DRAFT: PLEASE DO NOT QUOTE OR COPY

Independent components and hybrid models for asset prices *

Nicos Christofides, Maurice Klajman, Alexander Christofides

Centre for Quantitative Finance, Imperial College, London SW7 2BX
Corresponding author: n.christofides@imperial.ac.uk

September 2003

ABSTRACT

Modelling of asset 'values' (prices or returns) is an important problem in finance. We describe the orthogonal decomposition of asset value time series into principal and independent components and show that with fat-tailed asset value distributions, the principal components show high fourth order dependence (i.e. have high cokurtosis) whereas independent components do not. Fitting stochastic processes to independent components and implementing these processes with discrete state transition graphs, is therefore much simpler since these models can be considered separately for each asset. Reduction of dimensionality by using only a subset of independent components is investigated and the resulting errors analysed. These errors are themselves modelled by a nonlinear model of the chosen set of independent components. A neural network (NN) is used for this purpose in the present study. An overall hybrid model of asset values results by combining the parametric stochastic model of the independent components with the non-parametric NN.

Keywords: Principal & Independent components, Neural networks, Hybrid models

1 Introduction

One of the most basic and often encountered problems in quantitative finance is concerned with the modelling of asset 'values' (prices or returns). Asset values vary

*Support for this work was provided by the IST Framework 5 Programme of the European Union, Contract IST2000-29405, Eurosignal Project

stochastically with time, but the nature of the underlying price dynamics depends on the nature of the asset. Stock prices, for example, are reasonably approximated (to a first degree) by Geometric Brownian Motion (GBM) and Black and Scholes derived their famous option pricing formula by making this assumption, [Black and Scholes 1973]. A direct consequence of assuming GBM for stock prices is that asset returns at any future time t are normally distributed. Even a cursory examination of asset returns, however, reveals the existence of "fat tails" indicating that extreme returns happen with much greater frequency than predicted by the normality assumption. The observed returns distribution is leptokurtic (see figure 1).

If the "asset" is an interest rate (or a commodity like oil) then in the long run its

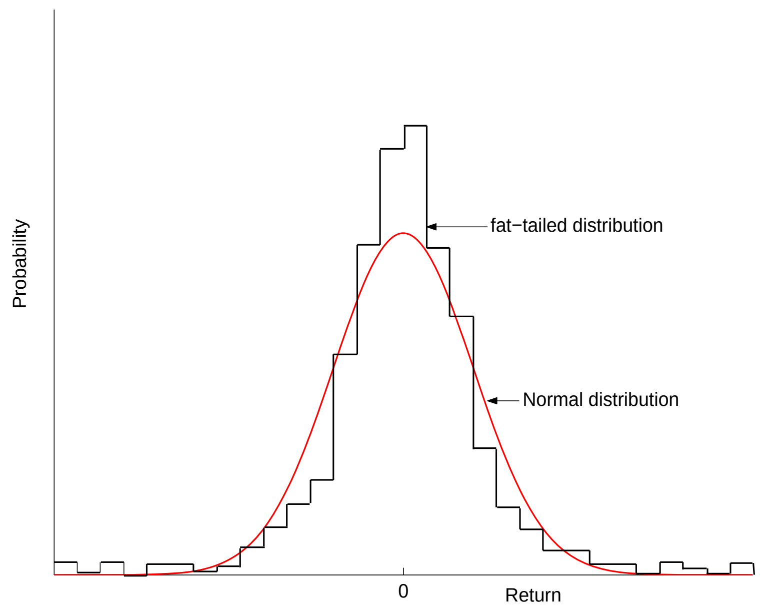

Figure 1: Asset returns are leptokurtic.

value (rate or price) does not exhibit growth, but tends to revert to a long term mean. Again, if the dynamics are modelled by a diffusion equation with a mean-reverting (MR) deterministic term, [Vasicek 1977], the probability of extreme events occurring are underestimated as compared with empirical observations [Cont 2000].

1.1 Direct modelling of asset values

Let us first consider an individual asset. The fat tails exhibited by the asset value dynamics can be modelled in a number of ways.

Adding jumps If we consider asset prices, then jumps can be introduced into the proposed equation so, for example, GBM+Jumps can be written as

$$\frac{dx_t}{x_t} = \mu dt + \sigma dW_t + \nu dq_t$$

where x_t is the asset price, μ and σ are the (constant) drift and volatility parameters, W_t is a random walk, ν is the size of a possible jump and q_t is a Poisson process, independent of W_t, so that in a time interval $[t, t+dt]$: $Prob(dq_t = 1) = \lambda \dot{d}t$, where λ is the (constant) jump intensity. The jump size ν can be either assumed to be a constant, or a Normally distributed variable, or chosen from a (small) set of equilikely jump sizes. Parameter estimation is (in general) quite simple.

Maximum Likelihood Estimation (MLE) is possible in the simple cases, because the pdf of GBM+Jumps can then be written analytically. (It is closely related to "the mixture of two Normals".) Maximizing the loglikelihood is straightforward with a sequential quadratic programming algorithm. We should note here, however, that MLE leads to an apparent overestimate of the jump intensity λ, with many small jumps being used to explain movements that can also be easily explained (almost as well) by the diffusion term.

The authors' experience is that recursive digital filter estimation of the parameters in GBM+Jumps is a superior method, declaring only large movements (that are unlikely to be produced by diffusion) as jumps and estimating the parameters accordingly.

Stochastic volatility It is well-known that fat tails can also be explained by assuming the volatility to be stochastic. Perhaps the simplest such model is GBM+GARCH, namely a model where x_t follows a GBM-type equation, except for the fact that the volatility in this equation is not constant but itself follows a GARCH(1,1) process. Thus (in discrete time) we have

$$\begin{aligned} \frac{\delta x_t}{x_t} &= \mu \delta t + \sqrt{v_t}\sqrt{\delta t}\epsilon_t \\ v_t &= \omega + \alpha \epsilon_t^2 + \beta v_{t-\delta t} \end{aligned}$$

where μ, ω, α and β are constant parameters. Again the pdf of x_t is known analytically and MLE can be used to calibrate the GBM+GARCH model. Other models that can be fitted and for which good calibration procedures exist are GBM+Jumps+GARCH and MR+Jumps+GARCH.

2 Modelling a basket of assets

Now consider the case of a basket of assets indexed $n = 1, \dots, \hat{n}$. The time series of the values (prices or returns) of each asset can be modelled as above. If we employ a

GBM+Jumps model (for example) we can use $dW_t(n)$ and $dq_t(n)$ for the stochastic processes corresponding to asset n. Now, however, we must also consider the correlations

Corr$(dW_t(n'), dW_t(n''))$ and Corr$(dq_t(n'), dq_t(n''))$; $n' \neq n'' = 1, \ldots, \hat{n}$.

Calibration now becomes complex, inaccurate and the number of parameters to be estimated becomes very large.

2.1 Principal components

A well-known and useful procedure for representing a basket of assets is by using principal components (PCs). Let $A_{\hat{\tau}\times\hat{n}}$ be a matrix whose column $n = 1, \ldots, \hat{n}$ is the time series of the values of asset n; i.e. $a_{\tau n}$ is the value of n at time $\tau = 1, \ldots, \hat{\tau}$. The principal components $B_{\hat{\tau}\times\hat{n}}$ can be written in the form

$$B_{\hat{\tau}\times\hat{n}} = A_{\hat{\tau}\times\hat{n}} \times V_{\hat{n}\times\hat{n}} \times D_{\hat{n}\times\hat{n}}$$

where $D_{\hat{n}\times\hat{n}}$ is a diagonal matrix of eigenvalues *in descending order* and $V_{\hat{n}\times\hat{n}}$ is a matrix of eigenvectors. The columns (PCs) of matrix $B_{\hat{\tau}\times\hat{n}}$ can be considered as time series themselves and have the property that they are orthogonal (uncorrelated). Given B, A can be reproduced exactly by a linear transformation $A_{\hat{\tau}\times\hat{n}} = B_{\hat{\tau}\times\hat{n}}\Phi_{\hat{n}\times\hat{n}}$ where Φ is a matrix of 'weights', i.e. ϕ_{pn} is the 'weight' of PC p towards asset n.

It is also well-known that A can be *approximated* by keeping and using only the first $\hat{p}$ (say) PCs and ignoring (setting to 0) PCs $\hat{p}+1, \ldots, \hat{n}$. The approximation is then given by

$$A_{\hat{\tau}\times\hat{n}} \approx B_{\hat{\tau}\times\hat{p}}\Phi_{\hat{p}\times\hat{n}} \qquad (1)$$

Note that

- If the asset returns are normal, then the PCs are normal.
- If the asset returns are normal, then the PCs being uncorrelated implies that that they are also independent.

Unfortunately, fat-tailed returns render PCs both *not* normally distributed *and* dependent. As an example, consider the (alphabetically) first 16 assets of the FTSE100 index and select the time series to cover the period 1 April 1996 to 31 March 2002, giving a total of over 900 observations for each time series. Consider the first 5 PCs in the PC-decomposition of these 16 time series. Although the second-order cross moments (covariances) among the PCs are zero, the fourth-order cumulants (cokurtosis) of the PCs are by no means zero. Figure 2 shows the projection of the 4-dimensional tensor of these (normalized) cumulants among the PCs onto a 2-D plane.

The peaks along the diagonal are the kurtosis terms of each PC and the off-diagonal peaks are the "co-kurtosis" terms which are clearly non-zero. These represent the co-movements of the fat-tails of the distributions of the PC time series.

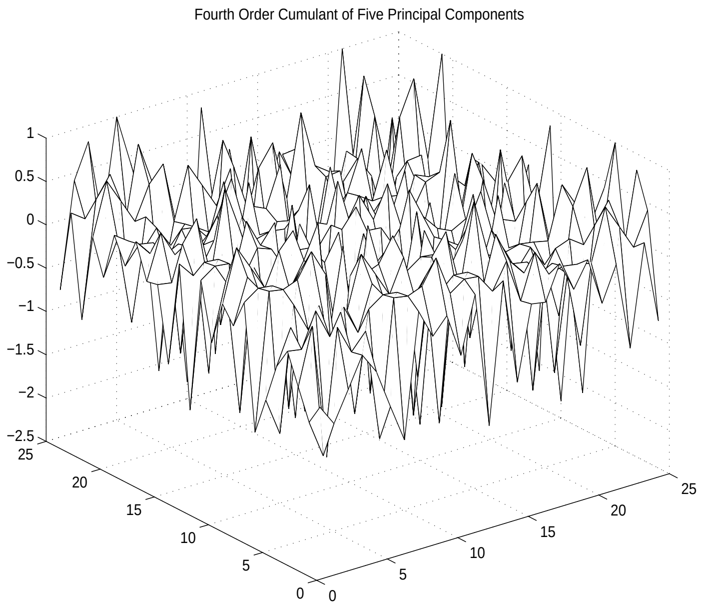

Figure 2: Fourth-order cumulants of 5 principal components projected onto a 2-D plane.

2.2 Independent components

Consider the principal component matrix $B_{\hat{\tau}\times\hat{p}}$ whose columns represent the time series of the first $\hat{p}$ principal components of the original time series matrix $A_{\hat{\tau}\times\hat{n}}$. The independent components (ICs) $C_{\hat{\tau}\times\hat{p}}$ of $B_{\hat{\tau}\times\hat{p}}$ can be written in the form

$$C_{\hat{\tau}\times\hat{p}} = \tilde{B}_{\hat{\tau}\times\hat{p}} \times M^{-1}_{\hat{p}\times\hat{p}}$$

where $\tilde{B}$ is B normalized so that the mean and variance of each column is 0 and 1 respectively and where M^{-1} is a transformation matrix. The columns of matrix $C_{\hat{\tau}\times\hat{p}}$ can again be considered as time series.
From equation 1 we can write

$$A_{\hat{\tau}\times\hat{n}} \approx C_{\hat{\tau}\times\hat{p}}\Psi_{\hat{p}\times\hat{n}} \tag{2}$$

where Ψ is an easy to compute matrix of ‘weights’. By a suitable choice of the transformation matrix M, ICs maintain the same second-order cross moments between themselves as found amongst the PCs, i.e. they are orthogonal to each other (with zero covariances). In addition, ICs can be made independent of each other, which in practice, means making the fourth-order cross moments (co-kurtosis) between them to be zero. Considering the earlier example, figure 3 shows the corresponding projection of the tensor of fourth-order (normalized) cumulants among the ICs onto a 2-D plane.

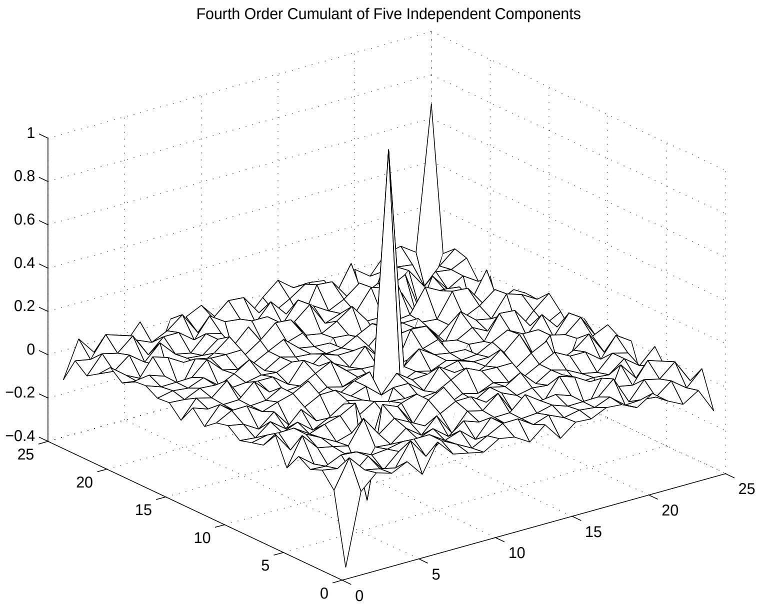

Figure 3: Fourth-order cumulants of 5 independent components projected onto a 2-D plane.

The peaks along the diagonal (i.e. the kurtosis terms of each IC - prior to normalization) are much larger than for the PC case, but the off-diagonal terms (the co-kurtosis) are now virtually zero.

2.3 Advantages of ICs over PCs

It is clear that we can model a basket of original assets by decomposing their value (return or price) time series into PCs or ICs and modelling these. The PCs/ICs can be modelled by stochastic processes, for example GBM+Jumps+GARCH and if these models are implemented as discrete state space transition graphs then the values of the original assets can be reconstructed at any vertex of this graph from the values of the PCs/ICs at that vertex. If the PCs/ICs are independent, then this approach has a number of advantages over modelling the asset values directly.

For example, implementing multidimensional correlated asset value dynamics by an arbitrage-free state space transition graph, is not a simple task. On the other hand, modelling the (independent) PCs/ICs would only require a separate one-dimensional graph to be constructed for each PC/IC, a very much easier task. Modelling the combined state space transition graph in this case, can then be done either by forming the *product-graph* of these separate graphs, or by other algorithms. It is worthwhile to note here that if the PC/IC values satisfy the no-arbitrage conditions on the combined graph then the reconstructed asset values on this graph also satisfy these conditions [Christofides, Christofides and Christofides 2000].

The above advantage of modelling PCs/ICs instead of modelling the assets directly, disappears for the case of PCs (but not for ICs) when the assets have pronounced fat-tailed distributions. This is due to the fact that the PCs can no longer be treated as independent, but the ICs can.

3 Reduction of dimensionality

Referring to the earlier example, figure 4 shows the time series of the 5 ICs and figure 5 shows the approximate reconstruction of price time series of the first two assets (Abbey-National and Amersham). Similar results apply for the other assets. [1]

Obviously, the more ICs one uses for the reconstruction, the better the approximation. With $\hat{n}$ ICs exact asset value replication results. Figure 6 shows the sum of errors for all 16 assets in the example (measured in terms of RMS value) remaining after using $1, \ldots, 16$ ICs.

3.1 A neural network (NN) error approximator

It is clear from figure 6 that as more and more ICs are used, their individual contribution towards error reduction decreases. A large part of the signal is accounted for by a linear combination of a few ICs. Although the remaining $\hat{n}$-dimensional error signal in the approximation by $\hat{p}$ ICs, is linear in the remaining $\hat{n}-\hat{p}$ ICs, $\hat{n}-\hat{p}$ may be very large. It may, therefore, be possible to consider non-linear combinations of the already chosen $\hat{p}$ ICs to explain part of the remaining error signal. The error time series remaining after the linear approximation by $\hat{p}$ ICs, are given by

$$E_{\hat{\tau}\times\hat{n}} = A_{\hat{\tau}\times\hat{n}} - C_{\hat{\tau}\times\hat{p}}\Psi_{\hat{p}\times\hat{n}}$$

We now construct a NN to be a nonlinear approximator to these error time series as follows.

There are $\hat{p}$ input vertices, the p^{th} one corresponding to the p^{th} independent component. The input into this vertex at time τ is $c_{\tau p}$ (see figure 7). There is one hidden layer of (say) $\hat{h}$ neurons, each of which has a radial basis function as its transfer function. There are $\hat{n}$ output vertices, the n^{th} one corresponding to the n^{th} asset. The output

[1] Note that the original asset prices are normalized (the price on 1 April 1996 is set to 100) before PC/IC analysis is performed.

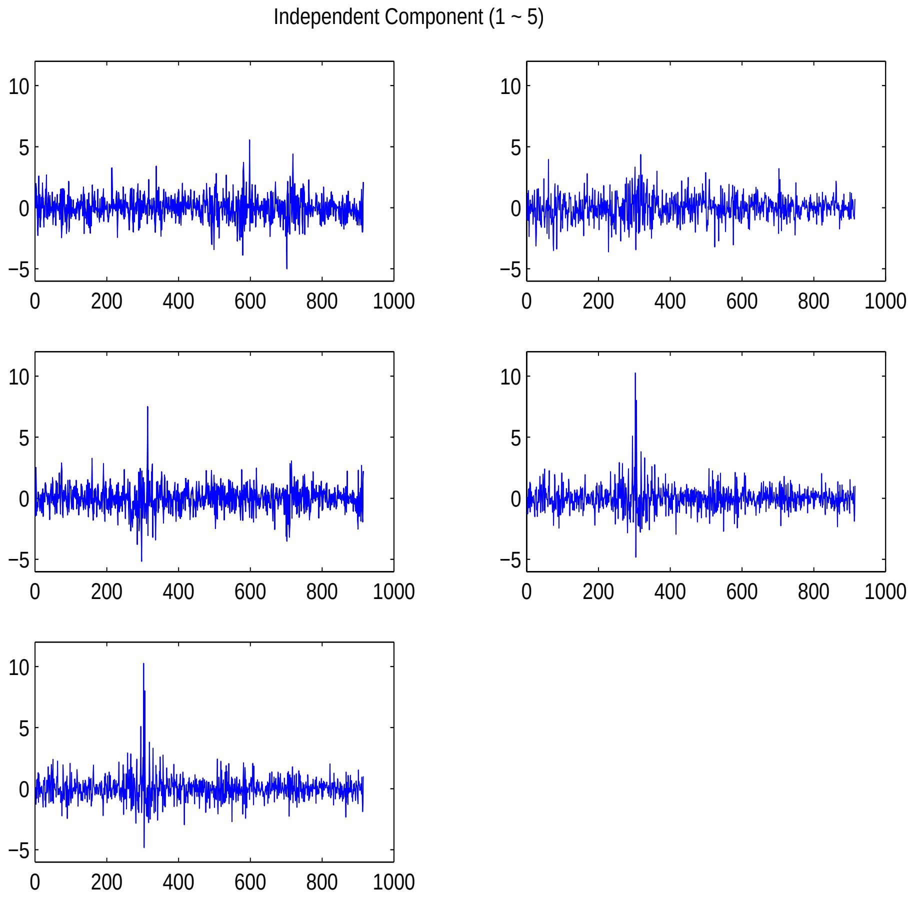

Figure 4: Time series of 5 independent components for the example.

from this n^{th} output vertex at time τ is the estimate $\tilde{e}_{\tau n}$ of the error $e_{\tau n}$. The predicted value of $a_{\tau n}$ is then given by $\sum_{p=1}^{\hat{p}} c_{\tau p}\psi_{pn} + \tilde{e}_{\tau n}$.

4 Forecasting results

We evaluated the forecasts resulting from the hybrid model applied to the FTSE-100, DJIA, Nasdaq, and DAX, baskets and two artificially constructed bond baskets of 20 bonds each. (See table 1). Basket 'Bonds1' is composed of US Treasury bonds and 'Bonds2' of corporate bonds of various ratings. Forecasts were made of prices/yields (for stocks/bonds) for *every* stock/bond in the basket and the average RMS forecast error of all the items in the basket is reported. Forecasts are also made for the volatility of these prices/yields. The hybrid model had 7 independent components and used GBM+Jumps+GARCH for the modelling of the IC time series. A NN with 5 neurons

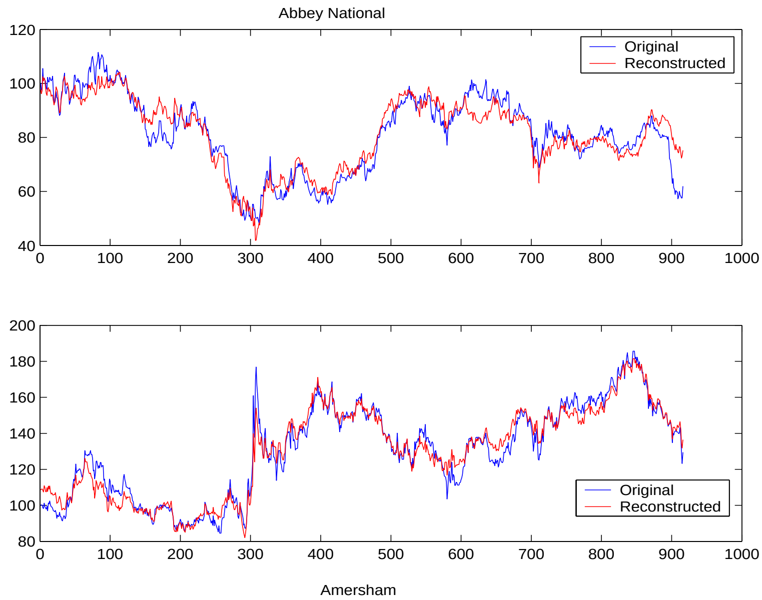

Figure 5: Approximate asset prices from 5 independent components.

in a single hidden layer was used in the hybrid model. The hybrid model was evaluated against "traditional" methods, in this case a GARCH(1,1) model applied directly to the asset prices/yields and calibrated by a maximum log-likelihood method. Starting 1 April 1998 and using daily data for the previous 3 years, a forecast was made 6 months ahead. These 6-month forecasts are made on a rolling-horizon basis, once a month until 1 April 2002. The results reported in table 1 are based on the averages of these 48 forecasts.

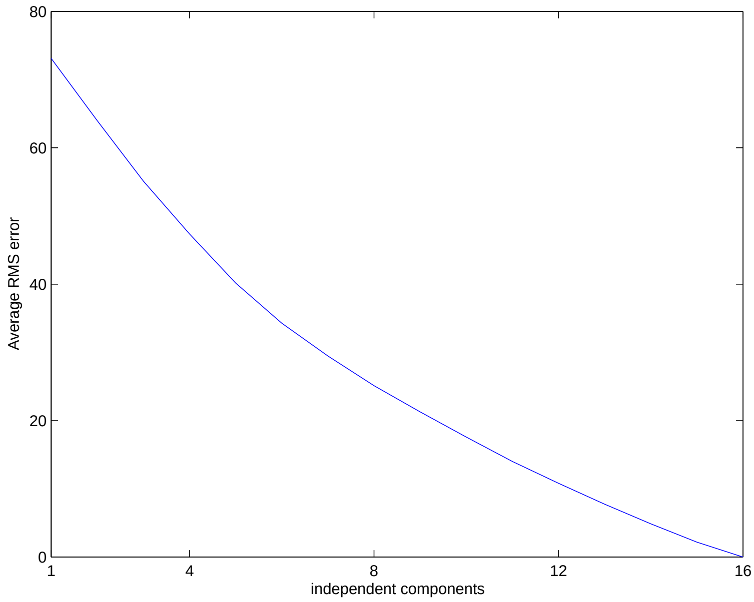

Figure 6: Average RMS error (percent) resulting from using $1, \ldots, 16$ independent components.

References

[Akaike 1973] Akaike H (1973). *Information theory and an extension of the maximum likelihood principle*. In Petrov B, Csaki F (Eds), Second international symposium on information theory. Academia Kiado, Budapest, Hungary, pp 267-281.

[Black and Scholes 1973] Black F. and Scholes M. (1973). *The pricing of options and corporate liabilities*. J. of Political Economy, vol. 81, pp 637-659.

[Bollerslev 1986] Bollerslev T. (1986). *Generalised autoregressive conditional heteroskedasticity*. J. of Econometrics, vol. 31, pp 307-327.

[Campbell, Lo and MacKinlay 1997] Campbell JY, Lo AW and MacKinlay AC (1997). *The econometrics of financial markets*. Princeton University Press, Princeton, USA.

[Christofides, Christofides and Christofides 2000] Christofides S., Christofides A. and Christofides N. (2000). *Graph representations for financial dynamics*. Report, Centre for Quantitative Finance, Imperial College, UK.

[Cont 2000] Cont R. (2000). *Empirical properties of asset returns*. Quantitative Finance, vol. 1, pp 223-236.

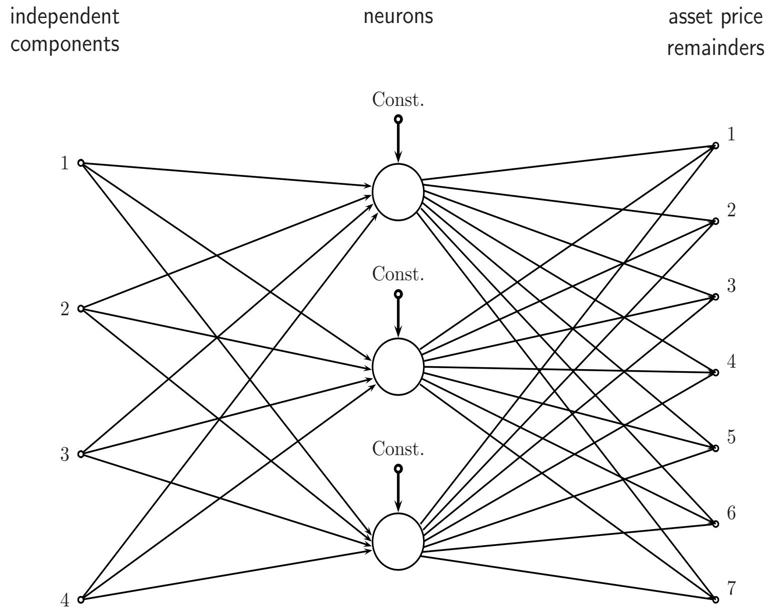

Figure 7: Neural network with independent component inputs.

[Embrechts, Kluppelberg and Mikosch 1997] Embrechts P., Kluppelberg C. and Mikosch T. (1997). *Modelling extremal events*. Springer, Heidelberg, Germany.

[Erb et al. 1994] Erb B., Campbell H.R. and Tadas V.E. (1994). *Forecasting international equity correlations*. Financial Analysts Journal, Nov-Dec, pp 32-44.

[Kon 1984] Kon S.J. (1984). *Models of stock returns: A comparison*. J. of Finance, vol. XXXIX, pp 147-165.

[LeBaron 2001] LeBaron B. (2001). *A builder's guide to agent-based financial markets*. Quantitative Finance, vol. 1, pp254-262.

[Nelson 1990] Nelson D B (1990). *ARCH models as diffusion approximations*. J. of Econometrics, vol. 45, pp 7-38.

[Vasicek 1977] Vasicek O.A. (1977). *An equilibrium characterization of the term structure* Journal of Financial Economics, vol. 5, pp 177-188.

Time-series	GARCH, % error		Ind. Comp., % error		Ind. Comp. + NN, % error	
	Value	Vol.	Value	Vol.	Value	Vol.
FTSE100	3.41	18.3	3.11	16.9	3.02	15.5
DJIA	3.10	14.7	2.85	13.0	2.71	12.3
Nasdaq	4.09	29.4	3.52	24.1	3.32	23.0
DAX	3.79	21.0	3.38	20.0	3.31	18.2
Bonds1	8.25	32.1	7.60	28.5	7.16	26.9
Bonds2	17.42	51.8	14.05	44.3	12.17	38.5

Table 1: Comparison of the performance of different models.
Errors shown are unweighted RMS averages of the errors in each of the assets forming the index.
Number of ICs used, 7. Number of neurons in hidden layer of NN, 5.
Predictions of values (prices for stocks, yields for bonds) after 6 months. Predictions of volatility is the average over the 6^{th} month.