Avoiding Model Errors and Pitfalls

PRESENTED TO: Quant Congress USA

BY: Martin Goldberg, Director ***martin.goldberg@citigroup.com***
Head of Model Validation
Risk Architecture
Citigroup
New York, New York

DATE: November 8, 2005

PLACE: New York City

citigroup

Caveats

1. These are my own personal opinions, and may not reflect the views of Citigroup. It may be that some of the Citigroup quants disagree with some of the material here.

2. This presentation is more in the nature of suggestions about how to build quant models than it is about any particular model or technique.

3. My last slide has the numbered references, and most of them are the URLs of pdf files.

Outline of this Talk

Good, bad, and pointless models

Testing assumptions and implementations and respecting your data

Beyond the Multivariate Gaussian Stationarity approximation

Right-sizing the number of parameters

Summary

Outline of this Talk

Good, bad, and pointless models

Testing assumptions and implementations and respecting your data

Beyond the Multivariate Gaussian Stationarity approximation

Right-sizing the number of parameters

Summary

Types of Model Error 1

- Mis-specified Market Dynamics
  - Oversimplified Assumptions
    - →neglecting jumps, skew, seasonality, etc.
    - →Too few stochastic processes (e.g. constant correlations for long-dated options)
  - Overcomplicated Assumptions
    - →Assuming calibrating instruments are all perfectly liquid
    - →Too many stochastic processes (separate skewness for each strike at each tenor)
    - →So complicated a model that it is analytically intractible or impossible to get to work
  - Assuming the dynamics are the same as for the index or the same as in the US market (Examples include single stocks vs. S&P, bespoke tranches vs. iTraxx, CAD vol skews vs USD vol skews)
- Misrepresentation of Security
  - (e.g. The contract states the herd-of-llamas feature has a triple axel but your model has only a double axel)

Types of Model Error 2

- Implementation errors and pitfalls
  - Example in C: if (X = 0) instead of if (0 == X)
  - Example in C++: Trying to use a yield curve object to store a vol surface
  - Example anywhere: The documentation is stale or missing, and the programmer is long gone
- Calibration Error
  - Deliberate Miscalibration
    - →"let's leave out that event – we'll never see anything like that again"
    - →"that must be a bad tick – the curve couldn't be that kinked"
    - →"the trader just sold at a **much** lower price – your model is overpricing; fix it!"
  - Local Minima in the fitting function
  - Unstable calibration - leads to wild swings in hedging prescription
  - Recalibrating too often or too infrequently
  - Calibrating to stale or fictitious market – if the model needs a price that doesn't exist then you should use a simpler, or at least a different, model

The Good, The Bad, and the Pointless 1

- No financial model is exactly correct – they are approximations to the behavior of many humans and a few computer programs buying and selling for heterogeneous reasons – some odd mix of stochastic calculus and psychology.
- A model is at best only as good as its assumptions. The assumptions depend on the underlying products, the intended purpose, liquidity, and your firm's preferences.
- It is rarely the case that one model is best for all uses.
- No matter how good the model is in some academic sense, if it never gets taken off the shelf and used it's pointless.
- Most of us have deadlines to meet. Very complex models are harder to implement. Remember Hofstadter's Rule, which states that everything takes longer than you think it will, even after you take Hofstadter's Rule into account.

The Good, The Bad, and the Pointless 2

- A good model is one that works for its intended purpose.
  - For an exotics customer facilitation book, a good model is one that has minimal hedge slippage throughout the lifetime of the deal. This can be tested with historical regression tests. As pointed out by Green and Figlewski[1], the regression test should recalibrate the model as often as the desk would do so in reality – usually daily, but sometimes weekly.
  - For a floor trader or spot FX or other very-high-frequency books where you usually don't need to cross-hedge, the key thing is to be fast, and to faithfully capture immediately all the transient glitches or spikes caused by large trades. These spikes can wreak havoc on an exotics hedge, which usually needs well-behaved first and second derivatives of the curves and surfaces.
  - For a proprietary statistical arbitrage book, the relative probabilities of all outcomes of the arb during the entire period from inception to close-out should be assessed with the least possible error. Any risk factors where the desk has no view should be hedged as described above for a flow book.

The Good, The Bad, and the Pointless 3

- A "pointless" model is one that never gets used. I am not in academia, so I am referring to the trading floor. Keep your audience in mind. If the parameters have no interpretation in terms of trader intuition they won't like them, and won't trust the model.
  - An example here is local vols versus stochastic vols. The stochastic vol parameters are non-intuitive and can't guarantee always pricing every vanilla option, but by construction the local vol surface ties exactly to vanillas at all strikes. Local vol gets the dynamics wrong, but it is much more popular since at least it gets all the vanillas right. The Heston parameters are not "floppy" enough, and how can a trader have confidence in an exotic price when the vanilla is wrong? However, a few traders prefer the hedge behavior from a stochastic vol model.
  - Trader intuition is about as good as quant modeling. If you cannot get your model to reflect qualitative trader intuitions about how the market behaves, try a different technique.
  - Every parameter should tell a story. Graphics are very useful here.
  - Don't be afraid of using a histogram or spline instead of a parametric representation - MC can handle lookup tables, and there usually isn't a good closed form solution anyway.

Rabin's Rules

(Mike Rabin was my boss in 1991)

Curiously, an electrician who installed an outlet in my basement had these same 3 rules for his work.

1. Pay Attention
   - What are the features you are trying to model? Did you use the right day-count conventions? What did the client actually ask for?
2. Think About What You Are Doing
   - You are going to dinner at Nobu in an hour, and the trading room PA system is broadcasting your favorite team's tie-breaking game. Neither of these should affect the nesting of parentheses on your if statement.
3. Double-Check Your Work
   - Limiting cases and paper trading simulations
   - Benchmarking against other models
   - Compiler warning messages, rereading the term sheet, etc.
   - A second set of eyes (My job at Citi)

Outline of this Talk

Good, bad, and pointless models

Testing assumptions and implementations and respecting your data

Beyond the Multivariate Gaussian Stationarity approximation

Right-sizing the number of parameters

Summary

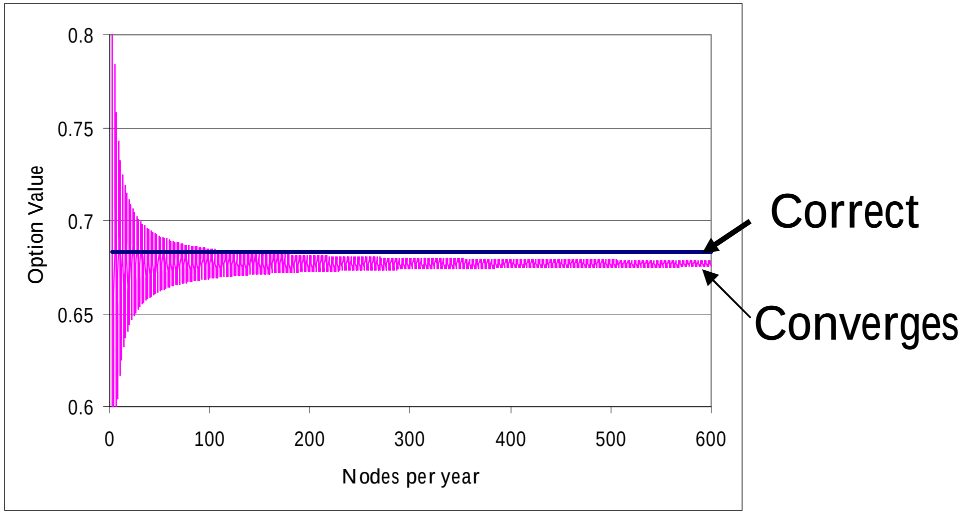
Testing Your Assumptions

- Getting a model to work right takes time. I will address here only models where you have the time to do it right, and not the ones where the trader demands a price before the customer comes back from lunch.
- Assumptions of the model under development should be verified at each stage. Know your clients!

1. Did you use some technique that your clients would object to?
   - The desk may always use some particular PDE solver or volatility paradigm.
   - There may be past bad experiences with some particular technique – e.g. the Green's function quant got divorced from the head trader last year.
2. Are your clients comfortable with estimating everything you have assumed was a trader input?
   - Price Verification policies need to be adhered to. This means conferring with back and middle office and not just the desk.
3. Document your model's assumptions. Do they look okay to your clients on paper?
   - Example: A model that assumes non-stochastic LIBOR is okay for traders of HY credits but not for investment grade desks.

Testing your Implementation

1. Does it converge to the right answer?
   - "right" may be a very subjective term.
   - Exotics prices are "fuzzy"
2. How quickly does it converge?



3. How well does it calibrate? Is the calibration getting stuck in a local minimum?
4. Is the hedge prescription relatively stable? Are there several areas of parameter space that fit the data about equally well(Broad shallow minima or multiple local minima)?
5. What stress events might cause the model to fail?
   - High Volatilities
   - Kinked or Spiky Curve / Surface, such as turn-of-year effects
6. Is there some alternate simple approximation to use as a benchmark? (This usually means Black-Scholes).

Respecting Your Data

1. Are there stylized facts that the model should reproduce (monotonicity, arbitrage-free, positive interest rates, etc)?
   - A more junior CDO tranche always has a higher loss probability than the more senior.
   - Smiles and skews flatten out for longer tenors.
2. Are there clever tricks in the literature to get your model to respect the above data that you may have overlooked?
3. Are there stylized facts that you were supposed to deliberately ignore?
   - Equity option vols decline for the week or so before expiry, but so erratically that any model of this effect is wrong.
   - Jumps / gap risks are handled separately by another model.
4. If you are using proxies, are the traders and control functions okay with your choices?
   - Historical vols / correlations proxying for implied
   - Liquid index smiles proxying for the single name vol curve shapes

Calibration and Dynamics Assumptions 1

- Are you calibrating the model to the appropriate instruments?
- Is the calibration set the same as the intended hedge instruments?
- Are you calibrating in the real world or in risk-neutral? Be careful.
  - Sometimes you have, in the same model, both liquid-market parameters calibrated risk-neutral and historically estimated parameters calibrated in the real world measure.
- Does your organization have a preferred framework for this to fit into?
  - Don't create unnecessary work for yourself by starting from scratch and conflicting with all the other models in your shop, unless you really have to.
  - If you create a new paradigm, does it extend to all the existing products done the old way?

Calibration and Dynamics Assumptions 2

- Does your minimization algorithm to calibrate account for the relative liquidities and bid-ask spreads of the various calibrating instruments?
- Is the algorithm robust and stable? Many of the statistical tests in the literature fall apart if anything is non-Gaussian or otherwise not i.i.d. (See for example the books [12] and [13])
- Does the model assume market dynamics that “smell wrong” to the trader?
  - Talk to the clients. Even quants need some people skills.
- What does the calibration error look like? Graphics are very useful. A recent example is [2]. It’s still very copyrighted so I am not going to cut and paste any of its pretty pictures.

Outline of this Talk

Good, bad, and pointless models

Testing assumptions and implementations and respecting your data

Beyond the Multivariate Gaussian Stationarity approximation

Right-sizing the number of parameters

Summary

Beyond the Multivariate Gaussian Stationarity Approximation

- The great success of Black-Scholes does not mean all the underlying assumptions were correct, nor does it mean that we cannot improve on it 31 years later. (BS assumes everything is lognormal with constant volatility in a perfectly liquid market with zero bid-ask spread, etc.)
- Lognormal processes have the advantage that they are easy to work with and never drop below zero.
- Ordinary Pearson correlations and Gaussian copulas are easy to work with and have very few parameters to calibrate.
- If you assume every new data point is a regime change you can't do any historical studies. If you assume no two assets have similar dynamics you can't do panel studies or use proxies. On the other hand regime changes can and do happen.
- Expediency and tractability can conflict with accuracy. This is a trade-off.

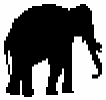
Beyond Gaussian for Univariate pdf's

- There are several alternatives to Gaussians and lognormals that I have used or seen used:
  1. Mixture-of-normals
  2. Generalized Gaussians
  3. Confluent Hypergeometric (Constant Elasticity of Variance), Bessel functions, Student-t, and other non-Gaussian "Special Functions" that integrate to one so you can use them as a probability density
  4. Empirical Histograms,usually smoothed using some kernel or spline
  5. Jump – Diffusions
- Don't settle on your approach based on only one day's data – the markets change and you want one that usually works okay, and not a model that's perfect sometimes and lousy other times.

Mixture Distributions: Linear Combination of Elephants

Mixture-of-normals or mixtures of Gaussians and lognormals is analytically tractable. You have Poisson jumps between several Gaussians of different widths with associated probabilities which must add to one. Each Gaussian has 2 parameters μ and σ. This approach is common, but I am not sure why. Taking this approach to extremes we get the **Linear Combination of Elephants Theorem:**

> Any closed curve on the plane can be represented to any desired degree of accuracy by

adding or subtracting sufficient numbers of elephant outlin  in different sizes and orientations. You use much smaller ones, rotated, to subtract off the legs, trunk, tail, and tusks that stick out too far, recursiv

This is obviously not a good basis function to use. Can the same be said of Gaussians? You decide.

Generalized Gaussians 1

- This is my personal preference. Take a Gaussian pdf Z, and non-linearly transform it using a strictly increasing transformation. These include:
  - (My favorite) Tukey g × h distribution (see references [3],[4] from U of Penn), which

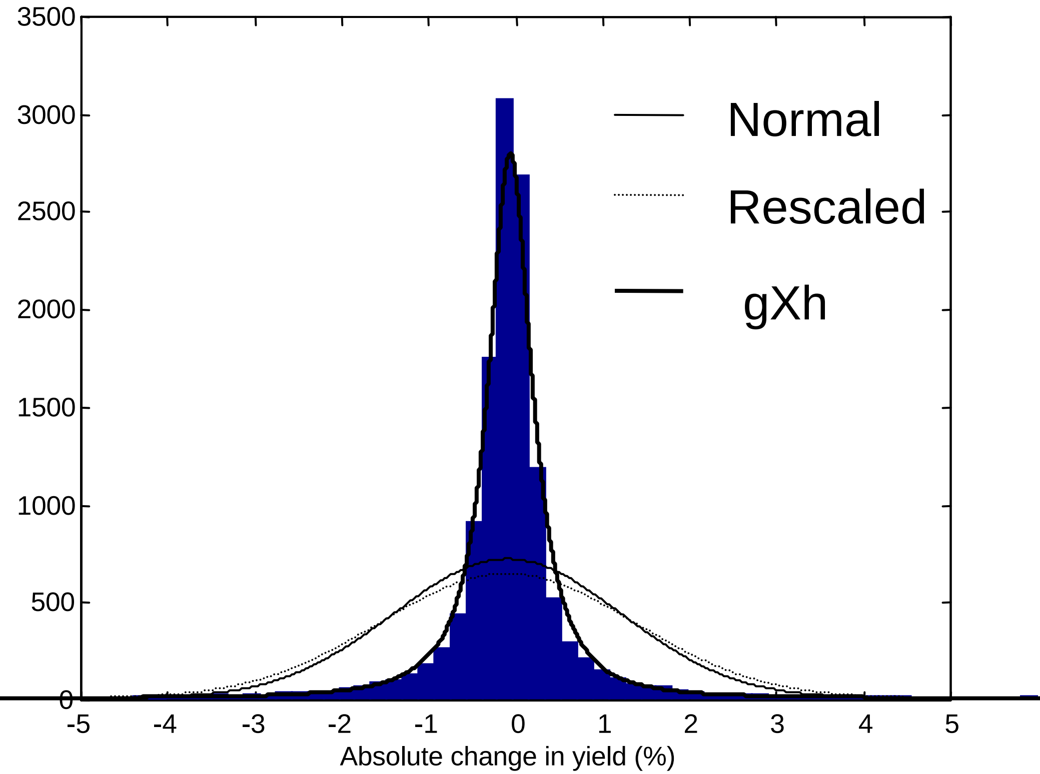
$$X_{gXh}(Z) = A + B\left(\frac{e^{gZ} - 1}{g}\right)e^{\frac{hZ^2}{2}}$$

    has g controlling skewness and h controlling kurtosis. g=h=0 is Gaussian, and g≠0,h=0 is lognormal. Options using this pdf have a closed-form solution. No theoretical justification, but it fits the data really well.
  - Fleishman transform[5] $X_{Fleish} = A + B\,Z + C\,Z^2 + D\,Z^3$ and its multivariate extension [6]. The parameters A, B, C, D do not have an intuitive interpretation.
  - The various Johnson, Tadikamalla, and Tadikamalla and Johnson transforms[7]. Any skewness and any kurtosis can be fit by this distribution, but not in an easy or intuitive way.

Generalized Gaussians 2

- Since most of the exotics you are working on probably have no closed form solution, or the closed form is so messy that you would do it numerically anyway, don't worry too much about solving the SDE analytically.
- Monte Carlo or quasiMC can of course handle any pdf you can specify.
- A binomial tree does not have enough degrees of freedom to converge to any arbitrary diffusion pdf.
- A trinomial tree probably doesn't either. Consider using a many-nomial tree [8] (explicit forward differences) or a many-nomial implicit finite difference grid – This is similar to a "willow tree". Dick Feynman and Mark Kaç reassure us that the numerically solved SDE fitting the pdf, and the numerically solved PDE, are the same thing. If you use implicit grids remember that you may need to impose artificial boundaries, either absorbing or reflecting. These false boundaries should be far enough out to not affect the results.
- Don't use trees or grids outer-producting 4 or more underlyings – the curse of dimensionality.

Fitting the Tukey distribution to single-B bond spreads



Comparison of normal, rescaled normal and (g **X** h) distribution fits to 10 day changes in idiosyncratic spread for single-B bonds using EJV data. Rescaled Cumulative Normal fits at 99th percentile.

Other special functions

Two cautionary notes here:

1. Explaining it to the client - why are you using such an unusual basis function, which the trader is not familiar with. You should have a compelling story.
2. Many such functions are quite time-consuming to calculate in the context of Monte Carlo simulations or grids. Especially if you need to use a numerical technique instead of a closed-form, be sure the run time doesn't make this model pointless.
   - If you have a closed-form solution, can you calibrate it?
   - Will it be arbitrage-free with respect to ALL the input rates, prices, vols, and intended hedge vehicles, or is it not flexible enough? How close does it need to be?
   - Can this function handle vol skews?

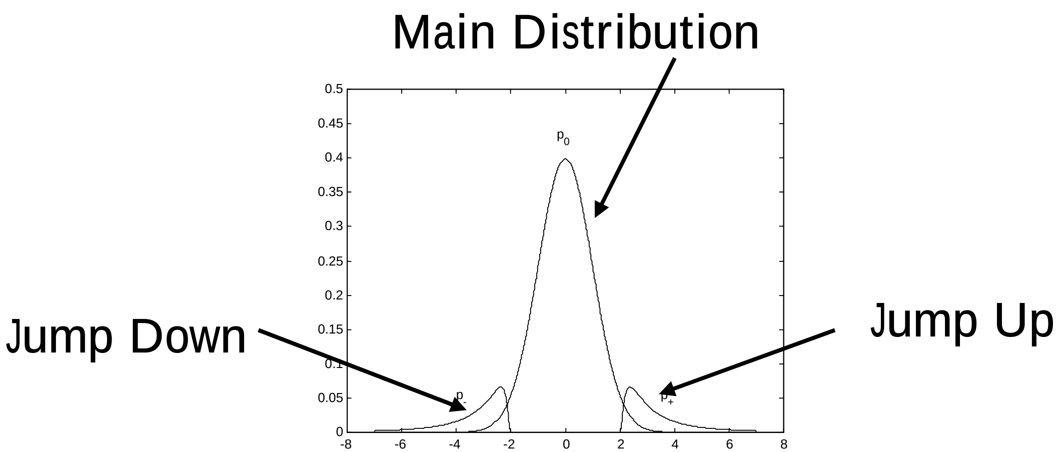
This is related to the distinction in between structural models (ab initio – find the true stochastic process) and reduced-form models (phenomenological - fit the data). Derman's book[10] dismisses structural models with "Many finance academics who should know better also seem to imagine that it can be done, but they don't live in the real world."

Smoothed Histograms

- The most extreme form of reduced-form model is to not have any equation at all for the functional form, but just use the data's actual shape. This is the rationale behind Historical Value-at-Risk.
- The cautionary note here is that any bad ticks or other data errors are exactly reproduced.
- Obviously no closed-form anything is likely to result from this, but Monte Carlo is ideally suited for this kind of table lookup.
- The discrete data may need to be smoothed / interpolated to a continuous function using your favorite spline or kernel or other method. Bu don't smooth away what should be jumps.
  - Familiarity with tricks in the academic literature is very helpful here.
  - There are many kernel smoothers besides the Gaussian kernel – Triangle, Epachenikov, and a host of others.
  - There are many different spline techniques – cubic, quintic, smoothing, monotone, taut/tense, etc.
  - various orthonormal basis sets to use for fitting, such as the Bernstein polynomial I will get to later.

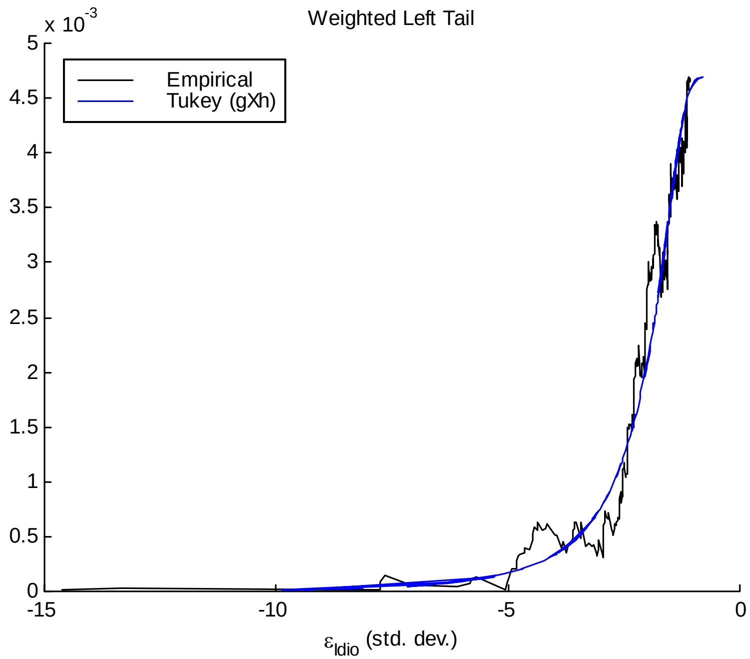
Jump Diffusions 1

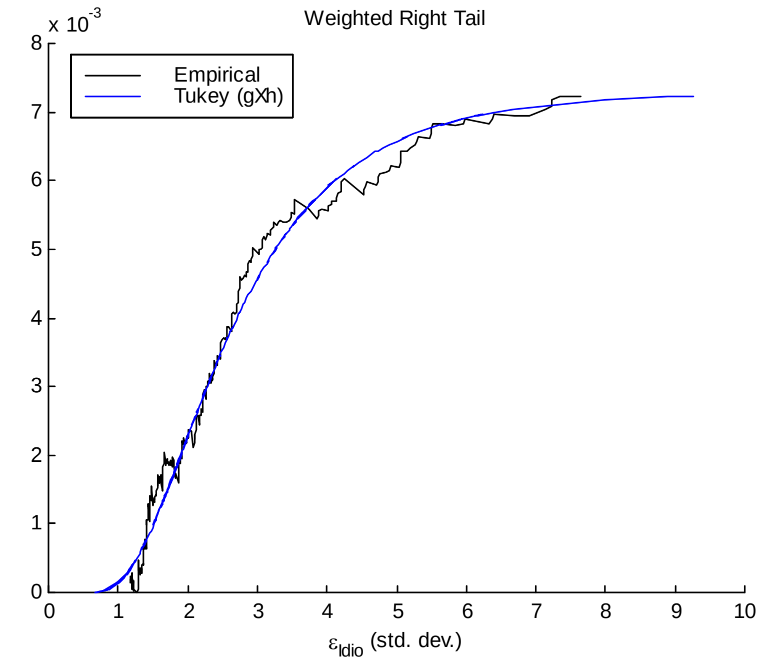
- Again, a possible description problem - when a jump occurs in the real world, there is a news story about it on the Bloomberg. Is your "jump" probability calibrating to the probability of actual events, which a trader has a gut feel for?
- In my opinion, you should not invoke jumps just to cover up deficiencies in the model of the diffusion part. This is called the principle of **basis set saturation** - be sure you have fully modeled the simpler part you are sure is there (diffusion) as well as can be done, before you add in higher order effects (jumps). Otherwise you're just adding "magic peanuts" to the elephant basis set.
- On the next slide is an example of real jumps. The idea here is that the jumps will cause spikes in the pdf tails.



Jump Diffusions 2 - Tukey g×h plus jumps

In this study of daily changes in idiosyncratic spread of single-B bonds (making the assumption that the historical data were stationary and ergodic), subtracting out a best-fit g×h diffusion left residual spikes in both tails, which we modeled as jumps distributed g×h, fitting jump amplitudes and g×h parameters to the data. You can see that the fit was really quite good.





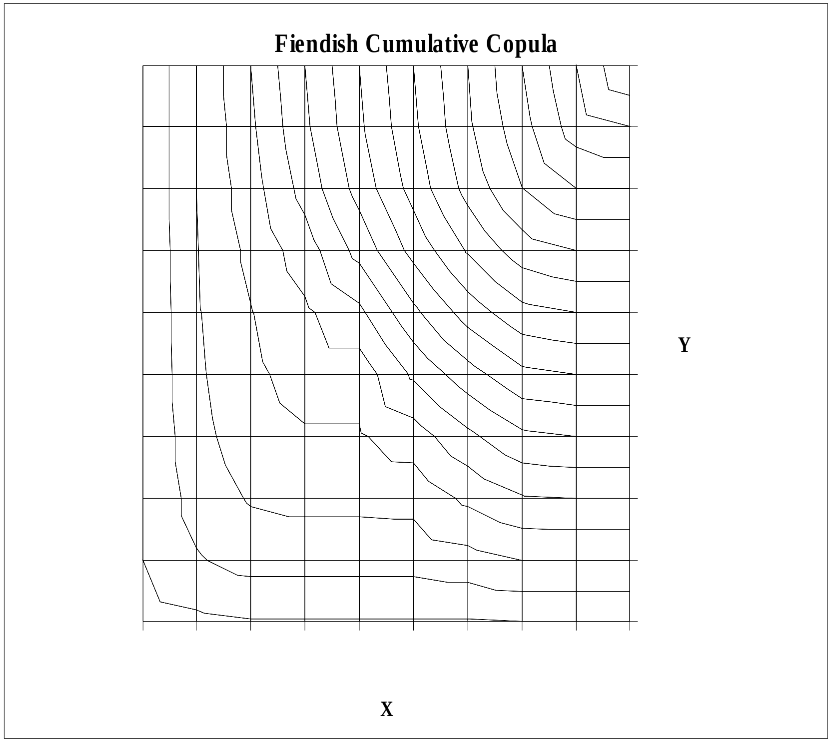
Stationarity

- Calibrating to live market data requires no assumptions of stationarity. However, if the market is very un-stationary, it has the same effect as unstable calibration - the hedge performance is lousy.
- Calibrating to a proxy or basket of proxies is a potential pitfall - how good is the proxy, and if you have too few data on the actual underlying, how can you tell if the proxy is good or not?
- When you use a time series estimate, you have several choices, including:
  - Assume stationarity and use all the data going back in time as far as possible.
  - Use a digital-signal-processing type filter. The most common one in finance is the exponentially weighted moving average used by RiskMetrics, where you don't actually calibrate the exponent, and there is no good reason to assume an exponential instead of cutting the filter with some other shape.
  - Statistical sampling error goes as $T^{-1/2}$. Assume the nonstationarity drift is linear (unless you have a better model for it) and then it goes as T. Then the total estimation error to minimize is $\lambda_1 T^{-1/2} + \lambda_2 T$. Estimate the λ from the data, and find the optimum data series length T^*. This is easiest if you have equal data weights from today back to T^* but you could modify this for some other filter shape. Preferably you would have some data-respecting reason to choose the particular filter shape. This technique is standard practice in atmospheric physics.

Multivariate Dependency 1

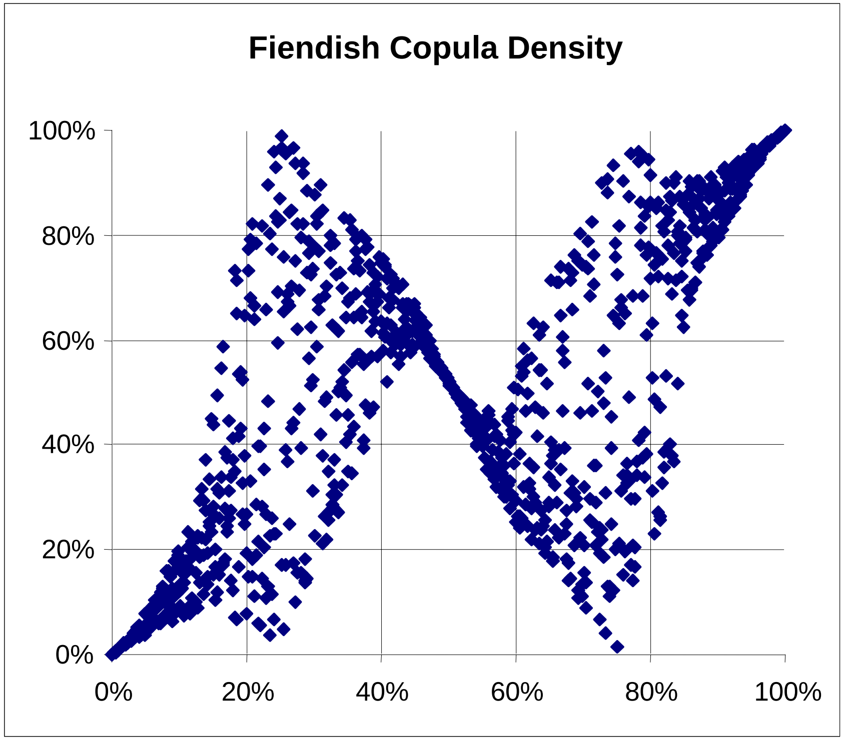
- The usual Pearson correlation is not a good measure for non-elliptical (skewed) distributions.
- Standard practice on the Street now is to use copulas, which can represent any possible association between variables, and don't depend on the marginal pdf's.
- The "usual" Gaussian copula is easy to work with, but with only one parameter it may not calibrate well.
- The econophysicists[14] have shown that in the equity markets, in good times spread trades are more popular, slightly lowering correlation, and in bad times a flight-to-quality raises correlations [9] - this is also called contagion.
- This is my personal counterexample showing what could go wrong. Here is the copula:

Fiendish Copula



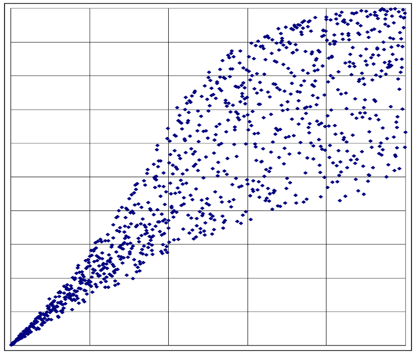
- It is not immediately obvious why this is so fiendish.
- Although all the theory is done using the cumulative distributions, the copula densities are more informative and make prettier pictures.
- Caveat: Copula theory is not nearly as well developed for more than 2 dimensions. The standard cheat scheme used in Credit Derivatives is to assume every underlying looks like all the others, which means all 2-dimensional slices look alike.

Fiendish Copula Density

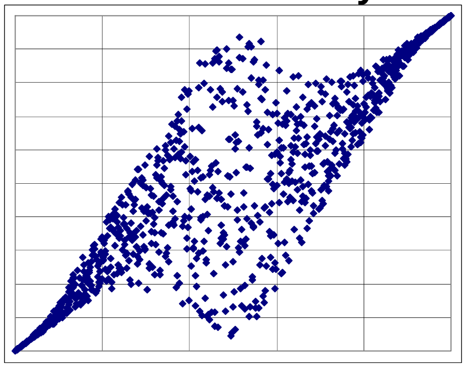


- Upper and lower tail dependence of 1; middle "local dependence" -1
- The rank correlation is constructed to be exactly zero.
- It is more pathological than what you will ever actually find, but it is a good stress test.
- You can find funnel-shaped and galaxy-shaped copula densities in real data, but in a less exaggerated form than below.

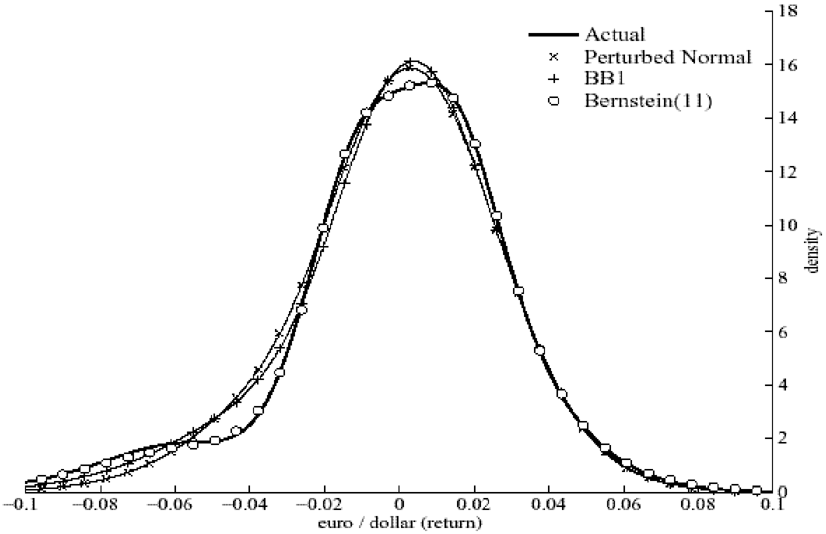
Extreme Funnel



Extreme Galaxy



A Copula which Respects the Data



- A recent example from the literature[11] is shown here. This April 2005 paper by Hurd, Salmon, and Schleicher is in my opinion well worth reading.
- In FX, triangular arbitrage gives you a liquid market in implied measures of association, so there is a unique correct copula that recovers the observed smiles.
- Here they fit the EUR/USD smile using EUR/GBP and GBP/USD smiles and various copulas.
- The 11-parameter Bernstein that fits the left tail does not have 11 story-telling parameters, but is more in the spirit of an interpolation that can be seen as graphics.
- This paper defers for future research going to more than 2 dimensions.

Outline of this Talk

Good, bad, and pointless models

Testing assumptions and implementations and respecting your data

Beyond the Multivariate Gaussian Stationarity approximation

Right-sizing the number of parameters

Summary

Right-sizing the number of parameters 1

- Flexible enough to fit data even in stress periods
- Appropriate model for the intended use
  - Perfect calibration vs. Smooth surface with smooth derivatives
  - Fast vs. Accurate
  - Illiquid actual data vs. Proxies
  - Historical vs. Implied
  - Parametrized vs. Histogram
  - Too few parameters and neglecting some effects vs. too many and fitting to noise
  - How many stochastic processes are needed for this product

    →FI example - a cap model probably only needs a 1-factor short rate model, but not so easy for an option on the difference in smiles between a 6-month straddle and a 1-year straddle
- Clients can believe the calibration

Right-sizing the number of parameters 2

- Every parameter tells a story
  - Parsimonious models with very few parameters are easier to understand, but every parameter needs a descriptive and convincing name
  - Curve fitters such as splines and kernels are acceptable even with zillions of calibration parameters, as long as you can produce good graphics, preferably interactive. These should fit the data much better than a simple model in order to be viable
  - Any representation between these 2 extremes is a harder sell.
- Graphical representation is extremely helpful – almost everyone likes good visuals
- Try a few functional forms to see which works best
- Understand what stresses will make your model collapse
  - Allowing for contagion may need too many parameters, but then if it happens you knew your model would go wrong
  - Vol term structures with a sharp enough decline imply imaginary forward vols in a simple model

Summary

- Pay Attention
- Think About What You are Doing
- Double Check Your Work
- Keep The Client Comfortable
- It's Only a Model, Not "Reality"
- Respect the Data
- Graphics Are Good
- Every Parameter should Tell a Story
- Backtesting, Especially During Historical Stress Periods, is a Good Idea
- There Are Lots of Useful Ideas in the Literature

References

1. Green, T.C. and S. Figlewski (1999), Market risk and model risk for a financial institution writing options, Journal of Finance, 54, 4, 1465-99. See also http://www.rotman.utoronto.ca/bicpapers/pdf/HullSuoIVF2.pdf
2. Sana ben Hamida and Rama Cont, J. Comp. Fin. 8(4) Summer 2005 pages 43 - 76
3. http://fic.wharton.upenn.edu/fic/papers/02/0225.pdf
4. http://fic.wharton.upenn.edu/fic/papers/02/0226.pdf
5. Fleishman, A.I. (1978) A method for simulating non-normal distributions,Psychometrika 43, 51-532.
6. http://swopec.hhs.se/hastef/papers/hastef0481.pdf
7. http://work.michalkaut.net/CV_and_study/SG_Heuristic_updates.htm and its references
8. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=259478
9. http://xxx.lanl.gov/abs/cond-mat?0107593 and his references 15 – 21
10. Emmanuel Derman, "My Life as a Quant" Wiley 2004, page 266
11. http://www2.warwick.ac.uk/fac/soc/wbs/research/wfri/wpaperseries/wf05-226.pdf
12. "Robust Statistics" by Peter Huber (Wiley 2003)
13. "Robust Nonparametric Statistical Methods (Kendall's Library of Statistics)" by T. P. Hettmansperger, J. W. McKean (Arnold 1998)
14. ArXiv.org and some physics journals see finance as a subset of condensed matter physics - see also http://www.econophysics.org/