Calibration and Data Analysis

In this lecture. . .

- the theoretical yield curve and the market yield curve, should they be the same?
- how calibration works, the pros and cons
- how to analyze short-term interest rates to determine the best model
- how to analyze the slope of the yield curve to get information about the market price of interest rate risk

By the end of this lecture you will

- know the meaning of 'calibration'

- appreciate the pros and cons of calibration

- be able to analyze data to find a good model

Introduction

We have seen the theory behind one-factor (and multi-factor) interest rate models.

These models will have as an output a 'theoretical' yield curve.

This theoretical yield curve will not be the same as the yield curve seen in the market.

Is this good or bad?!

It is bad if. . . your job is to price exotic, structured products which must be hedged with simple instruments. (How can you be expected to price exotics 'correctly' if you can't even match simple instruments?!)

It is good if. . . you believe your theoretical model and you are looking for arbitrage opportunities among simple instruments. (If your model output matched market prices then you'll never find any arbitrage!)

So in using any model we have to decide how to choose the parameters.

Should the parameters be chosen to match

- the market yield curve? Or

- historical interest rate data?

The former is **calibration** to a snapshot of the market at one instant in time.

The latter is fitting to time series data.

Let's start with calibration.

Calibration

Because of this need to correctly price liquid instruments, the idea of **yield curve fitting** or **calibration** has become popular.

When stochastic models are used in practice they are almost always fitted.

To match a theoretical yield curve to a market yield curve requires a model with enough degrees of freedom. (You are matching a curve, i.e. an 'infinite' number of points, so you need infinite degrees of freedom!)

This is done by making one or more 'parameters' time dependent.

- This functional dependence on time is then carefully chosen to make an output of the model, the price of zero-coupon bonds, exactly match the market prices for these instruments.

Ho & Lee

The Ho & Lee spot interest rate model is the *simplest that can be used to fit the yield curve*.

Now we don't necessarily say this model or idea is great, but we will go through the mathematics to see how it can be done.

Recap: In the Ho & Lee model the process for the *risk-neutral* spot rate is

$$dr = \eta(t)dt + c\,dX.$$

The standard deviation of the spot rate process, c, is constant, the drift rate η is time dependent.

As we've seen in an earlier CQF lecture, for this model the solution of the bond pricing partial differential equation for a zero-coupon bond is simply

$$Z(r, t; T) = e^{A(t;T) - r(T-t)},$$

where

$$A(t; T) = -\int_t^T \eta(s)(T-s)ds + \tfrac{1}{6}c^2(T-t)^3.$$

(Note that the variables are r and t, but we are also explicitly referring to the parameter T, the bond maturity.)

Working forwards: If we know $\eta(t)$ then the above gives us the theoretical value of zero-coupon bonds of all maturities. I.e. start with model ($\eta(t)$) and find answer (Z).

An inverse problem: But what if we know Z from the market, but don't know the unobservable η? Turn this relationship around and ask the question

- 'What functional form must we choose for $\eta(t)$ to make the theoretical value of the discount rates for all maturities equal to the market values?'

That is calibration!

(What about the parameter c?)

Suppose we want to calibrate our model today, time t^*. Today's spot interest rate is r^* and the discount factors *in the market* are $Z_M(t^*; T)$.

Call the special, calibrated, choice for η, $\eta^*(t)$.

To match the market and theoretical bond prices, we must solve

$$Z_M(t^*;T) = e^{A(t^*;T) - r^*(T-t^*)}.$$

Taking logarithms of this and rearranging slightly we get

$$\int_{t^*}^{T} \eta^*(s)(T-s)ds = -\log(Z_M(t^*;T)) - r^*(T-t^*) + \frac{1}{6}c^2(T-t^*)^3. \tag{1}$$

We know everything on the right-hand side. So this is an **integral equation** for $\eta^*(t)$.

(Luckily for us, it is quite easy to solve!)

Observe what happens if we differentiate the integral term with respect to T.

First differentiate once with respect to T

$$\frac{d}{dT}\int_{t^*}^{T}\eta^*(s)(T-s)ds = \int_{t^*}^{T}\eta^*(s)ds.$$

Differentiate again

$$\frac{d^2}{dT^2}\int_{t^*}^{T}\eta^*(s)(T-s)ds = \eta^*(T).$$

So, differentiating (1) twice with respect to T we get

$$\eta^*(t) = c^2(t - t^*) - \frac{\partial^2}{\partial t^2} \log(Z_M(t^*; t)).$$

The solution!

With this choice for the time-dependent parameter $\eta(t)$ the theoretical and actual market prices of zero-coupon bonds are the same.

Notes:

- Now that we know $\eta(t)$ we can price other fixed income instruments.

- We say that our prices are **consistent with the yield curve**.

- The same idea can be applied to other spot interest rate models.

- This is an inverse problem, and will typically be sensitive to input date (the Z).

Another calibrated model:
The extended Vasicek model of Hull & White

Most one-factor models have the potential for fitting, the more tractable the model the easier the fitting. If the model is not at all tractable then we can always resort to numerical methods.

The next easiest model to fit is the Vasicek model. The Vasicek model has the following stochastic differential equation for the risk-neutral spot rate

$$dr = (\eta - \gamma r)dt + c\,dX.$$

Hull & White extend this to include a time-dependent parameter

$$dr = (\eta(t) - \gamma r)dt + c\,dX.$$

Assuming that γ and c have been estimated statistically, say, we choose $\eta = \eta^*(t)$ at time t^* so that our theoretical and the market prices of bonds coincide.

Again, as covered in an earlier CQF lecture, under this risk-neutral process the value of a zero-coupon bond is

$$Z(r,t;T) = e^{A(t;T) - rB(t;T)},$$

where

$$A(t;T) = -\int_t^T \eta^*(s)B(s;T)ds$$

$$+\frac{c^2}{2\gamma^2}\left(T - t + \frac{2}{\gamma}e^{-\gamma(T-t)} - \frac{1}{2\gamma}e^{-2\gamma(T-t)} - \frac{3}{2\gamma}\right).$$

and

$$B(t;T) = \frac{1}{\gamma}\left(1 - e^{-\gamma(T-t)}\right).$$

As before, to fit the yield curve at time t^* we must make $\eta^*(t)$ satisfy

$$
\begin{aligned}
& -\int_{t^*}^{T} \eta^*(s) B(s;T) ds \\
& + \frac{c^2}{2\gamma^2}\left(T - t^* + \frac{2}{\gamma} e^{-\gamma(T-t^*)} - \frac{1}{2\gamma} e^{-2\gamma(T-t^*)} - \frac{3}{2\gamma}\right) \\
& = \log(Z_M(t^*;T)) + r^* B(t^*, T). \qquad (2)
\end{aligned}
$$

This is an integral equation for $\eta^*(t)$ if we are given all of the other parameters and functions, such as the market prices of bonds, $Z_M(t^*;T)$.

Equation (2) is easy to solve by differentiating the equation twice with respect to T. This gives

$$\eta^*(t) = -\frac{\partial^2}{\partial t^2}\log(Z_M(t^*;t)) - \gamma\frac{\partial}{\partial t}\log(Z_M(t^*;t)) + \frac{c^2}{2\gamma}\left(1 - e^{-2\gamma(t-t^*)}\right). \tag{3}$$

(Please don't get the idea from this that all models are easy to calibrate or all integral equations are easy to solve!)

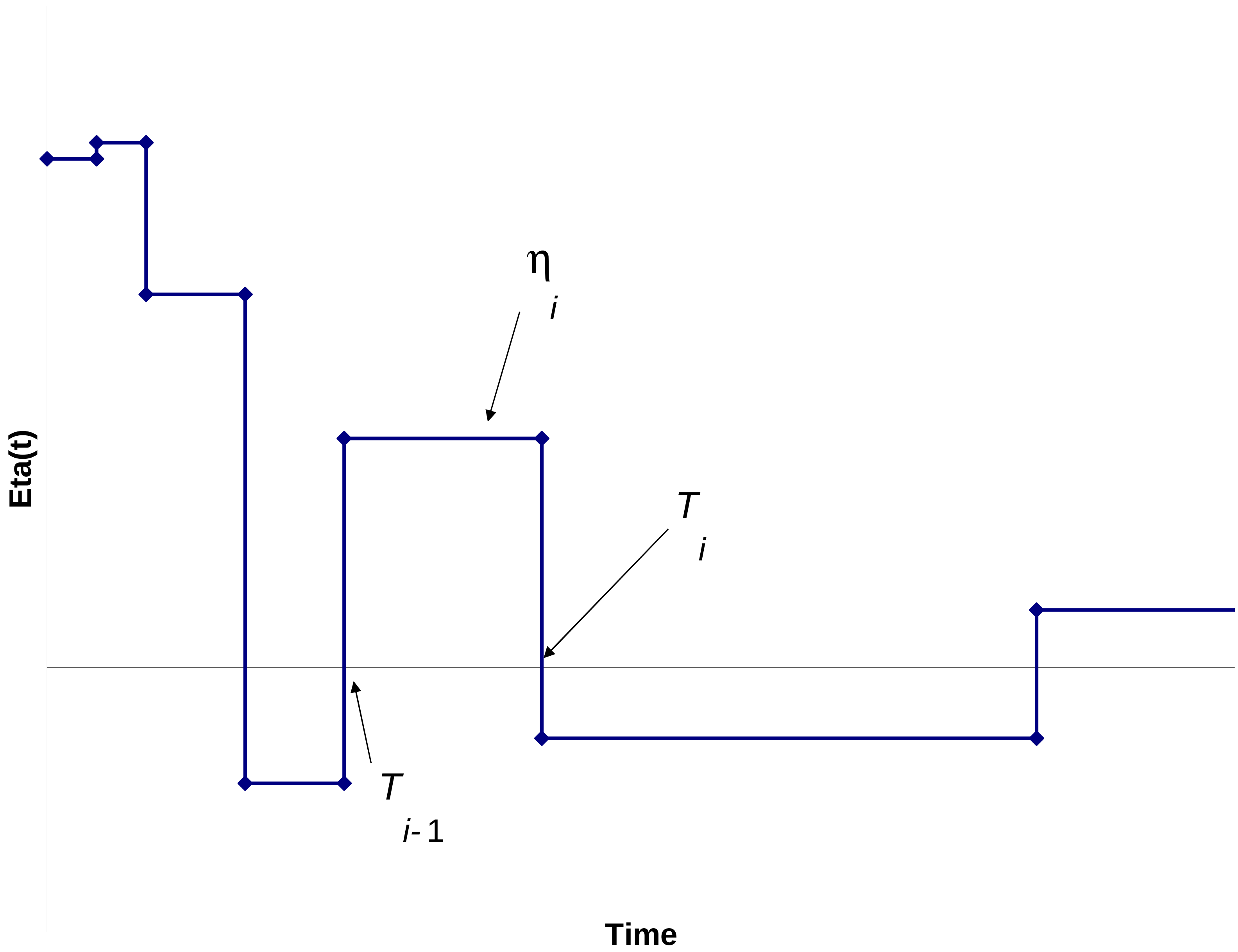
Back to Ho & Lee: Calibration in practice

In practice we do not have a differentiable yield curve, instead we have a finite number of bond prices. Just some dots, one per maturity.

To get a unique 'solution' for $\eta(t)$ we have to make some assumptions about its structure.

Examples: $\eta(t)$ is piecewise constant; $\eta(t)$ is piecewise linear and continuous; etc.

Let's assume that $\eta(t)$ is piecewise constant. It's the easiest to do the maths for!



Eta(t)
η_i
T_i
T_{i-1}
Time

$$\int_0^T \eta^*(s)(T-s)ds = -\log(Z_M(0;T)) - r^*T + \tfrac{1}{6}c^2T^3.$$

First η_1:

$$\int_0^{T_1} \eta_1(T_1-s)ds = \tfrac{1}{2}T_1^2\eta_1 = -\log(Z_M(0;T_1)) - r^*T_1 + \tfrac{1}{6}c^2T_1^3.$$

Therefore

$$\eta_1 = \frac{2}{T_1^2}\left(-\log(Z_M(0;T_1)) - r^*T_1 + \tfrac{1}{6}c^2T_1^3\right).$$

Next η_2:

$$\int_0^{T_1} \eta_1(T_2 - s)ds + \int_{T_1}^{T_2} \eta_2(T_2 - s)ds = -\log(Z_M(0;T_2)) - r^* T_2 + \tfrac{1}{6}c^2 T_2^3.$$

$$= \tfrac{1}{2}\eta_1 \left(T_2^2 - (T_2 - T_1)^2\right) + \tfrac{1}{2}\eta_2(T_2 - T_1)^2.$$

Therefore

$$\eta_2 = \frac{2}{(T_2 - T_1)^2}\Big(-\log(Z_M(0;T_2)) - r^* T_2 + \tfrac{1}{6}c^2 T_2^3$$
$$-\tfrac{1}{2}\eta_1 \left(T_2^2 - (T_2 - T_1)^2\right)\Big).$$

Generally, the left-hand side of equation (1) becomes

$$\sum_{i=1}^{j} \eta_i \int_{T_{i-1}}^{T_i} \left(T_j - s\right)\, ds$$

$$= \frac{1}{2} \sum_{i=1}^{j} \eta_i \left(\left(T_j - T_{i-1}\right)^2 - \left(T_j - T_i\right)^2 \right).$$

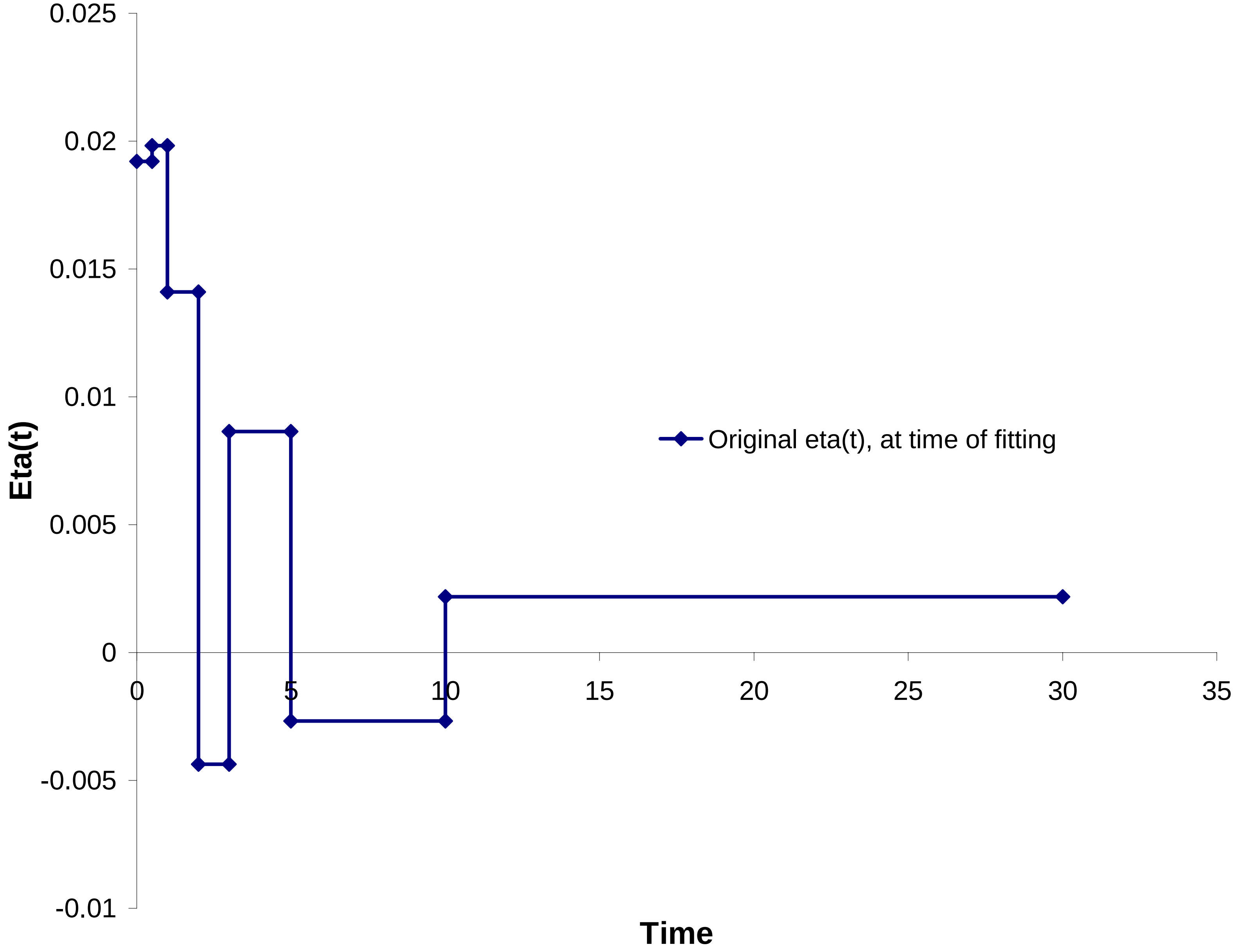
$$= \frac{1}{2}\eta_j \left(T_j - T_{j-1}\right)^2 + \frac{1}{2} \sum_{i=1}^{j-1} \eta_i \left(\left(T_j - T_{i-1}\right)^2 - \left(T_j - T_i\right)^2 \right).$$

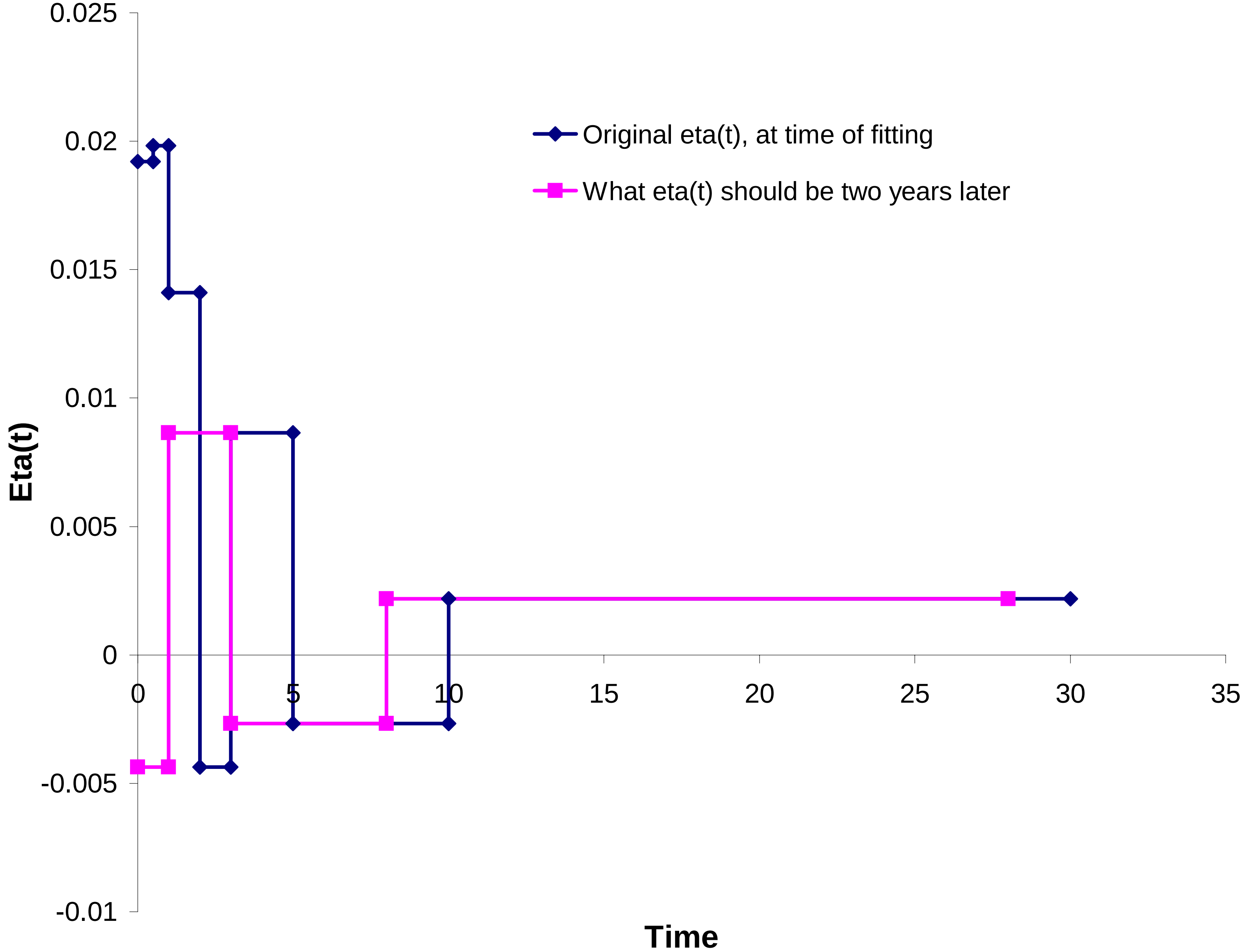
And so. . .

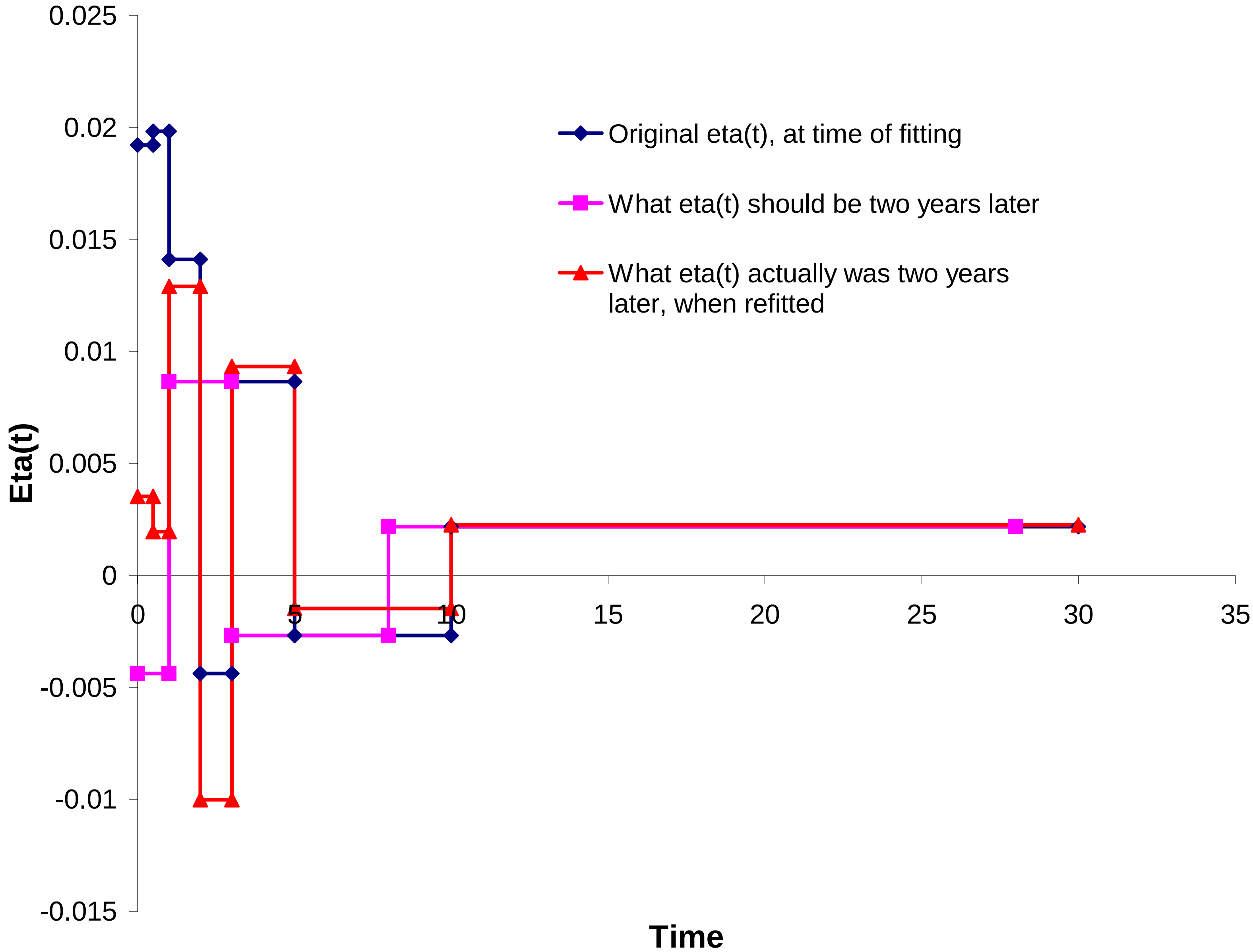
$$\eta_j = \frac{2}{\left(T_j - T_{j-1}\right)^2}\left(-\log\left(Z_M(0;T_j)\right) - r^* T_j + \tfrac{1}{6}c^2 T_j^3\right.$$

$$\left. -\tfrac{1}{2}\sum_{i=1}^{j-1}\eta_i\left(\left(T_j - T_{i-1}\right)^2 - \left(T_j - T_i\right)^2\right)\right)$$

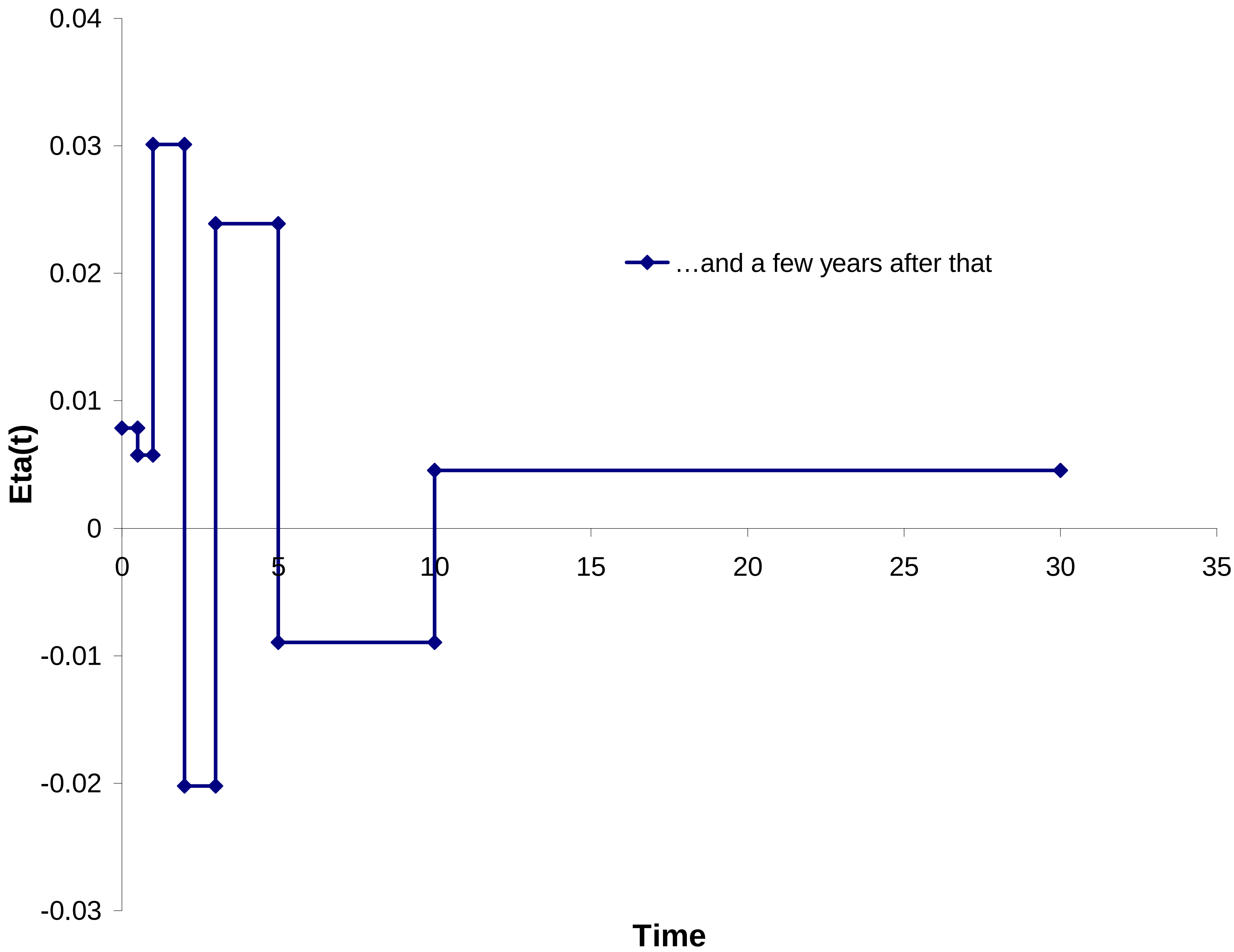
And this is what we find...


0.025
0.02
0.015
0.01
0.005
0
-0.005
-0.01
Eta(t)
0
5
10
15
20
25
30
35
Time
Original eta(t), at time of fitting


0.025
0.02
0.015
0.01
0.005
0
-0.005
-0.01
Eta(t)
0
5
10
15
20
25
30
35
Time
Original eta(t), at time of fitting
What eta(t) should be two years later



0.025
0.02
0.015
0.01
0.005
0
-0.005
-0.01
-0.015
Eta(t)
0
5
10
15
20
25
30
35
Time
Original eta(t), at time of fitting
What eta(t) should be two years later
What eta(t) actually was two years later, when refitted


0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
Eta(t)
0
5
10
15
20
25
30
35
Time
…and a few years after that

Yield-curve fitting: For and against

For

- The building blocks of the bond pricing equation are delta hedging and no arbitrage. If we are to use a stochastic model correctly then we must abide by the delta hedging assumptions. We must buy and sell instruments to remain delta neutral. The buying and selling of instruments must be done at the market prices. We *cannot* buy and sell at a theoretical price.

- Perhaps by hedging with other instruments the dependence of the model on its parameters and assumptions is reduced anyway.

Against

If the market prices of simple bonds were correctly given by a model calibrated at time t^* then, when we come back a week later, $t^* +$ one week, say, to refit the function $\eta^*(t)$, we would find that this function *had not changed* in the meantime.

This *never* happens in practice. We find that the function η^* has changed.

This means that the original model was incorrect.

One problem with calibration is that because it matches an instantaneous 'snapshot' of the market it is difficult to see how wrong it is!

And now for the opposite approach, analyze historical data. . . no reliance on a single 'snapshot.'

Data analysis to find the 'best' model

The one (or more)-factor models for the spot interest rate that we have seen were all chosen for their nice properties; for most of them we were able to find simple closed-form solutions of the bond-pricing equation.

Clearly, this means that the models are not necessarily a good description of reality.

Let's recap these models quickly. . .

Popular one-factor spot-rate models

The real spot rate r satisfies the stochastic differential equation

$$dr = u(r,t)dt + w(r,t)dX. \tag{4}$$

Model	$u(r,t) - \lambda(r,t)w(r,t)$	$w(r,t)$
Vasicek	$a - br$	c
CIR	$a - br$	$cr^{1/2}$
Ho & Lee	$a(t)$	c
Hull & White I	$a(t) - b(t)r$	$c(t)$
Hull & White II	$a(t) - b(t)r$	$c(t)r^{1/2}$
General affine	$a(t) - b(t)r$	$(c(t)r - d(t))^{1/2}$

Here $\lambda(r,t)$ denotes the market price of risk. The function $u - \lambda w$ is the risk-adjusted drift.

For all of these models the zero-coupon bond value is of the form $Z(r,t;T) = e^{A(t,T)-rB(t,T)}$.

The time-dependent coefficients in all of these models allow for the fitting of the yield curve and other interest-rate instruments.

From now on in this lecture we will see how to deduce a model for the spot rate from data; it is therefore unlikely to be nice and tractable!

The method

The method that we use assumes that

- the model is time homogeneous

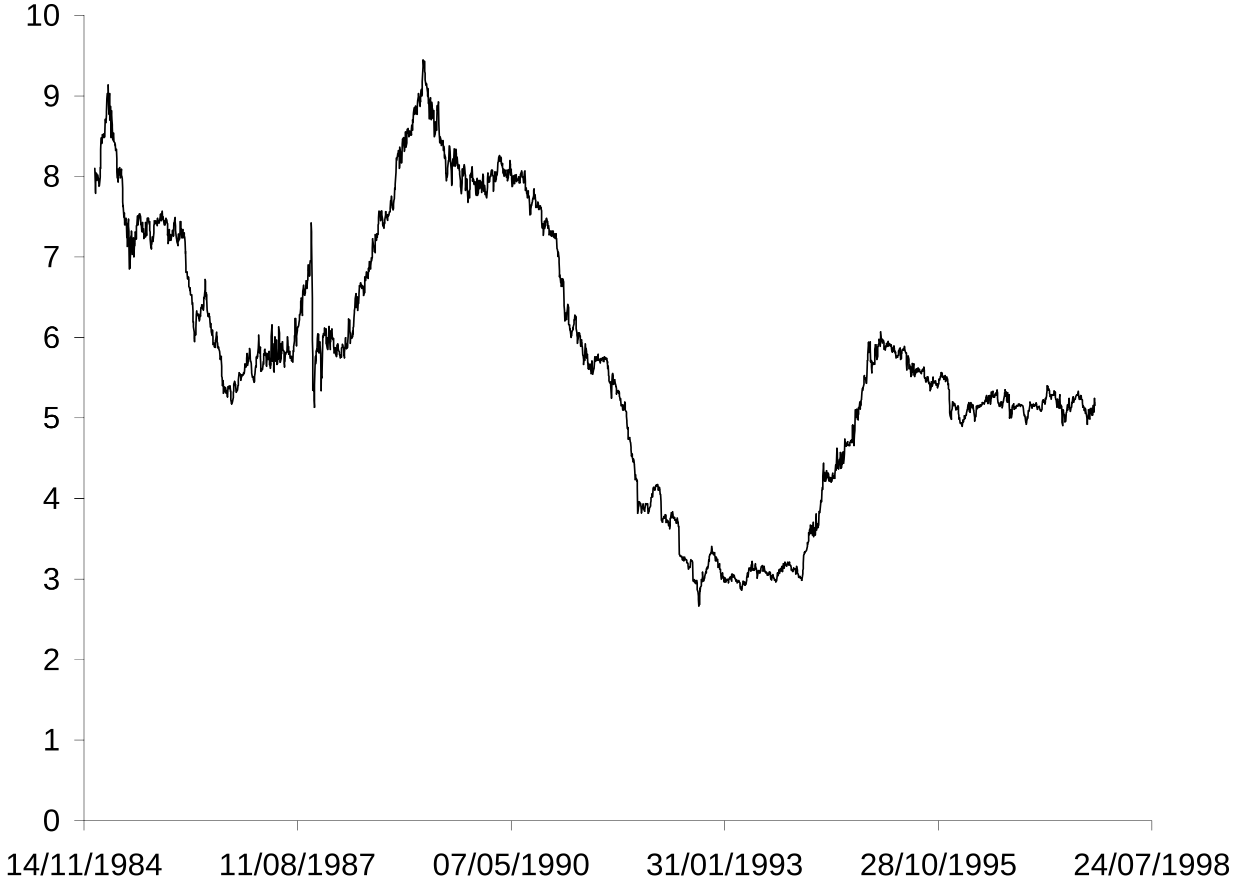
- the spot rate is well behaved (i.e. doesn't wander too far away, go to zero or infinity for example)

The downside to the resulting model is that we cannot find closed-form solutions for contract values, the risk-neutral drift and the volatility don't have a sufficiently nice structure.

In the figure are shown the US one-month LIBOR rates, daily, for the period 1985–1997, and is the data that we use in our analysis.

The ideas that we introduce can be applied to any currency, but here we use US data for illustration.

Aside: This method isn't specific to interest rates, it has also been used to model the gold price, equity and index volatility, and the rate of inflation.


10
9
8
7
6
5
4
3
2
1
0
14/11/1984
11/08/1987
07/05/1990
31/01/1993
28/10/1995
24/07/1998

There are three key stages:

1. By differencing spot rate time series data we determine the volatility dependence on the spot rate $w(r)$.

2. By examining the steady-state probability density function for the spot rate we determine the functional form of the drift rate $u(r)$.

3. We examine the slope of the yield curve to determine the market price of risk $\lambda(r)$.

The volatility structure

Our first observation is that many popular models take the form

$$dr = u(r)dt + \nu r^{\beta} dX. \tag{5}$$

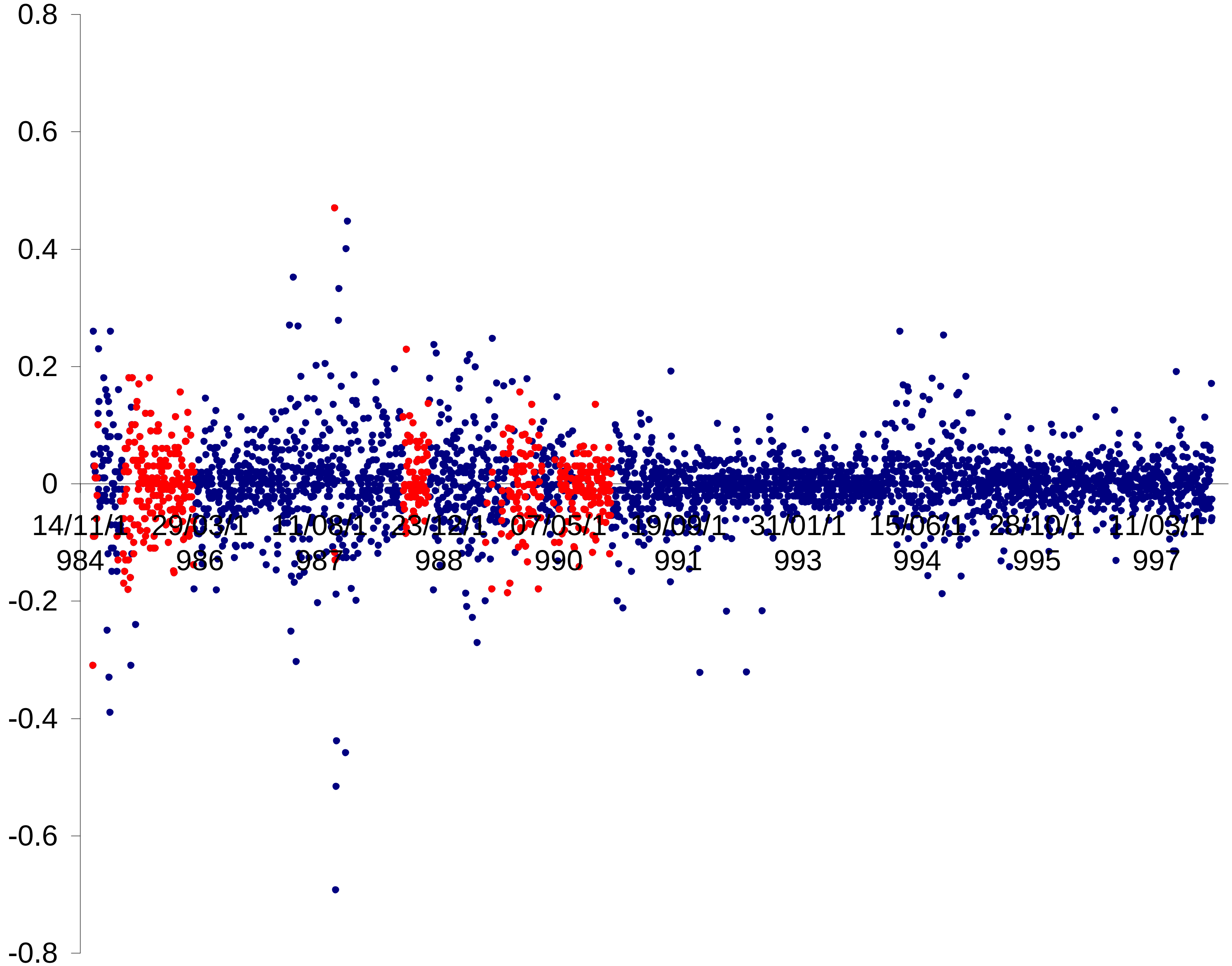
Examples of such models are the Ho & Lee ($\beta = 0$), Vasicek ($\beta = 0$) and Cox, Ingersoll & Ross ($\beta = 1/2$) models.

Using our US spot rate data we can estimate the best value for β using a very simple bucketing technique.

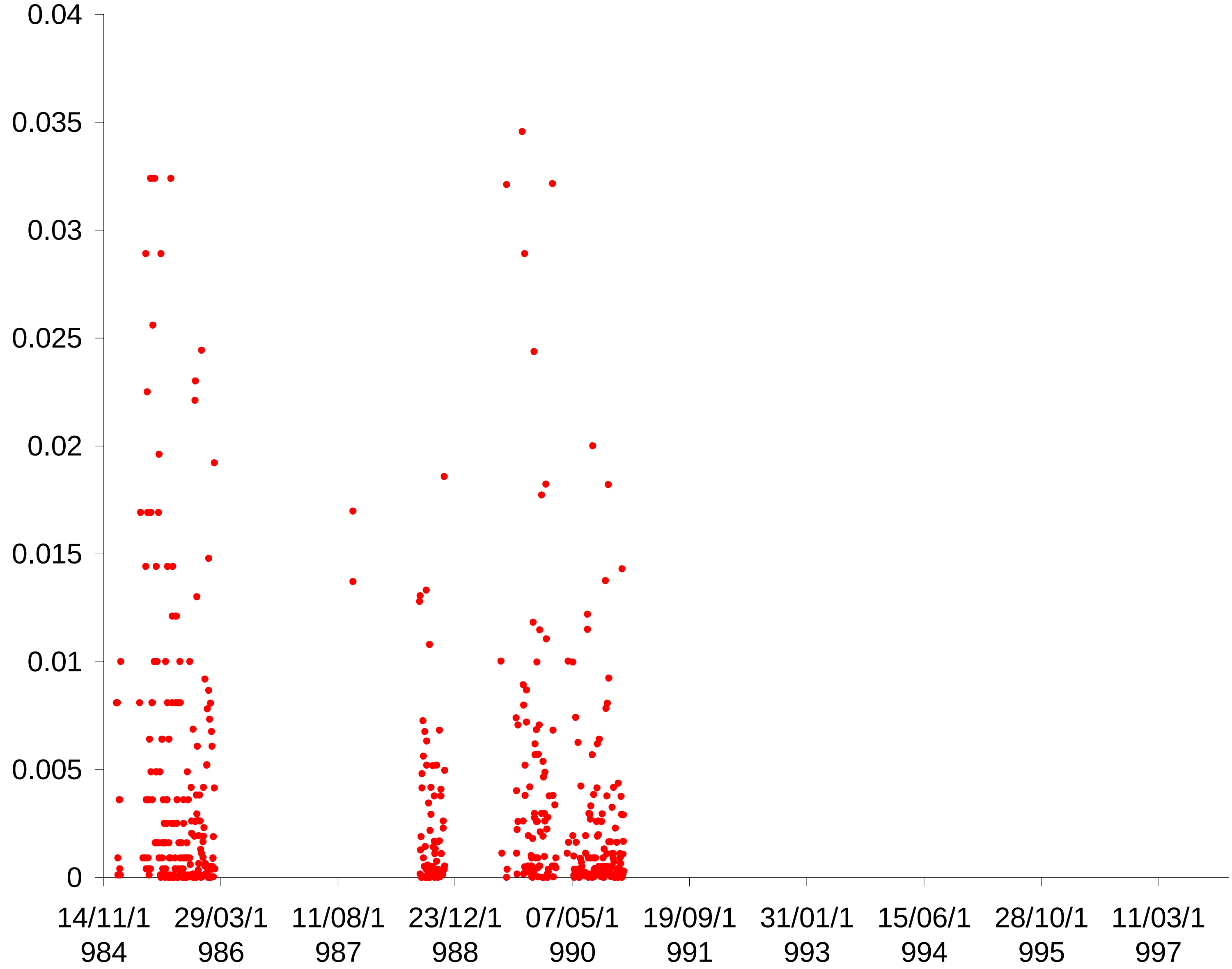
From the time-series data divide the changes in the interest rate, δr, into buckets covering a range of r values.


10
9
8
7
6
5
4
3
2
1
0
14/11/1984
11/08/1987
07/05/1990
31/01/1993
28/10/1995

Just examine the δrs associated with each bucket.


0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
14/11/1984
29/03/1986
11/08/1987
23/12/1988
07/05/1990
19/09/1991
31/01/1993
15/06/1994
28/10/1995
11/03/1997

Then calculate the average value of $(\delta r)^2$, for each bucket.


0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
14/11/1 984
29/03/1 986
11/08/1 987
23/12/1 988
07/05/1 990
19/09/1 991
31/01/1 993
15/06/1 994
28/10/1 995
11/03/1 997

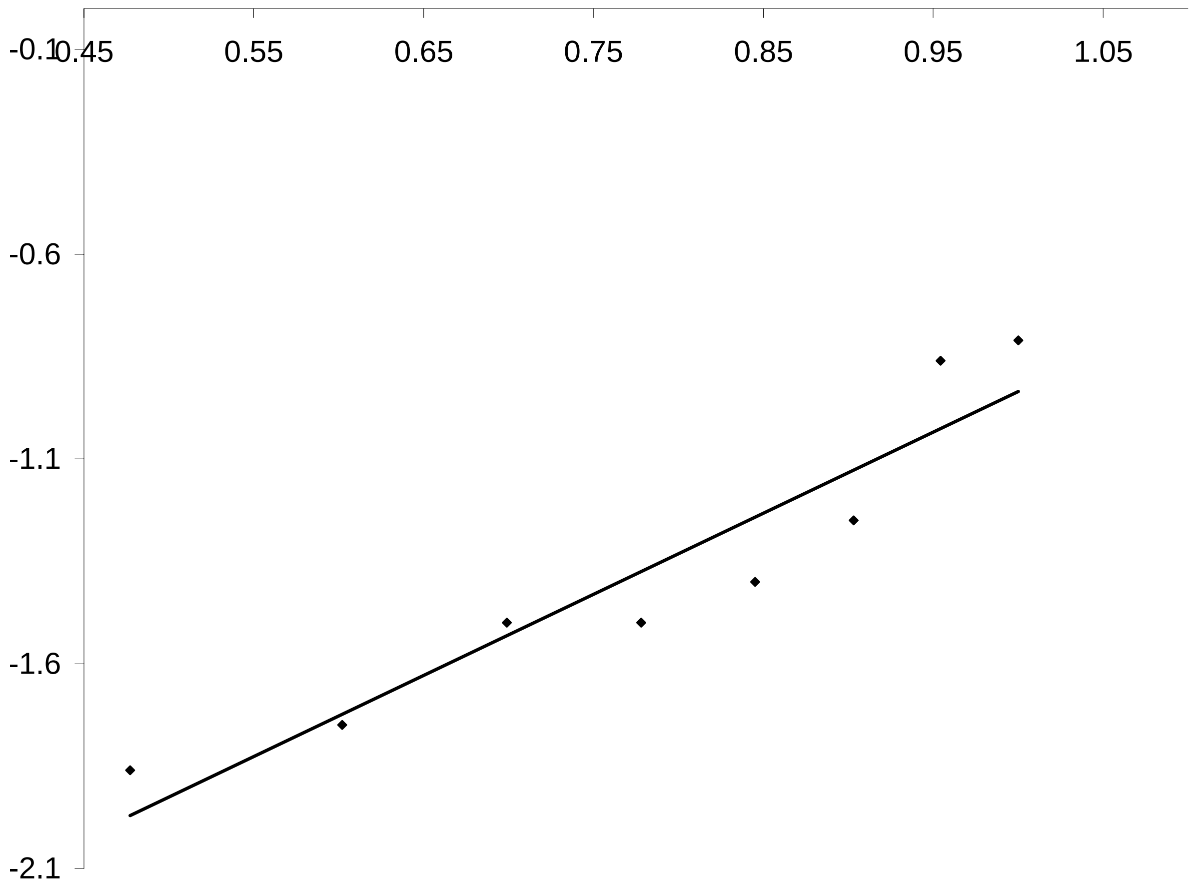
If the model (5) is correct we would expect

$$E[(\delta r)^2] = \nu^2 r^{2\beta} \delta t$$

to leading order in the time step δt, which for our data is one day.

Now plot $\log(E[(\delta r)^2])$ against $\log r$ using the data.

The slope of this 'line' gives an estimate for 2β and where the line crosses the vertical axis can be used to find ν.



Estimation of β

We can see that the line is very straight.

From this calculation it is estimated that

$$\beta = 1.13 \quad \text{and} \quad \nu = 0.126.$$

This confirms that the spot rate randomness increases as the spot rate increases, approximately linearly.

(And this rules out Vasicek, Ho & Lee, etc. etc.)

The drift structure

It is statistically harder to estimate the drift term from the data; this term is smaller than the volatility term and thus subject to larger relative errors.

Our approach to finding the drift function is via the empirical and analytical determination of the steady-state probability density function for r.

If r satisfies the s.d.e. (5) then the probability density function $p(r,t)$ for r satisfies the Fokker–Planck equation

$$\frac{\partial p}{\partial t} = \tfrac{1}{2}\nu^2 \frac{\partial^2}{\partial r^2}(r^{2\beta} p) - \frac{\partial}{\partial r}(u(r)p). \tag{6}$$

The steady state $p_\infty(r)$ will satisfy the time-independent version of (6):

$$\tfrac{1}{2}\nu^2 \frac{d^2}{dr^2}(r^{2\beta} p_\infty) - \frac{d}{dr}(u(r)p_\infty) = 0. \tag{7}$$

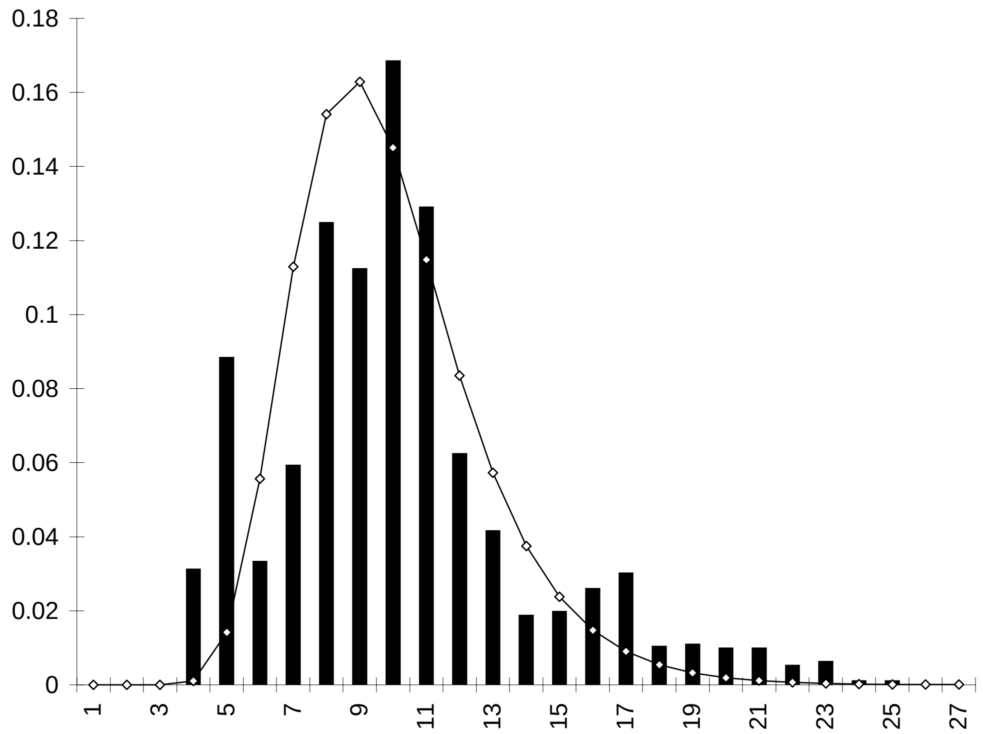
By integrating (7) we find the relationship between the steady-state probability density function and the drift function:

$$u(r) = \nu^2 \beta r^{2\beta-1} + \nu^2 \tfrac{1}{2} r^{2\beta} \frac{d}{dr} (\log p_\infty).$$

If we know one we can find the other.

Do we know $p_\infty(r)$?

We can determine a plausible functional form for $p_\infty(r)$ from one-month US LIBOR rates.



This figure shows empirical data, the bars, and a fitted function, the line.

Our choice for $p_\infty(r)$ is

$$\frac{1}{ar\sqrt{2\pi}}\exp\left(-\frac{1}{2a^2}(\log(r/\bar{r}))^2\right)$$

where $a = 0.4$ and $\bar{r} = 0.08$. From this we find that for the US market

$$u(r) = \nu^2 r^{2\beta-1}\left(\beta - \frac{1}{2} - \frac{1}{2a^2}\log(r/\bar{r})\right).$$

Advantages of working with the probability density function to find the drift function:

- more stable than other methods

- easy to see whether the probability density function ‘makes sense’

- spot rate cannot go to zero or infinity if probability density function zero there

The slope of the yield curve and the market price of risk

Now we have found $w(r)$ and $u(r)$, it only remains for us to find $\lambda(r)$.

We shall again allow λ to have a spot-rate dependence, but not a time dependence.

Note: There is no information about the market price of risk in the spot-rate process!

Such information is contained within instruments of finite (not infinitesimal) maturity.

We will examine the short end of the curve for this information.

Let us expand $Z(r,t;T)$ in a Taylor series about $t = T$, this is the short end of the yield curve.

We know that zero-coupon bonds satisfy

$$\frac{\partial Z}{\partial t} + \tfrac{1}{2}w(r,t)^2\frac{\partial^2 Z}{\partial r^2} + (u(r,t) - \lambda(r,t)w(r,t))\frac{\partial Z}{\partial r} - rZ = 0.$$

Look for a solution for small times to maturity of the form

$$Z \sim 1 + a(r)(T-t) + b(r)(T-t)^2 + \cdots.$$

Put this form into the bond pricing equation and equate powers of $(T-t)$ and you will find that

$$Z(r,t;T) \sim 1 - r(T-t) + \frac{1}{2}(T-t)^2(r^2 - u + \lambda w) + \dots \quad \text{as } t \to T.$$

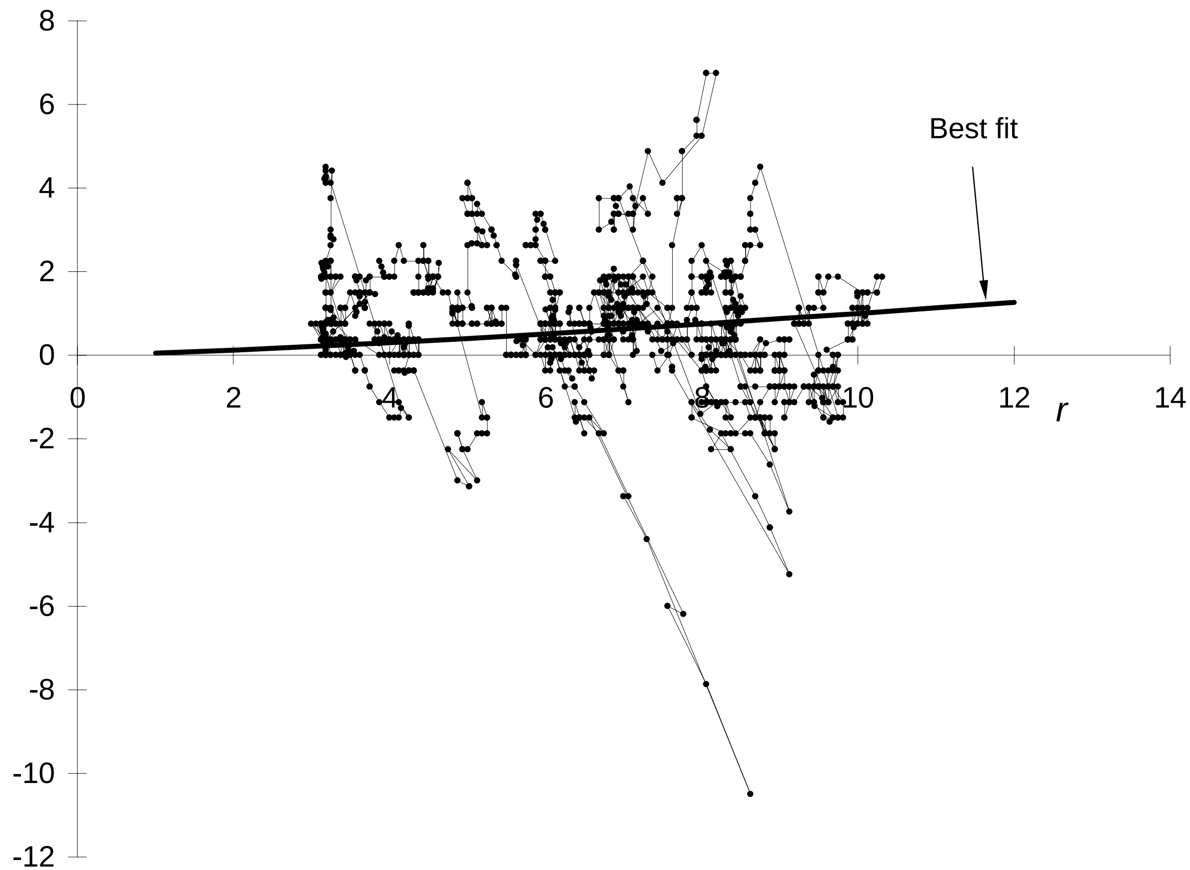
This is just a simple Taylor series approximation to the solution for a zero-coupon bond, *for any one-factor model*!

From this we can find the shape of the yield curve near the short end:

$$-\frac{\ln Z}{T-t} \sim r + \tfrac{1}{2}(u - \lambda w)(T-t) + \dots \quad \text{as } t \to T. \tag{8}$$

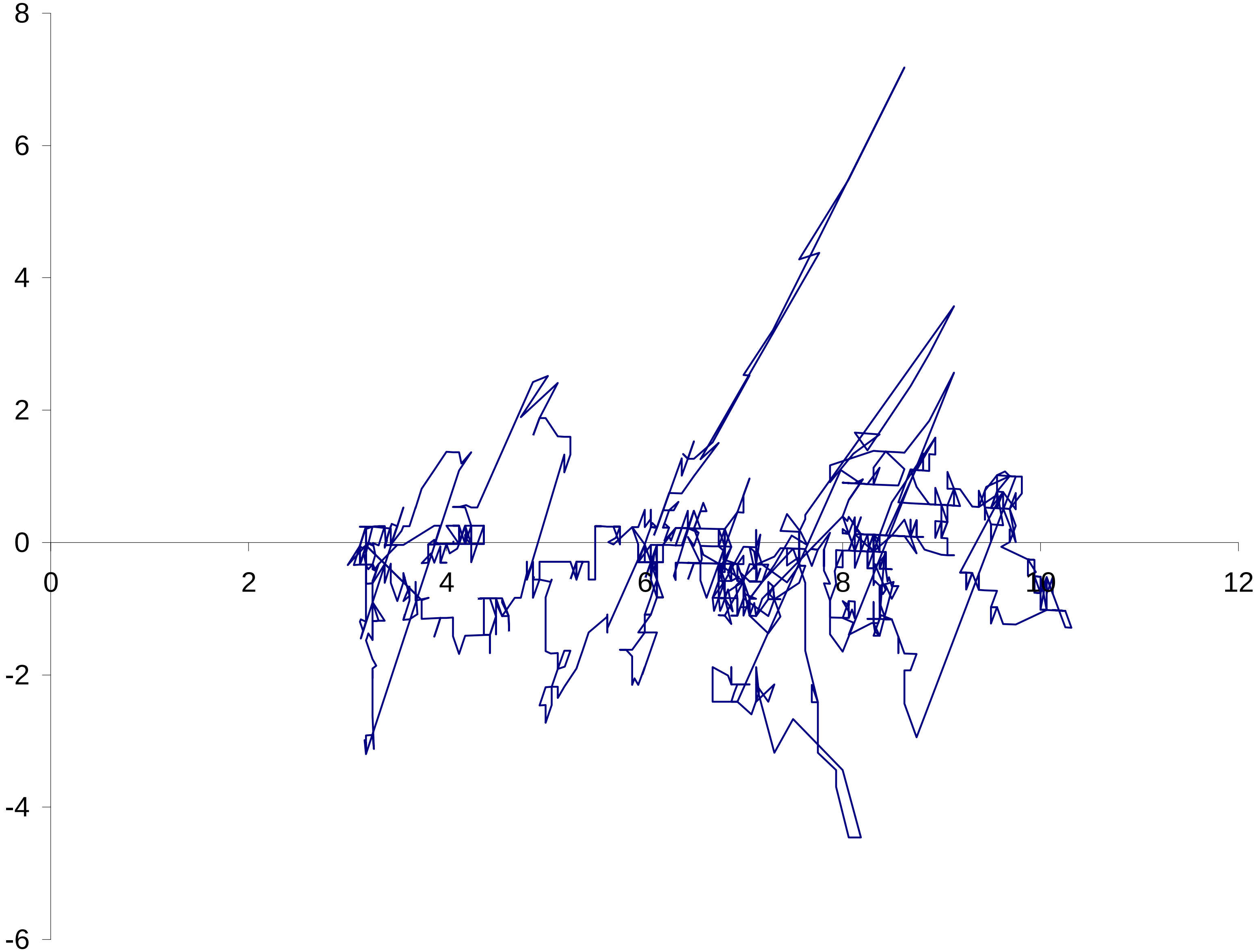
The first term says that the short end of the yield curve is r (obvious!), and the second term says that the slope of the yield curve at the short end in this one-factor model is simply $(u - \lambda w)/2$.

We can use this result together with time-series data to determine the form for $u - \lambda w$ empirically.

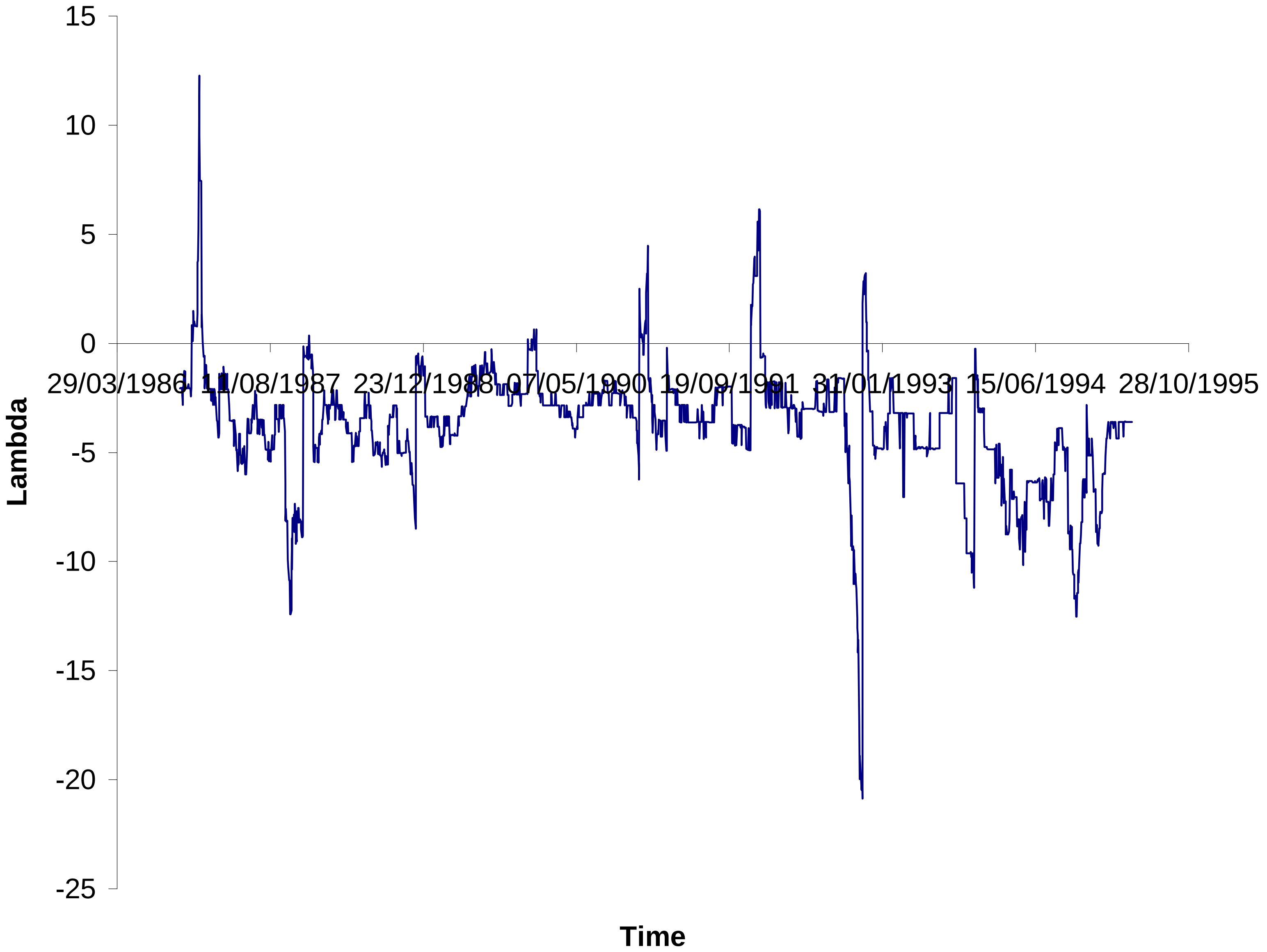


Slope of yield curve against spot rate and best fit, US LIBOR data

And so, the parameter λ as a function of r is . . .


8
6
4
2
0
-2
-4
-6
0
2
4
6
8
10
12

A mess!


15
10
5
0
-5
-10
-15
-20
-25
Lambda
29/03/1986
11/08/1987
23/12/1988
07/05/1990
19/09/1991
31/01/1993
15/06/1994
28/10/1995
Time

Possible conclusion from this:

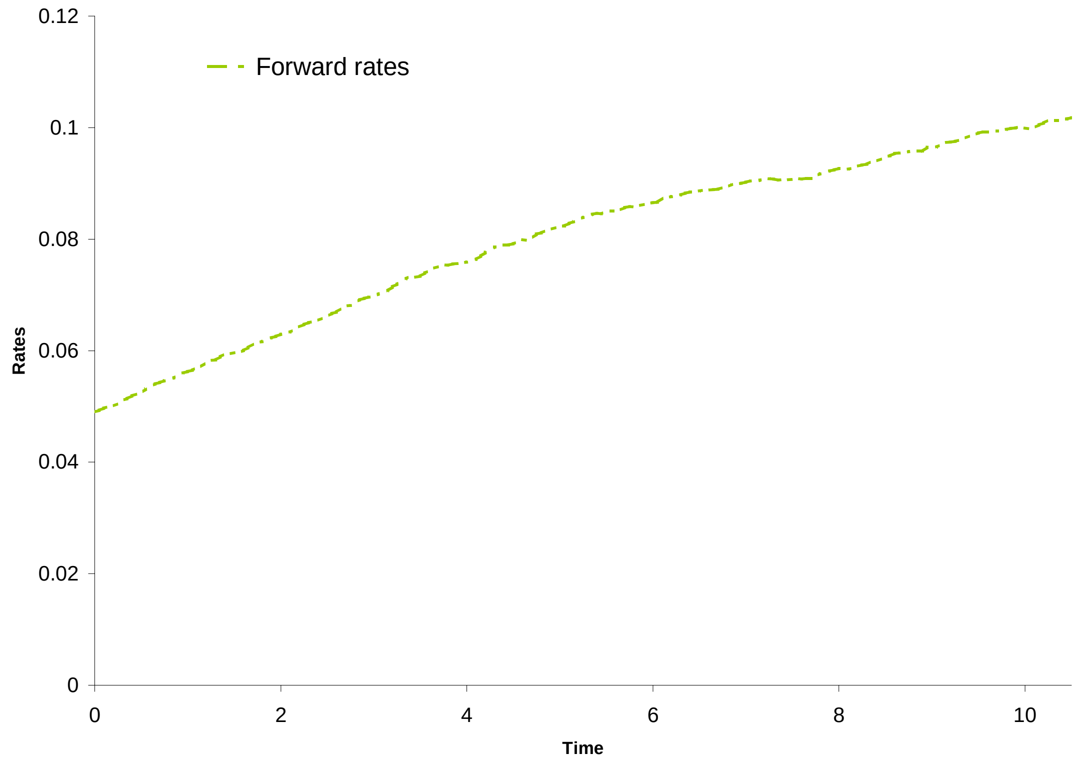
- it is ‘easy’ to model the spot interest rate!

- but difficult to model the market price of interest rate risk!!

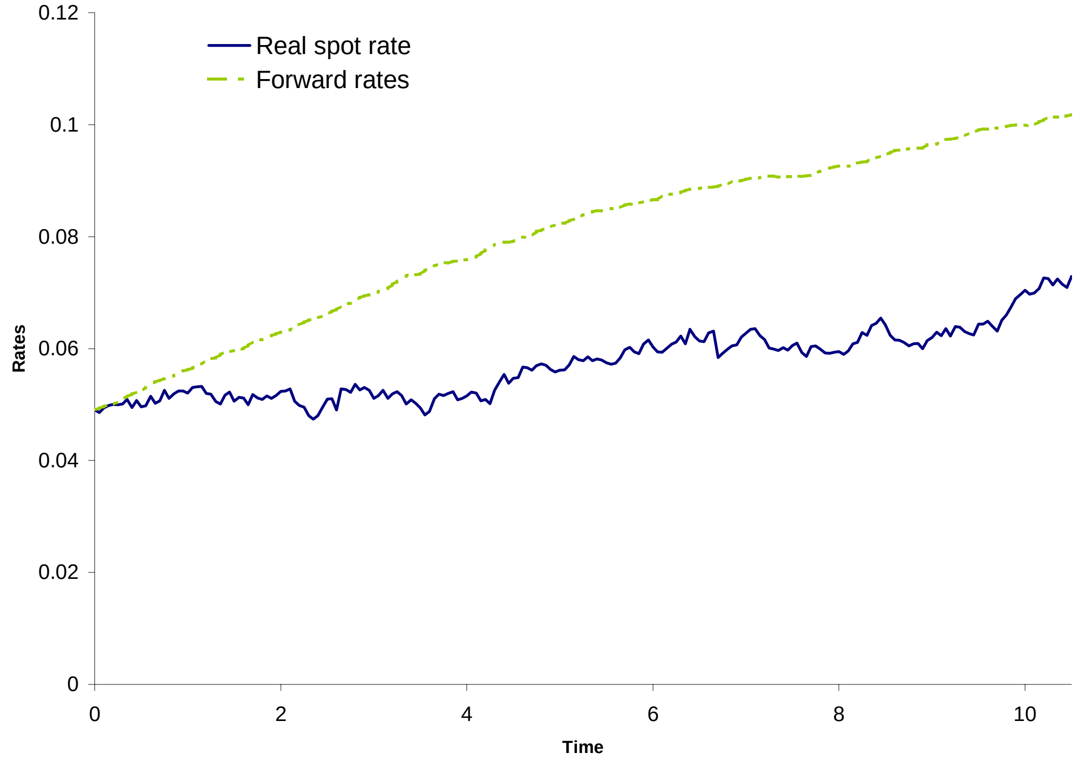
The relationship between forward rates and the real spot rate

What does the forward rate curve represent?

Question: Is the forward rate curve the market's expectation of the path of the future spot rate?



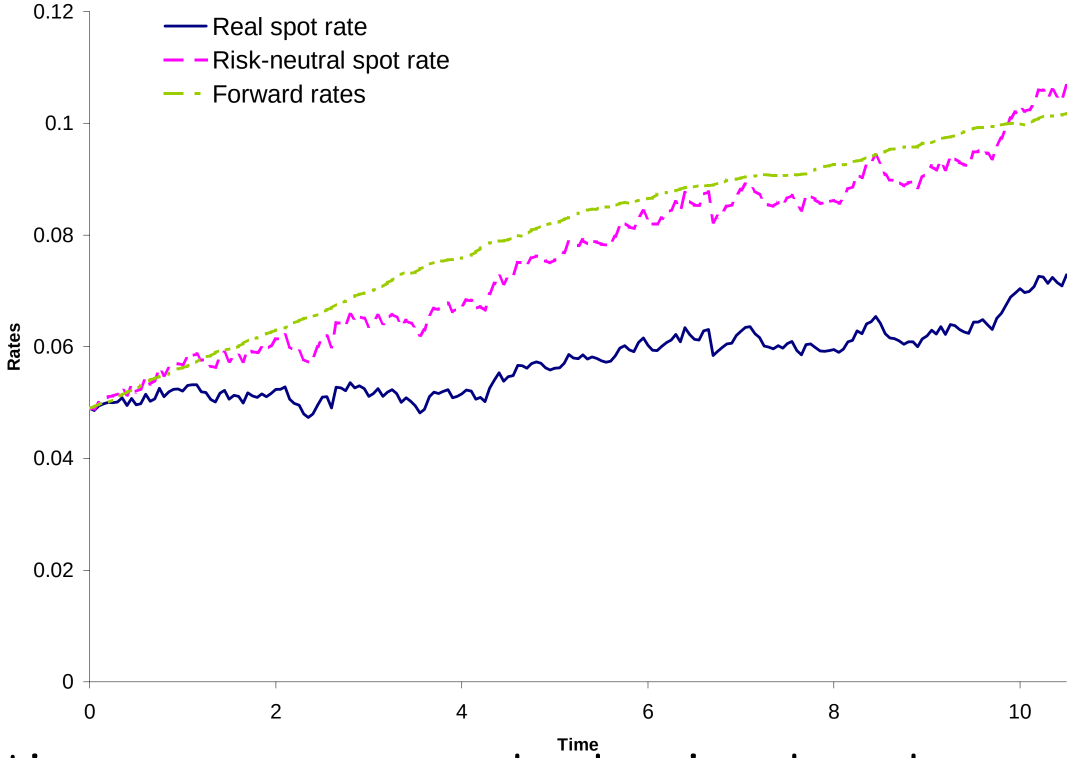
Well, no, because the spot interest rate tends to move more sideways than the typical forward curve.



How can we reconcile these two curves?

Easy. . . the forward curve is the expected future path of the *risk-adjusted* spot rate!

It allows for the valuation of risk!



If you tie your money up by buying bonds you expect a premium since you can't access your money in the meantime! (At least, not without risk.)

Summary

Please take away the following important ideas

- Spot interest rate models are usually calibrated to match market data, in particular the forward curve
- This calibration is in practice always inconsistent
- There are simple methods for examining interest rate data to find good models