Modelling Stock Market Excess Returns by Markov Modulated Gaussian Noise

Jonathan H. Manton [1]

Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Victoria 3052, Australia. jon@mullian.ee.mu.oz.au

Anton Muscatelli

Department of Political Economy, Adam Smith Building, University of Glasgow, Glasgow G12 8RT. gkaa22@udcf.gla.ac.uk

Vikram Krishnamurthy [2]

Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Victoria 3052, Australia. vikram@mullian.ee.mu.oz.au

Stan Hurn

Institute of Economics and Statistics, St. Cross Building, Oxford OX1 3UL. stan.hurn@economics.ox.ac.uk

Abstract

A basic analysis of stock market excess return data shows both linear and non-linear dependence present. Previous papers have used this to argue that it must therefore be possible to predict future values. However, this paper shows that the linear and non-linear dependence can be explained by simply allowing the mean and variance of Gaussian noise to be modulated by a (typically 3 state) hidden Markov model. Attempting to fit a Markov modulated AR process proved fruitless; the conclusion is that there is no AR-predictability present in excess return data.

JEL codes: G12, G14, C22

---

[1]Jonathan Manton acknowledges the support of the Australian Telecommunication and Engineering Research Board and the Cooperative Research Centre for Sensor Signal and Information Processing.

[2]Vikram Krishnamurthy acknowledges the support of an ARC large grant and the Cooperative Research Centre for Sensor Signal and Information Processing.

1 Introduction

Recent work on modelling stock market returns has focussed on the crucial questions of whether or not the returns are independently distributed over time and whether or not they are normally distributed. There is evidence to refute the claim that returns are linearly dependent [11, 27, 20, 23, 6, 10], and a growing literature documenting dependence in the squares of the returns [8, 18, 21]. It is also now accepted that many financial time series exhibit skewness and excess kurtosis and are therefore not normally distributed [13, 12]. The burgeoning literature on ARCH modelling (for good surveys, see [1, 3, 26]) and stochastic volatility [16] can be viewed as an attempt to reproduce these characteristics.

A perceived defect of many ARCH models is that persistence in the volatility of time series, which these models induce, can persist for too long. For example, [9] has argued that the volatility consequences in stock option prices of the 1987 crash disappeared more rapidly than suggested by simple ARCH models. An alternative way of characterising changes in volatility or in the autoregressive component of excess returns is through the use of Markov modulated models [14]. These models typically have less persistant volatility whilst also allowing fundamental characteristics like negative skewness and excess kurtosis to be modelled [15, 4].

In this paper, we apply this framework to the monthly excess returns on equities for the period 1960–1988 which was used in [6] for the USA, Japan, Germany, France, United Kingdom, Italy and Canada[3]. As with most data on stock returns, our excess returns series are found to exhibit negative skewness and kurtosis normally associated with asymmetric financial time series. We also find that linear dependence is much weaker than the observed non-linear dependence in the data[4]. Finally, we show that a simple Markov modulated Gaussian noise process is adequate to model our relatively low-frequency monthly data.

---

[3]In [6], the data was obtained from Morgan Stanley's Capital International Perspectives. We are grateful to Jim Poterba and Larry Summers for supplying us with the data used in their study.

[4]For reasons of space, we do not report summary statistics and test of linear and non-linear dependence on the raw data here. The stock return data is analysed fully in [6].

2 Markov-Modulated Gaussian Noise

2.1 The Model

Let y_t be our observed data on excess returns. Monthly excess returns in stock markets, y_t, are calculated as follows:

$$y_t = \frac{P_t + D_t}{P_{t-1}(1 + i_t)}$$

where P_t is the end-month equity price index, P_{t-1} is the lagged price, D_t is dividend payments, and i_t is the monthly short-term interest rate against which excess returns are measured (in this case short-term money market or treasury bill yields). Let s_t be the state (or regime) we are in at time t. s_t is modelled by a K-state Markov chain, which means the probability of switching from state i to state j at any time t (i.e., $p(s_t = j|s_{t-1} = i) = p_{ij}$) is a constant. Then we model each y_t as being independent Gaussian noise with mean $\mu(s_t)$ and variance $\sigma^2(s_t)$.

Our parameter vector is thus $\phi = [p_i, p_{ij}, \mu(s_k), \sigma^2(s_k)]$ where p_i is the probability of being in state i at time $t = 1$.

2.2 Maximum Likelihood Estimate of the Parameter Vector ϕ

Given the data set $\{y_1, \ldots, y_N\}$, computing the best (in the sense of maximum likelihood) value of ϕ is a nontrivial operation. The likelihood function is highly non-linear, making a direct attempt to find its maximum infeasible. Instead, the expectation maximization (EM) algorithm [7] can be used to converge iteratively to a local maximum of the likelihood surface. The fact that it is only a local maximum poses a degree of difficulty. In general, starting the EM algorithm with different initial parameter values can lead to convergence to different local maxima. For each local maximum, we explicitly calculate the likelihood. The best guess of the parameter vector ϕ will thus be the vector corresponding to the local maximum with the highest likelihood.

2.3 EM Algorithm

Let $S = \{s_1, \ldots, s_N\}$ be the (unknown) state sequence, and $Y = \{y_1, \ldots, y_N\}$ the observed output sequence. Let $f(S, Y; \phi)$ be the probability (density function) of obtaining a state sequence S and output Y given the parameter ϕ. By defining the function

$$Q(\phi^{(p)}, \phi) = \mathbf{E}\Big[\log f(S, Y; \phi) \mid Y; \phi^{(p)}\Big]$$

then it can be shown [7] that

$$\phi^{(p+1)} = \arg\max_{\phi} Q(\phi^{(p)}, \phi)$$

will result in a monotonically converging sequence

$$f(S, Y; \phi^{(p+1)}) \geq f(S, Y; \phi^{(p)})$$

of the likelihood function.

For the model at hand, the Q function can be derived as follows:

$$\begin{aligned}
\log f(S, Y; \phi) &= \log f(S; \phi) + \log f(Y \mid S; \phi) \\
\log f(S; \phi) &= \log p_{s_1} + \sum_{t=1}^{N-1} \log p_{s_t, s_{t+1}} \\
\log f(Y \mid S; \phi) &= \sum_{t=1}^{N} \log f(y_t \mid S; \phi) \\
&= \sum_{t=1}^{N} \left[\log \left(\frac{1}{\sqrt{2\pi}\sigma(s_t)} \right) - \frac{1}{2} \left(\frac{y_t - \mu(s_t)}{\sigma(s_t)} \right)^2 \right] \\
&= -\frac{N}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^{N} \log \sigma^2(s_t) - \frac{1}{2} \sum_{t=1}^{N} \left(\frac{y_t - \mu(s_t)}{\sigma(s_t)} \right)^2
\end{aligned}$$

For convenience, the functions

$$\begin{aligned}
\gamma_t^{(p)}(k) &= Pr\{s_t = k \mid Y; \phi^{(p)}\} \\
\xi_t^{(p)}(i, j) &= Pr\{s_t = i,\ s_{t+1} = j \mid Y; \phi^{(p)}\}
\end{aligned}$$

are introduced. They are standard definitions in hidden Markov models [28].

Let $g(s_t)$ and $h(s_t, s_{t+1})$ be any two functions which depend on the state of the Markov chain. Then it is clear that the following equations hold.

$$\begin{aligned}
\mathbf{E}\Big[g(s_t) \mid Y; \phi^{(p)} \Big] &= \sum_{i=1}^{K} g(i) \gamma_t^{(p)}(i) \\
\mathbf{E}\Big[h(s_t, s_{t+1}) \mid Y; \phi^{(p)} \Big] &= \sum_{i=1}^{K} \sum_{j=1}^{K} h(i, j) \xi_t^{(p)}(i, j)
\end{aligned}$$

Hence, the Q function is given by:

$$\begin{aligned}
Q(\phi^{(p)}, \phi) &= \mathbf{E}\Big[\log f(S, Y; \phi) \mid Y; \phi^{(p)} \Big] \\
&= \sum_{i=1}^{K} \gamma_1^{(p)}(i)\ \log p_i + \sum_{t=1}^{N-1} \sum_{i=1}^{K} \sum_{j=1}^{K} \xi_t^{(p)}(i, j)\ \log p_{ij} - \frac{N}{2} \log 2\pi \\
&\quad - \frac{1}{2} \sum_{t=1}^{N} \sum_{i=1}^{K} \gamma_t^{(p)}(i)\ \log \sigma^2(i) - \frac{1}{2} \sum_{t=1}^{N} \sum_{i=1}^{K} \gamma_t^{(p)}(i) \left(\frac{y_t - \mu(s_t)}{\sigma(s_t)} \right)^2
\end{aligned}$$

It can then be shown, by setting the partial derivatives of $Q(\phi^{(p)}, \phi)$ to zero to find the maximum, and using Lagrange multipliers to enforce the elements of p_i and the columns of p_{ij} to sum to one, that the update equations are:

$$
\begin{aligned}
p_i^{(p+1)} &= \gamma_1^{(p)}(i) \\
p_{ij}^{(p+1)} &= \frac{\sum_{t=1}^{N-1} \xi_t^{(p)}(i,j)}{\sum_{t=1}^{N-1} \xi_t^{(p)}(i)} \\
\mu^{(p+1)}(i) &= \frac{\sum_{t=1}^{N} \gamma_t^{(p)}(i)\, y_t}{\sum_{t=1}^{N} \gamma_t^{(p)}(i)} \\
\sigma^{(p+1)}(i)^2 &= \frac{\sum_{t=1}^{N} \gamma_t^{(p)}(i)\, \left(y_t - \mu^{(p+1)}(i)\right)^2}{\sum_{t=1}^{N} \gamma_t^{(p)}(i)}
\end{aligned}
$$

Note that $\gamma_t^{(p)}(i)$ and $\xi_t^{(p)}(i,j)$ are calculated using the old parameter vector $\phi^{(p)}$. The actual method of calculation is once again standard, and is given in [28].

2.4 Filtering the data

Once the maximum-likelihood estimate of the model parameters has been obtained, the well-known Viterbi algorithm [22] can be used to find the optimal state estimate in the Markov chain. In our case, the state estimate tells us whether the excess returns are in a high, medium or low state for any particular month.

In order to determine whether the model accurately fits the data or not, the model is used to calculate the prediction error. We subtract from each data point an estimate of that data point based on the previous data points and the model parameters. This gives us the model residuals. We divide by their expected variance to obtain the scaled residuals.

Standard tests for linear and non-linear dependence, and for normality, are then applied to these residuals.

3 Estimation Results

3.1 Model Estimates and Adequacy

Before applying our Markov modulated Gaussian noise model, we tried a Markov modulated AR(1) model, where we modelled the observed excess return data by $y_t = a(s_t)y_{t-1} + \sigma(s_t)w_t$ where $w_t \sim$ iid $N(0,1)$, $a(s)$ is the AR(1) coefficient for state s, and $\sigma^2(s)$ is the variance for state s.

However, the estimates of $a(s_t)$ were not significant, suggesting that an AR process is not present. This claim is substantiated by the fact that our Markov modulated Gaussian noise model produced residuals which contained no significant linear or non-linear dependencies and also passed the Jarque and Bera test for normality. In Table 1, we report the Jarque-Bera [17] test result along with the Ljung-Box [19] test for linear dependence. Table 2 lists the McLeod-Li [24] test for non-linear dependence.

The number of volatility states was chosen after some experimentation. A three state model (except for Japan, which turned out essentially to have only two states) was found to correctly characterise the excess returns data in that the residuals were generally found to be white noise. The only evidence of remaining serial dependence is in the case of Italy, where at lags 5 and beyond the Ljung-Box test appears significant. The normality tests are also passed in all cases except for the UK, where the Jarque-Bera statistic is significant at the 5%, but not at the 1% level. Generally these results are satisfactory, especially when compared to standard GARCH-type models. Using a model with more than three states would result in a single state splitting into two sub-states, of almost identical mean and variance. This also resulted in a noticeable sensitivity to initial parameter values, as might be expected.

In summary, the Markov modulated Gaussian noise model managed to explain the original linear and non-linear dependencies in the data, and showed that once the change in variance for the different states has been accounted for, the data is essentially white noise.

3.2 Discussion of Results

Our main results are shown in Figures 1, 2, 3 and 4. The figures show graphs of the mean and standard deviation predicted by our models. The first interesting feature to note concerns the so-called leverage effect originally noted in [2] in which volatility responds to past forecast errors in stock market returns[5]. As in recent attempts to model stock market volatility using ARCH-type models [25, 29, 5], the leverage effect is clearly visible in our results: excess return decreases lead to major increases in volatility, whilst increases in stock prices seem to have little impact.

For example, sharp falls in excess returns (see Figures 1 and 3) occur at the time of the 1973–75 recession in the US, France, the UK and Canada, at the time of the Mitterand experiment in France (1981), in Germany in 1960–1 when there was a tightening of monetary policy and a realignment within Bretton Woods, and in correspondence to the 1987 stock market crash in all countries, with the exception of Japan. In the case of 1987, the most marked effect is in the US and the UK model, whilst in continental Europe, as might be expected, the effect is much less

[5]The leverage effect has an intuitive explanation in terms of the riskiness of a firm's equity. If a firm's equity value unexpectedly rises (excess returns are positive) this will lower its leverage ratio and hence its riskiness: hence volatility tends to fall. The opposite happens in the case of an unexpected fall in equity values, which raises volatility.

marked. As expected, the events outlined above correspond to high volatility states (see Figures 2 and 4).

There are, however, more subtle features to the relationship between excess returns and volatility, which are apparent in our results and which *standard* ARCH-type models would not necessarily identify. Note for instance that for some prolonged periods, there is no inverse relationship between excess returns and volatility. This is particularly apparent for the late 1960s and 1970s in Germany and Japan, and for the 1980s in Italy. This suggests that the leverage effect is only visible following sudden sharp movements in equity values, i.e. that there is a "threshold effect". This suggests that any successful ARCH-type model would need to be "designed" very carefully to capture such asymmetries and threshold effects.

The problem of traditional ARCH models (particularly the popular GARCH(1,1) model) in allowing volatility to persist for too long has already been noted. The results reported here demonstrate that sudden changes in volatility, such as the increase in volatility induced by major stock market declines (1987 and 1974–75), do not persist beyond a couple of months in the US, with persistence in Germany, France and the UK only slightly longer. It is difficult to capture these transitions between different volatility states using standard parametric techniques.

In general, we would argue that the models tend to track actual movements in stock prices quite well over this period[6] confirming that these modelling techniques provide a useful adjunct to traditional ARCH models. The advantage over ARCH models is that state transitions allow greater flexibility in describing important, but short-lived, episodes, and interesting asymmetrics and threshold effects. To capture similar features using variants of ARCH models would require the use of complex and highly nonlinear ARCH schemes which may not be particularly stable. On the other hand, the modelling approach followed here has some important disadvantages vis-a-vis ARCH models. One key problem is that, by limiting the number of states considered, some features of the data may not be captured. For instance, our model fails to capture the 1969–70 recession or the very end of the 1981–82 recession (with the possible exception of the model for Canada). Whilst our approach is useful to capture major features in the data, it is less useful in analysing minor changes in excess returns or volatility.

Finally, it should be noted that we were unable to detect linear dependence in excess returns. In this regard, it is worth recognising that estimated transition probabilities describe each state as highly persistent, which explains why AR effects are difficult to fit. This tends to support those studies (e.g. [10]) which attribute apparent patterns in stock returns to changes in volatility [10] and confirms that any autoregressive elements present are extremely weak [6].

[6]For example, the progressive tightening of monetary policy in Canada in the 1980s is apparent in a number of downward adjustments in the mean value.

4 Conclusion

In this paper we have examined an alternative method of modelling stock returns, using Markov modulated Gaussian noise. Our results show that a three-state model can adequately characterise excess returns for the major G7 economies. Despite some evidence of linear dependence, allowing for changes in variance for different states fully accounts for this phenomenon.

A potential extension of our model would be to apply it to higher-frequency data, using more than 3 states in the analysis. It may well turn out that the resulting states may be grouped into three blocks, corresponding to the three main states found in this paper. The interpretation would be that a stock market can be in three main states in relation to the mean excess return, and that each of these states can be broken down into two sub-states, depending on whether volatility is high or low. This is something which should be explored in future work. However, as noted previously, it seems likely that such models would be more applicable to a higher-frequency data set, given that the simpler three state models employed here seem to work adequately on monthly data. Conceivably, one might also examine whether transitions between states in any one stock market depend on transitions in other stock markets, as one might expect.

References

[1] A. K. Bera and M. L. Higgins. ARCH models: Properties, estimation and testing. *Journal of Economic Surveys*, 7(4):305–362, 1993.

[2] F. Black. Studies of stock market volatility changes. In *Proceedings of the 1976 Business Meeting of the Business and Economic Statistics Section, American Economic Association*, pages 177–181. American Statistical Association, 1976.

[3] T. Bollerslev, R. Y. Chou, and K. F. Kroner. ARCH modelling in finance: A review of the theory and empirical evidence. *J. Econometrics*, 52:5–59, 1990.

[4] J. Cai. A Markov model of switching-regime ARCH. *Journal of Business and Economic Statistics*, 12:309–316, 1994.

[5] J. Y. Campbell and L. Hentschell. No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, 31:281–318, 1992.

[6] D. M. Cutler, J. M. Poterba, and L. H. Summers. Speculative dynamics. *Review of Economic Studies*, 58:529–546, 1991.

[7] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. B*, 39:1–38, 1977.

[8] Z. Ding, C.W.J. Granger, and R.F. Engle. A long-memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1:83–105, 1993.

[9] R.F. Engle and C. Mustafa. Implied ARCH models from options prices. *J. Econometrics*, 52:289–311, April 1992.

[10] E. F. Fama and K. R. French. Dividend yields and expected stock returns. *Journal of Financial Economics*, 22:3–25, 1988.

[11] E. F. Fama and K. R. French. Permanent and temporary components of stock prices. *Journal of Political Economy*, 96:246–273, 1988.

[12] A. R. Gallant, J. Rossi, and G. Tauchen. Stock prices and volumes. *Review of Financial Studies*, 5:199–242, 1992.

[13] A.R. Gallant, D.A. Hsieh, and G. Tauchen. On fitting a recalcitrant series: The pound/dollar exchange rate 1974–83. In J. Powell W. A. Barnett and G. Tauchen, editors, *Nonparametric and Semiparametric Methods in Economics and Statistics*. Cambridge University Press, 1991.

[14] J.D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–84, March 1989.

[15] J.D. Hamilton and R. Susmel. Autoregressive conditional heteroskedasticity and changes in regime. *J. Econometrics*, 64:307–33, 1994.

[16] A. Harvey, E. Ruiz, and N. Shephard. Multivariate stochastic variance models. *Review of Economic Studies*, 61:247–264, 1994.

[17] C. M. Jarque and A. K. Bera. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6:255–259, 1980.

[18] P. De Lima and N. Crato. Long range dependence in the conditional variance of stock returns. *Economic Letters*, 45:281–285, 1994.

[19] G. M. Ljung and G. E. P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65:279–303, 1978.

[20] A. W. Lo and A. C. McKinlay. Stock markets do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, 1:41–66, 1988.

[21] A.W. Lo. Long-term memory in stock market prices. *Econometrica*, 59:1279–1313, 1991.

[22] Hui-Ling Lou. Implementing the viterbi algorithm. *IEEE Signal Processing Magazine*, 12(5):42–52, September 1995.

[23] R. MacDonald and D. Power. Persistence in UK stock market returns: Aggregated and disaggregated perspectives. In M. P. Taylor, editor, *Money and Financial Markets*. Oxford: Blackwells, 1991.

[24] A. T. McLeod and W. K. Li. Diagnostic checking ARIMA time series models using squared-residual autocorrelations. *The Journal of Time Series Analysis*, 4:269–273, 1983.

[25] D. B. Nelson. Conditional heteroscedasticity in asset returns: A new approach. *Econometrica*, 59:281–292, 1991.

[26] A. Pagan. The econometrics of financial markets. Mimeo.

[27] J. Poterba and L. Summers. Mean reversion in stock prices: Evidence and implications. *Journal of Financial Economics*, 22:27–60, 1988.

[28] Lawrence R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, February 1989.

[29] E. Sentana and S. Wadhwani. Semi-parametric estimation and the predictability of stock market returns: Some lessons from Japan. *Review of Economic Studies*, 58:547–563, 1991.

	US	Germany	Japan	France	Canada	Italy	UK
Mean	0.0097	0.0077	0.0006	0.0115	0.0033	0.0009	0.0045
Variance	1.0554	0.9947	0.9227	1.1277	0.8534	1.1683	1.0517
J-B χ^2	2.4603	1.0766	1.4874	2.8760	1.1413	1.6153	8.6345
$Q(1)$	0.5072	2.8200	0.7913	1.5069	2.2558	4.7505	0.2642
$Q(2)$	1.2221	3.9996	0.8065	3.3136	2.7738	7.5195	4.2852
$Q(3)$	1.2909	4.0014	1.5630	3.8135	3.6025	7.5218	5.2748
$Q(4)$	1.6786	4.9236	2.1337	5.2043	5.8515	7.5828	5.5048
$Q(5)$	8.4121	4.9258	2.1337	5.2271	6.1390	12.0011	5.9603
$Q(6)$	9.1366	4.9280	2.1893	5.2592	6.1551	19.7027	6.2924
$Q(7)$	11.9272	5.9939	2.2047	7.4184	9.2809	20.8027	6.2979
$Q(8)$	14.3294	6.0491	2.2327	9.1464	14.1483	25.4442	6.9918
$Q(9)$	14.9102	6.0522	2.4988	10.7440	14.8296	25.6517	10.8575
$Q(10)$	15.9061	9.7040	3.2715	11.1543	16.9600	27.7705	10.8744

Table 1: Basic Statistics and test for Linear Dependence

	US	Germany	Japan	France	Canada	Italy	UK
$Q^2(1)$	1.7245	0.2989	4.3740	0.0365	0.4248	0.5835	0.2499
$Q^2(2)$	1.7270	3.3098	7.0219	0.0708	0.8206	4.7415	1.2252
$Q^2(3)$	2.3818	3.7697	7.1902	0.2131	4.7749	5.0107	2.4772
$Q^2(4)$	2.4592	3.7916	7.6161	6.0524	6.1589	6.1632	2.4888
$Q^2(5)$	3.4976	5.2456	7.6265	6.1251	7.5172	8.4698	2.7512
$Q^2(6)$	4.2055	5.3297	8.4183	6.1676	8.2656	8.5005	8.6316
$Q^2(7)$	7.1784	5.4018	8.9407	6.1883	10.9882	8.5067	8.8589
$Q^2(8)$	7.3979	5.5097	8.9581	6.2050	13.2059	8.9168	11.4294
$Q^2(9)$	9.2919	6.0408	9.4390	6.9452	14.1299	10.8959	11.5308
$Q^2(10)$	9.5223	6.3306	9.4428	7.0728	16.7391	10.9040	11.5417

Table 2: Test for Non-Linear Dependence

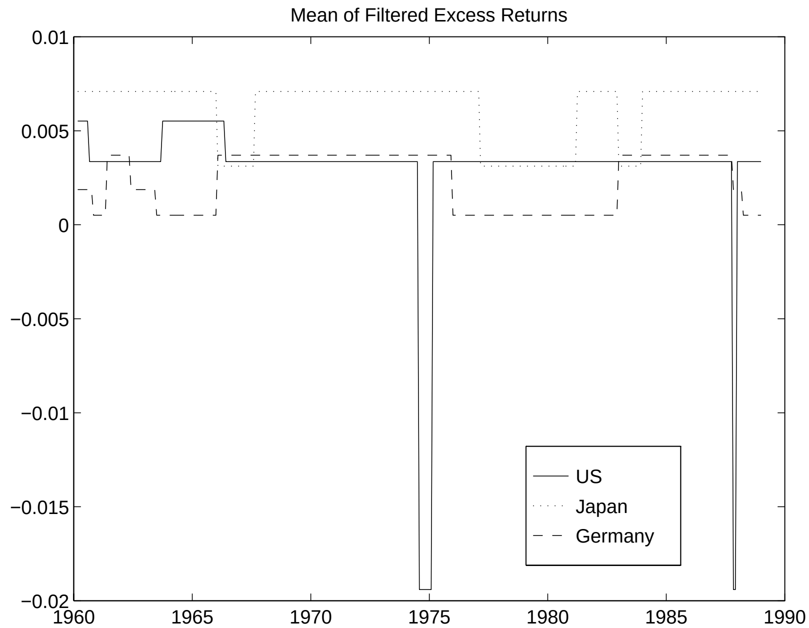

Figure 1: The mean value of excess returns for the three largest economies.

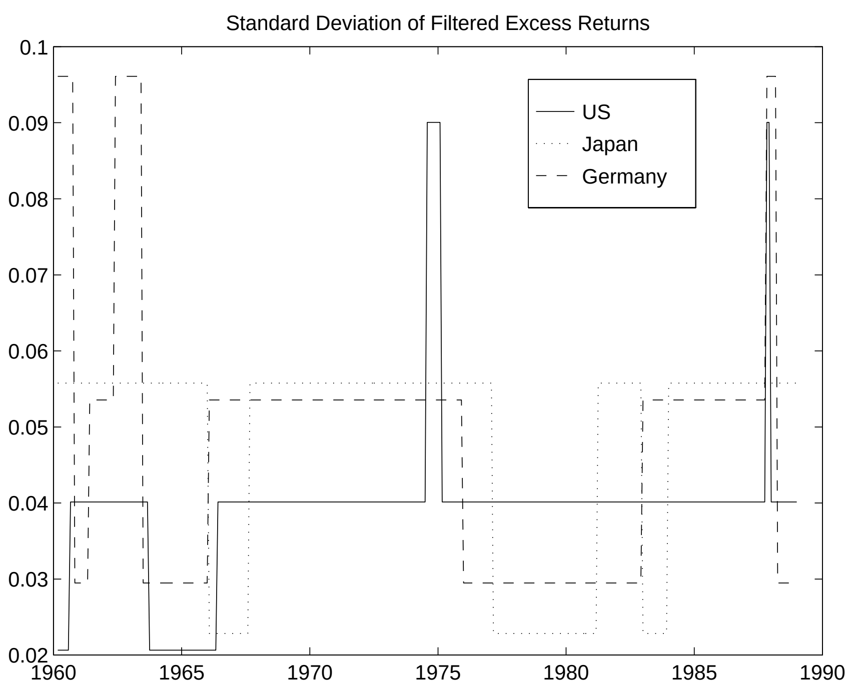

Figure 2: The standard deviation (volatility) of excess returns for the three largest economies.

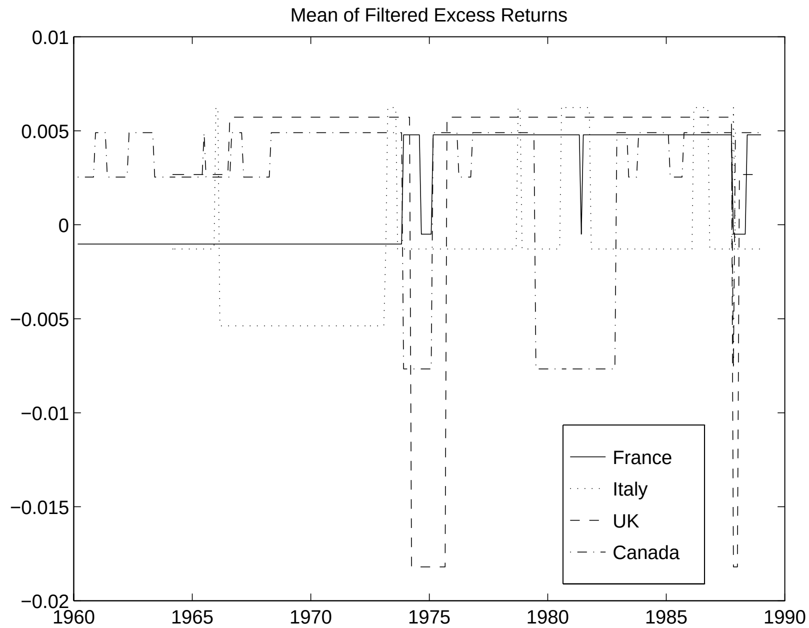

Figure 3: The mean value of excess returns for the remainder of the G7.

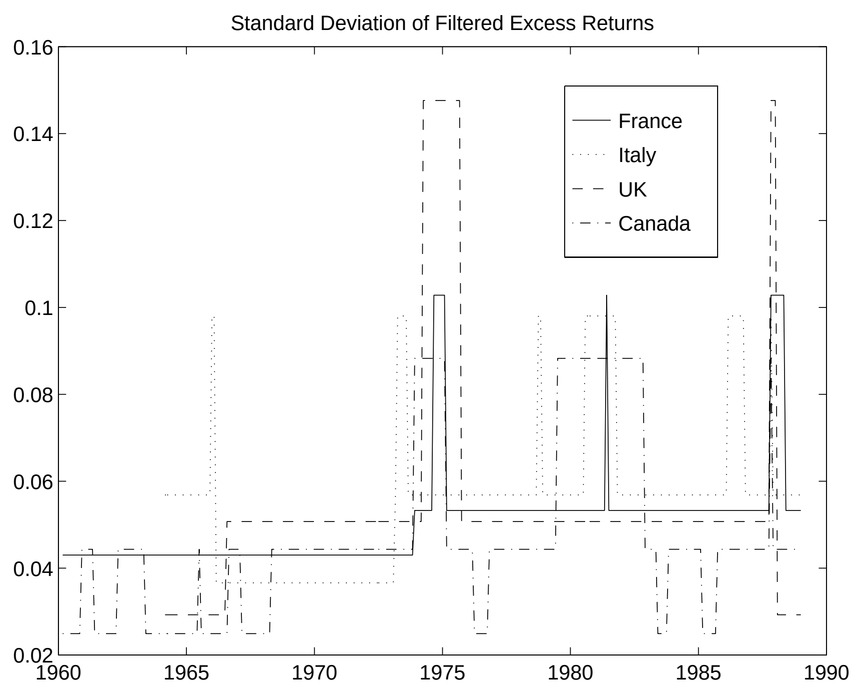

Figure 4: The standard deviation (volatility) of excess returns for the remainder of the G7.