Essays in Finance

by

Walter Toshihide Baily

A.B., Economics, Princeton University (1991)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1995

Author . . . .
Department of Economics
May 24, 1995

Certified by . . . .
Andrew Lo
Harris & Harris Group Professor of Finance
Thesis Supervisor

Certified by . . . .
Jeremy Stein
J.C. Penney Professor of Finance
Thesis Supervisor

Accepted by . . . .
Richard Eckaus
Chairman, Departmental Committee on Graduate Students

Essays in Finance

by

Walter Toshihide Baily

Submitted to the Department of Economics
on May 24, 1995, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis investigates three topics in financial economics. The first chapter develops a Bayesian procedure to select between various option pricing models based on historical data on the time series of the underlying asset as well as the options data. The exact form of the procedure is first developed. To better understand how this procedure chooses a model, I develop an approximation to the key quantities at issue and interpret the model selection procedure based on these approximations. The Bayesian methodology is demonstrated on foreign exchange options traded on the PHLX and used to compare the Black–Scholes model, Merton's jump–diffusion model and Cox's CEV model. When the complete data set of both the time series and options data is use, the jump–diffusion model is found to be the most probable model.

The second chapter uses option prices to examine whether crash fears in currencies explain the empirical failure of uncovered interest parity (UIP). In particular, we investigate the claim made by many traders that the high yield currency is subject to crashes. This has been given as a possible explanation of the failure of UIP and has come to be known as the peso problem explanation. By using observed option prices, we invert for jump expectation in the DM–Dollar and Yen–Dollar exchange rates from 1984 to 1993. We find that in the case of the DM, that indeed the high yield currency is generally expected crash while for the Yen, we find no relationship between the jump expectation and yield differential. We also find that option implied jump expectations weakly predict large movements. We find that exchange rates are expected to jump in the direction of purchasing power parity while the trade balance has no explanatory power in the formation of jump expectations.

The third chapter investigates how financing decisions can affect an entrepreneur's investment decision when the private information held by the entrepreneur is evolving over time. We find that when "standard securities"—securities whose value depend positively on the value of the firm—are issued, the entrepreneur will not wait for sufficient information and end up investing too early. For a stylized example, we show that with a risky project and limited liability, the entrepreneur can only issue such securities and consequently will always end up investing too early. A social planner faced with the same informational constraints as the financial market will want to impose a tax on issuing securities which will induce the entrepreneur to wait for sufficient information. If there are multiple firms with the same underlying private information, the firms with less cash on hand will "piggyback" on firms with more cash and issue their securities right after the cash rich firms do.

Thesis Supervisor: Andrew Lo
Title: Harris & Harris Group Professor of Finance

Thesis Supervisor: Jeremy Stein
Title: J.C. Penney Professor of Finance

Acknowledgments

Numerous people contributed to my development and the development of this thesis over the past few years. My advisors Andrew Lo and Jeremy Stein provided unwavering guidance and support. For that I owe them a debt I can never repay. I also owe Chi-fu Huang a heartfelt thanks for always being a constant source of inspiration and energy. David Scharfstein deserves a great thanks for his guidance during my second and early third year. I also thank him for being my third reader. I benefitted greatly from the friendship and guidance of many members of the Sloan faculty including John Heaton and Jiang Wang. They always gave of their time freely despite the fact that I was from the economics department. I owe Ricardo Caballero a thanks for giving me so much of his time during my second year. Thanks go to Dan Stroock and Donald Rubin for making probabilities come alive and making them live somewhere other than on $(\Omega, \mathcal{F})$.

I thank both the National Science Foundation and the U.S. Department of Education Jacob K. Javitz Fellowship Program for financial support. The NSF also provided funds for the RS6000 workstation known as Trader. I thank the World Economic Laboratory for summer funding during my second year.

Friends made Cambridge livable. I thank them for that. Li-Lan Cheng, Dan Lagattuta, Eric Lindblad, and Jojo Nyi Nyi made each day a little different and a little better. Peter Klibanoff gave me my sanity during my first year. Kiyohiko Minemoto made Cafe Mami the Cheers of Porter Square. Leo Kropywiansky helped in the trenches hunting for jumps in options. Jason Abrevaya, Ludwig Chincarini, Harrisson Hong and Nick Souleles helped burn the midnight oil and Trader CPU in E52-364. John Dippold single–handedly saved half the class, present party included, from themselves and the RS6000's. Chapter 2 would not exist without him. Michael Lee gave life to E52-243g. Marcello Cherchi helped ponder the graduate student existence. Mamoru Taniya gave me a home at Sali after MIT. And to all those friends not mentioned by name, thanks.

My parents deserve all the rest of my thanks. To them I dedicate this thesis.

To my parents

"So you're an economist. That means you study the stock market. What else?"

—The author, senior year in high school.

Contents

List of Figures

List of Tables

Chapter 1

A Bayesian Comparison of Option Pricing Models

1.1 Introduction—The Need For a Bayesian Framework

There has been a proliferation of option pricing models after the introduction of the original Black–Scholes model in 1973. Among them, several popular ones include Merton's jump-diffusion model and Cox's constant elasticity of volatility model. At the same time, the volume of trade in options and other derivative securities over the last 20 years has increased at a break-neck pace. Financial practitioners have found theoretical models essential to carry out even the most mundane transactions in the derivatives markets. But while the importance and number of option pricing models grows, there remains a gap between the theory and practice. How does one choose the model? Sophisticated statistical tests have been developed for testing option pricing models while numerous studies have compared various option pricing models in a wide range of markets. There has been relatively little emphasis, however, on the topic of optimal model selection. Because most studies have relied on the use of classical statistical methods, the conclusions reached are generally to accept or reject a model. If several models are accepted, there is generally no prescription by which to select among those models. An even worse scenario, which is actually the more common, is for all models to be rejected.

In this paper, we propose to tackle this problem by use of Bayesian methods. The

Bayesian approach will allow us to not only rank all models, but also will allow us, on the basis of the data, to impute implied probabilities to each model. This will in turn allow one to use decision theoretic methods by which to select among the models. Or if one prefers, it is even possible to take advantage of all models to price a new option. The Bayesian framework does not force one to choose just one model.

In our approach, we also take advantage of both the time series and option pricing implications of the respective model. In a typical option pricing model, dynamics are specified for the underlying asset. Together with preference restrictions it is shown that these dynamics imply an equilibrium price for an option on that stock. While some studies have focused on how well the stock price time series data fits the specified dynamic process, others have focused on how well the option pricing formulas fit the actual option prices. Few have examined both questions together. Our approach will simultaneously encompass both these issues.

To gain a better understanding of the Bayesian method of model comparison, Section 1.2 spends some time laying out and interpreting the basic methodology of comparing among models. Particular attention is paid to how the Bayesian method chooses between a parsimonious model and a more complicated one that may fit the current data set better, but generalize poorly. This analysis will clarify some issues that separate classical and Bayesian methods. We will see how Lindley's paradox can be interpreted in this light. Section 1.3 develops the basic setup and notation. Section 1.4 examines how one computes the implied probability of an option pricing model given the data. Section 1.5 derives the basic quantity which must be computed numerically to assign probabilities to models. Section 1.6 interprets this quantity by first deriving an approximation. This approximation reveals some of the key issues for comparing option pricing models. Section 1.7 summarizes three option models that we will compare empirically. Section 1.8 describes the PHLX foreign currency data source, while Section 1.9 specifies a set of priors that is applicable to the data and models considered. Section 1.10 gives the empirical results while Section 3.7 concludes.

1.2 Bayesian Model Comparison

Given several option pricing models, our objective will be to quantitatively rank the models on the basis of the data. The Bayesian framework provides a natural setting in which to analyze this problem. A key distinction of the Bayesian approach is that it can assign probabilities to different models on the basis of a given data set. A ranking based on these relative probabilities will automatically be a consistent method by which to evaluate the various models. The classical approach of ranking models, in comparison, suffers from the possibility that several models can be accepted without providing a ranking of those accepted models. An even worse scenario, especially relevant in the context of comparing option pricing models with large data sets, is that all models in consideration may be rejected.[1] This result will be troubling especially for financial practitioners who must select some option pricing model on which to base their pricing and hedging strategies.

In contrast, not only does the Bayesian approach assign consistent ordinal rankings to the various models, but it also does not force the practitioner to simply choose one model with which to work (since it also provides a cardinal ranking). If the practitioner desires, it is possible to utilize all the models and estimates to price an option. The following is just one such example. One can first determine the option price implied by each model. Next one can form a weighted average of these option prices using the probabilities of each model given the previous data. This yields an estimated option price that is the "expected option price". In this way, the Bayesian approach allows one to take maximum advantage of the models and data.

In this section, we will briefly examine the Bayesian approach to model selection paying particular attention to how the Bayesian procedure strikes a balance between selecting a model that fits the data well and selecting a parsimonious model that does not "over- fit" the data.

[1]Cox's [2] procedure, which is classical in nature, does not overcome this problem since it too must assume a null hypothesis. See Gaver and Geisel [5] for the case of regressor selection.

1.2.1 The Two Levels of Inference

Two levels of inference are generally involved in evaluating the consistency of a model with a given data set. First one must fit the model to the data. Because there are typically some free parameters in a model, one must first estimate these parameters from the given data by assuming the given model is correct. This first step is referred to as parameter estimation. The second level of inference is to rank the various models in consideration. In the case of a regression model, at the first stage, one must estimate the parameters of the regression model assuming that the model is correctly specified. At the second level one must express a preference over the various regression models.

Unfortunately, however, one cannot simply rank the models by how well they fit the data. It is a general fact that a more complex model will fit the data better. A linear regression function with two regressors will always fit the data better relative to another with only a subset of the regressors. A polynomial regression function with $N+1$ terms will always fit the data better than a polynomial regression function with only N terms. Selecting a model on the basis of its goodness of fit will thus consistently lead to accepting models that are overly parameterized and which generalize poorly. It is clear that there is a need for some formalization of the principle that *a complex model should not automatically be preferred to a simpler model even if the simpler model does not fit the current data set as well.* This has long been recognized as the basic building block of the non-parametric regression literature. What is not well known is that the Bayesian procedure for model comparison automatically and quantitatively captures this principle. And it does so without appealing to ad hoc penalties for model complexity.

1.2.2 The Bayesian Occam's Razor

At the first level of inference, namely parameter estimation, the Bayesian procedure yields a posterior distribution for the estimated parameter. If the data set is large enough and the prior distribution on the parameter does not have a strong peak, this posterior distribution will be approximately Gaussian. What is more, the mean of this Gaussian will be close to the maximum likelihood estimate (MLE) of the parameter. The variance of the Gaussian will be roughly equal to the MLE sampling theory variance.

In other words, the Bayesian procedure yields an estimate of the parameter in question similar to the classical approach. All that really seems to have changed in going from the classical to Bayesian procedure is the interpretation of the final distribution of the parameter. This fact, unfortunately, seems to have left some believing that the whole Bayesian approach only differs from the classical approach by the inclusion of subjective priors.

At the second level of inference where models are compared, however, the Bayesian procedure for model comparison turns out to be quite distinct from and perhaps more self-consistent when compared to the classical procedure of hypothesis testing. What is more, the Bayesian procedure naturally captures the principle of parsimony discussed earlier. This principle is often referred to as the Bayesian Occam's Razor (MacKay [14], Kashyap [9]). To see how it works, in what follows for the remainder of this subsection 1.2.2, we present a shortened version of the development in MacKay [14].

Consider again the two levels of inference. To set our bearing, we define some variables. Each model M_i has a vector of free parameters δ. A model is defined by its (1) parametric form (for example, the regression function and the distribution of the error term in a regression model), (2) the prior distribution of the free parameter δ, $P(\delta|M_i)$, and (3) the likelihood of the data D when δ is fixed $P(D|M_i, \delta)$.

At the first level of inference, the model M_i is assumed to be correct and we estimate the probable value of the parameter given the data. By Bayes' rule:

$$P(\delta|D, M_i) = \frac{P(D|\delta, M_i)P(\delta|M_i)}{P(D|M_i)} \tag{1.1}$$

To find the most probable value of δ, one maximizes $P(\delta|M_i, D)$. If the prior $P(\delta|M_i)$ is relatively flat, the most probable value of δ, δ_{MP} will coincide with the MLE estimate of δ. It can also be shown that $P(\delta|M_i, D)$ is approximately Gaussian centered at δ_{MP} when the sample D is large.

At the second level of inference, we would like to compare various models and assign probabilities to them on the basis of the data. Again, by Bayes rule:

$$P(M_i|D) = \frac{P(D|M_i)P(M_i)}{P(D)} = \frac{P(D|M_i)P(M_i)}{\sum_j P(D|M_j)P(M_j)} \tag{1.2}$$

If we like, we can assume flat priors over the various models, $P(M_i) = P(M_j)$ and simply evaluate:

$$P(M_i|D) = \frac{P(D|M_i)}{P(D)} = \frac{P(D|M_i)}{\sum_j P(D|M_j)} \tag{1.3}$$

Such a procedure would be analogous to maximum likelihood parameter estimation.

So to assign probabilities to the various models, all we need to do is evaluate the quantity $P(D|M_i)$. (It should be pointed out that this is *not* simply the classical likelihood function since the model does not specify the value of the free parameter δ.) The quantity $P(D|M_i)$ can be written as:

$$P(D|M_i) = \int P(D|M,\delta)P(\delta|M)d\delta \tag{1.4}$$

Noting that the integrand $P(D|M,\delta)P(\delta|M)$ is proportional to $P(\delta|M,D)$, it is clear that the integrand will have a strong peak at δ_{MLE} if the prior is not dogmatic. Suppose the peak of the integrand, loosely speaking, has a width of $\Delta\delta$. Then

$$P(D|M_i) \approx P(D|M,\delta_{MLE})P(\delta_{MLE}|M)\Delta\delta \tag{1.5}$$

The first term is the maximized likelihood function. The second and third piece constitute the Occam factor which is a factor less than one and penalizes the model M for having a free parameter. The larger $\Delta\delta$, the greater the posterior uncertainty in δ. It measures the "posterior accessible volume" of δ. On the other hand $\frac{1}{P(\delta_{MLE}|M)}$ is a rough measures of the "prior accessible volume" of δ. To see this, we note that if $P(\delta|M)$ is a relatively flat prior over a range of width $\Delta_0\delta$, then $\Delta_0\delta = \frac{1}{P(\delta|M)}$. So the Occam factor is equal to:

$$\text{Occam Factor} = \frac{\Delta\delta}{\Delta_0\delta} \tag{1.6}$$

This measures the ratio of the posterior accessible volume to the prior accessible volume of δ. A complex model with many parameters that are free to vary over a large range will be penalized with a large $\Delta_0\delta$. The Occam factor also penalizes models that have to be finely tuned to fit the data. In other words, the fact that $\Delta\delta$ is small, reflects the fact that only a few values of δ could have fit the data that well. Conversely, if any value of δ would have done as good a job in fitting the data, the model is probably a fairly good model. So even if the prior on δ is not very precise, if numerous values of δ would have fit the data just as well, the model will not be punished very much. The Occam factor thus measure the complexity of the model conditional on the data.[2]

1.2.3 Some Examples of Occam's Razor in the Literature

Probably the most famous example of Occam's factor coming into play is in Lindley's [10] paradox. There, a very simple model, a point null: $\delta = 0$, is compared to a very complex model where δ can take any value other than 0. As the sample becomes large, with the t-statistic kept constant for the point null, the simple model is preferred more and more by the Bayesian approach. This is because $P(D|M_{\delta=0})$ is kept fixed (the t-statistic is a measure of $P(D|M_{\delta=0})$) while the Occam factor for the complex model is shrinking with the sample size. The Occam factor is shrinking because the posterior accessible volume for δ given the complex model and the data is shrinking—note that there is no Occam factor for the simple model since it does not have a free parameter. The Occam factor for the complex model thus falls and causes $P(D|M_{complex})$ to drop. Thus the relative probability of the simple model rises. This shows that the Bayesian and classical approaches can be quite divergent even for relatively simple problems and that the Occam factor plays a key role in that distinction.

[2] Further details can be found in MacKay [14].

The Bayesian Occam factor also proved useful in analyzing the problem of regressor selection. Zellner [21] shows how the Akaike information criterion can be viewed as a first order approximation of the Bayesian prescription for regressor selection. Gull [6], [7], Skilling [20] and MacKay [14] applied the Bayesian approach to analyze the problem of non-parametric regression and the selection of bandwidth and basis functions. MacKay also applied this approach in [15] to the selection of network architectures in the context of back propagation neural networks.

1.3 The Setup for Comparing Option Pricing Models

In this section, we layout the basic setup of an option pricing model and develop the relevant notation.

We start with a complete filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$. There is an underlying asset s that follows an adapted process to be specified. An option y whose underlying security is s will typically promise its holder a payment of $g(s_T)$ at time T. A model of option pricing, denoted as M, will consist of the following components:

(1) The model M specifies the dynamics for s. Typically the process for s is characterized by a stochastic integral equation of the following form:

$$s(T) = s(T_0) + \int_{T_0}^{T} a_M(s,t;\delta_0)dt + \int_{T_0}^{T} b_M(s,t;\delta_0)dW(t) + \int_{T_0}^{T} c_M(s,t;\delta_0)dN_{\delta_1}(t) \quad (1.7)$$

where under P, $W(t)$ is a standard Wiener process and $N_{\delta_1}(t)$ is a standard Poisson jump process with jump intensity δ_1. We define $\delta \equiv \{\delta_0, \delta_1\}$ Note that δ is a free parameter whose value is left unspecified.

(2) The model M specifies an option pricing formula. In a perfectly competitive, complete information and frictionless environment, the model M specifies the price of an option y to satisfy $y(t) = f(s(t), z(t); \delta)$ where $z(t)$ is an $\mathcal{F}_t$ measurable vector variable. In the case of the Black–Scholes model, the vector $z(t)$ would include such elements as the time to maturity, the risk-free interest rate and the strike price. However, in the presence of market imperfections, such as transaction costs, bid–ask spreads and misinformed market participants, the price of the option $y(t)$ will deviate from $f(s(t), z(t); \delta)$ by an error term ϵ_t. Thus:

$$y(t) = f(s(t), z(t); \delta) + \epsilon_t \tag{1.8}$$

As for the distribution of ϵ, we will always restrict ϵ_t to have mean zero conditional on $s(t)$ and $z(t)$. In our initial analysis, we will also take ϵ_t to be iid Gaussian with variance σ^2.

(3) The model M specifies priors over the free parameters δ and σ. This is a non-standard element for an option pricing model. It will be necessary and informative, however, to specify the following priors:

$$P(\delta|M) = \text{ The prior distribution of } \delta \text{ given the Model } M \tag{1.9}$$

$$P(\sigma|M) = \text{ The prior distribution of } \sigma \text{ given the Model } M \tag{1.10}$$

Our initial analysis will leave the prior over δ unrestricted, but for tractability will examine the case where the prior over σ^2 is an inverted Gamma distribution. If one believes that the error term ϵ_t will generally be small, this can be captured in the prior distribution for σ. This should satisfy those who believe that market imperfections are small. This completes the model.

The data will be characterized as follows. The time series data on the underlying security, which we will call the stock for simplicity, consists of K pairs $\{s_k, \tau_k\}$, where $k = 1, \ldots, K$. The price of the stock at time τ_k is s_k. The option price data consists of N pairs, $\{y_n, \theta_n\}$, where $n = 1, \ldots, N$. The n-th pair, $\{y_n, \theta_n\}$ consists of the option price y_n at time θ_n. We will assume that $\{\theta_n\}_{n=1}^{N}$ is a subset of $\{\tau_k\}_{k=1}^{K}$ so that we know the price of the underlying security for each observation of an option price.

We define the following variables which will facilitate our discussion in the following sections:

$$
\begin{aligned}
Y &= \{y_n\}_{n=1}^{N} \text{ Option Price Data} \\
Z &= \{z_n\}_{n=1}^{N} \text{ Non-Price Arguments Entering the Option Formula} \\
S &= \{s_k, \tau_k\}_{k=1}^{K} \text{ Stock Price Time Series Data}
\end{aligned} \tag{1.11}
$$

1.4 The Probability of a Model Given the Data

We would like to evaluate the probability of a model M_i given the data Z, S, Y. Bayes' rule can be used to yield:

$$P(M_i|Z,S,Y) = \frac{P(Y|M_i,S,Z)P(M_i|S,Z)}{P(Y|S,Z)} \tag{1.12}$$

We can rewrite this equation using the fact that the model makes no prediction on the values of Z, which is equivalent to $P(M_i|S,Z) = P(M_i|S)$. So we obtain:

$$P(M_i|Z,S,Y) = \frac{P(Y|M_i,S,Z)P(S|M_i)P(M_i)}{P(Y|S,Z)P(S)} \tag{1.13}$$

When we are evaluating models using the same set of data the relative likelihood of the models will depend on the three quantities in the numerator. In order, these quantities reflect: (1) the probability of the option prices given the model and the quantities entering the pricing formula, (2) the likelihood of the time series data given the model, and (3) the prior probability of the model. We will concentrate on the first two quantities in this paper. One can take flat priors over the models, ie. $P(M_i) = P(M_j)$, if one does not have a good feeling of which model is more likely. Once one has evaluated these three quantities for each probable model, assigning a probability to the model is straightforward. One does not have to directly evaluate the denominator in (1.13), but can instead evaluate the following:

$$P(M_i|Z,S,Y) = \frac{P(Y|M_i,S,Z)P(S|M_i)P(M_i)}{\sum_j P(Y|M_j,S,Z)P(S|M_j)P(M_j)} \tag{1.14}$$

With flat priors over the models:

$$P(M_i|Z,S,Y) = \frac{P(Y|M_i,S,Z)P(S|M_i)}{\sum_j P(Y|M_j,S,Z)P(S|M_j)} \tag{1.15}$$

Now all we need to do is evaluate the two quantities $P(Y|M_i,S,Z)$ and $P(S|M_i)$. This is what we do now.

1.5 Evaluating $P(Y|M,S,Z)P(S|M)$

Before proceeding, we will find in it convenient to define $\beta = \sigma^{-2}$; β may be referred to as the precision of the options data. The probability of Y given M, S, and Z can be calculated as: (we leave off the subscript i from M for now)

$$\begin{aligned} & P(Y|M,S,Z) \\ = & \int\int P(Y,\beta,\delta|M,S,Z)d\beta d\delta \\ = & \int\int P(Y|\beta,\delta,M,S,Z)P(\beta|\delta,M_i,S,Z)P(\delta|M,S,Z)d\beta d\delta \end{aligned} \tag{1.16}$$

We now examine each quantity entering the above integral. First, note that $P(\delta|M,S,Z) = P(\delta|M,S)$. This is a result of the fact that Z is independent of M, S, and δ. As a consequence:

$$P(\delta|M,S,Z) = P(\delta|M,S) = \frac{P(S|\delta,M)P(\delta|M)}{P(S|M)} \tag{1.17}$$

We next examine $P(\beta|\delta,M,Z,S)$. We assume that δ, Z, and S tell us nothing about β. Consequently: $P(\beta|\delta,M,S,Z) = P(\beta|M)$. We take our prior distribution over β to be a gamma distribution with parameters ν and η:

$$P(\beta|\delta,M_i,S,Z) = P(\beta|M_i) = \frac{(\nu/2)^{(\eta/2)}\beta^{(\eta/2-1)}\exp(-\nu\beta/2)}{\Gamma(\eta/2)} \tag{1.18}$$

The remaining element of the integrand to be examined is $P(Y|\beta,\delta,M,S,Z)$. We assume that the error term ϵ is a conditionally mean zero iid Gaussian. So we can write the following (likelihood function):

$$P(Y|\beta,\delta,M,S,Z) = (2\pi/\beta)^{(-N/2)}\exp(-\beta\frac{1}{2}\sum_{n=1}^{N}(y_n - f(s_n,z_n,\delta))^2) \qquad (1.19)$$

Now define:

$$E_M(\delta,\nu) \equiv \frac{1}{2}\left(\sum_{n=1}^{N}[y_n - f(s_n,z_n,\delta)]^2 + \nu\right) \qquad (1.20)$$

Given the above, we can write equation (1.16) as:

$$\begin{aligned} P(Y|M,S,Z) &= \frac{(\nu/2)^{(\eta/2)}}{(2\pi)^{(N/2)}\Gamma(\eta/2)}\int\int \beta^{(N/2+\eta/2-1)}\exp[-\beta E_M(\delta,\nu)]P(\delta|M,S)d\beta d\delta \\ &= \frac{(\nu/2)^{(\eta/2)}}{(2\pi)^{(N/2)}\Gamma(\eta/2)}\int\left(\int \beta^{(\frac{N+\eta}{2}-1)}\exp[-\beta E_M(\delta,\nu)]d\beta\right)P(\delta|M,S)d\delta \\ &= \frac{(\nu/2)^{(\eta/2)}}{(2\pi)^{(N/2)}\Gamma(\eta/2)}\Gamma(\frac{N+\eta}{2})\int E_M(\delta,\nu)^{-\frac{N+\eta}{2}}P(\delta|M,S)d\delta \qquad (1.21) \end{aligned}$$

Using (1.17), we obtain:

$$P(Y|M,S,Z)P(S|M) = \frac{(\nu/2)^{(\eta/2)}\Gamma(\frac{N+\eta}{2})}{(2\pi)^{(N/2)}\Gamma(\eta/2)}\int E_M(\delta,\nu)^{-\frac{N+\eta}{2}}P(S|\delta,M)P(\delta|M)d\delta \qquad (1.22)$$

The quantity $P(S|\delta,M)$ is the likelihood function of the time series data S. Lo [12] provides a general method to calculate this likelihood function which we briefly discuss here. Following Lo, we have assumed that the model M specifies dynamics for $s(t)$ such that $s(t)$ satisfies the following stochastic integral equation (this is repeated here for clarity):

$$s(T) = s(T_0)+\int_{T_0}^{T} a_M(s,t;\delta_0)dt+\int_{T_0}^{T} b_M(s,t;\delta_0)dW(t)+\int_{T_0}^{T} c_M(s,t;\delta_0)dN_{\delta_1}(t) \quad (1.23)$$

where $W(t)$ is a standard Wiener process and $N_{\delta_1}(t)$ is a standard Poisson jump process with jump intensity δ_1. Also recall we defined $\delta \equiv \{\delta_0,\delta_1\}$. The process $s(t)$ is a Markov process and:

$$P(S|\delta,M) = P(s_1,\tau_1|\delta,M)\prod_{k=2}^{K} P(s_k,\tau_k|s_{k-1},\tau_{k-1},\delta,M) \qquad (1.24)$$

We will take $P(s_1, \tau_1|\delta, M) = 1$ as a convenient normalization. This will not affect our model comparisons as long as we are dealing with the same set of data across all models. Lo shows that under certain sufficient conditions $\rho_k \equiv P(s_k, \tau_k|s_{k-1}, \tau_{k-1}, \delta, M)$ satisfies the following Kolmogorov forward equation:

$$\frac{\partial}{\partial t}[\rho_k] = \frac{\partial}{\partial s}[a_M \rho_k] + \frac{1}{2}\frac{\partial^2}{\partial s^2}[b_M \rho_k] - \delta_1 \rho_k + \delta_1 \tilde{\rho}_k \left| \frac{\partial}{\partial s}[\tilde{c}_M^{-1}] \right| \tag{1.25}$$

subject to:

$$\rho_k(s, \tau_{k-1}) = \Delta(s - s_{k-1}) \tag{1.26}$$

and any other relevant boundary conditions, where $\tilde{c}_M(s,t,\delta) = s + c_M(s,t,\delta)$ and $\tilde{\rho}_k \equiv \rho_k(\tilde{c}_M^{-1}, t)$ and $\Delta(s - s_{k-1})$ is the Dirac-delta distribution centered at s_{k-1}. By solving (1.25), one can subsequently evaluate $P(S|\delta, M)$ via (1.24).

Once we specify a prior $P(\delta|M)$, we can evaluate the integral in (1.22) numerically. The simplest method by which to do this is to simply draw random samples of δ from the prior distribution and to average $E_M(\delta, \nu)^{-\frac{N+n}{2}}$ across these draws. Unfortunately, an analytic solution to this integral will not generally be feasible. To obtain a better understanding of what $P(Y|M, S, Z)P(S|M)$ captures, we will now turn to an analytic approximation. This approximation reveals some very interesting features of what the Bayesian method looks for in an option pricing model.

1.6 Interpreting $P(Y|M, S, Z)P(S|M)$ via an Approximation

Because an analytic solution is not available, an approximation to $P(Y|M, S, Z) \times P(S|M)$ will be crucial if we are to understand what the Bayesian procedure captures. For this reason, this section is perhaps the most important. We will approximate $P(Y|M, S, Z) \times P(S|M)$ analytically in terms of two well known estimates of δ. One can estimate δ by using the time series stock price data. This can be obtained by maximizing $P(S|\delta, M)$ over δ. We define this maximum likelihood time series estimate as:

$$\hat{\delta}_S \equiv \text{argmax}_\delta P(S|\delta, S, M) \tag{1.27}$$

where the subscript S notes the fact that $\hat{\delta}_S$ is estimated using S. The standard error of $\hat{\delta}_S$ can be estimated as:

$$\hat{V}_{\hat{\delta}_S} = \left(\frac{\partial^2 \ln P(S|\delta, S, M)}{\partial\delta\partial\delta'}\right)^{-1}_{\delta=\hat{\delta}_S} \tag{1.28}$$

One can also estimate δ by using the "option implied" value of δ. In the case of the Black–Scholes model, for example, one could estimate the "option implied" volatility parameter. One can thus obtain an estimate of $\hat{\delta}_Y$ by performing the non-linear optimization:

$$\hat{\delta}_Y = \text{argmin}_\delta \sum_{n=1}^{N} [y_n - f(s_n, z_n, \delta)]^2 \tag{1.29}$$

Since we assumed that $y_n = f(s_n, z_n, \delta) + \epsilon_n$ and that ϵ_n are iid Gaussian, $\hat{\delta}_Y$ will also be the MLE estimate of δ given only the options data. We will later also be interested in the sampling theory standard error estimate for $\hat{\delta}_Y$ which can be written as:

$$\hat{V}^2_{\hat{\delta}_Y} = \hat{\sigma}^2_{NLS} \left(E''_M(\hat{\delta}_Y)\right)^{-1} \tag{1.30}$$

where:

$$\hat{\sigma}^2_{NLS} = \frac{1}{N} \sum_{n=1}^{N} (y_n - f(s_n, z_n, \hat{\delta}_Y))^2 \tag{1.31}$$

To summarize the two sampling theory estimates of δ:

$$\begin{aligned} \hat{\delta}_S &= \text{Estimate of } \delta \text{ based on the Stock Time Series Data} \\ \hat{V}_{\hat{\delta}_S} &= \text{Standard Error of } \hat{\delta}_S \\ \hat{\delta}_Y &= \text{Estimate of } \delta \text{ implied by Option Prices} \\ \hat{V}_{\hat{\delta}_Y} &= \text{Standard Error of } \hat{\delta}_Y \end{aligned} \tag{1.32}$$

These sampling theory estimates will help us interpret the approximation of $P(Y|M, S, Z)P(S|M)$ which we derive next. It turns out that for purposes of approximation and interpretation, it is actually easier to analyze $P(Y|M, S, Z)$ and $P(S|M)$ separately.[3]

We analyze $P(Y|M, S, Z)$ first using (1.21). To evaluate the integral in the last line of (1.21), we will take a large sample approximation of $P(\delta|S, M)$ and also take a second order Taylor expansion of $E_M(\delta)^{-\frac{N+\eta}{2}}$. To evaluate $P(\delta|S, M)$, we take advantage of a standard Bayesian large sample approximation that says that the posterior distribution of a parameter given a sufficiently large data set is approximately Gaussian (cf. Zellner [22] p.31). This Gaussian will have mean equal to the maximum likelihood estimate of δ when it is estimated using the time series data S. The variance of the Gaussian will approximately equal the sampling theory variance estimate $\hat{V}_{\hat{\delta}_S}$. (Of course the interpretation of the mean and variance here are very different from the sampling theory view.) This approximation relies primarily on the sample size (the number of observations in S) being large—it does not necessarily assume that the prior over δ is flat.

For simplicity, we consider the case where δ is one dimensional. In this case, by the preceding discussion:

$$P(\delta|M, S) \approx (2\pi \hat{V}_{\hat{\delta}_S}^2)^{-1/2} \exp(-\frac{1}{2\hat{V}_{\hat{\delta}_S}^2}(\delta - \hat{\delta}_S)^2) \qquad (1.33)$$

where $\hat{\delta}_S$ is the maximum likelihood estimate of δ given the time series data S, and $\hat{V}_{\hat{\delta}_S}^2$ is the sampling theory variance estimate for the MLE estimator $\hat{\delta}_S$.

We approximate $E_M(\delta, \delta_0)^{-\frac{N+\eta}{2}}$ by taking its second order Taylor expansion around $\hat{\delta}_Y$.

$$\begin{aligned} E_M(\delta)^{-\frac{N+\eta}{2}} \equiv h(\delta) &\approx h(\hat{\delta}_Y) + (\delta - \hat{\delta})h'(\hat{\delta}_Y) + \frac{1}{2}(\delta - \hat{\delta}_Y)^2 h''(\hat{\delta}_Y) \\ &= h(\hat{\delta}_Y) + \frac{1}{2}(\delta - \hat{\delta}_Y)^2 h''(\hat{\delta}_Y) \end{aligned} \qquad (1.34)$$

[3]The reason for this is that it is easier to analyze $P(\delta|S, M)$ rather than $P(S|\delta, M)P(\delta|M)$.

where the second line follows because:

$$\begin{aligned}\hat{\delta}_Y &= \text{argmin}_\delta \sum_{n=1}^{N} [y_n - f(s_n, z_n, \delta)]^2 \\ &= \text{argmin}_\delta E_M(\delta, \nu) \\ &= \text{argmax}_\delta h(\delta) \qquad (1.35)\end{aligned}$$

So that:

$$P(Y|M,S,Z) \approx \frac{(\nu/2)^{(\eta/2)}\Gamma(\frac{N+\eta}{2})}{(2\pi)^{(N/2)}\Gamma(\eta/2)} \int \left(h(\hat{\delta}_Y) + \frac{1}{2} h''(\hat{\delta}_Y)(\delta - \hat{\delta}_Y)^2 \right) P(\delta|S,M) d\delta \qquad (1.36)$$

Using the large sample approximation of $P(\delta|M,S)$ given in (1.33), the above quantity can be shown to equal:

$$P(Y|M,S,Z) \approx \frac{(\nu/2)^{(\eta/2)}\Gamma(\frac{N+\eta}{2})}{(2\pi)^{(N/2)}\Gamma(\eta/2)} \left[h(\hat{\delta}) + \frac{1}{2} h''(\hat{\delta}_Y) \left[(\hat{\delta}_Y - \hat{\delta}_S)^2 + \hat{V}^2_{\hat{\delta}_S} \right] \right] \qquad (1.37)$$

From the definition of $h(\delta)$ and $\hat{\delta}_Y$, we know:

$$\begin{aligned} h(\hat{\delta}_Y) &= E_M(\hat{\delta}_Y)^{-\frac{N+\eta}{2}} \\ h''(\hat{\delta}_Y) &= -\left(\frac{N+\eta}{2}\right) E_M(\hat{\delta}_Y)^{-\frac{N+\eta}{2}-1} E''_M(\hat{\delta}_Y) \qquad (1.38)\end{aligned}$$

The expression for $h''(\hat{\delta}_Y)$ can be simplified by noting:

$$\begin{aligned}\left(\frac{N+\eta}{2}\right) E_M(\hat{\delta}_Y)^{-1} &= \left(\left[\frac{2}{N+\eta}\right] \left(\frac{1}{2}\right) \left[\sum_{n=1}^{N} [y_n - f(s_n, z_n, \hat{\delta}_Y)]^2 + \nu \right] \right)^{-1} \\ &= \left(\left[\frac{N}{N+\eta}\right] \left(\frac{1}{N} \sum_{n=1}^{N} [y_n - f(s_n, z_n, \hat{\delta}_Y)]^2 \right) + \frac{\nu}{N+\eta} \right)^{-1} \\ &= \left(\left[\frac{N}{N+\eta}\right] \hat{\sigma}^2_{NLS} + \frac{\nu}{N+\eta} \right)^{-1} \\ &\approx \left(\hat{\sigma}^2_{NLS}\right)^{-1} \qquad (1.39)\end{aligned}$$

where the last approximation holds when N is large. Thus:

$$\begin{aligned} h''(\hat{\delta}_Y) &\approx -E_M(\hat{\delta}_Y)^{-\frac{N+\eta}{2}} \left(\left(E''_M(\hat{\delta}_Y) \right)^{-1} \hat{\sigma}^2_{NLS} \right)^{-1} \\ &= -E_M(\hat{\delta}_Y)^{-\frac{N+\eta}{2}} \left(\hat{V}^2_{\hat{\delta}_Y} \right)^{-1} \end{aligned} \tag{1.40}$$

Using the previous approximations, we can write (1.37) as:

$$P(Y|M,S,Z) \approx c(N,\eta,\nu) \left(E_M(\hat{\delta}_Y) \right)^{-\frac{N+\eta}{2}} \left(1 - \frac{(\hat{\delta}_Y - \hat{\delta}_S)^2}{2\hat{V}^2_{\hat{\delta}_Y}} - \frac{\hat{V}^2_{\hat{\delta}_S}}{2\hat{V}^2_{\hat{\delta}_Y}} \right) \tag{1.41}$$

where:

$$c(N,\eta,\nu) \equiv \frac{(\nu/2)^{(\eta/2)}\Gamma(\frac{N+\eta}{2})}{(2\pi)^{(N/2)}\Gamma(\eta/2)} \tag{1.42}$$

This approximation of $P(Y|M,S,Z)$ has a nice interpretation. The term $\left(E_M(\hat{\delta}_Y) \right)^{-\frac{N+\eta}{2}}$ measures how well the model could fit the options data if δ were free to take any value. This follows since:

$$\begin{aligned} \left(E_M(\hat{\delta}_Y) \right)^{-\frac{N+\eta}{2}} &= \max_\delta \ (E_M(\delta))^{-\frac{N+\eta}{2}} \\ &= \max_\delta \left(\frac{1}{2} \right)^{-\frac{N+\eta}{2}} \left(\sum_{n=1}^{N} [y_n - f(s_n, z_n, \delta)]^2 + \nu \right)^{-\frac{N+\eta}{2}} \\ &= \left(\frac{1}{2} \right)^{-\frac{N+\eta}{2}} \left(\min_\delta \sum_{n=1}^{N} [y_n - f(s_n, z_n, \delta)]^2 + \nu \right)^{-\frac{N+\eta}{2}} \end{aligned} \tag{1.43}$$

The second piece in the last large parenthesis of (1.41) reflects the tension between the δ implied by the time series data and the δ implied by the option prices. The larger is the difference between these two estimates, normalized by $\hat{V}^2_{\hat{\delta}_Y}$, the smaller is $P(Y|M,S,Z)$.

The third piece in the last large parenthesis also captures an Occam factor which penalizes the model based on its complexity. By not making precise predictions before the data arrives, the model becomes less likely. However, if the model does not need to be finely tuned, or in other words $\hat{V}^2_{\hat{\delta}_Y}$ is large, the penalty is lessened. The ratio

$\frac{\hat{V}^2_{\hat{\delta}_S}}{\hat{V}^2_{\hat{\delta}_Y}}$ is an approximate measure of the ratio of the prior accessible volume for δ after the stock data arrived but before the options data arrive, and the posterior accessible volume for δ. A model is rewarded for starting with a small prior accessible volume. If in contrast $\hat{V}^2_{\hat{\delta}_S}$ is large, the combination of the model and time series data do not make very precise predictions about the option price data. If, however, many values of δ would have done as good a job fitting the data, as measured by the flatness of the likelihood at $\hat{\delta}_Y$ or equivalently the size of $\hat{V}^2_{\hat{\delta}_Y}$, the penalty is lessened. We should note that this approximation should not be used in place of the actual numerical evaluation of $P(Y|M,S,Z)P(S|M)$. Namely, it is possible that this approximation of a probability might be negative.

To approximate $P(S|M)$ we will follow MacKay [14] and rely on the previous approximation that $P(\delta|M,S)$ is nearly Gaussian with mean $\hat{\delta}_S$ and variance $\hat{V}^2_{\hat{\delta}_S}$. Now we recall that for all δ:

$$P(\delta|M,S) = \frac{P(S|\delta,M)P(\delta|M)}{P(S|M)} \tag{1.44}$$

So that by rewriting:

$$P(S|M) = \frac{P(S|\delta,M)P(\delta|M)}{P(\delta|M,S)} \tag{1.45}$$

for all δ, so that a fortiori:

$$P(S|M) = \frac{P(S|\hat{\delta}_S,M)P(\hat{\delta}_S|M)}{P(\hat{\delta}_S|M,S)} \tag{1.46}$$

However, by using the Gaussian approximation of $P(\delta|M,S)$ given in (1.33),

$$P(\hat{\delta}_S|M,S) \approx (2\pi\hat{V}^2_{\hat{\delta}_S})^{-\frac{1}{2}} \tag{1.47}$$

So that:

$$P(S|M) \approx P(S|\hat{\delta}_S,M)P(\hat{\delta}_S|M)(2\pi\hat{V}^2_{\hat{\delta}_S})^{\frac{1}{2}} \tag{1.48}$$

The first term $P(S|\hat{\delta}_S,M)$ is the maximized likelihood function and is a measure of how well the model could fit the data if δ were free to take any value. The last two terms represent the Occam factor. If $P(\delta|M)$ is relatively flat, $\frac{1}{P(\delta|M)}$ is a measure of the prior accessible volume for δ. On the other hand, $\hat{V}^2_{\hat{\delta}_S}$ is a measure of the posterior accessible volume for δ. Thus:

$$\frac{(2\pi\hat{V}^2_{\hat{\delta}_S})^{\frac{1}{2}}}{\frac{1}{P(\hat{\delta}_S|M)}} \tag{1.49}$$

is an approximate measure of the ratio of posterior to prior accessible volume for δ. If a model specified a prior that was not very precise, in other words $P(\hat{\delta}_S|M)$ is small, the model is penalized. If however, many values of δ other than $\hat{\delta}_S$ could have fit the data just as well, as measured by the flatness of the likelihood at $\hat{\delta}_S$ or equivalently the size of $\hat{V}^2_{\hat{\delta}_S}$, the smaller the penalty.

1.7 Three Option Pricing Models

In this section, I briefly layout the three option pricing models that we will compare in the next section using the Bayesian methodology: (1)the Black–Scholes model, (2)Merton's jump-diffusion model and (3)Cox's constant elasticity of volatility model. In our formulation, an option pricing model must specify three components: stock price dynamics, an option pricing formula and a prior distribution on the relevant free parameters. I leave the specification of the prior distribution until after we describe the data set we will be using.

1.7.1 The Black–Scholes Model

Stock Price Dynamics: The Black–Scholes model [1], $(M = BS)$, specifies the following stock price dynamics under the probability measure P:

$$s(T) = s(T_0) + \int_{T_0}^{T} \mu_{bs} s(t)dt + \int_{T_0}^{T} \sigma_{bs} s(t)dW(t) \tag{1.50}$$

where as before $W(t)$ is a standard Wiener process. The likelihood function for the time series data can be written:

$$P(S|\mu_{bs}, \sigma_{bs}, M = BS) =$$

$$P(s_1, \tau_1 | \mu_{bs}, \sigma_{bs}, M = BS) \prod_{k=2}^{K} P(s_k, \tau_k | s_{k-1}, \tau_{k-1}, \mu_{bs}, \sigma_{bs}, M = BS) \quad (1.51)$$

where $P(s_k, \tau_k | s_{k-1}, \tau_{k-1}, \mu_{bs}, \sigma_{bs}, M = BS)$ equals

$$\frac{1}{s_k(2\pi\sigma_{bs}^2 \Delta\tau_k)^{1/2}} \exp(-\frac{(Y_k - Y_{k-1} - (\mu_{bs} - \sigma_{bs}^2/2)\Delta\tau_k)^2}{2\sigma_{bs}^2 \Delta\tau_k}) \quad (1.52)$$

and $Y_k = \ln s_k$.

Option Pricing Formula: In the context of pricing options with relatively short maturities, people have found it convenient to assume that the spot interest rate remains roughly constant for the remaining life of the option. All models we examine here make this assumption. Further, we deal exclusively with European put and call options in our analysis. To simplify our exposition, we define ζ to equal 1 if the option is a call and -1 if it is a put. This will allow us to write down one formula for both calls and puts.

In the Black Scholes model, the time T_0 price of a European option is:

$$\zeta S e^{-d(T-T_0)} \Phi(\zeta d1) - \zeta K e^{-r(T-T_0)} \Phi(\zeta d2) \quad (1.53)$$

where d is the (constant) flow dividend rate of the underlying asset, $\Phi(.)$ is the gaussian cumulative distribution function, and:

$$\begin{aligned} d1 &= \frac{\ln(S/K) + (r - d + \sigma_{bs}^2/2)(T - T_0)}{\sigma_{bs}(T - T_0)^{1/2}} \\ d2 &= d1 - \sigma_{bs}(T - T_0)^{1/2} \end{aligned} \quad (1.54)$$

1.7.2 The Merton Jump Diffusion Model

Stock Price Dynamics: The Merton jump-diffusion model [30], $(M = JD)$, captures the possibility that the stock price process takes discrete jumps between times when the the process is evolving continuously. The specified dynamics for s are:

$$s(T) = s(T_0) + \int_{T_0}^{T} (\mu_{jd} - \lambda\bar{\kappa})s(t)dt + \int_{T_0}^{T} \sigma_{jd}s(t)dW(t) + \int_{T_0}^{T} \kappa_t s(t)dN_\lambda(t) \quad (1.55)$$

where N_λ is a poisson process with jump intensity λ and unit increments. For each poisson event, κ is distributed as an independent log-normal random variable with mean $\bar{\kappa}$. Specifically:

$$\ln(1+\kappa) \sim \mathrm{N}(a - \xi^2/2, \xi^2) \quad (1.56)$$

so that $\bar{\kappa} \equiv \mathrm{E}[\kappa] = e^a - 1$. This captures the possibility that stock prices take discontinuous jumps with a random magnitude. The rate of return on the stock is μ_m since the process $-\lambda\bar{\kappa}s(t)dt + \kappa s(t)dN_\lambda(t)$ and the natural filtration form a martingale. The likelihood function $P(S|\mu, \sigma, M = JD)$ can be written (See Rosenfeld for a derivation):

$$\begin{aligned} &P(S|\mu_{jd}, \sigma_{jd}, \lambda, \kappa, M = JD) = \\ &\quad P(s_1, \tau_1|\mu_{jd}, \sigma_{jd}, \lambda, \kappa, M = JD) \\ &\quad \times \prod_{k=2}^{K} P(s_k, \tau_k|s_{k-1}, \tau_{k-1}, \mu_{jd}, \sigma_{jd}, \lambda, \kappa, M = JD) \end{aligned} \quad (1.57)$$

where $P(s_k, \tau_k|s_{k-1}, \tau_{k-1}, \mu_{jd}, \sigma_{jd}, \lambda, \kappa, M = JD)$ equals:

$$\begin{aligned} &\sum_{j=0}^{\infty} P(j|\lambda, \Delta\tau_k)\frac{1}{(2\pi[\sigma_{jd}^2\Delta\tau_k + \xi^2 j])^{1/2}s_k} \\ &\times \exp\left(-\frac{\left(Y_k - Y_{k-1} - (\mu_{jd} - \lambda\bar{\kappa} - \frac{\sigma_{jd}^2}{2})\Delta\tau_k - (a - \frac{\xi^2}{2})j\right)^2}{2(\sigma_{jd}^2\Delta\tau_k + \xi^2 j)}\right) \end{aligned} \quad (1.58)$$

where $Y_k = \ln s_k$ and:

$$P(j|\lambda, \Delta\tau_k) = \frac{e^{-\lambda\Delta\tau_k}(\lambda\Delta\tau_k)^j}{j!} \quad (1.59)$$

Option Pricing Formula: In the Merton jump–diffusion model, the time T_0 price of a

European option ($\zeta = 1$ for call $\zeta = -1$ for put) maturing at time T is:

$$e^{-r(T-T_0)} \sum_{n=0}^{\infty} P(n|\lambda, T-T_0) \left(\zeta S e^{b_n(T-T_0)} \Phi(\zeta d1_n) - \zeta K \Phi(\zeta d2_n)\right) \tag{1.60}$$

where d is the flow dividend rate of the underlying asset and:

$$\begin{aligned} d1_n &= \frac{\ln(S/K) + b_n(T-T_0) + (\sigma^2(T-T_0) + n\xi^2)/2}{(\sigma^2(T-T_0) + n\xi^2)^{1/2}} \\ d2_n &= d1_n - (\sigma^2(T-T_0) + n\xi^2)^{1/2} \\ b_n &= r - d - \lambda\bar{\kappa} + n\ln(1+\bar{\kappa}) \end{aligned} \tag{1.61}$$

1.7.3 The Cox CEV Model

Stock Price Dynamics: The Cox constant elasticity of volatility model [3], $M = CEV$, captures the possibility that volatility will decrease (increase) by a constant percentage as the asset price increases by a percent. Cox's original model captures the possibility of a negative relationship between asset prices and volatility while Emanuel and MacBeth [4] considered the reverse situation. We incorporate both possibilities below. The specified dynamics are:

$$s(T) = s(T_0) + \int_{T_0}^{T} \mu_{cev} s(t)dt + \int_{T_0}^{T} \sigma_{cev} s(t)^{\gamma/2} dW(t) \tag{1.62}$$

The likelihood function $P(S|\mu_{cev}, \sigma_{cev}, \gamma, M = CEV)$ can be written:

$$\begin{aligned} &P(S|\mu_{cev}, \sigma_{cev}, \gamma, M = CEV) = \\ &\quad P(s_1, \tau_1|\mu_{cev}, \sigma_{cev}, \gamma, M = CEV) \\ &\quad \times \prod_{k=2}^{K} P(s_k, \tau_k|s_{k-1}, \tau_{k-1}, \mu_{cev}, \sigma_{cev}, \gamma, M = CEV) \end{aligned} \tag{1.63}$$

where $P(s_k, \tau_k|s_{k-1}, \tau_{k-1}, \mu_{cev}, \sigma_{cev}, \gamma, M = CEV)$ equals:

$$|2-\gamma|\mathcal{K}^{1/(2-\gamma)}(xw^{1-2\gamma})^{1/(4-2\gamma)} e^{-x-w} I_{1/|2-\gamma|}(2\sqrt{xw}) \tag{1.64}$$

where $I_q(.)$ is the modified Bessel function of the first kind of order q and:

$$\begin{aligned} \mathcal{K} &= \frac{2\mu_{cev}}{\sigma_{cev}^2(2-\gamma)[e^{\mu_{cev}(2-\gamma)\Delta\tau_k}-1]} \\ x &= \mathcal{K}(s_{k-1})^{2-\gamma}e^{\mu_{cev}(2-\gamma)\Delta\tau_k} \\ w &= \mathcal{K}(s_{k-1})^{2-\gamma} \end{aligned} \tag{1.65}$$

Option Pricing Formula: In the Cox CEV model, the time T_0 price of a European option ($\zeta = 1$ for call $\zeta = -1$ for put) maturing at time T is:

$$\zeta S e^{-d(T-T_0)}\left(\frac{1-\zeta}{2}+\zeta Q_1\right) - \zeta K e^{-r(T-T_0)}\left(\frac{1+\zeta}{2}-\zeta Q_2\right) \tag{1.66}$$

where d is the flow dividend rate of the asset, and for the case $\gamma < 2$:

$$\begin{aligned} Q_1 &= Q(2y; 2+2/(2-\gamma), 2x) \\ Q_2 &= Q(2x; 2/(2-\gamma), 2y) \end{aligned} \tag{1.67}$$

where Q() is the complementary non-central chi-squared distribution function[4] where:

$$y = \mathcal{K}K^{2-\gamma} \tag{1.68}$$

For the case $\gamma > 2$:

$$\begin{aligned} Q_1 &= Q(2x, 2/(\gamma-2), 2y) \\ Q_2 &= Q(2y, 2+2/(\gamma-2), 2x) \end{aligned} \tag{1.69}$$

For the case, $\gamma = 2$, this model reduces to the Black Schol·s model.

[4]See Schroder [19] for a derivation and efficient computational algorithm of this distribution. He also presents an approximation which is very accurate for the range where the exact technique is slow to converge. The approximation is that developed by Sankaran [18]. This approximation is also very useful and extremely accurate for computing the transition density when the time between observations is small and when γ is close to 2.

1.8 Data Source

I demonstrate the Bayesian methodology of selecting option pricing models using data on foreign exchange options traded on the Philadelphia Stock Exchange. I consider the three option pricing models laid out in the previous section; namely, the Black–Scholes model, Cox's constant elasticity of volatility model and Merton's jump–diffusion model. I lay out the data source here in this section and then discuss the formulation of priors in the next section. The section after the next one will summarize the empirical results.

The options data that I use are on European style Deutschemark put and call options that were traded between January and February of 1993. There are 1817 observations of option trades during this time: 1097 puts and 720 calls. The options are traded 24 hours a day during weekdays, but are not traded over the weekend. Listed with each option is the time of the trade (up to the minute), the Telerate quote of the underlying currency (DM) at the time of the trade, the strike price of the option and the maturity of the option.[5]

Summary statistics of moneyness and term to maturity are reported in Table 1.1 on page 39 and Table 1.2 on page 39. One sees that puts were more heavily traded during this period and that most of the options traded were either at the money or out of the money. This is typical since deep in the money options are relatively expensive.

1.9 Priors Over Parameters

One must put priors over the various parameters specified in a model. In particular, one must specify $P(\delta|M_i)$ and $P(\beta|M_i)$ where we recall that δ is the vector of parameters that specify the dynamics of the underlying process and β is the "precision" of the options data. Of course one could simply specify any distribution for the priors

[5]The U.S. interest rate is determined by finding the T-Bill rate with the maturity closest to that of the option while the U.S.–German interest rate differential is set to the differential in the Eurocurrency market.

Table 1.1: Moneyness of Puts and Calls

S/K	< 0.90	0.90-0.94	0.94-0.98	0.98-1.02	1.02-1.06	1.06-1.10	> 1.10
puts	9	30	178	540	281	52	7
1097	(1%)	(3%)	(16%)	(49%)	(26%)	(5%)	(1%)
calls	10	67	208	354	55	19	7
720	(1%)	(9%)	(28%)	(49%)	(8%)	(3%)	(1%)

The ratio S/K is reported separately for puts and calls with the percentage of puts or calls belonging to each range reported underneath.

Table 1.2: Term to Maturity of Puts and Calls

Months to Maturity	0-1	1-2	2-3	4-5	5-6	6+
puts	394	241	133	47	134	138
1097	(36%)	(22%)	(12%)	(4%)	(12%)	(13%)
calls	314	180	51	29	78	68
720	(44%)	(25%)	(7%)	(4%)	(11%)	(9%)

The term to maturity is reported separately for puts and calls with the percentage of puts or calls in each range reported below.

at this point and simply argue that these are their subjective beliefs. I will in this section present priors that I believe are realistic for each model, but at the same time try to specify priors for each model that are in some way comparable to each other.

1.9.1 Prior over the precision of the options data

We have already assumed in our earlier development that β has a gamma distribution with parameters ν and η. The notion that the observed trade price deviates from the "true" theoretical price comes from a belief that market imperfections cause the trade to take place at prices deviating slightly from the true value. We assume that the models do not differ in their assessment of the sizes of such market imperfections. This will allow us to assume one set of ν and η for all models. As long as these priors are not too dogmatic, the resulting probability of a model will not be strongly influenced by slight changes in the priors. To nail down a reasonable prior, note that because the error term is additive, we must have some idea of the scale in which option prices are quoted. A typical option is quoted in the range of 1 dollar. A deep in the money option can trade as high as 10 dollars, and out of the moneys can trade as low as a few cents. It seems reasonable to assume that a typical mispricing will be on the order of magnitude of about 5 to 40 cents. As shown in Hamilton (p.355-357), if one starts with a diffuse prior and observes one mispricing of 10 cents, then the posterior would yield a $\eta = 1$ corresponding to the number of observations and a $\nu = 0.01$ (dollars squared) corresponding to the sum of squared residuals of the sample. We display the plot of the distribution this implies for the regression standard error σ_{reg} in Figure 1-1 on page 41. If one thinks of this prior as coming from one observation, one quickly sees that this distribution is not dogmatic at all–i.e. not based a large previous sample.

1.9.2 Prior over the drift

In stating priors over the parameters relating to the underlying price dynamics $P(\delta|M_i)$, we will state our priors so that the mean and variance of the process are at least com-

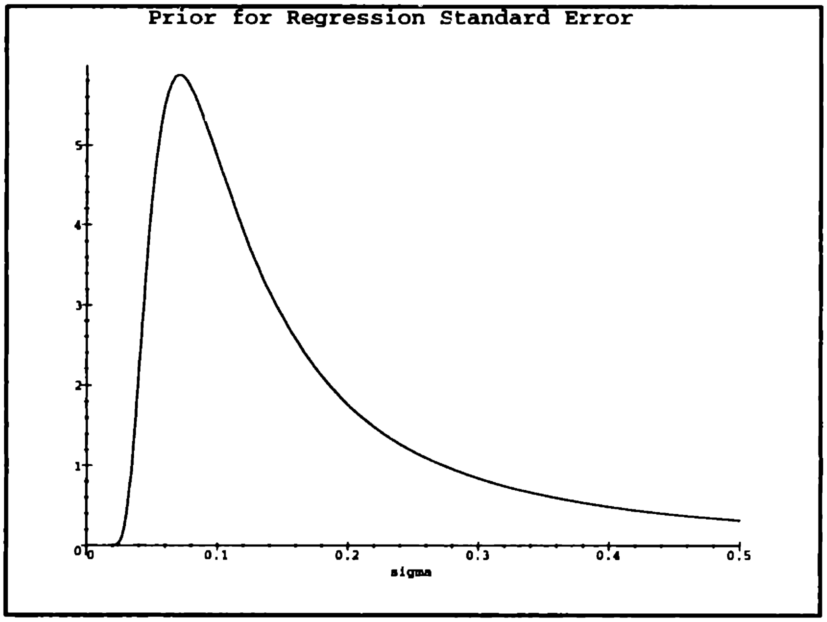

Figure 1-1: Prior for Regression Standard Error

parable. In fact, we will take the distribution of μ to be the same across all models and independent of the other parameters. The distribution for μ for all models is:

$$\mu \sim \text{Tr}(\mu_l, \mu_m, \mu_h) \tag{1.70}$$

where $\text{Tr}(l, m, h)$ denotes a distribution that has a triangular density function such that the support of the distribution is $[l, h]$ and the peak of the density is at m. See Figure 1-2 on page 42. This density will be very useful because it has a finite support, is easy to picture mentally and easy to simulate draws from. We take μ_m to equal the interest differential between the two countries at the beginning of the sample period, $r_0^{US} - r_0^{GER}$, which in our sample is -0.0561. This is motivated by the theoretical notion of uncovered interest parity. At the same time, the empirical evidence on uncovered interest parity does not support the theory very well so we let the support of the distribution to be fairly wide. For our benchmark, we take $\mu_l = \mu_m - 0.07$ and $\mu_h = \mu_m + 0.07$.

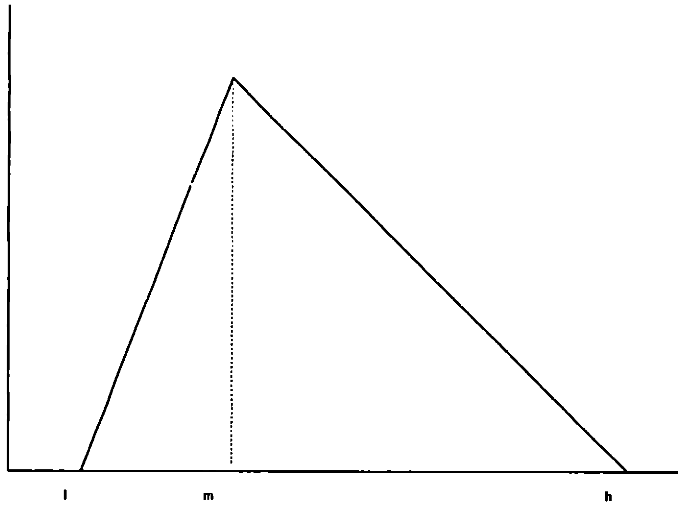

Figure 1-2: Triangular Density

1.9.3 Prior over the Black–Scholes parameter

The only parameter in the Black–Scholes model for which we have not yet specified a prior is the diffusion volatility parameter σ_{bs}. As a benchmark, we take:

$$\sigma_{bs}|(M = BS) \sim \text{Tr}(0.08, 0.14, 0.20) \tag{1.71}$$

This volatility is representative of the volatility estimates found in historical data by Jorion [19] and also in Baily and Kropywiansky. Baily and Kropywiansky found that volatility in any given month from 1984 to 1992 always fell in this range and that extremely high or low volatility in this range was fairly rare.

1.9.4 Prior over the Jump–Diffusion parameters

Defining:

$$v_{\text{jump}}^2 = \lambda[(\ln(1+\bar{\kappa}) - \xi^2/2)^2 + \xi^2], \tag{1.72}$$

the variance of $\ln(S_{t+T}/S_t)$, given that the underlying dynamics are given by a jump-diffusion process can be written as:

$$v^2T = \{\sigma^2 + v_{\text{jump}}^2\}T \tag{1.73}$$

So the instantaneous standard deviation of the returns process is given by v. To make the prior for the jump–diffusion model comparable to the Black–Scholes prior, we will impose the same prior on v in the jump–diffusion model as we imposed on σ in the Black–Scholes model. Thus we take:

$$v|(M = JD) \sim \text{Tr}(0.08, 0.14, 0.20) \tag{1.74}$$

Now that we have imposed the above prior on v, the full distribution I will use can be most easily stated by giving the following algorithm for drawing a sample set of parameters from this prior distribution. We will first draw v from the above specification and then draw λ, $\overline{\kappa}$ and ξ from three independent triangular distributions.

$$\begin{aligned} \lambda &\sim \text{Tr}(0.1, 1, 5) \\ \overline{\kappa} &\sim \text{Tr}(-0.05, 0, 0.05) \\ \xi &\sim \text{Tr}(0, 0.01, 0.025) \end{aligned} \tag{1.75}$$

If the following condition

$$v^2_{\text{jump}}/v^2 \leq 0.4 \tag{1.76}$$

is satisfied, we will keep the observation. If, however, the condition is not satisfied we will draw another set in the same way and keep repeating this procedure until a set of parameters satisfies this criteria. This will ensure that the variance due to the jump component is less than 40% of the total variance of the process. This is imposed so that we do not end up with a process that is almost a pure jump process with very little diffusion component. Once an acceptable set is found, σ_{jd} is set so that:

$$v^2 = v^2_{\text{jump}} + \sigma^2_{jd} \tag{1.77}$$

This prior ensures that the final distribution for v is the same as for σ in the Black–Scholes model while also allowing for a fairly straightforward interpretation of the prior over the other parameters. In this prior, there will be a negative correlation between each of λ, $\overline{\kappa}$, ξ, and σ_{jd}. For example, when the probability of a jump λ is high, it is less likely for the jump to be big: $\overline{\kappa}$ and ξ will be small.

1.9.5 Prior over the CEV parameters

The pure diffusion part of the CEV process can be written:

$$\sigma_{cev} S_t^{\gamma/2} dW = VOL_{cev}(S_t) S_t dW \tag{1.78}$$

where $VOL_{cev}(S_t)$ is the instantaneous volatility of the returns process when the stock price equals s and can be written:

$$VOL_{cev}(S_t) = \sigma_{cev} S_t^{\gamma/2-1} \tag{1.79}$$

To match the prior for the CEV model with that for the Black–Scholes model, I take the prior over $VOL_{cev}(S_0)$ to be the same as the prior over σ_{bs} in the Black–Scholes model.

$$VOL_{cev}(S_0)|(M = CEV) \sim \text{Tr}(0.08, 0.14, 0.20) \tag{1.80}$$

To complete the prior, I will specify priors over γ. This in combination with the above will generate a joint prior over δ and γ. The elasticity of the returns volatility is equal to:

$$\frac{\partial VOL_{cev}(S_t)}{\partial S_t} \frac{S_t}{VOL_{cev}(S_t)} = \gamma/2 - 1 \tag{1.81}$$

So that $\gamma = 0$ corresponds to an elasticity of -1 while $\gamma = 4$ corresponds to an elasticity of 1. I will take as a benchmark a triangular prior density over the elasticity with support from -1 to 1 and a peak at 0. This is equivalent to the following triangular prior density over γ:

$$\gamma \sim \text{Tr}(0, 2, 4) \tag{1.82}$$

1.10 Empirical Results

Once we have specified our priors for the various parameters in the different models, we are ready to impute probabilities to each of the models based on a given data set. Before proceeding, however, it is worthwhile to step back and examine the general set up we have given for comparing option pricing models.

1.10.1 A helpful perspective to interpret the results

An option pricing model essentially is composed of two components. First it specifies the dynamics for the underlying asset. Second, it specifies an option pricing formula. Although we have stated that one should consider both these specifications in analyzing an option pricing model, it is informative to consider what would happen if an option pricing model only specified one or the other. For each model, we can specify a "time series only" version or an "option pricing only" version. We will denote the version of the model that specifies only the dynamics of the underlying asset by superscripting the model with 'TS'. For example, the version of the Black–Scholes model that specifies only the underlying asset dynamics will be referred to as BS^{TS} while a generic time series model will be referred to as M_i^{TS}. The version that specifies only the option pricing will be superscripted with 'OP'–i.e. BS^{OP} or M_i^{OP}. The full model that specifies both time series dynamics for the underlying as well as the option pricing function will be referred to without any subscript–i.e. BS or M_i. We will assume that the "time series only" model and the "option pricing only" model agree with the complete model on the specification of priors for the parameters in the models.

We assume that of the 'time series only' models, one of our specified models BS^{TS}, JD^{TS} or CEV^{TS} is the correct model. Similarly, of the 'option pricing only' models we assume that one of the ones we specified BS^{OP}, JD^{OP} or CEV^{OP} is the correct model. As before, we also assume that of the complete models we specify BS, JD or CEV is the correct model. Although we could easily allow for different prior probabilities of the models being correct, we will for simplicity assume that within each class of models, each one is equally likely. Thus the probability of any model being true *a priori* is 1/3.

Given the above assumption and by using Bayes rule, one can show that:

$$P(M_i^{TS}|S,Y,Z) = \frac{P(S|M_i^{TS})}{\sum_{j=1}^{3} P(S|M_j^{TS})} \tag{1.83}$$

and that:

$$P(M_i^{OP}|S,Y,Z) = \frac{P(Y|M_i^{OP},S,Z)}{\sum_{j=1}^{3} P(Y|M_j^{OP},S,Z)} \tag{1.84}$$

As show in Section 1.4, one can calculate the probability of a complete model being the correct one using:

$$P(M_i|Z,S,Y) = \frac{P(Y|M_i,S,Z)P(S|M_i)}{\sum_{j=1}^{3} P(Y|M_j,S,Z)P(S|M_j)} \tag{1.85}$$

It should be noted that although $P(S|M_i^{TS}) = P(S|M_i)$, in general it will be the case that $P(Y|M_i^{OP},S,Z) \neq P(Y|M_i,S,Z)$. This is because:

$$\begin{aligned} P(Y|M_i^{OP},S,Z) &= \int P(Y|M_i^{OP},S,Z,\delta)P(\delta|M_i^{OP},S,Z)d\delta \\ &= \int P(Y|M_i^{OP},S,Z,\delta)P(\delta|M_i^{OP})d\delta \\ &= \int P(Y|M_i,S,Z,\delta)P(\delta|M_i)d\delta \end{aligned} \tag{1.86}$$

due to the fact M_i^{OP} makes no predictions on the value of S. On the other hand,

$$P(Y|M_i,S,Z) = \int P(Y|M_i,S,Z,\delta)P(\delta|M_i,S)d\delta \tag{1.87}$$

Where $P(\delta|M_i,S)$ is the posterior distribution of δ given the time series data. In words, $P(Y|M_i^{OP},S,Z)$ averages the likelihood function over the prior density for δ while $P(Y|M_i,S,Z)$ averages the likelihood over the posterior for δ given the time series data. So while $P(Y|M_i^{OP},S,Z)$ compares how well the options data fits with the prior beliefs, $P(Y|M_i,S,Z)$ compares how well the options data agrees with the time series data on the underlying process. This was shown through the approximations developed in Section 1.6.

So in summary, for each model we have four interesting quantities to examine. $P(S|M_i^{TS})$ examines how well the time series data fits with the model while $P(Y|M_i^{OP}, S, Z)$ measures how well the options data fits with the model. $P(Y|M_i, S, Z)$ measures how well the time series data agrees with the options data in the context of the given model while $P(Y|M_i, S, Z) \times P(S|M_i)$ measures how well both the time series and options data taken together fit with the given model.

1.10.2 Results

The method to calculate the relevant quantities above is given in Section 1.5. Table 1.3 on page 51 reports the results of the Bayesian model comparison procedure on the January to February 1993 DM options data set discussed in Section 1.8 using the benchmark set of priors that I specified. I report the results of comparing models based on the time series data, options data and then both. I also report $P(Y|M_i, S, Z)$ for each model. This quantity cannot really be used to compare option pricing models, but as pointed out above, is an informative number since it tells us the degree of agreement between the options and time series data. This quantity is very easy to compute since it is simply the quotient of $P(Y|M_i, S, Z) \times P(S|M_i)$ and $P(S|M_i)$, and since $P(S|M_i) = P(S|M_i^{TS})$ one has all the relevant quantities already. Because $P(Y|M_i^{OP}, S, Z)$, $P(S|M_i^{TS})$, $P(Y|M_i, S, Z)$ and $P(Y|M_i, S, Z) \times P(S|M_i)$ can all be very large, I report the logs of these quantities in the table. The basic result is that based on the time series data alone, the jump–diffusion model is the most likely model, while based on the options data alone the CEV model is the most likely. If one however considers both the time series and options data together, the conclusion is that the jump–diffusion model is the most likely model. It is informative to see that when we use the complete model specification, the jump diffusion model does not simply win out because it fit the time series data better. Based on $P(Y|M_i, S, Z)$, which as pointed out before reflects the degree of agreement between the time series and options data in the context of model M_i, we see that the option price data agrees better with the time series data in the context of the jump–diffusion model. Thus, the jump–diffusion model wins on both counts. It yields a higher $P(S|M_i)$ as well as a higher $P(Y|M_i, S, Z)$.

If one were simply interested in finding a model that gave the right price for the option, the CEV model may be sufficient. However, for the purposes of hedging it is important to not only have the correct pricing function, but also the correct dynamics. For this purpose, we find that the jump–diffusion model is perhaps the best model.

To see the effect of alternative priors, I report the results of the Bayesian model comparison procedure for different sets of priors. We will see that the above results are very robust to changes in the prior distributions. First, I examine the effect of specifying an alternative prior over the drift. Although our prior over the drift is identical across models, as pointed out by Lo and Wang [13], the specification of the drift can significantly affect the estimate of the diffusion (and jump) component.[6] Since the specification of the stochastic component is different across our models, this is a potentially important consideration. We take a prior over the drift which is more diffuse than the original. Specifically, we specify our alternative prior over μ to be uniform on $\mu_m - 0.14$ and $\mu_m + 0.14$ where as before μ_m is the U.S–German interest differential at the beginning of the sample -0.0561. This is in contrast to our previous prior for μ which was a triangular distribution centered at μ_m with support from μ_m -0.07 to μ +0.07. The results are given in Table 1.4 on page 51. The results as to which model is the most likely, are unchanged.

Since the volatility is perhaps the most important aspect of the process for option pricing, it is interesting to see the effect of starting with a prior that is less informative on its value. I specify an alternative prior on volatility that is uniform on 0.04 to 0.24. This is in contrast to the benchmark prior which was triangular with support 0.08 to 0.20 and a peak at 0.14. It is straightforward to specify te priors this implies for the Black–Scholes model. I use the same method as before in specifying the parameter values for the CEV and Jump–Diffusion based on this prior for volatility. The results are reported in Table 1.5 on page 52. Again, one sees that the results as to which

[6]Lo and Wang's key consideration is the issue of predictability which is not addressed in our setup.

model is the most likely, are unchanged.

It is also informative to examine the effect of specifying different priors for the parameters of the CEV and jump diffusion process. First we examine what happens when we allow a greater fraction of the jump–diffusion volatility to come from the jump component in the Merton model. Previously, we had allowed a maximum of 40% of the volatility to come the jump component. We increase this number to 60%. We also examine the effect of making the prior over the elasticity parameter in the CEV model more diffuse. Specifically, we take a prior over γ that is uniform, instead of triangular, on the domain 0 to 4. This is the same support as before, but we now express less certainty in the notion that this distribution of γ is concentrated near 2–an elasticity of 0. I report these two models along with the benchmark Black–Scholes model in Table 1.6 on page 52. The ranking of these models are not changed. Even if one were to compare each of the models with the modified prior against the two other models with their benchmark prior, the results would not change.

1.11 Conclusion

This paper developed a Bayesian framework for comparing option pricing models based on historical data on the time series of the underlying asset as well as the the options data. The general method of the procedure was first developed. Next, to obtain an intuitive feel for the factors at play, I developed an approximation to the relevant quantities and interpreted the Bayesian method using those approximations. As an application, I applied the Bayesian procedure to PHLX traded foreign exchange options to compare the Black–Scholes model, Merton's jump–diffusion model, and Cox's CEV model. I found that the jump–diffusion model was the most probable model based on the time series data while the based solely on the options data, the CEV model was the most probable. When one uses both the time series and option pricing implication of the models, the jump–diffusion model is the most likely model based on the data. We also showed that the success of the jump-diffusion model lay not only in how well it fit the time series data, but also in how well the time series data agreed with the options data in the jump–diffusion context.

Table 1.3: **Benchmark:** Probability of Model based on (a) time series only, (b) options only (c) both time series and options.

		BS	JD	CEV
Time Series Only	Probability	0.00	1.00	0.00
	$\log P(S\|M_i^{TS})$	2093.84	2120.93	2093.74
Options Only	Probability	0.00	0.00	1.00
	$\log P(Y\|M_i^{OP}, S, Z)$	959.08	958.77	964.25
	$\log P(Y\|M_i, S, Z)$	715.35	823.72	720.91
Complete Model	Probability	0.00	1.00	0.00
	$\log P(Y\|M_i, S, Z)P(S\|M_i)$	2809.19	2944.65	2814.65

The probability of a model given the 1993 DM options data collected from the PHLX. The log of the averaged likelihood is given underneath.

Table 1.4: **Flatter Prior over Drift:** Probability of Model based on (a) time series only,(b) options only (c) both time series and options.

		BS	JD	CEV
Time Series Only	Probability	0.00	1.00	0.00
	$\log P(S\|M_i^{TS})$	2093.83	2120.92	2093.72
Options Only	Probability	0.00	0.00	1.00
	$\log P(Y\|M_i^{OP}, S, Z)$	959.08	958.77	964.25
	$\log P(Y\|M_i, S, Z)$	715.33	823.74	720.91
Complete Model	Probability	0.00	1.00	0.00
	$\log P(Y\|M_i, S, Z)P(S\|M_i)$	2809.16	2944.66	2814.63

Given a flatter prior over μ, the probability of a model given the 1993 DM options data collected from the PHLX. The log of the averaged likelihood is given underneath.

Table 1.5: **Flatter Prior over the Volatility:** Probability of Model based on (a) time series only,(b) options only (c) both time series and options.

		BS	JD	CEV
Time Series Only	Probability	0.00	1.00	0.00
	$\log P(S\|M_i^{TS})$	2094.35	2121.15	2094.33
Options Only	Probability	0.00	0.00	1.00
	$\log P(Y\|M_i^{OP}, S, Z)$	958.05	957.72	963.18
	$\log P(Y\|M_i, S, Z)$	713.78	821.66	719.54
Complete Model	Probability	0.00	1.00	0.00
	$\log P(Y\|M_i, S, Z)P(S\|M_i)$	2808.13	2942.81	2813.87

Given a flatter prior over the instantaneous volatility, the probability of a model given the 1993 DM options data collected from the PHLX. The log of the averaged likelihood is given underneath.

Table 1.6: **Model Specific Changes:**Probability of Model based on (a) time series only,(b) options only (c) both time series and options.

		BS	JD	CEV
Time Series Only	Probability	0.00	1.00	0.00
	$\log P(S\|M_i^{TS})$	2093.84	2121.22	2093.79
Options Only	Probability	0.00	0.00	1.00
	$\log P(Y\|M_i^{OP}, S, Z)$	959.08	958.65	964.32
	$\log P(Y\|M_i, S, Z)$	715.35	823.89	721.18
Complete Model	Probability	0.00	1.00	0.00
	$\log P(Y\|M_i, S, Z)P(S\|M_i)$	2809.19	2945.11	2814.97

I specify a flatter prior over the CEV elasticity of volatility parameter and allow the JD model to have a greater proportion of the volatility come from jumps while keeping the BS model at its benchmark prior. The above is the probability of a model given the 1993 DM options data collected from the PHLX. The log of the averaged likelihood is given underneath.

Bibliography

[1] Black F. and Scholes M., "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 1973, **81**, 637–54.

[2] Cox D.R. "Test of Separate Families of Hypotheses," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability,* 1961, Vol. 1, University of California Press, Berkeley.

[3] Cox, J. "Notes on Option Pricing I: Constant Elasticity of Variance Diffusions," Unpublished note, Stanford University, Graduate School of Business, 1975.

[4] Emanuel, D. and MacBeth, J. "Further Results on the Constant Elasticity of Variance Call Option Pricing Model." *Journal of Financial and Quantitative Analysis,* 1982, **17**,533–554.

[5] Gaver and Geisel, "Discriminating Among Alternative Models: Bayesian and Non-Bayesian Methods," in Paul Zarembka, ed., *Frontiers of Econometrics,* Academic, New York, 1974, 49–80.

[6] Gull, S.F.,"Bayesian Inductive Inference and Maximum Entropy," in G.J. Erickson and C.R.Smith Eds., *Maximum Entropy and Bayesian Methods in Science and Engineering,* 1988, Vol. 1, 53–74, Kluwer, Dordrecht.

[7] Gull, S.F., "Developments in Maximum Entropy Data Analysis," in J.Skilling Ed. *Maximum Entropy and Bayesian Methods, Cambridge 1988,* 1989, 53–71, Kluwer, Dordrecht.

[8] Jorion, P., "On Jump Processes in the Foreign Exchange and Stock Markets," *Review of Financial Studies,* 1988, **1**, 427–445.

[9] Kashyap, R.L., "A Bayesian Comparison of Different Classes of Dynamic Models Using Empirical Data," *IEEE Transactions on Automatic Control* 1977, **AC–22:5**, 715–727.

[10] Lindley, D.V., "A Statistical Paradox," *Biometrika,* 1957,**44**, 187–192.

[11] Lo, A., "Statistical Tests of Contingent Claims Asset–Pricing Models, a New Methodology," *Journal of Financial Economics,* 1986, **17**, 143–174.

[12] Lo, A., "Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data," *Econometric Theory,* 1988, **4**, 231–247.

[13] Lo, A. and Wang J. "Implementing Option Pricing Models When Asset Returns are Predictable," *Journal of Finance,* 1995, **50**, 87–130.

[14] MacKay D., "Bayesian Interpolation" in *Bayesian Methods for Adaptive Models*, PhD. Dissertation, California Institute of Technology, Pasadena California. Chapter 2, 1992.

[15] MacKay D. "A Practical Bayesian Framework for Backprop Networks," in *Bayesian Methods for Adaptive Models*, PhD. Dissertation, California Institute of Technology, Pasadena California. Chapter 3, 1992.

[16] Merton, R., "Option Pricing When the Underlying Stock Returns are Discontinuous." *Journal of Financial Economics,* 1976, **5**, 125–144.

[17] Rosenfeld, E., *Stochastic Processes of Common Stock Returns: An Empirical Examination*, Ph.D. dissertation, Alfred P. Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, 1980.

[18] Sankaran, M., "Approximations to the Noncentral Chi–Squared Distribution." *Biometrika*, 1963, **50**, 199–204.

[19] Schroder, M.,"Computing the Constant Elasticity of Variance Option Pricing Formula," *Journal of Finance*, 1989, **44**, 211–219.

[20] Skilling, J., "On Parameter Estimation and Quantified MaxEnt," in J.Skilling Ed. *Maximum Entropy and Bayesian Methods, Cambridge 1988*, 1991, Kluwer, Dordrecht, 455–466.

[21] Zellner, A.,"Jeffreys–Bayes Posterior Odds Ratio and the Akaike Information Criterion for Discriminating Between Models," *Economics Letters,* 1978, **1**, 337-342.

[22] Zellner, A. *An Introduction to Bayesian Analysis in Econometrics*, Wiley, New York, 1987. This is the 1987 printing of a 1971 book.

Chapter 2

Determinants of Foreign Exchange Jump Expectations: Evidence From Option Prices (with L. Kropywiansky)

2.1 Introduction

Traders have claimed that the strategy of borrowing money from a low interest rate country's bank account and putting that money into a bank account of a high interest rate country will yield a modest profit during most periods, but that once in a while the high yield currency will crash and wipe out years worth of profits from this otherwise seemingly profitable strategy. They point out that many exchange traders in the past have actually followed this simple strategy and have been profitable in the short run, but that eventually many were stung by a collapse in the high yield currency to the point that they were forced to close out their position. Many traders have come to believe that the high yield currency is susceptible to sudden major devaluations. Some have characterized the higher interest rate in a country as simply a currency crash premium demanded by investors. Traders say that whether or not this strategy is profitable in the long run is still an open question.[1]

[1]These views were related to Walter Baily during extensive discussions with numerous traders and researchers at the New York and Tokyo offices of Goldman Sachs, Merrill Lynch and Salomon Brothers as well as the Salomon Brothers Hong Kong office.

Researchers starting with Krasker [21], Mussa [31] and Rogoff [33] have formally argued that this type of exchange rate dynamic makes it very difficult to test whether or not such a strategy makes money on average. Their arguments have been couched in terms of what is called the uncovered interest parity hypothesis, but is essentially equivalent to the question of whether the above strategy makes money on average. Uncovered interest parity simply states that if investors are risk neutral and have rational expectations, a foreign currency must on average depreciate (appreciate) by the amount that the foreign interest rate exceeds (is less than) the domestic interest rate. To see this, note that if a U.S. investor bought a dollar worth of Deutschemarks and invested it into a German bank account for a month, converting the total amount back into dollars at the end of the month, the return on the strategy would simply be the return on the Deutschemark over the month plus the interest accrued on the German bank account. Under risk neutrality, the expected return on this strategy must be equal to the U.S. interest rate. Or put differently, the expected return on the Deutschemark must simply be equal to the U.S. interest rate minus the German interest rate.

But if, as traders claim, it is the case that the majority of the expected depreciation in the high yield currency happens only in the form of rare but large crashes, over short horizons it is very likely that the currency will appear to not satisfy uncovered interest parity. If the bulk of the exchange rate movement is expected to come in the form of jumps, the tests become extremely sensitive to whether or not a jump occurred in the sample. The interest differential could appear over short horizons not to explain the expected return of the exchange rate. Even more deceptive is the fact that because any test conducted where no jumps occurred in the sample would underestimate the volatility of the process, the tests of uncovered interest parity could mistakenly reject the hypothesis even when it is true.

This type of problem is commonly referred to as the peso problem and is quite general in theory. It could arise in empirical tests of stock pricing or the interest rate term structure.[2] However, because exchange rates seem to be particularly susceptible

[2]Reitz [32] suggests that rare but catastrophic falls in aggregate consumption could explain the

to occasional large jumps, peso problems are most likely to arise in the context of international asset pricing models. The term "peso problem" has its roots in a puzzle that lasted for over 20 years from April 1954 to August 1976 when the Mexican interest rate was higher than the U.S. rate despite a fixed exchange rate of 0.080 dollars per peso over the entire period. One could seemingly make a pure arbitrage profit by borrowing money in the U.S. and lending to banks in Mexico. The puzzle can be resolved by noting that speculators may have been expecting the possibility of a collapse of the peso which finally occurred on August 31, 1976 when the Mexican government formally let the peso float. In other words, the higher interest rate in Mexico could simply have been a crash premium to compensate holders of Mexican bonds. The peso initially fell from 0.080 to 0.050 dollars per peso before finally settling down to a small range around 0.044 dollars per peso by March 1977.[3]

Peso problems may be just as likely to occur under floating exchange rate regimes as they are under fixed rates. It has long been speculated that floating exchange rates, like other assets, may be subject to "bubbles" which persist for some time before eventually bursting. It is also true that the post-1970's regime of floating is best characterized as a "dirty float," in which exchange markets are subject to occasional large interventions by national governments. Thus, it is a priori plausible that either bubbles or expected interventions could give rise to expectations of exchange rate jumps.[4] During 1980-85, the dollar appreciated an average of 33% against each of the major currencies despite higher interest rates in the U.S. The dollar's rise was reversed in the fall of 1985 when the G-7 countries staged a coordinated open market operation. It is plausible that the higher interest rate in the U.S. prior to 1985 reflected the market's expectation that the dollar's rise would eventually be reversed, either of its own accord or through intervention.[5] More recently, there has been speculation of

equity premium in stock returns. Lewis [23] gives a discussion of peso problems and the term structure while [25] gives a general discussion of peso problems in a variety of markets.

[3]Lizondo [27] gives a detailed account of the behavior of the forward discount before the Mexican devaluation.

[4]Note that bubbles and interventions need not be mutually exclusive events. The so-called "dollar problem" of the 1980's could be interpreted as a bubble which was eliminated by intervention.

[5]See Froot and Thaler [17]

a peso problem in the Deutschemark. Some believe that the Bundesbank's current high interest rate policy is not sustainable and it will eventually lower interest rates to stimulate the German economy and thereby send the Deutschemark tumbling.[6] Traders have also been eyeing the Thai baht for a few years now as a high yield currency play that could have disastrous consequences.[7]

The anecdotal evidence on the existence of occasional large jumps in foreign exchange markets is confirmed by the formal empirical analyses of Jorion [19], Akgiray and Booth [1], and Tucker and Pond [34]. These studies find that a jump-diffusion model dominates various forms of the pure diffusion model in explaining the behavior of the major trading currencies during the post-1974 floating rate period.[8] There is also an empirical option pricing literature which provides further evidence that jumps are important in currency markets. For example, Bodurtha and Courtadon [10] find that the simple Black-Scholes option pricing formula makes systematic errors when used to price out-of-the-money foreign exchange options, and that those errors are consistent with the market expectations of large jumps. In a series of papers, Bates[4], [5],[6] and [7] shows that options prices can be used to uncover market expectations of jumps and that such expectations are significant for the dollar-DM exchange rate in his sample of 1984–1987.[9]

In this paper we use option prices to uncover the Dollar-Deutschemark and Dollar–Yen jump expectations of market participants for the sample period of 1984 to 1993 and examine whether the high yield currency was in fact susceptible to crashes as claimed by traders. Although it has long been claimed that jump fears are present

[6]See for example Dornbusch[12].

[7]A view held by a trader at Salomon Brothers' Hong Kong office.

[8]Jorion's result holds even when he allows for conditional heteroskedasticity in the diffusion process. Akgiray and Booth and Tucker and Pond compare the jump-diffusion model to mixture-of-normals and stable Paretian models.

[9]In [4], [5], Bates shows that there was significant downward skewness in the dollar-DM exchange rate during the "dollar problem" prior to 1986, and that this skewness is consistent with the presence of jump expectations. In a later paper [7], he estimates the parameters of a jump–diffusion process with stochastic volatility under the assumption that the parameters stay fixed over his entire 1984-1991 sample. In a recently received working paper[8], he examines a longer time series for the DM and yen. In [6], he examines the behavior of options on S&P 500 futures during the time leading up to the '87 crash and concludes that the crash was expected.

in the high yield currency, there has been very little evidence to back up this claim. We follow Bates and take advantage of the fact that since option prices reflect the market's belief of the underlying asset's distribution, one can use observed options prices to recover the market's expectation of a crash in foreign exchange markets. We use a modified version of Merton's jump–diffusion option pricing formula as a filter to elicit the market's expectation of the size, direction, and probability of a jump in a currency. This is much like using Black–Scholes implied volatility as a filter of the market's belief of the underlying volatility.

We confirm in the case of the Dollar–Deutschemark rate the claim made by traders and the story forwarded by many proponents of the peso problem that the high yield currency is susceptible to crashes. In the case of the Deutschemark versus the dollar, a one percent decrease in the U.S.–German interest differential increases the expected depreciation in the Deutschemark due to jumps by roughly 0.18 percent. We, however, do not find support for the hypothesis in the case of the Dollar–Yen rate. We find that there were substantial fears of jump depreciation of the dollar versus the Deutschemark and the yen in the 1984-1989 period. In the 1989–1993 period, there are many more months where substantial dollar jump appreciations were expected against the Deutschemark than jump depreciations. The dollar exhibits expected jump appreciation against the yen from 1989 to early 1991 while during mid 1991 to mid 1992 there are dollar jump depreciations expected against the yen.[10] We also examine whether option implied jump expectations can predict actual jumps. We find that jump expectations do seem to predict jumps if only weakly. We further investigated whether jump expectations can be explained by fundamental economic forces such as purchasing power parity and the trade balance. We find evidence that when the currency is trading far away from purchasing power parity (PPP), there is a tendency for traders to think that there will be a jump of the currency in the direction toward PPP levels. We find that the trade balance has very little power in explaining jump beliefs.

[10] After mid–1992, trading in yen options decreased dramatically and leaves many months with insufficient observations.

The paper is organized as follows. Section 2.2 discusses the data, the option pricing methodology and the statistical method used to invert for the jump–diffusion parameters. Although the inversion process is straightforward in theory, there are many computational problems that had to be tackled. These are discussed in Section 2.3. Section 2.4 presents the estimated jump–diffusion parameters and conducts some robustness checks. Section 2.5 checks whether option implied jump expectations can predict actual jumps. Section 2.6 examines the relationship between the jump expectations, purchasing power parity and the trade balance. Section 2.7 concludes.

2.2 Data, Pricing and Statistical Issues

The options data used are from the Philadelphia Stock Exchange (PHLX) and consist of transactions data on foreign currency options traded between January 1984 and December 1993. PHLX currency options are written on the underlying cash market on the respective foreign currency—not the futures market as is the case in the Chicago markets. Each observation consists of the price paid for the option, the Telerate spot exchange rate at the time of the transaction, and the term to maturity and strike price of the option. The options data for June through November of 1985 cannot be used due to severe errors in reporting during that period.

To invert for jump–diffusion parameters, we will need to specify an option pricing formula that is used by the market to price options in the presence of jumps. For this purpose, we take a behavioral view of the way that options are priced in the market. Ultimately, we will use the option pricing formula as a filter to back out market expectations. To obtain such an option pricing formula, we will first need to specify the dynamics of the underlying process that the derivative traders believe the exchange rate follows in a continuous time setting. We assume options traders in pricing options take the exchange rate to follow the process given below between now and the time the option expires:

$$\frac{dS}{S} = [\mu - \lambda\kappa]dt + \sigma dz(t) + \kappa dq(t) \tag{2.1}$$

where $z(t)$ is a Wiener process and $q(t)$ is a poisson process with jump intensity λ and where κ is a non–stochastic jump size. We assume traders input their current belief of λ, κ, and σ when pricing options. The instantaneous expected rate of return of this process will be μ. Conditional on no jumps occurring however, the expected rate of return will be $\mu - \lambda\kappa$. To price the option, we follow Merton[30] and assume that the jump risk is not priced–i.e. that it can be diversified away.

Because the majority of currency option transactions on the PHLX are of American style options (96% of the Deutschemark option transactions are American—almost all transactions were American prior to 1991), we chose to work with the American options in order that we do not run out of observations in many of the years prior to 1991. A drawback of using American options is that there exists no analytic closed form formula for the price of an American option. One must either use a finite difference backward marching scheme or rely on an analytic approximation of the American option price. Because the non-linear least squares regression we use to invert the options prices is an iterative method, relying on finite difference methods would make the whole procedure prohibitively time consuming. Instead, we rely on an analytic approximation proposed by Bates [6] which extends the work of Barone–Adesi and Whaley [3] and MacMillan [29] to the case where the underlying price process follows a jump-diffusion process.

As pointed out by Bates [7], theoretical pricing of PHLX options must also take account of delivery lags. According to PHLX contractual agreements, an American option is settled in 5 days after expiration and 7 days (5 business days) after early exercise. Given these contractual specifications and the process described above, one can arrive at the following approximation to the American put or call price:

$$\Psi_A(pc) = \begin{cases} \Psi_E(pc) + \left(\frac{S}{S^*}\right)^q \left[pc \cdot \left(e^{-r^*\tau_2} S^* - e^{-r\tau_2} K\right) - \Psi_E(pc)\right] & \text{if } S < S^* \\ pc \cdot \left(e^{-r^*\tau_2} S^* - e^{-r\tau_2} K\right) & \text{if } S > S^* \end{cases} \tag{2.2}$$

where pc takes the value 1 if the option is a call and -1 if it is a put, S is the value of the underlying exchange rate, K is the strike price, τ_2 is the delivery lag after early exercise, and S^* is the level of the exchange rate at which it is optimal to exercise early and hence satisfies the following equation:[11]

$$S^* = \text{ argmax}_{S^*} \left(\frac{S}{S^*}\right)^q \left[pc \cdot \left(e^{-r^*\tau_2} S^* - e^{-r\tau_2} K\right) - \Psi_E(pc)\right] \tag{2.3}$$

while the parameter of curvature q satisfies:

$$\frac{1}{2}\sigma^2 q^2 + (r - r^* - \frac{1}{2}\sigma^2 - \lambda\kappa)q - \frac{r}{1 - e^{r\tau}} + \lambda(1+\kappa)^q - \lambda = 0 \tag{2.4}$$

and $\Psi_E(pc)$ is the corresponding European option price:

$$\Psi_E(pc) = e^{-r(\tau+\tau_1)} \sum_{n=0}^{\infty} P(n) \cdot pc \cdot \left[S e^{(r-r^*)\tau_1 + b_n\tau} \mathrm{N}(pc \cdot d_{1n}) - K\mathrm{N}(pc \cdot d_{2n})\right] \tag{2.5}$$

where:

$$P(n) = \frac{e^{-\lambda\tau}(\lambda\tau)^n}{n!} \tag{2.6}$$

and τ_1 is the delivery lag after expiration, N(.) denotes the cumulative normal distribution and:

$$b_n = r - r^* - \lambda\kappa + n\log(1+\kappa)/\tau \tag{2.7}$$

$$\begin{aligned} d_{1n} &= \frac{\log(S/K) + (r - r^*)\tau_1 + (b_n + \sigma^2/2)\tau}{\sigma\sqrt{\tau}} \\ d_{2n} &= \frac{\log(S/K) + (r - r^*)\tau_1 + (b_n - \sigma^2/2)\tau}{\sigma\sqrt{\tau}} \end{aligned} \tag{2.8}$$

[11] The first order condition of this maximization is equivalent to the smooth pasting condition derived in Bates.

A set of observed option prices $\{y_i\}_{i=1}^{M}$ will in all likelihood not fit the parametric specification. A set of three options along with the price of the underlying asset and the other observable inputs to the option pricing function will uniquely determine $\hat{\lambda}, \hat{\kappa}$ and $\hat{\sigma}$. Any other set of three options will probably produce another set of $\hat{\lambda}, \hat{\kappa}$ and $\hat{\sigma}$. This is a problem common to all option pricing models. The likely reasons for this is that (i) the assumed model is misspecified, (ii) some options are thinly traded and trades do not take place at "true values", (iii) trades take place at bid and ask prices instead of at the "true value," (iv) investors make errors in evaluating the "true value" of the options. All four reasons are likely play a role. We will assume that our specification is approximately correct and that the majority of the errors come from factors (ii), (iii) and (iv). We further assume that the deviations of the observed prices from the "true values" can be expressed in terms of a mean zero random variable that is independent of the arguments entering the option pricing formula.

By taking a set of options traded during a given day, we can "invert" these options to find the implicit parameters used by option traders during that day. A consistent estimate of λ, κ and σ is obtained by the following non-linear least squares regression:[12]

$$\{\hat{\lambda}, \hat{\kappa}, \hat{\sigma}\} = \text{argmin}_{\lambda,\kappa,\sigma} \sum_{i=1}^{M} (y_i - \Psi_i(pc))^2 \tag{2.9}$$

Due to the over-abundance of data in many months, we construct a sub-sample of the data as follows. Our objective was to pick one trading day in a given month and take options trading in that day that had the same expiration date. We also wanted at least 50 options. We start with the first trading day of a given month and find the expiration date that was less than 120 days away with the most options. If that set contained at least 50 options we used that set. Otherwise we proceeded to the next trading day.

[12]Consistency is achieved as the number of options $M \to \infty$.

2.3 Computational Considerations

In theory, the inversion process is straightforward given the setup above. The actual implementation of this procedure however presents many computational challenges and consumed most of our time. We spend some time here on the problems that came up and the method we used to tackle them. Some of the techniques we present should be useful even to options researchers who are not planning on inverting for jump diffusion parameters.

First, one must come to terms with the problem of calculating the European jump diffusion option price $\Psi_E(.)$ which is expressed in closed form only as an infinite series. This European formula is an integral part of the analytic approximations used for the American formula. The standard approach used to calculate this formula is to simply determine a cut off level for the number of terms in the sum based on the parameters. We found that such a method would require a very large number of terms and would make the inversion process extremely time consuming. We noticed, however, that the number of terms required for accuracy was much less for the puts compared to the calls when κ was positive. The reverse was true when κ was negative. One can take advantage of this to reduce the computation time for all options using put-call parity. We will return to this shortly. The reason why fewer terms are needed for the puts relative to the calls when $\kappa > 0$ can be understood by taking the simplified case of a pure jump process. The put price in this case (ignoring delivery lags) would be:

$$e^{-r\tau} \sum_{n=0}^{\infty} P(n) \left(K - e^{-\lambda\kappa\tau} S_0 (1+\kappa)^n\right)^+ \tag{2.10}$$

Each term of the sum inside the $(.)^+$ reflects the final payoff conditional on n jumps occurring. If κ is positive $1+\kappa$ is greater than one so that $(1+\kappa)^n$ is increasing in n—or in other words, each jump is pushing the option out of the money. So for sufficiently large M, $\left(K - e^{-\lambda\kappa\tau} S_0(1+\kappa)^n\right)^+$ will be equal to 0 for all $n > M$. It is also clear that deeper in the money puts require more terms since more jumps are needed until the put goes out of the money. On the other hand, the call price is equal to:

$$e^{-r\tau}\sum_{n=0}^{\infty}P(n)\left(e^{-\lambda\kappa\tau}S_0(1+\kappa)^n-K\right)^+ \tag{2.11}$$

Thus when $\kappa > 0$, each jump pushes the option deeper *into* the money. This means that the payoff is increasing in n. The infinite sum will still converge since the probabilities are dying out at the rate of $1/n!$, but many more terms will be required before the product is insignificant. With the put and the call, the price can be written:

$$\sum_{n=0}^{\infty}h_n i_n \tag{2.12}$$

where h_n is the product of the probability and $e^{-r\tau}$ and i_n is the payoff at expiration. When κ is positive, h_n and i_n are both decreasing for the call while for the put although h_n is decreasing, i_n is increasing. So clearly, calculating the put in this case requires much fewer terms. A similar logic can be applied to show that the call is faster to converge when κ is negative. A similar argument holds for the case when there is diffusion on top of the jump process.

So to calculate a call price when $\kappa > 0$, one can first calculate the corresponding put price and use put–call parity to find the call price. When $\kappa < 0$ one can do the reverse. The put–call parity relation with proportional dividends r^* and delivery lag τ_1 can be written:

$$C(S,K)+e^{-r(\tau+\tau_1)}K = P(S,K)+e^{-r^*(\tau+\tau_1)}S \tag{2.13}$$

We then selected the number of terms to equal the minimum of 8 and $6*\lambda*\tau$ plus 8 times the moneyness in excess of 0.05.[13] This procedure gave us values for the option price that were accurate to within 1×10^{-10} for all parameter values that we tried and typically required only 8 to 10 terms. We found that this technique had its greatest positive impact when we used the same basic technique to calculate the partial derivatives of the pricing formulas with respect to the jump diffusion parameters–a

[13]In other words, we added more terms if the option was more than five percent in the money. We did not take away terms if it was less than five percent in the money.

necessary step when actually doing the non-linear least squares minimization.

Second, to calculate the analytic approximation to the American price one must solve for S^* using equation 2.3. Again in theory this procedure is straightforward, but in practice requires great care. The second derivative of the objective function of S^*, (2.3), can be very erratic and necessitates the use of a bisection algorithm in combination with Newton-Raphson algorithm in solving for S^*. The first derivative is not even monotonic. In other words, the objective function (2.3) is not well approximated by a quadratic function. If one started with the standard Newton Raphson procedure without the bisection algorithm the procedure would not converge in many cases. The problem is especially acute for shorter term options.

Third, to do the actual parameter inversion, we use the Levenberg–Marquardt procedure that takes advantage of the approximate local quadratic property of the objective function around the maxima in a non-linear least squares regression.[14] This procedure however, requires that we be able to calculate the derivative of the function $\Psi_A(.)$ with respect to the parameters. Whether one takes this derivative analytically or calculates it numerically, one can save considerable time by noting the following. The partial derivative of Ψ_A with respect to the parameters is the sum of the derivative of Ψ_E with respect to the parameter and the partial derivative of the American premium, the second piece in the formula for Ψ_A, with respect to the parameters. The derivative of the American premium, call it AP, with respect to the generic parameter β is:

$$\frac{\partial AP}{\partial \beta} = \left.\frac{\partial AP}{\partial \beta}\right|_{\text{Keeping } S^* \text{ fixed}} + \frac{\partial AP}{\partial S^*}\frac{\partial S^*}{\partial \beta} \tag{2.14}$$

All the terms involved are straightforward to calculate except the very last one, $\frac{\partial S^*}{\partial \beta}$. However, one finds that this term need not be calculated because $\frac{\partial AP}{\partial S^*}$ is equal to zero

[14]To stem any possible confusion, note that here we are referring to the approximate quadratic property of the sum of squared residual objective function and not the objective function for S^* which we were referring to in the above paragraph.

since S^* is the maximizer of AP. Thus:

$$\frac{\partial AP}{\partial \beta} = \left.\frac{\partial AP}{\partial \beta}\right|_{\text{Keeping } s^* \text{ fixed}} \tag{2.15}$$

This is a consequence of the envelope theorem. This fact greatly simplifies the calculations of the derivative.

Fourth, one can drastically reduce the required computing time by noting as did Bates[6] that S^* for a fixed τ,r and r^* is homogeneous in K. In other words, given the same τ,r and r^*, if S_1^* is the solution to the objective function (2.3) with K equal to K_1, then S_2^* for K equal to K_2 is equal to $S_1^* \frac{K_2}{K_1}$. Since we are using only one day's worth of data and one maturity for each given month, τ,r, and r^* will all be the same. So we only need to calculate one S^* for the puts and one S^* for the calls for each set of parameters that we try.

Fifth, we use the innovation of Bates [6] and minimize the least squares objective function by searching over the transformed variables:

$$\{\beta_1, \beta_2, \beta_3\} = \{\log(\nu), \mathrm{N}^{-1}(f), \log(1+\kappa)\} \tag{2.16}$$

where:

$$\begin{aligned} \nu &= \lambda(\log(1+\kappa))^2 + \sigma^2 \\ f &= \lambda(\log(1+\kappa))^2/\nu \end{aligned}$$

These last two parameters ν and f are the total variance of the process and the proportion of the variance coming from jumps respectively. This effectively reduces the dimension of the search because the estimate of ν takes on roughly the same value regardless of the other two parameters. Because the objective function has multiple minima, six starting guesses were used in each monthly regression to ensure that the global minimum was found.

Lastly, all the programming was coded in Fortran to make the procedures as fast as possible. We initially coded the whole procedure in the matrix programming language Matlab, but found that one set of regressions for one country would take a month to run.

2.4 Jump–Expectation Estimates

The results of the monthly non-linear least squares regressions along with the asymptotic approximations to the standard errors are reported in Tables 2.2 through 2.8 starting on page 85 while the monthly time series for the jump expectation of the dollar–DM exchange rate $\lambda_i\kappa_i$ is shown in Figure 2-1 on page 69. The jump expectation is quoted in annualized terms. Hence, 0.05, or 5 per cent, corresponds to a expected depreciation of 0.42 per cent over a single month (λ is measured in a scale of probability per unit of time while κ has no time dimension). Although the estimates for $\lambda\kappa$ are of a fairly reasonable order of magnitude, the estimates of the individual components λ and κ are somewhat less believable. This is a result of the fact that in the space of λ and κ (for a fixed value of total variance[15]) there is an approximate hyperbolic ridge in the least squares objective function—the objective function is peaked for one value of $\lambda \times \kappa$.[16] The individual estimate of λ and κ, however, is not well estimated because the peak of this ridge is not very well identified—the objective function takes on similar value for all points on this ridge.[17]

The estimated jump expectations are of an economically reasonable magnitude and, moreover, the time series of $\hat{\lambda}\hat{\kappa}$ seems to agree with the qualitative stories told about the dollar in the 1980's and 1990's. During the mid-1980's when the dollar was thought to be overvalued it was indeed the case that participants in the options market were expecting dollar depreciations due to jumps. These results are in general accord with those of Bates [5],[4] who also found substantial expected jump depreciations for

[15]The objective function is maximized for one value of total variance, ν, almost regardless of the other parameters.

[16]The estimates for κ can sometimes be quite large especially when λ is small. This is not surprising since when λ is very small, κ does not much influence the pricing function. This is the reason why the standard errors can be very large for κ when λ is small.

[17]There were also two months out of the total 114 for which the non–linear least squares procedure did not converge. For the yen there was one month that did not converge.

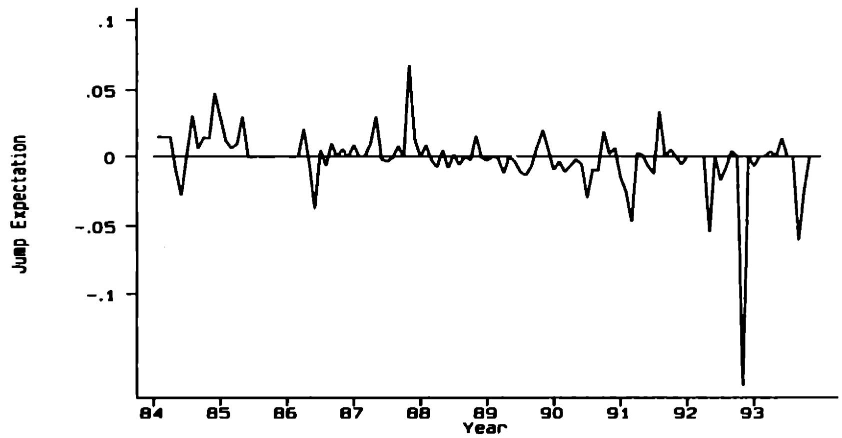

Figure 2-1: Jump Expectations $\lambda\kappa$ for the DM

his sample of 1984–87 using CME options on Deutschemark futures.[18] Interestingly, occasional fears of dollar depreciation due to jumps persisted until the late 1980's while the dollar was on a steady downward path during this entire period. January 1990 marks the start of a new period during which there are many months when the market was calm, with essentially no expected depreciation due to jumps. These calm months are interspersed with months when there are fairly dramatic expected appreciations.[19] It is also interesting to note that this later period coincides with the time that the economic problems associated with the reunification of Germany became apparent. People claimed that the tight monetary policy implemented by the German government could not be sustained given the severe recession in the former East Germany. The Bundesbank it was believed would eventually have to lower its rate and would consequently send the Deutschemark tumbling.[20].

[18]It is unfortunate that the PHLX options price data are completely contaminated and hence unusable for the months June–November 1985 which includes the post-Plaza Accord crash of the dollar. The expected depreciation for April 1985 is large, perhaps indicating anticipation of the eventual decision of the G-5 to bring the dollar down.

[19]Bates [7] estimates option-implied exchange rate parameters under an assumption of time-invariance of the parameters over the entire period 1984-1991 and concludes that there is no statistically significant jump component in the dollar-DM exchange rate. There are at least two ways to reconcile this result with ours and with Bates' results. First, because Bates assumes time-invariance of the parameters in [7], the fact that the jump expectations shown in Figure 2-1 are large and time-varying may be masked. Second, Bates allows for stochastic volatility of the continuous component of exchange rates while we do not, so that we may be imputing skewness of the exchange rate process to the jump component, when it is in fact due to stochastic volatility of the continuous component.

[20]See Dornbusch [12]

It is interesting to note that the two months with the most dramatic jump expectations are October 1987, when a crash occurred in the U.S. stock market, and October 1992, the period of the ERM crisis. In October 1987 there was a 6.7% expected dollar depreciation while in October 1992 there was a 16.7% expected Deutschemark depreciation. In both cases, the currency of the country in crisis was expected to jump down.

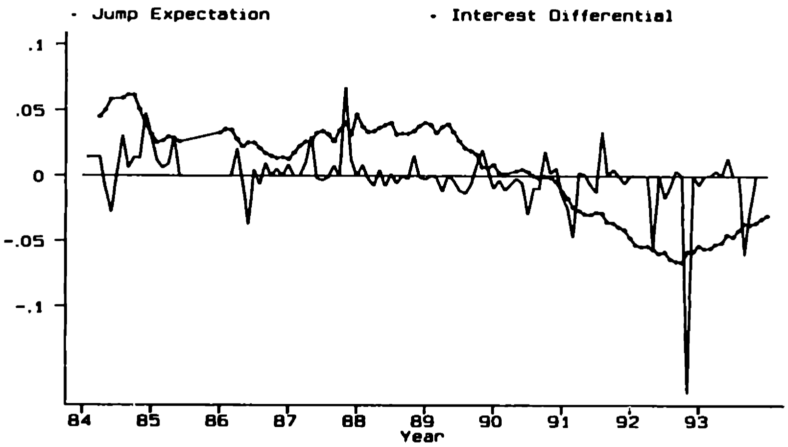

Figure 2-2: $\lambda\kappa$ for the DM & U.S.–German Interest Differential

The broad picture presented by the option-implied jump expectations is one of expected depreciation due to jumps during 1984–1989 and expected appreciation thereafter. As Figure 2-2 shows, this pattern of expected jump depreciation closely tracks the behavior of the U.S.–German one month interest differential. During the earlier period of expected depreciations, the interest differential is positive (i.e., a dollar forward *discount*), while in the latter period of expected appreciations the differential is negative. The correlation between the interest differential and the jump expectation is 0.30. A linear regression of the jump expectation on the interest differential yields a slope coefficient of 0.18 with an OLS standard error of 0.06. Even throwing out October 1992 and October 1987, the regression coefficient is 0.11 with an OLS standard error of 0.04.

Figure 2-3 on page 71 displays the jump expectations in the dollar–yen rate while Tables 2.9 through 2.15 starting on page 92 give the individual estimated jump diffusion parameters. There are some similarities here to the dollar–DM case. In the period from 1984 to 1985 there are fears of dollar depreciations from jumps. Again it is possible that market participants feared the coordinated attempt by the G7 to bring down the dollar that ultimately took place in late 1985. After this period, from 1986 to late 1988 there remained fears of a dollar depreciation although much smaller than what was expected earlier. From 1989 to late 1990 there are months with dollar expected depreciations interspersed with months of large expected dollar appreciations. After 1990, any similarity to the DM–dollar jump expectations end. From 1991 to mid 1992, there are consistent fears of dollar depreciation. Starting in mid–1992 the options trading on the yen became so thin on the PHLX that for many months there was not a single day where more than 50 options with the same maturity were traded. From July 1992 to December 1993, 10 of the 18 months did not have enough options. It is not surprising in this case that for the months for which we did have enough data that the estimates were so erratic. Perhaps the trading was so thin in the markets that options traded far from their "true" value.

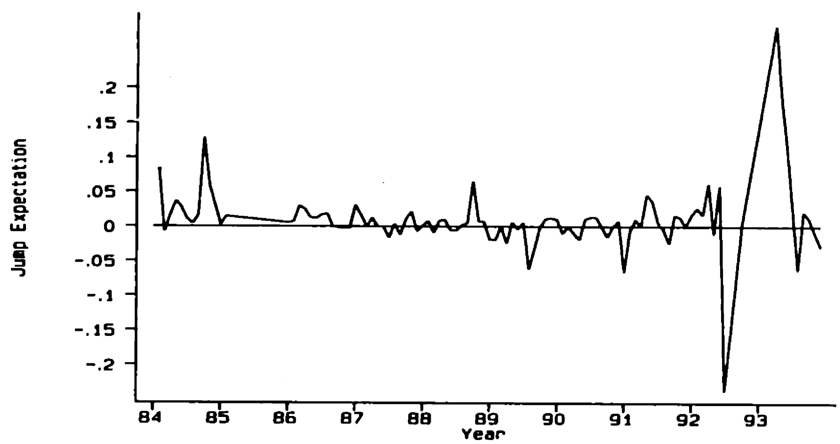

Figure 2-3: Jump Expectations $\lambda\kappa$ for the Yen

Figure 2-4 displays the time series of the yen jump expectation together with the U.S.–Japan one month interest differential. The dollar–yen rate presents a stark contrast to the case of the dollar–Deutschemark rate. The jump differential is not at all correlated with the interest differential. The correlation between the interest differential and the jump expectation is -0.03 while a regression of the jump expectation on the interest differential yields a slope coefficient of -0.06 and an OLS standard error of 0.24. Even using only the data prior to mid–1992 the correlation is -0.0012 and the regression coefficient is -0.0013 with an OLS standard error of 0.12. Thus the statement that crashes are feared in the high yield currency seems to not hold universally although it may hold for some currencies such as the Deutschemark. We will see later that jump expectations in the yen are very difficult to explain, especially if one includes the data after mid-1992.

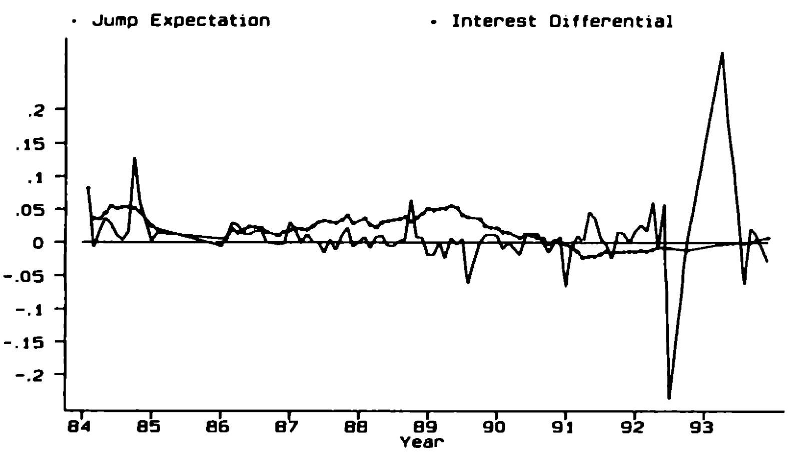

Figure 2-4: $\lambda\kappa$ for the Yen & U.S.–Japan Interest Differential

To complete the picture of the option implied parameters Figure 2-5 gives the monthly time series of σ_i. The volatility figure, which is annualized, varies from a high of about 18 per cent during the summer of 1985 to lows which are below 10 per cent during many months in the sample. The option implied estimates for the yen's continuous volatility are plotted in Figure 2-6. The continuous volatility in the yen falls to a low of 0.023 in September 1984. Because λ and κ are large during this month the total volatility including the jump component works out to a more realistic number of 0.092.

2.4.1 Robustness Checks

In this section, we test the sensitivity of our results to the given specifications. First we test to see whether the relative number of puts to calls could be significantly affecting our results. We do this by estimating the parameters from a data set that takes an equal number of puts and calls each month. Secondly, we also test whether

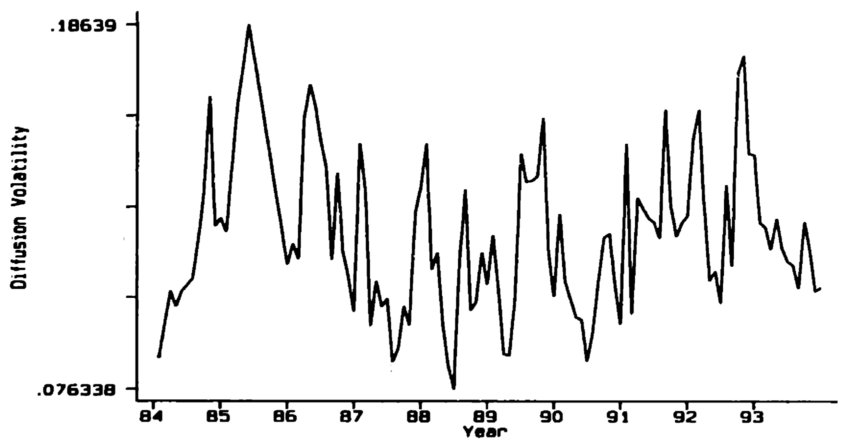

Figure 2-5: Diffusion Volatility for the DM

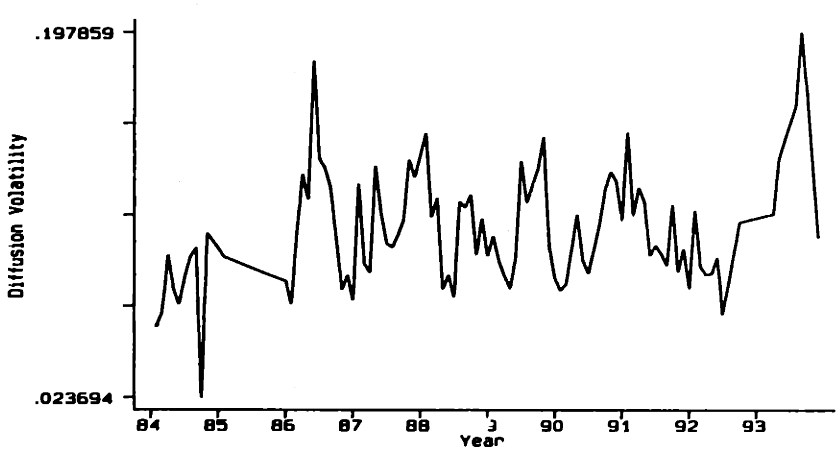

Figure 2-6: Diffusion Volatility for the Yen

our results are driven by slight errors in the analytic approximation that we use to price our American options. We do this by estimating the jump–diffusion parameters using only the out-of-the-money options.

Because the relative number of puts and calls in our data set varies from month to month, it may be of concern to make sure that this fact by itself is not affecting our results. If for example it is the case that all the options traded in a given month are puts, some may be worried that we would pick up a negative jump expectation even in the absence of a market belief that there would be a negative jump. First, we point out that there are months in which puts are more heavily traded and we find positive jump expectations. However, to allay any lingering fears, we invert our parameters using a balanced sample of puts and calls each month. We require that each day that we use have at least 50 total observations Figure 2-7 displays the jump expectations along with the interest differential for the DM. The regression coefficient of the jump expectation on the interest differential is 0.24 with an OLS standard error of 0.11. even throwing out the October 1992 negative jump expectation (ERM crisis) which has a $\lambda\kappa$=-0.39, the regression coefficient is 0.078 with an OLS standard error of 0.048. The plot of the yen jump expectations using a balanced sample of puts and calls is given in Figure 2-8. The regression coefficient of the jump expectation on the interest differential is 0.069 with an OLS standard error of 0.191.

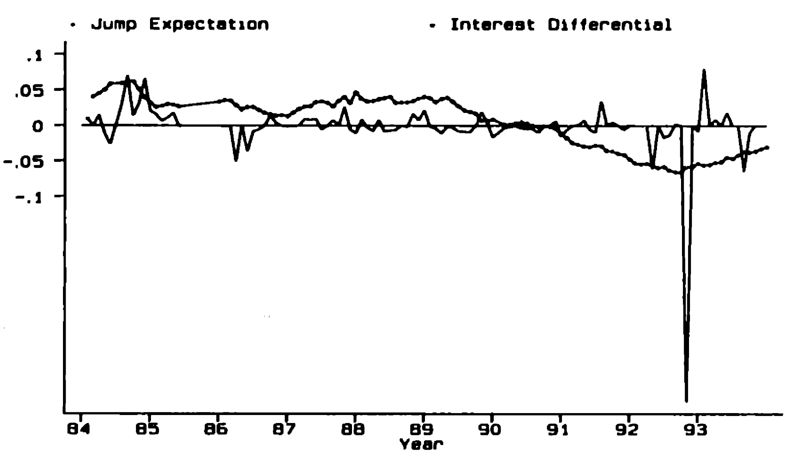

Figure 2-7: $\lambda\kappa$ and the interest differential for the DM using a balanced sample of puts and calls

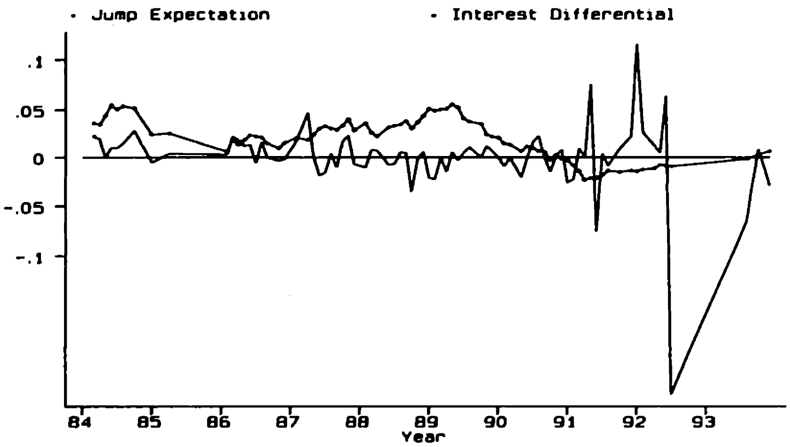

Figure 2-8: $\lambda\kappa$ and the interest differential for the Yen using a balanced sample of puts and calls

One might also be concerned that the analytic approximation that we used for the American option price might create a bias in our jump–diffusion estimates. We can partially address this concern by estimating our jump–diffusion parameters using only out-of-the-money options. Since the American premium, which is the part that we approximate, is essentially worthless for out of the money options, we should be able to obtain estimates that are immune to this potential problem by using only out-of-the-money options. The estimates of the jump expectations for the Deutschemark using only out-of-the-money options is given in Figure 2-9. The plot is very similar to what we obtained using both in and out of the money options. The major difference is the presence of a major negative jump expectation in the DM that was previously not there. Regressing the jump expectations on the interest differential now, we obtain a coefficient of 0.091 with an OLS standard error of 0.083. If we drop the the new massive negative jump expectation, we get a slope coefficient of 0.10 with an OLS standard error of 0.045. For the yen, the estimates are plotted in Figure 2-10. Here the plots actually look significantly different. The regression of the jump expectations on the interest differential is now -0.37 with an OLS standard error of 0.30.

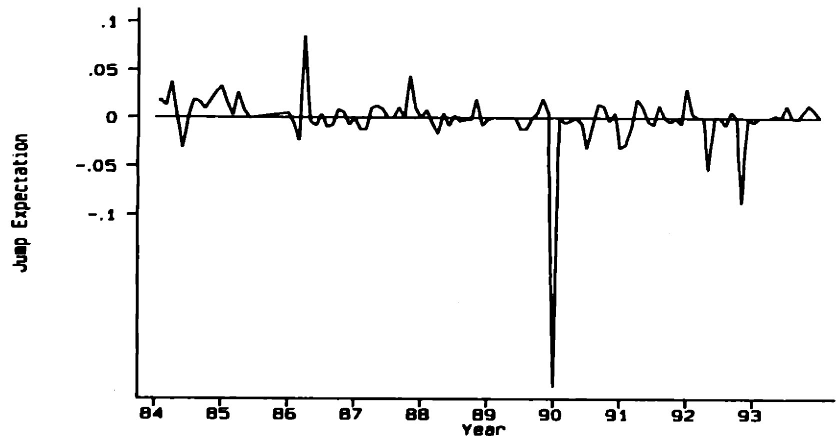

Figure 2-9: $\lambda\kappa$ for the DM with out-of-the-money options

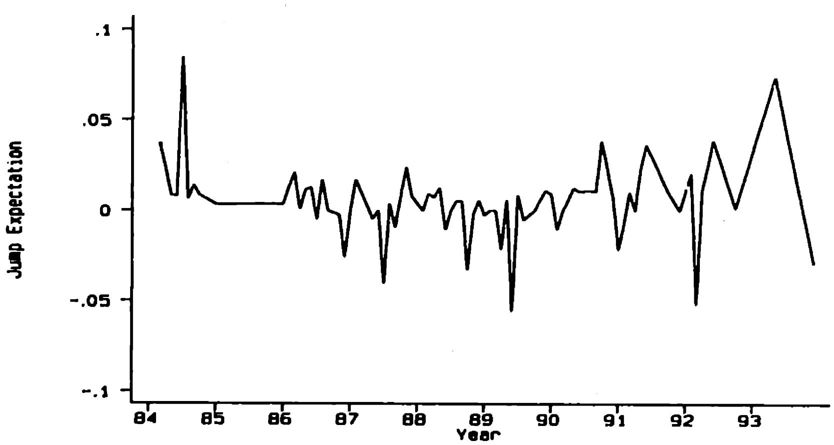

Figure 2-10: $\lambda\kappa$ for the Yen with out-of-the-money-options

2.5 Can Option Implied Jump Expectations Predict Jumps?

Given the option implied jump expectations we estimated in the previous section, it is a point of immediate interest to see whether these jump expectations can predict jumps. This is of interest for several reasons. First, if jumps can be predicted using option prices, this would clearly help traders in formulating their strategies. Second, the peso problem explanation of the empirical failure of uncovered interest parity says that jump expectations held by market participants, although rational, are generally not realized in small samples due to the underlying hypothesis that jumps are very rare. This would imply that we should not be able to predict jumps using the estimated jump expectations.

The question of whether option implied jump expectations predict jumps is complicated by the problem of identifying what are jumps. One possible way to do this is to estimate the jump diffusion parameters over a period following the trade date of the options that we used and then determine what was a jump and what wasn't based on these parameters. This approach will typically claim that a jump occurred over every sample and that the largest one day movement over that sample was a jump. See Kropywiansky [22] for details. We instead approach this problem in the following less structural manner. First we take the trading day of the options that were used to invert for the jump expectations in a given month. Next, we estimate the volatility $\hat{\sigma}$ and drift $\hat{\mu}$ of the process over the next thirty days after that date using the time series data and given an assumption that the process is following a geometric brownian motion.[21] Next we call a one day move a "large movement" if:

$$|\Delta \log S_t - (\hat{\mu} - \frac{1}{2}\hat{\sigma}^2)\Delta t| > M * \hat{\sigma}\sqrt{\Delta t} \tag{2.17}$$

This is akin to a classical hypothesis test of whether a particular one day move came from a pure diffusion process. All it is asking is if a detrended one day movement

[21]The estimates are maximum likelihood estimates.

was more than M standard deviations away from zero. Next we sum the movements $\Delta \log S_t - (\hat{\mu} - \frac{1}{2}\hat{\sigma}^2)\Delta t$ that were determined to be "large movements." We will refer to this quantity as the total large movements over the thirty day period. We try three quantities for M, 2.0, 2.5 and 3.0. We plot the total "large movement" for the each month with M=2.5 for the DM in Figure 2-11 and for the yen in Figure 2-12. One can see the similarity of the time series plot of the total jump movement series for the DM to the time series plot for $\hat{\lambda}\hat{\kappa}$. There are generally positive jumps early on and negative jumps later on. The yen series is not as obvious although one sees that there have been more positive jump moves than negative ones.

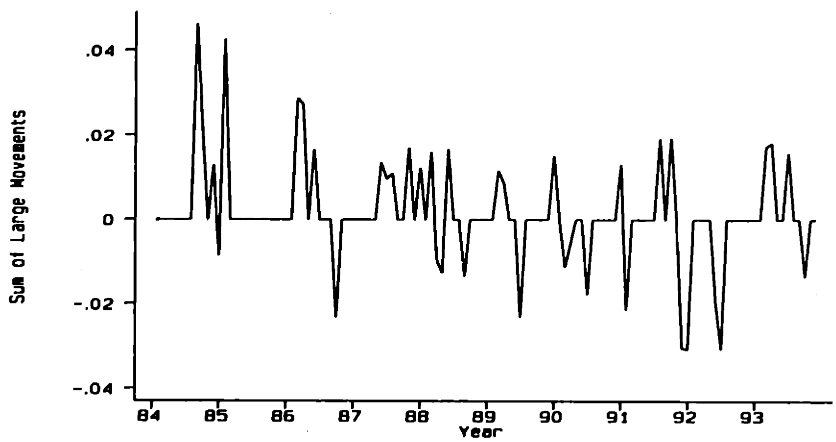

Figure 2-11: The sum of the DM "jump" movements in the 30 days following the day on which the option was traded

We now regress the sum of the "large movements" over the next 30 days on $\hat{\lambda}\hat{\kappa}$. The results are reported in Table 2.1 on page 79. Because of the possible concern that the post–June 1992 data is unrepresentative for the yen due to the thin trading in that period, we also report the numbers for the pre–June 1992 yen sample. The general result is that the although the sign is always correct, the coefficient is fairly small and the standard error large. We checked to see if these results were driven by outliers, we found that if anything there were outliers that worked against a positive slope coefficient. It is hard to conclude from this regression that the option implied jump expectations predict the large movements, but at the same time our positive coefficients are consistent with the hypothesis that they do.

Table 2.1: Regression of the sum of total jump movements on $\hat{\lambda}\hat{\kappa}$

	Slope	Intercept	% months with >0 jumps	avg num. jumps when >0 jumps
DM				
M=2.0	0.074 (0.079)	0.0006 (0.0018)	83 %	1.26
M=2.5	0.093 (0.049)	0.0016 (0.0011)	34 %	1.00
M=3.0	0.021 (0.029)	0.0001 (0.0006)	8 %	1.00
Yen				
M=2.0	0.033 (0.032)	0.0027 (0.0014)	72 %	1.29
M=2.5	0.036 (0.023)	0.0022 (0.0010)	33 %	1.00
M=3.0	0.004 (0.012)	0.0016 (0.0005)	10 %	1.00
Yen pre-6/92				
M=2.0	0.053 (0.056)	0.0029 (0.0015)	71 %	1.27
M=2.5	0.045 (0.041)	0.0022 (0.0011)	33 %	1.00
M=3.0	0.025 (0.023)	0.0015 (0.0006)	11 %	1.00

The total "large movement" for the 30 days following the option trading day used to invert for the jump expectation is regressed on the option implied jump expectations. We also report here the percentage of months that have at least one "jump" movement and the average number of jumps given that at least one occurred.

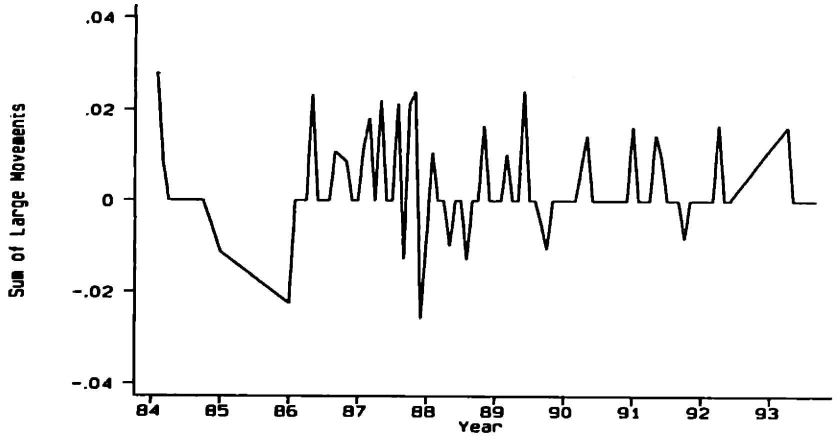

Figure 2-12: The sum of the yen "jump" movements in the 30 days following the day on which the option was traded

2.6 Jump Expectations, Purchasing Power Parity and the Trade Balance

We have seen that jump expectations are positively correlated with the interest differential in the case of the Deutschemark. It is interesting to see whether jump expectations are related to other variables such as the trade balance and the deviation of the currency from purchasing power parity for at least three reasons. First such an investigation is interesting in its own right since little is know about how jump expectations are formed. Second, one may be interested to determine what actually drives the yen jump expectations since we saw that the interest differential does not provide much guidance there. Third if the jump expectations are correlated with these variables in the way one would expect given a tendency to jump back to long run equilibrium, this will provide some independent reason to believe that our jump expectation estimates reflect market beliefs.

It is well known that currencies often trade at levels far away from their purchasing power parity levels. This is immediately made clear if one considers that the rate of change of the purchasing power parity implied exchange rate is equal to the difference in the inflation rates of the two countries in question. If the instantaneous inflation rate at time t in Germany is i_t^{DM} and i_t^{US} in the U.S., the purchasing power parity implied dollars per DM exchange rate S_t^{PPP} satisfies:

$$\log S_{t+1}^{PPP} - \log S_t^{PPP} = i_t^{US} - i_t^{DM} \tag{2.18}$$

But since the inflation rate in the countries we are considering have been fairly low and similar in the past ten year, the PPP rate cannot be moving very much. The actual dollar–DM and dollar–yen rate, however, has fluctuated quite drastically over the past 20 years. To construct the PPP implied exchange rate levels, we start with the 1988 estimate of the purchasing power parity level of the DM and yen estimated by Heston and Summers [18] and set that to the January 1988 PPP level. We construct estimates of the purchasing power parity exchange rate levels for the other months by updating with the consumer price index levels in the respective countries. The plot of the PPP rate for the DM along with the actual exchange rate is given in Figure 2-13 while the corresponding plot for the Yen is given in Figure 2-14.

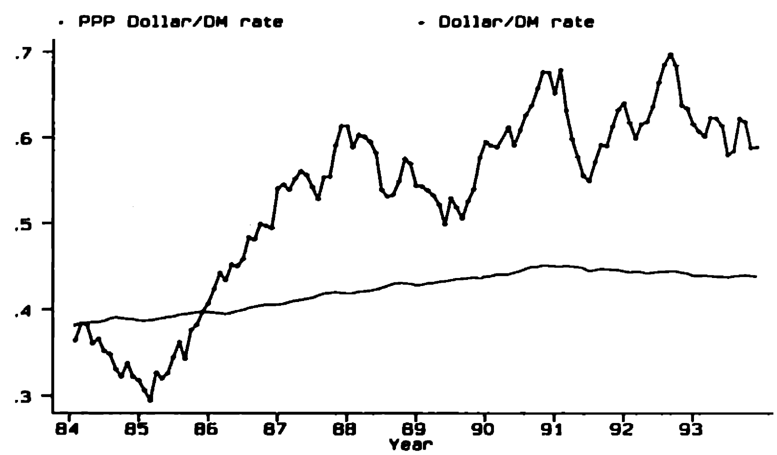

Figure 2-13: PPP Dollar/DM Exchange Rate and the actual Dollar/DM rate

As the graphs indicate, the dollar was trading close to PPP levels against the DM and yen in the beginning of 1984 and 1986. Relative to the dollar, the DM was undervalued relative to PPP levels between these two dates, but after 1986 steadily climbed. By late 1993, the DM was trading at roughly 150% of PPP levels. The case is even more dramatic for the yen. Between 1984 and 1986, the yen traded at roughly PPP levels, but has climbed dramatically since and by late 1993 was at roughly 210%

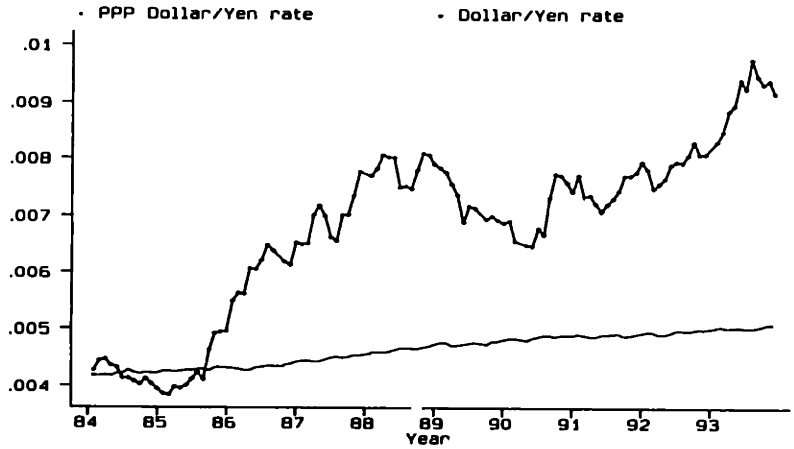

Figure 2-14: PPP Dollar/Yen Exchange Rate and the actual Dollar/Yen rate

PPP levels.[22]

It is of interest to see if traders expect the currency to jump in the direction of purchasing power parity. We regressed the jump expectations on the percentage overvaluation (the ratio of the actual exchange rate to the PPP implied exchange rate minus 1) of the respective currency. If one thought that jumps should be expected to return the currency toward PPP, the coefficient of this quantity should be negative—the greater the percentage markup over PPP, the more likely the currency is to jump down. For the DM, we found that the slope coefficient was -0.033 with an OLS standard error of 0.011. For the yen, the slope coefficient is -0.012 with an OLS standard error of 0.023. However, noting again the paucity of options data for the yen after June 1992 and the highly erratic behavior of the jump expectations thereafter, it is worth checking the result if we only use data prior to that date. In that case, the slope coefficient is -0.049 with an OLS standard error of 0.018. With both the yen and the DM, traders appear to think that a currency is more likely jump down if it is trading far above PPP levels. We can interpret our results as saying that if the DM is 50% overvalued relative to PPP, traders expect an annualized 1.65% jump depreciation of the DM.

Purchasing power parity is a long run equilibrium notion about the determination

[22]The yen has further appreciated since then and by mid–April 1995, was trading at 280% PPP levels.

of exchange rates based on the idea that the price of goods in two countries should be the same. Along the same lines, some have argued that in the long run, currencies will align themselves so that two countries will import as much as they export from a given trading partner. Whether one believes such an argument, there is no doubt that political pressures in the U.S. have leaned toward weakening the dollar against trading partners with whom the U.S. has trade deficits. This has been most apparent in the U.S. government's attempt to drive down the dollar against the yen over the past 10 years when record trade deficits were recorded with Japan. It is interesting to see how traders react to a higher trade deficit number. One would think that the more negative the trade balance, the more likely is the dollar–Yen rate to jump up. The bilateral trade deficit numbers are shown in Figure 2-15 for Germany and Figure 2-16 for Japan.

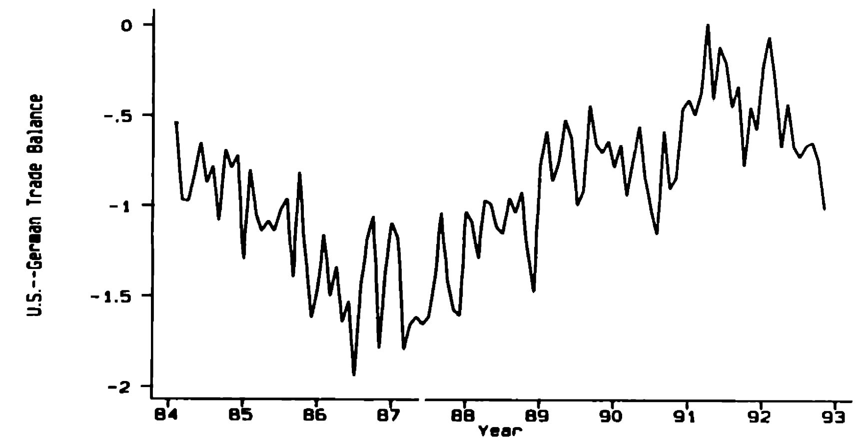

Figure 2-15: U.S.–German Monthly Trade Balance in Billions of 1990 Dollars

The pattern presented is roughly the same for Germany and Japan except that the trade deficit with Japan has been much larger. The U.S. trade deficit was growing for both countries until late 1987 after which the gap started to shrink. Although the deficit with Germany decreased steadily thereafter, the deficit with Japan started to worsen again in 1990. We regress the jump expectations on the trade deficits measured in Billions of 1990 U.S. dollars. For the DM we find a slope coefficient of -0.0075 with an OLS standard error of 0.005. For the yen using the whole sample we find a slope coefficient of 0.002 with an OLS standard error of 0.005. Using only the

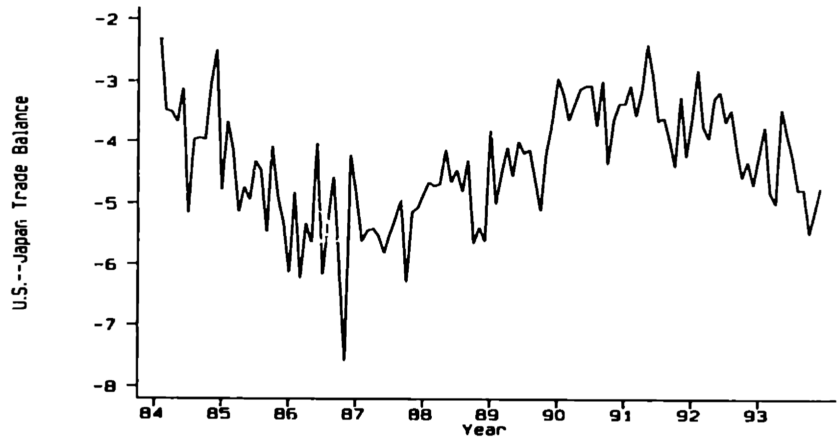

Figure 2-16: U.S.–Japan Monthly Trade Balance in Billions of 1990 Dollars

pre June 1992 data we get a slope coefficient of 0.0030 with an OLS standard error of 0.0027. In other words, the trade deficit does not seem to significantly impact jump expectations.

2.7 Conclusions

By using option prices, we confirmed the claim made by traders and proponents of the peso problem that the high yield currency is subject to negative jump fears in the case of the dollar–DM rate. We did not find any evidence to this effect for the dollar-yen rate. We further investigated whether the jump expectations inverted from option prices could predict jumps and found that it could do so if only weakly. We also found that when a currency trades far away from purchasing power parity levels, the currency is expected to jump in the direction of PPP. We however found that the trade balance did not affect jump expectations.

Table 2.2: Deutschemark Option Implied Parameters: 1/1984 - 6/1985

	$\hat{\lambda}\hat{\kappa}$	$\hat{\lambda}$	$\hat{\kappa}$	$\hat{\sigma}$	Obs.
1984: 1	.01484 (.00669)	.08719 (.07636)	.17016 (.07920)	.08613 (.00234)	90
1984: 2	Did Not Converge	Did Not Converge	Did Not Converge	Did Not Converge	186
1984: 3	.01498 (.00193)	.00476 (.01064)	3.14405 (6.68264)	.10575 (.00082)	232
1984: 4	-.00960 (.00315)	.04837 (.02140)	-.19847 (.02693)	.10107 (.00144)	98
1984: 5	-.02795 (.00804)	.21475 (.11855)	-.13017 (.03535)	.10552 (.00183)	74
1984: 6	Did Not Converge	Did Not Converge	Did Not Converge	Did Not Converge	50
1984: 7	.03088 (.00855)	.20895 (.10361)	.14778 (.03362)	.10959 (.00270)	66
1984: 8	.00617 (.00296)	.00863 (.02291)	.71437 (1.55910)	.12175 (.00120)	92
1984: 9	.01465 (.00464)	.05411 (.03883)	.27080 (.10986)	.13588 (.00163)	200
1984:10	.01362 (.00179)	.00027 (.00029)	51.30559 (61.07282)	.16473 (.00092)	180
1984:11	.04755 (.02218)	.46536 (.40677)	.10219 (.04227)	.12575 (.00408)	106
1984:12	.03054 (.00863)	.17536 (.08457)	.17418 (.03569)	.12805 (.00292)	70
1985: 1	.01180 (.00265)	.03493 (.02530)	.33777 (.17165)	.12397 (.00103)	178
1985: 2	.00613 (.00146)	.00059 (.00107)	10.46990 (17.15566)	.14278 (.00097)	161
1985: 3	.00925 (.00264)	.00023 (.00119)	39.69226 (194.66454)	.16270 (.00148)	108
1985: 4	.03016 (.01891)	.14128 (.18842)	.21344 (.15230)	.17445 (.00392)	208
1985: 5	-.00040 (.00249)	.00042 (.00058)	-.93878 (6.23759)	.18639 (.00112)	159
1985: 6	NA	NA	NA	NA	NA

Non-linear least squares regression results. The PHLX options data from June to November 1985 is completely contaminated and unusable.

Table 2.3: Deutschemark Option Implied Parameters: 7/1985 - 12/1986

	$\hat{\lambda}\hat{\kappa}$	$\hat{\lambda}$	$\hat{\kappa}$	$\hat{\sigma}$	Obs.
1985: 7	NA	NA	NA	NA	NA
1985: 8	NA	NA	NA	NA	NA
1985: 9	NA	NA	NA	NA	NA
1985:10	NA	NA	NA	NA	NA
1985:11	NA	NA	NA	NA	NA
1985:12	.00000	.00000	-.56572	.11402	52
	(.00163)	(.00299)	(>999.0)	(.00129)	
1986: 1	-.00043	.00135	-.31532	.12024	53
	(.00380)	(.01353)	(.34790)	(.00200)	
1986: 2	.00000	.00000	-.79293	.11568	72
	(.00278)	(.00350)	(39.26578)	(.00162)	
1986: 3	.02069	.00781	2.64869	.15837	80
	(.01839)	(.23834)	(78.56580)	(.00297)	
1986: 4	-.00403	.01936	-.20814	.16844	54
	(.00774)	(.05842)	(.24846)	(.00244)	
1986: 5	-.03729	.29739	-.12540	.16162	186
	(.01867)	(.29031)	(.06140)	(.00293)	
1986: 6	.00451	.00598	.75401	.15083	66
	(.01234)	(.10801)	(11.57873)	(.00318)	
1986: 7	-.00642	.02706	-.23729	.14274	142
	(.00282)	(.02083)	(.08263)	(.00109)	
1986: 8	.00953	.02432	.39158	.11554	104
	(.02321)	(.48734)	(6.91109)	(.00375)	
1986: 9	-.00023	.00027	-.84269	.14148	74
	(.00323)	(.00479)	(3.07188)	(.00116)	
1986:10	.00531	.00134	3.95616	.11797	135
	(.00137)	(.00324)	(9.00849)	(.00066)	
1986:11	-.00024	.00032	-.76092	.11013	120
	(.00149)	(.00068)	(4.30859)	(.00086)	
1986:12	.00849	.01332	.63757	.09958	54
	(.00722)	(.08385)	(3.48390)	(.00221)	

Non-linear least squares regression results.

Table 2.4: Deutschemark Option Implied Parameters: 1/1987 - 6/1988

	$\hat{\lambda}\hat{\kappa}$	$\hat{\lambda}$	$\hat{\kappa}$	$\hat{\sigma}$	Obs.
1987: 1	.00000 (.00135)	.00000 (.00143)	-.94836 (>999.0)	.15024 (.00100)	114
1987: 2	-.00001 (.00371)	.00001 (.00547)	-.83040 (79.36450)	.13676 (.00158)	68
1987: 3	.01001 (.01198)	.00317 (.26659)	3.15459 (261.41756)	.09536 (.00232)	64
1987: 4	.03031 (.02227)	.26422 (.42957)	.11470 (.10268)	.10875 (.00415)	74
1987: 5	-.00180 (.00221)	.00514 (.00778)	-.35018 (.10603)	.10106 (.00094)	64
1987: 6	-.00349 (.00473)	.01275 (.02132)	-.27387 (.09612)	.10354 (.00138)	106
1987: 7	-.00092 (.01066)	.00372 (.04996)	-.24731 (.46636)	.08463 (.00340)	96
1987: 8	.00772 (.01006)	.00183 (.24168)	4.22075 (552.29952)	.08862 (.00196)	59
1987: 9	.00000 (.00227)	.00000 (.00541)	-.43907 (177.24395)	.10107 (.00081)	85
1987:10	.06731 (.03825)	.96552 (.90560)	.06972 (.02595)	.09548 (.00660)	207
1987:11	.01234 (.01199)	.02030 (.19690)	.60778 (5.31675)	.12996 (.00248)	160
1987:12	.00000 (.00349)	.00000 (.00407)	-.87268 (550.46510)	.13706 (.00177)	56
1988: 1	.00830 (.01296)	.00359 (.16189)	2.30816 (100.44408)	.15023 (.00305)	75
1988: 2	-.00222 (.00584)	.00707 (.02299)	-.31387 (.20471)	.11247 (.00157)	124
1988: 3	-.00750 (.00351)	.03821 (.02674)	-.19636 (.04943)	.11763 (.00097)	54
1988: 4	.00451 (.00991)	.00325 (.16842)	1.38643 (68.78617)	.09590 (.00253)	68
1988: 5	-.00795 (.01452)	.12253 (.37374)	-.06487 (.08213)	.08375 (.00242)	64
1988: 6	.00141 (.01726)	.00151 (.43298)	.93783 (258.00673)	.07634 (.00349)	156

Non-linear least squares regression results.

Table 2.5: Deutschemark Option Implied Parameters: 7/1988 - 12/1989

	$\hat{\lambda}\hat{\kappa}$	$\hat{\lambda}$	$\hat{\kappa}$	$\hat{\sigma}$	Obs.
1988: 7	-.00573 (.00249)	.02338 (.01414)	-.24522 (.04644)	.11556 (.00105)	114
1988: 8	.00000 (.00271)	.00000 (.00335)	-.82776 (150.01432)	.13646 (.00101)	80
1988: 9	-.00240 (.02759)	.03317 (.58711)	-.07240 (.45082)	.09998 (.00487)	60
1988:10	.01577 (.01625)	.00641 (.37136)	2.46070 (140.05107)	.10255 (.00269)	112
1938:11	-.00065 (.00357)	.00271 (.01715)	-.23887 (.19786)	.11758 (.00117)	81
1988:12	-.00276 (.00389)	.00816 (.01395)	-.33858 (.10774)	.10791 (.00114)	80
1989: 1	-.00002 (.00571)	.00003 (.01053)	-.57049 (10.05064)	.12280 (.00142)	63
1989: 2	-.00113 (.00486)	.00412 (.02036)	-.27480 (.18302)	.10615 (.00124)	102
1989: 3	-.01204 (.00548)	.09562 (.06818)	-.12589 (.03608)	.08693 (.00163)	120
1989: 4	-.00004 (.00103)	.00004 (.00104)	-.98091 (.33548)	.08637 (.00052)	84
1989: 5	-.00345 (.00274)	.01170 (.01142)	-.29524 (.06005)	.10332 (.00084)	118
1989: 6	-.01138 (.00446)	.02998 (.01884)	-.37967 (.09755)	.14726 (.00120)	106
1989: 7	-.01358 (.01911)	.12088 (.32964)	-.11238 (.15023)	.13891 (.00280)	56
1989: 8	-.00694 (.00375)	.01366 (.01040)	-.50775 (.12397)	.13919 (.00110)	104
1989: 9	.00766 (.01267)	.01693 (.18861)	.45231 (4.29296)	.14053 (.00222)	160
1989:10	.01990 (.01263)	.05569 (.17253)	.35743 (.88607)	.15820 (.00195)	208
1989:11	.00514 (.00326)	.01170 (.03926)	.43905 (1.21001)	.11777 (.00078)	111
1989:12	-.00956 (.00926)	.00984 (.01054)	-.97222 (.11016)	.10422 (.00319)	62

Non-linear least squares regression results.

Table 2.6: Deutschemark Option Implied Parameters: 1/1990 - 6/1991

	$\hat{\lambda}\hat{\kappa}$	$\hat{\lambda}$	$\hat{\kappa}$	$\hat{\sigma}$	Obs.
1990: 1	-.00338 (.00289)	.00756 (.00897)	-.44744 (.16802)	.12941 (.00080)	171
1990: 2	-.01102 (.02397)	.11507 (.51201)	-.09578 (.21991)	.10894 (.00430)	57
1990: 3	.14049 (.04097)	1.62874 (1.23116)	.08626 (.04114)	.07188 (.00827)	104
1990: 4	-.00205 (.00191)	.00260 (.00132)	-.78637 (.90770)	.09786 (.00060)	50
1990: 5	-.00536 (.00344)	.01969 (.02479)	-.27206 (.18080)	.09711 (.00108)	90
1990: 6	-.02970 (.01229)	.31751 (.25365)	-.09353 (.03609)	.08474 (.00281)	124
1990: 7	-.00939 (.00258)	.01104 (.00454)	-.85019 (.13549)	.09320 (.00091)	63
1990: 8	-.01010 (.00575)	.06434 (.09625)	-.15697 (.15066)	.10858 (.00120)	79
1990: 9	.01908 (.02156)	.16466 (.42718)	.11587 (.17017)	.12206 (.00346)	132
1990:10	.00195 (.02960)	.01902 (.52756)	.10265 (1.29871)	.12339 (.00469)	57
1990:11	.00613 (.00268)	.00667 (.00820)	.91910 (.81890)	.10769 (.00075)	87
1990:12	-.01488 (.02426)	.07115 (.65924)	-.20918 (1.60055)	.09591 (.00358)	56
1991: 1	-.02524 (.01286)	.12348 (.18296)	-.20445 (.20026)	.15027 (.00248)	118
1991: 2	-.04675 (.07526)	1.20746 (3.05838)	-.03872 (.03600)	.09900 (.00749)	165
1991: 3	.00254 (.00241)	.00296 (.00386)	.85849 (.34841)	.13413 (.00103)	304
1991: 4	.00099 (.00336)	.00100 (.00425)	.99304 (.88923)	.13096 (.00115)	149
1991: 5	-.00687 (.00949)	.02617 (.16277)	-.26240 (1.27403)	.12813 (.00186)	184
1991: 6	-.01243 (.00865)	.07889 (.13996)	-.15756 (.17925)	.12677 (.00178)	73

Non-linear least squares regression results.

Table 2.7: Deutschemark Option Implied Parameters: 7/1991 - 12/1992

	$\hat{\lambda}\hat{\kappa}$	$\hat{\lambda}$	$\hat{\kappa}$	$\hat{\sigma}$	Obs.
1991: 7	.03408	.58149	.05860	.12208	101
	(.06329)	(1.76487)	(.06938)	(.00693)	
1991: 8	.00000	.00000	3.59082	.16069	233
	(.00198)	(.00062)	(756.39840)	(.00083)	
1991: 9	.00516	.02604	.19819	.13138	91
	(.00497)	(.03771)	(.10481)	(.00114)	
1991:10	.00002	.00001	1.96951	.12261	152
	(.00398)	(.00229)	(47.61608)	(.00127)	
1991:11	-.00556	.01114	-.49923	.12680	206
	(.00775)	(.11378)	(4.43558)	(.00190)	
1991:12	.00002	.00001	3.92008	.12889	87
	(.01118)	(.00330)	(350.78826)	(.00296)	
1992: 1	.00003	.00003	.94545	.15158	126
	(.00256)	(.00304)	(10.21535)	(.00101)	
1992: 2	.00000	.00000	2.36471	.16069	286
	(.00272)	(.00126)	(>999.0)	(.00096)	
1992: 3	.00000	.00000	1.06865	.13121	151
	(.00198)	(.00199)	(>999.0)	(.00069)	
1992: 4	-.05395	.74915	-.07202	.10904	122
	(.06067)	(1.47157)	(.06089)	(.00918)	
1992: 5	.00000	.00000	.71729	.11186	127
	(.00501)	(.00758)	(656.56762)	(.00142)	
1992: 6	-.01702	.34640	-.04913	.10228	62
	(.05288)	(1.71758)	(.09185)	(.00650)	
1992: 7	-.00791	.04123	-.19183	.13790	151
	(.01211)	(.18133)	(.55535)	(.00251)	
1992: 8	.00397	.01467	.27059	.11351	68
	(.00363)	(.01797)	(.08844)	(.00109)	
1992: 9	.00000	.00000	2.49346	.17208	482
	(.00319)	(.00140)	(>999.0)	(.00157)	
1992:10	-.16642	2.52705	-.06585	.17717	83
	(.25648)	(6.07613)	(.05708)	(.02382)	
1992:11	.00000	.00000	1.27870	.14722	121
	(.00300)	(.00258)	(333.24275)	(.00132)	
1992:12	-.00691	.01103	-.62631	.14696	74
	(.01379)	(.21915)	(11.20695)	(.00248)	

Non-linear least squares regression results.

Table 2.8: Deutschemark Option Implied Parameters: 1/1993 - 12/1993

	$\hat{\lambda}\hat{\kappa}$	$\hat{\lambda}$	$\hat{\kappa}$	$\hat{\sigma}$	Obs.
1993: 1	.00000 (.00329)	.00000 (.00194)	1.87472 (>999.0)	.12674 (.00154)	89
1993: 2	.00021 (.00554)	.00010 (.00296)	2.19512 (9.96525)	.12529 (.00215)	138
1993: 3	.00374 (.00320)	.01450 (.01610)	.25814 (.07016)	.11871 (.00101)	75
1993: 4	.00000 (.00365)	.00000 (.00182)	2.18133 (>999.0)	.12791 (.00122)	77
1993: 5	.C1398 (.00732)	.10291 (.09498)	.13582 (.05920)	.11908 (.00171)	71
1993: 6	.00000 (.00255)	.00000 (.00149)	1.85153 (>999.0)	.11491 (.00112)	69
1993: 7	.00000 (.00292)	.00000 (.00198)	1.59380 (>999.0)	.11363 (.00094)	95
1993: 8	-.05959 (.34878)	2.05367 (18.75430)	-.02902 (.09534)	.10670 (.02281)	116
1993: 9	-.02495 (.02896)	.19716 (.48754)	-.12654 (.17134)	.12706 (.00612)	75
1993:10	.00003 (.00382)	.00001 (.00243)	1.78686 (34.71422)	.11858 (.00132)	82
1993:11	.00000 (.00289)	.00000 (.00133)	2.35403 (>999.0)	.10579 (.00246)	53
1993:12	.00000 (.00273)	.00000 (.00252)	1.16309 (>999.0)	.10678 (.00105)	74

Non-linear least squares regression results.

Table 2.9: Yen Option Implied Parameters: 1/1984 - 6/1985

	$\hat{\lambda}\hat{\kappa}$	$\hat{\lambda}$	$\hat{\kappa}$	$\hat{\sigma}$	Obs.
1984: 1	.08212 (.13291)	3.15364 (7.79196)	.02604 (.02223)	.05810 (.01428)	56
1984: 2	-.00742 (.00262)	.04428 (.02426)	-.16765 (.05233)	.06398 (.00148)	56
1984: 3	.01694 (.00593)	.00312 (.01137)	5.42420 (18.81884)	.09149 (.00355)	144
1984: 4	.03663 (.03501)	.74849 (1.17786)	.04894 (.03048)	.07520 (.00534)	72
1984: 5	.02739 (.01553)	.39900 (.42303)	.06864 (.03512)	.06815 (.00334)	68
1984: 6	.01047 (.01165)	.11341 (.22428)	.09232 (.08069)	.08142 (.00265)	72
1984: 7	.00370 (.00472)	.00081 (.05402)	4.56073 (298.09819)	.09071 (.00177)	52
1984: 8	.01572 (.01365)	.03255 (.24142)	.48312 (3.17482)	.09501 (.00330)	56
1984: 9	.12838 (.00364)	1.97365 (.12750)	.06505 (.00241)	.02369 (.00599)	58
1984:10	.05890 (.07608)	.87614 (1.84122)	.06722 (.05477)	.10204 (.01177)	57
1984:11	Insufficient Data	Insufficient Data	Insufficient Data	Insufficient Data	0
1984:12	.00266 (.00254)	.00078 (.00597)	3.41530 (23.13685)	.09489 (.00101)	66
1985: 1	.01407 (.02233)	.20630 (.53834)	.06819 (.07006)	.09069 (.00383)	54
1985: 2	Insufficient Data	Insufficient Data	Insufficient Data	Insufficient Data	0
1985: 3	Did Not Converge	Did Not Converge	Did Not Converge	Did Not Converge	56
1985: 4	Insufficient Data	Insufficient Data	Insufficient Data	Insufficient Data	0
1985: 5	Insufficient Data	Insufficient Data	Insufficient Data	Insufficient Data	0
1985: 6	NA	NA	NA	NA	NA

Non-linear least squares regression results. The PHLX options data from June to November 1985 is completely contaminated and unusable.

Table 2.10: Yen Option Implied Parameters: 7/1985 - 12/1986

	$\hat{\lambda}\hat{\kappa}$	$\hat{\lambda}$	$\hat{\kappa}$	$\hat{\sigma}$	Obs.
1985: 7	NA	NA	NA	NA	NA
1985: 8	NA	NA	NA	NA	NA
1985: 9	NA	NA	NA	NA	NA
1985:10	NA	NA	NA	NA	NA
1985:11	NA	NA	NA	NA	NA
1985:12	.00520 (.00224)	.00088 (.00242)	5.94526 (14.62967)	.07909 (.00111)	73
1986: 1	.00690 (.00269)	.00179 (.00892)	3.85742 (18.2[illegible]340)	.06842 (.00130)	67
1986: 2	.02962 (.02348)	.08305 (.54621)	.35659 (2.06332)	.10324 (.00367)	183
1986: 3	.02500 (.00693)	.00990 (.05928)	2.52690 (14.77603)	.13005 (.00205)	86
1986: 4	.01261 (.01688)	.01837 (.30525)	.68611 (10.49026)	.11860 (.00295)	61
1986: 5	.01135 (.00834)	.00425 (.06009)	2.66946 (35.90711)	.18439 (.00192)	118
1986: 6	.01659 (.01960)	.06838 (.32138)	.24255 (.85664)	.13759 (.00336)	70
1986: 7	.01798 (.01115)	.00943 (.13700)	1.90733 (26.56515)	.13295 (.00239)	84
1986: 8	.00000 (.00055)	.00000 (.00084)	-.67848 (640.98433)	.12405 (.00051)	94
1986: 9	.00786 (.01193)	.00838 (.15237)	.93775 (15.66880)	.13427 (.00186)	59
1986:10	-.00255 (.00337)	.01015 (.01746)	-.25093 (.12168)	.07521 (.00115)	50
1986:11	-.00198 (.00363)	.00289 (.00860)	-.68369 (.84117)	.08195 (.00125)	81
1986:12	.03041 (.02630)	.34378 (.52862)	.08845 (.06020)	.07022 (.00674)	66

Non-linear least squares regression results.

Table 2.11: Yen Option Implied Parameters: 1/1987 - 6/1988

	$\hat{\lambda}\hat{\kappa}$	$\hat{\lambda}$	$\hat{\kappa}$	$\hat{\sigma}$	Obs.
1987: 1	.01696 (.00434)	.00298 (.02693)	5.69500 (50.13680)	.12558 (.00138)	67
1987: 2	.00000 (.00641)	.00001 (.01365)	-.49415 (56.30940)	.08749 (.00321)	51
1987: 3	.01201 (.00950)	.07528 (.13427)	.15953 (.15891)	.08354 (.00309)	63
1987: 4	.00000 (.00237)	.00000 (.00312)	-.77584 (>999.0)	.13402 (.00128)	71
1987: 5	.00000 (.00540)	.00001 (.00734)	-.75898 (29.84215)	.11185 (.00171)	57
1987: 6	-.01554 (.00773)	.08151 (.06710)	-.19066 (.07090)	.09712 (.00214)	95
1987: 7	.00384 (.00802)	.00232 (.12980)	1.65217 (88.93422)	.09550 (.00199)	87
1987: 8	-.01149 (.00454)	.06805 (.04984)	-.16882 (.06261)	.10124 (.00135)	114
1987: 9	.00954 (.00927)	.03510 (.10919)	.27164 (.58256)	.10827 (.00296)	61
1987:10	.02150 (.03350)	.00953 (.59411)	2.25621 (137.15446)	.13709 (.00531)	97
1987:11	-.00629 (.00385)	.01649 (.01440)	-.38148 (.11520)	.12894 (.00146)	71
1987:12	.00000 (.00217)	.00000 (.00321)	-.69553 (201.59369)	.12691 (.00123)	87
1988: 1	.00786 (.01487)	.00219 (.20083)	3.59400 (323.39196)	.14967 (.00317)	116
1988: 2	-.00839 (.00725)	.07202 (.10621)	-.11656 (.07669)	.11009 (.00159)	61
1988: 3	.00784 (.01467)	.00614 (.24201)	1.27671 (47.95901)	.11899 (.00230)	58
1988: 4	.00954 (.01054)	.05471 (.2539[illegible])	.17431 (.61913)	.07569 (.00256)	70
1988: 5	-.00521 (.00303)	.03709 (.03411)	-.14055 (.05042)	.08210 (.00071)	97
1988: 6	-.00588 (.00365)	.06010 (.06369)	-.09777 (.04470)	.07180 (.00094)	59

Non-linear least squares regression results.

Table 2.12: Yen Option Implied Parameters: 7/1988 - 12/1989

	$\hat{\lambda}\hat{\kappa}$	$\hat{\lambda}$	$\hat{\kappa}$	$\hat{\sigma}$	Obs.
1988: 7	.00173	.00054	3.22901	.11724	52
	(.00729)	(.07713)	(450.48254)	(.00192)	
1988: 8	.00473	.00877	.53974	.11456	123
	(.01371)	(.25469)	(14.13426)	(.00256)	
1988: 9	.06478	.03314	1.95459	.12064	96
	(.08370)	(3.88565)	(226.64914)	(.00599)	
1988:10	.00773	.00394	1.96253	.09211	86
	(.03057)	(.82853)	(405.14342)	(.00491)	
1988:11	.00658	.00933	.70476	.10909	95
	(.00722)	(.12720)	(8.84578)	(.00149)	
1988:12	-.01910	.24786	-.07705	.09135	88
	(.01099)	(.25040)	(.03447)	(.00175)	
1989: 1	-.01900	.25313	-.07506	.10074	68
	(.01699)	(.39895)	(.05172)	(.00238)	
1989: 2	.00000	.00000	-.44272	.08915	117
	(.00059)	(.00139)	(>999.0)	(.00040)	
1989: 3	-.02360	.29758	-.07931	.08171	94
	(.01414)	(.31677)	(.03721)	(.00294)	
1989: 4	.00598	.01189	.50306	.07589	64
	(.00629)	(.18072)	(7.12420)	(.00115)	
1989: 5	-.00374	.02721	-.13745	.08998	63
	(.00333)	(.03179)	(.04029)	(.00099)	
1989: 6	.00534	.00125	4.28192	.13639	85
	(.03231)	(.53078)	(>999.0)	(.00596)	
1989: 7	-.06112	1.24579	-.04906	.11725	111
	(.09608)	(2.91522)	(.03774)	(.01090)	
1989: 8	-.04394	.75621	-.05811	.12563	58
	(.06987)	(1.93430)	(.05651)	(.00748)	
1989: 9	-.00001	.00001	-.67954	.13282	87
	(.00490)	(.00750)	(15.19774)	(.00125)	
1989:10	.01087	.01270	.85611	.14775	78
	(.02830)	(.42155)	(26.21976)	(.00415)	
1989:11	.01156	.07254	.15942	.09635	67
	(.00225)	(.03561)	(.04911)	(.00075)	
1989:12	.00997	.08570	.11638	.08039	57
	(.01308)	(.20233)	(.12369)	(.00469)	

Non-linear least squares regression results.

Table 2.13: Yen Option Implied Parameters: 1/1990 - 6/1991

	$\hat{\lambda}\hat{\kappa}$	$\hat{\lambda}$	$\hat{\kappa}$	$\hat{\sigma}$	Obs.
1990: 1	-.00950	.07633	-.12446	.07486	100
	(.00343)	(.05173)	(.04025)	(.00094)	
1990: 2	.00000	.00000	-.87911	.07794	69
	(.00027)	(.00031)	(>999.0)	(.00039)	
1990: 3	.00346	.00634	.54601	.09514	65
	(.00068)	(.00400)	(.27202)	(.00050)	
1990: 4	-.01859	.19674	-.09451	.11107	56
	(.05183)	(1.25051)	(.33883)	(.00598)	
1990: 5	.01043	.03152	.33088	.08906	86
	(.01193)	(.26171)	(2.37447)	(.00229)	
1990: 6	.01308	.07740	.16893	.08321	56
	(.00928)	(.23832)	(.40099)	(.00190)	
1990: 7	.01283	.04299	.29853	.09446	50
	(.01682)	(.40429)	(2.41862)	(.00230)	
1990: 8	-.00010	.00013	-.78554	.10660	108
	(.00301)	(.00304)	(4.82423)	(.00099)	
1990: 9	-.01497	.15679	-.09550	.12318	66
	(.02092)	(.40698)	(.11601)	(.00372)	
1990:10	.00012	.00008	1.64992	.13129	57
	(.00243)	(.00176)	(6.38384)	(.00108)	
1990:11	.00798	.01273	.62663	.12692	59
	(.00501)	(.02515)	(.88381)	(.00118)	
1990:12	-.06525	1.18222	-.05519	.10876	52
	(.05253)	(1.48748)	(.02558)	(.00699)	
1991: 1	-.00701	.03341	-.20997	.14991	98
	(.00399)	(.03971)	(.13074)	(.00111)	
1991: 2	.00976	.10095	.09663	.11124	67
	(.02630)	(.56309)	(.28090)	(.00365)	
1991: 3	.00020	.00007	2.81099	.12399	82
	(.00263)	(.00116)	(8.97184)	(.00107)	
1991: 4	.04602	.84656	.05436	.11721	85
	(.09270)	(2.78979)	(.06987)	(.00980)	
1991: 5	.03627	.47701	.07605	.09150	59
	(.02345)	(.71177)	(.06458)	(.00320)	
1991: 6	.00619	.01847	.33513	.09635	62
	(.00412)	(.02294)	(.20759)	(.00107)	

Non-linear least squares regression results.

Table 2.14: Yen Option Implied Parameters: 7/1991 - 12/1992

	$\hat{\lambda}\hat{\kappa}$	$\hat{\lambda}$	$\hat{\kappa}$	$\hat{\sigma}$	Obs.
1991: 7	-.00370 (.00638)	.01282 (.11647)	-.28863 (2.13481)	.09232 (.00150)	50
1991: 8	-.02388 (.01700)	.29460 (.36249)	-.08106 (.04289)	.08697 (.00350)	62
1991: 9	.01572 (.01910)	.00899 (.04433)	1.74951 (6.83855)	.11575 (.00297)	106
1991:10	.01264 (.00891)	.10296 (.17847)	.12274 (.13266)	.08400 (.00175)	50
1991:11	.00001 (.00441)	.00002 (.00774)	.62737 (27.37423)	.09461 (.00136)	110
1991:12	.01637 (.00350)	.02120 (.01537)	.77220 (.41439)	.07605 (.00090)	77
1992: 1	.02642 (.00557)	.07248 (.04426)	.36449 (.16151)	.11287 (.00135)	67
1992: 2	.01677 (.01418)	.08418 (.23841)	.19923 (.41017)	.08618 (.00249)	53
1992: 3	.06118 (.04152)	1.46273 (1.60499)	.04182 (.01767)	.08236 (.00536)	60
1992: 4	-.00999 (.01806)	.08510 (.54355)	-.11742 (.53892)	.08291 (.00297)	63
1992: 5	.05827 (.08929)	1.07681 (3.22003)	.05411 (.07956)	.09023 (.01050)	78
1992: 6	-.23637 (.60844)	11.46564 (44.50849)	-.02062 (.02708)	.06379 (.05437)	79
1992: 7	Insufficient Data	Insufficient Data	Insufficient Data	Insufficient Data	0
1992: 8	Insufficient Data	Insufficient Data	Insufficient Data	Insufficient Data	0
1992: 9	.00147 (.01320)	.00020 (.00344)	7.36968 (60.87501)	.10741 (.00513)	91
1992:10	Insufficient Data	Insufficient Data	Insufficient Data	Insufficient Data	0
1992:11	Insufficient Data	Insufficient Data	Insufficient Data	Insufficient Data	0
1992:12	Insufficient Data	Insufficient Data	Insufficient Data	Insufficient Data	0

Non-linear least squares regression results.

Table 2.15: Yen Option Implied Parameters: 1/1993 - 12/1993

	$\hat{\lambda}\hat{\kappa}$	$\hat{\lambda}$	$\hat{\kappa}$	$\hat{\sigma}$	Obs.
1993: 1	Insufficient Data	Insufficient Data	Insufficient Data	Insufficient Data	0
1993: 2	Insufficient Data	Insufficient Data	Insufficient Data	Insufficient Data	0
1993: 3	.28880 (.18452)	9.98507 (11.88631)	.02892 (.01639)	.11136 (.01132)	50
1993: 4	.18450 (.08072)	3.49497 (3.53775)	.05279 (.03157)	.13775 (.00567)	73
1993: 5	.11460 (.35797)	.09908 (21.53421)	1.15672 (247.81432)	.14630 (.01511)	51
1993: 6	Insufficient Data	Insufficient Data	Insufficient Data	Insufficient Data	0
1993: 7	-.06242 (.11364)	.23227 (2.84634)	-.26875 (2.81383)	.16269 (.01232)	51
1993: 8	.02087 (.09179)	.06251 (1.76734)	.33388 (7.98880)	.19786 (.00852)	52
1993: 9	.01038 (.01497)	.00430 (.13419)	2.41485 (72.12596)	.16930 (.00188)	56
1993:10	Insufficient Data	Insufficient Data	Insufficient Data	Insufficient Data	0
1993:11	-.02878 (.01131)	.03301 (.04712)	-.87190 (.94528)	.10029 (.00157)	64
1993:12	Insufficient Data	Insufficient Data	Insufficient Data	Insufficient Data	0

Non-linear least squares regression results.

Bibliography

[1] Akgiray, V. and Booth, G.G., "Mixed Diffusion–Jump Process Modeling of Exchange Rate Movements," *Review of Economics and Statistics*, 1988,**70**, 638–647.

[2] Backus, D., Gregory, A. and Telmer, C., "Accounting for Forward Rates in Markets for Foreign Currency," *Journal of Finance*, 1993, **48**, 1887–1908.

[3] Barone–Adesi, G. and Whaley, R. E., "Efficient Analytic Approximation of American Option Values," *Journal of Finance*, 1987, **42**, 301–320.

[4] Bates, D., "The Crash Premium: Option Pricing Under Asymmetric Processes, with Applications to Options on Deutschemark Futures," Working Paper 36–88, University of Pennsylvania, Rodney L. White Center, 1988.

[5] Bates, D., "Crashes, Options, and International Asset Substitutability," Princeton University Ph.D. dissertation, January 1989.

[6] Bates, D., "The Crash of 87: Was It Expected? The Evidence from Option Markets," *Journal of Finance*, 1991, **46**, 1009–1044.

[7] Bates, D., "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in PHLX Deutschemark Options," National Bureau of Economic Research Working Paper No. 4596, December 1993.

[8] Bates, D., "Distributional Abnormalities Implicit in Foreign Currency Futures Options," University of Pennsylvania, Working Paper, August 1994.

[9] Bekaert, G. and Hodrick, R.J., "On biases in the measurement of foreign exchange risk premiums," *Journal of International Money and Finance*, 1993, **12**, 115–138.

[10] Bodurtha, J. and Courtadon, G., "Tests of the American Option Pricing Model in the Foreign Currency Option Market," *Journal of Financial and Quantitative Analysis*, 1987, **22**, 153–167.

[11] Cox, J., Ingersoll, J., Ross, S., "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, 1985,**36**, 363–384.

[12] Dornbusch, R., "Why the Mighty Mark is a Sitting Duck," *Business Week*, March 7, 1994, **30**, page 30.

[13] Evans, M.D.D., and Lewis, K., "Do Long–Term Swings in the Dollar Affect Estimates of the Risk Premium?" forthcoming *Review of Financial Studies.*

[14] Fama, E.F., "Forward and Spot Exchange Rates," *Journal of Monetary Economics*, 1984, **14**, 319–338.

[15] Frankel, J. and Froot, J., "The Forward Discount Bias: Is it an Exchange Risk Premium?" *Quarterly Journal of Economics*, 1989, **104**, 139-161.

[16] Froot, K., "Short Rates and Expected Asset Returns," NBER Working paper No. 3247, January 1990.

[17] Froot, K.A. and Thaler, R., "Anomalies: Foreign Exchange," *Journal of Economic Perspectives,* 1990, **4**, 179–192.

[18] Heston, A. and Summers, R., "Penn World Tables V," *Quarterly Journal of Economics*, 1991, **106**, 327–368.

[19] Jorion, P., "On Jump Processes in the Foreign Exchange and Stock Markets," *Review of Financial Studies*, 1988, **1**, 427–445.

[20] Jones, E.P., "Options Arbitrage and Strategy with Large Price Changes," *Journal of Financial Economics*, 1984, **13**, 91–113.

[21] Krasker, W., "The 'Peso Problem' in Testing the Efficiency of Forward Exchange Markets," *Journal of Monetary Economics,* 1980, **6**, 269–276.

[22] Kropywiansky, L., "Interventions and Jumps in the Dollar Exchange Rate," 1995 MIT working paper.

[23] Lewis, K., "Was There a 'Peso Problem' in the U.S. Term Structure of Interest Rates: 1979–1982?," *International Economic Review*, 1991, **32**, 159–173.

[24] Lewis, K., "Can Learning Affect Exchange–Rate Behavior? The Case of the Dollar in he Early 1980's," *Journal of Monetary Economics*, 1989, **23**, 79–100.

[25] Lewis, K., "Peso Problems," in J.Eatwell, M.Milgate, and P. Newman eds., *The New Palgrave Dictionary of Money and Finance,* Macmillan Press, London.

[26] Lewis, K., "Puzzles in International Financial Markets," Mimeo, The University of Pennsylvania, 1994.

[27] Lizondo, J.S., "Foreign Exchange Futures Prices Under Fixed Exchange Rates," *Journal of International Economics*, 1983, **14**, 69–84.

[28] Lyons, R., "Tests of the Foreign Exchange Risk Premium Using the Expected Second Moments Implied by Option Pricing," *Journal of International Money and Finance*, 1988, **7**, 91–108.

[29] MacMillan, L.W., "Analytic Approximation for the American Put," *Advances in Futures and Options Research*, 1986, **1**, 119–139.

[30] Merton, R., "Option Pricing When Underlying Stock Returns are Discontinuous," *Journal of Financial Economics*, 1976, **3**, 125–144.

[31] Mussa, M., "Empirical Regularities in the Behavior of Exchange Rates and Theories of the Foreign Exchange Market." In Brunner, K. and A.H.Meltzer, eds. *Policies for Employment Prices and Exchange Rates,* Vol 11. Carnegie--Rochester Conference Series on Public Policy, supplement to the *Journal of Monetary Economics*, 1979, 9-57.

[32] Reitz, T., "The Equity Risk Premium: A Solution," *Journal of Monetary Economics*, 1989, **22**, 117–131.

[33] Rogoff, K., *Essays on the Expectations and Exchange Rate Volatility,* Ph.D. Dissertation, Department of Economics, Massachusetts Institute of Technology, Cambridge, MA, 1979.

[34] Tucker, A.L. and Pond, L., "The Probability Distribution of Foreign Exchange Price Changes: Tests of Candidate Processes," *Review of Economics and Statistics*, 1988, **70**, 638–647.

Chapter 3

Financing Real Investments With Evolving Private Information

3.1 Introduction

Much of the recent response to Modigliani and Miller's proposition that financial structure does not affect a firm's value has made use of the assumption that the firm has information that the capital market does not. Whether it be the true value of the firm's assets in place or the profitability of a project in consideration, the implicit understanding behind these recent papers is that there is some aspect of the firm's operations that only managers can know. This leads to numerous interesting situations where managers may forgo profitable investments or undertake inefficient ones because the capital market cannot asses the true value of the securities issued by the firm that are needed to finance new projects. The asymmetry of information often makes it impossible for the manager and the suppliers of capital to come to an efficient solution which will induce the managers to invest optimally. A point that is mostly ignored, however, is the fact that the private information held by the manager is often changing, sometimes even very rapidly, over time. A firm's prospective project may appear to the managers as a dud one year, but become an attractive venture the next. A proposed area of expansion for the firm may all of a sudden appear less attractive to managers after further marketing research. The expected final payout from ongoing R&D projects may fluctuate wildly with each test and experiment. And generally any piece of private information held by the managers of a firm is likely to change significantly over time. Flipped on its head, the fact that this information changes so rapidly may be the best reason to suppose that the capital market cannot keep up with its development. In other words, the fact that a piece of information evolves rapidly makes it very costly for the capital market to keep itself informed of its evolution.

The nature of evolving private information, however, gives the investment decision an interesting twist, for no longer is the decision a static invest or not question, but rather a decision of when to invest in terms of the private information that's available to the firm. Suppose that the private information arrives in bundles of either good or bad news each period so that the accumulation of this information represents the total private information available to the firm. On the basis of the total information received, a firm will decide each period whether to invest today or wait for more information. In comparison, the information structure of standard adverse selection models can be thought of as follows. Initially, the capital market and firm are symmetrically informed, but the firm subsequently receives one bundle of information which is either good or bad. The firm then has to decide whether or not to invest. We will see that the evolving nature of the private information actually makes the problem one of moral hazard rather than adverse selection since the problem really lies in how much information a firm collects before investing. Depending on the type of securities that are issued, a firm will end up collecting too much or too little information. We will find that for most "standard" securities that are used, however, the firm will actually end up collecting too little information and end up investing too early.

Our basic prediction is that firms that require significant external financing will make hasty decisions and not take advantage of the information that will be revealed to them over time. For example, firms may not wait for a full analysis of market demand to be completed before investing. Even if this research is itself costless, as long as it takes time to complete, firms will be reluctant to wait. Or firms may go ahead with a project before they know that the technology needed is actually feasible. Absent the FDA, pharmaceutical firms may mass market a drug before they are sure of its side effects even knowing that subsequent legal suits may be costly. As long as acquiring information takes time, firms that need significant external financing will not have the right incentives to wait for it.

A social planner faced with the same informational constraint as the capital market can solve this problem by deliberately forcing the underpricing of the firm's securities. This shows, how from a social welfare perspective, a competitive capital market can be a mixed blessing. The interesting question raised by this, however, is not that government intervention is needed, but rather how we might already have "solutions" to take care of this problem in our economy. One example, is an oligopolistic credit market. In fact, firms may actually choose to enter into a relationship with a bank that underprices its securities if the bank in return offers services at a discount. An example that immediately comes to mind is that of the Japanese firm-banking relationships.

A third prediction is that if several firms have the same underlying private information with regard to their projects, a cash strapped firm may have an incentive to mimic the financing activities of a firm that is cash rich to signal to the market, that it too has collected its information. Cash strapped firms will thus issue their securities right after a cash rich firm.

The body of research most closely related to the current question at hand is that of the literature on investment under uncertainty that abstracts from corporate financing issues. Papers by Bernanke [2], McDonald and Siegel [9], Pindyck [11] and Dixit [5] have emphasized that when the per period profitability of an investment project evolves with ongoing uncertainty and the project requires an up front sunk cost to initiate, a firm will not invest as soon as the expected net present value of future profits exceeds the cost of the project.[1] The firm will instead want to wait for more information to see if the project is really a worthwhile investment. It will only invest when the foregone profits today are larger than the value of the information that will be revealed tomorrow. In this paper, we extend this basic intuition and

[1]See Dixit and Pindyck [6] for a comprehensive treatment of this topic. They have collected many of the techniques and results of the literature in this book.

apply it in analyzing how much information a firm will collect when its supplier of capital cannot observe the underlying source of uncertainty.

The analysis in this paper will mainly be focused on a small start-up firm managed by a risk-neutral owner. This will allow us to abstract from any principal agent problems and will also simplify interpreting what is meant by the objectives of the company. Namely, we will assume that the owner manager maximizes the net expected profits that accrue to him or her. If we wanted to extend our analysis to include diffusely held public corporations, we would have to assume that the current managers act selflessly on behalf of the current shareholders. This assumption has been questioned by Hart [8] among others.

The paper is organized as follows. In Section 3.2, I present a simplified approach to analyzing the standard investment decision which abstracts from financing issues. Section 3.3 develops a general model of investment with evolving private information. Section 3.4 provides a stylized example and shows how the presence of limited liability can cause the entrepreneur to invest too early. Section 3.5 discusses the solutions a social planner may implement to overcome the basic inefficiencies. This question is probably of most interest, not from a policy standpoint, but from how such "solutions" may already exist in our current economy. Section 3.6 asks what happens when there are several firms with the same underlying source of uncertainty. In this case, the firms with insufficient funds to finance their project may try to invest at the same time as a firm with sufficient funds and thus try to signal to the capital market that they are investing at the optimal time. Section 3.7 concludes.

3.2 The Discount Factor Approach

Starting with papers by Brennan and Schwartz [3], McDonald and Siegel [9], Pindyck [11] and Dixit [5], the literature on investment under uncertainty has relied on the use of stochastic dynamic programming. Although it is perhaps the most general framework in which to attack this set of problems, it invariably makes use of an optimality condition known as the "smooth pasting" or "high contact" condition which can be difficult to interpret economically. Consequently, departures from a first best world become difficult to analyze since they implicitly involve violations of some optimality condition. If the smooth pasting condition is difficult to interpret, any violation of it will surely be even more troublesome to deal with. To overcome this drawback of dynamic programming, I present an alternative approach to solve for the investment rule which can subsequently be analyzed graphically by use of marginal benefit and marginal cost curves. While the recursive nature of dynamic programming tends to obscure the relevant optimality conditions, our approach is "constructive" by design and will allow a simple economic interpretation of the optimality conditions at hand. In this section, I lay out the analytics of this approach and present a graphical interpretation.

We will initially consider the case where the entrepreneur has enough cash to finance the project without external financing. The general strategy will be to find the value of a future investment opportunity when we fix the investment rule, and then to maximize this value over the set of feasible rules. I consider a discrete project whose size is fixed so that the only decision ever made by the firm is when to invest. The firm must pay a sunk cost K to start the project. The value of the activated project, $V(X)$, will depend on a stochastically evolving state variable X which is always observable to the entrepreneur. One can for example take X to be an exogenous level of demand or the per period profit implied by that level of demand. We assume that $V(.)$ is monotonically increasing and differentiable in X. In addition, we assume for simplicity, a risk neutral economy. We take X to follow a continuous, time homogeneous Ito process of the form:

$$dX(t) = \mu(X)dt + \sigma(X)dB(t) \tag{3.1}$$

where $B(t)$ is a standard Brownian motion. Now, fix the level of X at which the firm will invest in the project at X_I. So at the point of investment, the firm will pay a cost K and receive a project worth $V(X_I)$ for a net gain of $V(X_I) - K$. The firm will not want to set X_I arbitrarily high since this would delay the time until it received the project. With a positive discount rate, the firm will want to balance the benefits of receiving a high value project against the time value of waiting for it. This is the fundamental trade-off of the modern theory of investments in real assets. Thus, all we need to do is find the relevant discount rate so that we can evaluate the present discounted value of $V(X_I) - K$. Clearly, this discount rate will be a function of X_I and the current level of X.

Put differently, we want to calculate the value of an asset, $q(X, X_I)$, which pays out one dollar when the process starting at X hits a level X_I. Thus this discount factor can also be considered as a contingent claim on the underlying factor X which can be priced in a risk neutral economy using the fact that the drift rate on the asset must equal $rq(X, X_I)$. By Ito's lemma, this can be written:

$$\frac{1}{2}\sigma^2(X)\frac{\partial^2 q}{\partial X^2} + \mu(X)\frac{\partial q}{\partial X} = rq \tag{3.2}$$

where $q(X, X_I)$ must also satisfy the boundary conditions:

$$\begin{aligned} q(X_I, X_I) &= 1 \\ \lim_{|X-X_I|\to\infty} &= 0 \end{aligned} \tag{3.3}$$

Also note that for $Y \in [X, Z]$, $q(X, Z) = q(X, Y)q(Y, Z)$ so that:

$$q_2(X, Z) = q(X, Y)q_2(Y, Z) \tag{3.4}$$

where the subscript 2 denotes the partial derivative with respect to the second argument. In the special case where $Y = Z$:

$$q_2(X, Z) = q(X, Z)q_2(Z, Z) \tag{3.5}$$

We can now value the investment option for a fixed investment rule X_I. It is simply:

$$q(X, X_I)\,(V(X_I) - K) \tag{3.6}$$

The objective is now to simply maximize the above with respect to X_I. This yields the first order condition:

$$-q_2(X, X_I)\,(V(X_I) - K) = q(X, X_I)V'(X_I) \tag{3.7}$$

Applying equation (3.5), we can rewrite this as:

$$-q(X, X_I)q_2(X_I, X_I)\,(V(X_I) - K) = q(X, X_I)V'(X_I) \tag{3.8}$$

Interpreting $q(X, X_I)$ as the discounting factor, the right hand side of equation (3.8) can be interpreted as the present discounted value of the marginal benefit of choosing a higher X_I, while the left hand side is the present discounted value of the marginal cost. This is seen more clearly by dividing equation (3.8) through by $q(X, X_I)$:

$$-q_2(X_I, X_I)\,(V(X_I) - K) = V'(X_I) \tag{3.9}$$

The right hand side is the marginal benefit of waiting for a higher X that results due to the project being more valuable at the point of exercise while the left hand side is the marginal cost the firm incurs due to the fact that it must wait longer to start the project and to receive $V(X_I) - K$.

Figure 3-1 on page 110 lays out the marginal cost and benefit of waiting as a function of X_I. The intersection of the two determines the optimal investment rule X_I. If the value of the firm doesn't depend on the underlying volatility of the process, increasing the volatility decreases the marginal cost of waiting while it does not affect the marginal benefit.[2] In other words, the option value of waiting is now greater due to the higher volatility of the underlying process so that investment is delayed. Figure 3-2 on page 111 shows how this increase in volatility will delay the investment. This comparative static result is arguably the most significant contribution of the recent literature on irreversible investment. The NPV rule simply states that the firm should invest when $V(X_I) - K = 0$ or equivalently when the marginal cost of

[2]This generally implies that the firm has no option values—say to exit for example.

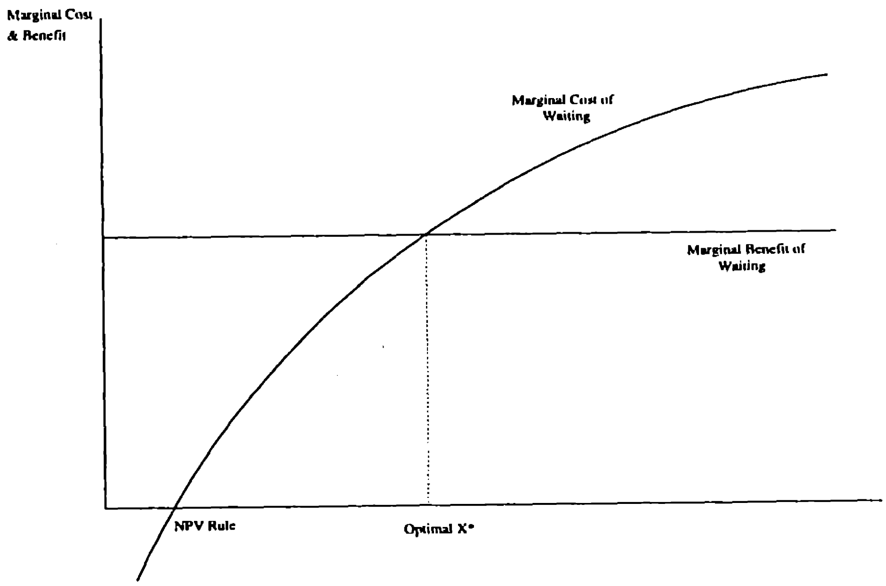

Figure 3-1: Optimal Investment Under Uncertainty

waiting, given by the left hand side of equation (3.9), is zero. Thus the NPV rule can be found in Figure 3-1 by locating the intercept of the marginal cost curve with the x-axis. Clearly, the NPV rule would dictate investing at a lower level of X than is truly optimal. One can also interpret the NPV rule as a rule that takes the marginal benefit of waiting to be zero. The NPV rule does not consider the value of the information that will be revealed in the future.

3.3 A Model of Evolving Private Information

3.3.1 Assumptions of the model

We examine the case where an entrepreneur does not have enough money to finance a project and must go to an uninformed financial market for funds. The key assumption we make is that:

ASSUMPTION 1: A state variable X which can be seen by the entrepreneur cannot be observed by the financial market and also cannot be inferred from observable variables when the entrepreneur issues securities.

It is best to think of X as a summary statistic of that portion of the firm's information that the capital market cannot observe or infer. We could also include

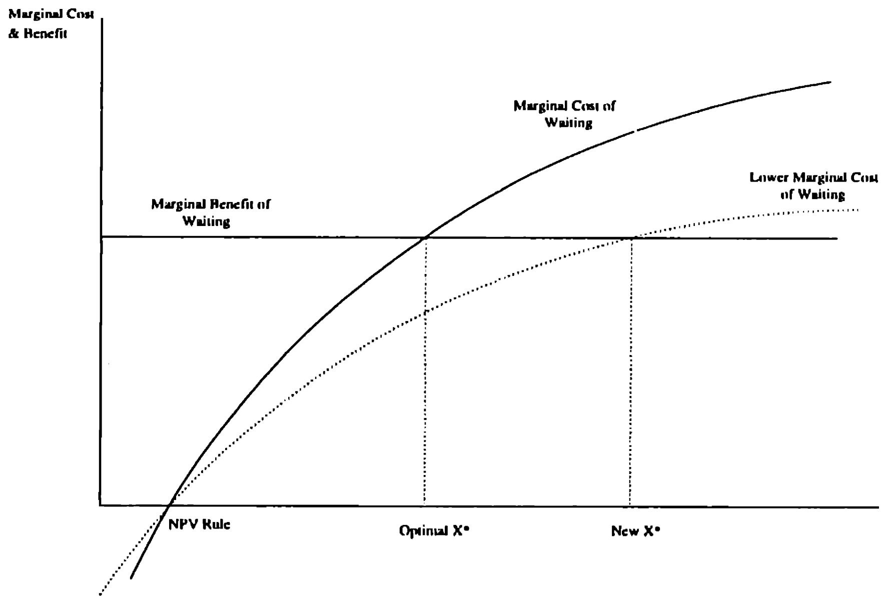

Figure 3-2: Increased Volatility Delays Investment

information that is observable to everybody, along with the evolving private information, but this would unnecessarily complicate the analysis. In papers that do not consider the possibility of private information, authors have assumed that the underlying source of uncertainty is the evolution of cash flow and that current cash flow is a sufficient statistic for deriving the conditional distribution of future cash flows. In our model, we can assume that cash flow is observable, but that there is some unobserved variable affecting its evolution. For example, X could include the possibility of future product development, the possibility of a law suit against a faulty product, or the likelihood of developing a technology that reduces costs. The main point is that X does not have to simply be current cash flow. So one might instead think of X as a variable that affects future cash flows rather than current. In this section, I lay out a general model of this form and give an example in the next section.

The entrepreneur has no cash on hand and must raise external funds of K to start the project. The entrepreneur can issue securities that are worth $O(X, A)$ where A represents the contractual agreements of the security. Although the market cannot observe X or its future values, the true value of the security will in most cases still depend on X. The amount the security market is willing to pay for the asset will be fixed, however, and so cannot depend directly on X. The amount the market will pay for the asset also cannot depend on time due to the stationarity of the problem. The only variable that the market's valuation will depend on is A, the contractual agreements of the asset. The value of the security given only the contractual agreements will be denoted by $S(A)$. There are two potential types of financial contracts that one can imagine. First, we can consider the case where the entrepreneur issues a security with the agreement that he/she will invest immediately. Second, one can also consider an entrepreneur who issues a security today with the agreement that he/she will invest at a future date with possible monetary transfers that will occur depending on when the entrepreneur finally invests relative to the date of the security issue. If the optimal configuration of the plant, such as input choices, depends on other observable variables that are changing over time (but which we will not model), the feasibility of this second contract depends critically on its ability to specify exactly what type of project will be undertaken at the future date as a function of the observable state variables. Most likely, writing and then enforcing a complete contingent contract will be very expensive. At the same time, if the entrepreneur can choose certain configurations of the business which are less profitable but which yield private benefits to the entrepreneur, this will result in a large deadweight loss in the potential value of the firm. This view of incomplete financial contracting is developed by Aghion and Bolton [1]. The main point is that if some decisions must be made at a future date and a contract cannot specify the action that should be taken for every realization of the world, the entrepreneur might not make that decision to maximize the value of the firm.

Although Aghion and Bolton's incomplete contracts approach utilizes the concept of moral hazard, a similar story can also be told from the perspective of adverse selection. One can imagine that there are other types of projects in the economy some of which are simply pet projects, which have negative NPV but yield some private benefits to the manager. It may be very difficult to write a contract to rule out such projects in advance if they are similar enough to the specified profitable projects. Then many "dishonest" entrepreneurs will approach the market pretending to have a project as characterized in equation (3.6). If instead the entrepreneur issues the security and simultaneously invests as the first type of contract would require, the capital market can observe the details of the plant and equipment that are being constructed and can determine whether the project is legitimate or simply a pet project. With enough "dishonest" entrepreneurs in the economy and limited liability, the second type of contract becomes impossible to implement. So the question then is whether it is so difficult to write a contract detailing what kind of project will be undertaken in the future so that a dishonest entrepreneur would refuse to sign on. Since the capital market must mostly rely on the entrepreneur for the details of the project to begin with, without actually observing the setup of the plant, it will be difficult indeed for the capital market to specify in writing (and enforce) what constitutes an undesirable project. This type of contract realistically does not appear to be feasible for the kind of small start-up companies we are considering. For these reasons we will focus on the contract that requires financing and investing to occur simultaneously:

ASSUMPTION 2: The entrepreneur issues securities and invests simultaneously.

Whether it be for the theoretical reasons cited above, this assumption at least seems to hold true empirically for small start-up firms who are trying to find their initial financing. It should also be noted that relaxing this assumption will in general not bring us back to the first best world. The problem will then lie in constructing an incentive compatible contract where the entrepreneur issues the securities at a given level of X. This will generally not be possible without imposing some ex-post inefficiencies. No matter what type of contract we allow, the problem will always lie in generating incentive compatibility.

On the technical side, to make the problem tractable, I assume that the private information X follows a stationary process (such as an Ornstein-Uhlenbeck process) with no inaccessible region on the real line[3] and that the market has steady state beliefs about the value of the process. This makes the beliefs of the market constant through time and consequently makes the entrepreneur's problem stationary. Although, I have assumed that X follows an Ito process, the subsequent calculations

[3]An arithmetic Ornstein-Uhlenbeck process for example .

will mainly rely on the continuity of the sample paths.

3.3.2 Preliminary Intuition

When the entrepreneur has enough cash to finance a project, he/she balances the tradeoff between forgoing profits that the entrepreneur could earn by entering immediately against the benefits of waiting for more information. However, in the presence of asymmetric information where the entrepreneur must raise external financing, the costs and benefits of waiting are not distributed proportionately between the entrepreneur and the financiers. The burden of lost profits is completely borne by the entrepreneur while the financiers reap part of benefit of the entrepreneur waiting for more information. Thus while external financing is costly from the entrepreneur's perspective, it causes over-investment that results because the entrepreneur finds it too costly to keep waiting up to the socially optimal level for investment. The entrepreneur will ex-ante choose the financial instrument that minimizes the transfer of the benefits of waiting to the financiers. In this section, I use the tools developed in Section 3.2 to present a simple analysis of the entrepreneur's investment decision under asymmetric information.

3.3.3 Equilibrium–its definition and calculation

We will define a pure strategy equilibrium for the contractual agreements of the financial security A. To allow comparison to a world with complete information, we take $S(\mathcal{I}, A)$ to be the value of a security with contractual agreements A and where the capital market has the information set $\mathcal{I}$. With complete information X_A, the level of X at which the entrepreneur invests, is included in the information set $\mathcal{I}$. We define X_A^* to be the level of X at which the entrepreneur invests. We define N to be the number of securities issued by the firm. The equilibrium is defined as follows:

Equilibrium

[E1] *The entrepreneur maximizes the following with respect to X_A:*

$$q(X, X_A)\,(V(X_A) - K + N\,[S(\mathcal{I}, A) - O(X_A, A)]) \tag{3.10}$$

[E2] *Total funds must equal K:*

$$N \cdot S(\mathcal{I}, A) = K \tag{3.11}$$

[E3] *Zero profit condition in the financial market:*

$$S(I, A) = \begin{cases} O(X_A, A) & \textit{if } X \in \mathcal{I} \\ O(X_A^*, A) & \textit{if } X \notin \mathcal{I} \end{cases} \tag{3.12}$$

Condition [E3] distinguishes between the case where the capital market can and cannot observe X. If the capital market can observe X, then the security will be priced correctly no matter what level of X_A is chosen. If the capital market cannot observe X, then the security will only be correctly priced in equilibrium, but not off of the equilibrium path.

Solving for N, the above conditions for the case of symmetric information is simply for the entrepreneur to maximize:

$$q(X, X_A)\,(V(X_A) - K) \tag{3.13}$$

with respect to X_A. This is the same maximization problem as faced by an entrepreneur with sufficient funds to finance the project without external financing. Thus with symmetric information, the entrepreneur invests at the optimal time regardless of the security issued. For the purpose of comparison to the case of asymmetric information, we note that the first order condition to the above problem can be written:

$$-\,q_2(X_A^*, X_A^*)\,(V(X_A^*) - K) = V'(X_A^*) \tag{3.14}$$

For the case of asymmetric information, the equilibrium conditions can be combined to yield the following maximization problem for the entrepreneur:

$$q(X, X_A)\left[V(X_A) - K\left(\frac{O(X_A, A)}{O(X_A^*, A)}\right)\right] \tag{3.15}$$

(It is important to note that the denominator of the last piece has as its argument X_A^* rather than X_A; if it indeed were simply X_A instead, we would end up with the equilibrium for the case of symmetric information.) The first order condition to this maximization problem is thus:

$$-q_2(X_A^*, X_A^*)\,(V(X_A^*) - K) = V'(X_A^*) - K\left(\frac{O'(X_A^*, A)}{O(X_A^*, A)}\right) \tag{3.16}$$

Thus we see that with external financing and incomplete information, an extra term on the right hand side of equation (3.16) causes the equilibrium to diverge from the case of complete information. This term represents the marginal increase in the value of the security resulting from an increase in X–it is the marginal benefit of waiting that is transferred from the entrepreneur to the security holders. This problem is the most severe when the security is very sensitive to the level of X or equivalently to the value of the firm. Because the entrepreneur is not compensated by the security market for this transferred benefit, this drives a wedge between the total marginal benefits of waiting and the marginal benefit faced by the entrepreneur. The marginal cost of waiting, or simply the foregone profits from delayed entry, are completely absorbed by the entrepreneur. Figure 3-3 on page 116 presents a diagrammatic summary of how

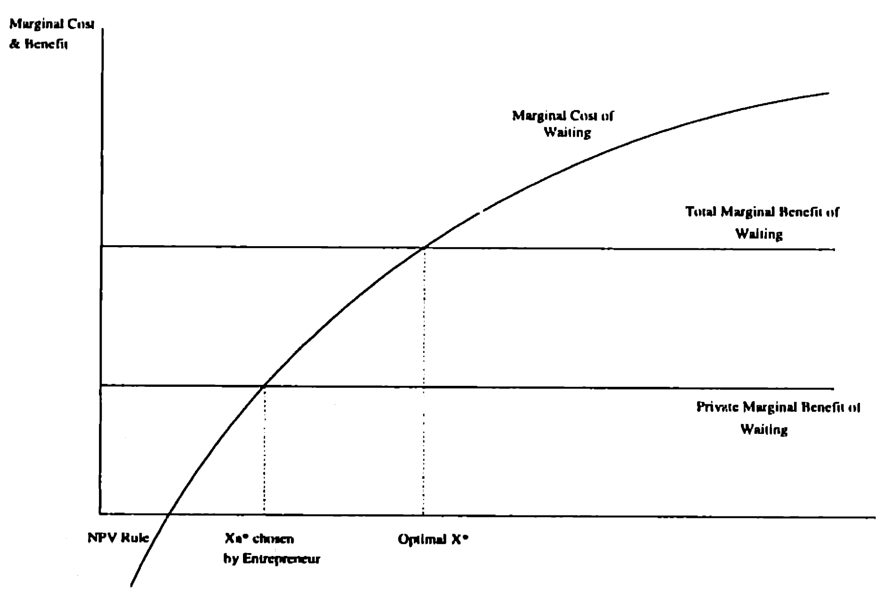

Figure 3-3: Investment with Asymmetric Information

the investment trigger is determined. If the value of the security depends positively on the value of the firm (call such securities standard securities) and consequently on X, waiting is a positive externality and will cause the entrepreneur to invest too early. There is a wedge between the total marginal benefit of waiting and the private marginal benefit of waiting. For standard securities, the total marginal benefit is higher than the private marginal benefit of waiting. The total marginal cost of waiting, however, is completely borne by the entrepreneur. This problem is the most severe for securities such as equity, but is still a problem for risky debt as well since the default risk decreases when the firm enters at a higher level of X.

Ultimately, one would like to issue a security that is unaffected by the value of the firm (and consequently by X), but this will not be possible in many cases. The general conclusion is that as long as the set of securities that can be issued depend positively on the value of the firm, the entrepreneur will invest too early. With evolving private information the equilibrium payoff to the entrepreneur is:

$$q(X, X_A^*)\,(V(X_A^*) - K) \tag{3.17}$$

where X_A^* is given implicitly in (3.16). The entrepreneur will choose A to maximize the above quantity subject to (3.16). This payoff will always be less than the payoff to an entrepreneur with enough cash to finance the project as long as $O'(X^*, A) \neq 0$ for all feasible A, where X^* is the first best level of X at which to invest and is equivalently the maximizer of (3.17). We will also see later that the zero profit condition in the financial market plays a crucial role in the inability of society to reach an optimal investment policy. This leaves open a role for the social planner to induce optimal investment by buying the entrepreneurs' securities at an "unfair" price or equivalently by imposing a tax on the issuing of securities.

3.4 An Example

In this section, we present a simple stylized example of the model developed above for the case where although the realized profits of the firm are observable to everybody, the entrepreneur observes the future expected growth rate of profits while the capital market does not. The objectives of this section are two fold. The first is to show an example of how limited liability causes all feasible securities to be "standard seccurities"–securities whose value depend positively on the value of the firm. By the preceeding section, this implies that the cash constrained entrepreneur invests too early relative to an entrepreneur with sufficient funds to pay for the project. The second objective of this section is to show an example of how it can be the case that the underlying variable X is not directly discernible to outsiders at the time of investment. As pointed out before, if X is current cash flow it will clearly be directly observable by outsiders. If however, the level of X affects future cash flow, but not current, it will clearly not be observable. The example we develop builds upon the latter structure.

The project we will consider earns a per period profit of R_0 when it is started. This will continue for a time, say T periods, at which point profits will either rise to $R_1^H > 0$ or fall to $R_1^L > 0$, where $R_1^L < R_1^H$, and subsequently remain there forever. At the time of investment, we assume that only the entrepreneur can observe the probability of either state occurring which is changing continuously over time. To put this into the context of the previously developed model, assume that the high state will only occur if $X_T > X^-$, where 0 is taken to be the time of investment, X_T is the value of a stationary process X at the time T and X^- is a constant. Now, define:

$$\pi_0 = \int_0^T e^{-rt} R_0 dt \qquad \begin{aligned} \pi_1^H &= \frac{R_1^H}{r} \\ \pi_1^L &= \frac{R_1^L}{r} \end{aligned} \tag{3.18}$$

Then, the value of the project as a function of X at the time of investment is:

$$V(X_0) \;=\; \pi_0 + \delta\left[\mathrm{Prob}(X_T > X^-|X_0)\pi_1^H + \left[1 - \mathrm{Prob}(X_T > X^-|X_0)\right]\pi_1^L\right]$$

$$= \pi_0 + \delta \left[\text{Prob}(X_T > X^- | X_0)(\pi_1^H - \pi_1^L) + \pi_1^L\right] \tag{3.19}$$

where $\delta = e^{-rT}$. Note that $V'(X) > 0$ since $\pi_1^H > \pi_1^L$ and:

$$\frac{\partial}{\partial X_0}\text{Prob}(X_T > X^- | X_0) > 0 \tag{3.20}$$

Thus this project demonstrates a case where the variable X cannot be observed directly by outsiders at the time of investment. To show how it can be the case that all securities issued by the entrepreneur end up depending positively on the value of the firm, we make the following two key assumptions.

Assumption We assume that $\pi_0 + \delta\pi_1^H > K$ while $\pi_0 + \delta\pi_1^L < K$ so that the project is a risky one, but potentially profitable in expectation.

Assumption There is limited liability. In other words, the entrepreneur's utility cannot be decreased beyond the point he/she has no money so that there is no imprisonment or other non–pecuniary punishment.

The first assumption rules out the possibility that the entrepreneur can simply issue riskless debt. If the project always at least paid for itself, there would be no problem in simply borrowing money from a bank at the risk-free rate. The second assumption, which is quite natural, is stated more for the purpose of emphasizing the difficulty of aligning the interests of the security holders with the entrepreneurs. We will show in what follows that the security issued by the entrepreneur will always depend postively on the value of the firm.

A financial contract can be summarized by what it agrees to pay out from the time of investment $t = 0$ to $t = T$ and what it pays out in the low and high states from time $t = T$ to $t = \infty$. Take a contract A. Denote by $d_0(t, A)$ the flow payment at time t for $t \in [0, T]$ and by $d_{1H}(t, A)$ the flow payment at time t for $t \in [T, \infty]$ when the high state is attained and $d_{1L}(t, A)$ the flow payment at time t for $t \in [T, \infty]$ when the low state is attained. Define:

$$D_0 = \int_0^T e^{-rt} d_0(t, A)dt \qquad \begin{aligned} D_1^H &= \textstyle\int_0^T e^{-rt} d_{1H}(t, A)dt \\ D_1^L &= \textstyle\int_0^T e^{-rt} d_{1L}(t, A)dt \end{aligned} \tag{3.21}$$

Then by simplifying as in (3.19), the value of the contract A as a function of X is:

$$O(X_0, A) = D_0 + \delta \left[\text{Prob}(X_T > X^- | X_0)(D_{1H} - D_{1L}) + D_{1L}\right] \tag{3.22}$$

The financial contract also must be feasible, in that it cannot promise more than the firm can pay out. Defining N to be the number of shares that are issued, it must be that: $N\ D_0 \leq \pi_0$, $N\ D_{1H} \leq \pi_{1H}$ and $N\ D_{1L} \leq \pi_{1L}$

For total funds to equal K, it must be that $N\ O(X_A^*, A) = K$. We can now show the following:

Proposition: $O'(X, A) > 0$ for all A where $N\ O(x, A) = K$ for some x.

Proof: First note that:

$$\begin{aligned} N\ O(X_0, A) &= N\ D_0 + \delta N\ D_{1L} + \delta \text{Prob}(X_T > X^- | X_0)\ N\ (D_{1H} - D_{1L}) \\ &\leq \pi_0 + \delta \pi_{1L} + \delta \text{Prob}(X_T > X^- | X_0)\ N\ (D_{1H} - D_{1L}) \\ &< K + \delta \text{Prob}(X_T > X^- | X_0)\ N\ (D_{1H} - D_{1L}) \end{aligned} \tag{3.23}$$

Since probabilities are always positive, $N\ O(X, A)$ will never equal K unless $D_1^H - D_1^L > 0$. But we also know that:

$$O'(X_0, A) = \delta \left(\frac{\partial}{\partial X_0} \text{Prob}(X_T > X^- | X_0) \right) (D_{1H} - D_{1L}) \tag{3.24}$$

and we know that the first 2 terms are positive as well as the third, which we showed by virtue of equation (3.23). Thus $O'(X, A) > 0$ if $N\ O(x, A) = K$ for some x. □

This proposition thus implies that the entrepreneur will always overinvest since we have seen in the previous section that $O'(X, A) > 0$ implies that the entrepreneur will invest at a level of X that is too low–i.e. we will end up with early investment. Essentially, the point is that the entrepreneur must pay more in the high statte than in the low state if te security holders are to break even. But if the security holders receive more in the high state than in the low state, part of the benefit of waiting transfered from the entrepreneur to the security holders. Thus waiting becomes a positive externality, and consequently there will not be enough of it.

3.5 The Social Planner's Problem and the Mispricing of Securities

As long as the securities that can be issued by the entrepreneur are sensitive to the value of the firm, or more specifically $O'(X^*, A) \neq 0$ for all feasible A, the entrepreneur will not invest at the optimal level of X. An interesting question to ask at this juncture is how a social planner, faced with the same informational constraints as the capital market, could induce the entrepreneur to invest at the optimal level of X? If we ignore distributional issues, there are two straightforward solutions. First, the social planner could simply give the entrepreneur K so that the entrepreneur now has enough cash to finance the project without raising external financing. Alternatively, the social planner can act as the sole provider of external finance and deliberately under-price all securities that the entrepreneur issues. To see this consider again equation (3.10), except replace $S(\mathcal{I}, A)$ with $S_P(A)$, the price that the social planner will pay for the securities A issued by the entrepreneur. Also simplify the equation using (3.5) and consider the case where $S_P(A)$ must be determined without observing X so that $\partial S/\partial X_A = 0$:

$$\begin{aligned} & -q_2(X_A^*, X_A^*)\left(V(X_A^*) - K + \left(\frac{K}{S_P(A)}\right)[S_P(A) - O(X_A^*, A)]\right) \\ = \;& V'(X_A^*) - \left(\frac{K}{S_P(A)}\right) O'(X_A^*, A) \end{aligned} \tag{3.25}$$

This can be rewritten:

$$\begin{aligned} & -q_2(X_A^*, X_A^*)\,(V(X_A^*) - K) \\ & \quad = V'(X_A^*) + \left(\frac{K}{S_P(A)}\right)(q_2(X_A^*, X_A^*)[S_P(A) - O(X_A^*, A)] - O'(X_A^*, A)) \end{aligned} \tag{3.26}$$

So the social planner can induce the entrepreneur to invest at the optimal X as long as the second piece on the right hand side is equal to zero. This can be achieved by setting:

$$S_P(A) = O(X^*, A) - \left(\frac{O'(X^*, A)}{-q_2(X^*, X^*)}\right) \tag{3.27}$$

subject to the constraint that $S_P(A) > 0$, or else the entrepreneur will not be able to raise any money. The denominator on the right hand side is positive since $q_2(X^*, X^*) < 0$. Consequently, we see that for standard securities where $O'(X, A) > 0$, the social planner will want to underprice the security to induce optimal investment. Under this regime, the entrepreneur will invest at X^*, the socially optimal level at which to invest. By under-valuing the entrepreneur's securities, the social planner reduces the marginal cost of waiting and thus induces optimal waiting. Forgoing an investment today is not as costly anymore simply because the investment is less attractive. By combining (3.27) and (3.10), we find that the expected payoff to the entrepreneur will be:

$$\begin{aligned} & q(X, X^*)\left[V(X^*) - K + \left(\frac{K}{S_P(A)}\right)\left(\frac{O'(X^*, A}{q_2(X^*, X^*}\right)\right] \\ = \; & q(X, X^*)\left[V(X^*) - K + K\left(q_2(X^*, X^*)\left[\frac{O(X^*, A)}{O'(X^*, A)}\right] + 1\right)^{-1}\right] \qquad (3.28) \end{aligned}$$

So that if the social planner allows the entrepreneur to choose which security he/she will issue, since X^* is fixed, the entrepreneur will simply choose A to minimize $\frac{O'(X^*,A)}{O(X^*,A)}$

Of course another way to implement this regime is to impose a tax on issuing securities creating a gap between the true value of securities and what the entrepreneur receives in return. If we re-define $S(A)$ as the amount that the entrepreneur receives for the securities net of taxes, then $S(A) = O(X^*_A, A) - \tau(A)$, where $\tau(A)$ is the tax imposed on issuing securities of type A. By setting $\tau(A) = -O'(X^*, A)/q_2(X^*, X^*)$, the social planner can achieve the same social optimum.

The solution to the social planner's problem is probably more interesting in the context of how such "solutions" may already exist in the economy without government intervention. For example, a monopolist in a local capital market, such as a rural community bank, may underprice it's clients securities. Other ways in which such monopoly power can arise is discussed in Rajan [12]. This monopolistic provider of capital may actually improve social welfare by discouraging hasty investments by the entrepreneurs. Alternatively a firm may voluntarily choose to enter into a exclusive relationship with a bank that underprices it's debt, but in return deliver a flow of services at a discount. This relationship forces the entrepreneur to wait for more information before investing and increases the total pie to be split among the bank and the firm. Of course, the firm will have to be compensated for allowing the bank to underprice its securities.

3.6 Multiple Firms and Piggyback Financing

Perhaps the most interesting prediction of the theory presented so far is that firms that need to raise the most external financing will be the ones that will jump in early into a market without waiting for much information, while firms that have sufficient cash on hand find it worthwhile to wait and see how their project fairs over time before committing to it. In empirically examining this idea, however, one may want to look at the order of entrance into an industry based on how much cash a firm has on hand (normalized in some manner). This implicitly assumes that each firm has a similar underlying process X which represents its evolving private information. So a firm with more cash on hand will invest after others have entered.

The above analysis, however, ignores the signaling role of the firm with sufficient cash to invest without outside financing. If other entrepreneurs in the industry know that a cash rich firm will enter later, they have an incentive to piggyback on this firm and wait until this firm has issued securities before issuing their own. They will wait for market sentiment on their industry to pick up. This sentiment will pick up significantly when a cash rich firm enters the industry and in effect signals that X is high. The other entrepreneurs will wait for this credible signal to be sent before entering themselves. Our prediction in this case is that a cash rich firm will be the first to issue and will be followed soon after by firms that must raise more external financing. This section thus should be considered as a caveat to the main thesis of the paper.

If some firms are not sure whether a cash rich firm will ever enter the industry, however, the piggyback equilibrium may break down. The cash strapped firms are not willing to wait for a cash rich firm if it may be in vain. The market sentiment on this industry's securities will never pick up if this cash rich firm doesn't arrive. In this case, the cash strapped firms may give up waiting and invest at the level of X otherwise optimal for them in the absence of a cash rich firm. It is in this equilibrium that we expect the predictions of our paper to be carried out. The big firms, that tend to have enough cash to invest, will wait longer before starting a project than small start-up companies that are financed almost exclusively with new outside capital.

We will show an example of how this could occur. First we need a lemma:

Lemma: If at time 0 it is common knowledge that $X = X^*$, where X^* is the first best level of X at which to invest, but that after time 0 the capital market cannot observe the further evolution of X, then the entrepreneur will still invest at time 0 with $X = X^*$.

Proof: Defining $\tau^* = 0 = \{\inf\ t | x(t) > X^*\}$, we know that:

$$\tau^* = 0 = \text{argmax}_\tau\ \mathrm{E}\left[e^{-r\tau}[V(X(\tau)) - K | X(0) = X^*\right] \quad (3.29)$$

The cash strapped entrepreneur, however, wants to maximize the following with respect to an optional time τ (the key to using optional times is that it allows strategies that are time dependent):

$$\mathrm{E}\left[e^{-r\tau}\left[V(X(\tau)) - K + \left(\frac{K}{S(\tau, A)}\right)(S(\tau, A) - O(X(\tau), A))\right] | X(0) = X^*\right] \quad (3.30)$$

This can be broken into two pieces:

$$\begin{aligned} &\mathrm{E}\left[e^{-r\tau}[V(X(\tau)) - K] | X(0) = X^*\right] \\ &\quad + \mathrm{E}\left[e^{-r\tau}\frac{K}{S(\tau, A)}(S(\tau, A) - O(X(\tau), A)) | X(0) = X^*\right] \end{aligned} \quad (3.31)$$

Because the capital market and the entrepreneur have symmetric information at time 0, the second piece, which is just the expected gain from financing activities, must be equal to zero. No matter what optional time strategy is chosen by the entrepreneur, in equilibrium, the net expected gain from financing activities will be zero as of time zero when both parties have symmetric information. However, choosing a τ such that $\text{Prob}(\tau = 0) < 1$ yields a strictly lower expected payoff via the first piece than choosing $\tau = 0$. □

Now, if an entrepreneur knows that another entrepreneur with the same underlying X, but who has enough cash to finance the project exists with probability 1, then the cash strapped entrepreneur will wait for the cash rich entrepreneur to invest and then follow suit immediately. When the cash rich entrepreneur enters, the capital market knows that $X = X^*$. The cash strapped firm then can issue his/her securities immediately for their fair value and invest at the optimal level of X—this assured by the lemma above.

If the existence of a cash rich firm is uncertain, then the cash strapped firm may invest as though the cash rich firm didn't exist. We assume that as in the example of Section 3.4, that all contracts are such that $O'(X, A) > 0$. Define the probability that there exists an entrepreneur with cash on hand greater than K to be P:

Proposition: There exists a $P > 0$ such that a firm with no cash will invest as though the cash rich firm did not exist. This is supported by the out of equilibrium belief that if a cash rich firm invests right before a cash strapped firm, that at that point in time $X = X^*$.

Proof: By not waiting for the cash rich entrepreneur to appear, the payoff to the entrepreneur is:

$$q(X, X_A^*)\,[V(X_A^*) - K] \tag{3.32}$$

where $X_A^* < X^*$ since we assumed that $O'(X, A) > 0$ for all A and X. The entrepreneur can deviate from this strategy by waiting until X^* to see if a cash rich firm arrives. If it does not, the market believes that the entrepreneur did not deviate from the no waiting strategy. If the cash rich firm does invest, however, the market will believe that $X = X^*$ at that time. So the expected payoff from deviating is:

$$\begin{aligned} &Pq(X, X^*)[V(X^*) - K] \\ &\quad +(1-P)q(X, X^*)\left[V(X^*) - K + K\left[\frac{S(A) - O(X^*, A)}{S(A)}\right]\right] \end{aligned} \tag{3.33}$$

where $S(A) < O(X^*, A)$ since the capital market thinks that $X = X_A^* < X^*$. We can simplify the above to obtain:

$$q(X, X^*) \left[V(X^*) - K + \left[\frac{K}{S(A)}\right] (S(A) - O(X^*, A))\right] + \Delta \tag{3.34}$$

where:

$$\Delta = P\left(\frac{K}{S(A)}\right)\left[\frac{O(X^*, A) - S(A)}{S(A)}\right] \tag{3.35}$$

where the second and third pieces of Δ are strictly positive. Since X_A^* is the maximizer of:

$$f(z) = q(X, z)\left[V(z) - K + \left[\frac{K}{S(A)}[S(A) - O(z, A)]\right]\right] \tag{3.36}$$

it must be that:

$$f(X_A^*) > f(X^*) \tag{3.37}$$

so that there exists an $\epsilon > 0$ such that $f(X_A^*) > f(X^*) + \epsilon$. The expression in (3.34) is simply $f(X^*) + \epsilon$, so that to show there exists a P such that deviating is not optimal, all we need to do is set P such that $\Delta < \epsilon$ so that $f(X^*) + \Delta < f(X_A^*)$. Because the second and third terms of (3.35) are positive, we can thus choose P to be positive as well. Since we need only choose $\Delta < \epsilon$, we can also choose $P < 1$. □

So if the probability of a cash rich firm arriving is 1, then the equilibrium will be for cash strapped firms to piggyback, while if the probability of a cash rich firm arriving is small, then cash strapped firms will invest as though the cash rich firm did not exist.

3.7 Conclusion

We have provided a first step in examining how the evolving nature of private information can affect investment and corporate financing decisions. With most standard financial instruments, a firm with no cash on hand will tend to make a hasty decision and invest too early. This results from the fact that while the entrepreneur bears all the cost of foregone profits today in waiting for more information, part of the benefit of waiting accrues to the new securities holders. A social planner may try to solve this problem by either underpricing the securities issued by a firm or imposing a tax on security issues. When there are multiple firms with the same underlying source of uncertainty, we can end up in a situation where firms with less cash on hand will invest at the same time as a cash rich firm. If on the other hand, the cash strapped firm does not think that a cash rich firm is likely to enter this market, we may see instances ex-post where the cash strapped firm enters a market earlier than a cash rich one.

Bibliography

[1] Aghion, P. and Bolton, P., "An Incomplete Contracts Approach to Financial Contracting," 1992, *Review of Economic Studies*, **59**, 473–494.

[2] Bernanke, B., "Irreversibility, Uncertainty, and Cyclical Investment," *Quarterly Journal of Economics*, 1983, **98** ,85–106.

[3] Brennan, M. and Schwartz, E., "Evaluating Natural Resource Investments," *Journal of Business*, 1985, **58**, 135–157.

[4] Choe, R. M. and Nanda, V., "Common Stock Offerings Across the Business Cycle: Theory and Evidence," Working Paper, Vanderbilt, 1989.

[5] Dixit, A., "Entry and Exit Decisions under Uncertainty," *Journal of Political Economy*, 1989, **97**, 620–638.

[6] Dixit, A. and Pindyck, R., *Investment Under Uncertainty*, Princeton University Press, 1994.

[7] Fleming, W. and Raymond R., *Deterministic and Stochastic Optimal Control.* New York: Springer Verlag, 1975.

[8] Hart, O., "Theories of Optimal Capital Structure: A Principal Agent Perspective," Working Paper, MIT, 1991.

[9] McDonald, R. and Siegel, D., "The Value of Waiting to Invest," *Quarterly Journal of Economics*, **101**, 1985, 707–727.

[10] Myers, S. and Majluf, N., "Corporate Financing and Investment Decisions When Firms Have Information That Investors Do Not Have," *Journal of Financial Economics*, 1984, **13**, 187–222.

[11] Pindyck, R., "Irreversible Investment, Capacity Choice, and the Value of the Firm," *American Economic Review*, 1988, **78**, 969–985.

[12] Rajan, R., "Insiders and Outsiders: The Choice Between Relationship and Arm's Length Debt," *Journal of Finance*, 1992, **47**, 1367–1400.