Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 597*

First Version, December 1997
Revised, March 1999

PITFALLS IN TESTS FOR CHANGES IN CORRELATIONS

Brian H. Boyer, Michael S. Gibson and Mico Loretan

PITFALLS IN TESTS FOR CHANGES IN CORRELATIONS

Brian H. Boyer, Michael S. Gibson and Mico Loretan[*]

Abstract: Correlations are crucial for pricing and hedging derivatives whose payoff depends on more than one asset. Typically, correlations computed separately for ordinary and stressful market conditions differ considerably, a pattern widely termed "correlation breakdown." As a result, risk managers worry that their hedges will be useless when they are most needed, namely during "stressful" market situations.

We show that such worries may not be justified since "correlation breakdowns" can easily be generated by data whose distribution is stationary and, in particular, whose correlation coefficient is constant. We make this point analytically, by way of several numerical examples, and via an empirical illustration.

But, risk managers should not necessarily relax. Although "correlation breakdown" can be an artifact of poor data analysis, other evidence suggests that correlations do in fact change over time.

Keywords: risk management, risk measurement, hedging, derivatives, correlation, conditional correlation, normal distribution, foreign exchange

1 Introduction

The correlations between financial time series are crucial to risk management and the pricing of portfolios of assets.[1] Of particular interest is whether the correlations are constant over time: Unstable correlations make it difficult (or impossible) to hedge exposure to one risk factor with an offsetting position in another asset. In discussions of the importance of correlation for risk management and hedging, one often encounters statements to the effect that one has to allow for "correlation breakdown," i.e., the empirical regularity that the correlations between the series differ between "quiet" (or "ordinary") periods and "hectic" (or "unusual") periods. To quote the global risk manager for a major securities trading firm (Bookstaber, 1997):

> During major market events, correlations change dramatically.

Correlation breakdowns, if they occur, call into question the usefulness of hedging operations based on correlations estimated from long time series of historical data, since they may be inaccurate precisely when they may be needed the most.

The purpose of our paper is to show that testing for changes in correlations is not as straightforward as one might think. Specifically, we demonstrate that splitting a sample of data according to the *ex post* realizations of a series, say between "large" and "small" values of one of the series, can yield very misleading results, because such a procedure is likely to suggest correlation breakdown regardless of whether the correlation coefficients have changed. We make this point analytically, by way of several numerical examples, and via an empirical illustration. Although it may not be obvious at first, our results are a direct consequence of selection bias, a phenomenon familiar to statisticians and econometricians.

In Section 2, we introduce some notation, provide an analytical expression for the conditional correlation between two independently and identically distributed (i.i.d.) bivariate normal random variables, and compute the value of the conditional correlations for a variety of methods of splitting a dataset into subperiods of interest. In each case, looking at the resulting conditional correlations would indicate "correlation breakdown" even though the data are, by construction, i.i.d. In Sec-

---

[1]See, for example, Wilson 1993; Sullivan, 1995; Campa and Chang, 1997.

tion 3, we provide several empirical illustrations of how "correlation breakdown" may be generated by innocuous-seeming conditioning on events of interest. We examine a simple bivariate dataset of daily changes in German and Japanese exchange rates (*vs.* the U.S. dollar). We find that the patterns of conditional correlations calculated from this dataset are very similar to what we would find if the data were known to be i.i.d. bivariate normal. Our empirical illustrations confirm that it would be improper to conclude that the (population) correlation between two series varies across observations based on sample-splitting exercises alone. In Section 4, we discuss what our findings imply for proper testing for changes in correlations.

2 The relationship between conditional and unconditional correlations

To test whether the correlation between two time series is constant or is changing over time, one could consider comparing sampling correlations between the two series calculated from subsets of the data.[2] If these *conditional* correlations are found to be statistically different from each other, one might be tempted to conclude that the population correlation is not constant. In this section we demonstrate both analytically and numerically that this intuitively attractive approach to testing for correlation breakdowns can be very misleading.

2.1 Analytical derivation for bivariate normal random variables

We begin by considering a pair of bivariate normal random variables x and y with (unconditional) correlation coefficient ρ. We are interested in studying the effect that various forms of conditioning events—placing restrictions on the support of the distribution of (x, y)—have on the correlation between x and y.

Empirical practice frequently proceeds by restricting only one of the two variables. Events of interest in this paper are therefore mostly of the form "$x \in \mathcal{A}$", where $\mathcal{A} \subset \mathbb{R}$.[3] When there is

---

[2]The focus on correlations and hence on linear dependence is entirely appropriate when the joint distribution of the data is multivariate normal or, more generally, multivariate elliptic. In empirical practice, the joint distributions of many asset price changes are frequently found to be fairly close to being multivariate elliptic.

[3]We do not treat events of the type "$x \in \mathcal{A} \subset \mathbb{R}$ *and* $y \in \mathcal{B} \subset \mathbb{R}$" explicitly in the present paper. Our aim is to demonstrate the ways in which (intentional or inadvertent) conditioning on events may affect the correlation between two variables, rather than provide a general analysis of all types of conditioning. Explicit formulæ are straightforward to derive along the lines suggested by Theorem 1 in this section.

no chance of confusion, we shall denote the events simply by "$\mathcal{A}$". We only consider events of nontrivial ($\neq 0, 1$) probability.

The following theorem, which is proven in Appendix A, states the relationship between conditional and unconditional correlations for bivariate normal random variables when conditioning places restrictions on one of the two variables:

Theorem 1. *Consider a pair of bivariate normal random variables x and y with variances σ_x^2 and σ_y^2, respectively, and covariance σ_{xy}. Put $\rho = \sigma_{xy}/(\sigma_x \sigma_y)$, the unconditional correlation between x and y. Consider any event $x \in \mathcal{A}$, where $\mathcal{A} \subset \mathbb{R}$ such that $0 < \Pr(\mathcal{A}) < 1$. The conditional correlation $\rho_{\mathcal{A}}$ between x and y, conditional on the event $x \in \mathcal{A}$, is equal to*

$$\rho_{\mathcal{A}} = \rho \left(\rho^2 + (1 - \rho^2) \frac{\mathrm{Var}(x)}{\mathrm{Var}(\, x \mid x \in \mathcal{A} \,)} \right)^{-1/2}. \tag{1}$$

Inspecting equation (1), we make the following observations:

1. $\mathrm{sign}(\rho_{\mathcal{A}}) = \mathrm{sign}(\rho)$. Conditioning, by itself, does not affect the sign of the correlation coefficient.

2. $\rho_{\mathcal{A}} = \rho$ if $\rho = 0$, i.e., if x and y are independent, or if $|\rho| = 1$, i.e., if the bivariate distribution is degenerate (a case which we do not consider further).

3. For $\rho \neq 0$, $|\rho_{\mathcal{A}}| \gtrless |\rho|$ if $\mathrm{Var}(\, x \mid x \in \mathcal{A} \,) \gtrless \mathrm{Var}(x)$. This is the result that is of primary interest to us. The dependence of $\rho_{\mathcal{A}}$ on the ratio $\mathrm{Var}(\, x \mid \mathcal{A} \,)/\mathrm{Var}(x)$ is illustrated graphically in Figure 1 for several values of ρ.

In practice, the parameters ρ and $\mathrm{Var}(x)$, whether conditional or unconditional, are usually estimated from time series data. In keeping with common usage, one would equate the full-sample estimates of moments with the corresponding "unconditional" moments, and sub-sample estimates—where the subsamples are selected based on the conditioning criterion—with the "conditional" moments.

To extend the logic of Theorem 1 to the case of time series observations, consider the bivariate time series $\{x_t, y_t\}$, $t = 1, 2, \ldots, n$, which has support $\mathbb{R}^2 \times \mathbb{R}^2 \times \cdots \times \mathbb{R}^2$ (n copies) or, equivalently,

3

$\mathbb{R}^n \times \mathbb{R}^n$. Formally, conditioning consists of restricting attention to those observations for which the x_t's fall into a subset of $\mathbb{R}^n$, i.e., to require $(x_1, x_2, \ldots, x_n)' \in \mathcal{A}$, where $\mathcal{A} \subset \mathbb{R}^n$. As an example of a sample-based conditioning criterion, we could study the correlation between those x_t's and y_t's for which x_t falls into the first or fourth quartile of the sampling distribution of x.

When the random variables x_t and y_t are i.i.d. bivariate normal with contemporanous correlation coefficient ρ, equation (1) holds exactly for conditioning events $\mathcal{A}$ defined over $(x_1, x_2, \ldots, x_n)'$. If the sequence $\{x_t, y_t\}$, $t = 1, 2, \ldots, n$, is not i.i.d., but is assumed to satisfy certain stationarity and weak dependence conditions, a more general version of equation (1) would apply. Both statements are proven in Appendix C.

Figure 1: Dependence of $\rho_{\mathcal{A}}$ on $\mathrm{Var}(x \mid \mathcal{A}) / \mathrm{Var}(x)$

4
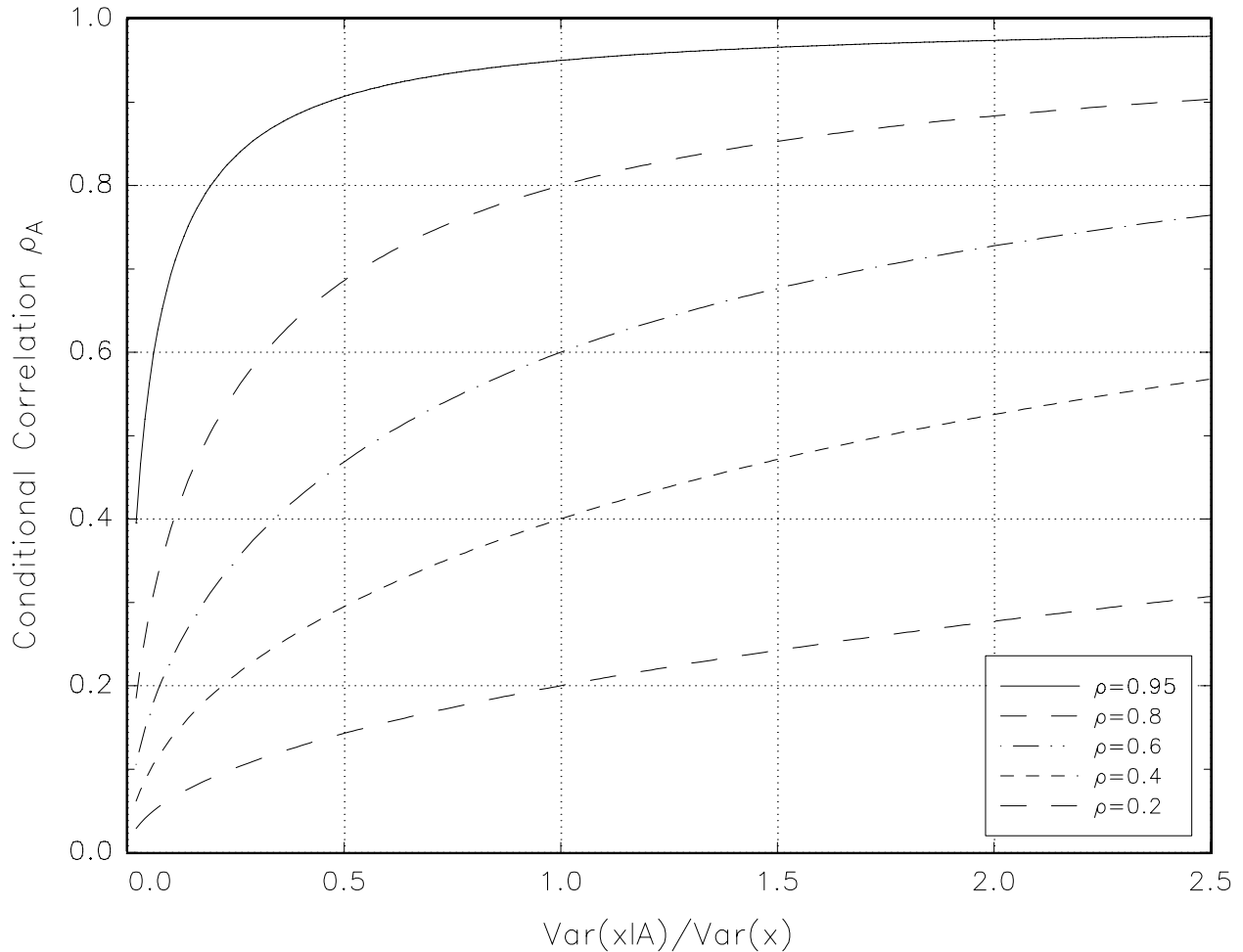
2.2 Numerically calculated conditional correlations

From the preceding discussion, we note that knowledge of $\text{Var}(\,x\mid\mathcal{A}\,)$ lets us determine whether $\rho_{\mathcal{A}}$ is less than or greater than ρ. In Appendix B, we show how one may compute the conditional variance of a normally distributed random variable x for various types of conditioning events of the form $x \in \mathcal{A} \subset \mathbb{R}$. We now provide three numerical illustrations of how much conditional correlations can differ from their unconditional counterparts.

As a first illustration of the dependence of the conditional correlation on the nature of the conditioning event, let x and y be i.i.d. mean-zero, unit-variance normal random variables. The only free parameter is ρ, the contemporaneous correlation coefficient. Let $D_1, D_2, \ldots, D_{10}$ represent the deciles of the (marginal) distribution of x. The conditioning events $\mathcal{A}$, in this case, are of the form "$x \in D_i$" for the deciles i, $i = 1, \ldots, 10$. We consider two values of ρ, one moderate ($\rho = 0.50$) and one high ($\rho = 0.95$).

We should expect the variance of $(\,x\mid x \in D_i\,)$ to be larger for those deciles that fall into the tails of the distribution, simply because the tail deciles are wider than the central deciles. Therefore, by inspection of equation (1), we would also expect the conditional correlation between x and y to be higher when x is in the tail of its distribution, irrespective of the value of ρ. These expectations are confirmed by Table 1, where the conditional variance of x and the conditional correlation between x and y are given for each of the ten deciles.[4] As is shown in the center column of the table, the conditional variance of x strongly depends on the chosen decile: $\text{Var}(\,x\mid x \in D_1\,)$ exceeds $\text{Var}(\,x\mid x \in D_5\,)$ by a factor of more than 30. As a result, the relationship between deciles and conditional correlations is distinctly "U-shaped"—for both values of ρ we consider—with the conditional correlations being largest in D_1 and D_{10}. Clearly, then, a U-shaped pattern need not indicate correlation breakdown, but may instead merely be a consequence of the *ex post* partitioning of the data—in this case into deciles.

The preceding illustration studied events which consist of single intervals of the data. In practice, we are often interested in *two-sided* events, such as "x is more than (less than) two/three/four standard deviations away from its mean." In Table 2, we present the relationship between two-sided

---

[4]$\text{Var}(\,x\mid x \in D_i\,)$ is calculated with equation (B.2); $\text{Corr}(x,y\mid x \in D_i)$ is calculated with equation (1).

Table 1: Conditional variances and correlations, decile delimited events, bivariate normal random variables

x and y are bivariate normal with zero mean and unit variance, and unconditional correlation ρ.

Decile	Interval	$\mathrm{Var}(\,x \mid x \in D_i\,)$	$\mathrm{Corr}(\,x, y \mid x \in D_i\,)$	
			$\rho = 0.50$	$\rho = 0.95$
1	$-\infty < x \le -1.282$	.169	.231	.781
2	$-1.282 < x \le -.842$	.0159	.0725	.358
3	$-.842 < x \le -.524$	.00834	.0526	.268
4	$-.524 < x \le -.253$	.00610	.0451	.231
5	$-.253 < x \le 0.00$	.00534	.0421	.217
6	$0.00 < x \le .253$	.00534	.0421	.217
7	$.253 < x \le .524$	.00610	.0451	.231
8	$.524 < x \le .842$	.00834	.0526	.268
9	$.842 < x \le 1.282$	.0159	.0725	.358
10	$1.282 < x < \infty$	.169	.231	.781

tail events $\mathcal{A}$ (and their complements, $\mathcal{A}^c$) and the resulting conditional correlation coefficients, again for random variables that are i.i.d. bivariate normally distributed. We consider four values for the unconditional correlation (.20, .50, .80, and .95) and two-sided events with probabilities of 50%, 10%, 5%, and 1%. (Each side has one half of the total probability of the event.) The case $\mathrm{Pr}(\mathcal{A}) = 50\%$ signifies that the event of interest consists of x falling into either the lowermost or uppermost quartile of its distribution. This corresponds to the case where we split the sample into subsamples of equal size, according to whether x is far away from its median or not, and wish to test whether the subsample correlations differ from each other. Cases of $\mathrm{Pr}(\mathcal{A}) = 10\%$ or less serve to compare the correlation between "tail" observations of the data with the correlation among "ordinary" observations.

For two-sided events we find $|\rho_{\mathcal{A}}| > |\rho| > |\rho_{\mathcal{A}^c}|$, since $\mathrm{Var}(\,x \mid \mathcal{A}\,) > \mathrm{Var}(x) > \mathrm{Var}(\,x \mid \mathcal{A}^c\,)$. We note that the pairs of conditional correlations are often far apart from each other, especially when the (population) correlation coefficient is relatively small. E.g., if $\rho = 0.5$ and $\mathrm{Pr}(\mathcal{A}) = 10\%$, $\rho_{\mathcal{A}} = 0.771$ whereas $\rho_{\mathcal{A}^c} = 0.415$—a difference of close to 100%! Surely an unsuspecting analyst might feel tempted to conclude that this discrepancy typifies a clear instance of "correlation breakdown"

6

Table 2: Conditional correlations, two-sided events, bivariate normal random variables

x and y are bivariate normal with zero mean and unit variance. Events $\mathcal{A}$ are two-sided "tail" events of the marginal distribution of x.

| | Two-Sided Event Probabilities | | | | | | | |
| | 50% | | 10% | | 5% | | 1% | |
$\mathrm{Corr}(x,y)$	ρ_A	ρ_{A^c}	ρ_A	ρ_{A^c}	ρ_A	ρ_{A^c}	ρ_A	ρ_{A^c}
.20	.268	.077	.393	.159	.434	.175	.510	.193
.50	.618	.213	.771	.415	.806	.449	.859	.485
.80	.876	.450	.942	.725	.953	.758	.968	.789
.95	.972	.754	.988	.923	.990	.936	.994	.946

between ordinary (quiet) and unusual (hectic) data observations. But, once more, the differences between the conditional correlations are caused by the choice of subsamples alone and not by any change in the parameters of the data generating process.

A third form of conditioning is to look at subsamples of the data that are characterized by "high volatility." Let a sample of n draws from (x, y) be divided into n_m equally sized subsamples ("months"; $m = 1, 2, \ldots, n_m$). The subset of "high-volatility months" can be defined as the set of months in which the ratio of the (within-month) variance of x to the overall (i.e., unconditional) variance of x exceeds some threshold k $(k \geq 1)$:

$$\mathrm{HVM} = \{m : \frac{\mathrm{Var}(\,x_t \mid t \in \text{month } m\,)}{\mathrm{Var}(x_t)} \geq k, \ m = 1, \ldots, n_m\}$$

Using this definition, the conditioning event $\mathcal{A}$ can be defined as

$$\mathcal{A} = \{(x_t, y_t) : t \in \text{month } m, \ m \in \mathrm{HVM}\}$$

Since the conditional variance of x is chosen directly in this case, the application of equation (1) is direct when x_t and y_t are assumed to be i.i.d. bivariate normal random variables. Conditioning on

Table 3: Effects of conditioning on "high volatility months"

x_t and y_t are i.i.d. bivariate normal with means equal to zero, variances equal to unity, and correlation 0.5.

"High volatility months" are defined as months where $\text{Var}(x)$ within the month is greater than or equal to a threshold k. A month is assumed to have 20 observations.

Variance threshold k	Fraction of months identifed as "high volatility months"	$\text{Corr}(x_t, y_t \mid t \in$ "high volatility month")
1.8	.01	0.615
1.7	.02	0.610
1.6	.04	0.603
1.5	.06	0.595
1.4	.10	0.585
1.3	.16	0.575
1.2	.24	0.564
1.1	.33	0.553
1.0	.45	0.541

"high volatility months" will cause the conditional correlation ρ_A to be greater in absolute value than the unconditional correlation ρ.

Table 3 illustrates the relationship between the conditional correlation and conditioning events of this type. It assumes that x_t and y_t are distributed i.i.d. bivariate normal with zero means, variances equal to unity, and correlation $\rho = 0.5$. We further assume each "month" has 20 observations. Because the x's are independent standard normal variables, the within-month conditional variance of x, which is a function of a sum of terms involving x^2, is distributed proportional to a χ^2 random variable with 20 degrees of freedom.

The first column of the table tabulates different values of the threshold k. The second column uses the cumulative distribution function for a $\chi^2(20)$ random variable to predict the fraction of months that will have a conditional variance greater than a certain threshold k. The third column

shows what the conditional correlation between x and y is for those months.[5] Again the pattern emerges that conditioning events characterized by high volatility of x are associated with high conditional correlation between x and y, even though the underlying data are i.i.d.

2.3 Conditional correlations for data that are not multivariate normal

The central argument of our paper is that computing correlations conditional on realizations of one variable, and observing that those correlations are different for different conditioning events, gives you no basis to conclude that the "true" correlation of the data-generating process is changing over time. Theorem 1 and Tables 1, 2, and 3 each illustrate this point. They are derived for the case of i.i.d. normally distributed random variables because the multivariate normal distribution is analytically tractable. However, Appendix C shows that our argument can be extended to treat the case of non-i.i.d. random variables, although no simple analytic expressions are available.

To illustrate this point, we reproduced the numerical illustrations in Tables 1, 2, and 3 for the case of x_t and y_t distributed as bivariate GARCH(1,1) with constant contemporaneous correlation, a model introduced by Bollerslev (1990). GARCH random variables are neither normal nor identically distributed, as their variance changes over time. However, like the case of i.i.d. bivariate normal random variables we used above, the GARCH model of Bollerslev (1990) does feature a constant contemporaneous correlation coefficient.[6]

The results are omitted to save space and because they are nearly identical to the results in Tables 1, 2, and 3.[7] "Correlation breakdown" is still observed when conditioning on deciles, two-sided events, or high variance months, even when the true data generating process has constant contemporaneous correlation and time-varying volatility.

---

[5]The numbers in the third column are

$$\int_{z \geq k} \rho_{\mathcal{A}}(z)\, \mathrm{d}G(z),$$

where $\rho_{\mathcal{A}}(z)$ is from equation (1) with $\mathrm{Var}(x \mid \mathcal{A}) = z$ and $G(z)$ is the cumulative distribution function of $V/20$, where V is a $\chi^2(20)$ random variable.

[6]Because no analytic formula is available for the conditional variance of a bivariate GARCH(1,1) process—the $\Sigma_{xx|\mathcal{A}}$ from Appendix C—the results for GARCH random variables were produced using Monte Carlo simulations of GARCH(1,1) data with constant contemporaneous correlation ρ.

[7]The tables are available from the authors on request.

3 Empirical illustrations

We now present conditional correlations for the three conditioning events discussed above using actual financial time series. The dataset comprises daily log changes in the German and Japanese spot exchange rates versus the U.S. dollar, from January 2, 1991 through December 31, 1998 (henceforth, German and Japanese exchange rates).[8] The full sample correlation between the two series is 0.504. A scatterplot of the data is provided in Figure 2. The data points are clustered around an upward-sloping line. However, the clustering appears to be slightly less tight in the tails of the data than in the central portion. It is therefore of interest to examine how the correlations differ between the subsets of the data.

The empirical values for the conditional correlations between the two series, by empirical decile, are presented in Table 4. The *theoretical* conditional correlations that would apply if the data were drawn from an i.i.d. bivariate normal distribution can be seen in Table 1, in the column labelled "$\rho = 0.50$." Rather than reproduce these numbers in Table 4, we provide a 90% confidence interval for the theoretical conditional correlation.[9] We observe that the empirical and theoretical conditional correlations follow virtually the same U-shaped pattern. The empirical conditional correlation is outside the 90% confidence interval for the theoretical conditional correlation only once, in decile 3. Hence, the U-shaped pattern of correlations present in the data cannot be used, by itself, to determine whether actual correlations differ across hectic and quiet subperiods.

Conditional correlation coefficients for two-sided tail events ("hectic periods") were also calculated for the exchange rate data and are presented in Table 5. Again, the theoretical conditional correlations which would apply if the data were i.i.d. and bivariate normal are nearly the same as those shown in Table 2 in the row "$\rho = 0.5$," so we again show 90% confidence intervals for the theoretical conditional correlations under the assumption of bivariate normality.

We observe, yet again, that the empirical and theoretical conditional correlations are quite similar. The empirical conditional correlations are never outside the 90% confidence interval for normally distributed data. We can therefore not conclude that the true correlation between the

---

[8]The exchange rates were collected by the Federal Reserve Bank of New York, at noon of each U.S. business day.

[9]All confidence intervals in this section were generated with a Monte Carlo simulation on simulated bivariate normal data with $\rho = 0.504$ and sample size $n = 2,000$.

Figure 2: German and Japanese exchange rates vs. U.S. dollar: Scatterplot of daily log changes × 100, January 1991–December 1998
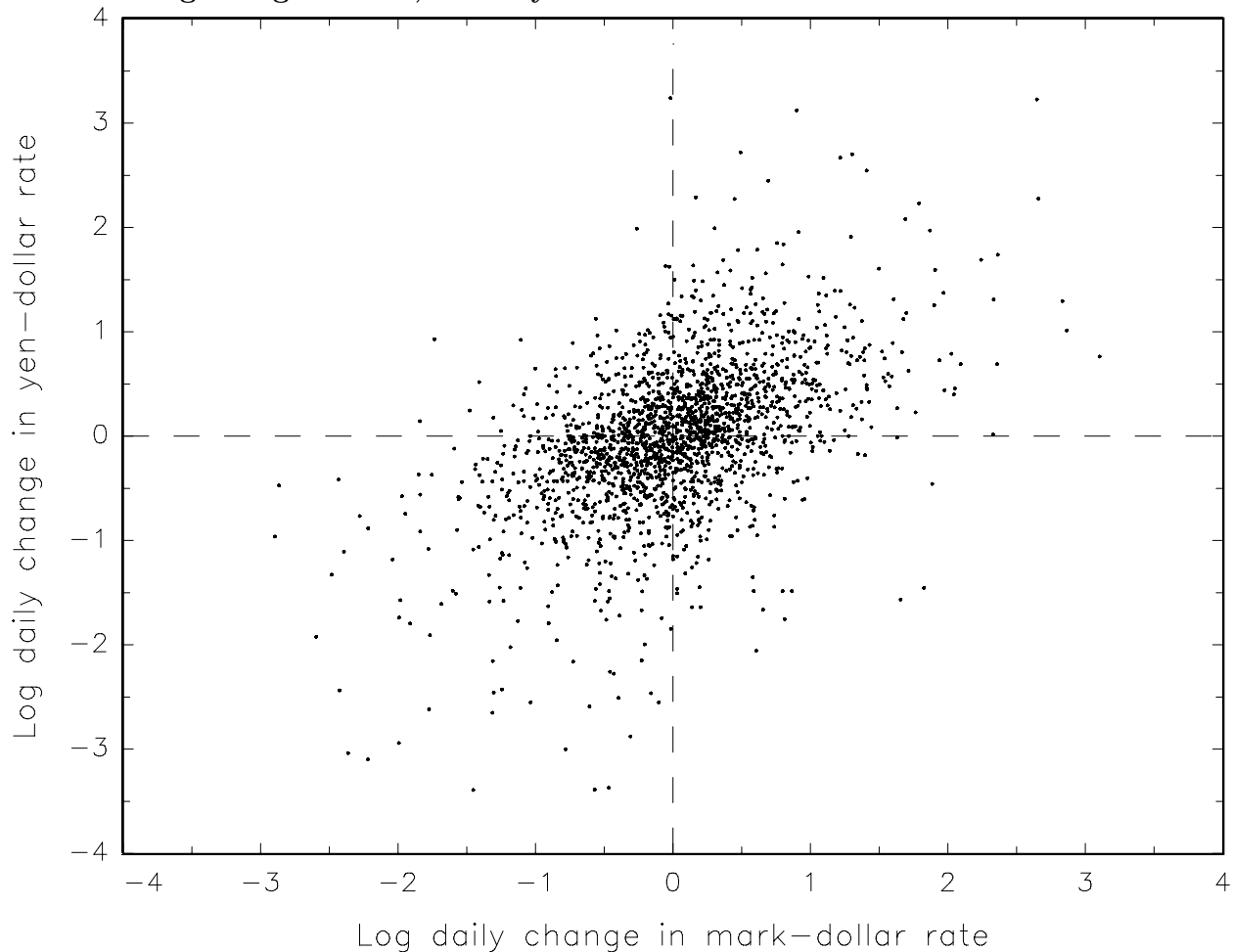
two empirical series is different in the "tails" of the distribution. Instead, we should conclude that the question of correlation breakdown for these series cannot be decided on the basis of this sample-splitting exercise.

Our third illustration is to compute the empirical correlation between the two exchange rates conditional on being in a month where the dollar-mark exchange rate exhibited "high volatility." Our data sample has 96 months. A scatterplot of the ratio of the within-month variance of daily dollar-mark returns (on the horizontal axis) against the within-month correlation between dollar-mark and dollar-yen returns (on the vertical axis) is presented in Figure 3. In addition to the scatterplot of empirical data, a curve representing the theoretical conditional correlation under the

Table 4: Empirical and theoretical conditional correlations, decile delimited events, exchange rate data

Data: Daily log changes of German (x) and Japanese (y) exchange rates relative to U.S. dollar, scaled by 100%. Sample period: January 2, 1991 to December 31, 1998. Full-sample correlation $\rho = 0.504$.

| Decile D_i of x	Range of decile D_i	Empirical Cond. Corr.	90% Conf. Interval for Theoretical Cond. Corr.
1	$-2.896\% < x \leq -.782\%$	.322	$(.117, .342)$
2	$-.782\% < x \leq -.482\%$	$-.026$	$(-.040, .192)$
3	$-.482\% < x \leq -.285\%$	.198	$(-.064, .168)$
4	$-.285\% < x \leq -.133\%$	.067	$(-.071, .159)$
5	$-.133\% < x \leq .006\%$	.038	$(-.076, .158)$
6	$.006\% < x \leq .150\%$	$-.009$	$(-.076, .158)$
7	$.150\% < x \leq .286\%$	.050	$(-.071, .159)$
8	$.286\% < x \leq .485\%$	$-.026$	$(-.064, .168)$
9	$.485\% < x \leq .797\%$	.040	$(-.040, .192)$
10	$.797\% < x < 3.103\%$	.285	$(.117, .342)$

Table 5: Empirical and theoretical conditional correlations, two-sided events, exchange rate data

Data: Daily log changes of German (x) and Japanese (y) exchange rates relative to U.S. dollar. Sample period: January 2, 1991 to December 31, 1998. Full sample correlation $\rho = 0.504$. Events $\mathcal{A}$ are two-sided "tail" events of the marginal distribution of x.

	Conditional Correlations							
	Two-Sided Event Probabilities							
	50%		10%		5%		1%	
	$\rho_{\mathcal{A}}$	$\rho_{\mathcal{A}^c}$	$\rho_{\mathcal{A}}$	$\rho_{\mathcal{A}^c}$	$\rho_{\mathcal{A}}$	$\rho_{\mathcal{A}^c}$	$\rho_{\mathcal{A}}$	$\rho_{\mathcal{A}^c}$
Empirical conditional correlation	.589	.224	.726	.386	.746	.448	.837	.477
90% conf. interval for theoretical cond. corr.	(.580, .665)	(.143, .286)	(.714, .829)	(.372, .464)	(.739, .871)	(.410, .497)	(.734, .954)	(.450, .530)

Figure 3: Relation between conditional variance and correlation, within-month data, daily-frequency German and Japanese exchange rate returns

Scatterplot of within-month variance of German exchange rate ($\mathrm{Var}(\,x \mid x \in \mathcal{A})$) against within-month correlation between German and Japanese exchange rate changes vs. the U.S. dollar (ρ_{month}), January 1991–December 1998 (96 months). Solid curve: theoretical conditional correlation as a function of conditional variance of x. Error bars: 90%-confidence intervals for sample size of 21.
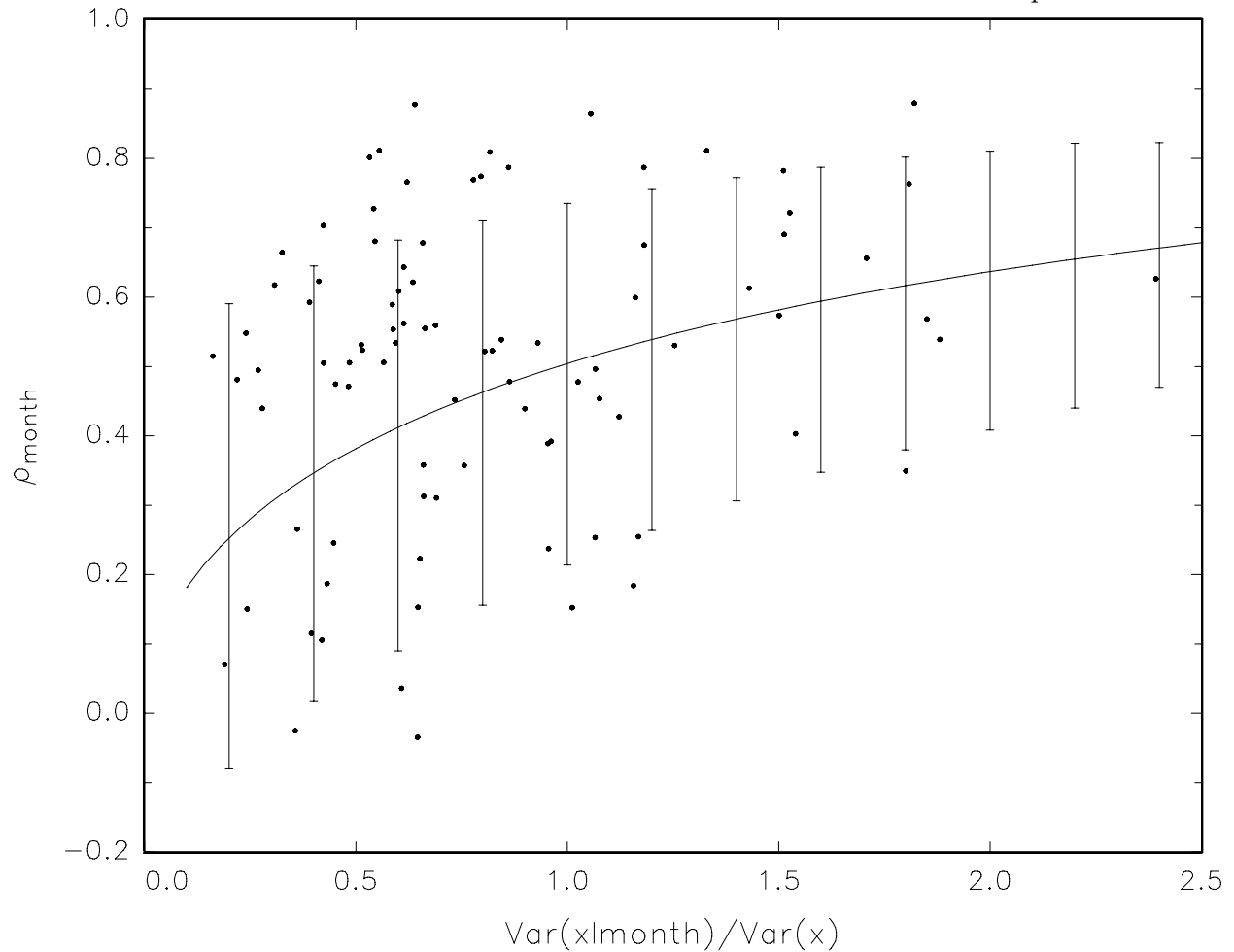
assumption of bivariate normality is plotted along with pointwise 90% confidence intervals (the vertical bars). The figure shows that the empirical within-month correlations are positively related to the within-month variance, as they would be if the data generating process were bivariate normal. Comparing the correlation coefficients in "high volatility" months and "low volatility" months could

13

Table 6: Empirical and theoretical conditional correlations, "high volatility month" events, exchange rate data

Data: Daily log changes of German (x) and Japanese (y) exchange rates relative to U.S. dollar. Sample period: January 2, 1991 to December 31, 1998. Full-sample correlation $\rho = 0.504$. "High volatility months" are defined as months where $\mathrm{Var}(\,x\mid\text{month}\,)/\mathrm{Var}(x)$ is greater than or equal to a threshold k.

Variance threshold k	Fraction of months identifed as "high volatility months"	Corr($x_t, y_t \mid t \in$ "high volatility month")	
		Empirical	Theoretical 90% confidence interval
1.8	.13	.662	(.425,.791)
1.7	.14	.660	(.427,.777)
1.6	.14	.660	(.442,.749)
1.5	.19	.639	(.467,.717)
1.4	.20	.638	(.485,.683)
1.3	.21	.644	(.496,.658)
1.2	.22	.642	(.498,.635)
1.1	.28	.622	(.498,.616)
1.0	.34	.562	(.494,.601)

give the illusion of "correlation breakdown," but again the same pattern would emerge if the data generating process were i.i.d.[10]

Table 6 shows how the empirical conditional correlation varies when conditioning on various definitions of a "high volatility month."[11] The table also shows the 90% confidence interval for the theoretical conditional correlations under bivariate normality. Evidence of fat tails in the distribution of the dollar-mark exchange rate is clear in the second column, which shows a wider spread of the distribution of within-month variance in Table 6 compared with the statistics in Table 3 based on the bivariate normal distribution. Still, the theoretical and empirical conditional correlations have the same basic pattern: highest when using a more extreme cutoff for the definition of a "high volatility month" and declining as the cutoff is reduced. For the thresholds $k = 1.2$

---

[10]The figure does show that the empirical data depart from bivariate normality. There is a clustering of points outside the 90% confidence interval in the "northwest" corner of the figure. This clustering would be unlikely to occur if the data were truly bivariate normal.

[11]Note that Table 6's conditioning events are of the form $\mathrm{Var}(\,x\mid\text{month}\,)/\mathrm{Var}(x) \geq k$ while Figure 3's are of the form $\mathrm{Var}(\,x\mid\text{month}\,)/\mathrm{Var}(x) = k$.

and $k = 1.1$, the empirical conditional correlation is outside the 90% confidence interval for bivariate normal data. For the remaining thresholds, it is not. Once again, although the correlation in "high volatility months" is greater than the unconditional correlation, that is no basis to conclude that the data generating process exhibits non-constant correlation.

4 Conclusion

We have shown that changes in correlations over time or across "regimes" cannot be detected reliably by splitting a sample according to the realized values of the data. The lack of reliability stems from the finding that correlation breakdowns will be "uncovered" by this method, irrespective of the actual stationarity properties of the data. This result is a direct consequence of the (implicit) selection bias that occurs when a sample is split according to the realized or observed values alone.

What valid alternative methods exist to detect changing correlations? In order to carry out a valid test, we argue that it is necessary that the researcher begin with a data-coherent model of the data generating process that builds in the possibility of structural changes, estimate the model's parameters, and only then decide whether the estimated parameters imply changing correlations (and possibly other structural breaks). For example, if the data were generated according to a Markov regime switching model with separate parameters for "quiet" and "hectic" time periods, one could estimate the model's parameters and then test whether the estimated correlations differ significantly between regimes. We are undertaking further research along these lines. For other valid approaches to testing the constancy of correlations, see Bera and Kim (1996), Karolyi and Stulz (1996), and Longin and Solnik (1995).

We caution that, in empirical practice, one must guard against subtle influences of data mining: The choice of model to represent the data generating process must be based on considerations that go beyond prior knowledge as to which model may "fit" the data best. Relying on such knowledge may reintroduce the problem of splitting the data by *ex post* criteria, and hence possibly invalidate the formal test of constancy of correlations across regimes.

Appendix A. Proof of Theorem 1

The proof relies on the well known property of bivariate normal random variables that each element may be expressed as a weighted sum of the other element and an independent component which is also normally distributed.[12]

Let u and v be two independent standard normal random variables. To create two random variables x and y such that x and y are distributed bivariate normal with means μ_x and μ_y and variances σ_x^2 and σ_y^2, respectively, and correlation coefficient $\rho = \sigma_{xy}/(\sigma_x \sigma_y)$, $|\rho| \leq 1$, the following operations can be performed on u and v:

$$x = \mu_x + \sigma_x \, u \tag{A.1}$$

$$y = \mu_y + \rho \sigma_y \, u + \sqrt{1 - \rho^2} \, \sigma_y \, v$$

$$= \mu_y + (\rho \sigma_y / \sigma_x)(x - \mu_x) + \sqrt{1 - \rho^2} \, \sigma_y \, v \tag{A.2}$$

Without loss of generality for what follows, we may assume that $\mu_x = 0$ and $\mu_y = 0$.

Consider any event $\mathcal{A}$ such that $0 < \Pr(\mathcal{A}) < 1$. The conditional correlation coefficient $\rho_\mathcal{A}$ between x and y, by definition, can be expressed as

$$\rho_\mathcal{A} = \frac{\text{Cov}(x, y \mid \mathcal{A})}{\sqrt{\text{Var}(x \mid \mathcal{A})}\sqrt{\text{Var}(y \mid \mathcal{A})}} \, . \tag{A.3}$$

We now replace both occurrences of y in equation (A.3) with the expression given in equation (A.2). The numerator of equation (A.3) may be rewritten as

$$\text{Cov}(x, y \mid \mathcal{A}) = \text{Cov}\big(x, (\rho \sigma_y / \sigma_x) \, x + \sqrt{1 - \rho^2} \, \sigma_y \, v \mid \mathcal{A}\big), \tag{A.4}$$

or equivalently,

$$\text{Cov}(x, y \mid \mathcal{A}) = \text{Cov}\big(x, (\rho \sigma_y / \sigma_x) \, x \mid \mathcal{A}\big) + \text{Cov}\big(x, \sqrt{1 - \rho^2} \, \sigma_y \, v \mid \mathcal{A}\big). \tag{A.5}$$

---

[12]See, e.g., Goldberger (1991), p. 75.

But since x and v are independent, the covariance between x and y conditional on the event $\mathcal{A}$ simplifies to

$$\operatorname{Cov}(x, y \mid \mathcal{A}) = (\rho \sigma_y / \sigma_x) \operatorname{Var}(x \mid \mathcal{A}). \tag{A.6}$$

Substituting equation (A.2) into the second part of the denominator of equation (A.3) and recalling that x and v are independent and v has unit variance, we obtain

$$
\begin{aligned}
\operatorname{Var}(y \mid \mathcal{A}) &= \operatorname{Var}\big((\rho \sigma_y / \sigma_x) x + \sqrt{1 - \rho^2}\, \sigma_y v \mid \mathcal{A}\big) \\
&= (\rho^2 \sigma_y^2 / \sigma_x^2) \operatorname{Var}(x \mid \mathcal{A}) + (1 - \rho^2) \sigma_y^2 \operatorname{Var}(v \mid \mathcal{A}) \\
&= (\rho^2 \sigma_y^2 / \sigma_x^2) \operatorname{Var}(x \mid \mathcal{A}) + (1 - \rho^2) \sigma_y^2. \tag{A.7}
\end{aligned}
$$

By substituting equations (A.6) and (A.7) back into equation (A.3), we may express the conditional correlation between x and y as

$$\rho_{\mathcal{A}} = \frac{(\rho \sigma_y / \sigma_x) \operatorname{Var}(x \mid \mathcal{A})}{\sqrt{\operatorname{Var}(x \mid \mathcal{A})} \sqrt{(\rho^2 \sigma_y^2 / \sigma_x^2) \operatorname{Var}(x \mid \mathcal{A}) + (1 - \rho^2) \sigma_y^2}}. \tag{A.8}$$

Finally, as was to be shown in this proof, we may simplify this expression to

$$\rho_{\mathcal{A}} = \rho \left(\rho^2 + (1 - \rho^2) \frac{\operatorname{Var}(x)}{\operatorname{Var}(x \mid \mathcal{A})}\right)^{-1/2}.$$

Incidentally, this result also demonstrates that—at least for the case when the joint distribution of x and y is normal—the conditional correlation coefficient does not depend on the variance of y directly.

Appendix B. Analytical calculation of conditional correlation coefficients

In this appendix, we provide some technical information on how one may calculate conditional correlations analytically when the conditioning events are of the form $x \in \mathcal{A}$, where $\mathcal{A} \subset \mathbb{R}$. Since we exclude zero-probability events from our analysis (*cf.* Theorem 1), events of interest cannot contain isolated points on the real number line. Candidate events $\mathcal{A}$ must therefore consist of either an interval or a set of non-overlapping intervals. Suppose first that $\mathcal{A}$ is a single interval, i.e., $\mathcal{A} = [a, b]$ with $a < b$.[13] In this case, the conditional variance of x is

$$
\begin{aligned}
\text{Var}\left(x \mid x \in [a, b]\right) &= \text{E}\left(x^2 \mid x \in [a, b]\right) - \text{E}\left(x \mid x \in [a, b]\right)^2 \\
&= \frac{\int_a^b x^2 f(x)\, \mathrm{d}x}{\int_a^b f(x)\, \mathrm{d}x} - \left[\frac{\int_a^b x f(x)\, \mathrm{d}x}{\int_a^b f(x)\, \mathrm{d}x}\right]^2 .
\end{aligned}
\tag{B.1}
$$

If x and y are bivariate normal and x has unit variance, equation (B.1) can be rewritten as, putting $\Pr(\mathcal{A}) = \int_a^b f(x)\, \mathrm{d}x$,

$$
\text{Var}\left(x \mid x \in [a, b]\right) = \frac{-b e^{-b^2/2} + a e^{-a^2/2}}{\sqrt{2\pi}\, \Pr(\mathcal{A})} + 1 - \left[\frac{-e^{-b^2/2} + e^{-a^2/2}}{\sqrt{2\pi}\, \Pr(\mathcal{A})}\right]^2 .
\tag{B.2}
$$

Alternatively, $\mathcal{A}$ may consist of a collection of mutually exclusive intervals, i.e., $\mathcal{A} = \cup \mathcal{A}_i$, $i = 1, \ldots, \ell$, with $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$ for all $i \neq j$. The conditional variance of x easily follows from straightforward modifications to equations (B.1) and (B.2). Equation (B.1), for instance, becomes

$$
\text{Var}(x \mid \mathcal{A}) = \frac{\int_{\mathcal{A}} x^2 f(x)\, \mathrm{d}x}{\Pr(\mathcal{A})} - \left[\frac{\int_{\mathcal{A}} x f(x)\, \mathrm{d}x}{\Pr(\mathcal{A})}\right]^2
\tag{B.1$'$}
$$

where $\int_{\mathcal{A}}$ signifies integration over the intervals $\mathcal{A}_i = [a_i, b_i]$ that comprise the set $\mathcal{A}$.

Thus, by substituting the applicable analytical expression for the conditional variance of x into equation (1), it is possible to calculate the conditional correlation between x and y numerically, for any conditioning set $\mathcal{A}$ of interest, as long as $\Pr(\mathcal{A}) > 0$.

---

[13]It is immaterial here whether this interval is open or closed since we are working with normal distributions, which have continuous probability density functions.

Appendix C. Extension of Theorem 1 to the multivariate, non-i.i.d. case

In this appendix, we show how the main result of Section 2, Theorem 1, may be extended to the case of time-series observations. Instead of assuming that (x, y) are a pair of bivariate normal random variables, we now assume that $x \equiv (x_1, x_2, \ldots, x_n)'$ and $y \equiv (y_1, y_2, \ldots, y_n)'$ are random vectors which satisfy:

$$x_{(n \times 1)} \sim N(\mu_x, \Sigma_{xx}), \tag{C.1}$$

$$y_{(n \times 1)} \sim N(\mu_y, \Sigma_{yy}), \tag{C.2}$$

$$\text{and} \quad \begin{pmatrix} x \\ y \end{pmatrix} \sim N \left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}{}' & \Sigma_{yy} \end{bmatrix} \right), \tag{C.3}$$

where $\Sigma_{xy} \equiv \mathrm{Cov}(x, y) = \mathrm{E}\big((x - \mu_x)(y - \mu_y)'\big)$.

If x and y are covariance stationary, the vectors μ_x and μ_y are vectors of constants, and Σ_{xx}, Σ_{yy}, and Σ_{xy} are banded matrices. More generally, we may allow some heterogeneity in both time series, as long as the heterogeneity and serial dependence satisfy conditions that permit the application of a weak law of large numbers (WLLN) such as the one proved by Andrews (1988).

The average correlation between x and y may be defined as

$$\rho = \frac{\mathrm{tr}(\Sigma_{xy})}{\sqrt{\mathrm{tr}(\Sigma_{xx}) \cdot \mathrm{tr}(\Sigma_{yy})}}, \tag{C.4}$$

where $\mathrm{tr}(\cdot)$ is the trace operator. An estimator of ρ is defined as the full-sample correlation between x and y:

$$\hat{\rho} = \frac{\frac{1}{n} \sum_t (x_t - \bar{x}_n)(y_t - \bar{y}_n)}{\sqrt{\frac{1}{n} \sum_t (x_t - \bar{x}_n)^2 \frac{1}{n} \sum_t (y_t - \bar{y}_n)^2}} \tag{C.5}$$

Since a suitable WLLN is assumed to hold, we observe that $p \lim \hat{\rho} = \lim_{n \to \infty} \rho$, and may therefore write $\hat{\rho} = \rho + $ sampling error. (Again, if x and y are covariance stationary, ρ is a constant in any finite sample.)

19

Since the random vectors x and y are multivariate normal, they may also be written as linear combinations of independent standard-normal random vectors U and V, as follows:

$$x = \mu_x + \Sigma_{xx}^{1/2} U, \tag{C.6}$$

$$y = \mu_y + \Sigma_{xy}\Sigma_{xx}^{-1}(x - \mu_x) + (\Sigma_{yy} - \Sigma_{xy}\Sigma_{xx}^{-1}\Sigma_{xy}')^{1/2} V$$

$$= \mu_y + \Sigma_{xy}\Sigma_{xx}^{-1/2}U + (\Sigma_{yy} - \Sigma_{xy}\Sigma_{xx}^{-1}\Sigma_{xy}')^{1/2} V \tag{C.7}$$

As in the discussion in Section 2, we are interested in the effects conditioning exerts on the correlation between x_t and y_t. We assume that conditioning consists of restricting the sample space to $(x \in \mathcal{A}, y \in \mathbb{R}^n)$, where $\mathcal{A} \subset \mathbb{R}^n$ such that $0 < \Pr(\mathcal{A}) < 1$. The random vectors $(x \mid x \in \mathcal{A})$ and $(y \mid x \in \mathcal{A})$ are, in general, not multivariate normal. Setting their first and second moments equal to $\mu_{x|\mathcal{A}}$, $\mu_{y|\mathcal{A}}$, $\Sigma_{xx|\mathcal{A}}$ and $\Sigma_{yy|\mathcal{A}}$, respectively, we define the conditional covariance matrix as

$$\mathrm{Cov}(x, y \mid \mathcal{A}) = \mathrm{E}\big((x - \mu_{x|\mathcal{A}})(y - \mu_{y|\mathcal{A}})' \mid \mathcal{A}\big) = \Sigma_{xy|\mathcal{A}}.$$

Setting $n_{\mathcal{A}}$ as the number of sample points for which $x_t \in \mathcal{A}$, the quantities $\rho_{\mathcal{A}}$ and $\hat{\rho}_{\mathcal{A}}$ may defined similarly to the corresponding unconditional moments (C.4) and (C.5):

$$\rho_{\mathcal{A}} = \frac{\mathrm{tr}(\Sigma_{xy|\mathcal{A}})}{\sqrt{\mathrm{tr}(\Sigma_{xx|\mathcal{A}}) \cdot \mathrm{tr}(\Sigma_{yy|\mathcal{A}})}} \tag{C.8}$$

$$\hat{\rho}_{\mathcal{A}} = \frac{\frac{1}{n_{\mathcal{A}}}\sum_{t \in \mathcal{A}}(x_t - \bar{x}_n^{\mathcal{A}})(y_t - \bar{y}_n^{\mathcal{A}})}{\sqrt{\frac{1}{n_{\mathcal{A}}}\sum_{t \in \mathcal{A}}(x_t - \bar{x}_n^{\mathcal{A}})^2 \cdot \frac{1}{n_{\mathcal{A}}}\sum_{t \in \mathcal{A}}(y_t - \bar{y}_n^{\mathcal{A}})^2}} \tag{C.9}$$

As in the unconditional case, we assume that a suitable WLLN applies to let the sample conditional moments converge to (finite) constants as $n \to \infty$ (and, of course, that $\rho_{\mathcal{A}}$ converges to the same limit).

Some algebra shows that

$$\Sigma_{yy|\mathcal{A}} = \mathrm{Var}(y \mid \mathcal{A})$$

$$= \mathrm{Var}\left(\Sigma_{xy}\Sigma_{xx}^{-1}(x - \mu_x) + \left(\Sigma_{yy} - \Sigma_{xy}\Sigma_{xx}^{-1}\Sigma_{xy}'\right)^{1/2} \cdot V \;\middle|\; \mathcal{A}\right)$$

$$= \mathrm{Var}\left(\Sigma_{xy}\Sigma_{xx}^{-1}(x - \mu_x) \mid \mathcal{A}\right)$$

$$+ (\Sigma_{yy} - \Sigma_{xy}\Sigma_{xx}^{-1}\Sigma_{xy}')^{1/2} \mathrm{Var}(V \mid \mathcal{A})(\Sigma_{yy} - \Sigma_{xy}\Sigma_{xx}^{-1}\Sigma_{xy}')^{1/2}$$

$$= \Sigma_{xy}\Sigma_{xx}^{-1}\mathrm{Var}(x \mid A)\Sigma_{xx}^{-1}\Sigma_{xy}' + \left(\Sigma_{yy} - \Sigma_{xy}\Sigma_{xx}^{-1}\Sigma_{xy}'\right)$$

$$= \Sigma_{xy}\Sigma_{xx}^{-1}\Sigma_{xx|\mathcal{A}}\Sigma_{xx}^{-1}\Sigma_{xy}' + \Sigma_{yy} - \Sigma_{xy}\Sigma_{xx}^{-1}\Sigma_{xy}'. \tag{C.10}$$

Similarly,

$$\Sigma_{xy|\mathcal{A}} = \mathrm{Cov}(x, y \mid \mathcal{A})$$

$$= \mathrm{Cov}\left(x, \Sigma_{xy}\Sigma_{xx}^{-1}x - \left(\Sigma_{yy} - \Sigma_{xy}\Sigma_{xx}^{-1}\Sigma_{xy}'\right)^{1/2}V \mid \mathcal{A}\right)$$

$$= \mathrm{Cov}\left(x, \Sigma_{xy}\Sigma_{xx}^{-1}x \mid \mathcal{A}\right) - \mathrm{Cov}\left(x, (\Sigma_{yy} - \Sigma_{xy}\Sigma_{xx}^{-1}\Sigma_{xy}')^{1/2}V \mid \mathcal{A}\right)$$

$$= \Sigma_{xy}\Sigma_{xx}^{-1}\mathrm{Var}(x \mid \mathcal{A})$$

$$= \Sigma_{xy}\Sigma_{xx}^{-1}\Sigma_{xx|\mathcal{A}}. \tag{C.11}$$

The relationship between $\rho_{\mathcal{A}}$ and ρ, defined in (C.8) and (C.4) above, can—in general—not be expressed in simple terms such as (1). However, in the case that (x_t, y_t) is i.i.d., we may simplify these equations, since $\Sigma_{xx} = \sigma_x^2 \cdot I_n$, $\Sigma_{yy} = \sigma_y^2 \cdot I_n$, and $\Sigma_{xy} = \sigma_{xy} \cdot I_n$. Setting $\rho = \sigma_{xy}/(\sigma_x\sigma_y)$, we may write:

$$\Sigma_{xx|\mathcal{A}} = \sigma_{x|\mathcal{A}}^2 \cdot I_n \tag{C.12}$$

$$\Sigma_{yy|\mathcal{A}} = \rho\sigma_x\sigma_y(\sigma_x^2)^{-1}\Sigma_{xx|\mathcal{A}}(\sigma_x^2)^{-1}\rho\sigma_x\sigma_y + \sigma_y^2 I_n - \frac{\rho^2\sigma_x^2\sigma_y^2}{\sigma_x^2}I_n$$

$$= \sigma_y^2\left(\rho^2\frac{\sigma_{x|\mathcal{A}}^2}{\sigma_x^2} + 1 - \rho^2\right)I_n \tag{C.13}$$

$$\Sigma_{xy|\mathcal{A}} = \frac{\rho\sigma_x\sigma_y}{\sigma_x^2}\sigma_{x|\mathcal{A}}^2 I_n$$

$$= \frac{\rho\sigma_y}{\sigma_x}\sigma_{x|\mathcal{A}}^2 I_n. \tag{C.14}$$

We thus obtain

$$
\begin{aligned}
\rho_{\mathcal{A}} &= \frac{\frac{\rho \sigma_y}{\sigma_x}\sigma^2_{x|\mathcal{A}}}{\sigma^2_{x|\mathcal{A}}\cdot\sigma^2_y\sqrt{\rho^2\frac{\sigma^2_{x|\mathcal{A}}}{\sigma^2_x}+1-\rho^2}} \\
&= \frac{\rho}{\frac{\sigma_x}{\sigma^2_{x|\mathcal{A}}}\sqrt{\sigma^2_{x|\mathcal{A}}\left(\rho^2\frac{\sigma^2_{x|\mathcal{A}}}{\sigma^2_x}+1-\rho^2\right)}} \\
&= \frac{\rho}{\sqrt{\rho^2+\frac{\sigma^2_x}{\sigma^2_{x|\mathcal{A}}}(1-\rho^2)}},
\end{aligned}
\tag{C.15}
$$

which is exactly equal to (1) in the bivariate case.

References

Andrews, Donald W.K., "Laws of Large Numbers for Dependent Non-identically Distributed Random Variables," *Econometric Theory* 4 (1988), 458–467.

Bera, Anil K. and Sangwhan Kim, "Testing Constancy of Correlation with an Application to International Equity Returns," CIBER Working Paper 96-107 (Champaign, IL: University of Illinois, 1996).

Bollerslev, Tim, "Modelling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH Model," *Review of Economics and Statistics* 72:3 (1990), 498–505.

Bookstaber, Richard, "Global Risk Management: Are We Missing the Point?," *Journal of Portfolio Management* 23:3 (1997), 102–107.

Campa, José M. and P. H. Kevin Chang, "The Forecasting Ability of Correlations Implied in Foreign Exchange Options," *Journal of International Money and Finance* 17:6 (1998), 855–880.

Goldberger, Arthur S., *A Course in Econometrics* (Cambridge, MA: Harvard University Press, 1991).

Karolyi, G. Andrew and René M. Stulz, "Why Do Markets Move Together? An Investigation of U.S.–Japan Return Comovements," *Journal of Finance* 51:3 (1996), 951–986.

Longin, François and Bruno Solnik, "Is the Correlation in International Equity Returns Constant: 1960–1990?," *Journal of International Money and Finance* 14:1 (1995), 3–26.

Sullivan, Greg, "Correlation Counts," *Risk* 8 (August 1995), 36–38.

Wilson, Thomas, "Infinite Wisdom," *Risk* 6 (June 1993), 37–45.