Pi_t\left(P_t - m_t\right)\right] \\ \qquad\qquad + \sum_{\tau=t+1}^{T} E_{t..T}\left[\Pi^*_\tau\left(P_t - m_t\right)\right]. \end{cases} \tag{3-14}$$

The first summation in each of these equations is zero, since at time $t-1$ each Π_τ for $\tau < t$ is constant and factors out of the expectation. Removing this summation and expanding the Π^*_τ in the second one by its definition (3-4), the first equation becomes

$$\begin{aligned} 0 = E_{t..T}\Big[\Pi_t\left(\bar{S}_{1,t} - S_{1,t}\right)\Big] &+ \sum_{\tau=t+1}^{T} E_{t..T}\Big[S_{1,\tau}S_{2,\tau}X_\tau \cdot \left(\bar{S}_{1,t} - S_{1,t}\right)\Big] \\ &+ \sum_{\tau=t+1}^{T} E_{t..T}\Big[(\bar{S}_{1,\tau} - S_{1,\tau})H^*_\tau\left(\bar{S}_{1,t} - S_{1,t}\right)\Big] \\ &+ \sum_{\tau=t+1}^{T} E_{t..T}\Big[(P_\tau - m_\tau)Z^*_\tau\left(\bar{S}_{1,t} - S_{1,t}\right)\Big], \end{aligned} \tag{3-15}$$

and similarly the second equation becomes

$$\begin{aligned} 0 = E_{t..T}\Big[\Pi_t\left(P_t - m_t\right)\Big] &+ \sum_{\tau=t+1}^{T} E_{t..T}\Big[S_{1,\tau}S_{2,\tau}X_\tau \cdot \left(P_t - m_t\right)\Big] \\ &+ \sum_{\tau=t+1}^{T} E_{t..T}\Big[(\bar{S}_{1,\tau} - S_{1,\tau})H^*_\tau\left(P_t - m_t\right)\Big] \\ &+ \sum_{\tau=t+1}^{T} E_{t..T}\Big[(P_\tau - m_\tau)Z^*_\tau\left(P_t - m_t\right)\Big]. \end{aligned} \tag{3-16}$$

We can evaluate the expectations in two stages, as $E_{t..T}[\quad] = E_{t..\tau-1}\big[E_{\tau..T}[\quad]\big]$. When we do this in the second summation of (3-15), the one involving H^*_τ, then because H^*_τ does not depend on any variables in $\omega_\tau \cup \ldots \cup \omega_T$, as was proved above, it can be factored out of the inner expectation. Also, $(\bar{S}_{1,t} - S_{1,t})$ can be factored out of the inner expectation. Thus

$$\begin{aligned} E_{t..T}\Big[(\bar{S}_{1,\tau} - S_{1,\tau})H^*_\tau(\bar{S}_{1,t} - S_{1,t})\Big] \\ &= E_{t..\tau-1}\Big[E_{\tau..T}\Big[(\bar{S}_{1,\tau} - S_{1,\tau})H^*_\tau(\bar{S}_{1,t} - S_{1,t})\Big]\Big] \\ &= E_{t..\tau-1}\Big[H^*_\tau(\bar{S}_{1,t} - S_{1,t}) \cdot E_{\tau..T}\Big[(\bar{S}_{1,\tau} - S_{1,\tau})\Big]\Big] \\ &= 0. \end{aligned} \tag{3-17}$$

A similar argument also shows that

$$\begin{cases} E_{t..T}\Big[(P_\tau - m_\tau)Z^*_\tau(\bar{S}_{1,t} - S_{1,t})\Big] &= 0 \\ E_{t.\ T}\Big[(\bar{S}_{1,\tau} - S_{1,\tau})H^*_\tau(P_t - m_t)\Big] &= 0 \\ E_{t..T}\Big[(P_\tau - m_\tau)Z^*_\tau(P_t - m_t)\Big] &= 0. \end{cases} \tag{3-18}$$

This transforms (3-14) into

$$\begin{cases} 0 = E_{t..T}\left[\Pi_t\left(\bar{S}_{1,t} - S_{1,t}\right)\right] + \sum\limits_{\tau=t+1}^{T} E_{t..T}\Big[S_{1,\tau}S_{2,\tau}\left(\bar{S}_{1,t} - S_{1,t}\right)\Big] \cdot X_\tau \\ 0 = E_{t..T}\left[\Pi_t\left(P_t - m_t\right)\right] + \sum\limits_{\tau=t+1}^{T} E_{t..T}\Big[S_{1,\tau}S_{2,\tau}\left(P_t - m_t\right)\Big] \cdot X_\tau\,. \end{cases} \tag{3-19}$$

This completes the induction, since (3-19) does not involve any of H_τ, Z_τ for $\tau \neq t$. □

We do a final rearrangement of the first-order conditions to make them look more like the one period case. We again expand the Π_t and rearrange terms to write (3-19) as (3-9), which we repeat here for convenience:

$$\begin{cases} \mathrm{Var}_t[S_{1,t}] \cdot H_t + \mathrm{Cov}_t[S_{1,t}, m_t] \cdot Z_t = \sum\limits_{\tau=t}^{T} \mathrm{Cov}_{t..\tau}[S_{1,\tau}S_{2,\tau}, S_{1,t}] \cdot X_\tau \\ \mathrm{Cov}_t[S_{1,t}, m_t] \cdot H_t + \mathrm{Var}_t[m_t] \cdot Z_t = \sum\limits_{\tau=t}^{T} \mathrm{Cov}_{t..\tau}[S_{1,\tau}S_{2,\tau}, m_t] \cdot X_\tau. \end{cases} \tag{3-9}$$

We introduce notations for the coefficients, analogous to equations (2-15) of the one period case:

$$
\begin{aligned}
&A_{11,t} = \mathrm{Var}[S_{1,t}], \quad A_{12,t} = A_{21,t} = \mathrm{Cov}[S_{1,t}, m_t], \quad A_{22,t} = \mathrm{Var}[m_t], \\
&B_{1,t,\tau} = \mathrm{Cov}_{t..\tau}[S_{1,\tau}S_{2,\tau}, S_{1,t}], \quad B_{2,t,\tau} = \mathrm{Cov}_{t..\tau}[S_{1,\tau}S_{2,\tau}, m_t],
\end{aligned}
\tag{3-20}
$$

so that

$$
\begin{cases}
A_{11,t}H_t + A_{12,t}Z_t = \displaystyle\sum_{\tau=t}^{T} B_{1,t,\tau}X_\tau \\
A_{21,t}H_t + A_{22,t}Z_t = \displaystyle\sum_{\tau=t}^{T} B_{2,t,\tau}X_\tau.
\end{cases}
\tag{3-21}
$$

The equations (3-21) are clearly linear in $Y = (X_1, \ldots, X_T)$, the sequence of cash flows. Consequently, the solutions will also be linear. This proves the following proposition.

Proposition 2. *Under quadratic utility, but with no distributional assumptions about the exchange rate stochastic process, the solutions to the hedging problem are linear. That is, suppose $Y = (X_1, \ldots, X_T)$, $Y' = (X_1', \ldots, X_T')$, and $Y'' = (X_1'', \ldots, X_T'')$ are any three cash flows, and suppose their optimal hedges are $\mathcal{H} = (H_1, Z_1, \ldots, H_T, Z_T)$, $\mathcal{H}' = (H_1', Z_1', \ldots, H_T', Z_T')$, and $\mathcal{H}'' = (H_1'', Z_1'', \ldots, H_T'', Z_T'')$, respectively. Then $Y = \alpha Y' + \beta Y''$ implies $\mathcal{H} = \alpha\mathcal{H}' + \beta\mathcal{H}''$ for any real numbers α and β.*

This will allow us to build a solution from simpler cases. The simplest hedging problem in the multi-period case is when there is a single cash flow in period h in the amount of one unit of currency. This cash flow is $X_\tau = \delta_{\tau,h}$, which is 1 when $\tau = h$ and is 0 when $\tau \neq h$. We represent the optimal hedge for this cash flow as $(\eta_{t,h}, \zeta_{t,h})$. In this case, for $t > h$, the righthand side of both equations in (3-21) becomes 0, so the solution is $(\eta_{t,h}, \zeta_{t,h}) = (0, 0)$. This makes sense. If there are no more cash flows due after time h, then there is no point in doing any further hedging. When $t \leq h$, it is the solution of the following simplified version of (3-21):

$$\begin{cases} A_{11,t}H_t + A_{12,t}Z_t = B_{1,t,h}, \\ A_{21,t}H_t + A_{22,t}Z_t = B_{2,t,h}. \end{cases} \tag{3-22}$$

Analogous to equation (2-17), the solution by Cramer's rule is

$$\begin{cases} \eta_{t,h} = (B_{1,t,h}A_{22,t} - B_{2,t,h}A_{12,t})/\Delta_t, \\ \zeta_{t,h} = (A_{11,t}B_{2,t,h} - A_{21,t}B_{1,t,h})/\Delta_t, \end{cases} \tag{3-23}$$

where

$$\Delta_t = A_{11,t}A_{22,t} - A_{21,t}A_{12,t} = \mathrm{Var}[S_{1,t}]\mathrm{Var}[m_t] - \mathrm{Cov}[S_{1,t}, m_t]^2. \tag{3-24}$$

These equations are also valid when $t > h$, for $S_{1h}S_{2h}$ is deterministic at time $t > h$, so $B_{1,t,h} = B_{2,t,h} = 0$, and we again have $\eta_{t,h} = \zeta_{t,h} = 0$. Putting in the explicit definitions of $A_{ij,t}$ and $B_{i,t,h}$, these are

$$\begin{cases} \eta_{t,h} = (\mathrm{Cov}[S_{1h}S_{2h}, S_{1t}]\mathrm{Var}[m_t] - \mathrm{Cov}[S_{1h}S_{2h}, m_t]\mathrm{Cov}[S_{1t}, m_t])/\Delta_t, \\ \zeta_{t,h} = (\mathrm{Cov}[S_{1h}S_{2h}, m_t]\mathrm{Var}[S_{1t}] - \mathrm{Cov}[S_{1h}S_{2h}, S_{1t}]\mathrm{Cov}[S_{1t}, m_t])/\Delta_t. \end{cases} \tag{3-25}$$

It's interesting to look at this solution in terms of regression coefficients. If X and Y are two random variables, then the regression of Y on X is the equation

$$Y - E[Y] = \beta(X - E[X]) + \epsilon, \tag{3-26}$$

where $E[\epsilon] = 0 = \mathrm{Cov}[X, \epsilon]$ and

$$\beta = \beta[Y, X] = \mathrm{Cov}[Y, X]/\mathrm{Var}[X]. \tag{3-27}$$

As in (3-18), we can write

$$\Delta_t = \mathrm{Var}[S_{1t}]\mathrm{Var}[m_t] \cdot (1 - R_t^2), \tag{3-28}$$

where R is the correlation coefficient between S_{1t} and m_t:

$$R_t^2 = \frac{\mathrm{Cov}[S_{1t}, m_t]^2}{\mathrm{Var}[S_{1t}]\mathrm{Var}[m_t]} = \beta[S_{1t}, m_t]\beta[m_t, S_{1t}]. \tag{3-29}$$

Putting these together, the equations for $\eta_{t,h}$ and $\zeta_{t,h}$ when $t \leq h$ become

$$\begin{cases} \eta_{t,h} = (\beta[S_{1h}S_{2h}, S_{1t}] - \beta[m_t, S_{1t}]\beta[S_{1h}S_{2h}, m_t])/(1 - R_t^2), \\ \zeta_{t,h} = (\beta[S_{1h}S_{2h}, m_t] - \beta[S_{1t}, m_t]\beta[S_{1h}S_{2h}, S_{1t}]/(1 - R_t^2). \end{cases} \tag{3-30}$$

We now express the solution to the hedging problem in the general case in terms of this simpler solution. Let $Y_h = (0, \ldots, 0, 1, 0, \ldots, 0)$ for each $h = 1, \ldots, T$ be the series of cash flows which is 1 in period h but is 0 in each period $\tau \neq h$, and let the optimal hedge sequence for Y_h be $\mathcal{H}_h^* = (\eta_{1,h}, \zeta_{1,h}, \ldots, \eta_{T,h}, \zeta_{T,h})$. If $Y = (X_1, \ldots, X_T)$ is any series of cash flows, then we have $Y = \sum_{h=1}^T X_h \cdot Y_h$, and the linearity proven in proposition 2 shows that the optimal hedge sequence for Y is $\mathcal{H}^* = \sum_{h=1}^T X_h \cdot \mathcal{H}_h^*$. This proves the following proposition.

Proposition 3. *Under the assumption of quadratic utility, the optimal hedge for the cash flow* $Y = (X_1, \ldots, X_T)$ *is* $\mathcal{H}^* = (H_1^*, Z_1^*, \ldots, H_T^*, Z_T^*)$*, where*

$$\begin{cases} H_t^* = \sum_{h=t}^{T} \eta_{t,h} \cdot X_h \\ Z_t^* = \sum_{h=t}^{T} \zeta_{t,h} \cdot X_h. \end{cases} \tag{3-31}$$

The optimal hedge is obtained by hedging each future cash flow separately, without any interaction and without any effects over time.

In the case treated by Chang and Wong, $B_{i,t,h}$ was independent of h, and therefore so were $\eta_{t,h}$ and $\zeta_{t,h}$, and in this case the solution simplifies further to

$$\begin{cases} H_t^* = \eta_t \cdot \sum_{h=t}^{T} X_h \\ Z_t^* = \zeta_t \cdot \sum_{h=t}^{T} X_h. \end{cases} \tag{3-32}$$

In this case, one simply adds up all future cash flows and then hedges them as though it were a one period hedge.

We have now developed the model for both the one period case and the multi-period case, as far as we can without having details of the underlying distributions. We now turn to consideration of the behavior of the model when the distributions of the exchange rates are bivariate lognormal distributions. In Chapter 4 we consider the one period model, and in Chapter 5 we look at the dynamic model.

CHAPTER 4

BIVARIATE LOGNORMAL EXCHANGE RATES: THE ONE PERIOD MODEL

In this chapter we develop the one period model under the assumption that the distributions of the exchange rates are bivariate lognormal distributions.

Before doing so, we want to discuss what models of currency exchange rates are considered to be reasonably realistic. For a discussion about the distributions for price series one can see Taylor (1986), in particular pages 12 and 14. It is the changes in the prices that we are interested in, rather than the absolute level of the price series, so we are concerned with the differences between periods. Next we observe that the behavior of the series should not depend on the unit of measurement chosen. For this reason it's best to use either the differenced logarithms $\log S_{i,t} - \log S_{i,t-1}$ or the relative change $(S_{i,t} - S_{i,t-1})/S_{i,t-1}$, rather than the differences $S_{i,t} - S_{i,t-1}$ of the series itself. Either of these will have the desired effect.

The next point of consideration is some stylized facts.

1) $\{x_t\}$ is (almost) uncorrelated.

2) $\{|x_t|\}$ is correlated.

3) Unusual events occur far more often than the normal distribution allows.

4) The distributions are skewed.

These are observations which seem to be approximately true for a broad range of market price series $\{x_t\}$.

Items 1) and 2) might seem, at first glance, to be contradictory. However, a simple example clarifies this. Let ϵ_t for $t = 1, \ldots, 2T$ be independent random variables with values ± 1, each occurring with probability $1/2$. Let $x_t = \epsilon_t$ for $t = 1, \ldots, T$ and $x_t = 2\epsilon_t$ for $t = T+1, \ldots, 2T$. Then $\{x_t\}$ is uncorrelated, but $\{|x_t|\}$ has values 1 for $t = 1, \ldots, T$ and values 2 for $t = T+1, \ldots, 2T$, which is definitely correlated. To make this kind of example work for a less trivial situation, we can construct a series $x_t = U_t \cdot V_t$, where U_t is a series of independent, but identical, distributions, and V_t is a series of variances that change over time, perhaps randomly. Taylor (1986) has further discussion of such series.

Item 3) is often resolved by using a distribution which has "fatter" tails than the normal distribution does, such as the Student t-distribution with 7 degrees of freedom. In conjunction with this, item 4) can be accommodated by squeezing the left side of the distribution and stretching the right side, thereby making it skewed to the right. A model which incorporates all of these variations is developed in Giot and Laurent (2003), and they seem to have considerable success in using their model for assessing value-at-risk.

The theoretical model developed in the previous chapters can accommodate all of these variations, but implementing them poses considerable technical difficulty. For this reason we shall address only the simplest variation.

In CW the simplest possible model was used. Their distributional assumption was that the differences of the direct exchange rate series $\Delta S_{1,t} = S_{1,t} - S_{1,t-1}$ and

$\Delta S_{2,t} = S_{2,t} - S_{2,t-1}$ have a bivariate normal distribution.

The simplest assumption that goes beyond this is that S_1 and S_2 have a bivariate lognormal distribution; i.e., their logarithms have a bivariate normal distribution:

$$(\log S_1, \log S_2) \sim N_2[\mu_1, \mu_2, \nu_1^2, \nu_2^2, \rho], \tag{4-1}$$

and for this section we assume S_1 and S_2 are so distributed. It is often easier to use standardized forms $z = (\log S_1 - \mu_1)/\nu_1$ and $w = (\log S_2 - \mu_2)/\nu_2$, which then satisfy

$$\begin{cases} \log S_1 = \mu_1 + \nu_1 z \\ \log S_2 = \mu_2 + \nu_2 w, \end{cases} \tag{4-2}$$

and whose joint distribution $(z, w) \sim N_2[0, 0, 1, 1, \rho]$ has the joint p.d.f.

$$f(z, w) = \tfrac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2}[z^2+w^2-2\rho z w]/(1-\rho^2)}. \tag{4-3}$$

We let $S_{i,0}$ be the value of the exchange rate at the beginning of the period. One does not usually expect exchange rates to have any upward or downward bias, barring specific information about the economies involved, so we want $E[S_i] = S_{i,0}$. By Jensen's inequality, we will then have to have $E[\log(S_i)] < \log(S_{i,0})$. By (4-4), below, we see that we can get $E[S_i] = S_{i,0}$ by taking $\mu_i = \log(S_{i,0}) - \nu_i^2/2$, and then we have $E[\log(S_i)] = \log(S_{i,0}) - \nu_i^2/2$. The assumption that the expected value of the exchange rate at the end of the period is the same as the exchange rate at the beginning of the period forces the logarithms to have a downward drift.

To evaluate H^* and Z^* from equation (2-17), we need to evaluate the variances and covariances involved in the A_{ij} and B_i, and these are obtained from moments of the form $E[S_1^k]$, $E[S_1^k S_2]$, $E[S_1^k m]$, and $E[S_1^k S_2 m]$. We state the relevant moment equations in (4-4), (4-5), and (4-6) and then prove them.

For convenience we set $\phi_i = e^{\nu_i^2}$, $\lambda = e^{\rho\nu_1\nu_2}$, and $\psi_k = e^{(k^2-k)/2\cdot\nu_1^2+k\rho\nu_1\nu_2}$. The moments without m are

$$\bar{S}_i = e^{\mu_i+\nu_i^2/2} = S_{i,0}$$

$$E[S_i^k] = \bar{S}_i^k \phi_i^{(k^2-k)/2} \tag{4-4}$$

$$E[S_1^k S_2] = \bar{S}_1^k \bar{S}_2\, \phi_1^{(k^2-k)/2} \lambda^k.$$

Let $\Phi(x)$ denote the cumulative density function of the normal distribution, let $c = -k\nu_1 - n\rho\nu_2$, and define the auxiliary functions

$$\Phi_{k,n} = \Phi_{k,n}(\nu_1, \nu_2, \rho) = e^{-k\nu_1^2 - kn\rho\nu_1\nu_2}\Phi(c + \tfrac{\nu_1}{2}) - \Phi(c - \tfrac{\nu_1}{2}). \tag{4-5}$$

Then the moments including m are

$$E[S_1^k S_2^n m] = E[S_1^{k+1} S_2^n]\, \Phi_{k,n}. \tag{4-6}$$

Although this is true for all n, it will only be used for $n = 0$ and $n = 1$.

Derivation of Lognormal Moments

First we note the following standard result, when z and w have a bivariate normal distribution, $(\log z, \log w) \sim N_2(0, 0, 1, 1, \rho)$:

$$E[e^{az+bw}] = e^{\frac{1}{2}(a^2+b^2)+\rho ab}. \tag{4-7}$$

This appears in many standard sources, including Johnson and Kotz (1972). We give a brief proof.

Proof of (4-7). We complete the square in the following by choosing $A = a + \rho b$, $B = b + a\rho$, and $C = A^2 + B^2 - 2\rho AB = (1-\rho^2)(a^2 + b^2 + 2\rho ab)$.

$$\begin{aligned} E[e^{az+bw}] &= \frac{1}{2\pi(1-\rho^2)} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{az+bw-\frac{1}{2}[z^2+w^2-2\rho zw]/(1-\rho^2)}\, dw\, dz \\ &= \frac{1}{2\pi(1-\rho^2)} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}[(z-A)^2+(w-B)^2-2\rho(z-A)(w-B)-C]/(1-\rho^2)}\, dw\, dz, \\ &= e^{\frac{1}{2}\frac{C}{(1-\rho^2)}} = e^{\frac{1}{2}(a^2+b^2)+\rho ab}. \end{aligned}$$ □

From this, since $S_1 = \mu_1 + \nu_1 z$ and $S_2 = \mu_2 + \nu_2 w$, we have

$$\begin{aligned} E[S_1^k S_2^n] &= e^{k\mu_1+n\mu_2} E[e^{k\nu_1 z+n\nu_2 w}], \\ &= e^{k\mu_1+n\mu_2}\left(e^{\frac{1}{2}[(k\nu_1)^2+(n\nu_2)^2]+\rho(k\nu_1)(n\nu_2)}\right) \\ &= e^{k(\mu_1+\nu_1^2/2)} e^{n(\mu_2+\nu_2^2/2)} e^{\left(\frac{1}{2}(k^2-k)\nu_1^2+\frac{1}{2}(n^2-n)\nu_2^2+kn\rho\nu_1\nu_2\right)} \\ &= \bar{S}_1^k \bar{S}_2^n \phi_1^{\frac{k^2-k}{2}} \phi_2^{\frac{n^2-n}{2}} \lambda^{kn}, \end{aligned} \tag{4-8}$$

and this includes equations (4-4).

Now we want to consider the moments involving m. In the following we use the notation $(z \le c)$ in an expression $E[F(z,w)(z \le c)]$ to denote the indicator function

$$I_c(z) = \begin{cases} 0, & \text{if } z > c \\ 1, & \text{if } z \le c. \end{cases} \tag{4-9}$$

Then since $\bar{S}_1 \ge S_1$ if and only if $z \le \nu_1/2$, we can write

$$\begin{aligned} E[S_1^k S_2^n m] &= E[S_1^k S_2^n (\bar{S}_1 - S_1)(z \le \tfrac{\nu_1}{2})] \\ &= \bar{S}_1 E[S_1^k S_2^n (z \le \tfrac{\nu_1}{2})] - E[S_1^{k+1} S_2^n (z \le \tfrac{\nu_1}{2})]. \end{aligned} \tag{4-10}$$

As in the proof of (4-4), we first prove a lemma, but this one is less standard.

Lemma. *For any real numbers a, b, and c*

$$E[e^{az+bw}(z \le c)] = E[e^{az+bw}]\Phi(c - a - b\rho), \tag{4-11}$$

where $\Phi(c) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c} e^{-z^2/2} dz$ *is the cumulative density function of the unit normal distribution.*

Proof. With $B = \rho z + b(1-\rho^2)$, we have

$$\begin{aligned}
E[e^{az+bw}(z \le c)] &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{c} \int_{-\infty}^{\infty} e^{az+bw-\frac{1}{2}[z^2+w^2-2\rho zw]/(1-\rho^2)}\, dw\, dz \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{c} e^{az-\frac{1}{2}\frac{z^2}{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}[w^2-2Bw]/(1-\rho^2)}\, dw\, dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c} e^{az-\frac{1}{2}\frac{z^2}{1-\rho^2}} \left\{ \frac{1}{\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}[(w-B)^2-B^2]/(1-\rho^2)}\, dw \right\} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c} e^{az-\frac{1}{2}\frac{z^2}{1-\rho^2}} \left\{ e^{\frac{1}{2}B^2/(1-\rho^2)} \right\} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c} \exp\left\{ -\frac{1}{2}\left[\frac{z^2-B^2}{1-\rho^2} - 2az \right] \right\} dz, \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c} \exp\left\{ -\frac{1}{2}(z-(a+b\rho))^2 + \frac{1}{2}(a^2+b^2) + \rho ab \right\} dz \\
&= e^{\frac{1}{2}(a^2+b^2)+\rho ab}\, \Phi(c-a-b\rho) \\
&= E[e^{az+bw}]\, \Phi(c-a-b\rho).
\end{aligned}$$

□

From (4-4) we see that $E[S^{(k+1)}S_2^n] = \bar{S}_1 E[S_1^k S_2^n]\phi_1^k \lambda^n$, and so

$$\begin{aligned}
E[S_1^k S_2^n m] &= E[S_1^k S_2^n \bar{S}_1 (z \le \tfrac{\nu_1}{2})] - E[S_1^{(k+1)} S_2^n (z \le \tfrac{\nu_1}{2})] \\
&= \bar{S}_1 E[S_1^k S_2^n]\Phi(\tfrac{\nu_1}{2} - k\nu_1 - n\rho\nu_2) - E[S_1^{(k+1)} S_2^n]\Phi(\tfrac{\nu_1}{2} - (k+1)\nu_1 - n\rho\nu_2) \\
&= E[S_1^{(k+1)} S_2^n]\left[\phi_1^{-k}\lambda^{-n}\Phi(\tfrac{\nu_1}{2} - k\nu_1 - n\rho\nu_2) - \Phi(\tfrac{\nu_1}{2} - (k+1)\nu_1 - n\rho\nu_2)\right] \qquad (4\text{-}12) \\
&= E[S_1^{(k+1)} S_2^n]\left[e^{-k\nu_1^2 - n\rho\nu_1\nu_2}\Phi(\tfrac{\nu_1}{2} - k\nu_1 - n\rho\nu_2) - \Phi(\tfrac{\nu_1}{2} - (k+1)\nu_1 - n\rho\nu_2)\right].
\end{aligned}$$

The quantity in brackets is $\Phi_{k,n}$, so (4-6) is proven.

For each k and n, we have $E[S_1^k S_2^n] = \bar{S}_1^k \bar{S}_2^n \cdot \varphi(\nu_1, \nu_2, \rho)$ for some function φ. Two things should be noted about these moments. First μ_1 and μ_2 occur only in the powers of $\bar{S}_1$ and $\bar{S}_2$. The second is a variety of homogeneity in that the powers of $\bar{S}_1$ and $\bar{S}_2$ match those of S_1 and S_2.

We now use these to evaluate the coefficients of the first-order equations (2-14) for H^* and Z^*:

$$\begin{aligned}
A_{11} &= \mathrm{Var}[S_1] &&= E[S_1^2] - \bar{S}_1^2 &&= \bar{S}_1^2 \cdot [\phi_1 - 1],\\
A_{12} &= \mathrm{Cov}[S_1, m] &&= E[S_1 m] - E[S_1]E[m] &&= \bar{S}_1^2 \cdot [\phi_1 \Phi_{1,0} - \Phi_{2,0}],\\
A_{22} &= \mathrm{Var}[m] &&= E[(\bar{S}_1 - S_1)m] - E[m]^2 &&= \bar{S}_1^2 \cdot \left[\Phi_{0,0} - \phi_1 \Phi_{1,0} - \Phi_{0,0}^2\right],\\
B_1 &= -E[S_1 S_2 (\bar{S}_1 - S_1)] && &&= \bar{S}_1^2 \bar{S}_2 \cdot \left[\phi_1 \lambda^2 \qquad - \lambda\right], \quad \text{and}\\
B_2 &= -E[S_1 S_2 (P - m)] && &&= \bar{S}_1^2 \bar{S}_2 \cdot \left[\phi_1 \lambda^2 \Phi_{1,1} - \lambda \Phi_{0,0}\right].
\end{aligned} \tag{4-13}$$

We define quantities a_{ij} and b_i by $A_{ij} = \bar{S}_1^2 \cdot a_{ij}$ and $B_i = \bar{S}_1^2 \bar{S}_2 \cdot b_i$ so that these are the expressions in brackets in (4-13):

$$\begin{aligned}
a_{11} &= \phi_1 - 1,\\
a_{12} = a_{21} &= \phi_1 \, \Phi_{1,0} - \Phi_{0,0},\\
a_{22} &= \Phi_{0,0} - \Phi_{0,0}^2 - \phi_1 \, \Phi_{1,0},\\
b_1 &= \phi_1 \lambda^2 \qquad - \lambda, \quad \text{and}\\
b_2 &= \phi_1 \lambda^2 \, \Phi_{1,1} - \lambda \Phi_{0,0}.
\end{aligned} \tag{4-14}$$

Each of a_{ij} and b_i is independent of μ_1 and μ_2. Because every coefficient of the equations (2-14) have the same factor $\bar{S}_1^2$, it can be factored out, and they become

$$\begin{cases} a_{11} H^* + a_{12} Z^* = b_1 \cdot \bar{S}_2 X \\ a_{21} H^* + a_{22} Z^* = b_2 \cdot \bar{S}_2 X. \end{cases} \tag{4-15}$$

These equations are linear in $\bar{S}_2 X$, so the solution of the hedging equations has the form

$$\begin{cases} H^* = \eta \cdot \bar{S}_2 X \\ \\ Z^* = \zeta \cdot \bar{S}_2 X, \end{cases} \tag{4-16}$$

where η and ζ are functions of only ν_1, ν_2, and ρ and are given by Cramer's rule:

$$\begin{cases} \eta = (b_1 a_{22} - b_2 a_{12})/\delta \\ \\ \zeta = (b_2 a_{11} - b_1 a_{21})/\delta, \end{cases} \tag{4-17}$$

where $\delta = a_{11}a_{22} - a_{12}a_{21}$.

This concludes the development of the one period model when the underlying distributions are the bivariate lognormal distributions. In the next chapter we do the same for the dynamic model.

CHAPTER 5

BIVARIATE LOGNORMAL EXCHANGE RATES: THE DYNAMIC MODEL

We now apply the previous section to the dynamic model. The assumption of CW (see the introduction), that the volatilities are nonstochastic, will be maintained here. This is not realistic and is known empirically to be false. There is considerable discussion of this topic in Taylor (1986), and there are many articles in the current literature about the observed patterns of volatility memory in price series. Two such examples are Ding and Granger (1996) and Zumbach (2004). In this chapter we deal with the following model, for which the volatility is known and does not vary across periods. We could, at the cost of more cumbersome notation, equally well handle the case in which each period has its own volatilities $\nu_{i,t}$ and correlation ρ_t, but the results would be essentially the same.

We assume $S_{1,t}$ and $S_{2,t}$, $t = 1, \ldots, T$ are related by

$$\begin{cases} \log S_{1,t} = \log S_{1,t-1} + \nu_1 z_t - \nu_1^2/2 \\ \log S_{2,t} = \log S_{2,t-1} + \nu_2 w_t - \nu_2^2/2, \end{cases} \tag{5-1}$$

where $(z_t, w_t) \sim N_2(0,0,1,1,\rho)$. We assume that each of z_t and w_t is independent of z_τ and w_τ when $\tau \neq t$. We can iterate $\log S_{i,t}$ from its definition to get

$$\begin{cases} \log S_{1,t} = \log S_{1,0} + \nu_1 \left(z_1 + \ldots + z_t\right) - t \cdot \nu_1^2/2 \\ \log S_{2,t} = \log S_{2,0} + \nu_2 \left(w_1 + \ldots + w_t\right) - t \cdot \nu_2^2/2\,. \end{cases} \tag{5-2}$$

The initial values are $\omega_0 = \{S_{1,0}, S_{2,0}\}$, and the random variable sets are $\omega_t = \{z_t, w_t\}$. From this we see that $S_{1,t}$ and $S_{2,t}$ are independent of ω_τ for all $\tau > t$. In addition, note that the variance of $S_{i,t}$, as seen from time $t = 0$, is $t \cdot \nu_i^2$, which grows linearly with time, as does the drift $-t \cdot \nu_i^2/2$.

Most of the work has been done in the previous section, and we only need to make slight alterations appropriate to this process. First note that for one period expectations,

$$E_t[S_{i,t}] = \bar{S}_{i,t} = S_{i,t-1}, \tag{5-3}$$

so that using equations (4-4) and, as before, $\lambda = e^{\rho\nu_1\nu_2}$, we have

$$E_t\Big[S_{1,t}S_{2,t}\Big] = \lambda\, S_{1,t-1}S_{2,t-1}. \tag{5-4}$$

Then by induction, for any $t < h$,

$$E_{t+1..h}\Big[S_{1,h}S_{2,h}\Big] = \lambda^{h-t}\, S_{1,t}S_{2,t}. \tag{5-5}$$

This lets us evaluate

$$\begin{aligned} B_{1,t,h} &= -E_{t..h}\Big[S_{1,h}S_{2,h}(\bar{S}_{1,t} - S_{1,t})\Big] \\ &= -E_t\Big[\, E_{t+1..h}[S_{1,h}S_{2,h}] \cdot (\bar{S}_{1,t} - S_{1,t})\Big] \\ &= -E_t\Big[\lambda^{h-t}\, S_{1,t}S_{2,t}(\bar{S}_{1,t} - S_{1,t})\Big] \end{aligned} \tag{5-6}$$

and a similar result holds for $B_{2,t,h}$. Hence, if we define

$$\begin{cases} B_{1,t} = -E_t[S_{1,t}S_{2,t}(\bar{S}_{1,t} - S_{1,t})] \\ \\ B_{2,t} = -E_t[S_{1,t}S_{2,t}(P_t - m_t)], \end{cases} \tag{5-7}$$

then we have

$$\begin{cases} B_{1,t,h} = \lambda^{h-t}\, B_{1,t} \\ \\ B_{2,t,h} = \lambda^{h-t}\, B_{2,t}. \end{cases} \tag{5-8}$$

This makes the first-order conditions

$$\begin{cases} A_{11,t}H^*_{t,h} + A_{12,t}Z^*_{t,h} = B_{1,t} \cdot \lambda^{h-t}X_h \\ \\ A_{21,t}H^*_{t,h} + A_{22,t}Z^*_{t,h} = B_{2,t} \cdot \lambda^{h-t}X_h. \end{cases} \tag{5-9}$$

In addition, the dependence on t can be factored out, since, with a_{ij} and b_i as in equations (4-15), we have

$$\begin{aligned} A_{ij,t} &= \bar{S}^2_{1,t} \cdot a_{ij}, \quad \text{and} \\ \\ B_{i,t} &= \bar{S}^2_{1,t}\bar{S}_{2,t} \cdot b_i. \end{aligned} \tag{5-10}$$

The factors $\bar{S}^2_{1,t}$ are common to all terms and so cancel out of the equations, and we are left with equations (4-16), but with $\bar{S}_2 X$ replaced by $\bar{S}_{2,t}\lambda^{h-t}X_h$, so we can write the solution in the form of equations (4-18).

$$\begin{cases} H^*_{t,h} = \eta \cdot \lambda^{h-t}\bar{S}_{2,t}X_h \\ \\ Z^*_{t,h} = \zeta \cdot \lambda^{h-t}\bar{S}_{2,t}X_h. \end{cases} \tag{5-11}$$

Because the volatilities and correlation were assumed to be constant, the one period solutions η and ζ have no dependence on t, and they are given by equations (4-18).

Summing these solutions gives the optimal hedge for the entire cash flow series:

$$\begin{cases} H^*_t = \eta \cdot \bar{S}_{2,t} \displaystyle\sum_{h=t}^{T} \lambda^{h-t}X_h \\ Z^*_t = \zeta \cdot \bar{S}_{2,t} \displaystyle\sum_{h=t}^{T} \lambda^{h-t}X_h. \end{cases} \tag{5-12}$$

The direct exchange rate between the domestic currency and the foreign currency is $S = S_1 S_2$ and, although both S_1 and S_2 have their end-of-period expectation equal to their beginning-of-period value, this is not true for S, as is seen in (5-4). It's not very much different, for all of ρ, ν_1, and ν_2 are relatively small numbers, and thus $\lambda = \exp(\rho\nu_1\nu_2)$ is very close to 1, but it is different, nonetheless. Usually ρ is negative, whereas ν_1 and ν_2 are always positive, so S experiences a slight deflation over one period due to the negative correlation of S_1 and S_2. Over $h-t$ periods, the deflation becomes λ^{h-t}, as we see in (5-5), and this is the factor which appears in the hedging formulas (5-12).

This concludes the theoretical discussion of the model, for one period or for many periods, assuming that the exchange rates have bivariate lognormal distributions. In the next three chapters we want to estimate the parameters and assess the performance of the competing models for a concrete case. We will consider the Japan–Taiwan–U.S. currencies over the period 1997–2001. This requires finding the maximum likelihood estimators, which are the solution of a system of equations. In order to solve these we first discuss in Chapter 6 an extension of Newton's method for a system of equations.

CHAPTER 6

NEWTON'S METHOD FOR A SYSTEM OF EQUATIONS

In the next section we will discuss the estimation of parameters for various models. This will sometimes require solving a system of several nonlinear equations in several unknowns. In this section we discuss the application of Newton's method in this setting. I haven't seen this anywhere before, so I don't have a reference, but I presume this is not the first exposition of this method.

Suppose we have a system of n equations in n unknowns:

$$\begin{cases} F_1(x_1, \ldots, x_n) = 0 \\ \qquad \vdots \\ F_n(x_1, \ldots, x_n) = 0\,. \end{cases} \tag{6-1}$$

Let's suppose that we can solve the last $n-k$ equations for each of $x_{k+1}, \ldots, x_n$ as a function of $x_1, \ldots, x_k$, i.e., $x_m = x_m(x_1, \ldots, x_k)$, for $m = k+1, \ldots, n$. Substituting these into the k-th equation, we obtain a function $f_k(x_1, \ldots, x_k) = F_k(x_1, \ldots, x_n)$. We want to find the partial derivative of this function with respect to x_k in terms of the partial derivatives of the original system of equations. The answer is

$$\frac{\partial f_k}{\partial x_k} = \frac{D_k}{D_{k+1}}, \tag{6-2}$$

where

$$D_k = \begin{vmatrix} F_{k,k} & \dots & F_{k,n} \\ \vdots & \ddots & \vdots \\ F_{n,k} & \dots & F_{n,n} \end{vmatrix}, \tag{6-3}$$

with each $F_{i,j}$ being the corresponding partial derivative, and where D_{k+1} is the subdeterminant obtained by striking out the first row and column of D_k.

Proof of (6-2). The values of $x_{k+1}, \dots, x_n$ are determined by the last $(n-k)$ equations of the system (6-1):

$$\begin{cases} F_{k+1}(x_1, \dots, x_n) = 0 \\ \vdots \\ F_n(x_1, \dots, x_n) = 0\,. \end{cases} \tag{6-4}$$

Provided each F_m is continuously differentiable, the implicit function theorem guarantees that each of $x_{k+1}, \dots, x_n$ is determined as a continuously differentiable function of $x_1, \dots, x_k$. Then we can implicitly differentiate each of these with respect to x_k to get a system of equations:

$$\begin{cases} F_{k+1,k} + F_{k+1,k+1}\dfrac{\partial x_{k+1}}{\partial x_k} + \dots + F_{k+1,n}\dfrac{\partial x_n}{\partial x_k} = 0 \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ F_{n,k} + F_{n,k+1}\dfrac{\partial x_{k+1}}{\partial x_k} + \dots + F_{n,n}\dfrac{\partial x_n}{\partial x_k} = 0\,. \end{cases} \tag{6-5}$$

This can be solved by Cramer's rule, and the answer is

$$\frac{\partial x_m}{\partial x_k} = \frac{D_{k+1,m}}{D_{k+1}}, \tag{6-6}$$

where $D_{k+1,m}$ is obtained by replacing the column $(F_{k+1,m}, \dots, F_{n,m})'$ of the determinant D_{k+1} with the column $(-F_{k+1,k}, \dots, -F_{n,k})'$. The minus sign can be factored out of the determinant and the column $(F_{k+1,k}, \ldots, F_{n,k})'$ moved to the front, resulting in

$$D_{k+1,m} = (-1)^{k+m} M_{k,m}, \tag{6-7}$$

where $M_{k,m}$ is the subdeterminant obtained from D_k by striking out its first row and the column $(F_{k,m}, \ldots, F_{n,m})'$. Note that $D_{k+1} = M_{k,k}$. We can now differentiate:

$$\begin{aligned}
\frac{\partial f_k}{\partial x_k} &= \frac{\partial}{\partial x_k} F_k(x_1, \ldots, x_n) \\
&= F_{k,k} + \sum_{m=k+1}^{n} F_{k,m} \frac{\partial x_m}{\partial x_k} \\
&= F_{k,k} + \sum_{m=k+1}^{n} F_{k,m} (-1)^{k+1} \frac{M_{k,m}}{D_{k+1}} \qquad (6\text{-}8) \\
&= \sum_{m=k}^{n} F_{k,m} (-1)^{m+k} M_{k,m} \Big/ D_{k+1} \\
&= \frac{D_k}{D_{k+1}}.
\end{aligned}$$

□

The solution of the entire system (6-1) by Newton's method will now be recursive. Supposing we know how to solve the subsystem (6-4) whenever necessary, we proceed as follows:

Step 0: Suppose we have some initial values $x_1, \ldots, x_k$.

Step 1: Solve the system (6-4) for the values of $x_{k+1}, \ldots, x_n$.

Step 2: Use the partial derivative (6-2) to make an iteration by Newton's method.

$$x_k' = x_k - \frac{F_k(x_1, \ldots, x_n)}{\partial f_k / \partial x_k} :$$

Replacing x_k with x_k' go back to step 0 and repeat until convergence is reached. When

When it converges, we have a set $x_k, \ldots, x_n$ which is the solution of

$$\begin{cases} F_k(x_1, \ldots, x_n) = 0 \\ \quad\quad\vdots \\ F_n(x_1, \ldots, x_n) = 0 \end{cases} \tag{6-9}$$

for the given values of $x_1, \ldots, x_{k-1}$.

This process involves iterations within iterations, and if the number of equations is large it could conceivably be quite time consuming, so it behooves us to consider an efficient choice for a stopping rule. The question is, how many iterations do we do at each level? In the applications done for the following section, I've found that the inner loops can be done with only one iteration, while the outer loop, for $k = 1$, must be done for five iterations to avoid loss of accuracy. Further iterations of the inner loops don't seem to add anything to the accuracy of the final approximation nor to reduce the number of iterations necessary at the outer level. With this stopping rule, we seem to have a fairly efficient technique. I can't say whether this will always be a good rule. It may be that it works for these examples only because of the nature of the equations involved or because the initial approximation is fairly good to start with. If this method is to be used in a different setting, I'd advise testing the convergence properties before trusting it too far.

In this chapter we have developed a method for solving a system of equations in several variables. In the next chapter we apply this method to finding the maximum likelihood estimators for the parameters of distributions.

CHAPTER 7

MAXIMUM LIKELIHOOD ESTIMATORS

We now want to do a numerical example, comparing the hedging results obtained for different models. In implementing these models we need to use estimates for the parameters, and the estimates we use will be the maximum likelihood estimators, adjusted for bias when possible. For the bivariate normal model the estimators will be the familiar ones used in ordinary regression. For the bivariate lognormal model, with the term $-\nu_i^2/2$ added to avoid inflation, we need to use a multidimensional Newton's method to find the estimates.

The Bivariate Normal Model

We start with the simplest model, the bivariate normal model with no inflation, which is specified by

$$\begin{cases} S_{1,t} = S_{1,t-1} + \sigma_1 z_t, \\ \\ S_{2,t} = S_{2,t-1} + \sigma_2 w_t, \end{cases} \tag{7-1}$$

where $(z_t, w_t) \sim N_2(0, 0, 1, 1, \rho)$. This is standard, but we'll have to deal with variations on it, so we treat it in detail here so that we can adapt the method for the other models.

The negative logarithm of the probability density function, in terms of the differences $\Delta S_{i,t} = S_{i,t} - S_{i,t-1}$, is

$$\begin{aligned}-\log f(\Delta S_{1,t}, \Delta S_{2,t}) = \log(2\pi) + \frac{1}{2}\log(1-\rho^2) + \log\sigma_1 + \log\sigma_2 \\ + \frac{1}{2(1-\rho^2)}(z_t^2 + w_t^2 - 2\rho w_t z_t),\end{aligned} \tag{7-2}$$

where $z_t = \Delta S_{1,t}/\sigma_1$ and $w_t = \Delta S_{2,t}/\sigma_2$. Summing over all $t = 1, \ldots, T$ we get the likelihood function to be minimized by the maximum likelihood estimators:

$$\begin{aligned}\Lambda &= -\sum_{t=1}^{T} \log f(\Delta S_{1,t}, \Delta S_{2,t}) \\ &= T\log(2\pi) + \frac{T}{2}\log(1-\rho^2) + T\log\sigma_1 + T\log\sigma_2 \\ &\qquad + \frac{1}{2(1-\rho^2)}\sum_{t=1}^{T}(z_t^2 + w_t^2 - 2\rho z_t w_t).\end{aligned} \tag{7-3}$$

Since this goes to infinity as $\rho \to \pm 1$ and $\sigma_1, \sigma_2 \to \infty$, or $\sigma_1, \sigma_2 \to 0$, there must be a minimum at the solution to the first-order equations. Before taking the derivatives with respect to σ_1 and σ_2, we note that $\partial z_t/\partial\sigma_1 = -z_t/\sigma_1$ and $\partial w_t/\partial\sigma_2 = -w_t/\sigma_2$. Setting the derivatives of Λ with respect to σ_1, σ_2 and ρ equal to 0 we have

$$\begin{cases} 0 = \dfrac{\partial\Lambda}{\partial\sigma_1} = \dfrac{T}{\sigma_1} - \dfrac{1}{\sigma_1(1-\rho^2)}\displaystyle\sum_{t=1}^{T}(z_t^2 - \rho z_t w_t), \\ 0 = \dfrac{\partial\Lambda}{\partial\sigma_2} = \dfrac{T}{\sigma_2} - \dfrac{1}{\sigma_2(1-\rho^2)}\displaystyle\sum_{t=1}^{T}(w_t^2 - \rho z_t w_t), \\ 0 = \dfrac{\partial\Lambda}{\partial\rho} = \dfrac{-T\cdot\rho}{(1-\rho^2)^2} + \dfrac{\rho}{(1-\rho^2)^2}\displaystyle\sum_{t=1}^{T}(z_t^2 + w_t^2) - \dfrac{1+\rho^2}{(1-\rho^2)^2}\displaystyle\sum_{t=1}^{T} z_t w_t. \end{cases} \tag{7-4}$$

We simplify these equations first by writing $J_{ij} = \sum_{t=1}^{T}\Delta S_{i,t}\Delta S_{j,t}$ so that $\sum z_t^2 = J_{11}/\sigma_1^2$, $\sum w_t^2 = J_{22}/\sigma_2^2$, and $\sum z_t w_t = J_{12}/(\sigma_1\sigma_2)$ and then clearing the denominators, which leaves us with

$$\begin{cases} \text{(a)} \quad 0 = T\rho(1-\rho^2)\sigma_1^2\sigma_2^2 - \rho J_{11}\sigma_2^2 + \rho^2 J_{12}\sigma_1\sigma_2, \\ \text{(b)} \quad 0 = T\rho(1-\rho^2)\sigma_1^2\sigma_2^2 - \rho J_{22}\sigma_1^2 + \rho^2 J_{12}\sigma_1\sigma_2, \\ \text{(c)} \quad 0 = T\rho(1-\rho^2)\sigma_1^2\sigma_2^2 - \rho J_{11}\sigma_2^2 - \rho J_{22}\sigma_1^2 + (1+\rho^2)J_{12}\sigma_1\sigma_2. \end{cases} \tag{7-5}$$

Subtracting (a) from (c) and (b) from (c) gives us

$$\begin{cases} \text{(a)} \qquad \rho J_{11}\sigma_2^2 = J_{12}\sigma_1\sigma_2, \\ \\ \text{(b)} \qquad \rho J_{22}\sigma_1^2 = J_{12}\sigma_1\sigma_2. \end{cases} \tag{7-6}$$

Multiplying (7-6-a) and (7-6-b) gives us

$$\hat{\rho} = \frac{J_{12}}{\sqrt{J_{11}J_{22}}}, \tag{7-7}$$

and substituting (7-6-a) into (7-5-a) and (7-6-b) into (7-5-b) gives us

$$\begin{cases} \hat{\sigma}_1 = \sqrt{\dfrac{J_{11}}{T}}, \quad \text{and} \\ \hat{\sigma}_2 = \sqrt{\dfrac{J_{22}}{T}}. \end{cases} \tag{7-8}$$

Since there is no mean in the original model, the estimate for the variance is unbiased, as is seen from

$$E[\hat{\sigma}_i^2] = \frac{1}{T}\sum_{t=1}^{T} E[\Delta S_{i,t}^2] = \sigma_i^2, \tag{7-9}$$

and the small sample bias correction is not needed.

The Bivariate Lognormal Model with No Drift

The simplest form of the bivariate lognormal model does not have the terms $-\nu_i^2/2$ that were added to the model in section 3, and its equations are

$$\begin{cases} \log S_{1,t} &= \log S_{1,t-1} + \nu_1 z_t, \\ \log S_{2,t} &= \log S_{2,t-1} + \nu_2 w_t. \end{cases} \tag{7-10}$$

These satisfy $E[\log S_{i,t}] = \log S_{i,t-1}$, rather than $E[S_{i,t}] = S_{i,t-1}$. The reason for considering this model is to follow the stylized fact that there is no statistically significant difference observed in markets between the average of $\log S_{i,t}$ and $\log S_{i,t-1}$.

The other model was chosen in the development of section 4 because the term $-\nu_i^2/2$ introduced there was too small to be statistically significant, so its presence doesn't really contradict the stylized fact, and theoretically it's more satisfying to assume $E[S_{i,t}] = S_{i,t-1}$.

Formally, with $\Delta \log S_{i,t}$ substituted for $\Delta S_{i,t}$ and ν_i substituted for σ_i, the analysis of the maximum likelihood estimators is identical to that given above for the bivariate normal model.

The Bivariate Lognormal Model with No Inflation

For this model we assume

$$\begin{cases} \log S_{1,t} = \log S_{1,t-1} - \dfrac{1}{2}\nu_1^2 + \nu_1 z_{1,t} \\ \log S_{2,t} = \log S_{2,t-1} - \dfrac{1}{2}\nu_2^2 + \nu_2 z_{2,t}, \end{cases} \tag{7-11}$$

where $(z_{1,t}, z_{2,t}) \sim N_2(0, 0, 1, 1, \rho)$. Then, since in general $E[e^{az_{i,t}+b}] = e^{b+\frac{1}{2}a^2}$, we will have $E[S_{i,t}] = S_{i,t-1}$, i.e., there is no expected inflation.

The maximum likelihood estimators (MLEs) for ν_1, ν_2, and ρ will minimize the function

$$\begin{aligned} \Lambda &= -\sum_{t=1}^{T} \log f(\Delta \log S_{1,t}, \Delta \log S_{2,t}) \\ &= T\log(2\pi) + \frac{T}{2}\log(1-\rho^2) + T\log\nu_1 + T\log\nu_2 \\ &\qquad + \frac{1}{2(1-\rho^2)}\sum_{t=1}^{T}(z_{1,t}^2 + z_{2,t}^2 - 2\rho z_{1,t}z_{2,t}). \end{aligned} \tag{7-12}$$

Each $z_{i,t} = (\Delta \log S_{i,t} + \frac{1}{2}\nu_i^2)/\nu_i$ is determined by ν_i and the data points, and the derivatives satisfy $\partial z_{i,t}/\partial \nu_i = 1 - z_{i,t}/\nu_i$. From this we find the first-order conditions for the minimum to be

$$\begin{cases} 0 = \dfrac{\partial \Lambda}{\partial \nu_1} = \dfrac{T}{\nu_1} + \dfrac{1}{2(1-\rho^2)} \displaystyle\sum_{t=1}^{T} \left(2z_{1,t}(1 - \frac{z_{1,t}}{\nu_1}) - 2\rho(1 - \frac{z_{1,t}}{\nu_1}) z_{2,t} \right), \\ 0 = \dfrac{\partial \Lambda}{\partial \nu_2} = \dfrac{T}{\nu_2} + \dfrac{1}{2(1-\rho^2)} \displaystyle\sum_{t=1}^{T} \left(2z_{2,t}(1 - \frac{z_{2,t}}{\nu_2}) - 2\rho z_{1,t}(1 - \frac{z_{2,t}}{\nu_2}) \right), \\ 0 = \dfrac{\partial \Lambda}{\partial \rho} = \dfrac{-T\rho}{(1-\rho^2)^2} + \dfrac{\rho}{(1-\rho^2)^2} \displaystyle\sum_{t=1}^{T} (z_{1,t}^2 + z_{2,t}^2) - \dfrac{1+\rho^2}{(1-\rho^2)^2} \sum_{t=1}^{T} z_{1,t} z_{2,t}. \end{cases} \tag{7-13}$$

These are not as tractable as the equations from the previous model, though we can do some similar simplifications, and we can solve for ρ explicitly in terms of ν_1 and ν_2. We set $Z_i = \sum_{t=1}^{T} z_{i,t}$ and $Z_{ij} = \sum_{t=1}^{T} z_{i,t} z_{j,t}$. Keep in mind that each $z_{i,t}$ is a function of ν_i and hence so also are the Z_i and Z_{ij}. Introducing these and clearing the denominators allows us to rewrite the first-order conditions as

$$\begin{cases} 0 = T(1-\rho^2) + \nu_1 Z_1 - Z_{11} - \rho\nu_1 Z_2 + \rho Z_{12}, \\ 0 = T(1-\rho^2) + \nu_2 Z_2 - Z_{22} - \rho\nu_2 Z_1 + \rho Z_{12}, \\ 0 = T\rho(1-\rho^2) - \rho[Z_{11} + Z_{22}] + (1+\rho^2) Z_{12}. \end{cases} \tag{7-14}$$

Comparing the first two of these equations, we see that

$$\nu_1 Z_1 - Z_{11} - \rho\nu_1 Z_2 = \nu_2 Z_2 - Z_{22} - \rho\nu_2 Z_1, \tag{7-15}$$

and this gives us ρ in terms of ν_1 and ν_2:

$$\rho = \frac{\nu_1 Z_1 - Z_{11} - \nu_2 Z_2 + Z_{22}}{\nu_1 Z_2 - \nu_2 Z_1}. \tag{7-16}$$

There seems to be no way to get explicit expressions for ν_1 and ν_2, so we shall have to use iterative techniques. For this system of equations we need to use Newton's method for a system of equations, as discussed in Chapter 5. For initial values we use the parameters from the lognormal model with no drift, since we expect the solutions to be fairly close.

The Bivariate Lognormal Model with Sensed Inflation

Another reasonable model is to assume the bivariate lognormal model, but with a drift which is to be determined from the data. This would be

$$\begin{cases} \log S_{1,t} = \log S_{1,t-1} + \nu_1 z_t + \mu_1, \\ \log S_{2,t} = \log S_{2,t-1} + \nu_2 w_t + \mu_2. \end{cases} \tag{7-17}$$

This is a standard regression equation, and we will obtain the standard estimates. The likelihood function is formally the same as (7-3), but with $z_t = (\Delta \log S_{1,t} - \mu_1)/\nu_1$ and $w_t = (\Delta \log S_{2,t} - \mu_2)/\nu_2$. The equations (7-4) for the partial derivatives are still valid, but to them we must add the two equations for the derivatives with respect to μ_1 and μ_2:

$$\begin{cases} \dfrac{\partial \Lambda}{\partial \mu_1} = \displaystyle\sum_{t=1}^{T} (2z_t(-\frac{1}{\nu_1}) - 2\rho w_t(-\frac{1}{\nu_1})) = 0, \quad \text{and} \\ \dfrac{\partial \Lambda}{\partial \mu_2} = \displaystyle\sum_{t=1}^{T} (2w_t(-\frac{1}{\nu_1}) - 2\rho z_t(-\frac{1}{\nu_1})) = 0. \end{cases} \tag{7-18}$$

Since $\rho^2 \neq 1$, these give us $\sum z_t = \sum w_t = 0$, and applying the above equations for z_t and w_t we get that $\hat{\mu}_1$ and $\hat{\mu}_2$ are the usual sample means:

$$\begin{cases} \hat{\mu}_1 = \dfrac{1}{T} \displaystyle\sum_{t=1}^{T} \Delta \log S_{1,t}, \quad \text{and} \\ \hat{\mu}_2 = \dfrac{1}{T} \displaystyle\sum_{t=1}^{T} \Delta \log S_{2,t}. \end{cases} \tag{7-19}$$

The equations (7-4)–(7-8) now follow exactly as before, with the difference that we take $J_{ij} = \sum_{t=1}^{T} (\Delta \log S_{i,t} - \hat{\mu}_i)(\Delta \log S_{j,t} - \hat{\mu}_j)$. There now is, however, a bias to the estimates of the variances ν_1^2 and ν_2^2. To see this, first note that $E[\Delta \log S_{i,t}^2] = \nu_i^2 + \mu_i^2$ and, for $s \neq t$, $E[\Delta \log S_{i,t} \Delta \log S_{i,s}] = \mu_i^2$. Then

$$\begin{aligned} T\hat{\nu}_i^2 &= \sum_{t=1}^{T} (\Delta \log S_{i,t} - \hat{\mu}^i)^2 \\ &= \sum_{t=1}^{T} (\Delta \log S_{i,t})^2 - \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \Delta \log S_{i,t} \Delta \log S_{i,s} \\ &= \left(1 - \frac{1}{T}\right) \sum_{t=1}^{T} (\Delta \log S_{i,t})^2 - \frac{1}{T} \sum_{t=1}^{T} \sum_{s \neq t} \Delta \log S_{i,t} \Delta \log S_{i,s} \end{aligned} \tag{7-20}$$

and so

$$\begin{aligned} E[T\hat{\nu}_i^2] &= T\left(1 - \frac{1}{T}\right)(\nu_i^2 + \mu_i^2) - \frac{1}{T} \cdot T(T-1)\mu_i^2 \\ &= (T-1)\nu_i^2. \end{aligned} \tag{7-21}$$

To get an unbiased estimate for the variance, we must use the small sample estimate

$$\widehat{\nu}_i = \sqrt{\frac{J_{ii}}{T-1}} = \sqrt{\textstyle\sum_{t=1}^{T} (\Delta \log S_{i,t} - \hat{\mu}_i)^2 / (T-1)}\,. \tag{7-22}$$

N.B. Although the variance estimate is unbiased, the standard deviation estimate $\widehat{\nu}_i$ cannot be unbiased at the same time because, due to the concavity of the square root function,

$$E\left[\widehat{\nu}_i\right] = E\left[\sqrt{\widehat{\nu}_i^2}\,\right] < \sqrt{E\left[\widehat{\nu}_i^2\right]} = \nu_1. \tag{7-23}$$

The Bivariate Normal Model with Sensed Inflation

To make a comparison between the assumptions of lognormal distributions versus normal distributions, we also estimate the hedge using a bivariate normal model which allows for inflation parameter, ξ_1, ξ_2, determined by the data. The specification for this is

$$\begin{cases} S_{1,t} = \xi_1 \cdot S_{1,t-1} + \sigma_1 z_t, \\ \\ S_{2,t} = \xi_2 \cdot S_{2,t-1} + \sigma_2 w_t, \end{cases} \tag{7-24}$$

where $(z_t, w_t) \sim N_2(0, 0, 1, 1, \rho)$ are independent, identically distributed pairs.

We set $\mathcal{D}S_{i,t} = S_{i,t} - \xi_i S_{i,t-1}$ and $J_{ij} = \sum_{t=1}^{T} \mathcal{D}S_{i,t}\mathcal{D}S_{2,t}$. Then $z_t = \mathcal{D}S_{1,t}/\sigma_1$ and $w_t = \mathcal{D}S_{2,t}/\sigma_2$ and the likelihood function $\Lambda = -\sum_{t=1}^{T} \log f(\mathcal{D}S_{1,t}, \mathcal{D}S_{2,t})$ is formally the same as in (7-3). Since the differentiation formulas $\partial z_t/\partial\sigma_1 = -z_t/\sigma_1$ and $\partial w_t/\partial\sigma_2 = -w_t/\sigma_2$ are still valid, all of the subsequent formulas are still good, and the maximum likelihood estimators $\hat{\sigma}_1$, $\hat{\sigma}_2$, and $\hat{\rho}$ are given by (7-7) and (7-8). Keep in mind that J_{ij} now incorporates the modified difference $\mathcal{D}S_{i,t}$, and that the maximum likelihood estimators $\hat{\xi}_1$ and $\hat{\xi}_2$ have not been determined yet.

We now examine the marginal equations for ξ_1 and ξ_2. Using the facts that $\partial z_t/\partial\xi_1 = -S_{1,t-1}/\sigma_1$ and $\partial w_t/\partial\xi_2 = -S_{2,t-1}/\sigma_2$, the corresponding first-order equations become

$$\begin{cases} \dfrac{\partial\Lambda}{\partial\xi_1} = \dfrac{1}{2(1-\rho^2)} \displaystyle\sum_{t=1}^{T} \Big(2z_t\big(-\frac{S_{1,t-1}}{\sigma_1}\big) - 2\rho w_t\big(-\frac{S_{1,t-1}}{\sigma_1}\big)\Big) = 0\,, \\ \dfrac{\partial\Lambda}{\partial\xi_2} = \dfrac{1}{2(1-\rho^2)} \displaystyle\sum_{t=1}^{T} \Big(2w_t\big(-\frac{S_{2,t-1}}{\sigma_2}\big) - 2\rho z_t\big(-\frac{S_{2,t-1}}{\sigma_2}\big)\Big) = 0\,. \end{cases} \tag{7-25}$$

These can be simplified to

$$\begin{cases} \displaystyle\sum_{t=1}^{T} z_t S_{1,t-1} = \rho \sum_{t=1}^{T} w_t S_{1,t-1}, \\ \displaystyle\sum_{t=1}^{T} w_t S_{2,t-1} = \rho \sum_{t=1}^{T} z_t S_{2,t-1}. \end{cases} \tag{7-26}$$

This expands to

$$\begin{cases} \sigma_2 \sum_{t=1}^{T} (S_{1,t} - \xi_1 S_{1,t-1}) S_{1,t-1} = \rho\sigma_1 \sum_{t=1}^{T} (S_{2,t} - \xi_2 S_{2,t-1}) S_{1,t-1}, \\ \sigma_1 \sum_{t=1}^{T} (S_{2,t} - \xi_2 S_{2,t-1}) S_{2,t-1} = \rho\sigma_2 \sum_{t=1}^{T} (S_{1,t} - \xi_1 S_{1,t-1}) S_{2,t-1}. \end{cases} \tag{7-27}$$

Now let $K^{ij} = \sum_{t=1}^{T} S_{i,t-1} S_{j,t-1}$, $K_j^i = \sum_{t=1}^{T} S_{i,t} S_{j,t-1}$ and $K_{ij} = \sum_{t=1}^{T} S_{i,t} S_{j,t}$. (Thus a superscript indicates a lagged variable $S_{i,t-1}$.) Then the equations (7-27) become

$$\begin{cases} \xi_1 \sigma_2 K^{11} - \rho\xi_2 \sigma_1 K^{12} = \sigma_2 K_1^1 - \rho\sigma_1 K_2^1, \\ -\rho\xi_1\sigma_2 K^{12} + \ \xi_2\sigma_1 K^{22} = \sigma_1 K_2^2 - \rho\sigma_2 K_1^2 . \end{cases} \tag{7-28}$$

These can be solved by Cramer's rule, and now we explicitly indicate that the σ_1, σ_2, and ρ involved are the estimators:

$$\begin{cases} \hat{\xi}_1 = \dfrac{K^{22}K_1^1 - \hat{\rho}^2 K^{12} K_1^2}{K^{11}K^{22} - \hat{\rho}^2 K^{12} K^{12}} - \hat{\rho}\dfrac{\hat{\sigma}_1}{\hat{\sigma}_2} \dfrac{K^{22}K_2^1 - K^{12}K_2^2}{K^{11}K^{22} - \hat{\rho}^2 K^{12}K^{12}} \\ \hat{\xi}_2 = \dfrac{K^{11}K_2^2 - \hat{\rho}^2 K^{12} K_2^1}{K^{11}K^{22} - \hat{\rho}^2 K^{12} K^{12}} - \hat{\rho}\dfrac{\hat{\sigma}_2}{\hat{\sigma}_1} \dfrac{K^{11}K_1^2 - K^{12}K_1^1}{K^{11}K^{22} - \hat{\rho}^2 K^{12}K^{12}}. \end{cases} \tag{7-29}$$

These, together with (7-7) and (7-8), give a system of five equations in five unknowns, whose solutions are the maximum likelihood estimators $\hat{\sigma}_1$, $\hat{\sigma}_2$, $\hat{\rho}$, $\hat{\xi}_1$, and $\hat{\xi}_2$. We can use the Newton's method for a system of equations to find the solutions.

In this chapter we have developed equations for the maximum likelihood estimators of the distributions we want to evaluate. In the next chapter we want to apply these, with the help of Newton's method for a system of equations, to a specific example.

CHAPTER 8

JAPAN–TAIWAN–USA 1997–2001: AN EXAMPLE

In this chapter we apply the methods we have developed to a specific historical situation, the Taiwanese dollar, Japanese yen, and U.S. dollar over the period 1997–2001. This is the same example used in CW, and we use this example for the sake of comparison. They used the Japanese yen (JPY) as the domestic currency, the Taiwanese dollar (NTD) as the foreign currency, and the United States dollar (USD) as the third currency. This is not perfectly realistic, since there are robust derivatives markets between all of these currencies, but it is convenient since the exchange rate data are easily available.

Before doing this, we want to point out that CW carried out this part of their paper improperly, and in doing so obtained far too optimistic a conclusion for the use of options. The proper fitting of the parameters for their model should have been obtained by a regression of $\Delta S_{2,t}$ on $\Delta S_{1,t}$. Instead, they incorrectly regressed $S_{2,t}$ on $S_{1,t}$ and applied the hedging model to the undifferenced series, where it makes no sense, instead of applying it to the differenced series, where it does make sense. As a result, they claim that in this situation there is a further 5% reduction in the variance of the cash flows when options and futures are used, over the reduction obtained by using futures alone. This is an amount which makes the use of options an appealing addition to the hedging strategy. This is, however, far too generous. We shall see here that the true reduction using options is a paltry 0.044% over the reduction obtained by using futures alone. With such a small improvement, the use of options no longer looks economically feasible. The improvement is so small that it is hard to imagine a situation where the additional cost or trouble of using options for currency hedging would be worthwhile.

We will compare five different hedging strategies. The first will be to make no hedge. The second will be to use only futures, assuming a bivariate normal process, with the hedge H^0, as in CW. The third will be to use only futures, assuming a bivariate lognormal process, with the hedge H^0 as given in Chapter 4. The fourth will be to use both futures and options, assuming a bivariate lognormal distribution, using H^* and Z^* as in CW. The fifth will be to use both futures and options, assuming a bivariate lognormal distribution, using H^* and Z^* as discussed in Chapter 4.

We assume there is a cash flow of X = NTD100 expected in one week for each week that is to be hedged. The parameters will be calibrated from data over the previous 52 weeks. Since the objective is to minimize the variance of the cash flow, we will calculate the sample variance of the differences of the actual cash flow CF with the expected cash flow $\mathrm{CF}_0 = S_{t-1}X$ for each of the models. This will be done over 200 weeks, from 17.10.1997 through 30.3.2001 (date convention is dd.mm.yyyy), and the variances will be reported in batches of 20 weeks, together with a cumulative total.

For a bivariate normal process $S_{i,t} = S_{i,t-1} + \nu_i z_i$, $t = 1, \ldots, T$, where $(z_1, z_2) \sim$

$N_2(0,0,1,1,\rho)$, we set $J_i = \sum \Delta S_{i,t}^2$ and $X = \sum \Delta S_{1,t} \Delta S_{2,t}$, and then the maximum likelihood estimates are $\hat{\nu}_i = \sqrt{J_i/T}$ and $\hat{\rho} = X/\sqrt{J_1 J_2}$. Since there is no constant term in the relation $\Delta S_{i,t} = \nu_i z_i$, this is unbiased, and we do not need to correct for bias by replacing T with $T-1$ in the estimates of $\hat{\nu}_i$. (The calculations in CW used a regression of $S_{2,t}$ on $S_{1,t}$, which is not appropriate in this model.) For the bivariate lognormal model, we do the same, replacing each $\Delta S_{i,t}$ with $\Delta \log S_{i,t}$. The results are reported in Table 1. The period analyzed is 11/1997 through 4/2002. This was chosen because it's the same period of time considered by CW. There is data from 1983, but the results over all of this time are fairly similar. In most periods the unhedged cash flow has by far the highest variance.

Table 1

Comparison of cash flow variances for five hedging strategies

Date	unhedged	NL ni,fo	LGNL ni,fo	LGNL nd,fo	NL si,fo	LGNL si,fo	NL ni	LGNL ni	LGNL nd	NL si	LGNL si
14.11.1997	60.355	31.580	31.545	31.540	31.502	31.465	31.605	31.556	31.546	31.523	31.558
3. 4.1998	49.682	39.602	39.907	39.945	39.922	40.249	39.640	39.923	39.951	39.952	38.665
21. 8.1998	81.277	3.768	4.033	4.020	4.281	4.791	3.801	4.052	4.064	4.332	5.799
8. 1.1999	158.130	7.582	7.303	7.266	6.954	6.798	7.242	7.179	7.172	6.696	5.822
28. 5.1999	57.270	4.554	4.962	4.959	4.560	4.950	4.591	4.981	4.965	4.597	3.679
15.10.1999	39.620	1.270	1.402	1.404	1.255	1.381	1.238	1.385	1.389	1.224	0.947
3. 3.2000	33.189	2.746	2.825	2.825	2.739	2.813	2.755	2.829	2.827	2.748	2.531
21. 7.2000	27.838	0.641	0.647	0.644	0.689	0.697	0.644	0.648	0.644	0.691	0.836
8.12.2000	7.752	3.756	3.777	3.777	3.769	3.798	3.752	3.775	3.784	3.765	3.543
27. 4.2001	23.327	2.392	2.321	2.319	2.317	2.260	2.401	2.324	2.321	2.324	2.020
Average	53.844	9.789	9.872	9.870	9.799	9.920	9.767	9.865	9.866	9.785	9.540

NL=normal, LGNL=lognormal, ni=no inflation, fo=futures only, nd=no drift, si=sensed inflation

We see a considerable reduction of the variance of the cash flows by using any one of the hedges, with the average reduction being a little over 80%.

When any of the hedging methods are employed, the amount of reduction in the average variance of the cash flows is very nearly the same. If we have to pick a best of the lot, it is the last one, the one which uses the bivariate lognormal model with an inflation term included, and this achieves a 2% or 3% improvement over the other models. However, there is considerable variablity in the numbers and this improvement is neither statistically nor economically significant.

A larger concern is the discrepancy between these averages for the variance of the cash flows, compared with the theoretical averages, if the assumed model were to match reality fairly well. If the bivariate lognormal model with no drift were assumed to be accurate, then the average of the variances of the cash flows should be around 3.60, and the values in Table 1 are considerably higher. This point is discussed further in the next chapter.

The above 200-week period was chosen to match the period considered in CW. There is data, however, from 1983, and in Figure 1 we show graphically the cash flow differences $CF - CF_0$ for each week in the roughly 900-week period from 1983 through 2002, where CF_0 is the cash flow which would have resulted if the exchange rates at the end of the period were the same as those at the beginning of the period. The results for three of these hedges are displayed. The first graph is for the unhedged cash flows, the second is for the cash flow differences resulting from the

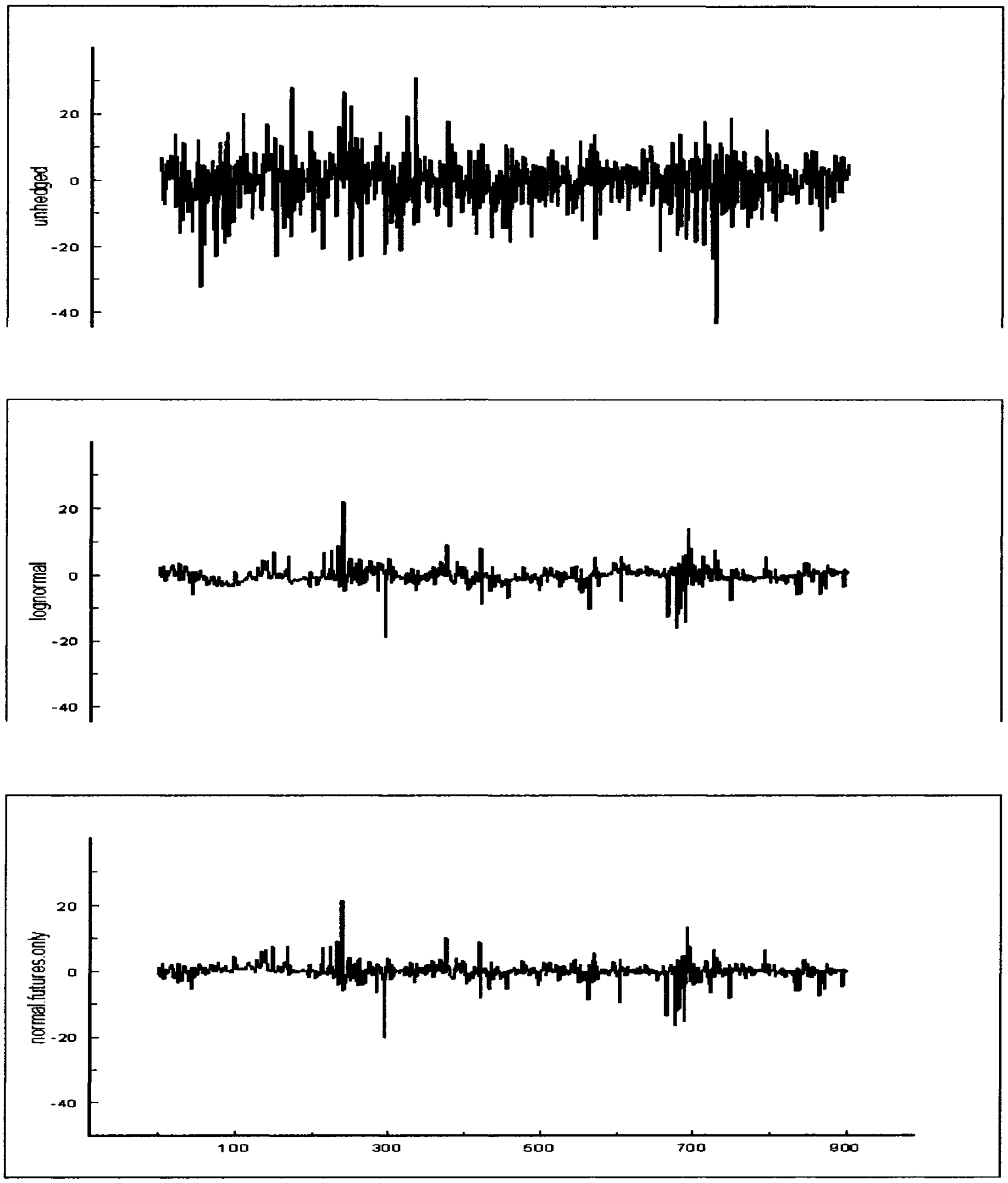


Figure 1. Cash flow deviations over a 900-week period 1983–2002. Panels are a) unhedged, b) lognormal, c) normal: futures only.

hedges derived from the bivariate lognormal model, and the third is the cash flows resulting from the hedges derived from the bivariate normal model without the use of options. The first thing to note is how very much similar the last two are. At first glance, they seem to be the same, although small differences can be detected. The graphs for the other seven hedges are just as similar and are omitted. The second thing to note is how big the improvement is by any hedge over the unhedged cash flows. In particular, any sort of hedge removes the huge negative spikes at around weeks 50 and 720.

We turn in the next chapter to a comparison of these real-life results with theoretically predicted results.

CHAPTER 9

SIMULATION RESULTS

In this chapter we wish to apply simulations to compare the real-life results of the previous chapter to those predicted by theory. In an earlier chapter we developed formulas for moments under the assumption of a bivariate lognormal model with no drift. If this is assumed to be the true underlying process, then using various hedge methods would lead to an expected variance of cash flows as shown in the first line of Table 2. There is a large discrepancy between these theoretical values and the actual averages shown in the last line of Table 1. The most likely explanation for this discrepancy is that the bivariate lognormal model is not a good representation of the real-life process. However, before reaching that conclusion, we wish to first rule out a couple of other possibilities.

Table 2

Comparison of theoretical and simulated cash flow variances

method:	unhedged	NL ni,fo	LGNL nd,fo	sample fo	NL ni	LGNL nd	sample
theoretical	27.49353562	3.59522822	3.59522818		3.59524320	3.59524320	
simulated	27.49266499	3.59694874	3.59694836	3.59694764	3.59697377	3.59693599	3.59693449

The simulation is 7 million samples from the bivariate lognormal model with no drift.
NL=normal, LGNL=lognormal, ni=no inflation, fo=futures only, nd=no drift, si=sensed inflation

The first possibility we wish to rule out is that the formulas used for the theoretical calculation are in error. To check this, we run a lengthy simulation of the bivariate lognormal model with no drift and compare the resulting variances of the cash flows. We use the values $\nu_1 = 0.01641526297$, $\nu_2 = 0.005990652048$, and $\rho = -0.2322515625$, which match those of the Taiwanese dollar, Japanese yen, and U.S. dollar exchange rates over the period 1996–2001. The results are shown in the second line of Table 2, and these show that there is no significant difference between the simulated values and the theoretical values. We conclude that the theoretical calculations are not the source of the discrepancy. The last column of Table 2 refers to a "sample" method. This is the *ex post* optimal hedge. That is, after all of the data are known, H and Z are calculated to minimize the cash flow variance given exactly those data. Thus no theoretical method can do better.

A second possible source of the discrepancy is the fact that estimators were used for the quantities ν_1, ν_2, R, σ_1, σ_2, and ρ. The theoretical values in Table 2 were calculated with the actual values for the parameters, and using estimators will introduce some discrepancy. To see how much effect this has, we run a simulation which generates sets of 252 weeks of simulated exchange rates. The hedges are then implemented for each of the last 200 weeks in the set, with estimators calculated from an entirely independent 52-week simulation for each week of the set. Thus the calculation of the cash flow variances requires 10,652 simulations for each set of 252 weeks, and so we cannot do anywhere near the 7 million iterations we did for Table 2. We repeat the simulation of such a period 2,100 times and compare the variances of the cash flows for several of the hedging methods. Table 3 reports the upper and lower limits for a 99% confidence interval for the mean cash flow variance based on this simulation. The lower limit is higher than the theoretical values given in Table 2, but not nearly so high as those encountered in Table 1.

Table 3

Cash flow variances based on estimated parameters

Method	unhedged	normal ni,fo	lognormal nd,fo	normal ni	lognormal nd
Upper	29.63354370	3.954476546	3.951799010	3.954641218	3.951773151
Lower	28.67119301	3.824129545	3.821534938	3.824266617	3.821501509

Upper and lower 99% confidence intervals for the cash flow variances.

From these simulation results, we can conclude that the difference between the real-life average cash flow variances shown in the last line of Table 1 and the theoretical mean cash flow variances calculated for the first line of Table 2 occurs because the model assumed, namely the bivariate lognormal model with no drift, is not a very good fit for the real-life exchange rates process.

CHAPTER 10

CONCLUSION AND DIRECTIONS FOR FUTURE RESEARCH

In this thesis we considered the situation of a firm receiving a payment or payments in the future denominated in a foreign currency, as was done in Chang and Wong (2003). The firm is assumed unable to directly hedge the foreign exchange risk and wishes to partially hedge the risk by selling futures or options in a third currency. Under the assumption of a quadratic utility function for the firm, Chang and Wong calculated the optimal hedging strategy under some assumptions about the distribution of the exchange rate processes. This thesis shows that, still using the assumption that the firm's utility function is quadratic, those assumptions can be removed and virtually the same results can be obtained. This allows for other distributional assumptions to be fitted, and the formulas for the optimal hedge were calculated for several variations, based on the bivariate lognormal model or the bivariate normal model. These formulas were then applied to the exchange rate series for Japan-Taiwan-United States over the period 1997-2001, and the variance of the cash flows after hedging were compared for these models. The best performance was obtained by the model which assumed a bivariate lognormal process with an inflation term, but the differences between the models were not statistically significant. Simulations showed that the considerable gap between the variances of the cash flows experienced by the real-life data and those predicted by the theory was not due to errors in the calculation nor by the use of estimators for the parameters, but it is rather due to the inadequacy of the models considered to explain the real-life data.

There are several points that future investigations can examine. First, data from other countries can be used. This should be fairly simple, as exchange rate data is readily available.

A second avenue for investigation is the use of different distributions. This is more difficult. Even in the simplest case considered in this thesis, there was considerable difficulty deriving the theoretical values for the hedging strategy. In order to use other distributions it will probably be necessary to develop approximating techniques for the moments necessary to implement the hedge. A computer algebra system might be able to evaluate some of these, but it might not; many integrals do not have a closed-form solution. Numerical methods of integration also have a difficulty in that the solution requires multiply iterated integrals, and approximations to these are difficult to make accurate. It is also possible to approximate the distributions by simulation, but that does not give more than a few digits of accuracy. Since the hedge formulas seem to be fairly sensitive to the values of the moments, this may not be a reliable method. Thus, there will be trouble adapting the results to more complicated distributions no matter what approximation technique we choose.

The underlying method, as developed in this thesis, used only options with a single at-the-money strike price. Allowing the firm to choose an optimal strike price would be more realistic. Implementing this modification would be a fairly straightforward extension of the techniques we have developed. It would also be possible to allow options at two or more strike prices. This introduces an extra variable into the first-order conditions for the optimum, but no further difficulties.

The most pressing need is to find out why there is such a large difference between the theoretical performance of the hedges and the real-life results. This is certainly due to the inadequacy of the bivariate lognormal model, and there are a number of better models discussed in the literature. It would not be too difficult to develop a simulation using one of these improved models, and then applying the hedge formulas from this thesis would show whether this closes the gap. If not, then more work would be necessary.

It is possible to introduce other variations in the model which would bring it closer to reality. The interest rate is always uncertain, and the result that the interest rate didn't matter was only true when that rate was certain. So an uncertain interest rate will change the nature of the solution. The amount that the firm expects to receive is often uncertain. Although one future payment might be known exactly, if we're applying this technique to a series of future payments, we can expect some of them to have uncertainty. There is a question, then, of how significant that uncertainty is.

BIBLIOGRAPHY

Black, Fischer & Myron J. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654.

Bodnar, Gordon M. & Günther Gebhardt (1999). Derivatives usage in risk management by U.S. and German non-financial firms: A comparative study. *Journal of International Financial Management and Accounting*, 10(3):153–187.

Bodnar, Gordon M., Gregory S. Hayt & Richard C. Marston (1996). Wharton 1995 survey of derivative usage by U.S. non-financial firms. *Financial Management*, 25(4):113–133.

Bodnar, Gordon M., Gregory S. Hayt & Richard C. Marston (1998). 1998 Wharton survey of financial risk management by US non-financial firms. *Financial Management*, 27(4):70–91.

Bodnar, Gordon M., Gregory S. Hayt, Richard C. Marston & Charles W. Smithson (1995). Wharton survey of derivative usage by U.S. non-financial firms. *Financial Management*, 24(2):104–114.

Bodnar, Gordon M., Abe de Jong & Victor MacRae (2003). The impact of institutional differences on derivatives usage: A comparative study of U.S. and Dutch firms. *European Financial Management*, 9(3):217–297.

Børsum, Øystein G. & Bernt Arne Ødegaard (2005). Currency hedging in Norwegian non-financial firms. *Economic Bulletin*, Q3:133–144.

Chang, Eric C. & Kit Pong Wong (2003). Cross-hedging with currency options and futures. *Journal of Financial and Quantitative Analysis*, 38(3):555–574.

Ding, Zhuanxin & Clive W.J. Granger (1996). Modeling volatility persistence of speculative returns. A new approach. *Journal of Econometrics*, 73(1)185–215.

Giot, Pierre & Sébastien Laurent (2003). Value-at-risk for long and short trading positions. *Journal of Applied Econometrics*, 18:641–664.

Jakoniuk, Gustavo, Mariano Tappatá & Eduardo Levy Yeyati (2000). El uso de instrumentos derivados en empresas no financieras: El caso de Argentina. Working paper.

Johnson, Norman L. & Samuel Kotz (1972). *Distributions in Statistics: Continuous Multivariate Distributions.* New York: John Wiley & Sons.

Reeve, Ted (1991). *Global Gold Hedge Survey* (First Boston Corporation, June 1991, Aug. 1991, Nov. 1991, Feb. 1992, June 1992, Sept. 1992, March 1993).

Reeve, Ted (1993). *Gold Hedge Survey* (Sanwa McCarthy Securities, Aug. 1993, Dec. 1993).

Reeve, Ted (1994). *Gold Hedge Survey* (Burns Fry Limited, Feb. 1994).

Taylor, Stephen (1986). *Modelling Financial Time Series.* Chichester: John Wiley and Sons.

Tufano, Peter (1996). Who manages risk? An empirical examination of risk management practices in the gold mining industry. *The Journal of Finance*, 51(4):1097–1137.

Zumbach, Gilles (2004). Volatility processes and volatility forecast with long memory. *Quantitative Finance*, 4(1):70–86.