Model complexity and model performance*

Stanislav Anatolyev†
New Economic School

Andrey Shabalin
University of North Carolina

Abstract

Applied researchers routinely construct dynamic time-series models for describing evolution and making forecasts. We investigate the trade-off between the model complexity and model performance using real financial data in order to determine what a reasonable compromise is, by analyzing the dependence of various model performance measures on the presence or absence of certain features in the model. The model features include mean persistence, volatility clustering, leverage effects, time-varying risk premia, heavy tails and skewness. The types of data we consider are individual stock prices, stock market indices, and exchange rates. We represent the results as panel regressions of performance measures on dummy variables representing model features, and as frequencies of events that inclusion of a feature improves or worsens a particular performance measure. Some of the results are expectable, but some are quite surprising.

* **Key words**: Conditional mean, conditional variance, goodness of fit, out-of-sample forecasting. **JEL classification**: C51.

† Corresponding author. Address: Stanislav Anatolyev, New Economic School, Nakhimovsky Prospekt, 47, Moscow, 117418 Russia. E-mail: sanatoly@nes.ru.

1 Introduction

Applied researchers routinely construct dynamic time-series models for describing evolution and making forecasts of series of interest. In doing this, a researcher is guided, on the one hand, by a need to incorporate certain stylized features of data into the model, and on the other hand, by how successfully the model fits the data in- and out-of-sample. Some model classes are specifically designed to take into account such stylized facts. Examples are augmenting an ARCH equation by an asymmetric term to take into account the leverage effect, and modeling the standardized innovations as non-normal in order to capture actual fat-tailedness of data distributions. At the same time, a model can be deemed successful only if it passes a battery of diagnostic and other tests, and eventually is judged by how it fits the data, both in-sample and in forecasting exercises.

In this paper, we investigate this trade-off between the model complexity and model performance using real financial data. To this end, we run a battery of estimation and forecasting exercises on various types of data to answer questions like: for such and such model performance measure, what are the critical features that should be incorporated in the model? how quickly does the performance improve as such and such features are incorporated in the model? what is a reasonable compromise between the model complexity and its performance for such and such type of data?

We try to shed light on these issues by looking at the dependence of various model performance measures on the presence or absence of certain features in the model. We represent the results in several forms. First, we run panel regressions of these measures on dummy variables representing model features, also including fixed individual effects to control for heterogeneity of different series of a certain type of data. The coefficients in such panel regression, together with their significance, indicate which model features are more and which are less important for a particular performance measure. Second, we report percentages of cases where inclusion of a feature improves or worsens a particular performance measure. Last, we describe which models are most successful for some of performance measures.

The types of data we consider are individual stock prices, stock market indices, and exchange rates, popular in the empirical finance literature. We experiment with 20 individual stocks that are included in the S&P100 index, 12 stock indices from major world stock exchanges, and 9 major foreign currencies. The data are weekly, and extend for 20–30 years totaling to approximately 800–1,600 observations, the last third of which are used for forecasting exercises, and the other two thirds for estimation. We present the results separately for each data type. The model features we consider allow for mean persistence, volatility clustering, leverage effects, time-varying risk premia, heavy tails, and skewness.

The results overall indicate that the tendencies in the model complexity–performance trade-off are quite different across the three types of data, but are quite uniform across different series of the same type. Some patterns can be expected, but some are surprising. As can be expected, it is easier to pin down an appropriate model if a performance measure is an in-sample one than if it is an out-of-sample one. For most performance measures concerning the mean, the presence of an AR or MA terms in the model is a main determinant of performance, even though coefficients belonging to these terms are statistically insignificant. On the contrary, for most performance measures concerning the variance, the presence of a GARCH equation in the model, and sometimes the presence of a heavy-tailed conditional distribution, are main determinants.

The paper is organized as follows. Section 2 describes our technology of constructing dynamic models. The description of the data is given in Section 3. Section 4 lists criteria used for judgement about model performance. The results are reported and analyzed in Section 5.

2 Dynamic models

We consider models where each feature f from a set of F features to be described shortly is either present or not. Some of the features may be present only if other features are, hence there are fewer than 2^F dynamic models in total. Any model has the following structure:

$$y_t = \mu_t + e_t,$$

where μ_t is (roughly) the mean, and e_t is (roughly) the error term. By default, $\mu_t = \bar{\mu}$, a constant, and $e_t = \varepsilon_t$, a martingale difference (relative to past data) innovation. Let

$$\varepsilon_t = \sqrt{\sigma_t^2}\eta_t, \quad \eta_t \sim i.i.d.\ D\left(0, 1, \tau\right),$$

where σ_t^2 is the conditional variance, η_t is the standardized innovation, D is the conditional distribution, τ is the vector of parameters of this distribution other than the mean and variance. By default, $\sigma_t^2 = \omega$, a constant, and $D\left(0, 1, \tau\right)$ is standard normal so that there is no τ.

The set of model features f contains, feature names appearing in parentheses:

(AR) Autoregressive component in μ_t (may be in effect only when there is no feature MA). When this feature is present, $\mu_t = \bar{\mu} + \phi y_{t-1}$.

(MA) Moving average component in e_t (may be in effect only when there is no feature AR). When this feature is present, $e_t = \varepsilon_t - \theta\varepsilon_{t-1}$.

(GARCH) ARCH effect in ε_t. When this feature is present, $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$.

(GJR) Asymmetry in σ_t^2 (may be in effect only when there is feature GARCH). When this feature is present, σ_t^2 in addition contains $\gamma\varepsilon_{t-1}^2\mathbb{I}\left[\varepsilon_{t-1} < 0\right]$.

(ArchM) ARCH-in-mean term in μ_t (may be in effect only when there is feature GARCH). When this feature is present, μ_t in addition contains $\delta\sigma_t^2$.

(Stud) Conditional fat-tailedness of η_t. When this feature is present, $D(0,1,\tau)$ is Student with $\tau = \nu$, where ν is the number of degrees of freedom.

(Skew) Conditional skewness of η_t (may be in effect only when there is feature Stud). When this feature is present, $D(0,1,\tau)$ is skewed Student with $\tau = (\nu, \lambda)'$, where λ measures the degree of skewness.

Thus, in total there are $F = 7$ features resulting in 45 dynamic models. The features we have included are, in our view, most important and popular in building a dynamic model for financial data that would serve general purposes of fitting the data and making short-run forecasts. At the same time we have not included those features that would serve more specific purposes, such as long memory in mean and/or variance, which would be clearly important for long-run forecasting, or nonlinear mean models of a regime switching type, which would be more relevant for capturing business cycle phenomena in macroeconomic data. We also do not cross the line beyond which the model becomes overly complicated and not widespread at the moment, like models with time-varying conditional skewness or conditional kurtosis.

The choice of a particular way to represent each feature when there is a variety of choices is made in favor of a simple and empirically popular model. For example, we have chosen the variance form for the ARCH-in-mean term, as, for example, in Lanne and Saikkonen (2004), and from a few variations of introducing asymmetric effects into the news impact curve we have selected the Glosten, Jagannathan and Runkle (1993) form. Next, among the fat-tailed distributions we have preferred Student's t-distribution as introduced in the GARCH context by Bollerslev (1987). Finally, from a few available skewed distributions that imply fat tails we have selected the skewed Student distribution (see, for example, Hansen, 1994) which is simpler in use than, for example, Exponential Generalized Beta of the second kind as in Wang, Fawson, Barrett and McDonald (2001), or the z-distribution as in Lanne and Saikkonen (2004). The orders of autoregressive, moving average and GARCH structures are set to low values thanks to the use of weekly data.

The method of estimation of all models is the method of conditional maximum likelihood. For simpler models, this reduces to simpler procedures, like OLS for an autoregressive model without ARCH effects.

3 Data

We use three types of financial data that are typically fitted with dynamic models: stock market indices from developed markets, individual stock prices from the New York Stock Exchange, and exchange rates of currencies from industrialized countries versus the US dollar. The data are weekly (Wednesdays) and cover twenty to thirty years thus amounting to about 800–1,600 observations. Each of the three types of data is represented by about a dozen of series that are routinely used in empirical time series studies. The raw data are converted into the form of returns by taking log differences.

Individual stock returns are represented by 20 stocks that were included in the S&P100 index in 2001 and have been traded since 1971. The symbols of these stocks are: AA, AEP, DD, DIS, EK, GE, GM, HON, IBM, IP, JNJ, KO, MCD, MMM, MRK, PG, S, T, UTX, XOM, and the data are taken from Yahoo!® Finance at `http://finance.yahoo.com`. The sample period is 01.01.1971–12.31.2000, with data on 10.21.1987, 10.28.1987, 11.04.1987 removed to avoid the influence of the 1987 stock market crash. This totals to 1548 observations in each series.

Exchange rates returns are represented by 9 currencies: BEF, CAD, CHF, DEM, FRF, GBP, ITL, JPY, SEK. The data are taken from the University of British Columbia's Pacific© Exchange Rate Service at `http://fx.sauder.ubc.ca`. The sample period is

01.02.1974–12.27.2000, totaling to 1390 observations in each series. The beginning of the sample period is chosen so as not to overlap with periods of fixed exchange rates.

The data for stock market index returns are represented by 12 indices, not all of which come from different markets. These samples are more heterogeneous across series; see the description in Table 1. The data are taken from Economy.com FreeLunch® at `http://www.economy.com/freelunch` except that FTSE and DJI are taken from Yahoo!® Finance, and CRSP is taken from the University of Chicago Center for Research in Security Prices (CRSP®). For all stock return indices, data on 10.21.1987, 10.28.1987, 11.04.1987 were also removed.

4 Performance criteria

We judge the model performance by two sorts of criteria: in-sample and out-of-sample. The sample of y_t for $t = 1, \cdots, T$ is divided into the estimation and prediction parts, the former running from 1 to R, the latter running from $R+1$ to $R+P(\equiv T)$.

The in-sample criteria we employ are:

- Quality of fit criteria: the Ljung–Box statistics of order 10 applied to residuals $\hat{\varepsilon}_t$ (LB) and squared standardized residuals $\hat{\eta}_t^2$ (LB^2). Recall that the critical values for $\chi^2(10)$ variable are: 15.99, 18.31 and 23.21 for 10%, 5% and 1% significance levels, respectively (for the sake of uniformity across models, we do not make a degrees of freedom adjustment for parameter estimation).

- Remaining non-linearity: the BDS test statistic (Brock, Dechert, Scheinkman and LeBaron, 1996) applied to standardized residuals $\hat{\eta}_t$ with the additional parameter 5 (BDS). The details of constructing the BDS test are contained, for example, in Hsieh (1989). Recall that the BDS statistic is asymptotically normal under the null of IID series, and it is most natural to use BDS test as one-sided. Hence, the critical values are: 1.28, 1.64 and 2.33 for 10%, 5% and 1% significance levels, respectively.

- Model selection criteria: we report results of using the Bayesian (BIC) information criterion (Schwarz, 1978).

- Stability criteria: Nyblom statistics for stability of individual coefficients and for entire model (Nyblom, 1987; Hansen, 1994). Let $l_t(\theta)$ be the loglikelihood for one observation, $\hat{\theta}$ be the ML estimate, K – number of parameters in the model. Then

$$\begin{aligned} \text{ModNyb} &= \frac{1}{R}\sum_{t=1}^{R} S_t' \hat{V}^{-1} S_t, \\ \text{IndNyb} &= \max_{k=1,\dots,K} \left(\frac{1}{R}\sum_{t=1}^{R} \frac{S_{kt}^2}{\hat{V}_{kk}} \right), \end{aligned}$$

  where

$$\hat{V} = \frac{1}{R}\sum_{t=1}^{R} \frac{\partial l_t(\hat{\theta})}{\partial \theta}\frac{\partial l_t(\hat{\theta})'}{\partial \theta}, \quad S_t = \sum_{i=1}^{t} \frac{\partial l_i(\hat{\theta})}{\partial \theta}.$$

  Unfortunately, the critical values for these tests depend on the parameter dimension which makes formal comparison across models of different degree of parsimony difficult.

When making forecasts, we use parameter values estimated only once from the data from 1 to R. Let $\hat{y}_{t|t-1}$ denote a forecast of y_t made at $t-1$. The out-of-sample criteria we employ are:

- Mean forecasting criteria: the mean squared and mean absolute prediction errors

$$\begin{aligned} MSPE &= \frac{1}{P}\sum_{t=R+1}^{R+P} \left(y_t - \hat{y}_{t|t-1} \right)^2, \\ MAPE &= \frac{1}{P}\sum_{t=R+1}^{R+P} \left| y_t - \hat{y}_{t|t-1} \right|. \end{aligned}$$

- Sign forecasting criteria: the proportion of times the sign is correctly predicted

$$SIGN = \frac{1}{P} \sum_{t=R+1}^{R+P} 1\left[y_t \hat{y}_{t|t-1} > 0\right].$$

- Volatility forecasting criteria: the mean squared prediction error for volatility (we use the abbreviation VSPE so that it is easily distinguished from the MSPE). Let $\hat{y}_{t|t-1}$ denote a forecast of y_t made at $t-1$, and $\hat{\sigma}^2_{t|t-1}$ be a model-based volatility estimate. Then

$$VSPE = \frac{1}{P} \sum_{t=R+1}^{R+P} \left((y_t - \hat{y}_{t|t-1})^2 - \hat{\sigma}^2_{t|t-1} \right)^2.$$

5 Results

For each series, we run maximum likelihood estimation of all 45 models over the estimation period, construct (fixed scheme) forecasts over the forecasting period, and compute values of performance measures.

To sift out the tendencies in a huge amount of output information, we aggregate the results in the following way which is reminiscent of the response surface methodology where linear regressions are used to make extrapolations. For every single performance measure and type of data, we regress its values on dummy variables representing the seven model features, also including individual effects pertaining to different series. That is, we run linear panel data regressions with fixed "series effects", but we do that separately for indexes, individual stocks and exchange rates because of a much greater heterogeneity across data types than across series of the same type. The output of interest is composed of regression coefficients and their significance showing a sign and impact of each model feature. A reader should keep in mind, however, that such regressions are not strictly in line with the panel data analysis because the statistics are evaluated over the same samples or different samples of the same sample period, so the actual regression error structure is not as ideal as is assumed in error component models. In addition, we document percentages of cases where inclusion of a feature improves or worsens a particular performance measure; we call these percentages "progress direction".

The results of panel data regressions together with progress direction numbers are presented in Table 2, where we have boldfaced the coefficients whose t-ratios are greater than 1.6 in absolute value, and the percentages that are either do not exceed 15% or do not fall short of 85%. We regard the boldfaced numbers as deserving attention most, and corresponding features as most critical ones. Graphical illustrations for selected series are provides in Figures 1–6. The models in figures are arranged by worsening performance from top to bottom. Below, we comment on general tendencies revealed by the numerical results, for each of performance measure separately.

LB The principal factor that plays a role for the LB characteristic is the presence of mean filtering, either of autoregressive or moving average type, while other features cannot seriously influence LB values. Interestingly, the AR or MA parameters (ϕ or θ), if present, may not be significant even at 10%.

LB^2 Most radically the LB^2 characteristic is influenced by the presence of volatility filtering by GARCH, and is absolutely insensitive to the presence of other features. The presence of pure GARCH alone is able to reduce LB^2 values from significant at 1% to insignificant at 10%. The GARCH parameters (α and β) are highly significant (an exception sometimes occurs in the presence of GJR when γ pulls significance away from β).

BDS The BDS statistic is able to point at no neglected nonlinearity after volatility filtering by GARCH, and it is rather insensitive to the presence of other features. However, even in the absence of a GARCH part BDS may not help detect nonlinearity. The presence of mean filtering and allowance for conditional thick tails may sometimes improves the BDS statistic.

BIC According to BIC, more parsimonious models containing one to two, rarely three features, are best, provided they contain a GARCH equation, while models not containing GARCH specifications are evident outsiders. Another important factor that significantly improves fit is the Stud feature that allows for conditionally heavy tails. For individual stocks, most often the best model is GARCH with conditional Student distribution (7 times out of 20), next comes conditionally normal GARCH with leverage (3 times). The latter model is best most often for stock indices (3 times out of 12). For exchange rates, *almost always* the best model is GARCH with conditional Student distribution (8 times out of 9).

Nyblom An unambiguously positive impact on the stability criteria makes the presence of the GARCH equation in the model. Another important factors are the heavy-tailedness and asymmetry in the conditional distribution, as well as the asymmetry in the variance equation, but these features tend to worsen the stability indicators.

MSPE and MAPE Confirming common wisdom, there is a general tendency that more parsimonious models tend to predict the mean more successfully. Further patterns are not very clear and are different for different types of data. On the whole, the MSPE and MAPE criteria are in consensus most of the time, but the two do not tend to agree on which models are best and which are worst. For individual stocks, the presence of ArchM is harmful for predicting the mean, while the presence of AR or MA filters has a favorable impact in most cases, even though corresponding coefficients are rarely significant. For stock indices, the presence of a thick-tailed conditional distribution worsens mean prediction, but allowing for skewness acts in the opposite direction, and the net effect on mean prediction is favorable. For exchange rates, the decisive and favorable factor happens to be heavy-tailed conditional distribution, and, perhaps surprisingly, mean filtering tends to worsen mean predictability.

SIGN For individual stocks and stock indices, it is easy to exceed the coin toss sign prediction of 50% using a model that is not among best; with best models one can achieve 59% for some stocks (XOM) and 62% for indices (CRSP and NYA). In contrast, it is much harder to predict signs of exchange rate movements; even with best models one cannot exceed 50% appreciably. There is no clear-cut pattern of which features impact SIGN most, but mean filtering seems to have greatest effect, negative in case of stocks or indices, and positive in case of exchange rates.

VSPE In volatility predictions, the presence of GARCH is important, if not decisive. Strangely, however, that the GARCH factor has a favorable impact in case of individual stocks or stock indices, but an adverse impact in case of exchange rates.

Overall, we observe quite appreciable difference in performance of the same dynamic models when they are fit to exchange rates compared to when they are fit to stock returns and indexes. A possible explanation is that the behavior of exchange rates may not be temporally stable during long periods.

References

Bollerslev, T. (1987) A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* 69, 542–547.

Brock, W., D. Dechert, J. Scheinkman and B. LeBaron (1996) A test for independence based on the correlation dimension. *Econometric Reviews* 15, 197–235.

Glosten, L.R., R. Jagannathan and D.E. Runkle (1993) On the relation between the expected value and the volatility of the nominal excess terurn on stocks. *Journal of Finance* 48, 1779–1801.

Hansen, B. (1994) Autoregressive conditional density estimation. *International Economic Review* 35, 705–730.

Hsieh, D. (1989) Testing for nonlinear dependence in daily exchange rates. *Journal of Business* 62, 339–368.

Lanne, M. and P. Saikkonen (2004) A skewed GARCH-in-mean model: An application to U.S. stock returns. Manuscript, University of Jyväskylä.

Nyblom, J. (1989) Testing the constancy of parameters over time. *Journal of the American Statistical Association* 84, 223–230.

Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics* 6, 461–464.

Wang, K.-L., C. Fawson, C.B. Barrett, and J.B. McDonald (2001) A flexible parametric GARCH model with an application to exchange rates. *Journal of Applied Econometrics* 16, 521–536.

Series	Description	Sample period	Sample size
		Stock index returns	
DJI	Dow Jones 30 industrials index	01.04.1971–12.26.2000	1562
SPX	Standard and Poor's 500 index	01.06.1971–12.27.2000	1562
NASD	Nasdaq composite index	02.10.1971–12.27.2000	1557
NYA	NYSE composite index	01.06.1971–12.27.2000	1562
CRSP	CRSP value-weighted index	01.03.1968–12.31.1997	1525
FTSE	UK FTSE-100 index	04.02.1984–11.22.2004	1076
CAC	France CAC-40 index	12.30.1987–11.17.2004	882
DAX	Germany DAX index	11.28.1990–11.17.2004	730
NIKKEI	Japan Nikkei-225 index	01.04.1984–11.17.2004	1087
TSE	Canada TSE-300 composite index	08.15.1984–11.17.2004	1055
HSI	Hong Kong Hang Seng index	12.31.1986–11.17.2004	931
STI	Singapore Straits Times index	01.06.1988–11.17.2004	881

Table 1. Description of data on stock indices.

Note: returns on 10.21.1987, 10.28.1987, 11.04.1987 were removed.

Features	In-sample						Out-of-sample			
	LB	LB^2	BDS	BIC	ModNyb	IndNyb	MSPE	MAPE	SIGN	VSPE
Panel regression: individual stocks										
	$\times 10^{0}$	$\times 10^{2}$	$\times 10^{-1}$	$\times 10^{-2}$	$\times 10^{-1}$	$\times 10^{-1}$	$\times 10^{1}$	$\times 10^{-2}$	$\times 10^{-3}$	$\times 10^{5}$
Stud	-0.042	0.008	-0.132	**-1.613**	**2.591**	0.738	0.053	0.773	0.730	-1.048
Skew	-0.009	0.001	-0.010	0.448	**2.453**	0.594	-0.003	-0.465	-0.349	0.122
AR	**-3.501**	0.004	-0.128	0.373	1.343	0.159	-0.371	-2.455	-1.814	-0.520
MA	**-3.528**	0.003	-0.128	0.367	1.415	0.179	-0.416	-2.583	-2.421	-0.552
ArchM	-0.234	0.003	0.000	**0.537**	1.038	-0.203	**1.256**	**8.248**	-1.474	1.934
GARCH	0.143	**-1.087**	**-8.135**	**-3.057**	**-3.715**	**-7.715**	-0.147	-1.663	-0.253	**-4.190**
GJR	0.000	0.007	-0.025	0.163	**2.625**	0.570	0.278	2.790	0.032	0.117
Progress direction: individual stocks										
Stud	55%	55%	65%	**95%**	**5%**	30%	50%	50%	55%	65%
Skew	55%	35%	65%	**5%**	**0%**	35%	65%	70%	35%	50%
AR	80%	50%	75%	**15%**	**0%**	35%	80%	60%	40%	65%
MA	**85%**	50%	75%	**15%**	**0%**	35%	**85%**	60%	45%	60%
ArchM	55%	35%	50%	**0%**	20%	65%	35%	25%	30%	40%
GARCH	20%	**100%**	**100%**	**90%**	70%	**95%**	60%	65%	50%	**100%**
GJR	65%	50%	60%	30%	**5%**	35%	35%	30%	60%	50%
Panel regression: stock indices										
	$\times 10^{0}$	$\times 10^{2}$	$\times 10^{0}$	$\times 10^{-2}$	$\times 10^{-1}$	$\times 10^{-1}$	$\times 10^{0}$	$\times 10^{-2}$	$\times 10^{-3}$	$\times 10^{5}$
Stud	0.141	-0.010	-0.039	**-1.880**	**4.729**	**2.189**	2.314	0.133	-0.014	0.125
Skew	-0.039	0.000	0.010	-0.068	**2.090**	-0.024	-3.329	-2.184	0.759	-0.487
AR	**-4.939**	-0.003	-0.019	0.473	1.309	-0.075	0.186	2.267	**-5.080**	-0.885
MA	**-4.434**	-0.002	-0.018	0.501	1.321	-0.071	0.348	2.074	**-6.071**	-0.806
ArchM	0.024	0.002	0.002	0.780	1.519	0.060	3.000	2.946	-2.111	-0.351
GARCH	-0.060	**-1.265**	**-1.240**	**-6.104**	-0.694	**-4.685**	1.125	-0.584	2.044	**-3.496**
GJR	-0.131	-0.004	-0.042	0.105	**2.430**	-0.153	0.196	1.749	-2.314	-0.966
Progress direction: stock indices										
Stud	25%	50%	67%	75%	**8%**	17%	25%	67%	50%	25%
Skew	83%	58%	42%	50%	**8%**	67%	**92%**	**92%**	50%	75%
AR	75%	75%	75%	**8%**	**8%**	83%	58%	42%	50%	67%
MA	75%	75%	75%	**8%**	**8%**	83%	58%	58%	25%	58%
ArchM	67%	25%	58%	**0%**	17%	33%	42%	33%	42%	67%
GARCH	58%	**100%**	**100%**	**100%**	33%	**92%**	50%	50%	50%	75%
GJR	33%	58%	**92%**	50%	**0%**	42%	58%	50%	42%	**92%**
Panel regression: exchange rates										
	$\times 10^{0}$	$\times 10^{1}$	$\times 10^{0}$	$\times 10^{-2}$	$\times 10^{0}$	$\times 10^{0}$	$\times 10^{0}$	$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{4}$
Stud	-0.467	-0.499	**-0.311**	-7.334	**1.703**	**1.109**	**-1.491**	**-3.988**	0.296	2.416
Skew	0.045	0.005	0.004	0.456	0.335	0.043	0.268	1.641	-0.379	0.209
AR	**-4.167**	0.070	-0.033	0.251	0.278	0.047	**1.326**	1.325	**1.281**	0.071
MA	**-3.907**	0.063	-0.028	0.317	0.287	0.037	1.062	1.236	**1.110**	0.008
ArchM	0.069	-0.022	0.005	0.648	0.163	-0.008	1.009	1.896	0.048	0.155
GARCH	0.101	**-6.823**	**-1.934**	-7.235	**-0.958**	**-1.977**	0.588	1.582	-0.505	**3.945**
GJR	-0.097	-0.072	0.003	0.546	0.328	0.035	-0.149	-0.678	0.144	0.793
Progress direction: exchange rates										
Stud	67%	44%	**89%**	**100%**	**0%**	**0%**	**100%**	**100%**	78%	22%
Skew	**11%**	**11%**	67%	22%	**0%**	44%	22%	**11%**	22%	33%
AR	**89%**	**11%**	**100%**	**11%**	**11%**	22%	**11%**	33%	78%	33%
MA	**89%**	**11%**	**100%**	**11%**	**11%**	22%	**11%**	44%	78%	33%
ArchM	44%	78%	**11%**	**0%**	**0%**	78%	33%	33%	67%	33%
GARCH	22%	**100%**	**100%**	**100%**	67%	**100%**	**0%**	**0%**	44%	33%
GJR	67%	78%	33%	**0%**	**0%**	33%	67%	67%	78%	**11%**

Table 2. Results of panel regressions and progress direction numbers.