RESPONSE to referee report 2 (dated 9 June 2008) of *Economics E-Journal* discussion paper DP 2008-18: *Forecast Evaluation of Explanatory Models of Financial Return Variability*

Response to **General remarks**:

1. The relationship between continuous- and discrete time models, are not mirrored in the simulation study in Section 3, which solely considers discrete time models. I interpret this comment to mean that—in the view of referee 2—section 3 should have considered continuous time models as well. And the fact that it does not means—in the view of referee 2—that there is a discrepancy between section 2 and 3. If this is the view of referee 2 (it is very possible that I am misinterpreting), then it suggests s/he has not fully understood section 2. The following sentence suggests this is the case: "Clarifying the link between Section 2 and the simulation study might also make it easier to understand the implications of the theoretical considerations in Section 2, which are (at least to me) not obvious". A consequence of the discussion in section 2 is exactly that it would be wrong to sample from continuous time models in section 3 for the objectives of the paper, and key to the understanding of this and many of the other choices in section 3 is an adequate understanding of section 2. Moreover, since referee 2 is not specific about which aspects of section 2 that are unclear to her/him, I find it difficult to address this first point. Additional expositions that may aid the understanding of section 2 are Bauwens et al. (2006, section 2.1), Sucarrat (2006, pp. 1-6, section 2.1 and section 4.4.1), Sucarrat (2007, section 4.2 in particular) and Bauwens and Sucarrat (2008).

2. One could fear that at least some of the conclusions depend heavily on the chosen model. It is always the case that, strictly speaking, simulation results apply only to situations where the simulation DGP is a congruent (or otherwise "true") representation of the DGP. Nevertheless, the simulation DGP can be seen as a simplified version of many specifications that belong to the ARCH and SV classes of discrete time models. Since the most frequently used models in the explanatory modelling of financial variability belong to the ARCH and SV classes of models, this suggests the simulation results hold in quite a few cases of practical interest. See also my response to point 3 below in the specific remarks part.

Response to **Specific remarks**:

1. In the last sentence of Section 2.1 (p. 4) it is stated that smaller absolute values of z_t *indicate that* $g(\mathbf{x}_t, \mathbf{b})$ *and* $h(\mathbf{y}_t, \mathbf{c})$ *successfully explain the variation in* r_t *and* e_t^2*. However, small values of* $|z_t|$ *can be obtained by simply choosing* $h(\mathbf{y}_t, \mathbf{c})$ *very large—some identification condition must be missing.* It is not correct that I say this, but possibly I should add a sentence or two in order to make the meaning of the sentence in question clearer. What I say is: "the better $g(\mathbf{x}_t, \mathbf{b})$ and $h(\mathbf{y}_t, \mathbf{c})$ explain the variation in r_t and e_t^2, respectively, the smaller z_t is likely to be in absolute value". Symbolically this might be written as something like "better $g(\mathbf{x}_t, \mathbf{b})$ and $h(\mathbf{y}_t, \mathbf{c})$" $\Rightarrow$ "smaller $|z_t|$". However, I do not say the converse, namely that "smaller $|z_t|$" $\Rightarrow$ "better $g(\mathbf{x}_t, \mathbf{b})$ and $h(\mathbf{y}_t, \mathbf{c})$", which seems to be the understanding of referee 2. Indeed, the fact that small $|z_t|$ can be arbitrarily achieved by large values of $h(\mathbf{y}_t, \mathbf{c})$ is the likely reason why kurtosis does so bad in ranking the models in the simulations. With respect to a possible "identification condition", one that naturally suggests itself—or at least within a GETS context—is variance-congruence.[1] For example, my guess is that Kurtosis might be useful in selecting between parsimonious, variance-congruent models obtained through simplification of the same general and unrestricted variance-congruent model (GUM). In other words, when multiple-path GETS specification search results in more than one model that parsimoniously encompass the GUM.

2. I think referee 2 meant to write: *On p. 8 it is stated that the condition $\alpha + \beta \geq 1$ implies non-stationarity. This is not true in general ($\alpha + \beta \geq 1$ implies non-covariance stationarity).* This is correct, I should have been more specific with respect to which type of stationarity I had in mind.

3. When introducing the DGP used for simulation in Section 3.1 it is regrettable that the choice of model is not justified any further. Has it for instance been employed in other studies? The simulation DGP nests the widely used (plain) GARCH(1,1) specification. Moreover, if we interpret the GARCH(1,1) part as proxying a persistence structure in the variance equation, the "jump" part as proxying a structure which most of the time is close to zero but sometimes large, and the IID series $\{x_t\}$ as proxying an approximately non-autocorrelated structure in the mean, then a vast class of models that are widely used is approximately nested within the simulation DGP. In the paper (first paragraph of section 3.1) I try to communicate this, but it seems there is room for improvement in the exhibition. Note also that a considerable amount of additional simulations, which are mentioned and referred to in the paper but not recounted in detail, were also undertaken. The additional simulations sought to establish to what extent parameter values that differed from the benchmark values affected the results. So for explanatory modelling purposes the results of the study are quite general.

4. The last few lines on p. 8 compare the value 0.45 to Table 1, but this table does not contain anything resembling this value. Indeed Table 1 seems to be a copy of Table 4. It is correct that table 1 is faulty. The correct table is reproduced below as table 1 after the references. An otherwise identical version of my paper, but with a correct table 1 is available via my webpage: `Http://www.eco.uc3m.es/sucarrat/`.

[1] A model is said to be a variance-congruent representation of a DGP if five properties are satisfied: (1) the standardised residual is an innovation with respect to the conditioning information-set, (2) the parameters are stable, (3) the conditioning variables are weakly exogenous with respect to the parameters, (4) the model is economically justifiable, and (5) the model is data-consistent. See Hendry (1995), Mizon (1995) and Sucarrat (2008, section 2) for more detailed discussions on the notion of congruency. The latter contains a discussion on the distinction between mean and variance congruency.

5. The parameters of model 1 (used for forecasting, see (6) on p. 10) are kept fixed at their true values, but the author admits these must be estimated in empirical applications. This seems like an unnecessary simplification and I would suggest that these parameters be estimated in the simulation study as well. The effect of estimation uncertainty is indeed an aspect of practical interest in forecast evaluation of models of financial variability. However, adding this ingredient to the current study would deviate from the main objectives of the paper. First, it is not necessary in order to achieve the objectives, and second the absence of estimation uncertainty is not just an "unnecessary simplification". Estimating the parameters at each simulation raises additional issues which complicates things considerably, and which are likely to produce misleading simulation results unless handled very carefully.

There are several additional issues to address if the parameters were to be estimated at each simulation. Most importantly, fitting (7) and (8) to series generated by the simulation DGP given by (6) gives rise to the question of whether the chosen estimation method produces fitted versions of (7) and (8) that are variance-congruent representations of the simulation DGP. The extent to whether this is the case or not depends on the chosen estimation method and can, in principle, be checked either analytically or numerically through simulation. In practice, however, analytical checking is not straightforward (or at least not for me) due to the complexity of the problem, and numerical simulation can be misleading due to approximation error. An example of the latter is suggested by tables 2 to 4 (below, after the references). Table 2 contains the averages of parameter estimates obtained by Gaussian maximum likelihood (ML) estimation of (6), (7) and (8), respectively, given that (6) is the simulation DGP with parameter values equal to the socalled benchmark values.[2] The results suggest the estimation algorithm of the arguably most popular commercial software on the market (EViews) provides reasonably accurate estimates of all the parameters of (6), except for ω which is negative and far from the right value. The estimate of ω should be close to 0.02 but is about -0.015. So how can we be sure that that the estimates of (7) and (8) produced by the algorithm are precise? This is a real issue because table 3 suggests that two properties that should be satisfied are often not satisfied by the average parameter estimates. The two properties are no first order serial correlation (AR(1)) and no first order autoregressive conditional heteroscedasticity (ARCH(1)) in the standardised residuals. Violation of these two properties are typically taken as an indication of the model(s) in question not being variance-congruent. Using the average parameter estimates in table 2 obtained for $T = 20\ 000$, table 3 contains the rejection frequencies of the nulls of no first order AR(1) and no first order ARCH(1) in the standardised residuals, using a nominal level of 5%. The rejection frequencies are close to the nominal level at all sample sizes for (6), but this cannot be taken as an indication of congruence since the estimate of ω is negative and therefore faulty. For the other models

[2] The results of table 2 are robust across a range of estimation settings. For example, using the Berndt-Hall-Hall-Hausman algorithm (a modified Newton-Raphson algorithm) instead of Marquardt (a modified Gauss-Newton algorithm), setting the value of b to a number close to zero, say, 0.01, or turning on backcasting produces virtually identical simulation results. In particular, in all cases the average estimate of ω is negative and around -0.015.

the rejection probabilities are not close to the nominal level for ARCH(1), since they tend to 1 as T goes to 20 000. Now, is the lack of variance-congruency for the fitted versions of (7) and (8) due to the estimation method, the numerical estimation algorithm, or is it because (7) and (8) cannot be variance-congruent representations of the simulation DGP for any estimation method? Table 4 suggests the problem lies with the numerical algorithm's estimate of ω_m. Table 4 contains the rejection frequencies of no AR(1) and no ARCH(1) in the standardised residuals using the "correct" parameter values for (7) and (8). That is, the (benchmark) parameter values used in the paper. For $T = 1000$ the rejection frequencies of both the AR(1) and ARCH(1) tests are close to the nominal level for all three models. As the sample size increases the frequencies remain close to the nominal level for model 1. For the fitted versions of (7) and (8), however, the ARCH(1) rejection frequencies increase to 11% and 16% as T reach 20 000. Now, this is not necessarily because the fitted versions of (7) and (8) are not congruent representations, since further investigation would be needed in order to identify the exact cause. Nevertheless, even if the fitted versions of (7) and (8) are not variance-congruent, the rejection frequencies suggest the "correct" parameter estimates used in the paper produce models that are much closer to congruence than when estimating the parameters with standard approaches. In econometric practice one typically searches for a variance-congruent model by changing specification, estimation method and/or estimation options before using the model for forecasting. Substantial effort would be needed in order to adequately incorporate and address all these issues in a simulation study.

Summarised, then, with respect to the suggestion of estimating the parameter values at each simulation, this might result in misleading simulation results. Alternatively, quite an effort might be necessary in order to identify the source of the bias, and in searching for an appropriate remedy. I agree that the effect of estimation uncertainty on forecast evaluation is of practical interest, but shedding light on this question is not the aim of the paper.

Madrid, 9 July 2008
Genaro Sucarrat

References

Bauwens, L., D. Rime, and G. Sucarrat (2006). Exchange Rate Volatility and the Mixture of Distribution Hypothesis. *Empirical Economics 30*, 889–911.

Bauwens, L. and G. Sucarrat (2008). General to Specific Modelling of Exchange Rate Volatility: A Forecast Evaluation. Forthcoming in the *International Journal of Forecasting*. Working Paper version available as `http://www.eco.uc3m.es/sucarrat/research/gets.pdf`.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.

Jarque, C. and A. Bera (1980). Efficient Tests for Normality, Homoskedasticity, and Serial Independence of Regression Residuals. *Economics Letters 6*, 255–259.

Ljung, G. and G. Box (1979). On a Measure of Lack of Fit in Time Series Models. *Biometrika 66*, 265–270.

Mizon, G. (1995). Progressive Modeling of Macroeconomic Time Series: The LSE Methodology. In K. D. Hoover (Ed.), *Macroeconometrics. Developments, Tensions and Prospects*. Kluwer Academic Publishers.

Sucarrat, G. (2006). *Essays in the Study and Modelling of Exchange Rate Volatility*. Ph. D. thesis, Department of Economics, Université catolique de Louvain, Louvain la Neuve, Belgium. Thesis Number 508/2006. Available as `http://edoc.bib.ucl.ac.be:81/ETD-db/collection/available/BelnUcetd-09152006-141220/unrestricted/these_sucarrat.pdf`.

Sucarrat, G. (2007). Econometric Reduction Theory and Philosophy: The First Stage in Hendry's Reduction Theory Revisited. Available as `http://www.eco.uc3m.es/sucarrat/research/hendry.pdf`.

Sucarrat, G. (2008). Forecast Evaluation of Explanatory Models of Financial Variability. Available as `http://www.eco.uc3m.es/sucarrat/research/volamodeval.pdf`.

Table 1: Descriptive statistics of interdaily (close, weekends excluded) exchange rate returns in percent from 26 September 2005 to 4 January 2008 ($T = 594$)

	USD/EUR	YEN/EUR	GBP/EUR	NOK/EUR
$S.E.$	0.446	0.564	0.313	0.356
$Kurtosis$	3.741	4.918	3.571	3.941
JB	14.825	118.750	18.965	25.487
	[0.00]	[0.00]	[0.00]	[0.00]
$AR(1)$	0.422	4.003	1.984	1.886
	[0.52]	[0.05]	[0.16]	[0.17]
$ARCH(1)$	0.406	56.119	0.938	3.557
	[0.52]	[0.00]	[0.33]	[0.06]

Note: $S.E.$ is the standard error of returns, $Kurtosis$ is the sample estimate of kurtosis, JB is the Jarque and Bera (1980) test for non-normality, and $AR(1)$ and $ARCH(1)$ are Ljung and Box (1979) tests for first order serial correlation in returns and squared returns, respectively. Values in square parentheses are the p-values associated with the tests.

Table 2: Average Gaussian ML estimates of models (6), (7) and (8)

T		Model (6)	Model (7)	Model (8)
1 000	$\hat{b}$	0.4466		
	$\hat{\omega}$	-0.0145	0.0701	0.1719
	$\hat{\alpha}$	0.1198	0.0613	0.0673
	$\hat{\beta}$	0.7737	0.7161	0.6403
	$\hat{c}$	0.0947	0.1006	
5 000	$\hat{b}$	0.4469		
	$\hat{\omega}$	-0.0159	0.0046	0.1009
	$\hat{\alpha}$	0.1204	0.0611	0.0640
	$\hat{\beta}$	0.7836	0.8316	0.7678
	$\hat{c}$	0.0906	0.0991	
10 000	$\hat{b}$	0.4474		
	$\hat{\omega}$	-0.0162	0.0036	0.1057
	$\hat{\alpha}$	0.1191	0.0602	0.0623
	$\hat{\beta}$	0.7863	0.8355	0.7611
	$\hat{c}$	0.0902	0.0977	
20 000	$\hat{b}$	0.4474		
	$\hat{\omega}$	-0.0160	0.0031	0.0871
	$\hat{\alpha}$	0.1199	0.0607	0.0636
	$\hat{\beta}$	0.7860	0.8359	0.7914
	$\hat{c}$	0.0895	0.0978	

Note: The Gaussian ML estimates are computed in EViews 6 using the default options (the Marquardt algorithm,[a] convergence criterion equal to 0.0001) but without backcasting, where each series is generated by means of the simulation DGP given by (6) with the benchmark values $(b, \omega, \alpha, \beta, c, p) = (5^{-1/2}, 0.02, 0.1, 0.8, 2, 0.1)$. The number of simulations is 1000, each with sample size equal to T and a prior burn-in sample equal to 100 observations in order to avoid initial value issues.

[a] A modified Gauss-Newton algorithm.

Table 3: Rejection probabilities of the nulls of no AR(1) and no ARCH(1) in the standardised residuals of models (6), (7) and (8), using the parameter estimates for $T = 20\ 000$ in table 2 and a nominal value of 5%

T		Model (6)	Model (7)	Model (8)
1 000	$AR(1)$	0.051	0.079	0.053
	$ARCH(1)$	0.052	0.685	0.157
5 000	$AR(1)$	0.053	0.090	0.069
	$ARCH(1)$	0.064	1.000	0.453
10 000	$AR(1)$	0.046	0.077	0.052
	$ARCH(1)$	0.049	1.000	0.720
20 000	$AR(1)$	0.048	0.082	0.068
	$ARCH(1)$	0.054	1.000	0.938

Note: AR(1) and ARCH(1) stand for the Ljung and Box (1979) test (Q-stat.) for first order serial correlation in the standardised residuals and in the squared standardised residuals, respectively. Computations are in EViews 6.

Table 4: Rejection probabilities of the nulls of no AR(1) and no ARCH(1) in the standardised residuals of models (6), (7) and (8), using the "correct" parameter values and a nominal value of 5%

T		Model (6)	Model (7)	Model (8)
1 000	$AR(1)$	0.047	0.041	0.043
	$ARCH(1)$	0.045	0.040	0.055
5 000	$AR(1)$	0.045	0.052	0.056
	$ARCH(1)$	0.042	0.064	0.074
10 000	$AR(1)$	0.049	0.041	0.047
	$ARCH(1)$	0.042	0.086	0.129
20 000	$AR(1)$	0.049	0.061	0.060
	$ARCH(1)$	0.048	0.108	0.165

Note: AR(1) and ARCH(1) stand for the Ljung and Box (1979) test (Q-stat.) for first order serial correlation in the standardised residuals and in the squared standardised residuals, respectively. Computations are in EViews 6.