The Korean Economic Review
Volume 28, Number 1, Summer 2012, 5-25.

On the Origins of Conditional Heteroscedasticity in Time Series

Richard Ashley*

The volatility clustering frequently observed in financial/economic time series is often ascribed to GARCH and/or stochastic volatility models. This paper demonstrates the usefulness of reconceptualizing the usual definition of conditional heteroscedasticity as the ($h = 1$) special case of h-step-ahead conditional heteroscedasticity, where the conditional volatility in period t depends on observable variables up through period $t - h$. Here it is shown that, for $h > 1$, h-stepahead conditional heteroscedasticity arises – necessarily and endogenously - from nonlinear serial dependence in a time series; whereas one-step-ahead conditional heteroscedasticity (i.e., $h = 1$) requires multiple and heterogeneously-skedastic innovation terms. Consequently, the best response to observed volatility clustering may often be to model the nonlinear serial dependence which is likely causing it, rather than 'tacking on' an ad hoc volatility model. Even where such nonlinear modeling is infeasible – or where volatility is quantified using, say, a model-free implied volatility measure rather than squared returns – these results suggest a re-consideration of the usefulness of lag-one terms in volatility models. An application to observed daily stock returns is given.

JEL Classification: C22, G12, G13
Keywords: nonlinearity, nonlinear serial dependence, conditional heteroscedasticity, ARCH models, GARCH models.

I. Introduction

In response to the seminal paper by Engle (1982), a vast literature has – and continues to grow – on the topic of autoregressive conditional heteroscedasticity.[1] This literature (and related work on stochastic volatility (SV) and random

Received: Dec. 11, 2011. Revised: Feb. 28, 2012. Accepted: March 5, 2012.

* Department of Economics, Virginia Tech, Blacksburg, Virginia 24061. Phone: (540) 231 6220. Fax: (540) 231 5097. E-mail: ashleyr@vt.edu. The author would like to thank Tim Bollerslev, C. Dahl, C. Kyrtsou, C. Parmeter, D. M. Patterson, J. S. Racine, and D. E. Rapach for helpful comments; all remaining errors and infelicities are his own. The latest version of this paper can be found at http://ashleymac.econ.vt.edu/working_papers/origins_of_conditional_heteroscedasticity.pdf.

[1] See Bollerslev, Chou, and Kroner (1992) and Bollerslev, Engle, and Nelson (1994) for surveys of the early work.

coefficient autoregressive (RCA) models)[2] are motivated by the empirical observation that many financial return time series exhibit apparent heteroscedasticity which is positively autocorrelated – typically, clustered episodes of relatively high variance. This literature is also frequently motivated by an intrinsic interest in modeling the volatility of returns for asset pricing purposes.

All of these approaches, however, are descriptive rather than explicative in nature: they provide no insight into *why* the variance in these financial return series varies over time in this way. Nor do these descriptive approaches in any fundamental way enhance our understanding of – or our ability to forecast – the process generating the returns series itself. Indeed, most ARCH/GARCH frameworks, for example, simply assume that the returns series is a martingale difference process.

This paper examines the proposition that conditional heteroscedasticity can arise endogenously, as a natural consequence of nonlinear serial dependence in the generating mechanism of the return time series. The key innovation here is to sharpen the usual notion of conditional heteroscedasticity, where the current variance of the time series is taken to be history-dependent, by defining "h-step-ahead conditional heteroscedasticity":

Definition: "h-step-ahead conditional heteroscedasticity"

A time series $y(t)$ is h-step-ahead conditionally heteroscedastic – i.e., exhibits "conditional heteroscedasticity at horizon h" – for a value of $h \geq 1$, if and only if

$$\mathrm{var}(y(t+h)\,|\,\{y(t)\})$$

depends on elements of the information set $\{y(t)\}$, where $\{y(t)\}$ consists simply of $\{y(t), y(t-1), y(t-2), \text{etc.}\}$.

The combination of this definition with a fairly general nonlinear generating mechanism for $y(t)$ – characterized by an additively separable innovation term with a fixed variance – is analyzed below. Two interesting new results emerge in Section 2.

First, one-step-ahead conditional heteroscedasticity ($h=1$) is shown to be impossible in this setting. Thus, the only way that one-step-ahead conditional heteroscedasticity can arise in a univariate process with an additive innovation term is through a separately-specified mechanism which is arbitrarily tacked onto the model for $y(t)$ itself in order to drive time evolution in the variance of the innovation term. The essentially *ad hoc* nature of this modeling maneuver is particularly apparent in the ARCH/GARCH and SV frameworks.

[2] Harvey, Ruiz, and Shephard (1994) extend the stochastic volatility framework to multivariate models and survey this literature. Tsay (1987) surveys (and extends) RCA models, showing that this class of models subsumes ARCH/GARCH as a special case.

One-step-ahead conditional heteroscedasticity in a time series thus requires something beyond such a univariate process with additive fixed-variance innovation term. For example, one might posit distinct "states," in some or all of which the innovation variance differs – as in the Markov switching model of Hamilton (1989), the self-exciting threshold autogressive (SETAR) models of Tong (1983), and the smooth transition threshold autoregressive (STAR) models of Teräsvirta and Anderson (1992). Or one might need to consider a multivariate framework, in which the conditional heteroscedasticity arises in a driving time series.

The second result which emerges is that, in contrast, multi-step-ahead conditional heteroscedasticity ($h > 1$) is shown to be a natural – indeed, a necessary – consequence of nonlinear serial dependence in the generating mechanism of $y(t)$. Thus, this kind of nonlinear serial dependence can provide an endogenous explanation for observed variance clustering in a time series. Moreover, this insight also suggests an alternative explanation for observed one-step-ahead conditional heteroscedasticity: this might be arising as a result of nonlinear serial dependence in the mean on a shorter time-scale than the interval at which the process is being sampled.

Note how distinct these results are from the older literature – e.g., Weiss (1986), Bera and Higgins (1997), and others – on choosing between an ARCH model or a nonlinear model for a time series. The present results show that these authors were examining a false dichotomy: with $h > 1$, conditional heteroscedasticity is a straightforward symptom of the nonlinear dependence in the time series.

In Section 3 these general results are illustrated using a one-term bilinear model proposed in Granger and Anderson (1978). Bilinear models have not been shown to be of practical use in modeling actual time series – and are not being proposed for that purpose here. The simple bilinear example used here does, however, provide a particularly clear theoretical illustration of exactly how non-linearity in a model endogenously generates multi-step-ahead conditional heteroscedasticy. An empirical application to daily returns data for Ford Motor Company is given in Section 4, which illustrates the equivalence of non-linear serial dependence and multi-step-ahead conditional heteroscedasticity in actual data. In this example, evidence is presented for both multi-step-ahead conditional heteroscedasticity and for non-linear serial dependence in the Ford Motor Company daily stock returns. Further, it is shown that the fitting errors for nonlinear models for these returns exhibit either substantially reduced evidence (or no evidence at all) of conditional heteroscedasticity.

Section 5 concludes the paper by suggesting that, in many cases, the ultimately most useful response to observed volatility clustering in a time series is to test for and model the non-linear serial dependence in the mean which is likely causing it.[3]

[3] E.g., see Hagiwara and Herce (1999).

In addition, even where an ARCH/GARCH or stochastic volatility model specification is preferred to non-linear modeling,[4] these results suggest a consideration of volatility models which omit the lag-one terms.

II. General Results

Consider a causal, stationary, single-equation dynamic model for a time series $y(t)$:

$$y(t) = \alpha + \sum_{j=1}^{k} \beta_j f_j(\{y(t-j)\})y(t-j) + u(t) \tag{1}$$

where the notation $\{y(t-j)\}$ denotes the information set consisting of $y(t-j)$, $y(t-j-1)$, $y(t-j-2), \ldots$ etc., so that $f_j(\{y(t-j)\})$ is just a general function of the values of the time series prior to and including period $t-j$.[5] The innovation series $u(t)$ is assumed to be generated by a zero-mean martingale difference process with a bounded (and fixed) variance, which is denoted σ^2; boundedness is also tacitly assumed for whatever higher moments of $u(t)$ are needed in order to ensure that $\mathrm{var}(y(t+h)\,|\,\{y(t)\})$ exists.

Note that the functions $f_1(\{y(t-1)\})\ldots f_k(\{y(t-k)\})$ are not assumed to be smooth, nor even necessarily available in closed-form. Thus, Equation 1 subsumes as special cases a wide variety of non-linear generating mechanisms – e.g., the semi-nonparametric models of Gallant and Nychka (1987), noisy Mackey-Glass models[6], and the neural network models analyzed by Teräsvirta, Lin and Granger (1993). It also subsumes switching models, such as SETAR and STAR models, but only those for which the innovation variance is the same in each state; that is, of course, commonly not assumed to be the case for SETAR and STAR models.

The dependence of $y(t)$ on $\{y(t-1)\}$ is parameterized in this way so as to admit of fairly general non-linear dependence in $y(t)$ while also allowing for a graceful restriction to a linearly dependent specification by merely omitting the functions $f_1(\{y(t-1)\})\ldots f_k(\{y(t-k)\})$. Indeed, the only real restrictions imposed by this specification are its univariate nature (although exogenous covariates could be easily added) and the assumption that the innovation process is additively

[4] Or where one quantifies volatility in a different way – e.g., using model-free implied volatility measures, such as the VIX and related measures advocated by Andersen and Bollerslev (1998).

[5] Stationarity is assumed here in both senses usually accorded this term: It is assumed that all of the model parameters and the joint distribution of model innovations are time-invariant and it is assumed that $y(t)$ has been transformed (if necessary) to be an I(0) time series.

[6] E.g., see Kyrtsou and Labys (2006) for a bivariate example or Kyrtsou and Terraza (2003) for a univariate example with ARCH errors tacked on. See also Kyrtsou (2009).

separable with constant variance.

The first point to make is that $y(t)$ generated by the model of Equation 1 can never exhibit any one-step-ahead conditional heteroscedasticity:

Proposition:

If $y(t+1)$ is generated by the model of Equation 1, then

$$\mathrm{var}(y(t+1)\,|\,\{y(t)\}) = \sigma^2 \tag{2}$$

Proof:

This proposition follows directly from the model of Equation 1 and the definition of h-step-ahead conditional heteroscedasticity, with h set to one. In particular, re-writing Equation 1 for $y(t+1)$, the information set $\{y(t)\}$ specifies all of the random variables on the right hand side of the equation except $u(t+1)$. Therefore, the deviation of $y(t+1)$ from its mean – conditional on $\{y(t)\}$ – is merely $u(t+1)$, which has fixed variance, σ^2.

Thus, one-step-ahead conditional heteroscedasticity in a process with an additively separable error term can only arise from either exogenous (unexplained) time variation in the variance of this error term – as in ARCH/GARCH and SV models – or from an assumed multiplicity of such error terms – as in Markov switching and SETAR/STAR models with state-dependent innovation variances. These formulations basically assume the result – time-varying variance – at the outset, but they do provide possible models for the empirically observed time-clustering of volatility in many financial/economic time series.

The new point here is that these formulations are not the only – nor even, perhaps, the most informative – way to model such time-clustered volatility. In particular, the following theorem demonstrates that any kind of non-linear serial dependence of the form given by the model of Equation 1 endogenously generates h-step-ahead conditional heteroscedasticity for one or more values of h greater than one. This h-step-ahead conditional heteroscedasticity can generate correlated clusters of high (or low) volatility without invoking any *ad hoc* exogenous volatility generating mechanism and without positing that the system is characterized by discrete states with distinct model error processes.

Theorem

If $\{y(t+h)\}$ is generated by the model of Equation 1, then

$$\begin{aligned}&\mathrm{var}(y(t+h)\,|\,\{y(t)\})\\ &\quad= \mathrm{var}\Big(\sum_{j=1}^{\min(h-1,k)} \beta_j f_j(\{y(t-j+h)\})y(t-j+h)\,\Big|\,\{y(t)\}\Big) + \sigma^2\end{aligned} \tag{3}$$

Proof:

Re-writing Equation 1 for time period $t+h$:

$$y(t+h)=\alpha+\sum_{j=1}^{k}\beta_j f_j(\{y(t-j+h)\})y(t-j+h)+u(t+h)$$

and grouping together the terms which are fixed and the terms which are stochastic under information set $\{y(t)\}$ yields:

$$\begin{aligned}y(t+h)=&(\alpha+\sum_{j=h}^{k}\beta_j f_j(\{y(t-j+h)\})y(t-j+h)+\sum_{j=1}^{h-1}\beta_j\mu_j)\\&+(\sum_{j=1}^{h-1}\beta_j[f_j(\{y(t-j+h)\})y(t-j+h)-\mu_j]+u(t+h))\end{aligned}$$

where $\mu_j\equiv E[f_j(\{y(t-j+h)\})y(t-j+h)]$.[7] Thus, $\text{var}(y(t+h)\,|\,\{y(t)\})$ is just:

$$E[(\sum_{j=1}^{h-1}\beta_j[f_j(\{y(t-j+h)\})y(t-j+h)-\mu_j]+u(t+h))^2]$$

which implies that

$$\text{var}(y(t+h)\,|\,\{y(t)\}=\text{var}(\sum_{j=1}^{h-1}\beta_j[f_j(\{y(t-j+h)\})y(t-j+h)]\,|\,\{y(t)\})+\sigma^2$$

since the fact that $u(t)$ is a martingale difference implies that the cross term

$$E[2u(t+h)\sum_{j=1}^{h-1}\beta_j[f_j(\{y(t-j+h)\})y(t-j+h)-\mu_j]]$$

is zero, proving the theorem.

Thus, the non-linear generating mechanism for $y(t)$ given in Equation 1 implies that the expression (Equation 3) for the conditional variance of $y(t+h)$ contains a term which might well depend on $\{y(t)\}$, implying h-step-ahead conditional heteroscedasticity. Corollary 1 makes this observation more explicit for the special case where h equals two:

[7] At this point one can either restrict attention to where $h-1$ does not exceed the value of k defined in Equation 1or simply let the function $f_j(\cdot)$ be zero for all values of j in excess of k.

Corollary 1

Where h equals two, Equation 3 reduces to:

$$\text{var}(y(t+2)\,|\,\{y(t)\} = \text{var}(\beta_1[f_1(\{y(t+1)\})y(t+1)]\,|\,\{y(t)\}) + \sigma^2 \tag{4}$$

Thus, if $\{y(t+2)\}$ is generated by the model of Equation 1, $\text{var}(y(t+2)|\{y(t)\})$ depends on elements of $\{y(t)\}$ if $f_1(\{y(t+1)\})$ depends on any elements of $\{y(t)\}$. If $f_1(\{y(t+1)\})$ depends on $\{y(t+1)\}$ but not on $\{y(t)\}$, then it must depend on $y(t+1)$. In that case – since the function $f_1(\cdot)$ is not zero – Equation 1 implies that $y(t+1)$ depends on $y(t)$, and hence on an element of the information set $\{y(t)\}$. Thus, in either case, $\text{var}(y(t+2)|\{y(t)\})$ depends on elements of $\{y(t)\}$ and $y(t)$ is therefore two-step-ahead conditionally heteroscedastic.

In contrast, if the function $f_1(\cdot)$ is a constant (γ) – so that $y(t)$ depends linearly on $y(t-1)$ – then

$$\begin{aligned}\text{var}(y(t+2)\,|\,\{y(t)\}) &= \beta_1^2\gamma^2\,\text{var}(y(t+1)\,|\,\{y(t)\}) + \sigma^2 \\ &= (\beta_1^2\gamma^2 + 1)\sigma^2\end{aligned}$$

where the second equality follows from Equation 2. Clearly, $y(t)$ is not two-step-ahead conditionally heteroscedastic in that case.

Thus, Corollary 1 implies that any time series $y(t)$ driven by additively separable constant-variance innovations exhibits two-step-ahead conditional heteroscedasticity if and only if $y(t)$ depends non-linearly on $y(t-1)$.

Corollary 2 extends the latter portion of Corollary 1 to horizons larger than two by noting that fully linear models – for which the functions $f_1(\{y(t-1)\})\ldots f_k(\{y(t-k)\})$ are all constants – can never exhibit conditional heteroscedasticity at any horizon. This result underscores the role that non-linear serial dependence plays in the origin of conditional heteroscedasticity:

Corollary 2

If $\{y(t)\}$ is generated by the model of Equation 1 but its serial dependence is constrained to be linear, so that $y(t)$ is the AR(k) process,

$$y(t) = \alpha + \sum_{j=1}^{m}\beta_j y(t-j) + u(t)$$

then $\{y(t)\}$ cannot display h-step-ahead conditional heteroscedasticity for any positive value of h. That is, $\text{var}(y(t+h)|\{y(t)\})$ is a constant which does not depend on any of the elements of $\{y(t)\}$.

Proof:

Noting that the restriction to linear serial dependence implies that the functions $f_1(\{y(t-1)\})\ldots f_k(\{y(t-k)\})$ are all constants, Corollary 2 follows directly from Equation 3. The $\mathrm{var}(y(t+h)\,|\,\{y(t)\})$ is essentially identical to the variance of the h-step-ahead forecast of $y(t)$, so this result is already long-known in the literature on ARMA models; a more detailed proof of Corollary 2 along these lines is thus relegated to an appendix.

Combining these results, it can be concluded that – at least for time series generated by models of the form of Equation 1, characterized by an additively separable innovation term with constant variance – conditional heteroscedasticity cannot arise at horizon one, nor for a linear model at any horizon, but will in naturally and endogenously arise in non-linear models for at least some horizons exceeding one. The next section illustrates this point by explicitly identifying the two-step-ahead conditional heteroscedasticity generated by a simple bilinear model.

III. An Illustrative Example: The Conditional Heteroscedasticity Generated by a Simple Bilinear Model

The Bilinear Model introduced by Granger and Anderson (1978) has not found wide use in non-linear time series modeling because of generic problems in showing that these models are invertible. This model is not being proposed for empirical use here, either. Nevertheless, a simple special case of this class of models is ideally suited for illustrating the general results obtained above.

Suppose, then, that $y(t)$ is generated by the particular bilinear model,

$$y(t) = \beta y(t-2)u(t-1) + u(t) \qquad u(t) \sim i.i.d.(0,1) \tag{5}$$

Granger and Anderson (1978) show that this model is invertible for $\beta^2 < \frac{1}{2}$; that is, for these values of β the model can be recast in the form of Equation 1for sufficiently large values of the parameter k. And it is straightforward to show that the $y(t)$ generated by this model are serially uncorrelated. Yet $y(t+1)$ is evidently (non-linearly) forecastable from its own past – $y(t)$, $y(t-1)$, $y(t-2)$, etc. Indeed, since the unconditional variance of $y(t+1)$ is known to be $1/(1-\beta^2)$ and the one-step-ahead forecast error variance – $\mathrm{var}(u(t+1))$ – equals one, this model can have an R^2 approaching one half.

What sort of conditional heteroscedasticity do the $y(t)$ generated by this model exhibit? It is troublesome to re-write Equation 5 in the form of Equation 1; this is why showing invertibility is generically problematic for bilinear models. But it is not difficult to directly calculate the h-step-ahead conditional heteroscedasticity of $y(t)$.

First, illustrating the general result of Theorem 1, note that the one-step-ahead conditional variance of $y(t)$ for this model is just the variance of $u(t+1)$:

$$\begin{aligned}
&\mathrm{var}(y(t+1)\,|\,\{y(t)\}) \\
&\qquad = E[(\beta y(t-1)u(t)+u(t+1)-E[(\beta y(t-1)u(t)+u(t+1)\,|\,\{y(t)\}])^2] \\
&\qquad = E[(\beta y(t-1)u(t)+u(t+1)-\beta y(t-1)u(t)^2] \\
&\qquad = E[(u(t+1))^2] \\
&\qquad = \mathrm{var}(u(t+1)) = 1
\end{aligned} \tag{6}$$

where this derivation uses the fact that $u(t) = y(t) - \beta y(t-2)u(t-1)$ is fixed under information set $\{y(t)\}$.

So the $y(t)$ generated by this model do not exhibit any one-step-ahead conditional hetereoscedasticity. In contrast, the $y(t)$ from this model do exhibit **two**-step-ahead conditional heteroscedasticity:[8]

$$\begin{aligned}
&\mathrm{var}(y(t+2)\,|\,\{y(t)\}) \\
&\qquad = E[(\beta y(t)u(t+1)+u(t+2)-E[(\beta y(t)u(t+1)+u(t+2)])^2] \\
&\qquad = E[(\beta y(t)u(t+1)+u(t+2))^2]
\end{aligned} \tag{7}$$

Note that $\beta y(t)u(t+1)$ is a zero-mean random variable under information set $\{y(t)\}$ and will (in the remainder of this derivation below) contribute to the conditional variance of $y(t+2)$. In contrast, the analogous term – $\beta y(t-1)u(t)$ – in Equation 6 above for the conditional variance of $y(t+1)$ was fixed under this information set.

Continuing the derivation,

$$\begin{aligned}
&\mathrm{var}(y(t+2)\,|\,\{y(t)\}) = E[(\beta y(t)u(t+1)+u(t+2))^2] \\
&\qquad = E[(\beta^2 y(t)^2 u(t+1)^2 + 2\beta y(t)u(t+1)u(t+2)+u(t+2)^2] \\
&\qquad = \beta^2 y(t)^2 E[u(t+1)^2] + 2\beta y(t)E[u(t+1)u(t+2)] + E[u(t+2)^2] \\
&\qquad = \beta^2 y(t)^2 + 1
\end{aligned} \tag{8}$$

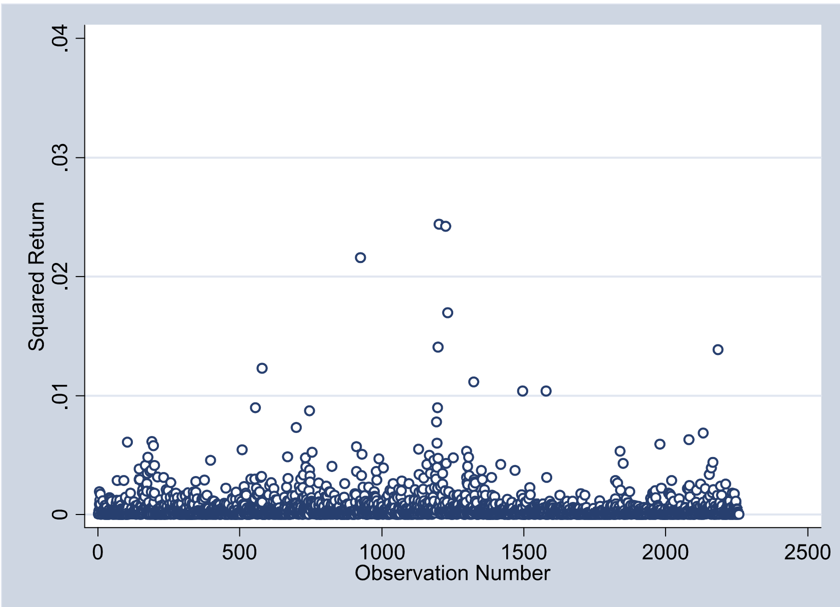
since $E[u(t+1)^2] = E[u(t+2)^2] = 1$ and $E[u(t+1)u(t+2)] = 0$. Thus, the conditional variance of $y(t+2)$ depends directly on the magnitude of $y(t)$ so long as β is non-zero – i.e., so long as $y(t)$ is non-linearly dependent on its own

[8] To simplify the notation, all expectations in the remainder of this section are taken to be conditional on the information set $\{y(t)\} = \{y(t), y(t-1), y(t-2), \ldots\}$.

past.[9]

IV. An Empirical Example of Multi-step-ahead Conditional Heteroscedasticity and its Relation to Non-linear Serial Dependence: Daily Returns to Ford Motor Company Stock

Autoregressive conditional heteroscedasticity is commonly observed in financial returns time series. Squared daily returns to Ford Motor Company stock, plotted below over a sample of 2258 consecutive trading days from January 2, 1998 to December 29, 2006, provide a typical example:[10]



Episodes of high volatility are evident in this time plot. So as to examine the degree to which this is indicative of autoregressive conditional heteroscedasticity and, in particular, of multi-stepahead conditional heteroscedasticity, an AR(p) model for the squared value of these daily Ford returns was identified and estimated,

[9] The $y(t)$ from this bilinear model also exhibit conditional heteroscedasticity at longer horizons, but these results are not particularly revealing. For example, $\text{var}(y(t+3)|\{y(t)\})$ equals $(1+\beta)^2 + \beta^4(y(t-1))^2(u(t))^2$.

[10] Returns data are from the Center for Research in Security Prices (CRSP) at the Graduate School of Business, University of Chicago.

yielding the results reported below in Table 1.[11]

[Table 1] Autoregressions of Squared Ford Motor Company Returns (OLS coefficient estimates)[12]

intercept	.00032**	2.391 (.202)
Lag 1	.076**	omitted
Lag 2	.071**	.077**
Lag 3	.053*	.059*
Lag 4	.084**	.088**
Lag 5	.057**	.064**
Lag 6	.021	.025
Lag 7	.097**	.099**
s^2	1.947×10^{-6}	1.957×10^{-6}
$\overline{R}^2$	.0495	.0444

As is typical of such data, the value of $\overline{R}^2$ (adjusted for degrees of freedom) is not very high for this model. But the evidence for conditional heteroscedasticity is very strong here: the null hypothesis that the coefficients on lags one through seven of the squared return series are all zero can be rejected with a p-value less than .00005.

Notably, omitting the lag-one term only slightly diminishes the adjusted R^2 of the relationship between squared returns and its own past; from .0495 to .0444. Evidently, the bulk of the conditional heteroscedasticity in the Ford Motor Company returns series is multi-step-ahead. On the other hand, since the coefficient on the squared return at lag one is statistically significant at the 1% level, the possibility of conditional heteroscedasticity at lag one cannot be ruled out in this case.

Based on the theorem proven in Section 2, then, one would expect that the Ford Motor Company returns series exhibits significant non-linear serial dependence, even though this returns series is only very weakly autocorrelated. And indeed this is the case. Table 2 lists the results from several tests for non-linear serial dependence and, for each of these tests, the null hypothesis that only linear serial dependence is present in the returns series can be convincingly rejected.

[11] As is typical with daily stock returns, Ford Motor Company returns themselves – not the squared returns – are very weakly autocorrelated. Identification of an AR(p) model for this return series yields a value of p equal to one, with an adjusted R2 of just .008 for the estimated model. As will be evident from the results given in Table 3, this model is so weak that the results given in Table 1 are not materially altered if the squares of the ("prewhitened") fitting errors from this AR(1) model are used instead.

[12] A single asterisk indicates that a coefficient estimate is significantly different from zero at the 5% level; a double asterisk indicates that a coefficient estimated is significantly different from zero at the 1% level.

[Table 2] Results of Formal Tests for Non-linear Serial Dependence in Ford Motor Company Returns

Test	p-value for rejecting H_o: pre-whitened Ford returns ~ i.i.d.
BDS	< .001
Hinich Bicovariance	< .001
Tsay	.001

These particular tests for non-linear serial dependence in the mean are fairly standard – all are documented in Ashley and Patterson (2006, Appendix 1) – but a few comments on them are in order here. Each of these tests is actually a test for serial independence and hence is applied to the ("pre-whitened") fitting errors from the weak AR(1) model for the Ford Motor Company daily returns series.[13] The Brock-Dechert-Scheinkman (BDS) test is known to be poorly-sized even in samples of this length, so the p-values for all three tests were evaluated using the bootstrap.

The BDS test is a nonparametric test based on an estimate of the correlation integral of the time series. In essence, for a time series x_t with standard deviation ε, the correlation integral counts up the number of m-histories – each defined as the sequence $\{x_t, x_{t-1} \dots x_{t-m}\}$ – lying within an m-dimensional hypercube of size $j\varepsilon$. The BDS test is typically done for j equal to .5, 1.0, and 2.0 and for embedding dimensions (m) equal to 2, 3, and 4. Simulation results in Ashley and Patterson (2006) indicate that – bootstrapping notwithstanding – the BDS test is still missized for embedding dimensions exceeding 2.[14] That does not actually matter here, however, as this time series is so non-linearly dependent (and lengthy) that all nine parameterizations of the BDS test yield essentially the same result.

The Hinich bicovariance test – Hinich and Patterson (1995) – examines the squared returns series, x_t, for non-linear serial dependence in a different way. This test estimates bicovariances – $E[(x_1 - \overline{x})(x_{t-r} - \overline{x})(x_{t-s} - \overline{x})]$ – for values of the integers r and s less than or equal to ℓ; here ℓ was set to five. These bicovariances should all equal zero if the time series is serially independent.

The Tsay test – Tsay (1986) – examines the squared returns for non-linear serial dependence in yet another way, basically by looking for quadratic serial dependence using ordinary multiple regression methods. As implemented for Table 2, these regressions incorporate lags of up to five days; again, see Ashley and Patterson (2006, Appendix 1) for implementation details.

[13] See footnote 10 above.

[14] Values of $m>2$ are problematic because of occasional under-estimates of the order of the AR(p) pre-whitening model; the bootstrapping is done using the resulting fitting errors, so it cannot properly account for this contribution to the chance of incorrectly rejecting the null hypothesis. See Ashley and Patterson (2006, Appendix 2) for more details on this point.

Many additional tests for non-linear serial dependence in the mean have been proposed, but the results Table 2 suffice to show that, with this sample length, there is very strong evidence for non-linear serial dependence in the Ford returns data – exactly as the theoretical results obtained here indicate ought to be the case (for a generating mechanism driven by an additive, homoscedastic innovation), in view of the evident multi-step-ahead conditional heteroscedasticity observed in this time series.

Table 3 reports results for autoregressions of squared Ford Motor Company returns in which the returns have been filtered (prior to squaring) so as to reduce or remove the serial dependence in the returns series – weak linear serial dependence in the case of the first column of the table, substantial non-linear serial dependence in the case of the remaining columns. Thus, the first column of Table 3 reports results on an autoregression of the squared residuals from a linear model for the Ford Motor Company returns.[15] These results confirm that, as is common with daily returns series, the linear serial dependence in the Ford Motor Company returns is so weak that the choice to use squared returns rather than squared pre-whitened returns in Table 1 was inconsequential. The remaining columns of Table 3 report results on autoregressions of the squared fitting errors from three nonparametric models for the non-linear serial dependence in the Ford returns. Each of these models is discussed, in turn, below.

The first of these nonparametric models is a standard kernel regression model for the Ford returns using eight lags in the Ford returns and a gaussian kernel with a fixed bandwidth chosen by the least-squares cross validation method.[16] The computational burden for these estimations was very heavy and the resulting model fits the data fairly poorly (considering the effort applied) with a raw R^2 of only .107. On the other hand, the kernel estimation procedure is numerically robust and the results given in Table 3 indicate that the degree of conditional heteroscedasticity remaining in its fitting errors is greatly reduced, albeit not completely eliminated.

The second nonparametric model for the Ford Motor Company returns was identified and estimated using local-polynomial regression methods. Here a quadratic polynomial is fit to the data in a window around each observation using a weighting function and a bandwidth chosen to minimize a generalized cross validation score; these latter are analogous to the similarly named constructs in kernel regression. For a detailed description of the local polynomial regression

[15] The best ARMA specification for Ford returns is an AR(1) model with an adjusted R^2 of only .008.

[16] The estimation was done using the "np" package in R – see Li and Racine (2007), Hayfield and Racine (2008) for details. The model with eight lags was clearly optimal on either the AIC, SIC, or adjusted R^2 criteria. Only fixed-bandwidth results are reported because the algorithms for calculating adaptive and generalized nearest neighbors bandwidths do not converge properly.

methodology (and the LOCFIT package for R implementing it) see Loader (1999). The local-polynomial regression model identified and estimated here used five lags in the Ford returns and a smoothing parameter (α) equal to .87. With a raw R^2 of .217 this model fits the Ford returns data noticeably better than did the kernel regression model, although still not all that well considering the flexibility of this model family. LOCFIT is without a doubt the best multivariate implementation of local-polynomial regression which is available – and the computational burden it imposes is vastly smaller than for kernel regression – but the reader is warned that LOCFIT mis-behaves somewhat for models with six lags and fails altogether for models with seven lags (in both Linux and Windows versions, with these and other data), so this five-lag model specification must be viewed with a good deal of skepticism. Be that as it may, the results in Table 3 show that the fitting errors from this model exhibit no evidence of conditional heteroscedasticity at lags larger than one and, overall, greatly reduced conditional heteroscedasticity compared to that observed in the returns series itself.

The third nonparametric model for the Ford returns was identified and estimated using penalized spline regression methods. The penalized spline method approximates the mean of the dependent variable by a spline function of the set of explanatory variables – i.e., by a piecewisecontinuous patchwork of polynomials – so as to minimize an objective function which is the sum of the squared fitting errors plus a penalty depending on an estimate of the average smoothness of the spline function. This smoothness penalty is parameterized as a linear function of the average value of the m^{th} derivative of the spline function; the value of m must exceed half the number of explanatory variables in the model. Most implementations of penalized spline regression – e.g., Kauermann, Krivobokova and Semmler (2008) and Krivobokova and Kauermann (2007) – restrict the analysis to either bivariate regression models or to additively separable models; the penalized spline regression implementation used here (procedure TPSPLINE in SAS) imposes neither of these restrictions.

The effective number of degrees of freedom consumed in the penalized spline fitting process increases sharply as the number of explanatory variables rises, however. This is reflected in an adjusted residual variance reported by the routine and in a resulting limitation on how many explanatory variables can be included in the model for a given sample length. For the Ford Motor Company returns a maximum of nine lags was therefore feasible; a seven-lag model minimizes the adjusted residual variance and yields an adjusted R^2 of .419. Reference to Table 3 shows that its fitting errors exhibit no signs of conditional heteroscedasticity at lags larger than one.[17]

[17] At first glance it is a bit surprising that the penalized spline model appears to even eliminate the (apparently) significant conditional heteroscedasticity in the Ford returns at lag one. This is likely due

While the penalized spline model appears to do the best job of modeling the non-linear serial dependence leading to the conditional heteroscedasticity observed in the Ford Motor Company returns, the broader issue of which of these nonparametric models is the "best" one is an open question. For example, the penalized spline model fits the sample data best, but quite likely it is the most severely over-fitted model. Post-sample forecasting could help resolve this issue, but that issue is beside the point here: the essential result is that the conditional heteroscedasticity in the fitting errors from all three non-linear models for the Ford Motor Company returns is either eliminated or substantially reduced compared to the conditional heteroscedasticity observed in the returns themselves.

[Table 3] Autoregressions of Squared Fitting Errors From Models of Ford Motor CompanyReturns (coefficient estimates)[18]

	Linear Model	Non-Linear Models[19]		
	AR(p) Model {p=1}[20]	Kernel Regression	Local Polynomial Regression	Penalized Spline Regression
intercept	.0003283**	.0004408**	.0004183**	.0002361**
Lag 1	.068**	.066**	.063**	.028
Lag 2	.080**	.074**	.024	-.013
Lag 3	.049*	.003	.014	-.014
Lag 4	.081**	-.036	-.001	-.033
Lag 5	.053**	.003	.006	-.025
Lag 6	.025	-.038	-.007	-.033
Lag 7	.095**	-.017	.116**	-.026
s^2	1.929 × 10-6	1.239 × 10-6	1.466 × 10-6	1.990 × 10-6
$\bar{R}^2$	.0456	.011	.016	.002

In summary, then, most of the observed conditional heteroscedasticity in the

---

to sampling error in the estimated coefficient at lag one in the autoregression of the squared errors from the penalized spline model reported in Table 3, as the estimated standard error for this coefficient estimate (.021) is substantial.

[18] A single asterisk indicates that a coefficient estimate is significantly different from zero at the 5% level; a double asterisk indicates that a coefficient estimated is significantly different from zero at the 1% level. Returns data are from the CRSP tapes.

[19] See text above for a description of how these three nonparametric models for the Ford returns were identified and estimated. Note that the autoregressions reported in this table are linear regressions of the squared fitting errors from these non-linear models against lags of these squared fitting errors – these are not the non-linear models for the Ford Motor Company returns series!

[20] The results in this column are comparable to (and not statistically different from) those given in Table 1. Table 1 reports results on autoregressions of the squared Ford Motor Company returns themselves; this column reports analogous results using the squared fitting errors from the weak AR(1) linear model for these returns.

Ford Motor Company daily returns is at lags greater than one; and there is strong evidence of non-linear serial dependence in the these returns over this sample period. Finally, the fitting errors for smooth non-linear models which more-or-less successfully remove this non-linear serial dependence in the returns series largely or completely eliminate its conditional heteroscedasticity, confirming and illustrating the result of the theorem proven in Section 2.

V. Conclusions

The theoretical results obtained here show that multi-step-ahead conditional heteroscedasticity is a natural – indeed, a necessary – result of non-linear serial dependence in a time series driven by an additive innovation term with fixed variance. In view of the fact that many time series (notably including many financial return time series) exhibit substantial evidence of such non-linear serial dependence, this result suggests that non-linear serial dependence in the model for the series itself might typically be the underlying cause of the volatility clustering observed in a number of such time series, in addition to dependence on driving co-variates which are themselves conditionally heteroscedastic.[21]

In those particular instances where correlated volatility fluctuations are observed, but where non-linearity tests – such as those reviewed in Barnett, et al. (1997), Kyrtsou and Serletis (2006), and in Ashley and Patterson (2006) – fail to detect any evidence of non-linear serial dependence in the mean, then there would seem to be little alternative to an *ad hoc* specification of a separate process driving this serial dependence in the volatility of the time series.[22]

However, where – as is common – such testing does indicate the presence of statistically significant non-linear serial dependence in the mean, the results obtained here indicate that the ARCH/GARCH or stochastic volatility modeling approaches would appear to be descriptive stop-gaps. Indeed, in such instances, the observed evidence for conditional heteroscedasticity – even that obtained using tests allowing for non-linearity, such as Blake and Kapetanios (2007) – is indicating that explicit modeling of this non-linear serial dependence in the time series is likely crucial to correctly understanding its underlying generating mechanism. This non-linear generating mechanism is then recognizably the source of the time variation in both the conditional mean and the conditional variance of the series. This would suggest that it might often be more fruitful in such instances to model the non-

[21] In other words, there is no need to 'tack on' an *ad hoc* ARCH/GARCH or SV model to generate the observed conditional heteroscedasticity in such time series.

[22] As noted in Section 1, if one-step-ahead conditional heteroscedasticity is the most important feature of the data, one might also (if possible) decrease the sampling interval or consider multivariate modeling.

linear serial dependence in the time series itself, rather than to content oneself with an auxiliary ARCH/GARCH or SV model of the model innovation dispersion.

Interestingly, analysts with a strong preference for making ARCH/GARCH or stochastic volatility models can also benefit from the insights obtained above. In particular, the result obtained here – that nonlinear serial dependence in the mean can cause conditional heteroscedasticity only at horizons exceeding one period – implies that the status of the lag-one terms routinely included in ARCH/GARCH and stochastic volatility model specifications differs from that of terms at longer lags: these terms at lag one correspond to a form of conditional heteroscedasticity which cannot arise from a non-linear generating mechanism driven by a single innovation with a unique, time-independent variance.

References

Andersen, T. G. and T. Bollerslev (1998), "Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts," *International Economic Review*, 39, 885-905.

Ashley, R. and D. M. Patterson (2006), "Evaluating the Effectiveness of State-Switching Models for U.S. Real Output," *Journal of Business and Economic Statistics*, 24(3), 266-77.

Barnett, W. A., A.R. Gallant, M.J. Hinich, J.A. Jungeilges, D.T. Kaplan, and M.J. Jensen (1997), "A Single-Blind Controlled Competition Among Tests for Nonlinearity and Chaos," *Journal of Econometrics*, 82, 157-92.

Bera, A. K. And M. L. Higgins (1997), "ARCH and Bilinearity as Competing Models for Nonlinear Dependence," *Journal of Business and Economic Statistics*, 15(1), 43-50.

Blake, A. P. and G. Kapetanios (2007), "Testing for ARCH in the Presence of Nonlinearity of Unknown Form in the Conditional Mean," *Journal of Econometrics*, 137, 472-488.

Bollerslev, T., R. Y. Chou, and K. F. Kroner (1992), "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence," *Journal of Econometrics*, 52, 5 - 59.

Bollerslev, T., R. F. Engle, and D. B. Nelson (1994), "ARCH Models" in R. F. Engle and D. L. McFadden (eds.) *Handbook of Econometrics*, 4, pp. 2959-3038, Amsterdam: Elsevier.

Brock, W. A., Dechert W., and Scheinkman J. (1996), "A Test for Independence Based on the Correlation Dimension," *Econometric Reviews*, 15, 197-235.

Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflations," *Econometrica*, 50, 987-1008.

Gallant, A. R. and D. W. Nychka (1987), "Seminonparametric Maximum Likelihood Estimation," *Econometrica*, 55, 363-90.

Granger, C. W. J. and A. A. Anderson (1978), *An Introduction to Bilinear Time Series Models*, Vandenheur and Ruprecht: Gottingen.

Hagiwara, M. and M. A. Herce (1999), "Endogenous Exchange Rate Volatility, Trading Volume and Interest Rate Differentials in a Model of Portfolio Selection," *Review of International Economics*, 7(2), 202-218.

Hamilton, James (1989), "A New Approach to the Economic Analysis of Non-Stationary Time Series and the Business Cycle," *Econometrica*, 57, 357-84.

Harvey, A. C., E. Ruiz, and N. Shephard (1994), "Multivariate Stochastic Volatility Models," *The Review of Economic Studies*, 61(2), 247-264.

Hayfield, Tristen and J. S. Racine (2008), "np: Nonparametric Kernel Smoothing Methods for Mixed Datatypes," R package version 0.14-3.

Hinich, M. and Patterson D. M. (1995), "Detecting Epochs of Transient Dependence in White Noise," unpublished manuscript, University of Texas at Austin.

Hinich, M. and Patterson D. M. (1998), "The Episodic Behavior of Dependencies in High Frequency Stock Returns," unpublished manuscript, Department of Finance, Virginia Tech.

Kauermann, G., T. Krivobokova and W. Semmler (2008), "Filtering Time Series with Penalized Splines," unpublished manuscript.

Krivobokova, T. and G. Kauermann (2007), "A Note on Penalized Spline Smoothing With

Correlated Errors," *Journal of the American Statistical Association*, 102(440), 1328-1337.

Kyrtsou, C. (2009), "Re-examining the Sources of Heteroskedasticity: the Paradigm of Noisy Chaotic Models," *Physica A*, forthcoming.

Kyrtsou, C. and W. C. Labys (2006), "Evidence for Chaotic Dependence Between US Inflation and Commodity Prices," *Journal of Macroeconomics*, 28, 256-266.

Kyrtsou, C. and A. Serletis (2006), "Univariate Tests for Nonlinear Structure," *Journal of Macroeconomics*, 28, 154-168.

Kyrtsou, C. and M. Terraza (2003), "Is it Possible to Study Chaotic and ARCH Behaviour Jointly? Application of a Noisy Mackey-Glass Equation with Heteroscedastic Errors to the Paris Stock Exchange Returns Series," *Computational Economics*, 21, 257-276.

Li, Q. and J. S. Racine (2007), *Nonparametric Econometrics: Theory and Practice,* Princeton University Press: Princeton.

Loader, C. (1999), *Local Regression and Likelihood*, Springer-Verlag: New York.

Patterson, D. M and R. Ashley (2000), *A Nonlinear Time Series Workshop,* Kluwer Academic Publishers: Boston.

Teräsvirta, T. and H. Anderson (1992), "Characterising Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models," *Journal of Applied Econometrics*, 7, 119-36.

Teräsvirta, T., C. F. Lin, and C. W. J. Granger (1993), "Power of the Neural Network Linearity Test," *Journal of Time Series Analysis*, 14, 209-20.

Tong, H. (1983), *Threshold Models in Nonlinear Time Series Analysis*, Springer-Verlag: New York.

Tsay, R. S. (1986), "Nonlinearity Tests for Time Series," *Biometrika*, 73, 461-6.

Tsay, R. S. (1987), "Conditional Heteroscedastic Time Series Models," *Journal of the American Statistical Association*, 82, 590-604.

Weiss, A. A. (1986), "ARCH and Bilinear Time Series Models: Comparison and Combination," *Journal of Business and Economic Statistics*, 4(1), 59-70.

Appendix: Proof of Corollary 2

If $\{y(t)\}$ is generated by the model of Equation 1 but its serial dependence is constrained to be linear, so that $y(t)$ is the AR(m) process,

$$y(t) = \alpha + \sum_{j=1}^{k} \beta_j y(t-j) + u(t) \tag{A1}$$

then $\{y(t)\}$ cannot display h-step-ahead conditional heteroscedasticity for any positive value of h. That is, $\mathrm{var}(y(t+h)\,|\,\{y(t)\})$ is a constant which does not depend on any of the elements of $\{y(t)\}$.

Proof:

Since $\mathrm{var}(y(t+h)\,|\,\{y(t)\})$ is essentially identical to the variance of the h-step-ahead forecast of $y(t)$, this result is already long known in the literature on ARMA models. Equation A1 can be re-written as

$$\varphi(B)y(t) = \alpha + u(t)$$

where

$$\varphi(B) \equiv 1 - \sum_{j=1}^{k} \beta_j B^j$$

The infinite-order lag polynomial $\Psi(B) \equiv [\varphi(B)]^{-1}$ exists, with weights Ψ_o, Ψ_1, Ψ_2 which eventually decline in magnitude sufficiently quickly so that $\sum_{j=0}^{\infty} \Psi_j^2$ is finite because the AR(k) process given by Equation A1 is assumed stationary. Multiplying both sides of Equation A1 by $\Psi(B)$ then yields the MA(∞) process,

$$y(t) = \alpha \sum_{j=0}^{\infty} \Psi_j + \sum_{j=0}^{\infty} \Psi_j u(t-j)$$

so that,

$$\begin{aligned} y(t+h) &= \alpha \sum_{j=0}^{\infty} \Psi_j + \sum_{j=0}^{\infty} \Psi_j u(t+h-j) \\ &= \alpha \sum_{j=0}^{\infty} \Psi_j + \sum_{j=h}^{\infty} \Psi_j u(t+h-j) + \sum_{j=0}^{h-1} \Psi_j u(t+h-j) \end{aligned}$$

Note, however, that Equation A1 implies that $\{u(t)\}$ is fixed under the information set $\{y(t)\}$.

Therefore,

$$y(t+h) = E[y(t+h) \mid \{y(t)\}] + \sum_{j=0}^{h-1} \Psi_j u(t+h-j)$$

and hence

$$\mathrm{var}(y(t+h) \mid \{y(t)\}) = \sigma^2 \sum_{j=0}^{h-1} \Psi_j^2$$

Thus, the h-step-ahead conditional variance is a constant which does not depend on any of the elements of $\{y(t)\}$, proving the corollary.