Three Essays on Causality

by Christos Bouras

Dissertation Submitted in Partial Fullfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

Department of Banking and Financial Management

University of Piraeus

December 2013

Acknowledgements

I would like to thank the many people who supported me in writing this thesis. First of all, I wish to thank my supervisor, Dr. C. Christou, for giving me the opportunity to start the Ph.D., her guidance throughout this thesis project, her steady encouragement to finish this thesis, and her patience. I am also indebted to Prof. N. Pittis for his helpful suggestions and inspiring me with his innovative spirit. I am also grateful to Prof. A. Antzoulatos for his helpful comments and assistance, in addition to creating a positive working atmosphere which encouraged collaboration and stimulating discussions. My thanks also go to my former colleagues.

Finally, I owe special thanks to my parents for giving me the chance to do this dissertation and my brother for his steady encouragement. Last but not least, I would like to thank all my friends who stood by my side during the last years.

Contents

Chapter 1

Introduction

Investigating causal relationships between a set of variables is one of the most important tasks in economics and finance because it helps to explain past data and experience and to provide reliable predictions of future outcomes. It is now well documented in the literature of economics that the presence of high correlation between economic time series cannot securely establish a causal relationship, since correlation is a measure of linear association only. If we consider a ‘universe’ entirely linear, then high correlation can be equated to causality. However, a correlation coefficient gives no indication about which variable is causing which. Wiener (1956) and Granger (1969) presented a fundamental probabilistic concept to analyze the dynamic relationships between economic data based on the direction of the time flow. The notion of Granger causality is defined in terms of predictability: the cause should precede the effect; and as a consequence, the cause should enable us to predict the effect.

According to Granger (1969, 1980, 1988), we say that the random variable x_t does not Granger cause the random variable y_t with respect to *E* if

$$f_{y_t}(g|E) = f_{y_t}(g|E_1), \quad \forall g \in \Re$$

where $E_1 = \sigma(y_{t-d}; d \geq 0)$ and $E = \sigma(y_{t-d}, x_{t-d}; d \geq 0)$ are the information sets generated by the past realizations of x_t and y_t up to time *t*-1 and $f_{y_t}(.|D)$ is the conditional distribution of y_t given the information set *D*. Hence, the definition of non-causality implies that the past information of x_t does not have an impact on the conditional distribution of y_t. When the equality does not hold, the variable x_t is said to Granger cause the random variable y_t with respect to *E*

$$f_{y_t}\left(g \middle| E\right) \neq f_{y_t}\left(g \middle| E_1\right), \quad \forall g \in \Re$$

1. The concept of Granger non-causality-in-mean

The above definition of non-causality, originally presented in Granger (1980), is too general to be applicable in practice. Granger introduces a weaker form of non-causality, namely Granger non-causality in mean, and it refers to the conditional mean of the random variables rather than their entire conditional distributions. We say the random variable x_t does not Granger cause the random variable y_t in mean with respect to *E* if

$$E\left(y_t \middle| E\right) = E\left(y_t \middle| E_1\right),$$

where $E\left(y_t \middle| D\right)$ is the conditional mean of y_t given the information set *D*. If the equality does not hold, the variable x_t is said to Granger cause the random variable y_t in mean with respect to *E*,

$$E\left(y_t \middle| E\right) \neq E\left(y_t \middle| E_1\right).$$

Hence, Granger causality in mean from x_t to y_t means that the past information of x_t is useful to help predict y_t at time *t*. This concept implies that the information embedded in the past values of x_t improves the 1 horizon ahead predictions of y_t because it yields better variances of the forecast error. Feedback or bi-directional Granger causality –in-mean between x_t and y_t occurs when x_t Granger causes y_t and y_t Granger causes x_t. Granger, as well as Pierce and Haugh (1977), also define instantaneous causality as contemporaneous correlation between the random variables. Evidence of non- causality-in-mean does not necessarily imply general non-causality, but rejection of the non- causality-in-mean hypothesis is sufficient to imply the presence of some general causality between the time series.

Tests for Granger noncausality-in-mean are straightforward for stationary time series. A test method involves estimating an autoregressive model for the variable enhanced with lagged terms of the other variable, calculating and retaining the sum of squared residuals from the regression, and comparing it to the sum of squared residuals obtained from fitting a univariate autoregressive specification for the other variable. The comparison can be performed by using an asymptotic chi square test. Haugh (1976) derives conditions for Granger non-causality-in-mean in terms of the cross-correlation function (CCF). In particular, the author shows that testing for independence between two covariance stationary time series via the cross-correlation function is equivalent to testing for Granger noncausality-in-mean. He proposes an asymptotic chi square test based on the sample cross-correlations at different lag orders of two separate innovation series obtained from fitting autoregressive moving average (ARMA) models to stationary time series. The Monte Carlo simulation results of Geweke, Meese and Dent (1983) show that the finite sample performance of these test procedures depends mainly on the selection of the lag truncation. Other Granger causality-in-mean tests are surveyed in Pierce and Haugh (1977) and Geweke, Meese and Dent (1983).

The widespread use of the Granger non-causality concept demonstrates why it is considered one of the major notions in the analysis of economic time series. The listing of applications of the concept includes various fields, such as finance, economics, marketing, physics and neuro-science among others.

2. The notion of Granger multistep non-causality

Granger' s original probabilistic concept is determined in terms of predictability one period ahead. Sims (1972, 1980, 1982) is the first author who addressed the issue of Granger causality *h* periods ahead in terms of a multivariate VAR model. Sims presents a similar causality concept with Granger. The identification of a causal relation does not require the specification of a econometric model. This is particularly appealing because it allows the development of empirical model strategies based on vector autoregressive models. Then, he presents general conditions for noncausality at any time period ahead in the bivariate case. In particular, in a bivariate process he establishes that a necessary condition for Granger non-causality from x_t to y_t requires that the corresponding coefficients of the innovations of x_t in the moving average representation of y_t are zero (Sims 1972, Theorems 1 and 2, or Pierce and Haugh 1972, Theorem 4.2). Therefore, the estimated impulse response coefficients of a moving average model provide a natural framework to test Granger non-causality at various time horizons. According to Sims, nonzero *h*-period ahead impulse responses coefficients derived from a VAR model, are interpreted as evidence of the presence of causality *h* time horizons ahead in the sense of Granger.

Moreover, the author proposes a measure of linear dependence between the multivariate time series. Sims measures Granger causal priority from x_t to y_t as the variance proportion of the *h*-period ahead forecast error of y_t attributed to the innovations of x_t (1982, pp. 131 – 132). Applications of Sims 's causality concept based on *h*-step ahead impulse responses and variance decompositions can be found in McMillin (1988), Faroque and Veloce (1990), Kyereme (1991), Stam et al (1991) and Tegene (1991).

Dufour and Tessier (1993) extend Sims's linear moving average characterization of Granger non-causality between two random variables in the multivariate case. They show that Sims 's conditions to achieve Granger non-causality between two random variables in terms of a moving average representation are not necessary and sufficient when a vector of auxiliary variables is also considered. Using a numerical illustration with a trivariate AR model, the authors show that evaluating whether the innovations of one variable has an effect on the other variable by means of impulse responses or the variance proportions of the forecast errors, is not equivalent to testing whether one variable has predictive ability for another (i.e., non-causality in the Granger sense).

Based on Sims's notion of causal priority, Geweke (1982, 1984 a, b) proposes measures, which enable the researcher to quantify the degree of dependence between two vectors of variables one period ahead. Thus, beyond testing for the presence of Granger causality between the variables, Geweke introduces criterions, which evaluate the intensity of a causal relation in terms of forecasting performance. The proposed measure equates to the sum of the mean square measure of unidirectional causality from x_t to y_t, the measure of unidirectional causality from y_t to x_t, and a measure of instantaneous causality. To estimate the relevant quantities, simple OLS linear regressions are used, the measures are calculated based on the estimated variances of the forecast errors, while inference is conducted by implementing asymptotic chi-square test criterions.

Lütkepohl (1994) extends Granger 's concept of one-step non-causality to multi-step non-causality in the multivariate case. Similarly to Granger, he considers all time series to be covariance stationary, and takes expected squared forecast error as a criterion for predictive accuracy. A random variable x_t is said to cause the random variable y_t h – steps ahead, where *h* denotes the forecast horizon, when the past information of x_t helps predict y_t up to any time point in period *h*. Thus, the past information of x_t may not be useful to improve predictions of y_t at time *t*, but it may be used to yield better forecasts of y_t at subsequent time periods.

3. Granger noncausality-in-mean for multivariate time series

Hsiao (1982) presents a generalization of Granger causal relationships among the random variables in a trivariate VAR process by introducing the notions of spurious and indirect causality. Hsiao defines two types of spurious causality, namely Type I and II respectively. According to Hsiao, there is Type I spurious Granger causality from z_t to y_t, if

$$V(y_t|E_1, E_2) = V(y_t|E_1)$$

and

$$V(y_t|U) < V(y_t|U - E_2)$$

where $E_1 = \sigma(y_{t-d}; d \geq 0)$, $E_2 = \sigma(z_{t-d}; d \geq 0)$ and $U = \sigma(y_{t-d}, x_{t-d}, z_{t-d}; d \geq 0)$ are the information sets generated by the past realizations of z_t, y_t and $\{x_t, y_t, z_t\}$up to time *t*-1,respectively, while $V(y_t|D)$ is the mean square error of the minimum mean square linear prediction error of y_t given information set *D*. The set $U - E_2$ denotes all the information in *U* apart the information in E_2. Type I spurious causality tell us that Granger causality occurs from z_t to y_t with respect to information set $U - E_3$ $(i.e., V(y_t|U) < V(y_t|U - E_2))$; however, this is a spurious result when the reference information set is reduced to $\{E_1, E_2\}$ $(i.e., V(y_t|E_1, E_2) = V(y_t|E_1))$. This type of spurious causality refers to situations where z_t appears to cause y_t, because it helps to reduce the noise between the interaction of the variables x_t and y_t, through the association of y_t with x_t.

Moreover, we say that there is indirect Granger causality from z_t to y_t, if

$$V(y_t|U) = V(y_t|U - E_2) < V(y_t|U - E_3) < V(y_t|U - E_2 - E_3)$$

$$V(x_t|U) < V(x_t|U - E_2)$$

and

$$V(x_t|E_2, E_3) < V(x_t|E_3)$$

where $E_3 = \sigma(x_{t-d}; d \geq 0)$ is the information set generated by the past realizations of x_t up to time $t - 1$. Indirect causality implies that although the random variable z_t does not Granger cause y_t $\left(i.e., V(y_t|U) = V(y_t|U - E_2)\right)$, it may still predict y_t through their relation with the random variable x_t. If z_t Granger causes x_t and x_t Granger causes y_t, then there is an indirect effect from z_t to y_t.

Hsiao argues that exclusion of variables, which contain valuable information for analyzing the causal structure of a multivariate time series, may lead to misleading characterization of the causality relations between the variables. Thus, Type II spurious Granger causality from z_t to y_t occurs, if

$$V(y_t|U) = V(y_t|U - E_2) < V(y_t|U - E_3) < V(y_t|U - E_2 - E_3)$$

$$V(z_t|U) < V(z_t|U - E_3)$$

and

$$V(z_t|E_2, E_3) < V(z_t|E_2)$$

The author presents necessary conditions for Granger non-causality, which involve coefficients of the moving average, as well as the autoregressive representation of the trivariate VAR process (theorems 1, 2 3 and 4, pages 9-11).

Tjostheim (1981) introduces a general framework for testing Granger one step ahead non-causality in a multivariate setting. He also proposes a test procedure based on multivariate regressions. Boudjellaba, Dufour and Roy (1992) present a new definition of Granger one step ahead non-causality for a stationary multivariate ARMA model. In particular, they first derive necessary and sufficient conditions for non-causality between two random vectors based on components of a stationary multivariate linear invertible process. A component-wise characterization of non-causality means that causality between the components of the random vector can be evaluated by considering causality between the corresponding scalar random variables. Their result generalizes Kang ‘s (1981) sufficient condition for Granger non-causal relations in a bivariate setting. Then, they investigate the implications of these general conditions on multivariate stationary ARMA processes. They establish a characterization of Granger one –step ahead non-causality between two vectors in terms of autoregressive and moving average coefficients for two different cases: first when all random variables participating the system are considered in the analysis and secondly when all scalar variables of the original vector process are partitioned in two subvectors (Theorems 1, 2 and 3, pp 1084). This two-sided characterization is then used to develop a causality test procedure for multivariate VARMA models.

4. Short and long horizon causality

For multivariate systems with more than three variables, establishing complete Granger causal relations based on Hsiao 's definitions becomes practically implausible. Dufour and Renault (1998) extend Hsiao 's general framework by introducing the concept of short-run and long-run non-causality. Their framework considers multivariate time series and is based on one hand on Lütkepohl' s notion of multi-step causality and on the other hand on Hsiao 's concept of indirect causality. Non-causality is evaluated at a specific point in time *h*, i.e., the forecast period, and this is appealing in empirical applications because it allows to differentiate between short run and long run causality. According to Dufour and Renault, short (or long) – run causality refers to situations where a random variable, say *X*, does not cause variable *Y* in period 1 in the Granger sense but it may still have predictive content for *Y* up to a subsequent time horizon, due to the indirect effects induced by the relation of *X* and *Y* with a vector of auxiliary variables, say *Z*.

First, without making any assumptions on stationarity or imposing an explicit functional form governing the dynamics of the multivariate process, i.e., such as an autoregressive representation, they present conditions (i.e., exhaustivity and separation conditions) which ensure the equivalence between non-causality at horizon one and non-causality at all horizons. Conditions for non-causality at all forecast horizons between vectors (random variables) are established in view of corresponding components of the vectors.

Second, they impose an autoregressive parametric form of possible infinite order governing the dynamics of the random variables, i.e., the linear invertible processes, in order to achieve a full characterization of the causal structure of the multivariate process at various forecast horizons. Characterization of non-causality at different *h* is obtained by setting specific zero restrictions on the autoregressive coefficients of the process. These coefficient restrictions reflect the indirect effects in terms of causality chains running between the random variables in the system at various forecast horizons (Theorem 3.2, page 1109). Defining non-causality by means of causality chains offers a better description of both *h*-step ahead causality and indirect causality. They also show that non-causality at various forecast horizons (characterized by zero restrictions on the coefficients of the linear invertible process) is equivalent to zero restrictions of the impulse response coefficients (of a moving average presentation) as in Sims and Lütkepohl. However, the authors argue that such causality properties fail to capture indirect causalities between the random variables. Instead, they show that generalized impulse response coefficients, defined as the lagged variables in forecasts at different horizons ahead, provide a solid characterization of the causal structure of the multivariate process. To make the hypotheses of interest empirically testable, they extend their results to the case of a finite order VAR process.

4.1. Testing for short and long horizon non-causality

Dufour and Renault' s conditions on non-causality between two variables at prediction horizon greater than one, with more than two time series included in the multivariate system, implies an increasing number of zero restrictions on products of VAR coefficients. The task of testing such hypotheses using likelihood ratio or Lagrange multiplier tests involves complicated computation given the difficulty of estimating conditional specifications under the null hypothesis. On the other hand, inference can be performed by using a standard Wald test because it only requires estimating the unrestricted model.

A regularity condition states that the asymptotic distribution of a standard Wald test is valid only when the matrix of the first partial derivatives of the coefficient restrictions is of full rank. Lütkepohl and Burda (1997) argue that the matrix of the first partial derivatives of Dufour and Renault 's VAR coefficient restrictions may be of reduced rank because these restrictions have a multilinear form. Therefore, the Wald statistic may fail to be asymptotically distributed as chi square under the null, and the use of the asymptotic chi square critical values may lead to misleading inference on Granger *h*-step ahead non-causality. The results from Monte Carlo simulation experiments provide evidence in support of their argument. Lütkepohl and Müller (1994) and Lütkepohl and Burda (1997) propose modified Wald statistics to test the *h*-step ahead non-causality hypothesis. Degenerate Wald statistics are shown to have a valid asymptotic chi square distribution under the null of non-causality at all forecast horizons. Nevertheless, these Wald statistics have poor finite sample power.

Dufour, Pelletier and Renault (2006) present an estimation and inference procedure for testing *h*-step ahead non-causality hypothesis in multivariate VAR models, stationary or nonstationary. They introduce the method of (*p*,*h*) autoregressions, which requires the estimation by least squares of long horizon vector autoregressions, in which some vector process V_{t+h} is regressed on $(V_1,..,V_t)$. The proposed parameterization enables the researcher to conduct inference in a simple efficient fashion by testing linear zero coefficient restrictions on the parameters of the

‘(p,h)- autoregressions’ via an asymptotic chi-square Wald test, after taking into account the serial correlation in the innovations using a Newey and West (1987) estimator. They also propose a lag-augmented version of their ‘(p,h)-autoregressions’ method to test non-causality at an arbitrary time horizon, in which the vector process V_t may be integrated of an unknown order. This extension is similar in spirit to that proposed by Toda and Yamamoto (1995) on testing the Granger one-period-ahead non-causality hypothesis in a VAR framework with levels ignoring the integration and cointegration properties of the time series. Since their simulation results show that the Wald statistic faces severe size distortions, the authors present a parametric Monte Carlo method to calculate p-values. Their simulation method ensures an enhanced finite sample size and power of the test procedure.

Hill (2007) argues that the test procedure of Dufour, Pelletier and Renault (2006) cannot be used for classification of causality chains over a range of time horizons because it focuses on testing the noncausality hypothesis at a time for a single prediction horizon, and therefore is not suitable for performing cross-horizon causality comparisons. According to Hill, their test method also fails to disentangle absence of causal linkages between the series and causal neutralization, in which several indirect effects offset each other. Moreover, the Dufour, Pelletier and Renault (2006) test results may present logical inconsistency with the original theory of Dufour and Renault (1998). In particular, Dufour and Renault state that causality at horizon h (where $h > 2$) from X to Y preceded by noncausality over the horizons 1 to h - 1, requires the presence of at least one indirect causal route between X, Y and the vector of auxiliary variables Z. However, the author shows that there may be situations where the outcome of the test implementation may violate this condition. Therefore, he proposes a recursive parametric representation of causality chains for trivariate processes defined in Hilbert spaces. Hill also presents a strategy which involves testing sequentially a set of simple linear restrictions in terms of trivariate VAR models by using an asymptotic Wald test statistic. The overall significance level of the test procedure at each step is controlled by means of Bonferroni bound adjustments. However, Bonferroni based procedures yield conservative tests with low finite sample power.

Eichler (2007) presents a graph-theoretic concept for analysis of *h*-step ahead causality relations of high-dimensional time series. Path diagrams are employed to visualize the autoregressive dynamics of multivariate weakly stationary processes, and then to derive general Granger causality relations among the random variables of the system. The author' s main concern is the theoretical graphical representation of general causality relations, and he does not consider an estimation and inference econometric procedure of multi-step causality in a time series framework.

Al-Sadoon (2010) shows that Dufour, Pelletier and Renault (2006) test may reject the horizon-specific noncausality hypothesis in situations where the causal effect is limited and confined to certain subspaces of the variations of the variables. Therefore, he presents a modified version of Dufour, Pelletier and Renault ‘s method for testing Granger noncausality up to a specific forecast horizon. Al-Sadoon argues that in order to reveal the full causal structure of the multivariate time series, rank restrictions rather than zero coefficient restrictions should be preferred on testing the relevant hypotheses. His test method involves fitting (p,h) autoregressions, and then using Kleibergen and Paap (2006) Wald type test statistic to conduct inference, after tackling the problem of serial correlation in the innovations by means of a Newey and West (1987) estimator. Following Robin and Smith (2000), estimation of the rank used in the test calculations is performed by implementing a sequential procedure. Similar to Dufour, Pelletier and Renault (2006), the author proposes a Monte Carlo method to compute the *p*-value of the Wald test.

Several limitations of testing *h*-step ahead non-causality were mentioned previously in this section. These procedures require the estimation of parametric mean regression and therefore are not designed to detect nonlinear causal associations or causalities in high order moments.

4.2. Measurement of short and long horizon causality

Causality tests provide in-sample evidence on whether a relation between economic time series is statistical significant, but they do not quantify the degree of their dependence. Econometric advances on the field present statistical tools for measurement of the magnitude of causality relations between economic data.

Dufour and Taamouti (2010) generalize Geweke 's (1982) one-period ahead causality concept to *h*-period ahead causality for multivariate time series. The authors argue that in multivariate systems the presence of indirect causal influences between the random variables, if there are any, may be evident only after several subsequent time periods. Granger non-causality at *h*-period ahead is defined through projections on Hilbert spaces, similarly with Dufour and Renault (1998). However, non-causality at forecast period *h* is characterized in terms of the variance-covariance matrix of the forecast errors (Proposition 3.1, pp. 45). The authors also present a characterization of what they define as 'unconditional non-causality' at a specific period *h*, i.e., the vector with auxiliary variables is dropped from the information set. Dufour and Taamouti propose a series of causality measures between two vector processes at a specific forecast period *h*, namely a mean square measure, an unconditional mean square measure, a measure of instantaneous causality, and a dependence measure. Dependence measure helps to determine in situations where bi-directional and instantaneous causality occurs, in which direction the causal effect is more intense. Hence, their concept enables the researcher to evaluate the strength of the causal relationship. For the case where the multivariate time series have a VARMA or VAR representation (or the system evolves as a general linear invertible process), explicit parametric forms of the causality measures which involve impulse responses are introduced by the authors (theorems 5.1, 5.2, 5.3, pp. 48-49). Asymptotically valid bias- corrected bootstrap confidence intervals are also proposed in order to evaluate the statistical significance of these measures.

Taamouti, Bouezmarni, and El Ghouch (2013) propose a nonparametric estimator and an asymptotic normal test for measures of Granger *h*-step-ahead causality between random variables. Measures of Granger causality, as originally introduced by Dufour and Taamouti (2010), are redefined in terms of copula densities. The measures are estimated in a nonparametric fashion using the Bernstein copula density. These criterions and tests can be used to quantify both linear and nonlinear causal relationships, as well as causalities of high order moments. The authors also propose a bootstrap bias-corrected estimator for the implementation of the measures and tests which yields good finite sample properties.

5. The concept of Granger non-causality-in-variance

Recent research on causal relations between the economic time series addresses the issue of non- causality-in-variance or second order non-causality. Granger et al. 's (1988) definition in a bivariate setting gave rise to a rapid growth of general econometric procedures for testing the non-causality-in-variance hypothesis between asset returns or economic variable growth rates (an early review of the literature includes that of Gagnon and Karolyi (2006)).We say the random variable x_t does not Granger cause the random variable y_t in variance with respect to *E* if

$$E\left(\left[y_t - E\left(y_t \middle| E\right)\right]^2 \middle| E\right) = E\left(\left[y_t - E\left(y_t \middle| E\right)\right]^2 \middle| E_1\right)$$

where $E\left(y_t \middle| D\right)$ is the conditional mean of y_t given the information set *D*.

Evidence of non- causality-in-mean and variance does not necessarily imply general non-causality, but rejection of either the non- causality-in-mean or the non-causality-in-variance hypothesis is sufficient to characterize general causality between the time series. Causality-in-variance is a sufficient condition for general causality even when the hypothesis of non-causality-in-mean holds.

5.1. Testing noncausality-in-variance in a univariate framework

Cheung and Ng (1996) generalize Haugh 's (1976) concept of testing independence between two covariance stationary series to variance non-causality. Variance non-causality is characterized in terms of the cross-correlation function (CCF) between the two separate squared whitened series (i.e., innovations derived from a univariate ARMA process) properly standardized by their conditional volatilities (it is assumed that each innovation series has a univariate GARCH representation). They propose an asymptotic chi-square test based on the sample cross-correlations at different lag orders of two separate squared standardized innovation series obtained from fitting univariate finite ARMA models with GARCH type errors to the data. Several other researchers have adopted Cheung and Ng 's concept of Granger variance non-causality; see, for example, Hu et al. (1997), Kanas and Kouretas (2002), Constantinou et al. (2005), Inagaki (2007).

Hong (2001) presented enhanced versions of the CCF tests by employing a weighting scheme on the cross-correlation estimates. Similar to Cheung and Ng (1996), his method can be applied to a bivariate covariance stationary process, while it involves fitting univariate ARMA (or a bivariate VAR specification) and GARCH specifications for each separate time series. Hong proposes asymptotic standard normal tests, which are calculated as the sum of the weighted T-1 (T denotes the length of the innovations) sample cross-correlations between the two squared standardized innovations, properly rescaled by some constant terms. Several widely used in the literature kernels such as the truncated, quadratic-spectral, Bartlett, etc, are implemented as weighting functions. All kernels allocate larger weight to a lower lag order (non-uniform kernels), except the truncated kernel (uniform kernel), which gives equal weighting to all lags selected. Moreover, kernels such as Daniell and Quadratic-Spectral use all T -1 cross-correlations but assign more weight to the N more recent lag time periods, while other kernels, like the truncated, Bartlett, Parzen and Tukey-Hanning, ignore sample cross-correlations larger than N. The parameter N controls the amount of smoothing applied to the cross-correlations and is denoted as bandwidth. The Monte Carlo experiments of Hong demonstrate that his tests are favorably compared to Cheung and Ng 's (1996) test in terms of empirical size and power.

In addition to the CCF based tests, Hafner and Herwartz (2006) propose a Lagrange Multiplier (LM) test which constitutes an adaptation of the misspecification testing framework in univariate GARCH models introduced by Lundbergh and Terasvirta (2002). Their method only tests for noncausality at a particular lag order, while its implementation is based on the estimation of univariate ARMA and GARCH specifications. Their test procedure involves regressing the squared innovations standardized by the GARCH conditional volatilities of one variable on quasi-maximum likelihood derivatives of the GARCH model estimation of the same variable and a bivariate set, which includes the innovations and conditional volatilities of the other variable. They show that their test has better power properties than Cheung and Ng 's test under a series of local alternatives.

Van Dijk, Osborn and Sensier (2005) prove that inference on Granger causality-in-variance may be misleading when both volatility series experience simultaneous structural breaks which are left unaccounted for. The asymptotic analysis of Rodrigues and Rubia (2007) also shows that the CCF based causality tests will not have their usual asymptotic distribution only when simultaneous structural changes occur in the dynamics of both volatility processes.

Testing noncausality-in-variance within the univariate framework is convenient because it only requires fitting univariate ARMA-GARCH models to the data. This is particularly appealing because these conditional specifications do not involve simultaneous modeling of intra and inter-series dynamics. On the other hand, the CCF test method does not preclude the possibility of causality relations that produce zero cross-correlations, such as potential nonlinear causalities or causality at high order moments. Moreover, these procedure are implemented on bivariate sets of data, and as a consequence, these tests ignore by construction potential indirect effects with other possibly important variables.

5.2. Testing noncausality-in-variance in a multivariate framework

Multivariate GARCH (MGARCH) models provide a natural framework for testing causality in variance (not all representations). A vast amount of empirical research is concerned with second -order causal relations between multivariate economic time series based on MGARCH representations (Karolyi (1995); Booth and Kootmos (1995); Booth et al. (1997); Jeantheau (1998); Ng (2000); Caporale et al. (2002), Ling and McAleer (2003)). However, little attempt has been made to derive conditions for general Granger causality-in-variance relations in MGARCH processes.

Comte and Lieberman (2000) are the first who addressed this issue formally by introducing a general theoretical framework, which involves the estimation of VARMA models with multivariate GARCH type errors. They present two definitions of Granger second-order non-causality: a conventional Granger second-order non-causality definition and a more general definition of linear Granger second order non-causality based on projections on Hilbert spaces. The authors argue that Granger et al 's definition equates to second order non-causality rather than variance non-causality because of the nonlinearities induced by the squared terms $[y_t - E(y_t|D)]^2$. They derive equivalence relations between the notions of Granger causality-in-mean, variance and second order non-causality (Proposition 1, pp. 538). They establish a characterization of second order Granger non-causality, which involves zero restrictions on specific coefficients of a MGARCH-BEKK representation. Likelihood based tests are also proposed to test the relevant hypotheses.

Hafner and Herwartz (2004) extend Comte and Lieberman' s results by proposing parametric characterizations of second order non-causality for strong, semi-strong, and weak MGARCH processes. Caporin (2007) presents general conditions for Granger non-causality-in-variance in the framework of MGARCH processes with in-mean effects.

Causality tests within the MGARCH framework are expected to have good power properties, provided that a sensible lag length selection is made for the model specification. However, these conditional volatility specifications face the 'curse of dimensionality' ; the number of coefficients to be estimated grows rapidly with increasing lag order and the number of variables in the system. Fitting restricted MGARCH models with a long lag structure involves tedious computation. Therefore, in empirical applications low order MGARCH models are usually employed. These limitations raise a reasonable amount of concern about the reliability of inference.

Lack of differentiation between causality-in-mean and causality-in-variance is likely to present serious challenges in empirical applications. Vilasuso (2001) demonstrates that inference on Granger causality-in-mean based on least squares test procedures- even when a heteroscedastic autocorrelation consistent (HAC) covariance matrix estimator is used - may be misleading under the presence of a causality-in-variance relation between the series. Hence, pre-testing for causality-in-variance is recommended before conducting inference on causality-in-mean. Moreover, the Monte Carlo simulation results of Pantelidis and Pittis (2004) show that, when there is causality-in-mean, conclusions drawn from implementing causality-in-variance tests without an explicit parameterization of such influences may lead to an erroneous claim that a statistical significant Granger causality-in-variance relation exists.

A brief enumeration of the applications of Granger causality-in-mean tests clearly highlights the widespread use of the concept in different fields. On the other hand, most (if not all) of the applications of Granger causality-in-variance tests are listed only in the field of finance. Despite the recent contribution of Vilasuso (2001) who emphasizes on the necessity of adopting the variance or second-order noncausality concept in time series analysis, a limited number of empirical efforts is documented.

Careful scrutiny of the literature presented above reveals some interesting aspects. First, existing literature focuses mostly on the existence and the direction of causality-in-variance. In many applications of these tests, however, the researcher's interest may be also centered on the causal lag structure of the time series. For instance, in applications with financial time series the researcher expects to find statistical significant volatility spillovers (if there any) at low lags. On the other hand, when both macroeconomic and financial time series are involved, one might anticipate that the volatility of asset prices will take more time to be transmitted to the volatility of macroeconomic variables. A typical example in the financial macroeconomics literature is the relationship between monetary policy and output growth. Several researchers, including Dufour, Renault and Pelletier (2006), Hill (2007), Dufour and Taamouti (2010) among others, document that monetary policy indicators, such as the short term interest rates and nonborrowed reserves, Granger cause GDP growth rates in US over a range of long prediction horizons. It is possible that similar causation patterns may occur between the second-order dynamics of the economic time series. Therefore, it is clearly desirable to know how sensitive are these tests in finite samples against alternative hypotheses of different causal lag structures. So far only limited simulation evidence is reported in the literature.

Second, the behavior of these tests in finite samples is evaluated via Monte Carlo simulation experiments using exclusively high persistent volatility data generating processes. In particular, these studies allow for alternative degrees of volatility persistence but their analysis is restricted to the cases of near Integrated and Integrated GARCH processes. These data generating mechanisms describe the dynamics of financial volatility. Empirical illustrations, however, indicate that macroeconomic volatility exhibits low degrees of persistence. It is not known a priori whether these test methods perform satisfactorily under low volatility dynamics in finite samples.

Third, inferential biases associated with the application of the test procedures may raise some concern among researchers and academics. The simulation results of Hong (2001) show that the finite sample properties of the Cheung and Ng (1996) test statistic present great sensitivity to arbitrary selections of the lag length. On the other hand, to date, how best to choose the bandwidth parameter N used in the calculations of Hong 's (2001) kernel based tests, remains unclear. The bandwidth parameter controls the degree of smoothing applied to the cross-correlations. The simulation results of Hong demonstrate that the choice of non-uniform weighting does not affect the finite sample properties of the kernel type tests. On the other hand, as with any kernel based semi-parametric test procedure, arbitrary selections of the bandwidth parameter have a significant impact on the power properties of the tests in finite samples. Therefore, it is clearly demanded a data-driven method of bandwidth selection under which formal statistical inference can be conducted, and helps to avoid possible inferential biases.

6. Testing Granger non-causality in cointegrated systems

Tests for Granger non-causality for stationary time series are straightforward. On the other hand, in cointegrated systems testing non-causality involves complicated computations given the presence of an uncertain number of unit roots. Sims, Stock and Watson (1990) demonstrate that, in a trivariate VAR model framework with levels, the Wald statistic may fail to be asymptotically distributed as standard chi square under the null of non-causality (in-mean) if some variables are found to be integrated and there is no cointegration between the time series. Under these circumstances, the use of standard chi-square critical values for evaluating the Wald type criterion may lead to misleading inference on causality. Moreover, they find that the Wald test is asymptotically distributed as standard chi square if the variables are cointegrated and if the long-run relationship involves the variable that is excluded under the null hypothesis of non-causality.

Mosconi and Giannini (1992) propose a method of testing jointly for Granger non-causality and the existence of cointegrated variables in the error correction framework. Coefficient restrictions on the parameters of levels VAR model are imposed jointly in order to test both hypotheses. To achieve this goal, the model is redefined as a error correction specification. Their approach requires pretesting for cointegrating ranks using Johansen' s (1988) procedure. Then, asymptotic chi-square likelihood ratio statistics are used to test non-causality in cointegrated systems. To compute the tests, a sequential procedure conditional on the cointegration rank is implemented to get estimates and log-likelihood values of both the restricted and unrestricted models.

Toda and Phillips (1993) study the asymptotic chi-square distribution of the Wald statistic when used to test Granger non-causality in multivariate VAR models with levels and error correction representations. They demonstrate that pretesting for a unit root, and cointegration in the economic time series is necessary before conducting statistical inference using the asymptotic chi-square Wald statistic. Moreover, the authors show that even if we have explicit knowledge of the previous, the usual asymptotic distribution of a Wald test may be invalid when testing non-causality in the levels VAR framework under the presence of a unit root or cointegrated variables. Specifically, the authors derive necessary conditions for non-causality in terms of the coefficients of a subset of the variables of a error correction model and a VAR model, and provide asymptotic theory for Wald criteria. Then, they show that the Wald statistic when implemented to test non-causality in the multivariate VAR framework, not only is distributed under the null as nonstandard chi-square but also depends on the nuisance parameters if some variables are integrated or cointegrated. To overcome these difficulties, the authors propose solutions which involve demanding computations but still yield disputable results. On the other hand, their analysis indicates that the asymptotic distribution of the Wald test based on the error correction representation is both nonstandard chi-square and dependent from nuisance parameters, except the cases where either the coefficient submatrices or the submatrices of cointegrating coefficients that are relevant under the null are of full rank. If one of the two conditions hold, the usual chi-square asymptotics are valid.

Toda and Yamamoto (1995) propose a method on testing the Granger non-causality hypothesis in a simple efficient fashion ignoring the integration and cointegration properties of the time series. The implementation of their procedure does not require the time series to be stationary, integrated of an arbitrary order, or cointegrated of an arbitrary order. Their method involves fitting a VAR model of order $p = k + d_{max}$ to the levels using the ordinary least squares estimation method, where k is the optimal lag order and d_{max} represents the maximal order of integration that the researcher believes might occur in the process. The authors imply overfitting the VAR specification with additional arbitrary d_{max} lags. Then, to conduct inference on causality one has to evaluate the statistical significance of usual zero coefficient restrictions on the VAR parameters up to lag k (d_{max} lagged coefficient matrices are assumed to be zero) using a standard Wald statistic.

Bruneau and Jondeau (1999) present a concept of long-run Granger non-causality between two variables based on the Error Correction Model (ECM) representation of cointegrated systems. According to the authors, a variable X is causal for another variable Y in the long-run, only if the past information of X helps to predict in the long-run the variable Y. Characterization of long-run Granger non-causality is achieved by setting a bilinear zero restriction on the parameters of a VAR model with levels and the long-run dynamic multipliers (Proposition 1, p. 547). This condition is not affected by the level of the cointegration rank. Hence, testing for long-run non-causality requires estimating a function of the long-run dynamic multipliers. However, the long-run dynamic multipliers cannot be estimated directly; they are estimated in terms of the restricted VAR (RVAR) model, as proposed by Campbell and Shiller (1988), and then inverted at a later stage. In particular, in a first step the parameters of the ECM representation and their asymptotic distribution are estimated using Johansen 's (1988) maximum likelihood method. The second step involves obtaining the estimators of the RVAR parameters and their asymptotic covariance matrix based on the ECM parameter estimators. Once the RVAR parameters are estimated, deriving the long-run dynamic multipliers is straightforward. The authors also propose an asymptotic chi-square test of Granger non-causality (Proposition 4, p. 555).

Yamamoto and Kurozumi (2006) extend Bruneau and Jondeau' s (1999) concept of long-run Granger non-causality to general long-run block non-causality, which we can make use to investigate for the presence of long-run causality from one block of variables to another block of variables. They define long-run non-causality similar to Bruneau and Jondeau (1999). Testable conditions for non-causality are derived based on their definition using the least squares prediction of detrended series (Proposition 1, p.708). The authors argue that short-term causality does not necessarily imply long-run causality and vice versa due to the presence of possible indirect causal effects. To make the hypotheses of interest empirically plausible, the authors first analyze the limiting distribution of the maximum likelihood estimator of the coefficient matrix on the error correction representation. In particular, they present the rank of the covariance matrix of the ML estimator in the limiting distribution (Proposition 2, p. 710). Their analysis indicates that the rank of the covariance matrix might be degenerate, making impossible to test the relevant hypotheses by means of a

Wald statistic because it is required that the covariance matrix is of full rank. The degeneracy problem is likely to appear in the long-run causality test of Bruneau and Jondeau (1999). To overcome this difficulty, the authors propose asymptotic chi-square Wald test statistics based on the generalized inverse of the covariance matrix. Implementation of these statistics requires determining the ranks of submatrices of the cointegrating matrix and its orthogonal matrix. The authors use sequentially the method of Kurozumi (2005) to test for the appropriate ranks.

Testing Granger non-causality in levels vector autoregressions and error correction models requires testing for unit roots and cointegration in the time series (we exclude the test method of Toda and Yamamoto (1995)). This strategy is not particularly appealing in empirical investigations due to the complicated computation. In addition, standard unit roots tests, such as Dicky and Fuller (1979) and Phillips and Perron (1988), are known to have poor power properties against the alternative hypothesis of stationarity. Furthermore, the finite-sample performance of tests for cointegrating ranks in Johansen 's ECM representation are found to depend on the values of the nuisance parameters. These observations raise some concern about the reliability of the statistical inference in finite samples.

7. Testing Granger non-causality in the frequency-domain

While traditional Granger causality tests have been the workhorse in time series analysis of economic data for more than three decades, they do not preclude the possibility of variation of the direction and strength of a causality relation over different frequencies. Recent researchers argue that it is crucial to distinguish between different causation patterns across different periodicities. Granger (1969) is the first to highlight the importance of using the spectral density approach to gain more insight in the dynamics of a causal relationship between two variables.

Geweke (1982) proposed a causality measure at a particular frequency based on the spectral density representation of a bivariate VAR process (for more see earlier section of this introduction). Hosoya (1991) builds on the Geweke framework, and introduces a causality test at a particular frequency. Yao and Hosoya (2000) presented a more sophisticated procedure for testing the noncausality null hypothesis at a particular frequency for a bivariate VAR process. Their test method involves evaluating a set of nonlinear coefficient restrictions of the vector autoregressive model parameters. To perform hypothesis testing, the Geweke causality measure based on VAR estimates is expanded via the delta method. Using this result, they propose a Wald test to test noncausality. Derivatives appearing in their Wald test statistic are approximated using numerical differentiation instead of deriving the exact analytical expression. Breitung and Candelon (2006) propose a simpler method to test for Granger causality measure between random variables in the Geweke framework. In particular, the null hypothesis of noncausality at a specific frequency is tested by using an F test statistic to evaluate the statistical significance of specific linear restrictions on VAR coefficients. These conditions are derived from the Geweke causality measure, when the latter is rewritten as a function of the estimated VAR coefficients. Their simulation results show that their test has good size and power in finite samples.

Lemmens, Croux, and Dekimpe (2008) propose a nonparametric test for the frequency-domain causality measure of Pearce (1979). Specifically, the cross-spectrum between two white-noise processes is estimated nonparametrically using a weighted scheme on the sample cross-covariances at different lag orders between two innovations obtained from fitting ARMA models to the data. Then, the coefficient of coherence is calculated based on the cross-spectrum estimates. According to the authors, the null hypothesis of Granger noncausality at a specific frequency is equivalent to testing whether the corresponding coefficient of coherence equates to zero. Inference is conducted by using an asymptotic chi-square test which is calculated based on the estimated squared coefficient of coherence at a certain frequency properly rescaled. Their results from Monte Carlo experiments show that the F test of Breitung and Candelon (2006) is more powerful in finite samples than their chi-square statistic.

8. Alternative notions of Granger non-causality

8.1. Out-of-sample Granger noncausality

Several researchers argue that the result from the implementation of a Granger in-sample causality test on the data must be in accord with the out-of-sample predictive performance of the corresponding forecasting model. Chao, Corradi and Swanson (2001) propose out-of-sample linear and nonlinear tests of the Granger one-step ahead noncausality hypothesis. In particular, based on sequences of forecast errors obtained from fitting autoregressive models to the data, the authors introduce an asymptotic normal test of noncausality. The authors also consider a more general test function including a nuisance parameter unidentified under the null hypothesis. Consequently, they develop a nonlinear causality test based on a approach followed in the field of testing for neglected nonlinearities.

8.2. Granger noncausality-in-risk

Modern portfolio theorists focus on measuring the probability of extreme downside asset price movements rather than estimating volatility, which is based on modeling the whole distribution of the asset returns. Hong, Liu, and Wang (2009) present the probabilistic concept of Granger non-causality-in-risk, which analyzes whether the occurrence of a large loss in one asset can help to predict the occurrence of a large loss in another asset. Their causality concept is defined in terms of the cross-spectrum between the left tails of the conditional distributions of the asset returns. The authors propose an asymptotic standard normal test, which is based on the weighted lagged sample cross-correlations between two risk indicators. The indicators are obtained from fitting univariate autoregressive quantile specifications on the asset returns.

9. Chapters of thesis

9.1. First Chapter

This thesis consists of three chapters. In the first chapter, we first evaluate the finite sample properties of recently proposed Granger causality-in-variance test procedures via extensive Monte Carlo simulations and then we introduce some modifications that improve their practical implementation to financial and economic time series. We focus our attention on four testing procedures: the Likelihood Ratio (LR) tests in the framework of a GARCH-BEKK(1,1) model as employed by Comte and Lieberman (2000); Cheung and Ng's sample cross-correlation (hereafter denoted as CCF) based S test (1996); the semiparametric CCF Q tests proposed by Hong (2001); and the Lagrange Multiplier (LM) test of Hafner and Herwartz (2006). Our investigation assesses and compares the size and power of these tests under various alternative hypotheses of practical importance. First, the sensitivity of the CCF based causality tests to the choice of the lag truncation and the bandwidth parameter is evaluated. Second, the credibility of statistical inference is assessed for different causal time lag structures, usually met in financial and macroeconomic applications. Third, we utilize different degrees of volatility persistence, reflecting the variations in the volatility dynamics between the financial and economic time series.

Our results show that Comte and Lieberman's likelihood ratio test as well as the Hafner and Herwartz's LM test suffer from severe size distortions, while they demonstrate very low power, under long horizon causality alternatives. Both cross correlation tests are reasonably well sized. However, Hong's kernel tests demonstrates less sensitivity to arbitrary choices of the weighting scheme and alternative volatility dynamics, when compared to Cheung and Ng's S test. Furthermore, cross correlations tests are favorably compared to LR and LM tests in terms of empirical power under a sequence of local alternatives. However, our results reveal that the power performances of Q and S tests greatly depend on the choice of bandwidth and lag truncation respectively.

Motivated by these findings, we introduce three simple methods for automatic bandwidth selection used in Hong's Q test calculations. The first method, the 'naïve' bandwidth selection criterion, chooses the bandwidth where the kernel CCF- based test value is maximum. Using the critical values of the standard normal distribution for the test statistic $Q = \max(Q(N))$ may result in misleading inference. Instead, our method uses the extreme value distribution to obtain the asymptotic critical value.

Two more satisfactory methods from a theoretical point of view are optimal bandwidth selection rules under which formal statistical inference can be conducted. The first method estimates the optimal bandwidth parameter as a structural change in the distributional behavior of the test. In particular, bandwidth is selected as the product of the comparison of the differences between the empirical distributions before and after a hypothetical change-point, for a sequence of hypothetical change-points. Dumbgen (1991) type estimators are implemented to determine optimal bandwidth parameter while three different semi-norms are used in the calculation of the measures. The second selection rule estimates the optimal bandwidth by applying the cross-validation method in kernel regressions of the squared standardized innovations. The procedure is based on choosing the bandwidth to minimize approximately the mean integrated squared error of the regression function. Härdle, Hall and Marron (1992) demonstrate that cross-validation yields bandwidths that are asymptotically consistent.

Simulation results show that the three methods for automatic bandwidth selection improve the test performance in finite samples. More importantly, under the first two bandwidth selection rules, the tests have high empirical power irrespective of the weighting scheme, the degrees of volatility persistence, or, whether causality is present at short or long horizon.

We also examine the causal relationship between the stock price volatility and industrial production volatility in the United States, Japan, Germany and Italy. Previous empirical evidence is inconclusive. Schwert (1989) finds that in US the volatility of stock returns in-sample predicts industrial production growth volatility only in two sub-samples out of four. He documents, however, weak evidence that industrial production growth volatility in-sample predicts stock returns volatility. On the other hand, Diebold and Yilmaz (2007) show that output growth fluctuations cause stock market volatility based on a large panel of economies. Hong's tests with bandwidths selected automatically by the 'naïve' criterion, present clear cut evidence that volatility spillovers occur from stock returns to industrial production growth rates in US and Japan. Moreover, the estimated bandwidth parameters are always large numbers suggesting long horizon causality dynamics. Hence, distant past changes in the volatility of stock returns appear to influence the recent fluctuations of industrial production growth rates in two out of four economies. Under optimal bandwidth parameter selection, these tests also suggest statistical significant short horizon causality-in-variance from industrial production growth to stock returns in US and Italy. The other tests yield mixed results.

9.2. Second Chapter

In the second chapter, we introduce a nonparametric method for testing unidirectional Granger causality-in-variance between two covariance stationary time series. Our approach is based on a model-free volatility proxy, as opposed to existing tests in the literature, whose implementation requires the estimation of a conditional volatility model via a GARCH type representation. These methods impose an explicit functional form on the evolution of the second order dynamics. However, these tests are suitable to be implemented to financial data, where the presence of the so-called 'volatility clustering' is well documented. On the other hand, Hamilton and Lin (1998) argue that evidence of ARCH effects in macroeconomic data is weak, casting doubts on whether GARCH specifications are suitable representations of second-order dynamics of the data. Hence, there is some concern about the reliability of statistical inference on second-order non-causality when GARCH based test procedures are applied to macroeconomic time series.

To our knowledge, no attempt to date has been made to test for causality-in-variance within an unconditional volatility framework. The methods of Schwert (1989) and Diebold and Yilmaz (2007) are the most relevant to the one proposed here. Schwert (1989), in particular, fits autoregressions to the series, and takes the absolute values of the associated residuals as volatility proxies. However, the proxies are transformed to conditional volatilities by estimating a VAR model. Hypothesis testing is performed by examining the significance of specific VAR coefficients zero restrictions via an F test. Diebold and Yilmaz (2007) follow Schwert' s approach and use the absolute values of the residuals from autoregressive models to estimate panel regressions.

Our nonparametric procedure compares simultaneously a set of p-values, which result from the implementation of a asymptotic standard normal test statistic at multiple single lag periods. The test calculations are based on the sample cross-correlations between the absolute values of innovations resulting from fitting autoregressions to the series of interest. By using a similar reasoning with Haugh (1976) and Cheung and Ng (1996), we establish that the test is asymptotically distributed under the null hypothesis of non-causality at an arbitrary lag order as standard normal. Joint inference is conducted by simultaneous comparison of the CCF-based test values at multiple lag periods, via the procedure developed by Rom (1990). In particular, the *p*-values are ordered and then contrasted to some corresponding critical levels of significance. The overall null hypothesis of non-causality is rejected if at least one *p*-value is found to be below the critical level of significance. The latter emerge as adjustments of the nominal level of significance to the lag truncation, so as to control the overall size of the test.

Monte Carlo experiments have been performed to evaluate the finite sample properties of the proposed test. The performance of our test is compared to two conventional tests, namely Cheung and Ng's (1996) *S* test and Hong's (2001) *Q* tests, under alternative models regarding the causal lag structure, the distributional characteristics of the series and the degree of fractional integration of the volatility process. Our simulation results show that the proposed test is well sized. In addition, the empirical test size exhibits less sensitivity than the *S* test to arbitrary choices of the lag selection parameter and the distribution of the error term. Interestingly the implementation of the proposed test yields surprisingly high finite sample power not only against the alternative hypothesis of short horizon causality, but also against the alternative hypothesis of long horizon causality. This great gain in empirical power holds for diverse selections of the lag truncation as well as for different sample sizes. By comparison, our test never performs worse than the *Q* and *S* tests and in fact outperforms both tests when dealing with short horizon causalities, especially so when the sample size is not very large. In addition, the power of both tests appears to depend greatly on the lag truncation and for the case of the Q test, on the weighting scheme used as well. Our findings also indicate that all test procedures have poor power under the presence of long-range dependence in the underlying volatility processes. Nevertheless, for large sample sizes there seems to be an advantage in using our proposed test.

Our test procedure is also used to examine the relationship between the volatilities of output growth and real stock returns. We used data for four develop economies, namely U.S., United Kingdom, Italy and Canada, covering a period from January 1973 to May 2011. Our test results show that industrial production volatility Granger cause real stock volatility in three economies out of four at level of significance 5%. This result is robust with respect to the implemented lag truncation. On the other hand, we find no significant relationship in the opposite direction. In comparison Cheung and Ng 's (1996) S test and Hong 's (2001) Q tests (with a few exceptions) do not reject the non-causality hypothesis in either direction.

9.3. Third Chapter

In the third chapter, our goal is to present a complete analysis of the multiple horizon causal linkages between the volatility of stock returns and industrial production growth rates, when a set of auxiliary processes, such as various monetary and macroeconomic variables, is included in the setting. We attempt to consider the classic question of whether fluctuations in the stock return volatility anticipate long-run changes in industrial production growth rates as the present value model predicts.

The studies of Whitelaw (1994), and Campbell, et al. (2001) are the most relevant to ours. Whitelaw (1994) applied different empirical formulations to identify leadership patterns between stock return volatility and stock returns over the course of the business cycle in US. He finds that the commercial paper- Treasury bill spread has predictive ability for stock return volatility, while he links the latter to the business cycle. Campbell, et al. (2001) document that volatility measures at the firm, industry and market level are highly correlated with GDP growth rates and NBER recession dates up to a lead and lag of a year in US. Their results from estimating OLS regressions with GDP growth as the dependent variable, and as independent variables, lagged GDP growth and the lagged return on the value-weighted CRSP index in addition to combinations of lagged aggregate volatility measures at different levels, indicate that stock return volatility has predictive power for short-term changes of real activity. Neither study, however, evaluates the relationship between output growth and stock return volatility at long predictive horizons, and both ignore possible influences from monetary policy factors on their joint behavior.

Multiple horizon Granger non-causality from the volatility of stock returns to industrial production growth rates is tested using the econometric procedure of Dufour, Renault and Pelletier (2006) on data from four developed economies, namely US, Germany, Japan, and Italy. Monetary policy indicators and economic variables such as money supply (M2) growth rates, inflation, and short-term interest rates operate as auxiliary variables. We document a large number of highly significant direct unidirectional causalities from the volatility of stock returns to output growth at both short and long horizons in all four economies. Our findings are consistent with the results of Whitelaw (1994), and Campbell, et al. (2001). More importantly, our results reveal that stock return volatility presents an enhanced ability to predict production growth rates at distant forecast periods. Interestingly, we also find that stock return volatility indirectly causes monthly growth rates of industrial production at long horizons through (i) nominal short-term interest rates in US, and Germany, (ii) money supply growth rates in Japan and (iii) inflation in Italy. These results have not been previously shown in the literature. Our findings also confirm the finding of Dufour, Pelletier and Renault (2006), Hill (2007), and Dufour and Taamouti (2010) that monetary policy is causal for real activity growth at both short and long horizons.

Second, the multi-step ahead causality measure introduced by Dufour and Taamouti (2010) is implemented on the time series to quantify the intensity of forecast improvement. Asymptotic valid confidence intervals are also constructed using a bootstrap technique presented by the authors. The results of the criterion largely confirm our hypothesis test findings. Except Japan, we document that stock market return volatility causes more strongly than the other variables real economy at long prediction horizons in Germany, and Italy. Moreover, the volatility of US stock returns appears to have a statistical significant indirect impact on output growth at distant horizons through money supply growth rates.

Third, a forecasting exercise reveals that the combination of stock return volatility, short-term interest rates, money supply growth, and inflation in a single regression model generates more accurate forecasts of output growth than the autoregressive model in the long term. Our forecasting results validate the causality linkages between stock market volatility, monetary policy and output growth obtained in our empirical investigations. We also show that single indicator specifications based on the lag structure revealed by our in-sample analysis fares well at predicting short term industrial production growth rates relative to the benchmark model when pooling the forecasts across estimation windows of different lengths or using a large simulation scheme to offset small sample estimation bias.

Chapter 2

Investigating the finite sample properties of causality-in-variance tests: A Monte Carlo approach

1. Introduction

Causality-in-variance has become an important element in risk management, asset pricing and the development of economic policy. A good understanding of the volatility spillovers is necessary for optimal asset allocation and the construction of better hedging techniques. For policy makers, the ability of measuring the transmission of volatility across markets is crucial for the development of various regulatory requirements, such as capital requirements or capital controls.

Granger's (1988) definition gave rise to a rapid growth of general econometric procedures for testing the noncausality-in-variance hypothesis between assets or economic variables (Cheung and Ng, 1996; Koutmos and Booth, 1995; Comte and Lieberman, 2000; Hong, 2001; Hafner and Herwartz, 2006). Recent econometric literature on testing causality-in-variance can be categorized into two strands. The first one focuses on testing for causality in the framework of multivariate models such as multivariate GARCH (MGARCH) models. MGARCH models provide the natural framework for testing causality in variance. Comte and Lieberman (2000) introduced a general theoretical framework, which involves the estimation of VARMA models with multivariate GARCH type errors. In their framework the null of non-causality-in-variance can be carried out by imposing linear restrictions on the GARCH parameters and testing their significance through likelihood based tests. The second strand focuses on testing for causality-in–variance by employing univariate GARCH models. Cheung and Ng (1996) proposed a simple two-step procedure based on the cross-correlation function (CCF) of the univariate squared standardised innovations obtained from univariate GARCH models. Hong (2001) presented enhanced versions of the CCF tests by employing a weighting scheme on the cross-correlation estimates[1]. In addition to the CCF based tests, Hafner and Herwartz (2006) propose a Lagrange Multiplier (LM) test which constitutes an adaptation of the misspecification testing framework in univariate GARCH models introduced by Lundbergh and Terasvirta (2002).

There is a considerable body of literature on investigating the causal relationships between the volatilities of international asset prices, including stock markets (Koutmos and Booth, 1995; Hu et al, 1997), exchange rates (Hong, 2001; Caporale et al., 2002), and exchange rates and derivatives (Cheung and Ng, 1996). However applications on economic time series are relatively sparse. To our knowledge, Vilasuso (2001) examined the causal relationship between money and prices, while Caporale and Spagnolo (2003) investigated whether there is evidence of volatility transmission between stock market and output growth in developed and emerging markets.

Despite the extensive empirical evidence and the recent theoretical developments in the field[2], so far only limited simulation evidence is reported about the finite sample performance of recently proposed causality-in-variance tests. Likelihood based tests within the MGARCH framework are expected to be more efficient, given that the parametric model is well specified. However, due to dimensionality problems this is never possible in practice; in empirical research low order MGARCH models are usually employed. On the other hand, testing procedures within the univariate framework are easier to implement, as they do not involve simultaneous modelling of intra and inter-series dynamics.

Furthermore, various major practical and theoretical questions are still left open to puzzle the academic researchers and the market practitioners. First, Cheung and Ng's (1996) tests involve calculating the sum of estimated squared cross-correlations. The practitioner must choose a number of cross-correlations to use in computing the test statistic. A similar decision must be made for the bandwidth parameter of the Hong's

[1] The reader can find many financial applications of the 'multivariate' (Karolyi ,1995; Booth and Kootmos ,1995; Booth et al. ,1997; Ng ,2000; Caporale et al. ,2002) and 'univariate' methods (Hu et al. ,1997; Kanas and Kouretas, 2002; Constantinou et al., 2005; Inagaki, 2007)

[2] Comte and Lieberman (2003) provided evidence on the asymptotic properties of the multivariate GARCH models.

(2001) tests. The lack of a procedure that helps the researcher to make an appropriate choice on the lag selection parameter or the bandwidth raises some uncertainty concerning the reliability of the statistical inference. So far, there is only limited evidence on the empirical performance of these tests related to different arbitrary selections of the lag truncation or the bandwidth. Second, the Cheung and Ng (1996) and Hong (2001) tests have not been compared against the Comte and Lieberman's (2000) likelihood based tests or the recently proposed Hafner and Herwartz's (2006) Lagrange Multiplier test. Third, it is interesting to examine the performance of these tests under different degrees of persistence, since they can be implemented not only to high persistent financial data, but also to macroeconomic data of moderate or low persistence. Fourth, it is of great importance to investigate the performance of causality-in-variance tests in the presence of long horizon causality dynamics. For example, in the case of financial data, one expects to detect causality at low lags (short horizon causality dynamics). However, when these tests are applied to detect causality between financial and macroeconomic variables causality patterns can be detected at distant lags, that is long horizon dynamics are present.

The purpose of our paper is to evaluate the finite sample properties of recently proposed Granger causality-in-variance testing procedures and to introduce some modifications that improve their practical implementation to financial and economic time series. We focus our attention on four testing procedures: the Likelihood Ratio (LR) tests in the framework of a GARCH-BEKK(1,1) model as employed by Comte and Lieberman (2000); Cheung and Ng's S test (1996); the semiparametric CCF Q tests proposed by Hong (2001); and the Lagrange Multiplier (LM) test of Hafner and Herwartz (2006). The causality-in-variance tests are compared in an extensive simulation study under different degrees of persistence, alternative causal structures and for a wide range of the lag truncation and the bandwidth parameter values. The simulation results show that Comte and Lieberman's LR as well as the Hafner and Herwartz's LM tests suffer from severe size distortions, while they demonstrate very low power, under long horizon causality alternatives. Both cross correlation tests are reasonably well sized. However, Hong's Q test demonstrates less sensitivity to arbitrary choices of the weighting scheme and alternative volatility dynamics, when compared to Cheung and Ng's S test. Furthermore, cross correlations tests are favorably compared to LR and LM tests in terms of empirical power under a sequence of local alternatives. However, our results reveal that the power performances of Q and S tests greatly depend on the choice of bandwidth and lag truncation respectively.

Motivated by these findings, we introduce simple methods for automatic bandwidth selection used in Hong's *Q* test calculations. Since the researcher has no apriori knowledge of the exact causal lag structure of the two series, it is evident that the appropriate bandwidth must be determined endogenously by the data. First, we follow an idea developed by Zivot and Andrews (1992) and we propose a naive approach which detects the particular bandwidth that assigns the most weight to the alternative. Since large values of the test statistic lead to rejection of the null of non-causality, bandwidth is chosen to maximize the causality test. As a consequence, critical values are data dependent and the use of standard normal distribution may lead to misleading inference. Our approach is to obtain critical values using the extreme value distribution as proposed by Berman (1964). A second method estimates the optimal bandwidth parameter as a structural change in the distributional behavior of the test. Dumbgen (1991) type estimators are used to determine optimal bandwidth parameter while three different semi-norms are implemented in the measure calculations. A third method estimates optimal bandwidth in a nonparametric regression of the squared standardized innovations based on the cross-validation method. Under this rule, bandwidth is selected by minimizing the integrated squared error criterion of the kernel regression, which is a classical measure of the closeness of the kernel estimator to its target parameter value. Simulation results show that the implementation of our procedures ensures high finite sample power.

In a brief empirical illustration, we examine the causal relationship between the stock price volatility and industrial production volatility in the United Kingdom, Japan, United States and Italy. The results suggest that the volatility of stock market returns is a leading indicator for the conditional variance of industrial production for the United States, United Kingdom and Japan at level of significance 5% while the volatility of industrial production help to predict the future stock returns volatility only in the case of the United Kingdom (at level 5%) and US and Italy (at level 10%).

The remainder of the chapter is organized as follows. In the next section we briefly present the four causality-in-variance tests. Section 3 provides the details about the Monte Carlo design. The simulation results are described in Section 4. Section 5 contains the empirical application and a final section concludes

2. Econometric testing procedures

In this section we briefly discuss the four methodologies for testing the no causality-in-variance null hypothesis.

2.1. Likelihood Ratio Tests in the MGARCH framework

Testing within the MGARCH framework has great empirical appeal because such specifications allow for interaction between the time varying volatilities of the series. Comte and Lieberman (2000) delivered a general theoretical framework with empirical implications for testing non causality-in-variance within VARMA models where the conditional covariance matrix is modeled as MGARCH.

Several MGARCH models have been employed in the literature[3]. A major problem with these models is that the researcher has to estimate a large number of parameters. Another problem is the positive definiteness of the conditional covariance matrix[4]. One of the most popular MGARCH specifications, the BEKK model defined in Engle and Kroner (1995), guarantees by construction the positive definiteness of the time-varying variance-covariance matrix.

Let us now briefly present the testing procedure in the framework of a BEKK model.

Consider a bivariate stochastic stationary and ergodic process $y_t = (y_{1t}, y_{2t})'$. Mean dynamics can be estimated by a VARMA model:

$$(1 - A_1 L - ... - A_r L^r) y_t = (1 - B_1 L ... - B_n L^n) u_t \quad (1)$$

[3] The most popular are the VECH model of Bollerslev, Engle, and Wooldridge (1988) , the BEKK model introduced in Engle and Kroner (1995), the Constant Correlation Model of Bollerslev (1990), and the Dynamic Correlation Model introduced by Engle (2002). See Bauwens, L., Laurent, S., Rombouts, J. V. K. (2006) for a broad review of the MGARCH models

[4] See Ding and Engle (1994)

Where L is the lag operator and A_l, $l = 1,2,...,m$ and B_v, $v = 1,2,...,n$ are 2x2 matrices of coefficients.

If $u_t = (u_{1t}, u_{2t})'$ is the vector of innovations, with $u_t / \psi_{t-1} \sim N(0, H_t)$, where ψ_{t-1} is the information set available at time $t-1$, the BEKK form of a GARCH(p,q) process is specified as:

$$H_t = D'D + \sum_{k=1}^{p} (G_k' u_{t-k} u_{t-k}' G_k') + \sum_{m=1}^{q} (F_m' H_{t-m} F_m). \qquad (2)$$

In (2) D is a 2x2 upper triangular matrix and G_k, $k = 1,2,..., p$, and F_m, $m = 1,2,..., q$ are 2x2 parameter matrices. One of the disadvantages of the BEKK(p,q) model is that the number of parameters increases rapidly with p and q. Conducting a testing procedure, which relies on the estimation of large number of parameters, raises a reasonable amount of concern about the precision of the estimates and the reliability of the statistical inference. Consequently, a low order model is preferred in the empirical research. Following the literature, we set $p = q = 1$ and we consider the BEKK(1,1) model for testing causality-in-variance in our analysis:

$$H_t = D'D + G_1' u_{t-1} u_{t-1}' G_1 + F_1' H_{t-1} F_1 \qquad (3)$$

As reported by Comte and Lieberman (2000), such a specification of the conditional covariance matrix allows us to test the null hypothesis of no causality-in-variance by setting the relevant off-diagonal coefficients of the parameter matrices equal to zero For example, the null hypothesis that the *i*th variable does not Granger-cause the *j*th variable in variance is formulated as $H_0 : g_{ij} = a_{ij} = 0$, $i, j = 1,2$, $i \neq j$.

The Likelihood Ratio test (LR) is employed to test for the validity of the null hypothesis of no causality-in-variance. Under the null, the LR test statistic has a chi-square distribution with γ degrees of freedom, where γ is the number of the restrictions imposed. It should be noted that under the usual regularity assumptions a sufficient condition for the asymptotic distribution of the LR test is the asymptotic normality of the quasi-maximum likelihood (QML) estimator of the MGARCH (p,q) specification. Not until recently, Comte and Lieberman (2003) derived asymptotic normality of the QML within the BEKK model.

The three alternative testing approaches that follow overcome the dimensionality problem because they involve estimation of univariate GARCH models.

2.2. Cheung and Ng 's (1996) CCF tests

Cheung and Ng (1996) present a simple and convenient two-step framework, with the important advantage that avoids the complexity of the direct estimation of the time-varying variance-covariance matrix. The CCF approach is an extension of the procedures developed by Haugh (1976) and McLeod and Li (1983). Their test statistics are asymptotic and they are not based on any innovation distribution assumption like normality.

In the first step, a VAR(r) model with GARCH type errors is fitted to the bivariate process $y_t = (y_{1t}, y_{2t})'$:

$$(1 - A_1 L - ... - A_r L^r) y_t = u_t , \tag{4}$$

with $u_t = (u_{1t}, u_{2t})'$, $u_{it} = \xi_{it} (h_{it})^{1/2}$, and $\{\xi_{it}\}$, $i = 1,2$, two independent sequences of iid random variables with mean zero and unit variance[5] . Conditional variances are given by

$$h_{it} = a_{i0} + \sum_{k=1}^{p} a_{ik} u_{i(t-k)}^2 + \sum_{m=1}^{q} b_{im} h_{i(t-m)} . \tag{5}$$

The second step involves the calculation of the sample cross-correlation function $\hat{\rho}_{ij}(\tau) = \{\hat{C}_{ii}(0)\hat{C}_{jj}(0)\}^{-\frac{1}{2}} \hat{C}_{ij}(\tau)$ ($i \neq j$) between the squared standardized residuals $\{\hat{e}_{it} = \hat{u}_{it}^2 / \hat{h}_{it}\}$, with $\hat{C}_{ij}(k) = T^{-1} \sum_{t=1}^{T-\tau} \hat{e}_{it} \hat{e}_{j(t+\tau)}$, $i, j = 1,2$, and $\tau \geq 0$.

[5] Lundbergh and Terasvirta (2002) assume $E\xi_{it}^3 = 0$ to ensure block diagonality of the information matrix of the log-likelihood function.

The null hypothesis that y_{it} does not Granger-cause in variance y_{jt} up to lag *N* can be tested using the test statistic

$$S = T\sum_{\tau=1}^{N}\hat{\rho}_{ij}^{2}(\tau), \tag{6}$$

which is asymptotically distributed as chi-square with *N* degrees of freedom. Alternatively, when the sample size is small a modified chi-square test can be used:

$$S^{*} = T\sum_{\tau=1}^{N}\omega_{\tau}\hat{\rho}_{ij}^{2}(\tau), \tag{7}$$

where $\omega_{\tau} = T/(T-\tau)$ or $\omega_{\tau} = (T+2)/(T-\tau)$

The empirical performance of Cheung and Ng s' tests presents great sensitivity to arbitrary selections of the lag truncation parameter *N*.

2.3. Hong's (2001) kernel-based CCF tests

Hong (2001) suggested a CCF based test which can be viewed as a generalization of the Cheung-Ng test. Specifically, he proposed a normalized test, the Q test, which utilizes a weighted sum of sample cross correlations. Under the null of no causality-in-variance, the Q test defined as

$$Q = \left\{T\sum_{\tau=1}^{T-1}w^{2}(\tau/N)\hat{\rho}_{ij}(\tau) - E(\tau)\right\}\Big/\{2V(\tau)\}^{1/2}, \tag{8}$$

follows asymptotically the standard normal distribution, where $w(.)$ is the weighting (kernel) function, N is called the bandwidth parameter, $E(\tau) = \sum_{\tau=1}^{T-1}(1-\tau/T)w^{2}(\tau/N)$,

and $V(\tau) = \sum_{\tau=1}^{T-1}(1-\tau/T)\{1-(\tau+1)/T\}w^{4}(\tau/N)$.

Hong (2001) considers several widely used in the literature kernels as weighting functions, all assigning larger weight to a lower lag order (non-uniform kernels), except the truncated kernel (uniform kernel), which gives equal weighting to all lags up to *N*. Kernels of compact support, like the truncated, Bartlett, Parzen and Tukey-Hanning assign to squared sample cross-correlations zero weight for lags higher than *N*, while the Daniell and Quadratic-Spectral kernels have unbounded support. These kernels include all *T*-1 sample cross-correlations. The Q test is one-sided, because under the alternative it diverges to positive infinity in probability as $T \to \infty$.

Hong (2001) provides also modified versions of the *Q* test for small samples, i.e.

$$Q_a = \left\{ T \sum_{\tau=1}^{T-1} (1 - \tau / T)^{-1} w^2 (\tau / N) \hat{\rho}_{ij}(\tau) - E^*(\tau) \right\} \Big/ \left\{ 2V^*(\tau) \right\}^{1/2}, \tag{9}$$

where $E^*(\tau) = \sum_{\tau=1}^{T-1} w^2(\tau / N)$ and $V^*(\tau) = \sum_{\tau=1}^{T-1} \left\{ 1 - (T - \tau)^{-1} \right\} w^4 (\tau / N)$.

The main disadvantage of the Hong's Q tests is the absence of a specific rule for the selection of the appropriate bandwidth *N*. Hong (2001) argues that the use of non – uniform kernel functions such as Quadratic Spectral, Parzen, Bartlett and Daniell, make the power robust to the choice of the bandwidth parameter N. However, the issue of how many cross-correlations should be employed in computing the Q, and Q^* test statistics remains an open question.

2.4. The Lagrange Multiplier test of Hafner and Herwartz (2006)

Hafner and Herwartz (2006) proposed a Lagrange Multiplier test statistic for causality in variance based on the estimation of univariate GARCH (1,1) models. Their framework is an adaptation of the general Lagrange Multiplier misspecification test introduced by Lundbergh and Terasvirta (2002). Let us assume for simplicity that $y_t = u_t$, where $y_t = (y_{1t}, y_{2t})'$ and $u_t = (u_{1t}, u_{2t})'$. To test the null that y_{jt} does not

Granger-cause y_{it} in variance for $i, j = 1,2$, $i \neq j$, Hafner and Herwartz (2006) consider the model:

$$h_{it} = a_i + b_i u_{it-1}^2 + c_i h_{it-1}, \tag{10}$$

$$u_{it} = \xi_{it} \left(h_{it} \lambda_t \right)^{1/2}, \tag{11}$$

$$\lambda_t = 1 + s'_{jt} \theta, \tag{12}$$

$$s_{jt} = \left(u_{jt-1}^2, h_{jt-1} \right)', \ \theta \geq 0, \quad \text{and} \tag{13}$$

$\{\xi_{it}\}$ a sequence of iid random variables with zero mean and unit variance.

The null and alternative hypotheses of the LM test are $H_0 : \theta = 0,\ H_1 : \theta \neq 0$. If the null of no causality-in-variance is valid, the model given in (10) disintegrate to a GARCH (1,1) model, and a LM test statistic can constructed as follows:

$$LM = (1/4T) \left(\sum_{t=1}^{T} \left(\xi_{it}^2 - 1 \right) s'_{jt} \right) V^{-1}(\beta_i) \left(\sum_{t=1}^{T} \left(\xi_{it}^2 - 1 \right) s_{jt} \right), \tag{14}$$

where $V(\beta_j) = (l/4T) \left(\sum_{t=1}^{T} s_{jt} s'_{jt} - \sum_{t=1}^{T} s_{jt} z'_{it} \left(\sum_{t=1}^{T} z_{it} z'_{it} \right)^{-1} \sum_{t=1}^{T} z_{it} s'_{jt} \right), \quad l = (1/T) \sum_{t=1}^{T} \left(\xi_{it}^2 - 1 \right)^2$,

$z_{it} = (1/h_{it})(\partial h_{it} / \partial \beta_i),\ \beta_i = (a_i, b_i, c_i)'$.

The LM statistic follows asymptotically the $\chi^2(2)$ distribution. The LM test of Hafner and Herwartz can be also easily computed as T times the R^2, where R^2 is the centered coefficient of determination of the regression of $(\xi_{it}^2 - 1)$ on s'_{jt} and z'_{it}.

3. Monte Carlo experiments

We conduct Monte Carlo experiments in order to assess and compare the size and power properties of the econometric tests outlined in the previous section in finite samples and under alternative assumptions of practical importance.

First, we evaluate the sensitivity of the CCF tests to the different choices of the lag truncation and the bandwidth parameter. CCF causality-in-variance tests proposed by Cheung and Ng (1996) and Hong (2001) are easy to implement, however, their performance in finite samples can depend greatly on the choice of the lag truncation and the bandwidth parameter respectively.

Second, the credibility of statistical inference is being assessed for different causal time lag structures[6], usually met in financial and economic applications. Even though the existing literature on testing the non-causality-in-variance hypothesis focuses mostly on the existence and the direction of causality, we believe that the causal lag structure of the time series is important. For example, in financial applications the researcher expects to detect spillover effects at low time lags; however that will not be the case when both macroeconomic and financial time series are involved. In such a case, one might anticipate that the volatility of asset prices takes more time to be transmitted to the volatility of macroeconomic variables. Furthermore, Dufour and Taamouti (2010) argue that a random variable X may cause a variable Y at long horizon if the causality is transmitted from X to Y indirectly through auxiliary variables.

In order to empirically reason this idea we conduct a preliminary analysis on two pairs of time series; the first includes financial and macroeconomic series while the second concerns only financial data. The sample covers a period from February of

[6] Hong (2001) provided some limited evidence on the performance of the CCF tests to different causal patterns. In particular, he included in his experiments cases of one and a fourth lag period causality in variance.

1973 through September of 2008[7]. In Figure1 we display graphically the sample cross-correlations for 36 lags between the monthly stock returns volatility and volatility industrial production growth[8] in the case of Japan. In the direction of positive lags, the cross-correlations exceed the two confidence bounds at distant lag periods, suggesting a long horizon causal structure in the volatility processes. In Figure 2 we plot the sample cross-correlations for 36 lags between the monthly squared stock returns of UK and US. The cross-dependence pattern changes thoroughly; only recent changes of the US squared stock returns seem to lead those of the UK. Motivated by these considerations, we extend the simulation design by taking into account also a second type of models where the conditional variances of the two processes are related through a distant time lag causal structure. We refer to this type of causality patterns as long horizon causality.

In addition to the previous, in our design we assume alternative volatility dynamics, reflecting the variations in the persistence of the volatility processes between financial and economic time series. In similar research (Hong (2001), Pantelidis and Pittis (2004), Dijk et al. (2005), and Hafner and Herwartz (2006)) the parameter values in the Monte Carlo experiments yield Integrated MVGARCH models and highly persistent volatility processes[9]. Although these papers pay close attention to high volatility persistence because it usually causes standard inference methods to break down, we broaden the evidence by including models with volatility dynamics usually met in real macroeconomic applications.

3.1. Data Generating Mechanism

The zero mean bivariate stochastic process $y_t = (y_{1t}, y_{2t})'$ is generated by the following data generating process (DGP):

$$y_t = u_t, \ t = 1,2,...,T, \tag{15}$$

[7] In Section 4 we describe the data with more details.
[8] The realised volatility is measured as the squared values of the (real) returns and growths.
[9] Hafner and Herwartz (2006) allow for alternative degrees of volatility persistence but their analysis is restricted to the cases of a near Integrated GARCH process and an Integrated GARCH process. A similar evaluation is performed by Pantelidis and Pittis (2004) focused on the size properties; note also that Hong (2001) employs low persistent univariate GARCH processes associated through exogenous lagged conditional variance terms.

$$u_t = \xi_t (H_t)^{1/2}, \quad (16)$$

$$H_t = D'D + \sum_{k=1}^{p} (G'_k u_{t-k} u'_k G_k) + \sum_{m=1}^{q} (F'_m H_{t-m} F_m), \quad (17)$$

$$\xi_t \sim NIID(0,1), \quad (18)$$

where $u_t = (u_{1t}, u_{2t})'$, $D = [d_{ij}]$, $G_k = [{}^k g_{ij}]$, $k = 1,2,..., p$, and $F_m = [{}^m f_{ij}]$, $m = 1,2,..., q$, are square parameter matrices .

Following Engle and Kroner' s (1996) definition of a covariance stationary BEKK process, we characterize the bivariate GARCH process defined in (15)-(18) as highly persistent if the maximum eigenvalue of $\sum_{k=1}^{p} (G_k \otimes G_k) + \sum_{m=1}^{q} (F_m \otimes F_m)$ is near to one in modulus.

To investigate the size as well as the power of the tests under a series of empirically plausible alternatives, we consider the following set of parameters under the DGPs defined in (15)-(18):

NULL1: [No Granger causality-in-variance, high persistence in the volatility process]

$$\begin{cases} p = q = 1 \\ ({}^1g_{11}, {}^1g_{12}, {}^1g_{21}, {}^1g_{22}) = (0.9,0,0,0.25) \\ ({}^1f_{11}, {}^1f_{12}, {}^1f_{21}, {}^1f_{22}) = (0.25,0,0,0.1) \end{cases}$$

NULL2: [No Granger causality-in-variance, low persistence in the volatility process]

$$\begin{cases} p = q = 1 \\ ({}^1g_{11}, {}^1g_{12}, {}^1g_{21}, {}^1g_{22}) = (0.5,0,0,0.35) \\ ({}^1f_{11}, {}^1f_{12}, {}^1f_{21}, {}^1f_{22}) = (0.3,0,0,0.1) \end{cases}$$

ALTER1: [Unidirectional, short horizon Granger causality-in-variance, high persistence in the volatility process]

$$\begin{cases} p = q = 1 \\ ({}^1g_{11}, {}^1g_{12}, {}^1g_{21}, {}^1g_{22}) = (0.9, 0, 0.25, 0.35) \\ ({}^1f_{11}, {}^1f_{12}, {}^1f_{21}, {}^1f_{22}) = (0.25, 0, 0.3, 0.1) \end{cases}$$

ALTER2: [Unidirectional, short horizon Granger causality-in-variance, low persistence in the volatility process]

$$\begin{cases} p = q = 1 \\ ({}^1g_{11}, {}^1g_{12}, {}^1g_{21}, {}^1g_{22}) = (0.5, 0, 0.25, 0.35) \\ ({}^1f_{11}, {}^1f_{12}, {}^1f_{21}, {}^1f_{22}) = (0.3, 0, 0.3, 0.1) \end{cases}$$

ALTER3: [Unidirectional, long horizon Granger causality-in-variance, high persistence in the volatility process]

$$\begin{cases} p = q = 12 \\ ({}^1g_{11}, {}^1g_{12}, {}^1g_{21}, {}^1g_{22}) = (0.9, 0, 0, 0.25) \\ ({}^1f_{11}, {}^1f_{12}, {}^1f_{21}, {}^1f_{22}) = (0.25, 0, 0.1) \\ ({}^{12}g_{11}, {}^{12}g_{12}, {}^{12}g_{21}, {}^{12}g_{22}) = (0, 0, 0.25, 0) \\ ({}^{12}f_{11}, {}^{12}f_{12}, {}^{12}f_{21}, {}^{12}f_{22}) = (0, 0, 0.3, 0) \\ ({}^kg_{11}, {}^kg_{12}, {}^kg_{21}, {}^kg_{22}) = (0, 0, 0, 0), \quad k = 2, 3, .., 11 \\ ({}^kf_{11}, {}^kf_{12}, {}^kf_{21}, {}^kf_{22}) = (0, 0, 0, 0), \quad m = 2, 3, .., 11 \end{cases}$$

ALTER4: [Unidirectional, long horizon Granger causality-in-variance, low persistence in the volatility process]

$$\begin{cases} p = q = 12 \\ ({}^1g_{11}, {}^1g_{12}, {}^1g_{21}, {}^1g_{22}) = (0.5,0,0,0.25) \\ ({}^1f_{11}, {}^1f_{12}, {}^1f_{21}, {}^1f_{22}) = (0.3,0,0,0.1) \\ ({}^{12}g_{11}, {}^{12}g_{12}, {}^{12}g_{21}, {}^{12}g_{22}) = (0,0,0.25,0) \\ ({}^{12}f_{11}, {}^{12}f_{12}, {}^{12}f_{21}, {}^{12}f_{22}) = (0,0,0.3,0) \\ ({}^kg_{11}, {}^kg_{12}, {}^kg_{21}, {}^kg_{22}) = (0,0,0,0), \quad k = 2,3,..,11 \\ ({}^mf_{11}, {}^mf_{12}, {}^mf_{21}, {}^mf_{22}) = (0,0,0,0), \quad m = 2,3,..,11 \end{cases}$$

For all the cases, we set $(d_{11}, d_{12}, d_{21}, d_{22}) = (0,0,0,0)$. We denote NULL1, ALTER1, and ALTER3 as highly persistent since in these models the maximum eigenvalue in modulus is 0.88. On the other hand, NULL2, ALTER2, and ALTER4 correspond to low persistent volatility processes; the maximum eigenvalue in modulus in these cases equals 0.45. There is no causality-in-variance under the high persistent NULL1, and the low persistent NULL2. This set up allows investigating the empirical size for different degrees of volatility persistence. Under ALTER1 and ALTER2 there exists short horizon causality-in-variance from y_{2t} to y_{1t} while in the case of ALTER3 and ALTER4 volatility spillovers from y_{2t} to y_{1t} occur at long horizon of time lag length 12. To assess the empirical performance of the CCF-based tests under different weighting schemes we use six kernel functions: the Bartlett, Parzen, Quadratic Spectral, Daniell, Tukey-Hanning and truncated kernels. Moreover, we calculate empirical size and power for a wide range of values for the bandwidth and the truncation lag parameter *N* (*N* = 5, 10, 20, 30, 40, 50, 60, 70, 80). The sample size *T* is 200, 500, and 1000[10] while for each case we generate 2500 replications. The start-up value for the conditional variance process is set equal to the unconditional variance H_t^*, $H_t^* = \left[I - \sum_{k=1}^{p} \left(G_k \otimes G_k \right)' - \sum_{m=1}^{q} \left(F_m \otimes F_m \right)' \right]^{-1} vec(D'D)$. After we have produced the simulated series, we apply the testing procedures described in Section 2.

[10] Following the literature we generate 4500 observations for each replication and we discard the first 2000 observations to reduce the start up effect.

4. Monte Carlo results

First, we examine the empirical size of the causality-in-variance tests described in section 2 under the null hypothesis that y_{2t} does not Granger-cause in variance y_{1t}. For space reasons we only present the results for Q and S tests, while the results for the Q_1^* and S^* tests are qualitatively very similar tests and are available upon request. Table 1 reports the size of the CCF tests at the 10% and 5% levels under the highly persistent NULL1. It is evident that Q tests perform quite similar for different sample sizes *T*, while their performance is relatively invariant to different non-uniform weighting schemes (e.g. Daniell, Tukey-Hanning, Quadratic-spectral, Bartlett and Parzen), and to the choice of bandwidth *N*. Specifically, the kernel based tests are slightly undersized for small *N*, but their empirical size is very close to the nominal for bandwidths larger than 10. On the other hand, *S* statistic faces under-sizing problems at both levels of significance, which become more pronounced as the lag truncation *N* increases. Its performance is very similar to that of the Q statistic when the truncated kernel is used. As far as the LM and the LR tests are concerned, Table 3 reveals that they face severe size distortions. Table 2 reports the empirical size of the CCF tests under the NULL2, i.e. when the underlying volatility process is low-persistent. All size patterns are very similar to those under NULL1. However, under NULL2 the under-sizing of the Q tests at low bandwidths is slightly larger. Furthermore, the LM and LR tests remain significantly undersized. In sum, under both NULL1 and NULL2 the size properties of Hong's tests are relatively invariant to the choice of the kernel function and bandwidth, while Cheung and Ng's *S* statistics appear to be more sensitive to the choice of the lag truncation parameter. The Comte and Lieberman's LR as well as Hafner and Herwartz's LM tests face significant size distortions.

Next, we examine the power performance of the causality-in-variance tests under evaluation, for a sequence of alternatives. The power performance of the CCF-based tests is evaluated for several values of the bandwidth/lag truncation parameter *N*, and different weighting schemes.

In the graphs of Panel I we display the power of the CCF-based tests, under the high persistent ALTER1, where there exists short horizon causality from y_{2t} to y_{1t}. The graphs in the left column present the performance of the tests at level 5% while those in the right column present the performance at level 10%. The power patterns present in the graphs of both columns are similar. On the other hand, the reader may observe that when the sample size *T* increases the cross-correlation tests appear to perform better at both levels of significance. Note that the power reaches a maximum value around *N* = 5 and then decreases gradually; for example, for *T* = 500 the rejection rates of Q which is based on the Quadratic Spectral kernel with *N* =5 and *N* = 50, are approximately close to 75% and 40% respectively at level 5%. Q tests with non-uniform weighting are more powerful than tests with uniform weighting. Specifically, the Q test which is based on the Parzen kernel is the most powerful, while the S test has the worst power performance. The results provide some support for use of certain kernels like the Parzen or the Bartlett and indicate that a careful choice of the number of cross-correlations is important for the reliability of the statistical inference. Particularly, in the case of low horizon causality large values of the bandwidth/lag truncation parameter *N* result in a significant loss of power.

Let us now explore the power performance of the CCF-based tests when the underlying volatility process is low persistent. The results for the tests against ALTER2, where there exists short horizon causality from y_{2t} to y_{1t}, are presented in the graphs of Panel II. The power performance is qualitatively similar to that observed against the high persistent ALTER1. Specifically, the power reaches a maximum value around *N* = 5 and then monotonically decreases. Furthermore, the performance of all CCF based tests is much better at large sample sizes than medium or small. For example, at 5% level of significance for *T* = 200 and *N* = 5 the power of Q tests when the Quadratic Spectral kernel is used, is close to 35%, while for *T* = 1000 is approximately equal to 95%. However, there is a significance difference in the performance of the CCF-based tests against ALTER1 and against ALTER2. All CCF-based tests are more powerful when the underlying volatility process is highly persistent, for all the values of the bandwidth/lag truncation parameter *N*, and irrespective of the weighting scheme employed. For example, for *T*= 500 the rejection rates of the Q test with N = 30, which is based on the quadratic-spectral kernel, are approximately equal to 86% and 66% against ALTER1 and ALTER2 respectively. As N increases the differences in power do not become more pronounced; the rejection rates of the Q test with N = 80, which is based on the quadratic-spectral kernel, are close to 65% and 46% against ALTER1 and ALTER2 respectively.

To investigate the power of causality-in-variance tests on detecting long horizon volatility spillovers, we examine their rejection probabilities against the high persistent ALTER3, and the low persistent ALTER4. Under both ALTER3 and ALTER4, there exists long horizon causality (at lag 12) from y_{2t} to y_{1t}. In the figures of Panel III we display the power performance of the CCF-based tests against ALTER3. The picture is quite different from that observed against the short horizon ALTER1 (Figure of T = 500 and level of significance 5%). Specifically, the power is very low for bandwidths/lag truncations less than 12; tests with non-uniform weighting have rejection rates steadily below 10%. However, as the bandwidth/lag truncation N approaches a critical value N^*, power exhibits a sudden increase and takes values between 95% and 100%. The critical value N^* equals 12 in the case of the tests with uniform weighting, while is somewhat larger for the tests with non-uniform weighting. The power performance of the tests appears to be very similar when applied at levels of significance 5% and 10% and at medium and large sample sizes.

Similar results are obtained against the low persistent ALTER4. The results presented in the graphs of Panel VI confirm the fact that the all power patterns are similar to those under the high persistent ALTER3. We observe the same tendency of under-rejecting the null of no causality-in-variance for N less than 12, while all tests reach high power at larger values of N. However, the power levels of the CCF-based tests are lower than those observed under the high persistent ALTER3 (figures of Panel III).

Finally, we examine the performance of the LM and LR tests of non causality-in-variance. In Table3 we report their empirical power against ALTER1-4. LR test establishes itself as being the trustworthiest empirical tool against short horizon alternatives irrespective of the alternate degrees of volatility persistence or the sample size. On the other hand, when applied to medium and large samples (T = 500 and 1000) the LM test has good power against ALTER2 (67% and 93.88% respectively) but is less powerful against ALTER1 (23.04% and 25.08% respectively) at level 5%, what indicates that the LM test loses power when the underlying volatility process is highly persistent. At levels 5% and 10%, both tests severely under-reject under the long horizon alternatives, ALTER3 and ALTER4 when applied to small, medium or large samples.

In summary, the empirical size of the CCF-based tests is close to the nominal size and appears to be relatively invariant to different non-uniform weighting schemes and to the choice of the bandwidth parameter N. On the other hand, LM and LR tests exhibit considerable size distortions. The persistence of the underlying volatility process does not seem to have any significant impact on the empirical size. However, the empirical power of the CCF-based tests, seem to be higher against high persistent alternatives, while the opposite holds for the LM test. CCF-based tests exhibit good power properties but their performance crucially depends on the choice of the bandwidth/lag truncation parameter N. Specifically, the horizon of causality determines how power depends on the choice of N. Simulation results support the belief that LR tests have the best power[11], but only under the presence of recent volatility spillovers. Our findings clearly indicate that the power performance of LR and LM tests against long horizon causality alternatives is very poor.

Because of their poor performance against some of the alternative models considered, LR and LM tests cannot be considered reliable tools for statistical inference. On the contrary, the CCF-based tests have better power properties provided that the bandwidth parameter is accurately selected. Among the CCF-based tests Hong's Q test should be preferred for empirical work since they exhibit better size properties. To enhance the applicability of these tests, we suggest a practical rule for selecting the optimal bandwidth[12]. In the next section, we describe a simple procedure that automatically selects the optimal bandwidth for the calculation of the Q tests.

[11] See Hafner and Herwartz (2006)
[12] Optimal bandwidth is considered the bandwidth that maximises the power of the tests

5. **Automatic Bandwidth Selection**

In this section, we describe two data driven method for selecting the optimal bandwidth parameter. As is shown by the simulations results presented in the previous section the bandwidth has little impact on the size of the Q tests while it significantly affects their empirical power.

The investigation of the empirical power as a function of the bandwidth N revealed an interesting feature: at a certain value of the bandwidth parameter, N^*, power exhibits some sort of episodic variation. Moreover, this shift occurs at a bandwidth that is associated with the true horizon k^* of the causality pattern. For bandwidths less than the true causality horizon k^* the tests yield very low power. Specifically, at these bandwidths the empirical power is very close to the nominal size of the tests. This is not surprising since for $N < k^*$ kernels of compact support (truncated, Parzen, Bartlett, Tukey-Hanning) attach non-zero weight only to cross-correlations $\rho(N)$ which equal to zero. This happens because in the case of a long horizon causality DGP, causality occurs at lag k^* greater than the selected bandwidth N. In the case of kernels of unbounded support (Daniell and Quadratic Spectral) the power behaves similarly since zero cross-correlations $\rho(N)$, $N < k^*$, receive much larger weights. Consequently, in the case of long horizon causality, Q statistics for bandwidths N less than the true causality horizon k^* behave as if the null hypothesis of no causality was true, and the power is close to the nominal size. At a certain bandwidth $N^* \geq k^*$ the statistics capture the long horizon causality dynamics and yield high power. This particular bandwidth N^* can be regarded as change-point in the distributional behaviour of the Q test, and the stability of the power performance of the test.

At these change-points Q tests yield maximum power. Thus, the optimal bandwidth can be defined by selecting among the various potential optional bandwidths the specific value that generates a permanent regime shift in the power performance of the Q tests.

Our primary objective is to test the null of non causality-in-variance against the alternative of causality of unknown horizon. To start with, we suggest a naive way to proceed, which can be seen as a rough approach. Since the size of the Q tests is approximately equal to the nominal significance level for a variety of arbitrary bandwidth selections, we suggest to detect the particular bandwidth that assigns the most weight to the alternative. Since large values of the test statistic lead to rejection of the null of no causality-in-variance, N is chosen to maximize the one sided Q test. This idea is quite similar in spirit to the one used by Andrews and Zivot (2002) to test the unit root hypothesis.

If we denote N^* the optimal bandwidth parameter, then we can utilize a grid search to determine its value:

$$\hat{N}^* = \arg\max_{2 < N \leq N_k} Q(N) \,. \tag{18}$$

$\hat{N}^*$ is an estimator of the optimal bandwidth and the upper bound N_k is any large positive integer. The proposed procedure determines the optimal bandwidth to be the point at which the Q test is maximized for a selected grid of N_k bandwidth values. The application of this procedure is quite simple and computationally attractive.

The resulting distribution of the test statistic $B_q = \max_{2<N<q} Q(N)$ for some N chosen using the data is different than the standard normal. The following theorem suggests using the critical values of the extreme value distribution of Type I to conduct inference.

Theorem. Let $Q(2),...,Q(q)$ a sequence of Q test statistics as defined in (8) with q any integer greater than two. Then

$$P\{c_q(B_q - d_q) \leq x\} \rightarrow \exp(-e^{-x}), \tag{19}$$

in distribution

where

$c_q = (2\log q)^{1/2}$, $B_q = \max_{2 \le i \le q} Q(i)$, and $d_q = c_q - 0.5(2\log q)^{-1/2}(\log\log q + \log 4\pi)$.

Our result is a straightforward application of the limit theorem for the maxima of independent and stationary random variables proposed by Berman (1964).

A strategy more satisfactory from a theoretical point of view could be based on the estimation of the optimal bandwidth N^* as a change-point in the distributional behaviour of the Q test. Darkhovskh (1976) and Carlstein (1988) are among the first who considered the problem of change-point estimation for a sequence of random variables in a nonparameric setting. Dumbgen (1991) generalized their results and showed that the rate of convergence is $O_p(N_k^{-1})$.[13] The concept behind Carlstein's approach is to compare the differences between the empirical distributions before and after a hypothetical change-point, for a sequence of hypothetical change-points. Specifically, given the sequence of statistics $\{Q(N), \quad 1,2,\dots,N_k\}$ we estimate the optimal bandwidth N^* using the estimator:

$$\hat{N}^* = \frac{1}{N_k}\left(\underset{1<N\le N_k}{\arg\max}\{\Lambda(D_N)\}\right), \tag{20}$$

where $D_N = \left[\frac{N}{N_k}\left(1-\frac{N}{N_k}\right)\right]\left(\frac{1}{N}\sum_{i=1}^{N}\delta_{Q(i)} - \frac{1}{N_k - N}\sum_{i=N+1}^{N_k}\delta_{Q(i)}\right)$, $\delta_{Q(i)}$ is the delta measure and $\Lambda(\cdot)$ is a semi-norm on the space of signed finite measures. Following Carlstein (1988) we employ the following three semi-norms:

$$\Lambda_h(\nu) := \left(\frac{1}{N_k}\sum_{i=1}^{N_k}|d_i|^h\right)^{1/h}, \quad h = 1,2, \text{ and } \Lambda_3(\nu) := \sup_{1\le i\le N_k}|d_i|, \text{where } d_i = \nu(\mathbb{1}_{\cdot < x_i}).$$

[13] Darkhovskh(1976), Calrlstein (1988) as well as Dumbgen (1991) considered independent sequence of random variables. Recently, Ben Hariz et al. (2007) considered a very general class of dependent sequences and proved the consistency of Dumbgen-type estimators.

The upper bound N_k can be any large number depending on the sample size. Our simulation results show that the selection of N_k has only minor effects on the performance of the Q tests.

A third selection rule estimates the optimal bandwidth by implementing the method of cross-validation in a kernel regression of $\{\hat{e}_{2t}\}$on $\{\hat{e}_{1t}\}$. A standard criterion for evaluating an estimator is to determine how close it is to the true parameter value. For nonparametric regression estimation by the kernel method, integrated squared error gives a measure of the distance between the regression and the underlying 'smoothed' curve. The problem of bandwidth parameter selection is to determine the bandwidth that is optimal in the sense that the integrated squared error criterion is minimized with respect to all bandwidth selections.

Let $(\hat{e}_{1t}, \hat{e}_{2t})$ be a realization of the bivariate process (X_t, Y_t). Consider the regression

$$Y_t = g(X_t) + u_t,\ t = 1,..,T \tag{21}$$

where $g(.): \Re^d \to \Re$ represents a functional of X_t which minimizes $E|Y - r(X)|^2$ with respect to $r(.): \Re^d \to \Re$ with $E|r(X)|^2 < \Delta < \infty$, while $\{u_t, t = 1,...,T\}$ are independent and identical distributed random variables with distribution $F(.)$.

The Nadaraya-Watson kernel regression estimator is defined as

$$\hat{g}(x) = \frac{\sum_{t=1}^{T} Y_t W\left((X_t - x)/N\right)}{\sum_{t=1}^{T} W\left((X_t - x)/N\right)} \tag{22}$$

where $W : \Re^d \to \Re_+$ is the kernel function.

The cross-validation (CV) procedure, as suggested by Rudemo (1982) and Bowman (1984), selects the optimal bandwidth N^* by minimizing the Integrated

Squared Error $\Lambda = \int (\hat{g}(x) - g(x))^2 dx$. Consider $\hat{g}_N(.)$ the estimate of the regression function $\hat{g}\ (.)$ which corresponds to a specific bandwidth *N*. For each observation *t*, the method evaluates the prediction error $\{Y_t - \hat{g}_{N,-t}(X_t)\}$ using the regression (21) with that observation removed from the modeling process. The term $\hat{g}_{N,-t}(.)$ denotes the 'leave-one-out' estimator, where the t^{th} observation is dropped from (22)

$$\hat{g}_{N,-t}(x) = \frac{\sum_{s\in\{1,..,T\}\setminus\{t\}} Y_s W((X_s - x)/N)}{\sum_{s\in\{1,..,T\}\setminus\{t\}} W((X_s - x)/N)},\ t = 1,...,T \qquad (23)$$

Optimal bandwidth is estimated using a grid search

$$\hat{N}^* = \arg\min_{2<N\leq T} CV(N) \qquad (24)$$

where

$CV(N) = T^{-1}\sum_{t=1}^{T}(Y_t - \hat{g}_{N,-t}(X_t))^2$ is the weighted average of squared errors.

Härdle, Hall and Marron (1992) demonstrate that cross-validation yields bandwidths that are asymptotically consistent.

Details about the finite sample performance of these estimation procedures are reported in the next subsection.

5.1. Simulation results

In this section we report the results of the simulation study conducted to evaluate the empirical performance of the Q tests when the bandwidth parameter is automatically selected based on the proposed in the previous section procedures. We also examine the sensitivity of these tests to the selection of the upper bound N_k. The

Monte Carlo design remains identical to the one described in section 2, and examines the size and the power properties of the Q tests and the modified statistics B_q.

Table 4 reports the empirical size under NULL1 and NULL2 when the "naïve" optimal bandwidth selection procedure is implemented. N_k is set equal to 100. Let us first discuss the results of the tests when the critical values of the extreme value distribution are used. It is evident that under the low-persistent NULL2, Q tests with non-uniform weighting perform quite well; their empirical size is close to the nominal. For example, for T = 500 at the 5% level the empirical size of the B_q test with Daniell, Quadratic Spectral, Parzen, Bartlett and Tukey-Hanning weighting is, 3.04%, 3%, 2.92%, 2.92% and 2.96% respectively. Contrary, B_q test based on standard normal critical values suffers from size distortions; for 5% level, the empirical size of the B_q test with Daniell and Quadratic Spectral weighting is 12.76% and 13.36% respectively. The empirical size performance of the B_q tests when the underlying volatility process is high-persistent (NULL1) is similar to that observed under the low-persistent NULL2. However, it is clear that in the former case the rejection rates are higher especially at the 5% level when using the critical values of the standard normal distribution. For example, for T = 1000 at the 5% level the empirical size of the B_q test with Daniell, Quadratic Spectral, Parzen, Bartlett and Tukey-Hanning weighting is, 19.36%, 19.56%, 17.56%, 17.56% and 18.56% respectively. At level 10% the results are very similar to those reported for NULL2. However, we need to highlight the fact that uniform weighting yields more profound over-rejection at both levels and for both null hypotheses; for example, under NULL2 the 5% empirical size of the B_q test with the truncated kernel when implemented to a sample of size 1000 is 30.64% and 40.92% at levels 5% and 10%, respectively.

Let us now report the empirical size performance of the Q tests when the optimal bandwidth is estimated using a Dumbgen-type estimator. The results reported in Table 5 show that the tests are well sized; the empirical size is very close to the nominal in most cases for both hypotheses. For example, under NULL2 for T = 500 the 5% empirical size of the Q test with Daniell, Quadratic Spectral, Parzen, Bartlett and Tukey-Hanning weighting is, 4.84%, 4.8%, 4.64%, 5.08% and 4.88%

respectively. Overall, the size distortions are less profound comparing to those reported above for the case of the “naïve” optimal bandwidth selection procedure.

Table 6 and Table 7 present the power simulation results for the proposed B_q tests with the “naïve” optimal bandwidth selection procedure. The case of short horizon alternatives, ALTER1 and ALTER2, are reported in Table 6. The proposed tests appear to have very good properties against both alternatives at levels of significance 5% and 10% for medium and large sample sizes; for instance when T = 1000, the power of the tests is always beyond 80%. Note that when these tests are implemented to small samples (T = 200) they achieve a bad performance against ALTER1 and ALTER2 ranging approximately from 10% to 50% at both levels. However, we should stress out that the reported power is the maximum power that can be achieved for this sample size for the range of bandwidths from 1 to 100 (see figures in Panel I and Panel II). The choice of the weighting scheme does not appear to have any effect on the performance of the tests.

Table 7 reports the power performance of the tests when our “naive” method of automatic bandwidth selection is implemented against the long-horizon alternatives, ALTER3 and ALTER4. The results suggest that B_q tests achieve maximum power against ALTER3, that is when the underlying volatility process is highly persistent. Specifically, when B_q tests are implemented on data with T = 500 and 1000 the power exceeds 90% irrespective of the kernel function used. Again, as in ALTER1 and ALTER2, the tests have a relatively satisfactory performance when applied to small sample sizes. Persistence in the volatility process does not seem to have any significant effect on the rejection probabilities in the presence of long horizon causality. In the case of low persistence, tests continue to have very good power against long horizon causality patterns; for T = 500, 1000 the power of the B_q tests against ALTER4 is between 72% and 99.6% at the 5% significance level.

Let us now discuss the power performance of the Q tests when the optimal bandwidth is estimated using a Dumbgen-type estimator. The results are calculated for three semi-norm choices, Λ_1 , Λ_2, and Λ_3. Table 8 reports the empirical power under the short horizon alternatives, ALTER1 and ALTER2. In general, we can the empirical power is lower than the one observed when the “naïve” optimal bandwidth selection was used. This difference is more pronounced when the sample size is small. In the case of T=200 the power is relatively low. For example, under the high persistent ALTER1 the Daniell kernel yields power at the 5% and 10% levels near 24% and 30%, respectively. The power performance is getting better as the sample size increases. In the case of non-uniform kernels the power at the 10% level is between 61% and 74%, and 88% and 97%, for T=500 and T=100, respectively. Under the low persistent ALTER2 the power is slightly lower. Except for this, all power patterns are similar to those under ALTER1. Non- uniform kernels give similar power, while the choice of the semi-norm has no significant effect on the power performance of the tests. Under the long-horizon alternatives, ALTER3 and ALTER4, the results reported in Table 9 show that most of the kernels give high power. The exceptions is the Parzen and truncated kernels which give much lower power especially in small and medium sample sizes. All the other kernels give high power both under ALTER3 and ALTER4. For example, for T=1000 the empirical power is around 99% at both levels of significance. In the case of moderate sample sizes the power remains high; for T=500 at the 10% level the power is 96%-99% under ALTER1, and 89%-98% under ALTER2. Our procedure yields good power even in the case of a small sample. For example, in the case of T=200, at the 10% level the empirical power is 60%-85% and 43%-63% under ALTER1 and ALTER2, respectively.

Tables 8, 9 and 10 report the size and power performance of the tests when the cross-validation method of automatic bandwidth selection is implemented at levels 5% and 10%. Overall, the tests are well sized at both levels of significance. For instance, for T = 500 at the 5% level the empirical size of the Q test under NULL1 with Daniell, Quadratic Spectral, Parzen, Bartlett and Tukey-Hanning weighting is, 5.44%, 5.48%, 5.44%, 5.40% and 5.40% respectively. The size properties of the Q tests when implemented to high-persistent volatility processes (NULL1) are similar to those observed under the low-persistent NULL2. The Q tests appear to be oversized tests at level 5% under both NULL1 and NULL2 when applying the method of cross-validation on data of sample size T =1000. For example, for T = 1000 at the 5% level the empirical size of the Q test under NULL1 with Daniell, Quadratic Spectral, and Tukey-Hanning weighting is, 9.88%, 9.96%, and 9.84% respectively.

Table 9 and Table 10 present the power simulation results for the proposed Q tests with the cross-validation based optimal bandwidth selection procedure. The case of short horizon alternatives, ALTER1 and ALTER2, are reported in Table 9. The proposed tests appear to have very good properties against ALTER2 at levels of significance 5% and 10% for all sample sizes; for instance when T = 1000, the power of the tests is always 100%. On the other hand, tests with non-uniform weighting achieve a low performance against ALTER1 ranging approximately from 10% to 30% at both levels (excluding quadratic spectral based test). Note that the power of the truncated Q test ranges from 40% to 98%.

Table 10 reports the rejection frequencies of the tests when the cross-validation method of automatic bandwidth selection is implemented against the long-horizon alternatives, ALTER3 and ALTER4. The tests have a relatively satisfactory performance against both alternative hypotheses at 5% and 10% when implemented to moderate and large sample sizes. In particular, the power against both models ranges from 94% to 97% . On the other hand, the rejection frequencies of the Q tests against ALTER3 and ALTER4 when implemented to samples of small size (T = 200) range from 69% to 83% and 32% to 46% respectively.

In summary, the proposed procedures for selecting the optimal bandwidth yield tests with reasonable size and very good power against a sequence of alternatives of practical importance. When the optimal bandwidth is estimated using a Dumbgen-type estimator or cross-validation in terms of a kernel regression the empirical size of the Q tests are very close to the nominal size, whereas the "naïve" optimal bandwidth selection procedure yields tests which slightly over-reject especially the processes are highly persistent. What concerns the empirical power performance, the "naïve" optimal bandwidth selection procedure outperforms in the case of short-horizon alternatives. Simulation results show that the "naïve" procedure consistently behave very well in terms of power no matter whether the causality is present at short horizon or long horizon. Under long-horizon alternatives all procedures ensure high power. The upper bound N_k has only minor effects on the size performance of the proposed procedures. However, smaller values of N_k seem to reduce the size distortions in the high-persistent case.

6. Empirical illustration

In our empirical analysis we examine the relationship between stock returns volatility and output growth volatility. The evidence provided in this section is complementary to those presented by Schwert (1989) and Diebold and Yilmaz (2007). An explicit view in finance is that stock prices reflect in the present time the discounted future expected earnings of all firms in a specific economy. So, one would consider that the changes of the conditional variance of the stock prices are proportional to the changes of the conditional variance of the future discount rates and expected future earnings. However, volatilities of both the future earnings and discount rates change whenever there are variations in the volatility of real economy. On the other hand, the stock prices are forward looking; reaction of speculative investors to anticipate events about the future economic fundamentals yield shifts in current stock price volatility.

Schwert (1989) provides some weak evidence on the predictive ability of the US stock market volatility for the future volatility of industrial production. He tested the non-causality-in-variance hypothesis in both directions by using monthly (and daily) US stock returns and industrial production growths covering a period from 1857 to 1987. Recently, by employing a panel data framework on quarterly data for a large number of countries, Diebold and Yilmaz (2007) have found that output growth Granger causes-in-volatility the stock market returns.

Our analysis is restricted to four developed countries: the UK, the US, Italy, and Japan. The data consist of monthly observations of the aggregate stock price index, the industrial production index, and the consumer price index (CPI). We define monthly stock returns and output growth rate as the logarithmic differences of stock indices and industrial production, respectively. Real stock returns are computed by subtracting CPI inflation from nominal stock returns. Stock market data is taken from Datastream. Industrial production and CPI are retrieved from IFS and OECD database, respectively. The sample period spans from January 1973 to September

2008. The standardized squared residuals are calculated by fitting the data to VAR-GARCH(2,2) models[14].

Hong's Q tests are calculated using a fixed bandwidth N = 12 as well as optimal bandwidth estimated by the "naive" optimal bandwidth selection procedure, B_q, the cross-validation (denoted as $Q(\hat{N}^*_{CV})$) and the Dubgen based method (denoted as $Q(\hat{N}^*_{DB})$) presented in the previous section. We employ four kernel functions, the Bartlett, the Parzen, the Daniell and the Quadratic Spectral. We exclude the truncated kernel since simulations showed that its performance is inferior relative to the non-uniform kernels. The empirical analysis commences with investigating the null hypothesis of no Granger causality-in-variance from stock market returns to industrial production growth. The results presented in Table 13 provide evidence that when the optimal bandwidth is utilized, Q tests strongly reject the null hypothesis for the UK, the US and Japan. Interestingly, in most cases the estimated optimal bandwidths are large numbers suggesting long horizon causality dynamics (if any). In other words, distant past changes in the volatility of real stock market returns seem to have an "impact" on the recent changes of the volatility of industrial production growth. For example, in the case of the US Dubgen type and cross-validation bandwidth estimation method yield bandwidth values $\hat{N}^* = 51$ and 45 for the Tukey-Hanning kernel, respectively. For these bandwidths the tests statistics values are 1.6519 and 1.8519 respectively, suggesting the existence of statistically significant distant period volatility spillovers; the corresponding *p*-values are 0.0493 and 0.0320, respectively. Not surprisingly, Hong's Q tests with $N = 12$ fail to detect any significant causality relationships for all countries in our sample, excluding US. Cheung and Ng's S and S tests with $N = 12$ indicate significant causality link only in the case of the US. One possible explanation is the poor power performance of the CCF tests under long horizon causality alternatives when the bandwidth/lag truncation parameter is not accurately chosen as documented in section 3.2 (Fig5 and Fig6). The LR test indicates that stock market returns Granger cause-in-variance the

[14] To save space the VAR(r)-GARCH(p,q) estimated coefficients are not reported, but are available from the author upon request.

industrial production growth only in the UK (the p-value is 0.0001). The later result is consistent with the conservatism of the LR test found in the simulation analysis. The LM rejects the null hypothesis in two countries, the US and Italy.

In Table 14, we present the results for testing the presence of causality from industrial production volatility to stock returns volatility. The B_q tests show statistically significant volatility spillovers from output growth to real stock returns in the US and Italy at the 10%, and in UK at the 5%. Moreover, cross-validation optimal bandwidth selection procedure indicates rejection of the null hypothesis in UK and Japan at levels 5% and 10% respectively, while the Dubgen type bandwidth selection rule indicates the presence of causality in UK at level 5%. In both countries the optimal bandwidth parameter values range from 2 to 20. Contrary, the empirical results provide evidence that neither Q tests with given bandwidth of 12, nor the S and S tests support the presence of volatility spillovers from industrial production to stock returns. However, these results could be expected given that our simulation analysis highlights the fact that in the case of short horizon causality the selection of a bandwidth greater than 5 results in a loss of power. On the contrary, LR test indicates highly significant volatility spillovers from output growth to real stock returns in Italy and the UK. The LM test rejects the null hypothesis only in the case of Italy.

To sum up, Q tests along with an optimal bandwidth selection procedure provide strong evidence of volatility transmission from real stock market returns to industrial production growth for the UK, the US and Japan. They also suggest the existence of volatility spillovers in the opposite direction at the 5% in the case of the UK, and at the 10% in case of US and Italy. Contrary, the applications of Q tests with an exogenously given bandwidth, and S tests fail to detect any volatility spillovers in any direction. LM and LR tests's results lead to mixed conclusions. The empirical results highlight the practical importance of the proposed optimal bandwidth estimation methods.

7. Conclusions

In this paper we evaluate the finite sample properties of four causality-in-variance tests recently proposed in the literature. We focus on the LR test within the framework of multivariate GARCH models proposed by Comte and Lieberman (2000), the LM test of Hafner and Herwartz (2006), and the cross correlation function based tests of Hong (2001) and Cheung and Ng (1996). Our findings indicate that LR and LM tests face severe size distortions while they demonstrate poor power performance against long horizon (distant-lag) causality dynamics. However, the LR test is the most powerful among the tests under study in the presence of short horizon causal links while is not affected by the degree of persistence in the volatility processes. Hong's Q tests are well sized, while their performance in the sense of size is relatively invariant to different non-uniform weighting schemes, and to the choice of bandwidth. Cheung and Ng's S tests are reasonably sized but their empirical size performance presents great sensitivity to arbitrary selections of the lag truncation parameter. Most importantly, lack of accuracy in the choice of the bandwidth/lag truncation parameter has a substantial effect on the power properties of the cross correlation function based tests. Furthermore, the finite sample properties of both Hong's and Cheung and Ng's tests are found to demonstrate moderate sensitivity to alternative degrees of persistence in the underlying volatility process.

To enhance the empirical performance of the kernel-based tests proposed by Hong (2001), we suggest three procedures for automatic optimal bandwidth selection. The first method determines the optimal bandwidth as the one at which the Q test is maximized for a selected grid of bandwidth values. The second method utilizes a Dumbgen-type change-point estimator. The third selects bandwidth in a kernel regression of the squared standardized innovations by using the cross-validation method. The simulation results demonstrate that the implementation of our procedures yield powerful tests with only minor size distortions. Most importantly, the resulting tests seem to be robust to different causality horizons and to alternative degrees of volatility persistence. Robustness and computational simplicity makes our procedures very attractive in practice.

We also applied the optimal bandwidth selection methods to examine whether there are statistically significant volatility spillovers between real stock returns and industrial production growth for the United Kingdom, the United States, Italy and Japan. Hong's Q tests with bandwidths selected by our automatic bandwidth selection procedure provide strong evidence in favour of volatility spillovers from stock market returns to output growth in the case of the UK, the US and Japan. In particular, distant period volatility spillovers are found to be significant in these countries. These tests also suggest statistically significant causality-in-variance from output growth to real stock returns in the UK, the US and Italy. Contrary, neither (original) Hong's Q tests nor Cheung and Ng's S tests support the presence of volatility spillovers between industrial production and stock returns in any direction. This empirical evidence could be easily explained since according to our simulation results, both testing procedures demonstrate poor power performance when the bandwidth/lag truncation parameter is not accurately selected.

Appendix of Chapter 2

Table 1: Size of the Cheung and Ng (1996) and Hong (2001) causality-in-variance tests at the 5% and 10% levels against NULL1

	Level 5%							Level 10%						
	Q_{Dan}	Q_{TH}	Q_{QS}	Q_{Bart}	Q_{Trun}	Q_{Parz}	S	Q_{Dan}	Q_{TH}	Q_{QS}	Q_{Bart}	Q_{Trun}	Q_{Parz}	S
N														
							T = 200							
5	3.56	3.16	3.76	3.52	5.12	3.12	3.56	5.20	5.00	5.76	5.24	7.80	4.40	6.72
10	4.52	4.40	5.08	4.28	5.68	3.84	3.96	7.00	6.52	7.80	6.68	8.04	5.60	7.40
20	5.12	5.20	5.20	5.04	5.64	4.88	3.08	7.36	7.56	7.72	7.40	8.64	7.52	6.08
30	5.24	5.12	5.56	5.00	5.64	5.16	3.04	7.76	7.68	8.16	7.52	8.56	7.40	5.08
40	5.48	5.32	5.40	5.12	6.04	5.12	2.24	8.32	7.72	8.48	7.80	9.44	7.60	3.92
50	5.44	5.56	5.80	5.24	6.32	5.24	1.52	8.56	8.24	8.68	8.00	10.0	7.60	3.16
60	5.76	5.32	5.92	5.36	7.08	5.20	1.24	8.52	8.48	8.96	8.20	10.56	7.96	2.16
70	5.96	5.56	5.88	5.60	7.64	5.48	0.88	8.76	8.48	9.52	8.20	12.08	8.24	1.88
80	6.00	5.80	6.36	5.68	8.40	5.32	0.64	9.00	8.72	10.08	8.52	12.52	8.52	1.28
							T = 500							
5	3.28	3.00	3.44	3.00	4.16	3.20	2.72	5.20	5.08	5.52	5.20	7.00	4.64	6.28
10	3.72	3.52	4.12	3.40	5.20	3.40	3.96	5.92	5.88	6.56	5.80	8.92	5.44	8.00
20	4.88	4.16	5.08	4.16	5.80	3.92	3.44	7.60	7.08	8.36	7.16	9.48	6.44	7.72
30	5.12	5.00	5.00	4.92	6.12	4.36	3.84	8.76	8.08	8.64	7.68	9.96	7.40	7.32
40	5.12	5.16	5.20	4.76	5.56	4.92	3.28	8.64	8.40	9.12	8.08	9.88	7.76	6.52
50	5.32	5.12	5.56	4.84	6.32	5.16	3.48	9.24	8.64	9.44	8.56	10.24	8.44	6.32
60	5.48	5.08	5.92	5.04	6.32	5.04	2.80	9.36	9.12	10.04	9.04	10.84	8.36	5.32
70	5.84	5.40	6.04	5.36	5.96	5.08	2.28	9.52	9.52	10.32	9.12	10.44	8.72	4.44
80	5.84	5.64	6.04	5.52	6.68	5.12	1.92	10.16	9.44	10.36	9.80	10.96	8.96	4.20
							T = 1000							
5	2.68	2.76	2.88	2.44	4.16	2.44	2.44	4.32	3.96	5.32	4.24	7.48	4.04	6.72
10	3.56	3.28	3.68	3.08	4.48	2.76	3.40	6.16	5.80	6.96	5.60	8.00	4.88	7.12
20	4.36	3.96	4.64	4.00	5.52	3.48	3.80	7.20	7.32	7.36	6.84	8.60	6.60	7.48
30	5.00	4.72	4.68	4.48	5.12	4.12	3.60	7.80	7.44	8.16	7.20	8.68	7.32	7.12
40	4.80	4.84	4.68	4.56	4.64	4.64	3.36	8.32	7.64	8.28	7.64	8.40	7.32	6.28
50	4.76	4.88	4.72	4.64	5.52	4.84	3.36	8.28	8.32	8.60	7.80	9.28	7.60	6.88
60	4.76	4.72	4.88	4.52	5.44	4.68	3.36	8.32	8.32	8.68	7.96	9.16	7.72	6.12
70	4.68	4.76	4.96	4.48	5.04	4.80	3.16	8.56	8.36	8.64	8.16	9.04	8.32	5.84
80	5.04	4.2	5.20	4.60	5.32	4.84	3.04	8.72	8.60	8.64	8.32	9.32	8.28	5.36

Notes: NULL1: [No Granger causality-in-variance, high persistence in the volatility process]. N is the lag truncation/ bandwidth parameter . A Normal bivariate BEKK (1,1) process is simulated (2500 replications) for different. S and Q are the Cheung and Ng's and Hong's tests, respectively. Dan, TH , QS, Bar , Trun, and Par stand for Daniell, Tukey-Hanning , Quadratic-spectral, Bartlett, truncated, and Parzen kernels

Table 2: Size of the Cheung and Ng (1996) and Hong (2001) causality tests at the 5% and 10% levels against NULL2

	Level 5%							Level 10%						
N	Q_{Dan}	Q_{TH}	Q_{QS}	Q_{Bart}	Q_{Trun}	Q_{Parz}	S	Q_{Dan}	Q_{TH}	Q_{QS}	Q_{Bart}	Q_{Trun}	Q_{Parz}	S
								T = 200						
5	1.08	0.92	1.48	0.96	3.12	0.72	2.00	1.88	1.52	2.36	1.68	4.88	1.32	4.32
10	2.28	1.76	2.80	1.72	3.72	1.28	2.68	4.08	3.36	4.56	3.40	5.76	2.28	4.84
20	3.64	3.20	3.88	3.08	4.56	2.40	2.56	5.08	4.96	5.68	4.68	7.44	4.44	4.92
30	3.92	3.60	4.44	3.52	5.20	3.24	2.12	6.04	5.60	6.68	5.44	7.76	5.08	4.32
40	4.44	4.08	4.76	4.04	4.92	3.64	1.92	6.76	6.32	6.96	6.16	8.00	5.52	3.36
50	4.84	4.80	4.60	4.20	5.00	3.84	1.24	7.00	6.84	7.20	6.48	8.16	6.00	2.24
60	4.68	4.92	4.72	4.36	5.24	4.32	0.92	7.28	7.00	7.52	6.52	8.56	6.48	1.72
70	4.72	4.52	4.64	4.28	6.04	4.76	0.60	7.28	7.20	8.00	6.84	9.28	6.68	1.20
80	4.60	4.64	5.00	4.44	6.84	4.80	0.40	7.60	7.24	8.40	6.88	9.88	6.84	0.88
								T = 500						
5	0.72	0.48	0.96	0.56	2.48	0.36	1.44	1.28	0.96	2.16	1.08	4.40	0.64	3.72
10	1.64	1.32	2.00	1.24	3.48	0.72	2.72	3.28	2.92	4.04	2.44	6.12	1.68	5.36
20	2.80	2.48	3.08	2.36	4.40	1.96	2.52	4.72	4.60	5.24	4.04	7.16	3.84	5.84
30	3.20	3.00	3.32	2.84	4.48	2.56	2.52	5.64	5.20	6.04	4.64	8.28	4.76	5.76
40	3.40	3.36	4.00	3.08	5.08	3.00	2.64	6.20	5.76	6.96	5.28	8.84	4.84	5.64
50	3.80	3.48	4.52	3.24	5.76	3.20	2.64	6.64	6.16	7.80	6.04	8.80	5.52	5.64
60	4.48	4.00	4.88	3.60	5.12	3.36	2.60	7.64	6.72	8.20	6.44	9.72	5.84	4.56
70	4.64	4.32	5.04	4.08	5.96	3.52	2.36	8.16	7.28	8.44	6.96	9.76	6.12	4.76
80	4.80	4.56	5.28	4.24	7.04	3.80	2.36	8.12	7.84	8.72	7.60	10.48	6.56	4.60
								T = 1000						
5	0.68	0.44	1.08	0.44	2.92	0.08	2.08	1.12	0.72	1.80	0.88	4.92	0.12	4.40
10	1.96	1.48	2.24	1.32	3.60	0.92	2.48	2.92	2.56	3.96	2.44	6.32	1.56	5.56
20	3.24	2.72	3.52	2.64	4.24	2.16	3.12	5.44	4.72	6.08	4.40	7.36	3.48	6.36
30	3.84	3.60	4.16	3.00	4.04	3.00	2.88	6.52	5.88	6.52	5.28	6.88	5.20	5.84
40	4.16	3.88	3.96	3.32	4.80	3.44	3.28	6.72	6.52	6.84	5.84	7.68	5.60	6.16
50	3.76	3.92	3.96	3.40	4.84	3.48	3.16	6.64	6.60	6.76	6.04	9.00	6.24	6.40
60	3.96	3.96	4.48	3.48	5.48	3.92	3.72	6.76	6.60	7.40	6.32	8.92	6.52	6.48
70	4.16	3.88	4.68	3.60	5.64	3.92	3.28	7.00	6.64	7.88	6.56	9.24	6.44	6.00
80	4.48	4.00	4.76	3.76	6.12	3.92	3.24	7.64	6.84	8.04	6.64	9.68	6.76	6.16

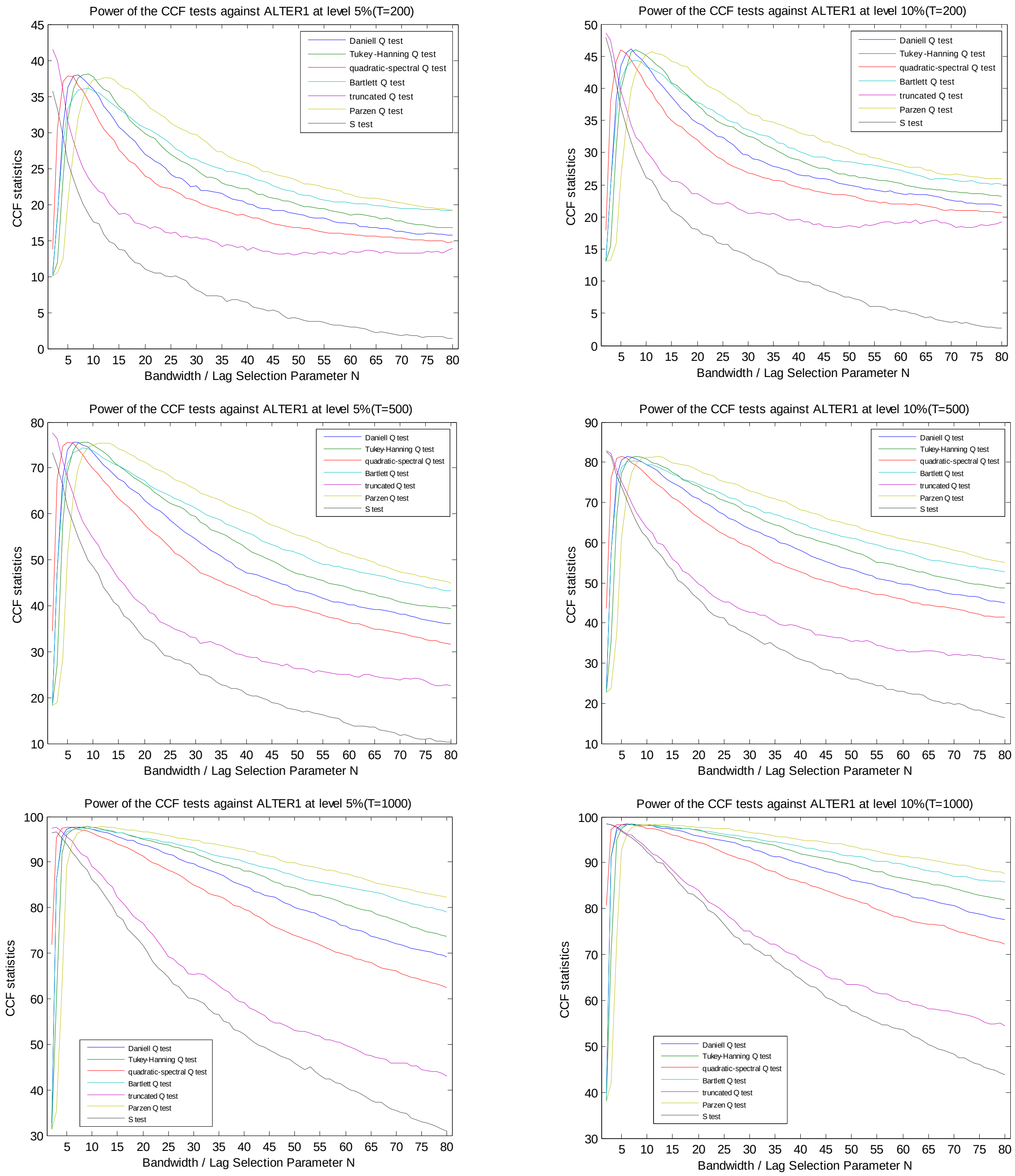
Notes: NULL2: [No Granger causality-in-variance, low persistence in the volatility process]. *N* is the lag truncation/ bandwidth parameter . A Normal bivariate BEKK (1,1) process is simulated (2500 replications) for different. S and Q are the Cheung and Ng's and Hong's tests, respectively. Dan, TH , QS, Bar , Trun, and Par stand for Daniell, Tukey-Hanning , Quadratic-spectral, Bartlett, truncated, and Parzen kernels

Table 3: size and power of the Comte and Lieberman ‘s (2000) and Hafner and Herwartz ‘s (2006) causality tests at the 5% and 10% levels respectively.

	Test Procedures			
	LR	LM	LR	LM
	Level 5%		Level 10%	
Sample sizes				
	NULL1: [No Granger causality-in-variance, high persistence in the volatility process]			
T = 200	0.1	11.04	0.1	18.24
T = 500	0.0	10.36	0.0	17.48
T = 1000	0.0	10.44	0.0	16.32
	NULL2: [No Granger causality-in-variance, low persistence in the volatility process]			
T = 200	0.0	11.60	0.1	20.00
T = 500	0.0	10.00	0.0	17.52
T = 1000	0.0	8.88	0.0	15.88
	ALTER1: [Unidirectional, short horizon Granger causality-in-variance, high persistence in the volatility process]			
T = 200	100	18.68	100	26.04
T = 500	100	23.04	100	29.68
T = 1000	100	25.08	100	32.60
	ALTER2: [Unidirectional, short horizon Granger causality-in-variance, low persistence in the volatility process]			
T = 200	100	33.80	100	45.80
T = 500	100	67.00	100	80.04
T = 1000	100	93.88	100	97.36
	ALTER3: [Unidirectional, long horizon Granger causality-in-variance, high persistence in the volatility process]			
T = 200	5.30	7.80	5.30	10.48
T = 500	8.10	8.88	8.10	10.92
T = 1000	2.70	17.20	2.70	20.40
	ALTER4: [Unidirectional, long horizon Granger causality-in-variance, low persistence in the volatility process]			
T = 200	11.00	9.36	11.00	16.24
T = 500	23.2	8.12	23.2	14.28
T = 1000	36.20	7.48	36.40	13.76

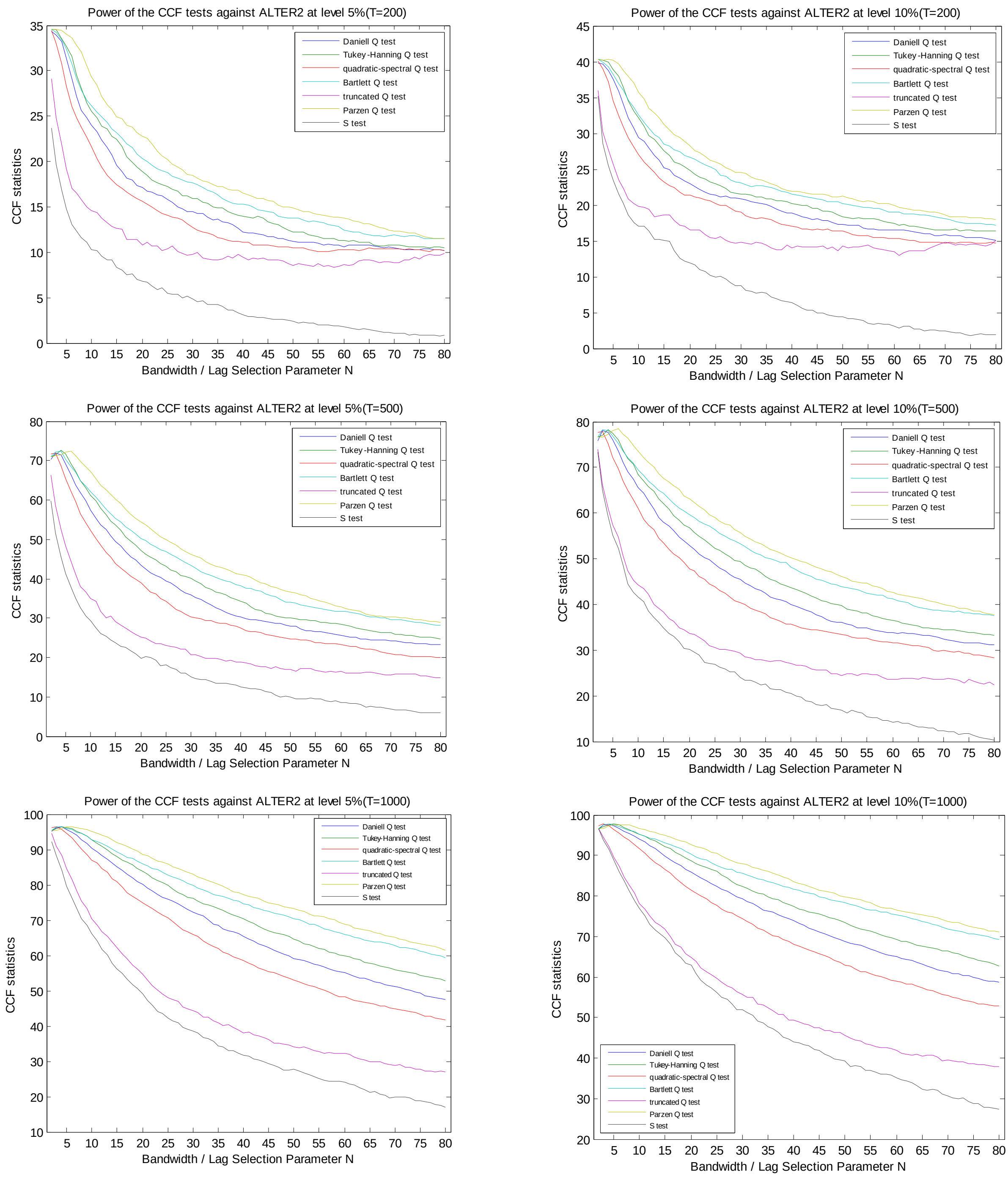
Note: LR and LM stand for Comte and Lieberman’s (2000) Likelihood Ratio and Hafner and Herwartz’s (2006) Lagrange Multiplier tests respectively (see section 2). A Normal bivariate BEKK (1,1) process is simulated (2500 replications) for In the NULL1, NULL2, ALTER1, ALTER2, while for ALTER3 and ALTER4 a Normal bivariate BEKK (12,12) process is simulated**.** 1000 replications are used in the LR test simulations.

Panel 1: Power of the CCF based tests against ALTER1 at levels of significance 5% and 10%



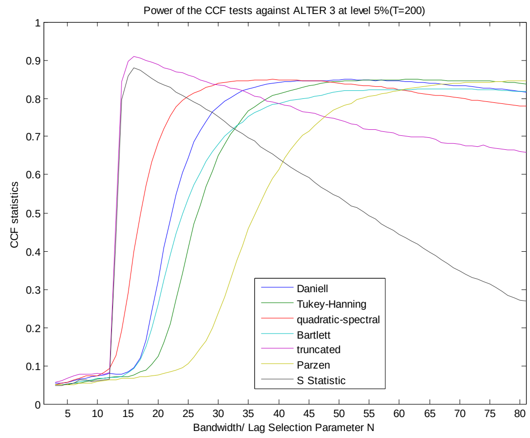
Notes: ALTER1: [Unidirectional, short horizon Granger causality-in-variance, high persistence in the volatility process]. A Normal bivariate BEKK (1,1) process is simulated (2500 replications) for different sample sizes T.

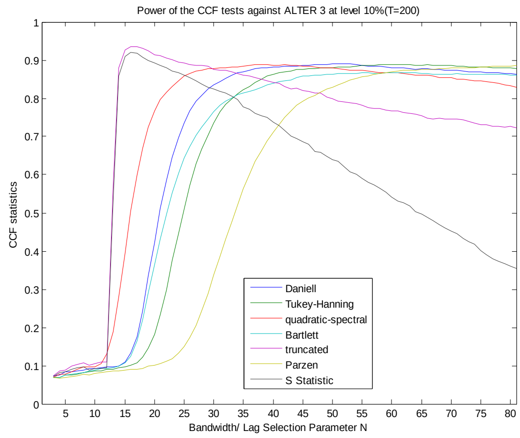
Panel II: Power of the CCF based tests against ALTER2 at levels of significance 5% and 10%

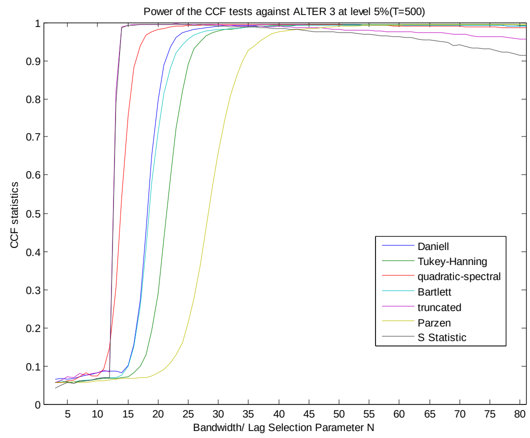


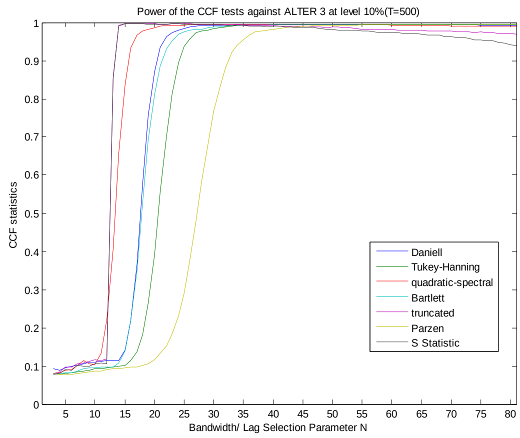
Notes: ALTER2: [Unidirectional, short horizon Granger causality-in-variance, low persistence in the volatility process]. A Normal bivariate BEKK (1,1) process is simulated (2500 replications) for different sample sizes *T*.

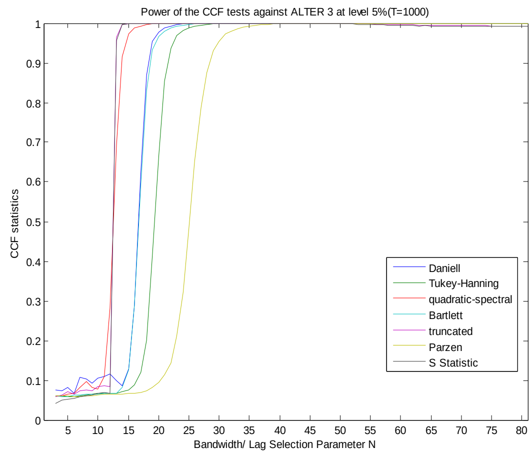
Panel III: Power of the CCF based tests against ALTER3 at levels of significance 5% and 10%

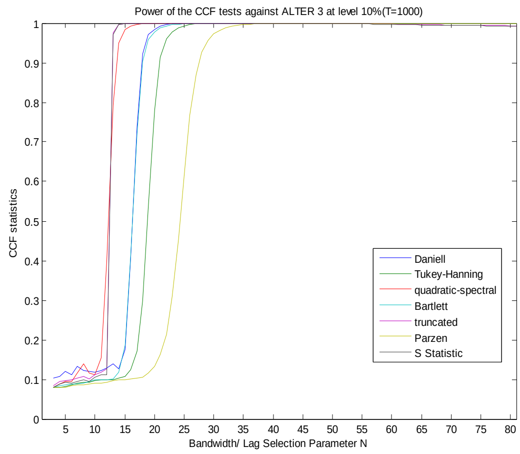






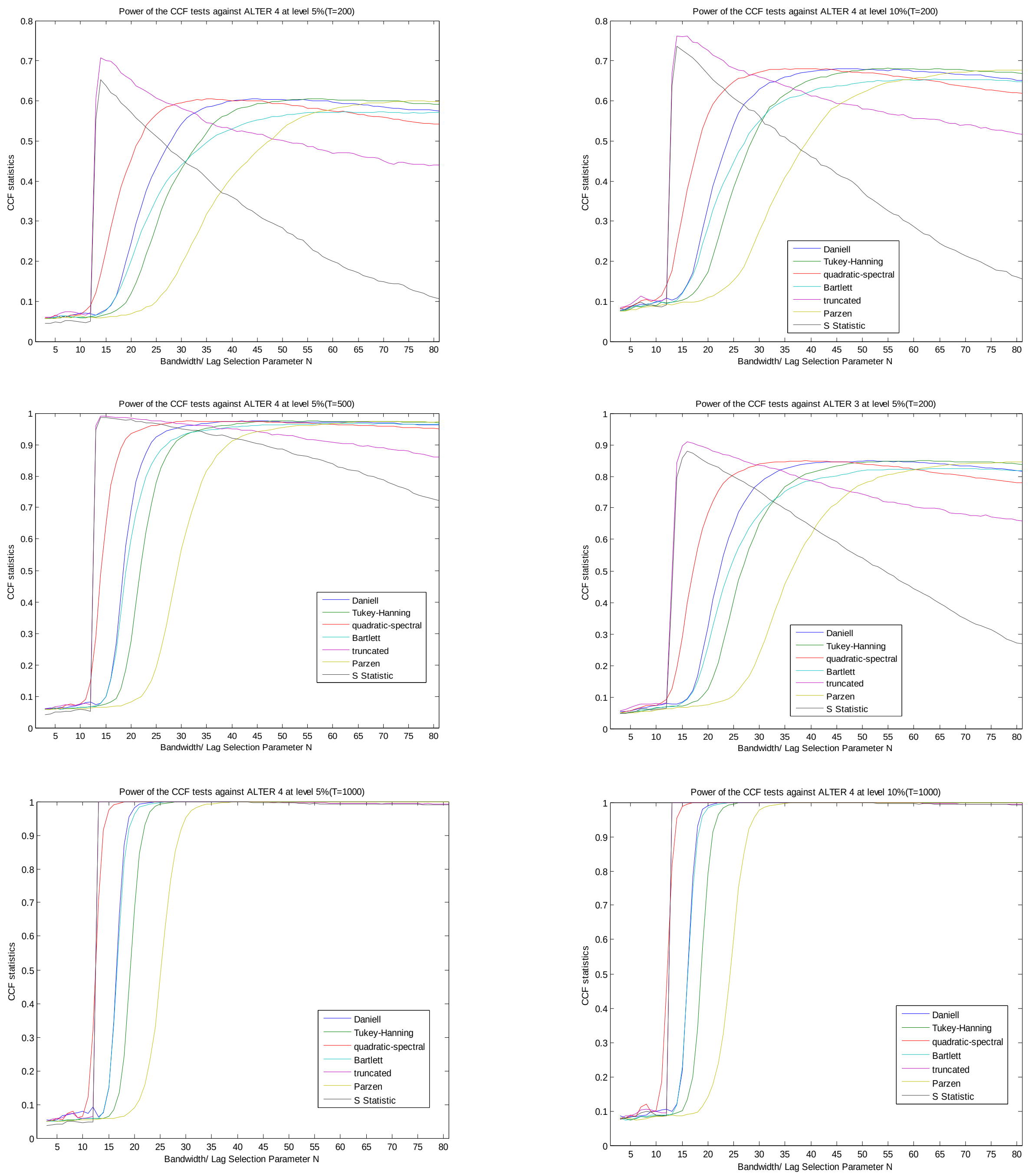






Notes: ALTER3: [Unidirectional, long horizon Granger causality-in-variance, high persistence in the volatility process]. A Normal bivariate BEKK (12,12) process is simulated (2500 replications) for different T .

Panel IV: Power of the CCF based tests against ALTER4 at levels of significance 5% and 10



Notes: ALTER4: [Unidirectional, long horizon Granger causality-in-variance, low persistence in the volatility process]. A Normal bivariate BEKK (12,12) process is simulated (2500 replications) for different sample sizes T .

Table 4: The empirical size of the Hong (2001) tests with the "naïve" automatic optimal bandwidth selection procedure at the nominal levels of 5% and 10%

	Levels of Significance			
	5%		10%	
	EVD	N(0,1)	EVD	N(0,1)
	NULL1: [No Granger causality-in-variance, high persistence in the volatility process]			
		T = 200		
Daniell	1.92	15.40	15.04	21.88
Tukey-Hanning	1.88	14.92	14.48	20.88
Quadratic-spectral	1.96	16.52	15.96	22.96
Bartlett	1.64	13.84	13.48	20.00
Truncated	3.28	27.32	26.40	37.72
Parzen	1.68	13.60	13.20	19.16
		T = 500		
Daniell	1.92	16.28	15.84	23.44
Tukey-Hanning	1.92	15.56	15.08	22.84
Quadratic-spectral	1.88	16.68	16.20	24.56
Bartlett	1.84	14.20	13.72	21.00
Truncated	2.64	29.44	28.60	41.16
Parzen	1.88	14.24	13.72	20.96
		T = 1000		
Daniell	5.88	19.36	19.00	25.96
Tukey-Hanning	5.84	18.56	18.16	25.36
Quadratic-spectral	5.80	19.56	19.08	26.64
Bartlett	5.68	17.56	17.36	24.20
Truncated	5.64	31.40	30.60	42.56
Parzen	5.80	17.56	17.12	23.52
	NULL2: [No Granger causality-in-variance, low persistence in the volatility process]			
		T = 200		
Daniell	1.32	12.56	12.28	18.60
Tukey-Hanning	1.12	12.24	11.84	18.28
Quadratic-spectral	1.20	13.24	13.04	19.88
Bartlett	1.04	11.24	10.96	16.68
Truncated	2.08	23.76	23.16	32.52
Parzen	1.08	11.20	10.92	16.56
		T = 500		
Daniell	3.04	13.12	12.76	18.44
Tukey-Hanning	2.96	12.44	11.84	17.80
Quadratic-spectral	3.00	13.84	13.36	19.96
Bartlett	2.92	11.04	10.44	16.28
Truncated	3.84	26.24	25.60	35.96
Parzen	2.92	10.88	10.36	15.92
		T = 1000		
Daniell	6.80	17.32	17.20	23.20
Tukey-Hanning	6.76	16.72	16.44	22.00
Quadratic-spectral	6.72	18.08	17.72	23.72
Bartlett	6.68	15.48	15.40	20.56
Truncated	7.24	30.64	30.00	40.92
Parzen	6.64	15.60	15.44	20.16

Notes: The optimal bandwidth parameter is automatically selected based on the algorithm: $\hat{N}^* = \arg \max_{2<N\leq N_k} Q(N)$ where N_k is set equal to 100. Normal bivariate BEKK (1,1) process is simulated (2500 replications) for different sample sizes *T*. Critical values are calculated based on the extreme value distribution (EVD) and the standard normal.

Table 5: The empirical size of the Hong (2001) tests with the optimal bandwidth estimated using a Dumbgen-type estimator at the nominal levels of 5% and 10%

	Levels of significance					
	5%			10%		
	Λ_1	Λ_2	Λ_3	Λ_1	Λ_2	Λ_3
	NULL1: [no Granger causality-in-variance; high persistence in the volatility process]					
	T = 200					
Daniell	4.680	4.680	4.680	7.640	7.600	7.480
Tukey-Hanning	4.920	4.960	4.760	7.520	7.640	7.600
Quadratic-spectral	4.760	4.840	4.880	7.560	7.640	7.520
Bartlett	5.120	5.120	5.120	7.720	7.920	7.840
Truncated	2.680	2.680	2.880	4.840	5.000	5.200
Parzen	5.040	5.080	5.040	7.160	7.280	7.040
	T = 500					
Daniell	5.200	5.240	5.000	8.280	8.400	8.240
Tukey-Hanning	5.320	5.520	5.480	8.320	8.400	8.480
Quadratic-spectral	4.880	5.040	4.920	8.600	8.520	8.160
Bartlett	5.600	5.720	5.480	8.400	8.480	8.640
Truncated	2.440	2.600	2.800	5.320	5.520	5.680
Parzen	4.800	4.760	4.840	7.400	7.640	7.400
	T = 1000					
Daniell	7.720	7.640	7.720	11.240	11.160	11.080
Tukey-Hanning	8.000	8.080	8.040	11.240	11.320	11.320
Quadratic-spectral	7.200	7.240	7.160	11.320	11.320	11.280
Bartlett	8.480	8.480	8.480	11.720	11.960	11.560
Truncated	4.760	4.720	4.880	7.320	7.480	7.920
Parzen	8.520	8.520	8.360	11.440	11.520	11.520
	NULL2: [no Granger causality-in-variance; low persistence in the volatility process]					
	T = 200					
Daniell	4.040	4.160	4.040	6.360	6.360	6.400
Tukey-Hanning	4.000	3.960	4.000	6.520	6.520	6.560
Quadratic-spectral	3.840	3.960	3.800	6.920	6.800	6.600
Bartlett	4.240	4.240	4.120	6.560	6.600	6.600
Truncated	2.360	2.400	2.640	4.440	4.560	4.800
Parzen	4.040	4.080	4.040	6.000	6.080	6.040
	T = 500					
Daniell	4.840	4.880	4.760	7.120	7.200	7.160
Tukey-Hanning	4.880	4.880	4.720	7.280	7.280	7.080
Quadratic-spectral	4.800	4.880	4.880	7.560	7.600	7.840
Bartlett	5.080	5.080	4.960	7.320	7.400	7.320
Truncated	2.800	3.080	3.160	4.640	4.600	4.720
Parzen	4.640	4.640	4.640	6.920	6.880	6.840
	T = 1000					
Daniell	9.000	9.000	9.000	12.080	12.080	11.920
Tukey-Hanning	9.440	9.400	9.360	11.760	11.920	11.720
Quadratic-spectral	8.800	8.800	8.680	11.880	11.800	11.680
Bartlett	9.480	9.400	9.280	11.720	11.800	11.760
Truncated	6.800	7.120	7.040	9.000	9.040	9.000
Parzen	8.920	9.000	9.040	11.120	11.280	11.240

Notes: A Normal bivariate BEKK (1,1) process is simulated (2500 replications) for different T . The optimal bandwidth parameter is estimated using a Dumbgen-type estimator. Λ_1, Λ_2 and Λ_3 are the semi-norms used in the estimation.

Table 6: Power of the of the Hong (2001) tests with with the “naïve” automatic optimal bandwidth selection procedure at the nominal levels of 5% and 10% under ALTER1 and ALTER2.

	Levels of Significance			
	5%		10%	
	EVD	N(0,1)	EVD	N(0,1)
	ALTER1: [short horizon Granger causality-in-variance, high persistence in the volatility process]			
	T = 200			
Daniell	14.24	46.20	46.52	54.76
Tukey-Hanning	14.16	46.04	46.48	54.04
Quadratic-spectral	14.16	46.60	47.32	54.92
Bartlett	12.96	43.36	44.12	52.16
Truncated	20.60	57.40	58.80	66.24
Parzen	13.64	44.84	45.12	52.52
	T = 500			
Daniell	43.48	80.08	80.12	86.16
Tukey-Hanning	43.44	80.12	80.08	85.88
Quadratic-spectral	43.24	80.40	80.32	86.08
Bartlett	40.52	78.72	78.80	84.56
Truncated	53.52	86.56	86.00	91.40
Parzen	42.56	79.60	79.32	85.36
	T = 1000			
Daniell	85.76	98.16	98.12	98.80
Tukey-Hanning	85.56	98.16	98.16	98.72
Quadratic-spectral	85.60	98.20	98.20	98.72
Bartlett	83.96	98.08	98.04	98.64
Truncated	90.60	98.76	98.64	99.08
Parzen	85.24	98.12	98.08	98.68
	ALTER2: [short horizon Granger causality-in-variance, low persistence in the volatility process]			
	T = 200			
Daniell	16.08	42.28	43.44	50.52
Tukey-Hanning	16.28	42.24	43.20	50.40
Quadratic-spectral	15.96	42.64	43.84	50.44
Bartlett	16.12	41.40	42.44	49.52
Truncated	11.84	44.44	46.04	54.44
Parzen	16.04	41.52	42.44	49.36
	T = 500			
Daniell	43.20	78.76	76.24	84.32
Tukey-Hanning	43.32	79.32	76.44	84.60
Quadratic-spectral	42.60	78.60	76.04	84.00
Bartlett	43.12	78.92	76.04	84.32
Truncated	33.00	75.40	72.64	83.60
Parzen	43.20	79.12	76.08	84.24
	T = 1000			
Daniell	79.80	97.44	94.48	98.16
Tukey-Hanning	80.00	97.44	94.52	98.40
Quadratic-spectral	79.68	97.40	94.44	98.28
Bartlett	79.68	97.40	94.32	98.36
Truncated	72.08	96.20	94.36	98.12
Parzen	79.92	97.44	94.36	98.36

Notes: The optimal bandwidth parameter is automatically selected based on the algorithm: $\hat{N}^* = \arg \max_{2<N\leq N_k} Q(N)$ where N_k is set equal to 100. Normal bivariate BEKK (1,1) process is simulated (2500 replications) for different sample sizes *T*. Critical values are calculated based on the extreme value distribution (EVD) and the standard normal.

Table 7: Empirical power of the Hong (2001) tests with with the "naïve" automatic optimal bandwidth selection procedure at the nominal levels of 5% and 10% under ALTER3 and ALTER4.

	Levels of Significance			
	5%		10%	
	EVD	N(0,1)	EVD	N(0,1)
	ALTER3: [long horizon Granger causality-in-variance, high persistence in the volatility process]			
		T = 200		
Daniell	50.60	86.20	86.00	89.84
Tukey-Hanning	47.88	85.08	84.88	88.68
quadratic-spectral	51.28	86.52	86.32	90.00
Bartlett	41.08	82.48	81.88	87.00
Truncated	73.08	94.48	94.80	96.44
Parzen	33.76	78.68	78.20	84.08
		T = 500		
Daniell	96.44	99.48	99.48	99.56
Tukey-Hanning	96.28	99.48	99.52	99.52
Quadratic-spectral	96.44	99.52	99.56	99.52
Bartlett	94.48	99.32	99.36	99.52
Truncated	99.12	99.68	99.68	99.76
Parzen	92.24	99.08	99.16	99.32
		T = 1000		
Daniell	99.84	99.88	99.88	99.88
Tukey-Hanning	99.80	99.88	99.88	99.88
Quadratic-spectral	99.80	99.88	99.88	99.88
Bartlett	99.76	99.88	99.88	99.88
Truncated	99.88	99.92	99.92	99.96
Parzen	99.48	99.88	99.88	99.88
	ALTER4: [long horizon Granger causality-in-variance, low persistence in the volatility process]			
		T = 200		
Daniell	23.48	64.32	64.08	71.20
Tukey-Hanning	22.44	62.68	62.44	69.88
Quadratic-spectral	24.12	65.56	65.48	72.36
Bartlett	19.24	58.80	58.52	66.84
Truncated	41.24	80.44	80.56	85.36
Parzen	16.04	55.96	56.00	63.96
		T = 500		
Daniell	81.80	97.80	97.68	98.92
Tukey-Hanning	81.12	97.64	97.52	98.60
Quadratic-spectral	82.08	97.92	97.92	98.88
Bartlett	76.00	96.56	96.40	97.92
Truncated	92.92	99.60	99.56	99.84
Parzen	72.44	95.76	95.80	97.68
		T = 1000		
Daniell	99.60	99.92	99.92	99.92
Tukey-Hanning	99.60	99.92	99.92	99.92
Quadratic-spectral	99.60	99.92	99.92	99.92
Bartlett	99.28	99.92	99.92	99.92
Truncated	99.92	99.96	99.96	100.00
Parzen	99.08	99.92	99.92	99.92

Notes: The optimal bandwidth parameter is automatically selected based on the algorithm: $\hat{N}^* = \arg\max_{2<N\leq N_k} Q(N)$ where N_k is set equal to 100. Normal bivariate BEKK (12,12) process is simulated (2500 replications) for different sample sizes *T*. Critical values are calculated based on the extreme value distribution (EVD) and the standard normal.

Table 8: The empirical power of the Hong (2001) tests with the optimal bandwidth estimated using a Dumbgen-type estimator at the nominal levels of 5% and 10% under ALTER1 and ALTER2.

	Levels of significance					
	5%			10%		
	Λ_1	Λ_2	Λ_3	Λ_1	Λ_2	Λ_3
	ALTER1: short horizon Granger causality-in-variance. high persistence in the volatility process					
	T = 200					
Daniell	23.760	23.840	24.080	30.320	30.560	31.080
Tukey-Hanning	25.000	25.240	25.720	32.920	32.880	33.240
Quadratic-spectral	21.640	21.800	21.720	28.280	28.400	28.680
Bartlett	26.680	26.880	27.360	34.000	34.120	34.200
Truncated	14.320	14.160	14.480	19.720	19.960	20.560
Parzen	28.320	28.560	28.800	35.240	35.240	35.640
	T = 500					
Daniell	56.560	56.800	57.280	65.920	66.000	66.280
Tukey-Hanning	60.040	60.320	60.920	68.520	68.560	68.880
Quadratic-spectral	51.120	51.080	51.920	62.040	61.920	61.960
Bartlett	61.840	62.040	61.920	70.280	70.440	70.640
Truncated	38.080	38.280	38.720	47.760	47.760	48.840
Parzen	64.120	64.440	64.920	72.760	72.880	73.200
	T = 1000					
Daniell	91.320	91.360	91.640	94.520	94.600	94.720
Tukey-Hanning	92.880	92.880	92.960	95.560	95.600	95.800
Quadratic-spectral	88.240	88.240	88.240	92.400	92.400	92.280
Bartlett	93.840	93.880	93.920	96.000	95.920	96.120
Truncated	74.920	75.160	75.720	83.280	83.440	84.120
Parzen	94.960	95.040	95.120	96.760	96.760	97.000
	ALTER2: short horizon Granger causality-in-variance. low persistence in the volatility process					
	T = 200					
Daniell	15.360	15.400	15.360	21.160	21.040	21.240
Tukey-Hanning	16.920	16.920	17.040	22.240	22.240	22.280
Quadratic-spectral	13.360	13.440	13.400	19.840	19.880	19.840
Bartlett	18.520	18.440	18.360	24.080	24.040	24.080
Truncated	8.040	8.040	8.080	12.320	12.400	12.480
Parzen	19.640	19.680	19.800	25.480	25.440	25.640
	T = 500					
Daniell	39.840	39.880	40.000	48.160	48.280	48.520
Tukey-Hanning	42.280	42.320	42.560	52.040	51.920	51.720
Quadratic-spectral	34.440	34.400	34.360	44.000	44.080	44.040
Bartlett	45.960	46.000	46.000	54.600	54.560	54.720
Truncated	22.480	22.480	22.960	30.320	30.720	31.520
Parzen	48.480	48.520	48.720	57.920	58.040	58.160
	T = 1000					
Daniell	73.920	73.960	74.040	80.200	80.320	80.320
Tukey-Hanning	77.440	77.440	77.560	83.120	83.200	83.160
Quadratic-spectral	69.040	69.080	69.160	76.480	76.480	76.560
Bartlett	79.960	80.040	80.040	85.000	84.960	85.000
Truncated	52.160	52.160	52.840	62.400	62.680	63.160
Parzen	82.360	82.400	82.480	87.200	87.200	87.240

Notes: A Normal bivariate BEKK (1,1) process is simulated (2500 replications) for different T . The optimal bandwidth parameter is estimated using a Dumbgen-type estimator. Λ_1, Λ_2 and Λ_3 are the semi-norms used in the estimation.

Table 9: The empirical power of the Hong (2001) tests with the optimal bandwidth estimated using a Dumbgen-type estimator at the nominal levels of 5% and 10% under ALTER3 and ALTER4.

	Levels of significance						
	5%				10%		
	Λ_1	Λ_2	Λ_3		Λ_1	Λ_2	Λ_3
	ALTER3: long horizon Granger causality-in-variance. high persistence in the volatility process						
				T = 200			
Daniell	69.120	70.440	68.760		77.360	78.360	77.280
Tukey-Hanning	48.640	51.800	47.520		59.080	62.320	57.080
Quadratic-spectral	80.400	80.320	80.080		85.440	85.320	85.080
Bartlett	58.040	60.960	57.840		67.760	70.360	67.640
Truncated	35.640	34.440	31.640		42.040	40.360	37.840
Parzen	14.160	15.600	14.160		20.320	22.240	19.920
				T = 500			
Daniell	98.360	98.480	98.320		99.040	99.040	98.960
Tukey-Hanning	93.840	95.040	93.080		96.720	97.280	96.040
Quadratic-spectral	99.320	99.320	99.280		99.440	99.400	99.360
Bartlett	96.840	97.440	96.840		98.160	98.440	98.120
Truncated	65.120	63.600	60.040		69.120	67.720	64.560
Parzen	31.480	37.040	28.280		42.960	48.920	38.800
				T = 1000			
Daniell	99.840	99.840	99.840		99.840	99.840	99.840
Tukey-Hanning	99.520	99.720	99.360		99.800	99.800	99.720
Quadratic-spectral	99.880	99.880	99.880		99.880	99.880	99.880
Bartlett	99.800	99.800	99.800		99.840	99.840	99.840
Truncated	82.720	81.400	78.040		84.560	83.360	80.160
Parzen	74.440	79.080	65.360		84.200	86.480	77.000
	ALTER4: long horizon Granger causality-in-variance. low persistence in the volatility process						
				T = 200			
Daniell	46.200	47.320	46.160		56.800	57.480	56.640
Tukey-Hanning	33.680	35.680	33.280		43.560	45.320	43.240
Quadratic-spectral	55.360	55.160	54.720		63.360	63.640	62.840
Bartlett	38.680	39.960	38.400		48.280	49.880	48.080
Truncated	13.640	13.440	12.720		19.280	19.040	18.400
Parzen	12.560	13.760	12.520		18.480	20.120	18.360
				T = 500			
Daniell	93.240	93.600	93.240		95.600	95.880	95.600
Tukey-Hanning	83.240	85.600	82.600		89.640	91.280	89.280
Quadratic-spectral	96.160	96.120	96.000		97.840	97.800	97.760
Bartlett	88.720	89.720	88.640		92.560	93.320	92.560
Truncated	17.400	16.800	15.600		20.760	20.120	18.840
Parzen	28.800	33.360	27.080		40.480	45.480	37.120
				T = 1000			
Daniell	99.920	99.920	99.920		99.920	99.920	99.920
Tukey-Hanning	99.560	99.720	99.520		99.880	99.880	99.840
Quadratic-spectral	99.920	99.920	99.920		99.920	99.920	99.920
Bartlett	99.840	99.880	99.840		99.880	99.920	99.880
Truncated	12.760	12.080	11.160		15.720	15.040	14.120
Parzen	71.920	76.920	64.440		82.640	85.640	76.760

Notes: A Normal bivariate BEKK (12,12) process is simulated (2500 replications) for different T . The optimal bandwidth parameter is estimated using a Dumbgen-type estimator. Λ_1, Λ_2 and Λ_3 are the semi-norms used in the estimation.

Table 10: The empirical size of the Hong (2001) tests with the optimal bandwidth estimated using the cross-validation method at the nominal levels of 5% and 10%

	Levels of Significance	
	5%	10%
	NULL1: [No Granger causality-in-variance, high persistence in the volatility process]	
	T = 200	
Daniell	5.44	6.88
Tukey-Hanning	5.36	7.00
Quadratic-spectral	5.36	7.36
Bartlett	5.28	7.08
Truncated	5.96	8.80
Parzen	4.84	6.76
	T = 500	
Daniell	5.44	7.28
Tukey-Hanning	5.40	7.44
Quadratic-spectral	5.48	7.52
Bartlett	5.40	7.44
Truncated	5.68	8.40
Parzen	5.44	7.44
	T = 1000	
Daniell	9.88	11.88
Tukey-Hanning	9.84	11.96
Quadratic-spectral	9.96	12.04
Bartlett	9.88	11.96
Truncated	9.20	12.04
Parzen	9.92	11.96
	NULL2: [No Granger causality-in-variance, low persistence in the volatility process]	
	T = 200	
Daniell	3.60	5.36
Tukey-Hanning	3.76	5.64
Quadratic-spectral	3.80	5.24
Bartlett	3.76	5.60
Truncated	3.68	5.40
Parzen	3.80	5.64
	T = 500	
Daniell	5.40	7.32
Tukey-Hanning	5.68	7.52
Quadratic-spectral	5.28	7.20
Bartlett	5.68	7.52
Truncated	4.32	6.56
Parzen	5.68	7.52
	T = 1000	
Daniell	8.68	9.60
Tukey-Hanning	8.68	9.76
Quadratic-spectral	8.72	9.60
Bartlett	8.68	9.76
Truncated	9.52	11.08
Parzen	8.68	9.76

Notes: A Normal bivariate BEKK (12,12) process is simulated (2500 replications) for different T . The optimal bandwidth parameter is estimated using cross-validation in terms of a kernel regression.

Table 11: The empirical power of the Hong (2001) tests with the optimal bandwidth estimated using the cross-validation method at the nominal levels of 5% and 10% under ALTER1 and ALTER2

	Levels of Significance	
	5%	10%
	ALTER1: [short horizon Granger causality-in-variance, high persistence in the volatility process]	
	T = 200	
Daniell	11.48	14.32
Tukey-Hanning	10.84	13.76
Quadratic-spectral	14.40	19.04
Bartlett	11.08	14.12
Truncated	42.12	48.76
Parzen	10.44	13.28
	T = 500	
Daniell	19.76	24.32
Tukey-Hanning	19.08	23.28
Quadratic-spectral	34.80	44.16
Bartlett	19.20	23.52
Truncated	77.52	82.64
Parzen	18.76	23.04
	T = 1000	
Daniell	34.20	40.44
Tukey-Hanning	32.76	39.16
Quadratic-spectral	73.20	81.76
Bartlett	32.84	39.16
Truncated	97.72	98.52
Parzen	32.68	39.00
	ALTER2: [short horizon Granger causality-in-variance, low persistence in the volatility process]	
	T = 200	
Daniell	87.56	89.84
Tukey-Hanning	87.28	89.80
Quadratic-spectral	88.64	90.60
Bartlett	87.44	89.84
Truncated	86.40	89.32
Parzen	87.32	89.76
	T = 500	
Daniell	68.28	73.56
Tukey-Hanning	68.76	74.12
Quadratic-spectral	69.48	74.76
Bartlett	68.76	74.08
Truncated	64.08	71.40
Parzen	68.84	74.08
	T = 1000	
Daniell	100	100
Tukey-Hanning	100	100
Quadratic-spectral	100	100
Bartlett	100	100
Truncated	100	100
Parzen	100	100

Notes: A Normal bivariate BEKK (12,12) process is simulated (2500 replications) for different T . The optimal bandwidth parameter is estimated using cross-validation in terms of a kernel regression.

Table 12: The empirical power of the Hong (2001) tests with the optimal bandwidth estimated using the cross-validation method at the nominal levels of 5% and 10% under ALTER3 and ALTER4

	Levels of Significance	
	5%	10%
	ALTER3: [long horizon Granger causality-in-variance, high persistence in the volatility process]	
	T = 200	
Daniell	79.36	83.24
Tukey-Hanning	78.56	82.24
Quadratic-spectral	78.64	82.88
Bartlett	76.00	80.36
Truncated	69.52	75.68
Parzen	72.36	77.24
	T = 500	
Daniell	95.52	95.76
Tukey-Hanning	94.96	95.16
Quadratic-spectral	95.28	95.52
Bartlett	95.00	95.40
Truncated	94.12	94.64
Parzen	94.36	94.60
	T = 1000	
Daniell	97.12	97.28
Tukey-Hanning	96.88	96.96
Quadratic-spectral	97.20	97.32
Bartlett	97.00	97.04
Truncated	97.12	97.20
Parzen	96.76	96.80
	ALTER4: [long horizon Granger causality-in-variance, low persistence in the volatility process]	
	T = 200	
Daniell	39.84	45.72
Tukey-Hanning	38.36	43.68
Quadratic-spectral	40.96	46.60
Bartlett	37.00	42.80
Truncated	38.28	43.56
Parzen	32.16	38.80
	T = 500	
Daniell	95.56	95.80
Tukey-Hanning	95.00	95.20
Quadratic-spectral	95.32	95.56
Bartlett	95.04	95.44
Truncated	94.12	94.68
Parzen	94.40	94.64
	T = 1000	
Daniell	96.92	97.04
Tukey-Hanning	96.40	96.60
Quadratic-spectral	97.16	97.28
Bartlett	96.68	96.84
Truncated	97.20	97.28
Parzen	96.00	96.32

Notes: A Normal bivariate BEKK (12,12) process is simulated (2500 replications) for different T . The optimal bandwidth parameter is estimated using cross-validation in terms of a kernel regression.

Table 13: Testing the null hypothesis that real stock market returns do not Granger cause-in-variance the industrial production growth in Italy, Japan, the United Kingdom, and the United States.

	Countries							
	Italy		Japan		United Kingdom		United States	
Kernel	B_q	$Q(N=12)$	B_q	$Q(N=12)$	B_q	$Q(N=12)$	B_q	$Q(N=12)$
Daniell	-0.1389	-0.3319	**3.8326**	-0.3693	**3.8966**	0.6821	**3.8869**	**1.9154**
	(0.1360)	(0.1888)	**(0.0422)**	(0.1863)	**(0.0414)**	(0.1581)	**(0.0416)**	**(0.0277)**
Tukey Hanning	-0.1470	-0.2241	**3.7935**	-0.6653	**4.0506**	1.1633	**3.9032**	**2.3636**
	(0.1363)	(0.1945)	**(0.0427)**	(0.1599)	**(0.0396)**	(0.1014)	**(0.0414)**	**(0.0090)**
Quadratic Spectral	-0.1501	-0.1770	**4.0964**	-0.4248	**3.9010**	0.1750	**3.8931**	**2.0945**
	(0.1364)	(0.1964)	**(0.0390)**	(0.1823)	**(0.0414)**	(0.1964)	**(0.0415)**	**(0.0181)**
Bartlett	-0.1686	-0.2559	**3.7474**	-0.6355	**4.0506**	1.2067	**3.8948**	**2.4299**
	(0.1372)	(0.1930)	**(0.0433)**	(0.1630)	**(0.0396)**	(0.0963)	**(0.0415)**	**(0.0076)**
	$Q(\hat{N}^*_{CV})$	$Q(\hat{N}^*_{DB})$	$Q(\hat{N}^*_{CV})$	$Q(\hat{N}^*_{DB})$	$Q(\hat{N}^*_{CV})$	$Q(\hat{N}^*_{DB})$	$Q(\hat{N}^*_{CV})$	$Q(\hat{N}^*_{DB})$
Daniell	-0.1597	-1.2054	**3.4873**	**3.7007**	-1.2627	-2.2635	1.3431	1.1087
	(0.5634)	(0.8860)	**(0.0000)**	**(0.0000)**	(0.8967)	(0.9882)	(0.0896)	(0.1338)
	[20]	[100]	**[75]**	**[84]**	[39]	[100]	[45]	[51]
Tukey-Hanning	-0.1470	-1.0167	**2.8739**	**3.5127**	-0.8418	-1.9000	**1.8519**	**1.6519**
	(0.5584)	(0.8454)	**(0.0020)**	**(0.0002)**	(0.8000)	(0.9713)	**(0.0320)**	**(0.0493)**
	[20]	[100]	**[75]**	**[99]**	[39]	[100]	**[45]**	**[51]**
Quadratic Spectral	-0.3888	-1.6103	**3.6687**	**3.7228**	-1.5381	-2.5292	0.8185	0.5616
	(0.6513)	(0.9463)	**(0.0000)**	**(0.0000)**	(0.9380)	(0.9943)	(0.2065)	(0.2872)
	[20]	[100]	**[75]**	**[99]**	[39]	[100]	[45]	[51]
Bartlett	-0.1996	-1.1022	**2.7034**	**3.1141**	-0.7900	-1.8572	**1.7734**	1.5833
	(0.5791)	(0.8648)	**(0.0034)**	**(0.0009)**	(0.7852)	(0.9684)	**(0.0381)**	(0.0567)
	[20]	[100]	**[75]**	**[99]**	[39]	[100]	**[45]**	[51]
LM	**32.7259**		0.7327		3.7457		**7.3128**	
	(0.0007)		(0.6933)		(0.1537)		**(0.0258)**	
LR	3.6663		-1.9174		**17.7346**		3.0253	
	(0.1599)		(1.000)		**(0.0001)**		(0.2203)	
S (*N* = 12)	12.7348		10.4654		7.8440		**31.6969**	
	(0.3886)		(0.5752)		(0.7972)		**(0.0015)**	
*S** (*N* = 12)	12.9591		10.6443		7.8746		**32.4309**	
	(0.3720)		(0.5596)		(0.7949)		**(0.0012)**	

Notes: *p*-values are in parentheses. Numbers in brackets are the optimal bandwidths computed according to cross-validation method ($\hat{N}^*_{CV}$) and Dubgen type estimators ($\hat{N}^*_{DB}$). $B_q = \max_{2<N\leq q} Q(N)$ with q = 200. The *p*-values of the B_q tests are based on the extreme value distribution. LR, LM, Q and S stand for the Comte and Lieberman's (2000), Hafner and Herwartz's (2006), Hongs (2001) and Cheung and Ng's (1996) tests, respectively. $Q(\hat{N}^*)$ and $Q(N=12)$ denote the Hong's statistics calculated at the optimal bandwidth and at the exogenously given bandwidth 12, respectively.

Table 14: Testing the null hypothesis that industrial production growth does not Granger cause-in-variance the real stock market returns in Italy, Japan, the United Kingdom, and the United States.

	Countries							
	Italy		Japan		United Kingdom		United States	
Kernel	B_q	$Q(N=12)$	B_q	$Q(N=12)$	B_q	$Q(N=12)$	B_q	$Q(N=12)$
Daniell	1.7361 (0.0789)	1.0979 (0.1092)	1.5636 (0.4014)	0.8061 (0.1441)	2.5959 (0.0598)	0.2364 (0.1940)	1.5934 (0.0823)	0.2625 (0.3965)
Tukey – Hanning	1.5206 (0.0841)	1.1669 (0.1010)	1.4451 (0.4331)	0.9078 (0.1321)	2.4124 (0.0646)	0.2610 (0.1928)	1.8168 (0.0770)	0.5886 (0.2781)
Quadratic Spectral	1.9213 (0.0747)	0.9145 (0.1313)	1.3722 (0.4533)	0.5704 (0.1695)	**2.4021** **(0.0254)**	0.4750 (0.1782)	1.4687 (0.0854)	-0.0136 (0.5054)
Bartlett	1.5288 (0.0839)	1.1875 (0.0986)	1.4451 (0.4331)	0.8368 (0.1406)	2.1180 (0.0604)	0.1202 (0.1980)	1.6879 (0.0800)	0.5715 (0.2838)
	$Q(\hat{N}^*_{CV})$	$Q(\hat{N}^*_{DB})$	$Q(\hat{N}^*_{CV})$	$Q(\hat{N}^*_{DB})$	$Q(\hat{N}^*_{CV})$	$Q(\hat{N}^*_{DB})$	$Q(\hat{N}^*_{CV})$	$Q(\hat{N}^*_{DB})$
Daniell	0.7887 (0.2152) [19]	1.0870 (0.1385) [120]	1.5636 (0.0590) [2]	0.1940 (0.4231) [33]	1.3403 (0.0901) [19]	1.5356 (0.0623) [21]	0.5813 (0.2805) [10]	-1.8246 (0.9660) [50]
Tukey-Hanning	0.8407 (0.2002) [19]	0.9184 (0.1792) [130]	1.4451 (0.0742) [2]	0.1379 (0.4452) [42]	0.7307 (0.2325) [19]	1.0179 (0.1544) [22]	0.8531 (0.1968) [10]	-1.4564 (0.9274) [50]
Quadratic Spectral	0.4665 (0.3204) [19]	1.2999 (0.0968) [106]	1.3722 (0.0850) [2]	0.1630 (0.4353) [134]	**2.0951** **(0.0181)** **[19]**	**2.1917** **(0.0142)** **[136]**	0.3394 (0.3672) [10]	-2.2907 (0.9890) [50]
Bartlett	0.8439 (0.1994) [19]	1.0156 (0.1549) [124]	1.4451 (0.0742) [2]	0.2000 (0.4207) [42]	1.1240 (0.1305) [19]	1.1240 (0.1305) [20]	0.8447 (0.1991) [10]	-1.4794 (0.9305) [50]
LM	**13.0732** **(0.0014)**		4.5083 (0.1050)		1.1624 (0.5592)		3.7979 (0.1497)	
LR	**14.0429** **(0.0009)**		-3.3742 (1.000)		**33.1281** **(0.0006)**		3.6129 (0.1642)	
S (*N* = 12)	13.4158 (0.3396)		12.4986 (0.4065)		10.5674 (0.5663)		8.2881 (0.7622)	
*S** (*N* = 12)	13.5687 (0.3291)		12.6553 (0.3946)		10.6807 (0.5565)		8.3721 (0.7554)	

Notes: *p*-values are in parentheses. Numbers in brackets are the optimal bandwidths computed according to cross-validation method ($\hat{N}^*_{CV}$) and Dubgen type estimators ($\hat{N}^*_{DB}$). $B_q = \max_{2<N\leq q} Q(N)$ with q = 200. The *p*-values of the B_q tests are based on the extreme value distribution. LR, LM, Q and S stand for the Comte and Lieberman's (2000), Hafner and Herwartz's (2006), Hongs (2001) and Cheung and Ng's (1996) tests, respectively. $Q(\hat{N}^*)$ and $Q(N=12)$ denote the Hong's statistics calculated at the optimal bandwidth and at the exogenously given bandwidth 12, respectively.

Chapter 3

A model-free test for causality-in-volatility, with an application to the output growth and stock return volatility relationship

1. Introduction

Modeling volatility spillovers across different assets or markets has been very important in the finance and macroeconomics literature since Morgenstern (1959) and more recently Granger et al. (1986), Schwert (1989), Baillie and Bolleslev (1990), Cheung and Ng (1990), Engel et al. (1990), Lin et al. (1994), Billio and Pelizzon (2003). Most of these papers estimate parametric models to examine specific formulations for the spillover effects, while Cheung and Ng (1996) and Hong (2001) develop general causality-in-variance tests within this framework.

Investigating whether two financial time series exhibit mutual dependence in variance is important for a wide range of applications, including risk management, asset pricing and the development of policies for economic and financial stability. Since the concept of Granger causality (Granger 1969, 1980,1988) there is a considerable amount of research that tests the casual relationship between economic and financial variables; see Geweke (1984) and Hoover (2001) for surveys. While most of the original studies focus on the mean, recent literature allows for a deeper analysis of the causality topic by extending it to second order moments.

The first definition of variance causality is due to Granger et al. (1986). Comte and Lieberman (2000) extended the original Granger idea by testing the null of non-causality-in-variance and by setting linear restrictions on the parameters of a multivariate GARCH model. Alternatively, Cheung and Ng (1996) present a two step procedure by first prewhitening two time series and then testing whether the two squared residual series, properly standardized by their respective conditional volatility estimates, are independent. Hong (2001) modified Cheung and Ng's approach by proposing a class of kernel based cross-correlation tests. Hafner and Herwartz (2006) adapt a Lagrange test for noncausality in variance of financial returns and show that their LM test outperforms the Cheung and Ng (1996) test in terms of empirical power.

While these tests have a good finite sample performance against a sequence of local alternatives, implementation of these test procedures would require estimation of a conditional volatility specification via a GARCH type representation, thus imposing the assumption that the conditional second moments follow an explicit functional form. However, while it is widely recognized that this class of statistical models are more than adequate to account for the 'volatility clustering' effects, usually met in financial return series, it is uncertain whether they represent the actual data generating mechanism. See for example Harvey and Siddique (1999) who argue that the presence of excess unconditional leptokurticity and skewness in financial returns, can affect the time series properties of the conditional variance and consequently, may lead to misleading inference on causality.

To our knowledge, no attempt to date has been made to test for causality-in-variance within an unconditional volatility framework[15]. The purpose of this paper is to propose a simple method for testing unidirectional Granger causality-in-variance. Our approach, as opposed to existing causality-in-variance procedures which impose an explicit functional form on the evolution of the second order dynamics, is based on a model-free volatility proxy. Our procedure compares simultaneously a set of p-values, which result from the implementation of an asymptotic standard normal test at multiple single lag periods. The test statistic employed in these evaluations examines the null of non-causality in volatility between the two time series for an individual lag period. The test calculations use the sample cross-correlations between the absolute values of innovations resulting from fitting autoregressions to the bivariate set of returns. Joint inference is conducted by applying the multiple comparison procedure

[15] While, in a related work, Schwert (1989) uses the absolute values of prewhitened residuals as volatility proxies or growth rates of different financial and macroeconomic time series, at a second stage these measures work as inputs for the estimation of a conditional volatility model. A considerable amount of recent work also emphasizes the use of empirical measures of asset return variability that involve squared and absolute returns. French, Schwert, and Stambaugh (1987) sum squared daily stock returns to compute monthly standard deviations. Subsequent researchers, such as Andersen, et al. (2001 a and b, 2003), employ high-frequency intradaily returns to calculate daily variability measures, which are considered to be unbiased and efficient estimates of financial volatility.

developed by Rom (1990). In particular, the p-values are ordered and then contrasted to some corresponding critical levels of significance. The overall null hypothesis of no-causality is rejected if at least one p-value is found to be below the critical level of significance. The latter emerge as adjustments of the nominal level of significance to the lag truncation, so as to ensure a firm control of the joint type I error probability[16]. Monte Carlo experiments have been performed to evaluate the finite sample properties of the proposed test. The performance of our test is compared to two conventional tests, namely Cheung and Ng's (1996) *S* test and Hong's (2001) *Q* tests, under alternative models regarding the causal lag structure, the distributional characteristics of the series and the degree of fractional integration of the volatility process.

Our results show that the proposed test is well sized. In addition, the finite sample size of the test is less sensitive than the *S* test to arbitrary choices of the lag selection parameter and the distribution of the error term. Interestingly the implementation of the proposed test yields surprisingly high finite sample power, even in the presence of long horizon causalities. This remarkable performance is robust with respect to the implemented lag truncation and it holds for different sample sizes.

By comparison, our test never performs worse than the *Q* and *S* tests and in fact outperforms both tests when dealing with short horizon causalities, especially so when the sample size is not very large. In addition, the power of both tests appears to depend greatly on the lag truncation and for the case of the *Q* test, on the weighting scheme used as well. Our findings also indicate that all test procedures have poor power under the presence of long-range dependence in the underlying volatility processes. Nevertheless, for large sample sizes there seems to be an advantage in using our proposed test.

We apply our tests to study volatility spillover between the real stock returns and industrial production growth in U.S., United Kingdom, Italy and Canada. Previous research on whether changes in industrial production growth can predict future stock returns has been inconclusive and there is very little evidence on the

[16] Nonparametric multiple comparison procedures have been applied to several econometric problems. For instance, Chow and Denning (1993) proposed a joint test for the martingale difference hypothesis, resulting from simultaneous comparisons of the variance ratios at multiple time horizons.

relationship between the volatilities of industrial production growth and real stock returns. We used data from Datastream, IFS and OECD database covering the period from January 1973 to May 2011. Our results show that fluctuations of industrial production growth generate extreme variability in the real stock return patterns but not the other way around. This result is robust with respect to the implemented lag truncation and holds for US, United Kingdom and Italy. In comparison, Cheung and Ng's (1996) S test and Hong's (2001) Q tests (with a few exceptions) do not reject the non-causality hypothesis for either direction.

The remainder of the paper is structured as follows. Section 2 briefly describes the Cheung and Ng (1996) and Hong (2001) tests. Section 3 describes our procedure. Section 4 reports some of our representative Monte Carlo simulation results. Section 5 describes the data. Section 6 is the empirical application of our test, where we study the volatility spillover between the U.S. real stock returns and real earnings growth. Section 7 provides a summary and concludes.

2. Causality-in-variance tests based on GARCH type conditional volatility models

In this section we briefly describe the Granger causality-in-variance test of Cheung and Ng (1996) and an enhanced version of their test procedure proposed by Hong (2001). Most empirical studies on volatility spillover use a Granger type test, namely regressing the square residuals of one variable on its own lagged square residuals and on other lagged variables in the framework of multivariate GARCH models. A different approach has been taken by Cheung and Ng (1996) who propose a new test for volatility spillover using the sample cross-correlation function between two squared residuals standardized by their conditional variance estimators. More specifically, their test is based on the sum of finitely many squared sample cross-correlations, which has a null asymptotically χ^2_M distribution. Hong (2001) proposes a class of new tests for volatility spillover. He basically tests for causality in variance in the sense of Granger (1969, 1980) who introduces the concept of causality in terms of incremental predictive ability between two time series, instead of the more conventional approach of cause and effect. Hong's test does not assume any specific distribution, such as normality and it applies to time series that exhibit conditional heteroskedasticity and may have infinite unconditional variances. He also introduces a weighting scheme for the sample cross correlation at each lag, in contrast to Cheung and Ng (1996) test which gives uniform weighting to each lag. This idea of non uniform weighting was also introduced by Engle (1982) to improve the power of his Lagrange Multiplier test for ARCH effects.

Let the bivariate stationary process $\varepsilon_{it} = \zeta_{it}(h_{it})^{1/2}, i = 1,2$, with $\zeta_{it} \sim iid(0,1)$, $E(\varepsilon_{it}/\Phi_{it-1}) = 0$ and $Var(\varepsilon_{it}/\Phi_{it-1}) = h_{it}$, where Φ_{it-1} denotes the σ–field generated by past realizations of ε_i available at time t-1 and $h_{it} = \kappa_i + \sum_{j=1}^{p} a_{ij}\varepsilon_{it-j}^2 + \sum_{j=1}^{q} b_{ij}h_{it-j}$. Following Cheung and Ng (1996), the independent series $\left\{\hat{\zeta}_{it} = \varepsilon_{it}^2 \big/ \hat{h}_{it}\right\}$, i = 1,2, are used for the sample cross-correlation function calculation at lag k, defined as

$$\hat{r}_{\zeta_1\zeta_2}(k) = \hat{C}_{\zeta_1\zeta_2}(k)\left[\hat{C}_{\zeta_1}\hat{C}_{\zeta_2}\right]^{-1/2} \quad (1)$$

where $\hat{C}_{11}$ and $\hat{C}_{22}$ are the sample variances of $\hat{\zeta}_{1t}$ and $\hat{\zeta}_{2t}$ respectively, and $\hat{C}_{12}(k)$ is defined as

$$\hat{C}_{\zeta_1\zeta_2}(k) = T^{-1}\sum_{t=1}^{T-k}\hat{\zeta}_{1t+k}\hat{\zeta}_{2t},\ k > 0 \quad (2)$$

The Portmanteau statistic S can be applied to test the null hypothesis that ε_{2t} does not Granger cause in variance ε_{1t} for all k = 1,…,N jointly[17]:

$$S = T\sum_{k=1}^{N}\hat{r}_{\zeta_1\zeta_2}^2(k) \overset{a}{\sim} \chi^2(N). \quad (3)$$

A drawback of this procedure is that inference is sensitive to the choice of N, i.e., the number of lagged cross-correlations used in the computation of the S statistic.

On the other hand, Hong (2001) proposed a class of normalized tests, the Q tests, which utilize a weighted scheme on the sample cross correlations[18]:

$$Q = \left\{T\sum_{k=1}^{T-1}v^2(k/N)\hat{r}_{\zeta_1\zeta_2}(k) - E\right\}\Big/\left\{2G\right\}^{1/2} \overset{a}{\sim} N(0,1) \quad (4)$$

[17] Throughout this paper, the symbol $\overset{\alpha}{\sim}$ will denote 'asymptotically distributed'.
[18] Hong(2001), as well as Cheung and Ng (1996), have presented small sample versions of their test statistics. Due to space limitations, we only report the main test statistics.

where $v(.)$ is a kernel function (for example the Daniell), $E = \sum_{k=1}^{T-1} (1 - k/T) v^2 (k/N)$ and $G = \sum_{k=1}^{T-1} (1 - k/T) \{1 - (k+1)/T\} v^4 (k/N)$.

The finite sample power of this class of tests may depend greatly on the choice of the bandwidth parameter *N*. Hong (2001) argues that the selection of the kernel function is of secondary importance.

3. Econometric procedure

Consider the bivariate ergodic and covariance stationary stochastic process $z_t = (z_{1t}, z_{2t})'$. Suppose that each separate $z_{it}, i = 1,2$ evolves as an ARMA (p,q) process

$$\phi_i(B) z_{it} = \delta_i(B) \varepsilon_{it}, \quad i = 1,2 \tag{5}$$

where

$$\phi_i(B) = 1 - \phi_{i1} B - ... - \phi_{ip} B^p,$$

$$\delta_i(B) = 1 - \delta_{i1} B - ... - \delta_{iq} B^q,$$

while *B* is the Backshift operator on *t*, and $\varepsilon_{it}, i = 1,2$ are the residual series. The terms $\phi_i(B)$ and $\delta_i(B)$ are assumed to have all roots outside the unit circle. Let the σ-fields $F_1 = \sigma(z_{1t-d}; d \geq 0)$ and $F = \sigma(z_{1t-d}, z_{2t-d}; d \geq 0)$ be two information sets generated by the past realizations of the series $\{z_{it-d}; d \geq 0\}$. The null hypothesis, that z_{2t} does not Granger cause-in-variance z_{1t} with respect to information set *F*, can be formulated as

$$H_0 : E(|\varepsilon_{it}| \mid F_1) = E(|\varepsilon_{it}| \mid F).$$

To investigate the previous hypothesis we consider the cross-correlation function between X_{1t} and X_{2t} at lag *k*, where $X_{it} = |\varepsilon_{it}|, i = 1,2$, defined as

$r(k) = d_{|\varepsilon_1\varepsilon_2|}(k)\left[d_{|\varepsilon_1|}(0)d_{|\varepsilon_2|}(0)\right]^{-1/2}$,

where $d_{|\varepsilon_i|}(0) = Var(X_{it}), i = 1,2$, and $d_{|\varepsilon_1\varepsilon_2|}(k) = Cov(X_{1t}, X_{2t-k})$ represent the variances of each X_{it} and the cross-covariance of X_{1t} and X_{2t} at lag *k* with *k* ≥ 0, respectively.

Denote $\hat{X}_{it} = |\hat{\varepsilon}_{it}|$, where $\{\hat{\varepsilon}_{it}\}$ the vector of residuals obtained from estimating regression (5) by OLS. The cross correlation function at lag *k* of X_{it} and X_{jt} can be estimated non-parametrically by

$$\hat{r}(k) = \hat{d}_{|\varepsilon_1\varepsilon_2|}(k)\left[\hat{d}_{|\varepsilon_1|}\hat{d}_{|\varepsilon_2|}\right]^{-1/2} \qquad (6)$$

where $\hat{d}_{|\varepsilon_i|}, i = 1,2$, the sample variance of X_{it} and

$$\hat{d}_{|\varepsilon_1\varepsilon_2|}(k) = T^{-1}\sum_{t=1}^{T-k}\left(\hat{X}_{1t} - \hat{\mu}_{|\varepsilon_1|}\right)\left(\hat{X}_{2t-k} - \hat{\mu}_{|\varepsilon_2|}\right), k \geq 0. \qquad (7)$$

Here, $\hat{\mu}_{|\varepsilon_i|}$, *i* = 1,2, represent the sample mean of X_{it}.

We can state the following theorem.

Theorem. Let $W = \sqrt{T}(\hat{r}\ (1),...,\hat{r}\ (k))$, where 1,2,…,*k* are fixed integers. Assume that the derivatives $\frac{\partial \hat{d}_{|\varepsilon_1\varepsilon_2|}(k)}{\partial \varphi_i}$ exist and are bounded in probability for $\varphi_i \in \varphi$, where φ is the parameter space, while the condition $E\left|\varepsilon_{it}^2\right|^{1+\lambda} < \Lambda < \infty$ holds for some $\lambda > 1$ and all *t*. Under the null hypothesis that z_{2t} does not Granger cause-in-variance z_{1t} at the specific lags 1, 2, …, *k*, *W* is asymptotically distributed as $N(0, I_k)$ (for the proof see the Appendix).

Under the null hypothesis that ε_{it} does not Granger cause-in-variance ε_{jt} up to lag *N*, the relationship $\rho(m) = 0$ holds for each *m* = 1,…, *N*. Consider now a set of *N* statistics $W(1),...,W(N)$ with corresponding p-values $P_1,...,P_N$ examining the validity of each individual hypothesis $H_{01},...,H_{0N}$, where $H_{ol}:\rho(l)=0$. The null hypothesis of Granger noncausality-in-variance is equivalent to a global hypothesis, which is a combination of N sub-hypotheses: $H_o=\cap\{H_{0l}\}_{l=1}^{N}$. A rejection of any $\{H_{0l}\}_{l=1}^{N}$ can, therefore, lead to the rejection of the global hypothesis for a specific level of significance α. Hence, a way to test the null hypothesis of noncausality is to use the W tests and the corresponding p-values and reject the null if any p-value P_l is less than α.

Miller (1966), argues that the overall significance level α must be controlled in situations where an individual test is implemented simultaneously at multiple lag periods. This result has implications for testing the joint significance of the cross-correlations. Comparing the p-values of the W test with the significance level α can lead to invalid inference. As suggested by Miller, the level α must be adjusted in order to account for the number of sub-hypotheses comprising the global hypothesis. An approach that has been followed in the statistics literature is the implementation of a Bonferroni inequality adjustment, i.e., an upper bound on the overall significance level is set by dividing the test size α with the number of individual hypotheses. Nevertheless, this method yields very low finite sample power. A number of modified versions of the Bonferroni method have been proposed in order to increase the test power while ensuring a firm control of the joint probability of type I error (Hochberg (1988), Holm (1979), Hommel (1988), Rom (1990)).

Multiple comparisons between the cross-correlation based test statistic are performed by applying the method developed by Rom (1990). Rom's procedure is an improvement on the Hochberg (1988) method. This method is implemented in two simple steps. First the elements of $\{H_{0l}\}_{l=1}^{N}$ are ordered by their p-values and adjusted significance levels are then calculated for their evaluation. Starting from the hypothesis with the largest p-value, this method sequentially contrasts the p-values with the corresponding adjusted significance levels. The overall null hypothesis of noncausality-in-variance is rejected only when one p-value is found to be smaller or equal to its corresponding level of significance. Otherwise, the null hypothesis of noncausality-in-variance is accepted. A detailed description of the procedure follows in the next subsection.

Rom's method of adjustment of the significance level:

The p-values are ordered from smallest to largest: $P_{[1]},...,P_{[N]}$. Let $a^*_{[1]},...,a^*_{[N]}$ be the adjusted significance levels employed in the multiple evaluations of the ordered individual hypotheses $H_{0[1]},...,H_{0[N]}$. The adjusted levels $a^*_{[l]}, l = 1,.., N$ are calculated recursively as:

$$a^*_{[N-i+1]} = \left\{ \sum_{h=1}^{i-1} a^h - \sum_{h=1}^{i-2} \binom{i}{h} a^{*(i-h)}_{[N-h]} \right\} / i, \tag{8}$$

where $i = 2, \ldots, N$ and $a^*_{[N]} = a$.

Inference is conducted by sequential pairwise comparisons of the p-values $P_{[.]}$ with the equivalent critical levels $a^*_{[.]}$. Rom' s method starts the testing by contrasting the largest p-value $P_{[N]}$ with the corresponding adjusted significance level $\alpha_{[N]}$. If $P_{[N]} \leq a^*_{[N]}$, then the overall null hypothesis is rejected; otherwise, the evaluation moves on to the next pair $\left(P_{[N-1]}, \alpha^*_{[N-1]}\right)$. The procedure is repeated sequentially until at least one l ($l = 1,\ldots, N$) is found, such that $P_{[l]} \leq a^*_{[l]}$ will hold. If not, then the global hypothesis is accepted.

The main steps of the proposed testing procedure can be summarized as follows:

1. The absolute values of the residuals are used in the computation of the W statistics and the corresponding p-values for a series of successive lag periods $l = 1,\ldots, N$.
2. The p-values are ordered from lowest to largest. Define α and use the formula proposed by Rom to obtain the N-1 adjusted levels of significance. If there is at least one p-value smaller than the corresponding adjusted significance level, reject the null hypothesis up to lag N; otherwise accept the null.

4. Monte Carlo simulations

This section presents our Monte Carlo simulation results conducted to evaluate the finite sample performance of our tests as compared with the two conventional causality-in-variance tests presented in Section 2. We begin by the description of the Monte Carlo set up, followed up by the presentation of our main results.

3.4.1. Experimental design

The simulated series $y_t = (y_{1t}, y_{2t})'$ are generated from a bivariate BEKK (p,q) process given by:

$$y_t = u_t, \quad t = 1,2,...,T, \tag{9}$$

$$u_t = \xi_t (H_t)^{1/2}, \tag{10}$$

$$H_t = D'D + \sum_{i=1}^{p} \left(G'_i\, u_{t-i} u'_{t-i} G_i\right) + \sum_{i=1}^{q} \left(F'_i H_{t-i} F_i\right), \tag{11}$$

where $u_t = (u_{1t}, u_{2t})'$, $\xi_t \sim IID(0,1)$, $D = [d_{ij}]$, $G_k = [g_{ij}^k]$, $k = 1,2,..., p$, and $F_k = [f_{ij}^k]$.

The parameter values in equations (9):(11) describe the following data generating processes (DGPs):

DGP1:

$$\begin{cases} p = q = 1 \\ (g_{11}^1, g_{12}^1, g_{21}^1, g_{22}^1) = (0.9, 0, 0, 0.25) \\ (f_{11}^1, f_{12}^1, f_{21}^1, f_{22}^1) = (0.35, 0, 0, 0.1) \\ \xi_t \sim NIID(0,1) \end{cases}$$

Under this setting we evaluate the size of the test since there is no causality-in-variance between ε_{1t} and ε_{2t}.

DGP2:

$$\begin{cases} p = q = 1 \\ (g_{11}^1, g_{12}^1, g_{21}^1, g_{22}^1) = (0.9, 0, 0, 0.25) \\ (f_{11}^1, f_{12}^1, f_{21}^1, f_{22}^1) = (0.35, 0, 0, 0.1) \\ \xi_t \sim Student's\ t(6) \end{cases}$$

The parameter values are identical to DGP1, although now the ξ_t are distributed as Student t with 6 degrees of freedom.

DGP3:

$$\begin{cases} p = q = 1 \\ (g_{11}^1, g_{12}^1, g_{21}^1, g_{22}^1) = (0.9, 0, 0.2, 0.25) \\ (f_{11}^1, f_{12}^1, f_{21}^1, f_{22}^1) = (0.35, 0, 0.2, 0.1) \\ \xi_t \sim NIID(0,1) \end{cases}$$

We allow ε_{2t} to Granger cause-in-variance ε_{1t} at first lag period, in order to investigate the power of the test.

DGP4:

$$\begin{cases} p = q = 10 \\ (g_{11}^1, g_{12}^1, g_{21}^1, g_{22}^1) = (0.9, 0, 0, 0.25) \\ (g_{11}^{12}, g_{12}^{12}, g_{21}^{12}, g_{22}^{12}) = (0, 0, 0.2, 0) \\ (f_{11}^1, f_{12}^1, f_{21}^1, f_{22}^1) = (0.35, 0, 0, 0.1) \\ (f_{11}^{12}, f_{12}^{12}, f_{21}^{12}, f_{22}^{12}) = (0, 0, 0.2, 0) \\ (g_{11}^k, g_{12}^k, g_{21}^k, g_{22}^k) = (0, 0, 0, 0),\ k = 2, 3, .., 9 \\ (f_{11}^k, f_{12}^k, f_{21}^k, f_{22}^k) = (0, 0, 0, 0),\ k = 2.3, .., 9 \\ \xi_t \sim NIID(0,1) \end{cases}$$

Under this model, there is a long lag causal structure given that ε_{2t} Granger cause-in-variance ε_{1t} at the 10th lag period.

Empirical results, reported in Taylor (1986), Ding, Granger and Engle (1996), Baillie et al. (1996), have demonstrated that the autocorrelations of realized volatility proxies, such as absolute or squared daily stock or exchange rates returns, exhibit a slow hyperbolic rate of decay over long lags. This property is characterized as long memory in financial volatility and has been the object of controversy by many researchers. Bollerslev and Mikkelsen (1996), Breidt et al. (1998), Henry and Payne

(1997), Eibens (1999) provide evidence on the presence of long term dependence in volatility dynamics. On the other hand, Ryden et al. (1998), Mikosch and Starica (1998), Granger and Hyung (1999) argue that the high degree of persistence is caused by unaccounted structural breaks or potential nonlinearities in the conditional second moments. Recent empirical work, such as Morana and Beltratti (2004), show that the hyperbolic memory in the exchange rate realized volatility is attributed to structural changes only to some extent. Motivated by these considerations, we investigate whether the presence of long memory effects may have some impact on the performance of the proposed test. To achieve this goal we use a different data generating mechanism which is presented in equations (12) to (15). In particular, each y_t evolves as a univariate Fractional Integrated GARCH (p, d, q) process as introduced by Baillie et al. (1996), enhanced by a volatility spillover term:

$$y_t = u_t,\ t = 1,2,...,T, \qquad (12)$$

$$u_t = \xi_t (h_{it})^{1/2},\ \xi_t \sim NIID(0,1) \qquad (13)$$

$$h_{it} = k_i + b_i h_{it-1} + (1 - b_i L - (1-\varphi_i)(1-L)^{d_i})u_{it}^2 + \delta_i h_{jt-1} \qquad (14)$$

$$(1-L)^{d_i} = 1 - d\sum_{k=1}^{1000}\Gamma(k-d_i)\Gamma(1-d_i)^{-1}\Gamma(k+1)L^k \qquad (15)$$

where $(1-L)^{d_i}$ is the Fractional differencing operator, d_i is the Fractional Differencing Parameter, Γ is the Gamma function, b_i is the GARCH parameter coefficient and δ_i is the coefficient of the volatility spillover term. When parameter d_i approaches zero any shocks will cause persistent effects on the volatility dynamics, which is a characteristic of long memory processes. On the other hand, the influence of a shock decays with a geometrical rate when d_i approaches unity.

Under the mechanism described above, the volatility processes are assumed to have fractional degrees of integration. We consider two hypotheses when studying the power properties of our tests. First we evaluate the effects of the presence of hyperbolic memory on the test performance. This is denoted as the long memory hypothesis. Then, we examine the impact of geometric memory. This is denoted as the short memory hypothesis.

DGP4:

$$\begin{cases} p = q = 1 \\ d_1 = d_2 = 0.4 \\ (k_1, b_1, \varphi_1, \delta_1) = (0.03, 0.3, 0.3, 0.1) \\ (k_2, b_2, \varphi_2, \delta_2) = (0.03, 0.3, 0.3, 0) \end{cases}$$

DGP5:

$$\begin{cases} p = q = 1 \\ d_1 = d_2 = 0.6 \\ (k_1, b_1, \varphi_1, \delta_1) = (0.03, 0.3, 0.3, 0.1) \\ (k_2, b_2, \varphi_2, \delta_2) = (0.03, 0.3, 0.3, 0) \end{cases}$$

The parameter values in DGP4 and DGP5 describe a long and a short memory model, respectively. In both models, unidirectional volatility spillovers occur from y_{2t} to y_{1t} at the first lag period.

To assess the empirical performance of the Q tests under different weighting schemes we use three kernel functions: the Quadratic Spectral, Daniell, and Tukey-Hanning kernels. Each of the statistics is calculated with N = 10, 20, 30, 40, and 50. Samples of size T = 200, 500, 1000 and 4000 are simulated while we use 2500 replications. To eliminate the effect of the start-up value[19], we generate T + 2500 observations and then we remove the first 2500 observations. A univariate GARCH (1,1) model is fitted to each simulated series in order to compute the series $\left\{\hat{\zeta}_{it} = y_{it}^2 / \hat{h}_{it}\right\}$, used in the Q and S test calculations.

[19] The start-up value was set to be the unconditional covariance of the BEKK process, i.e., $H_t^* = \left[I - \sum_{i=1}^{p} (G_i \otimes G_i)' - \sum_{i=1}^{q} (F_i \otimes F_i)' \right]^{-1} vec(D'D)$. For more details see Engle and Kroner (1993).

3.4.2. Simulation results

We evaluate first, the empirical size of the causality tests under the null hypothesis that y_{2t} does not Granger-cause in variance y_{1t}. Table 1 reports the size of the tests at the 5% and 10% levels of significance under the null hypothesis of a Normal GARCH-BEKK(1,1) process. The empirical size of the *W* test is close to the nominal levels of significance 5% and 10% for all sample sizes. For example, for *T* = 500 the size of *W* test at nominal level 5% is 4.00%, 4.52%, 5.12%, 5.48% and 5.52%, implemented for lag selections 10, 20, 30, 40, and 50 respectively. For *T* = 1000 the *W* test tends to slightly over-reject the null hypothesis. Overall, the size of our test does not appear to depend on the choice of the lag truncation *N*. On the other hand, the *Q* tests are slightly under-sized for *N* = 10; when *N* increases the estimated probability of Type I error tends to be near the 5% level. The choice of *N* has the opposite effect on the *S* test size since the under-sizing becomes more severe when *N* increases, especially for small sample sizes. The finite sample size of the tests, when implemented to volatility processes where ξ_t are distributed as Student *t* with 6 degrees of freedom (DGP2), is presented in Table 2. Overall, the W test has good size for all *T*. Compared to the previous results, our test statistic now appears to be slightly under-sized. On the other hand, leptokurtic errors cause the opposite effect on the size of the *Q* and *S* tests.

Table 3 presents the power of the tests at nominal level 5% and 10% under DGP3, where volatility spillovers from y_{2t} to y_{1t} occur at the first lag period. For all sample sizes *T*, the *W* test has great power against DGP3; with a few exceptions, the estimated Type II error probabilities are always 100% at both levels of significance. Furthermore, our results show that the test power is robust to arbitrary choices of the lag selection parameter. On the other hand, the power properties of the competing test procedures seem to depend on the choice of *N*. For instance, for *T* = 500 the rejection rates of *Q* test, which is based on the Tukey- Hanning kernel, when *N* takes the values 10 and 50, are 43.68% and 24.84% respectively at level 5%; on the other hand, the power of the *W* test is 100% for both selections of *N*. Under DGP3, the superiority of the *W* test performance at small, medium and large sample sizes (e.g. 200, 500 and 1000) over the competing test procedures is evident. It is interesting to note that for very large *T* (e.g. 4000) the *Q* and *S* tests share similar power with the *W* test.

Table 4 reports the power of the tests at nominal level of significance 5% and 10% under DGP4, where y_{2t} Granger causes-in-variance y_{1t} at the 10^{th} lag period. The *W* test has a satisfactory performance against long horizon causalities. For small and moderate sample sizes (e.g. 200 and 500) the rejection rates range from 94.96% to 100% at both nominal levels of significance. The power of our test for larger sample sizes (e.g. 1000 and 4000) is always 100%. The gain in power from implementing the *W* test is now small since the competing tests appear to have also good power. However, our test seems to have relatively more power in detecting direct changes in the causal structure of the data, since it always yields high power for $N = 10$.

Table 5 reports the power of the tests at nominal level of significance 5% and 10% under DGP5, where y_{2t} Granger causes-in-variance y_{1t} at the first lag period, while there is a long range dependence in the underlying volatility process. The results show that the *W* test has good power only for very large sample sizes. For $T = 4000$, the rejection rates at nominal level 10% are 56.92%, 74.08%, 73.80%, 74.76% and 74.60% when *N* equals 10,20,30,40 and 50 respectively. However, for smaller sample sizes all test procedures appear to perform below 50%. Moreover, the simulation results show that the power of the tests decreases with *T*. Note that for $T = 200$ the *W* test yields maximum power 13.72% and 18.32% at nominal levels 5% and 10% respectively. Interestingly, the power of our test seems to increase with *N* at both levels of significance. For $T = 1000$, when a lag truncation 10 and 50 is used, the test yields power 21.80% and 39.56% respectively. Nevertheless, comparing the *W*, *Q* and *S* simulation results, we observe that the first performs relatively better when applied to large and very large sample sizes (e.g. $T = 1000$ and 4000).

Table 6 presents the power of the tests at nominal level of significance 5% and 10% when the underlying process exhibits short range dependence (DGP6). The performance of our test is very poor under the presence of short memory in volatility, especially when implemented to medium and small sample sizes. However, a substantial improvement in the power of the test seems to have been achieved, when compared to the results of Table 5. The *Q* and *S* tests share similar power with the *W* test. It appears that there is not much to be gained by using the *W* test.

In conclusion, the simulation results show that the finite sample size of the *W* test is more robust than the S test to leptokurtic errors and the lag truncation, for different sample sizes. Moreover, it has excellent power under a series of local alternatives. Under these models, the choice of the lag selection parameter and the size of the sample have a relatively small effect on the power of the test. On the other hand, the *Q* and *S* tests have good power only when implemented to very large sample sizes, while their performance appears to depend greatly on the choice of the weighting scheme. The simulation results also reveal that all tests perform very poor when implemented to volatility processes that are characterized by long memory.

5. Data and empirical analysis

In the next section, the relationship between stock returns volatility and industrial production growth volatility is investigated. Our analysis is restricted to four developed countries: the US, the United Kingdom, Italy, and Canada. The data consist of monthly observations of the aggregate stock price index, the industrial production index, and the consumer price index (CPI). Stock market data is taken from Datastream. Industrial production and CPI are retrieved from IFS and OECD database, respectively. The sample period spans from January 1973 to May 2011. We define monthly stock returns and output growth rate as the logarithmic differences of stock indices and industrial production, respectively. Real stock returns are computed by subtracting CPI inflation from nominal stock returns.

In our empirical analysis we examine the relationship between stock returns volatility and output growth volatility. The evidence provided in this section is complementary to those presented by Schwert (1989), Beltratti and Morana (2006) and Diebold and Yilmaz (2007). Standard models in asset pricing imply that stock prices reflect in the present time the discounted future expected earnings of all firms in a specific economy. So, changes of the conditional variance of the stock prices are determined by the changes of the conditional variance of the future discount rates and/or the expected future earnings. However, volatilities of both the future earnings and discount rates are affected by unexpected fluctuations of the real economy. On the other hand, the stock prices are forward looking; reactions of speculative investors to anticipate events about the future economic fundamentals yield shifts in current stock price volatility.

In his seminal work, Schwert (1989) investigates the association between the volatilities of US stock returns and industrial production growths by using monthly and daily data covering a period from 1857 to 1987. He provides some weak evidence on the predictive ability of the US stock market volatility for the future volatility of industrial production while he finds even less for the reverse direction of the causal relationship. Beltratti and Morana (2006) show that the volatilities of US stock returns and industrial production growth are not related in the long-run. However, they demonstrate that the dynamics of stock return volatility can be decomposed into a persistent and a non persistent component, and that industrial production volatility accounts for a significant proportion of these two components.Recently, by employing a panel data framework on quarterly data for a large number of countries, Diebold and Yilmaz (2007) have found that output growth Granger causes-in-volatility the stock market returns.

Our analysis is restricted to four developed countries: the US, the United Kingdom, Italy, and Canada. The data consist of monthly observations of the aggregate stock price index, the industrial production index, and the consumer price index (CPI). We define monthly stock returns and output growth rate as the logarithmic differences of stock indices and industrial production, respectively. Real stock returns are computed by subtracting CPI inflation from nominal stock returns. Stock market data is taken from Datastream. Industrial production and CPI are retrieved from IFS and OECD database, respectively. The sample period spans from January 1973 to May 2011.

Van Dijk, Osborn and Sensier (2005) have shown, through Monte Carlo simulation, that inference on Granger causality-in-variance may be misleading if both time series experience ignored simultaneous structural changes in the volatility dynamics. Further theoretical evidence on this result is provided by Rodrigues and Rubia (2007). Their analysis shows that the cross-correlation based tests will fail to converge asymptotically to the standard normal distribution under the presence of structural breaks in both volatility series at the same time. Therefore, before the null hypothesis of noncausality-in-variance is tested, we implement the Kokoszka and Leipus (1998, 2000) (hereafter denoted as KL) test on the squared (and absolute) stock returns and industrial production growths in order to detect for possible simultaneous volatility structural breaks[20]. The KL test is calculated as $\frac{\max V_T(j)}{s}$, where $V_T(j) = T^{-1/2}\left(\sum_{i=1}^{j} u_i - \frac{j}{T}\sum_{i=1}^{T} u_i\right)$, u represents the series of squared (or absolute) returns (or growths), and s is the standard deviation of the realized volatility processes, estimated by a VARHAC model[21]. The structural break test results and the dates of the change-points, for the squared and absolute returns of the time series, are presented in Table 7. Statistical significant structural changes are found only in two cases: the squared growth rates of industrial production of United Kingdom at level of significance 10% and the real squared (and absolute) stock returns of Italy at levels of significance 5% and 10%. However, no simultaneous structural changes have been documented. Furthermore, there is no close relation between the dating of the statistical insignificant change points in the volatility of both variables. Thus, the pre-testing allows us to proceed to our main testing objective.

Once the variables are made stationary by taking logarithmic first differences, VAR (p) models are fitted to each pair of growth/return rates for these four economies[22]. Akaike's criterion is used to specify the order of the fitted VAR (p) models. We find that the criterion is minimized at order p = 3 for US, United Kingdom, and Italy, and at order p = 5 for Canada. Monthly volatility proxies have been constructed as the absolute values of the residuals $|\varepsilon_{it}|$ generated by the estimation of the four VAR(p) models. Then these proxies are used to calculate our test statistic $W = \sqrt{T}(\hat{r}(1),...,\hat{r}(N))$ for different lag orders N as described in Section 3 (see equations (6):(7)). As far as it concerns the Q and S test calculations, a univariate GARCH (1,1) model is fitted to each separate residual series, and the attained

[20] The simulations results of Andreou and Ghysels (2001) indicate that the Kokozska and Leipus test has a reasonable good finite sample performance when applied to highly persistent GARCH processes.
[21] VARHAC refers to the Vector Autoregression Heteroscedasticity and Autocorrelation Consistent estimator proposed by den Haan and Levin (1997). The AIC criterion is used for the selection of the autoregressive lag length order.
[22] Pantelidis and Pittis (2004) have shown that neglected causality-in-mean may lead to spurious detection of second order causality. To this aim, a VAR model is estimated for the conditional mean process rather a standard ARMA specification.

conditional volatilities $\hat{h}_{it}$ are used for the computation of the series $\left\{\hat{\zeta}_{it} = \varepsilon_{it}^2 / \hat{h}_{it}\right\}$. Then, the test statistics are calculated as presented in equations (1): (4) of Section 2.

The empirical analysis commences with investigating the hypothesis of no Granger causality-in-variance from stock market returns to industrial production growth. Table 8 demonstrates the results of the application of our test procedure, as well as the two causality tests described in Section 2, on the four economies mentioned previously. The tests are computed for different lag selections and bandwidths (N = 10, 20, 30 and 40). The p-values of the W test results are ordered from lowest to largest value, and then contrasted to the levels of significance 5% and 10% when adjusted for each lag order as described in equation (8). For each N we only report the specific W test values, where the p-value is smaller than the adjusted significance level α^*; otherwise, the W test results with the minimum p-value are presented. The corresponding adjusted significance levels for original levels 5% and 10% are also reported for each N. The W test results presented in Table 8 demonstrate that there is no evidence of the presence of causality linkages for the specific direction in these four countries at levels of significance 5% and 10%. For instance, in US for N = 10, 20, 30 and 40, the minimum p-value 0.0315 of the W test is always larger than the corresponding adjusted significance level, i.e., α^* = 0.0051, 0.0026, 0.0017 and 0.0013 at original level $\alpha = 5\%$. Similar results are obtained when our test procedure is implemented in the other three economies. For example when N =10 and 20 is selected the minimum p-values 0.0435, 0.0385, and 0.0450 of the W test in United Kingdom, Italy and Canada respectively, are larger than the corresponding levels α^* = 0.0051 and 0.0026 at original level of significance $\alpha = 5\%$. In the same way, according to the Q and S test results the null of no causality of stock return volatility for industrial production growth volatility is not rejected at any lag order at the 5% and 10 % significance levels.

In Table 9, we present the results for testing the presence of causality from industrial production volatility to stock returns volatility. The results of the W test show that there is clear-cut evidence of the presence of volatility spillovers from industrial production growth to stock return series. Specifically, in three cases out of four the null of no causality for the specific direction is rejected at almost all lag orders at both levels of significance 5% and 10%. For instance, in U.S. the p-value

0.0016 is smaller than the related α^* = 0.0051, 0.0026, and 0.0017 at original level $\alpha = 5\%$, when considering lag orders N = 10, 20, and 30, respectively. This result holds for the economies of United Kingdom and Italy for different choices of the lag truncation. In Italy, for all lag length selections the no causality null hypothesis is rejected since the p-value of the W statistic 0.0008 is always much smaller than the corresponding adjusted level of significance (α^*=0.0057, 0.0027, 0.0018 and 0.0013 at original level $\alpha = 5\%$, while α^*=0.0116, 0.0055, 0.0036 and 0.0027 at original level $\alpha = 10\%$). On the other hand, the Q and S tests do not find statistical significant causal linkages from industrial production volatility to stock returns volatility in the majority of the cases. For example, note that only when the bandwidth parameter N equals 10 and 20 respectively, the Tukey-Hanning Q test values 1.5391 and 1.4902 in U.S. and Italy respectively show rejection of the null hypothesis. A possible explanation for these contradictory results could be that Q and S test statistics lack in power when applied to moderate sample sizes. Moreover, the power of these tests demonstrates great sensitivity to the selection of the weighting scheme and the lag truncation.

To sum up, the W test provide strong evidence of volatility transmission from industrial production growth to real stock market returns for the US, the UK and Italy but not the other way around. Contrary, the applications of Q and S tests fail to detect any volatility spillovers in any direction with some minor exceptions. The empirical results highlight the practical importance of the proposed test procedure.

6. Conclusions

Implementation of standard causality-in-variance test procedures requires the estimation of parametric ARCH volatility models for the underlying return processes. In this paper, we introduce a new volatility spillover test which is based on model free volatility measure. In particular, our approach uses the absolute residuals, which result from fitting autoregressions to the return series, as inputs to calculate an individual cross-correlation based test statistic. This test statistic is shown to be asymptotically distributed under the null hypothesis as standard normal. A multiple comparison scheme is used on a set of the individual test values, so as to perform joint hypothesis testing.

Extended Monte Carlo simulations are conducted in order to evaluate the finite sample size and power of the proposed test. The evaluations focus on the comparative performance of our test against two conventional test procedures, namely Cheung and Ng ‘s (1996) Portmanteau statistic *S*, and Hong ‘s (2001) kernel *Q* tests. The results show that all tests yield good finite sample size. However the size of our test is less sensitive than the *S* test to the lag truncation, the sample size or the distributional characteristics of the error term. Moreover, the implementation of the proposed test ensures remarkable gain in power, with power diminishing when the underlying volatility processes exhibit long range dependence. The lag truncation and the sample size have no effect on the test power, apart from the situations where the volatility processes display long range dynamic dependence. Note that even so, our test never performs worse than the *Q* and *S* tests. Moreover, the power of the tests depends greatly on the lag truncation, or the weighting scheme as far as it concerns the kernel tests, when applied to samples of small and moderate size. In addition, the performance of these tests is very poor under the presence of long memory.

Our test procedure is used to examine the relationship between the volatilities of output growth and real stock returns. Previous research on whether changes in industrial production can predict future stock returns is inconclusive and there is very little evidence on the relationship between the volatilities of industrial production and real stock returns. We used data for four develop economies, namely U.S., United Kingdom, Italy and Canada, covering a period from January 1973 to May 2011. Our results show that industrial production volatility predicts real stock volatility in three economies out of four. This result is robust with respect to the implemented lag truncation. On the other hand, we find no significant relationship in the opposite direction. In comparison Cheung and Ng ’s (1996) S test and Hong ’s (2001) Q tests (with a few exceptions) do not reject the non-causality hypothesis in either direction.

Appendix of Chapter 3

Table 1: size of the tests under DGP1 at nominal levels of significance 5% and 10%.

Tests	Lag Selection Parameter / Bandwidth Parameter *N*									
	10		20		30		40		50	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
					***T* = 200**					
W	3.36	5.76	3.88	6.76	3.80	6.84	3.96	6.52	3.72	6.48
Q Dan	4.68	7.00	5.12	7.32	5.24	7.72	5.48	8.32	5.48	8.52
Q T-H	4.48	6.52	5.20	7.52	5.20	7.64	5.28	7.72	5.60	8.24
Q QS	5.08	7.80	5.20	7.64	5.56	8.04	5.44	8.44	5.80	8.56
S	3.96	7.40	3.04	6.08	2.96	5.00	2.24	3.88	1.52	3.16
					***T* = 500**					
W	4.00	6.36	4.52	7.32	5.12	7.88	5.48	8.56	5.52	8.72
Q Dan	3.88	5.88	4.96	7.64	5.16	8.80	5.16	8.64	5.36	9.32
Q T-H	3.48	5.84	4.20	7.08	5.04	8.12	5.20	8.36	5.12	8.60
Q QS	4.16	6.52	5.08	8.36	5.04	8.60	5.32	9.12	5.68	9.48
S	3.96	8.00	3.40	7.72	3.84	7.32	3.28	6.56	3.44	6.24
					***T* = 1000**					
W	4.28	7.60	4.92	7.88	5.72	8.88	5.72	9.56	5.80	9.60
Q Dan	3.64	6.16	4.36	7.20	5.04	7.76	4.80	8.32	4.80	8.28
Q T-H	3.24	5.80	3.96	7.36	4.76	7.44	4.92	7.64	4.88	8.32
Q QS	3.72	6.92	4.72	7.36	4.68	8.16	4.68	8.32	4.68	8.64
S	3.40	7.16	3.76	7.48	3.60	7.12	3.32	6.28	3.32	6.96
					***T* = 4000**					
W	6.60	10.12	5.72	9.64	6.16	9.48	6.20	9.72	6.48	10.56
Q Dan	3.84	6.00	5.04	7.96	4.80	8.36	4.76	8.68	4.60	8.56
Q T-H	3.32	5.60	4.52	7.60	4.96	8.28	4.68	8.44	4.68	8.60
Q QS	4.28	7.00	4.88	8.36	4.72	8.56	4.80	8.28	4.64	8.44
S	4.20	8.40	4.12	8.08	4.00	7.96	3.80	8.04	3.76	7.64

Notes: a Normal bivariate BEKK (1,1) process is simulated (2500 replications); there is no causality -in -variance between y_{1t} and y_{2t}. Dan, T-H and QS stand for Daniell, Tukey-Hanning and quadratic-spectral kernels (function $v(.)$) respectively; the one-sided Q tests are defined as: $Q = \left\{ T \sum_{k=1}^{T-1} v^2(k/N)\hat{r}_{\zeta_1\zeta_2}(k) - E \right\} \Big/ \{2G\}^{1/2} \overset{a}{\sim} N(0,1)$, where E and G some constants (see Section 2).

The S and W tests are defined respectively as: $S = T\sum_{k=1}^{N} \hat{r}^2_{\zeta_1\zeta_2}(k) \overset{a}{\sim} \chi^2(N)$ and $W = \sqrt{T}(\hat{r}\ (1),...,\hat{r}\ (N)) \overset{a}{\sim} N(0,1)$. As far as it concerns the implementation of the W test, the p-values are ordered from lowest to largest. Then, the Rom procedure (described in Section 3) is used to obtain the adjusted levels of significance. The decision to accept or reject H_0 is based on the pairwise comparisons between the ordered p-values and the adjusted levels of significance

Table 2: size of the tests under DGP2 at nominal levels of significance 5% and 10%.

Tests	Lag Selection Parameter / Bandwidth Parameter N									
	10		20		30		40		50	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
					***T* = 200**					
W	3.12	5.84	3.20	5.52	3.72	5.96	3.24	6.24	3.16	5.96
Q Dan	4.48	6.80	4.56	7.84	5.04	8.12	5.32	8.64	5.40	8.96
Q T-H	4.16	6.72	4.60	7.76	4.92	8.20	5.20	8.36	5.36	8.52
Q QS	4.56	7.40	5.04	8.20	5.28	8.52	5.40	9.00	6.00	9.44
S	3.56	7.56	2.92	5.88	2.60	5.12	2.12	4.04	1.76	3.44
					***T* = 500**					
W	3.96	6.84	4.60	7.00	4.72	7.68	5.08	8.60	5.36	8.76
Q Dan	4.20	6.88	4.44	7.56	4.68	8.00	4.84	8.24	4.64	8.84
Q T-H	3.84	6.12	4.44	7.64	4.76	7.36	4.56	8.12	4.84	7.96
Q QS	4.44	7.72	4.44	7.44	4.72	8.00	4.72	8.80	5.32	9.04
S	3.36	6.84	3.76	7.36	3.12	6.52	3.20	5.92	2.36	5.44
					***T* = 1000**					
W	4.32	7.60	4.20	7.72	4.80	8.28	5.20	8.28	5.56	8.68
Q Dan	3.92	6.36	4.52	7.56	5.00	8.76	5.20	9.24	5.40	9.36
Q T-H	3.44	5.76	4.56	7.20	4.68	8.24	5.32	9.20	5.44	9.36
Q QS	4.36	6.72	4.68	8.48	5.40	9.48	5.44	9.36	5.40	9.12
S	3.84	7.92	4.24	8.60	4.04	8.16	3.52	7.28	3.24	6.96
					***T* = 4000**					
W	6.40	10.40	5.80	9.84	5.92	9.76	5.76	9.76	6.40	10.12
Q Dan	3.80	6.24	4.72	7.64	4.72	8.32	4.96	8.28	5.00	8.44
Q T-H	3.76	5.84	4.44	7.16	4.60	8.08	4.86	8.24	5.04	8.52
Q QS	4.04	6.92	4.68	8.20	5.04	8.52	5.00	8.64	5.04	8.84
S	3.84	7.52	3.96	8.52	4.28	8.16	3.88	8.40	4.60	8.20

Notes: a Student's t (6 d.f.) bivariate BEKK (1,1) process is simulated (2500 replications); there is no causality -in -variance between y_{1t} and y_{2t} . Dan, T-H and QS stand for Daniell, Tukey-Hanning and quadratic-spectral kernels (function v(.)) respectively; the one-sided Q tests are defined as: $Q = \left\{ T \sum_{k=1}^{T-1} v^2(k/N) \hat{r}_{\zeta_1\zeta_2}(k) - E \right\} \Big/ \{2G\}^{1/2} \overset{a}{\sim} N(0,1)$ **,** where E and G some constants (see Section 2).

The S and W tests are defined respectively as: $S = T \sum_{k=1}^{N} \hat{r}^2_{\zeta_1\zeta_2}(k) \overset{a}{\sim} \chi^2(N)$ and $W = \sqrt{T}(\hat{r}(1),...,\hat{r}(N)) \overset{a}{\sim} N(0,1)$ **.** As far as it concerns the implementation of the W test, the p-values are ordered from lowest to largest. Then, the Rom procedure (described in Section 3) is used to obtain the adjusted levels of significance. The decision to accept or reject H_0 is based on the pairwise comparisons between the ordered p-values and the adjusted levels of significance

Table 3: power of the tests under DGP3 at nominal levels of significance 5% and 10% .

Tests	Lag Selection Parameter / Bandwidth Parameter N									
	10		20		30		40		50	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
					***T* = 200**					
W	98.00	98.68	99.32	99.44	99.0	99.32	98.88	99.32	98.68	99.24
Q Dan	19.28	25.08	15.48	20.04	13.32	18.00	12.00	16.96	11.32	15.80
Q T-H	20.28	26.20	16.48	21.72	14.48	18.96	13.00	17.72	11.92	16.84
Q QS	17.96	23.84	13.92	18.84	12.24	17.04	11.36	15.64	11.12	14.96
S	9.52	15.24	6.72	11.04	4.96	8.56	3.68	6.72	2.44	5.08
					***T* = 500**					
W	100	100	100	100	100	100	100	100	100	100
Q Dan	43.68	52.28	35.60	42.92	30.36	38.52	27.08	35.04	24.84	32.40
Q T-H	46.24	53.64	37.72	46.60	32.96	41.16	29.28	37.68	26.64	34.68
Q QS	40.08	49.76	32.04	40.40	27.04	35.32	24.56	31.80	21.88	29.88
S	25.04	35.76	17.08	26.20	13.52	21.64	11.48	18.24	9.4	15.24
					***T* = 1000**					
W	100	100	100	100	100	100	100	100	100	100
Q Dan	77.72	83.00	67.96	75.44	60.08	68.52	53.24	62.56	48.44	58.24
Q T-H	79.20	85.00	71.24	78.32	65.00	71.88	57.80	66.72	52.68	61.96
Q QS	74.68	80.80	63.88	70.56	53.88	62.96	47.64	57.28	43.04	52.84
S	54.12	67.20	39.08	51.44	29.84	42.12	25.32	36.52	22.16	32.88
					***T* = 4000**					
W	100	100	100	100	100	100	100	100	100	100
Q Dan	99.92	99.92	99.88	99.88	99.76	99.84	99.68	99.72	99.60	99.68
Q T-H	99.92	99.92	99.88	99.88	99.84	99.88	99.76	99.80	99.72	99.72
Q QS	99.88	99.88	99.84	99.88	99.72	99.76	99.48	99.72	99.00	99.52
S	99.80	99.84	99.28	99.64	98.12	98.92	96.24	98.16	94.04	96.52

Notes: a Normal bivariate BEKK (1,1) process is simulated (2500 replications); y_{2t} Granger cause-in-variance y_{1t} at the first lag period. Dan, T-H and QS stand for Daniell, Tukey-Hanning and quadratic-spectral kernels (function v(.)) respectively; the one-sided Q tests are defined as: $Q = \left\{ T \sum_{k=1}^{T-1} v^2(k/N)\hat{r}_{\zeta_1\zeta_2}(k) - E \right\} \Big/ \{2G\}^{1/2} \overset{a}{\sim} N(0,1)$ **,** where E and G some constants (see Section 2).

The S and W tests are defined respectively as: $S = T\sum_{k=1}^{N} \hat{r}^2_{\zeta_1\zeta_2}(k) \overset{a}{\sim} \chi^2(N)$ and $W = \sqrt{T}(\hat{r}\ (1),...,\hat{r}\ (N)) \overset{a}{\sim} N(0,1)$ **.** As far as it concerns the implementation of the W test, the p-values are ordered from lowest to largest. Then, the Rom procedure (described in Section 3) is used to obtain the adjusted levels of significance. The decision to accept or reject H_0 is based on the pairwise comparisons between the ordered p-values and the adjusted levels of significance

Table 4: power of the tests under DGP4 at nominal levels of significance 5% and 10%.

Tests	Lag Selection Parameter / Bandwidth Parameter *N*									
	10		20		30		40		50	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
					***T* = 200**					
W	94.96	96.28	97.32	98.24	97.00	97.92	96.48	97.48	96.20	97.20
Q Dan	6.68	9.32	62.64	71.36	83.32	88.12	85.32	89.36	84.96	88.68
Q T-H	5.72	8.16	35.80	47.24	78.44	84.16	84.40	88.84	85.40	89.44
Q QS	10.56	15.40	80.16	86.08	85.24	89.44	84.44	88.68	82.28	86.80
S	41.40	48.72	80.72	87.52	68.24	78.16	57.36	67.32	47.28	57.04
					***T* = 500**					
W	100	100	99.96	100	99.96	99.96	99.96	99.96	99.96	99.96
Q Dan	8.32	12.00	97.20	98.36	99.56	99.80	99.56	99.80	99.48	99.68
Q T-H	6.32	8.84	87.04	92.48	99.28	99.68	99.68	99.80	99.68	99.80
Q QS	27.72	38.52	99.40	99.76	99.72	99.80	99.44	99.68	99.24	99.48
S	75.68	82.44	99.68	99.88	98.84	99.08	98.12	98.92	96.48	97.80
					***T* = 1000**					
W	100	100	100	100	100	100	100	100	100	100
Q Dan	10.16	15.00	99.72	99.76	99.88	99.88	99.88	99.88	99.88	99.88
Q T-H	6.68	9.56	98.48	99.04	99.88	99.88	99.88	99.88	99.88	99.88
Q QS	64.48	76.80	99.88	99.88	99.88	99.88	99.88	99.88	99.88	99.88
S	95.00	96.44	99.88	99.88	99.88	99.88	99.72	99.80	99.60	99.68
					***T* = 4000**					
W	100	68.20	100	100	100	100	100	100	100	100
Q Dan	41.28	55.56	99.96	99.96	99.96	99.96	99.96	99.96	99.96	99.96
Q T-H	9.84	14.96	99.88	99.88	99.96	99.96	99.96	100	99.96	99.96
Q QS	99.28	99.56	99.96	100	99.96	100	99.96	99.96	99.96	99.96
S	99.84	99.92	99.96	100	99.96	99.96	99.96	99.96	99.96	99.96

Notes: a Normal bivariate BEKK (10,10) process is simulated (2500 replications); y_{2t} Granger cause-in-variance y_{1t} at 10th lag period; Dan, T-H and QS stand for Daniell, Tukey-Hanning and quadratic-spectral kernels (function ν(.)) respectively; the one-sided Q tests are defined as: $Q=\left\{T\sum_{k=1}^{T-1} v^2(k/N)\hat{r}_{\zeta_1\zeta_2}(k)-E\right\}\Big/\{2G\}^{1/2} \overset{a}{\sim} N(0,1)$ **,** where *E* and *G* some constants (see Section 2).

The *S* and *W* tests are defined respectively as: $S=T\sum_{k=1}^{N}\hat{r}^2_{\zeta_1\zeta_2}(k)\overset{a}{\sim}\chi^2(N)$ and $W=\sqrt{T}(\hat{r}\ (1),...,\hat{r}\ (N))\overset{a}{\sim}N(0,1)$ **.** As far as it concerns the implementation of the *W* test, the p-values are ordered from lowest to largest. Then, the Rom procedure (described in Section 3) is used to obtain the adjusted levels of significance. The decision to accept or reject H_0 is based on the pairwise comparisons between the ordered p-values and the adjusted levels of significance

Table 5: power of the tests under DGP5 at nominal level of significance 5% and 10%.

Tests	Lag Selection Parameter / Bandwidth Parameter N									
	10		20		30		40		50	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
					***T* = 200**					
W	7.52	11.56	12.88	17.20	13.64	18.32	13.72	18.28	13.60	17.48
Q Dan	11.32	15.80	12.60	16.96	12.68	17.84	12.52	18.48	13.00	18.44
Q T-H	10.84	15.16	12.52	16.28	12.64	17.36	13.12	17.80	12.80	18.04
Q QS	12.20	16.28	12.92	17.56	12.72	18.08	12.76	18.40	13.08	19.20
S	9.20	14.60	8.68	13.36	7.04	11.68	5.92	9.84	4.56	7.56
					***T* = 500**					
W	14.16	18.56	23.08	27.60	24.48	29.60	24.84	30.28	25.44	30.88
Q Dan	15.12	19.56	16.44	21.40	16.76	21.72	16.56	22.28	16.48	22.36
Q T-H	14.32	18.48	16.00	21.68	16.56	21.48	16.52	22.04	16.60	21.88
Q QS	15.76	21.00	16.80	21.48	16.52	22.04	16.64	22.16	16.76	22.52
S	12.40	18.16	12.32	18.56	11.88	18.04	11.00	16.52	10.24	15.64
					***T* = 1000**					
W	21.80	27.16	34.96	41.32	37.12	42.88	38.64	43.96	39.56	44.88
Q Dan	19.36	24.28	21.56	28.12	22.92	28.80	22.76	29.24	22.76	29.28
Q T-H	17.96	22.36	21.12	26.64	21.76	28.12	22.64	28.80	22.56	29.04
Q QS	20.16	25.28	22.00	28.44	22.40	28.88	23.04	29.00	23.36	28.88
S	17.12	25.16	17.12	24.64	17.32	24.60	17.20	24.40	16.48	23.84
					***T* = 4000**					
W	52.80	56.92	70.12	74.08	70.16	73.80	71.24	74.76	71.72	74.60
Q Dan	44.64	50.48	48.52	55.16	49.44	56.32	49.52	56.84	48.96	56.88
Q T-H	41.64	48.04	47.52	54.16	48.32	55.40	48.96	56.00	48.64	56.28
Q QS	46.64	52.52	48.32	55.68	48.92	56.20	48.32	56.28	48.36	55.92
S	40.84	51.28	40.40	51.20	40.44	50.00	40.36	48.80	40.12	50.20

Notes: two Normal FIGARCH (1,1) processes are simulated (2500 replications); the degree of fractional integration for both processes is 0.4 (long memory); y_{2t} Granger cause-in-variance y_{1t} at the 1[st] lag period. Dan, T-H and QS stand for Daniell, Tukey-Hanning and quadratic-spectral kernels (function ν(.)) respectively; the one-sided Q tests are defined as:
$Q = \left\{ T \sum_{k=1}^{T-1} v^2(k/N)\hat{r}_{\zeta_1\zeta_2}(k) - E \right\} \Big/ \{2G\}^{1/2} \overset{a}{\sim} N(0,1)$, where E and G some constants (see Section 2).

The S and W tests are defined respectively as: $S = T\sum_{k=1}^{N} \hat{r}^2_{\zeta_1\zeta_2}(k) \overset{a}{\sim} \chi^2(N)$ and $W = \sqrt{T}(\hat{r}\ (1),...,\hat{r}\ (N)) \overset{a}{\sim} N(0,1)$. As far as it concerns the implementation of the W test, the p-values are ordered from lowest to largest. Then, the Rom procedure (described in Section 3) is used to obtain the adjusted levels of significance. The decision to accept or reject H_0 is based on the pairwise comparisons between the ordered p-values and the adjusted levels of significance

Table 6: power of the tests under DGP5 at nominal level of significance 5% and 10%.

Tests	Lag Selection Parameter / Bandwidth Parameter N									
	10		20		30		40		50	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
					***T* = 200**					
W	13.52	16.76	24.80	28.84	26.36	31.04	27.44	32.04	27.56	32.36
Q Dan	18.16	23.24	20.48	25.32	21.24	26.40	20.72	26.40	20.20	26.36
Q T-H	16.68	21.68	20.24	24.60	21.24	26.28	20.88	26.44	20.52	26.28
Q QS	19.28	24.36	21.40	26.64	20.80	26.44	20.44	26.48	20.24	26.72
S	15.96	22.04	13.96	20.96	11.80	17.68	9.72	15.24	8.00	11.88
					***T* = 500**					
W	24.68	28.88	41.52	46.84	43.20	48.40	44.80	49.56	45.64	50.24
Q Dan	26.92	31.40	30.80	35.64	31.36	36.76	30.76	37.28	30.64	37.16
Q T-H	24.68	29.16	29.88	35.04	31.16	36.28	30.84	36.84	31.08	36.84
Q QS	28.72	33.12	31.04	36.24	30.84	37.00	30.72	37.12	30.32	36.68
S	24.68	32.64	24.68	32.12	23.76	31.12	22.64	28.80	20.12	27.36
					***T* = 1000**					
W	37.12	41.52	57.80	61.92	59.92	63.52	61.12	64.88	61.96	65.96
Q Dan	41.04	46.12	45.04	51.04	45.04	51.20	45.00	50.84	44.92	51.04
Q T-H	37.84	43.28	43.92	49.68	44.84	51.40	44.72	50.76	44.60	50.84
Q QS	42.88	48.00	44.52	51.48	44.76	50.68	44.64	50.80	44.32	50.72
S	38.24	46.84	37.20	45.88	36.80	44.52	36.28	43.92	35.48	42.68
					***T* = 4000**					
W	71.52	74.76	87.56	88.96	87.52	89.08	87.68	89.32	87.92	89.48
Q Dan	83.52	86.16	86.84	89.12	87.08	89.52	86.36	89.16	85.52	88.64
Q T-H	80.76	84.00	86.44	88.40	87.00	89.36	86.72	89.20	86.28	88.92
Q QS	85.04	87.64	86.96	89.24	86.20	89.00	85.24	88.52	84.60	88.00
S	82.48	87.04	81.16	86.28	80.20	84.76	79.48	84.20	78.40	83.32

Notes: two Normal FIGARCH (1,1) processes are simulated (2500 replications); the degree of fractional integration for both processes is 0.6 (short memory); y_{2t} Granger cause-in-variance y_{1t} at the 1st lag period. Dan, T-H and QS stand for Daniell, Tukey-Hanning and quadratic-spectral kernels (function ν(.)) respectively; the one-sided Q tests are defined as: $Q = \left\{ T \sum_{k=1}^{T-1} \nu^2(k/N) \hat{r}_{\zeta_1\zeta_2}(k) - E \right\} \Big/ \{2G\}^{1/2} \overset{a}{\sim} N(0,1)$, where E and G some constants (see Section 2).

The S and W tests are defined respectively as: $S = T \sum_{k=1}^{N} \hat{r}^2_{\zeta_1\zeta_2}(k) \overset{a}{\sim} \chi^2(N)$ and $W = \sqrt{T}(\hat{r}(1),...,\hat{r}(N)) \overset{a}{\sim} N(0,1)$. As far as it concerns the implementation of the W test, the p-values are ordered from lowest to largest. Then, the Rom procedure (described in Section 3) is used to obtain the adjusted levels of significance. The decision to accept or reject H_0 is based on the pairwise comparisons between the ordered p-values and the adjusted levels of significance

Table 7: results of the structural break pre-testing on the volatility of real stock returns and industrial production growth.

	Real Stock Returns		Industrial Production Growth	
	Squared returns	Absolute Returns	Squared Growth	Absolute Growth
	US			
KL test	0.3359	0.1005	0.7427	0.1678
Change-point date	01/06/2008	01/09/1996	01/01/1984	01/01/1984
	United Kingdom			
KL test	0.7959	0.8006	**3.0494**	**2.5196**
Change-point date	01/10/1992	01/01/1993	01/07/1987	01/07/1990
	Italy			
KL test	**1.2661**	0.9943	0.7270	0.2908
Change-point date	01/10/1986	01/01/1999	01/02/1985	01/01/1988
	Canada			
KL test	0.3821	0.7216	0.8635	0.3569
Change-point date	01/01/1988	01/05/1983	01/01/1985	01/01/1987

Notes: the KL test is computed as $\max V_T(j)/s$, where $V_T(j) = T^{-1/2}\left(\sum_{i=1}^{j} u_i - \frac{j}{T}\sum_{i=1}^{T} u_i\right)$, u equals to r^2 or $|r|$, r is the series of logarithmic returns/growths, and s is the standard deviation of the realized volatilities. A VARHAC estimator is implemented for the calculation of s. Under the null hypothesis of homogeneity in volatility the KL converges to the sup of a Brownian Bridge. The critical values at levels of significance 5% and 10% are 1.36 and 1.22 respectively. Values of the test statistic in bold are statistical significant at level 5% or /and 10%.

Table 8: testing the null hypothesis that real stock market returns do not Granger cause-in-variance industrial growth rates in USA, United Kingdom, Italy and Canada at levels of significance 5% and 10%.

	Test Statistics						
	W			*Q*			*S*
		Level α^*		Dan	TH	QS	
N		5%	10%				
				USA			
10	2.1512 (0.0315)	0.0051	0.0105	-0.6076 (0.7283)	-0.3466 (0.6356)	-0.7086 (0.7607)	4.5652 (0.9183)
20	2.1512 (0.0315)	0.0026	0.0053	-0.9453 (0.8277)	-0.8053 (0.7897)	-1.0463 (0.8523)	13.1465 (0.8710)
30	2.1512 (0.0315)	0.0017	0.0035	-1.2335 (0.8913)	-0.9975 (0.8407)	-1.2723 (0.8984)	20.4133 (0.9053)
40	2.1512 (0.0315)	0.0013	0.0026	-1.3779 (0.9159)	-1.1595 (0.8769)	-1.5102 (0.9345)	22.5635 (0.9881)
				United Kingdom			
10	2.0189 (0.0435)	0.0051	0.0105	-1.3521 (0.9118)	-1.1936 (0.8837)	-1.5060 (0.9340)	2.6119 (0.9891)
20	2.0189 (0.0435)	0.0026	0.0053	-1.8211 (0.9657)	-1.6699 (0.9525)	-1.8787 (0.9699)	9.6934 (0.9734)
30	2.0189 (0.0435)	0.0017	0.0035	-1.8131 (0.9651)	-1.8626 (0.9687)	-1.9963 (0.9771)	13.2534 (0.9965)
40	2.0189 (0.0435)	0.0013	0.0026	-1.9106 (0.9720)	-1.9291 (0.9731)	-2.1005 (0.9822)	16.8101 (0.9995)
				Italy			
10	2.0694 (0.0385)	0.0051	0.0105	-0.2080 (0.5824)	-0.2790 (0.6099)	-0.0103 (0.5041)	13.8889 (0.1781)
20	2.0694 (0.0385)	0.0026	0.0053	0.4515 (0.3258)	0.1976 (0.4217)	0.3906 (0.3480)	20.0015 (0.4578)
30	2.0694 (0.0385)	0.0017	0.0035	0.3023 (0.3812)	0.3823 (0.3511)	0.4567 (0.3240)	31.5874 (0.3870)
40	**3.4688** **(0.0005)**	0.0013	0.0026	0.3734 (0.3544)	0.3710 (0.3553)	0.7900 (0.2148)	**52.9506** **(0.0825)**
				Canada			
10	2.0049 (0.0450)	0.0051	0.0105	-0.2349 (0.5929)	-0.1344 (0.5535)	-0.4460 (0.6722)	5.9571 (0.8189)
20	2.0049 (0.0450)	0.0026	0.0053	-0.6080 (0.7284)	-0.5732 (0.7167)	-0.7846 (0.7837)	13.9838 (0.8313)
30	2.8489 (0.0044)	0.0017	0.0035	-0.9039 (0.8170)	-0.7531 (0.7743)	-0.7866 (0.7842)	25.6187 (0.6944)
40	2.8489 (0.0044)	0.0013	0.0026	-0.7632 (0.7773)	-0.8268 (0.7958)	-0.7429 (0.7712)	33.0217 (0.7749)

Notes: p-values are in parentheses. Numbers in bold indicate rejection of the null hypothesis at levels of statistical significance 5% and 10%. *N* refers to the lag selection parameter used in the *W* and *S* test calculations, and the bandwidth parameter for the *Q* tests calculations. Dan, T-H and QS stand for Daniell, Tukey-Hanning and quadratic-spectral kernels (function ν(.)) respectively. The Q tests and the *S* test are defined in Section 2 (equations (1)- (4)). The *W* test is defined as $W = \sqrt{T}(\hat{r}\ (1),...,\hat{r}\ (N)) \overset{a}{\sim} N(0, I_N)$. We only report the specific value of *W* test, $W_k = \sqrt{T}\hat{r}\ (k)$, where $P_k \le \alpha_K^*$, with P_k representing the p-value of the test and α_k^* the corresponding adjusted significance level. Levels α^* are calculated as $a_{N-i+1}^* = \left\{ \sum_{h=1}^{i-1} a^h - \sum_{h=1}^{i-2} \binom{i}{h} a_{N-h}^{*(i-h)} \right\} / i$, where $i = 2, \ldots, N$ and $a_{[N]}^*$ is set to be either 5% or 10%. The *W* test value with the smaller p-value is presented in the cases where the previous inequality does not hold.

Table 9: testing the null hypothesis that industrial production growth rate does not Granger cause-in-variance real stock market returns in USA, United Kingdom, Italy and Canada at levels of significance 5% and 10%.

	Test Statistics						
	W			Q			S
		Level α^*		Dan	TH	QS	
N		5%	10%				
				USA			
10	**3.1579**	0.0051	0.0105	1.1339	**1.5391**	0.7407	8.4687
	(0.0016)			(0.1284)	**(0.0619)**	(0.2295)	(0.5832)
20	**3.1579**	0.0026	0.0053	-0.1626	0.3102	-0.5001	11.7652
	(0.0016)			(0.5646)	(0.3782)	(0.6915)	(0.9239)
30	**3.1579**	0.0017	0.0035	-0.8043	-0.4026	-1.1057	16.4788
	(0.0016)			(0.7894)	(0.6564)	(0.8656)	(0.9783)
40	**3.1579**	0.0013	0.0026	-1.2325	-0.8309	-1.5464	20.1144
	(0.0016)			(0.8911)	(0.7970)	(0.9390)	(0.9963)
				United Kingdom			
10	**4.2144**	0.0051	0.0105	0.3545	0.2898	0.2815	10.3139
	(0.0000)			(0.3615)	(0.3860)	(0.3891)	(0.4134)
20	**3.6779**	0.0027	0.0055	0.2615	0.2687	0.2214	19.0271
	(0.0002)			(0.3969)	(0.3941)	(0.4124)	(0.5201)
30	**3.6779**	0.0018	0.0036	0.1686	0.2434	0.0092	26.6353
	(0.0002)			(0.4331)	(0.4038)	(0.4963)	(0.6423)
40	**3.6779**	0.0013	0.0027	-0.0303	0.1098	-0.1044	35.5267
	(0.0002)			(0.5121)	(0.4563)	(0.5416)	(0.6718)
				Italy			
10	**3.3664**	0.0057	0.0116	1.2343	0.9708	**1.4751**	**16.5214**
	(0.0008)			(0.1085)	(0.1658)	**(0.0701)**	**(0.0856)**
20	**3.3664**	0.0027	0.0055	**1.4585**	**1.4902**	1.2071	23.4596
	(0.0008)			**(0.0724)**	**(0.0681)**	(0.1137)	(0.2668)
30	**3.3664**	0.0018	0.0036	1.1287	1.2466	1.0826	38.7730
	(0.0008)			(0.1295)	(0.1063)	(0.1395)	(0.1309)
40	**3.3664**	0.0013	0.0027	1.1393	1.0959	1.0501	45.1971
	(0.0008)			(0.1273)	(0.1366)	(0.1468)	(0.2639)
				Canada			
10	1.5872	0.0051	0.0105	0.3291	0.3807	0.1490	7.5421
	(0.1125)			(0.3711)	(0.3517)	(0.4408)	(0.6735)
20	1.5872	0.0026	0.0053	-0.2898	-0.0946	-0.7164	10.9541
	(0.1125)			(0.6140)	(0.5377)	(0.7631)	(0.9474)
30	1.5872	0.0017	0.0035	-0.6784	-0.6614	-1.2886	14.8635
	(0.1125)			(0.7512)	(0.7458)	(0.9012)	(0.9905)
40	2.5957	0.0013	0.0026	-1.2859	-1.1077	-1.0517	32.1075
	(0.0094)			(0.9008)	(0.8660)	(0.8535)	(0.8085)

Notes: p-values are in parentheses. Numbers in bold indicate rejection of the null hypothesis at levels of statistical significance 5% and 10%.. *N* refers to the lag selection parameter used in the *W* and *S* test calculations, and the bandwidth parameter for the *Q* tests calculations. Dan, T-H and QS stand for Daniell, Tukey-Hanning and quadratic-spectral kernels (function ν(.)) respectively. The Q tests and the *S* test are defined in Section 2 (equations (1)- (4)). The *W* test is defined as $W = \sqrt{T}(\hat{r}(1),...,\hat{r}(N)) \overset{a}{\sim} N(0, I_N)$ **.** We only report the specific value of *W* test, $W_k = \sqrt{T}\hat{r}(k)$, where $P_k \leq \alpha_K^*$, with P_k representing the p-value of the test and α_k^* the corresponding adjusted significance level. Levels α^* are calculated as $a_{N-i+1}^* = \left\{ \sum_{h=1}^{i-1} a^h - \sum_{h=1}^{i-2} \binom{i}{h} a_{N-h}^{*(i-h)} \right\} / i$ **,** where $i = 2, \ldots, N$ and $a_{[N]}^*$ is set to be either 5% or 10%. *W* test value with the smaller p-value is presented in the cases where the previous inequality does not hold.

Proof of Theorem

A similar reasoning to Haugh (1976) and Cheung and Ng (1996) is used to prove Theorem 1. Consider that two time series $z_{it}, i = 1,2, t = 1,...,T,$ are generated separately by two autoregressive processes of order *p*

$$\varphi_i(B)z_{it} = \varepsilon_{it}, \tag{A.1}$$

where

i) $\varphi_i(B)$ is the polynomial in the lag operator B of length *p*, i.e., $\varphi_i(B) = 1 - \varphi_{i1}B - ... - \varphi_{ip}B^p$, and

ii) $\{\varepsilon_{it}\}$ are two sequences of white noise random variables, with $E(\varepsilon_{it}) = 0$, $E(\varepsilon_{it}\varepsilon_{is}) = 0, \forall t \neq s$, and $E(\varepsilon_{it}^4) < \infty$ for all *t*.

iii) independence between z_{1t} and z_{2t} is required.

Let $\tilde{\varphi}_i$ be some arbitrary coefficient values, estimated by OLS or maximum likelihood, while $\tilde{\varepsilon}_{it}$ denote the residual series which correspond to these coefficient values.

The sample cross-correlation function of $|\tilde{\varepsilon}_{it}|$ for some $k > 0$ is defined as

$$\tilde{r}(k) = \frac{\tilde{d}_{|\varepsilon_1\varepsilon_2|}(k)}{\left[\tilde{d}_{|\varepsilon_1|}(0)\tilde{d}_{|\varepsilon_2|}(0)\right]^{1/2}}, \tag{A.2}$$

where $\tilde{d}_{|\varepsilon_1\varepsilon_2|}(k) = T^{-1}\sum_{t=k+1}^{T}\left(|\tilde{\varepsilon}_{1t}| - \mu_{|\varepsilon_1|}\right)\left(|\tilde{\varepsilon}_{2t-k}| - \mu_{|\varepsilon_{21}|}\right)$, $\tilde{\mu}_{|\varepsilon_i|} = T^{-1}\sum_{t=1}^{T}|\tilde{\varepsilon}_{it}|$, and $\tilde{d}_{|\varepsilon_i|}(0) = T^{-1}\sum_{t=1}^{T}\left(|\tilde{\varepsilon}_{it}| - \mu_{|\varepsilon_i|}\right)^2$.

By using a similar notation, we define the terms $\hat{r}(k)$, $r(k)$, $\hat{d}_{|\varepsilon_1\varepsilon_2|}(k)$, $d_{|\varepsilon_1\varepsilon_2|}(k)$, $\hat{d}_{|\varepsilon_i|}(0)$, $d_{|\varepsilon_i|}(0)$, $\hat{\mu}_{|\varepsilon_i|}$, and $\mu_{|\varepsilon_i|}$.

Note that

$$\tilde{d}_{|\varepsilon_1\varepsilon_2|}(k) = T^{-1}\sum_{t=k+1}^{T}\left(\left|\tilde{\varepsilon}_{1t}\right| - \mu_{|\varepsilon_1|}\right)\left(\left|\tilde{\varepsilon}_{2t-k}\right| - \mu_{|\varepsilon_{21}|}\right) = T^{-1}\sum_{t=k+1}^{T}\left|\tilde{\varepsilon}_{1t}\tilde{\varepsilon}_{2t-k}\right| - T^{-2}(2-T^{-1})\sum_t\left|\tilde{\varepsilon}_{1t}\right|\sum_t\left|\tilde{\varepsilon}_{2t}\right|$$

$$= T^{-1}\sum_{t=k+1}^{T}\left|\tilde{\varepsilon}_{1t}\tilde{\varepsilon}_{2t-k}\right| - (2-T^{-1})\tilde{\mu}_{|\varepsilon_1|}\tilde{\mu}_{|\varepsilon_2|}$$

To prove Theorem 1, we first show that

$$\frac{\partial\tilde{d}_{|\varepsilon_1\varepsilon_2|}(k)}{\partial\tilde{\varphi}_{il}} = O_p(1/T^{1/2})$$

for each $\tilde{\varphi}_{il}, l = 1,.., p$

Proof:

The sample cross-covariance function $\tilde{d}_{|\varepsilon_1\varepsilon_2|}(k)$ is written as:

$$\tilde{d}_{|\varepsilon_1\varepsilon_2|}(k) = T^{-1}\sum_{t=k+1}^{T}\left|\tilde{\varepsilon}_{1t}\tilde{\varepsilon}_{2t-k}\right| - (2-T^{-1})\tilde{\mu}_{|\varepsilon_1|}\tilde{\mu}_{|\varepsilon_2|} =$$

$$= T^{-1}\sum_{t=k+1}^{T}\left|\tilde{\varphi}_1(B)z_{1t}\tilde{\varphi}_2(B)z_{2t-k}\right| - (2-T^{-1})T^{-1}\sum_{t=1}^{T}\left|\tilde{\varphi}_1(B)z_{1t}\right|T^{-1}\sum_{t=1}^{T}\left|\tilde{\varphi}_2(B)z_{2t}\right|$$

For simplicity in the calculations, the previous expression is written as

$$\tilde{d}_{|\varepsilon_1\varepsilon_2|}(k) = \tilde{d}_{E_1} + (T^{-1}-2)\tilde{d}_{E_2}$$

where

$$\tilde{d}_{E_1} = T^{-1}\sum_{t=k+1}^{T}\left|\tilde{\varepsilon}_{1t}\tilde{\varepsilon}_{2t-k}\right| = T^{-1}\sum_{t=k+1}^{T}\left|\tilde{\varphi}_1(B)z_{1t}\tilde{\varphi}_2(B)z_{2t-k}\right|$$

and $\tilde{d}_{E_2} = T^{-1}\sum_{t=1}^{T}\left|\tilde{\varepsilon}_{1t}\right|T^{-1}\sum_{t=1}^{T}\left|\tilde{\varepsilon}_{2t}\right|$.

First, we compute $\partial\tilde{d}_{E_1}/\partial\varphi_{1l}$. Given the realization $\{z_{1T}, z_{1T-1},..., z_{11}\}$, each term $\left|\tilde{\varepsilon}_{1t}\right|$ in the sum $\sum_{t=k+1}^{T}\left|\tilde{\varepsilon}_{1t}\tilde{\varepsilon}_{2t-k}\right|$ is substituted recursively,

$$\left|\tilde{\varepsilon}_{1T}\tilde{\varepsilon}_{2T-k}\right| = \left|z_{1T} - \tilde{\varphi}_{11}z_{1T-1} - \tilde{\varphi}_{12}z_{1T-2}... - \tilde{\varphi}_{1p}z_{1T-p}\right|\left|\tilde{\varepsilon}_{2T-k}\right|$$

$$\left|\tilde{\varepsilon}_{1T-1}\tilde{\varepsilon}_{2T-k-1}\right| = \left|z_{1T-1} - \tilde{\varphi}_{11}z_{1T-2} - \tilde{\varphi}_{12}z_{1T-3}... - \tilde{\varphi}_{1p}z_{1T-p}\right|\left|\tilde{\varepsilon}_{2T-k-1}\right|$$

$$\left|\tilde{\varepsilon}_{1T-2}\tilde{\varepsilon}_{2T-k-2}\right| = \left|z_{1T-2} - \tilde{\varphi}_{11}z_{1T-3} - \tilde{\varphi}_{12}z_{1T-4}... - \tilde{\varphi}_{1p}z_{1T-p}\right|\left|\tilde{\varepsilon}_{2T-k-2}\right|$$

.

.

$$\left|\tilde{\varepsilon}_{1k+1}\tilde{\varepsilon}_{21}\right| = \left|z_{1k+1} - \tilde{\varphi}_{11}z_{1k} - \tilde{\varphi}_{12}z_{1k-1} \dots - \tilde{\varphi}_{1p}z_{1k-(p-1)}\right|\left|\tilde{\varepsilon}_{21}\right|$$

Derivatives $\partial\tilde{d}_{|\varepsilon_1\varepsilon_2|}(k)\big/\partial\varphi_{il}$ exist by assumption. Calculating the derivatives of $\tilde{d}_{E_1} = \sum_{t=k+1}^{T}\left|\tilde{\varepsilon}_{1t}\tilde{\varepsilon}_{2t-k}\right|$ with respect to each parameter $\tilde{\varphi}_{1l}, l = 1,.., p$, yields

$$\frac{\partial\tilde{d}_{E_1}}{\partial\tilde{\varphi}_{11}} = -z_{1T-1}\frac{\tilde{\varphi}_1(B)z_{1T}}{\left|\tilde{\varphi}_1(B)z_{1T}\right|}\left|\tilde{\varepsilon}_{2T-k}\right| - z_{1T-2}\frac{\tilde{\varphi}_1(B)z_{1T-1}}{\left|\tilde{\varphi}_1(B)z_{1T-1}\right|}\left|\tilde{\varepsilon}_{2T-k-1}\right| - z_{1T-3}\frac{\tilde{\varphi}_1(B)z_{1T-2}}{\left|\tilde{\varphi}_1(B)z_{1T-2}\right|}\left|\tilde{\varepsilon}_{2T-k-2}\right| - \dots =$$
$$= -z_{1T-1}\frac{\tilde{\varphi}_1(B)z_{1T}}{\left|\tilde{\varphi}_1(B)z_{1T}\right|}\left|\tilde{\varphi}_2(B)z_{2T-k}\right| - z_{1T-2}\frac{\tilde{\varphi}_1(B)z_{1T-1}}{\left|\tilde{\varphi}_1(B)z_{1T-1}\right|}\left|\tilde{\varphi}_2(B)z_{2T-k-1}\right| - z_{1T-3}\frac{\tilde{\varphi}_1(B)z_{1T-2}}{\left|\tilde{\varphi}_1(B)z_{1T-2}\right|}\left|\tilde{\varphi}_2(B)z_{2T-k-2}\right| - ..$$

$$\frac{\partial\tilde{d}_{E_1}}{\partial\tilde{\varphi}_{12}} = -z_{1T-2}\frac{\tilde{\varphi}_1(B)z_{1T}}{\left|\tilde{\varphi}_1(B)z_{1T}\right|}\left|\tilde{\varepsilon}_{2T-k}\right| - z_{1T-3}\frac{\tilde{\varphi}_1(B)z_{1T-1}}{\left|\tilde{\varphi}_1(B)z_{1T-1}\right|}\left|\tilde{\varepsilon}_{2T-k-1}\right| - z_{1T-4}\frac{\tilde{\varphi}_1(B)z_{1T-2}}{\left|\tilde{\varphi}_1(B)z_{1T-2}\right|}\left|\tilde{\varepsilon}_{2T-k-2}\right| - \dots$$
$$= -z_{1T-2}\frac{\tilde{\varphi}_1(B)z_{1T}}{\left|\tilde{\varphi}_1(B)z_{1T}\right|}\left|\tilde{\varphi}_2(B)z_{2T-k}\right| - z_{1T-3}\frac{\tilde{\varphi}_1(B)z_{1T-1}}{\left|\tilde{\varphi}_1(B)z_{1T-1}\right|}\left|\tilde{\varphi}_2(B)z_{2T-k-1}\right| - z_{1T-4}\frac{\tilde{\varphi}_1(B)z_{1T-2}}{\left|\tilde{\varphi}_1(B)z_{1T-2}\right|}\left|\tilde{\varphi}_2(B)z_{2T-k-2}\right| - .$$

etc

Thus, for each l =1, 2, …,p,

$$\frac{\partial\tilde{d}_{E_1}}{\partial\tilde{\varphi}_{1l}} = -\sum_t z_{1t-2}\frac{\tilde{\varphi}_1(B)z_{1t}}{\left|\tilde{\varphi}_1(B)z_{1t}\right|}\left|\tilde{\varphi}_2(B)z_{2t-k}\right|$$

Similarly, to calculate the derivatives of $\tilde{d}_{E_2} = T^{-1}\sum_{t=1}^{T}\left|\tilde{\varepsilon}_{1t}\right|T^{-1}\sum_{t=1}^{T}\left|\tilde{\varepsilon}_{2t}\right|$ first consider that

$$\tilde{d}_{E_2} = T^{-1}\left(\left|\tilde{\varepsilon}_{1T}\right| + \left|\tilde{\varepsilon}_{1T-1}\right| + \left|\tilde{\varepsilon}_{1T-2}\right| \dots\right)T^{-1}\sum_{t=1}^{T}\left|\tilde{\varepsilon}_{2t}\right|$$

By recursive substitution,

$$\tilde{d}_{E_2} = T^{-1}\left(\left|z_{1T} - \tilde{\varphi}_{11}z_{1T-1} - \tilde{\varphi}_{12}z_{1T-2} - \dots\right| + \left|z_{1T-1} - \tilde{\varphi}_{11}z_{1T-2} - \tilde{\varphi}_{12}z_{1T-3} - \dots\right| + \dots\right)T^{-1}\sum_{t=1}^{T}\left|\tilde{\varepsilon}_{2t}\right|$$

Thus, the derivatives for each $\tilde{\varphi}_{1l}, l = 1,.., p$, are

$$\frac{\partial \tilde{d}_{E_2}}{\partial \tilde{\varphi}_{11}} = T^{-1}\left(-z_{1T-1}\,\frac{\tilde{\varphi}_1(B)z_{1T}}{|\tilde{\varphi}_1(B)z_{1T}|} - z_{1T-2}\,\frac{\tilde{\varphi}_1(B)z_{1T-1}}{|\tilde{\varphi}_1(B)z_{1T-1}|} - z_{1T-3}\,\frac{\tilde{\varphi}_1(B)z_{1T-2}}{|\tilde{\varphi}_1(B)z_{1T-2}|} - \dots \right) T^{-1}\sum_{t=1}^{T}|\tilde{\varepsilon}_{2t}|$$

$$\frac{\partial \tilde{d}_{E_2}}{\partial \tilde{\varphi}_{12}} = T^{-1}\left(-z_{1T-2}\,\frac{\tilde{\varphi}_1(B)z_{1T}}{|\tilde{\varphi}_1(B)z_{1T}|} - z_{1T-3}\,\frac{\tilde{\varphi}_1(B)z_{1T-1}}{|\tilde{\varphi}_1(B)z_{1T-1}|} - z_{1T-4}\,\frac{\tilde{\varphi}_1(B)z_{1T-2}}{|\tilde{\varphi}_1(B)z_{1T-2}|} - \dots \right) T^{-1}\sum_{t=1}^{T}|\tilde{\varepsilon}_{2t}|$$

etc.

Therefore, for each $\tilde{\varphi}_{1l}, l = 1,.., p$,

$$\frac{\partial \tilde{d}_{E_2}}{\partial \tilde{\varphi}_{1l}} = -T^{-1}\sum_t |\tilde{\varepsilon}_{2t}| T^{-1}\sum_t z_{1t-1}\,\frac{\tilde{\varphi}_1(B)z_{1t}}{|\tilde{\varphi}_1(B)z_{1t}|}$$

or

$$\frac{\partial \tilde{d}_{E_2}}{\partial \tilde{\varphi}_{1l}} = -T^{-1}\sum_t |\tilde{\varphi}_2(B)z_{2t}| T^{-1}\sum_t z_{1t-1}\,\frac{\tilde{\varphi}_1(B)z_{1t}}{|\tilde{\varphi}_1(B)z_{1t}|}$$

Since $\frac{\partial \tilde{d}_{|\varepsilon_1\varepsilon_2|}(k)}{\partial \tilde{\varphi}_{1l}} = \frac{\partial \tilde{d}_{E_1}}{\partial \tilde{\varphi}_{1l}} + (T^{-1}-2)\frac{\partial \tilde{d}_{E_2}}{\partial \tilde{\varphi}_{1l}}$, we show that

$$\frac{\partial \tilde{d}_{|\varepsilon_1\varepsilon_2|}(k)}{\partial \tilde{\varphi}_{1l}} = T^{-1}\sum_t M_{1t}M_{2t-k}$$

where $M_{1t} = -\frac{\tilde{\varphi}_1(B)z_{1t}}{|\tilde{\varphi}_1(B)z_{1t}|}z_{1t-1}$,

$$M_{2t-k} = |\tilde{\varphi}_2(B)z_{2t-k}| - (2-T^{-1})T^{-1}\sum_t |\tilde{\varphi}_2(B)z_{2t}|$$

Under the assumption that $\{M_{it}\}, i = 1,2$, are two independent, and stationary processes with finite fourth order moments, Theorem 14 of Hannah (1970, p.228) states that $\sqrt{T}\left(T^{-1}\sum c_t\right)$, where $c_t = M_{1t}M_{2t-k}$, is asymptotically distributed as normal, with mean zero and variance $p = \rho_1\rho_2$, where ρ_1 and ρ_2 are the autocorellation functions of processes M_{1t} and M_{2t-k} respectively. Processes

$|\varepsilon_{1t}|$ and $|\varepsilon_{2t-k}|$ in M_{1t} and M_{2t-k} respectively, are independent, and stationary because they are functions of the independent, and stationary processes ε_{1t} and ε_{2t}. Thus, from Hannah 's theorem it follows that $\partial\tilde{d}_{|\varepsilon_1\varepsilon_2|}(k)\big/\partial\tilde{\varphi}_{1l}$ is $O_p(1/T^{1/2})$. The derivatives $\partial\tilde{d}_{|\varepsilon_1\varepsilon_2|}(k)\big/\partial\tilde{\varphi}_{2l}$ can be treated in the same way. Therefore, for each l =1,…,p, and i =1,2, $\dfrac{\partial\tilde{d}_{|\varepsilon_1\varepsilon_2|}(k)}{\partial\tilde{\varphi}_{il}} = O_p(1/T^{1/2})$.

Then, we approximate $\tilde{d}_{|\varepsilon_1\varepsilon_2|}(k)$ with a Taylor expansion of first order about $\tilde{\varphi} = \hat{\varphi}$

$$\hat{d}_{|\varepsilon_1\varepsilon_2|}(k) = d_{|\varepsilon_1\varepsilon_2|}(k) + \sum_i \sum_l \left(\hat{\varphi}_{li} - \varphi_{li}\right)\frac{\partial\tilde{d}_{|\varepsilon_1\varepsilon_2|}(k)}{\partial\varphi_{li}}$$

We saw previously that $\partial\tilde{d}_{|\varepsilon_1\varepsilon_2|}(k)\big/\partial\tilde{\varphi}_{il} = O_p(1/T^{1/2})$, while it is established that $\hat{\varphi}_{il} = \varphi_{il} + O_p(1/T^{1/2})$, thus,

$$\hat{d}_{|\varepsilon_1\varepsilon_2|}(k) = d_{|\varepsilon_1\varepsilon_2|}(k) + O_p(1/T) \ .$$

Next, we approximate $\tilde{d}_{|\varepsilon_1|}(0)$ with a Taylor expansion of first order about $\tilde{\varphi} = \hat{\varphi}$

$$\hat{d}_{|\varepsilon_1|}(0) = d_{|\varepsilon_1|}(0) + \sum_i \sum_l \left(\hat{\varphi}_{li} - \varphi_{li}\right)\frac{\partial\tilde{d}_{|\varepsilon_1|}(0)}{\partial\varphi_{li}}$$

Next, we show that

$$\frac{\partial\tilde{d}_{|\varepsilon_1|}(0)}{\partial\tilde{\varphi}_{1l}} = O_P(1/T^{1/2})$$

Proof:

Again, sample variance estimator $\tilde{d}_{|\varepsilon_1|}(0) = T^{-1}\sum_{t=1}^{T}\left(|\tilde{\varepsilon}_{1t}| - T^{-1}\sum_t |\tilde{\varepsilon}_{1t}|\right)^2$ is calculated by recursive substitution. Hence, each term of the sum is written as

$$\left(|\tilde{\varepsilon}_{1T}|-T^{-1}\sum_t|\tilde{\varepsilon}_{1t}|\right)^2=\left(|z_{1T}-\tilde{\varphi}_{11}z_{1T-1}-\ldots|-T^{-1}\left(|\tilde{\varepsilon}_{1T}|+|\tilde{\varepsilon}_{1T-1}|+\ldots\right)\right)^2$$

$$\left(|\tilde{\varepsilon}_{1T-1}|-T^{-1}\sum_t|\tilde{\varepsilon}_{1t}|\right)^2=\left(|z_{1T-1}-\tilde{\varphi}_{11}z_{1T-2}-\ldots|-T^{-1}\left(|\tilde{\varepsilon}_{1T}|+|\tilde{\varepsilon}_{1T-1}|+\ldots\right)\right)^2$$

.

.

or

$$\left(|\tilde{\varepsilon}_{1T}|-T^{-1}\sum_t|\tilde{\varepsilon}_{1t}|\right)^2=\left(|z_{1T}-\tilde{\varphi}_{11}z_{1T-1}-\ldots|-T^{-1}\left(|z_{1T}-\tilde{\varphi}_{11}z_{1T-1}-\ldots|+|z_{1T-1}-\tilde{\varphi}_{11}z_{1T-2}-\ldots|+\ldots\right)\right)^2$$

$$\left(|\tilde{\varepsilon}_{1T-1}|-T^{-1}\sum_t|\tilde{\varepsilon}_{1t}|\right)^2=\left(|z_{1T-1}-\tilde{\varphi}_{11}z_{1T-2}-\ldots|-T^{-1}\left(|z_{1T}-\tilde{\varphi}_{11}z_{1T-1}-\ldots|+|z_{1T-1}-\tilde{\varphi}_{11}z_{1T-2}-\ldots|+\ldots\right)\right)^2$$

.

.

Assuming that $\partial\tilde{d}_{|\varepsilon_i|}/\partial\tilde{\varphi}_{1l}$ exists for each l = 1, 2,…,p, straightforward calculations yield

$$\frac{\partial\tilde{d}_{|\varepsilon_1|}(0)}{\partial\tilde{\varphi}_{11}}=2*\left(|\tilde{\varepsilon}_{1T}|-T^{-1}\sum_t|\tilde{\varepsilon}_{1t}|\right)\left(-z_{1T-1}\frac{\tilde{\varphi}_1(B)z_{1T}}{|\tilde{\varphi}_1(B)z_{1T}|}-T^{-1}\left(-z_{1T-1}\frac{\tilde{\varphi}_1(B)z_{1T}}{|\tilde{\varphi}_1(B)z_{1T}|}-z_{1T-2}\frac{\tilde{\varphi}_1(B)z_{1T-1}}{|\tilde{\varphi}_1(B)z_{1T-1}|}-\ldots\right)\right)+$$
$$+2*\left(|\tilde{\varepsilon}_{1T-1}|-T^{-1}\sum_t|\tilde{\varepsilon}_{1t}|\right)\left(-z_{1T-2}\frac{\tilde{\varphi}_1(B)z_{1T-1}}{|\tilde{\varphi}_1(B)z_{1T-1}|}-T^{-1}\left(-z_{1T-1}\frac{\tilde{\varphi}_1(B)z_{1T}}{|\tilde{\varphi}_1(B)z_{1T}|}-z_{1T-2}\frac{\tilde{\varphi}_1(B)z_{1T-1}}{|\tilde{\varphi}_1(B)z_{1T-1}|}-\ldots\right)\right)+\ldots$$

$$\frac{\partial\tilde{d}_{|\varepsilon_1|}(0)}{\partial\tilde{\varphi}_{12}}=2*\left(|\tilde{\varepsilon}_{1T}|-T^{-1}\sum_t|\tilde{\varepsilon}_{1t}|\right)\left(-z_{1T-2}\frac{\tilde{\varphi}_1(B)z_{1T}}{|\tilde{\varphi}_1(B)z_{1T}|}-T^{-1}\left(-z_{1T-2}\frac{\tilde{\varphi}_1(B)z_{1T}}{|\tilde{\varphi}_1(B)z_{1T}|}-z_{1T-3}\frac{\tilde{\varphi}_1(B)z_{1T-1}}{|\tilde{\varphi}_1(B)z_{1T-1}|}-\ldots\right)\right)+$$
$$+2*\left(|\tilde{\varepsilon}_{1T-1}|-T^{-1}\sum_t|\tilde{\varepsilon}_{1t}|\right)\left(-z_{1T-3}\frac{\tilde{\varphi}_1(B)z_{1T-1}}{|\tilde{\varphi}_1(B)z_{1T-1}|}-T^{-1}\left(-z_{1T-2}\frac{\tilde{\varphi}_1(B)z_{1T}}{|\tilde{\varphi}_1(B)z_{1T}|}-z_{1T-3}\frac{\tilde{\varphi}_1(B)z_{1T-1}}{|\tilde{\varphi}_1(B)z_{1T-1}|}-\ldots\right)\right)+\ldots$$

etc.

Thus, for each l = 1, 2,…,p we result in

$$\frac{\partial\tilde{d}_{|\varepsilon_1|}(0)}{\partial\tilde{\varphi}_{1l}}=T^{-1}\sum_t K_{1t}K_{1t-1}$$

where $K_{1t} = -2\left(|\tilde{\varphi}_1(B)z_{1t}| - T^{-1}\sum_t |\tilde{\varphi}_1(B)z_{1t}| \right),$

$$K_{1t-1} = z_{1t}\frac{\tilde{\varphi}_1(B)z_{1t+1}}{|\tilde{\varphi}_1(B)z_{1t+1}|} + T^{-1}\sum_t z_{1t-1}\frac{\tilde{\varphi}_1(B)z_{1t}}{|\tilde{\varphi}_1(B)z_{1t}|}$$

Using a similar reasoning as previously, by Theorem 14 of Hannah, $\partial\tilde{d}_{|\varepsilon_1|}(0) \big/ \partial\tilde{\varphi}_{1l} = O_p(1/T^{1/2})$.

Thus,

$$\hat{d}_{|\varepsilon_1|}(0) = d_{|\varepsilon_1|}(0) + O_p(1/T)$$

In the same way we prove that $\hat{d}_{|\varepsilon_2|}(0) = d_{|\varepsilon_2|}(0) + O_p(1/T)$. Then, implementing a technique as in Cheung and Ng (1996),

$$\begin{aligned}\hat{d}_{|\varepsilon_1|}(0)\hat{d}_{|\varepsilon_2|}(0) &= \left(d_{|\varepsilon_1|}(0) + O_p(1/T)\right)\left(d_{|\varepsilon_2|}(0) + O_p(1/T)\right) = \\ &= d_{|\varepsilon_1|}(0)d_{|\varepsilon_2|}(0) + O_p(1/T\ \ )\end{aligned}$$

Similarly,

$$\begin{aligned}\hat{d}_{|\varepsilon_1\varepsilon_2|}(k)\left[\hat{d}_{|\varepsilon_1|}(0)\hat{d}_{|\varepsilon_1|}(0)\right]^{-1/2} &= (d_{|\varepsilon_1\varepsilon_2|}(k) + O_p(1/T))\left(d_{|\varepsilon_1|}(0)d_{|\varepsilon_2|}(0) + O_p(1/T\ \ )\right)^{-1/2} = \\ &= d_{|\varepsilon_1\varepsilon_2|}(k)\left(d_{|\varepsilon_1|}(0)d_{|\varepsilon_2|}(0)\right)^{-1/2} + O_p(1/T^{1/2})\end{aligned}$$

or equivantely,

$$\sqrt{T}\hat{d}_{|\varepsilon_1\varepsilon_2|}(k)\left[\hat{d}_{|\varepsilon_1|}(0)\hat{d}_{|\varepsilon_1|}(0)\right]^{-1/2} = \sqrt{T}d_{|\varepsilon_1\varepsilon_2|}(k)\left(d_{|\varepsilon_1|}(0)d_{|\varepsilon_2|}(0)\right)^{-1/2} + O_p(1)$$

Thus,

$$\sqrt{T}\hat{r}(k) = \sqrt{T}d_{|\varepsilon_1\varepsilon_2|}(k)\left(d_{|\varepsilon_1|}(0)d_{|\varepsilon_2|}(0)\right)^{-1/2} + O_p(1)$$

By Hannah's Theorem, $\sqrt{T}d_{|\varepsilon_1\varepsilon_2|}(k)\left(d_{|\varepsilon_1|}(0)d_{|\varepsilon_2|}(0)\right)^{-1/2} \xrightarrow{d} N(0,1)$

since $d_{|\varepsilon_1|}(0)$ and $d_{|\varepsilon_2|}(0)$ converge in probability to the variances of $|\varepsilon_{1t}|$ and $|\varepsilon_{2t}|$ respectively.

Therefore, since

$\sqrt{T}\hat{r}(k) \xrightarrow{p} \sqrt{T}r(k)$, where $r(k) = d_{|\varepsilon_1\varepsilon_2|}(k)\left(d_{|\varepsilon_1|}(0)d_{|\varepsilon_2|}(0)\right)^{-1/2}$

and $\sqrt{T}r(k) \xrightarrow{d} N(0,1)$

by Lemma 4.7 of White (1984, page 67),

$\sqrt{T}\hat{r}(k) \xrightarrow{d} N(0,1)$

Hence, we establish that $\sqrt{T}\hat{r}(k)$ is distributed asymptotically under the null hypothesis of non-causality-in-variance as standard normal.

Chapter 4

Predicting output growth at long horizons: stock return volatility and monetary policy

1. Introduction

Stock prices are equal to the discounted values of the infinite stream of future expected firms' cash flows. Schwert (1989) states that under the rational expectations hypothesis, volatility of stock returns should reflect changes in the conditions of real activity growth at future distant periods. To date, the literature has come to a general consensus that stock return volatility may affect consumption spending via a household wealth effect, business fixed investment by means of capital cost effect, and government spending through taxes. Of greater concern is the cost of capital channel. Changes in stock return volatility influence the level of compensation that risk-averse investors request for bearing extra systematic risk. Hence, movements on expected returns may affect the cost of equity capital which, in turn, has an impact on investment. Moreover, Bernanke and Gertler (1999) argue that, variation of stock returns affects the condition of the balance sheets of firms, households and financial institutions, and as a consequence their ability to borrow and lend. Extreme rises or falls in credit flows creates distortions in consumption and investment, leading to severe fluctuations in real economic activity growth in the long run.

So far there is significant empirical evidence linking stock returns to output growth and stock returns to stock return volatility. For the former, stock returns entail expectations about future firms ' cash flows and discount rates because investors are forward looking; see Fama (1980, 1981, 1990), Geske and Roll (1983), Mandelker and Tandon (1985), James, Koreisha and Partch (1985), Schwert (1990), Barro (1990), Chen(1991), Lee (1992), Peiro (1996), Estrella and Mishkin (1998), Forni, et al. (2001) among others[23]. For the latter, Merton 's (1973) Intertemporal Capital Asset Pricing model states that there is a positive relation between stock returns and volatility because investors demand extra compensation for bearing extra risk; Whitelaw (1994) demonstrates that in the volatility of US stock returns leads expected returns through the business cycle.

Given these results, it is primarily an empirical question whether stock return volatility and real activity growth are associated. Schwert (1989) and Hamilton and Lin (1996) show that US aggregate stock return volatility increases in economic contractions. Whitelaw (1994) finds evidence that the commercial paper-Treasury bill yield spread anticipates changes of aggregate stock return volatility in US, which in turn anticipates changes of the expected returns. Campbell, et al. (2001) document that US monthly market, industry and firm-level volatility measures predict recessions up to one year-ahead. They also show that these measures have significant in-sample predictive ability for GDP growth rate one quarter ahead.

While the ability of stock return volatility for predicting real activity growth at short forecast horizons is well documented, the nature of their intertemporal relation remains unexplored. Hence our interest is whether information about subsequent industrial production growth rates is spread across current and recent periods in monthly stock return volatility. In our evaluations, macroeconomic and monetary factors, such as the short-term interest rates, the money supply growth rates, and inflation, are allowed to operate as intermediate processes which in theory[24] should help to communicate this information over the course of time. Our investigation is motivated by the seminal contributions of Fama (1990), Schwert (1990) and Lee (1992) that show how stock returns can help to predict output growth at subsequent time periods ahead. Our goal is to reconcile the results of Whitelaw (1994) and Campbell, et al. (2001) with earlier empirical findings on the intertemporal association between stock returns and real activity growth.

Fama (1990) regresses monthly US industrial production growth rates on stock returns at increasing time lags. He finds that lagged stock returns explain to some extent the variability of industrial production growth rates at distant periods.

[23] Stock and Watson (2003) surveyed the out-of-sample forecasting performance of both US stock returns' mean and volatility for output growth.

[24] Fama (1980) and Geske and Roll (1983) associate the linkage between the stock returns and real activity growth with inflation and monetary policy variables based on money demand and supply arguments.

Moreover, Fama documents that annual stock returns have an enhanced predictive ability for annual production growth rates at long horizons. These findings have largely been confirmed by Schwert (1990) who implements Fama's long horizon regressions on US data extended back to 1889.Lee (1992) uses impulse response functions and innovation accounting (i.e., forecast error variance decompositions) for long horizon causality testing in the Sims sense[25] between US monthly stock returns and industrial production growth rates, including short-term interest rates and inflation in his multivariate vector-autoregression (VAR) specification. His evidence suggests that there is Granger causal priority from stock returns to industrial production growth rates 24 months ahead, while the latter responds positively to shocks from the former up to 12 months ahead.

A weakness in Fama's work is that he does not explicitly model the joint dynamics of stock returns and real activity growth rates with monetary policy indicators. Since Lee's (1992) study, there have been many advances in the theoretical and empirical evaluation of long horizon causality relations between multivariate economic time series. For instance, Dufour and Tessier (1993) show that Sims's conditions for Granger non-causality between two random variables in terms of impulse responses coefficients (and consequently, the innovation accounting) are not sufficient when more than two time series are included in the setting. According to this result, for multivariate time series the use of the impulse response function and the innovation accounting up to a certain horizon may lead to misleading conclusions about the presence of a Granger causality relation.

Dufour and Renault (1998) are the first to present a theoretical multivariate framework on the notion of *h*-step ahead non-causality[26], referred as long (or short) horizon non-causality. The authors provide definitions and a set of conditions which ensure the equivalence between standard Wiener - Granger (1956, 1969) type one-step ahead non-causality and non-causality at any forecast period. Characterization of non-causality between two variables at a specific horizon *h* is achieved by setting zero

[25] Sims (1972, 1980a,b, 1982) is the first author who addressed the issue of Granger causality in terms of predictability several- periods ahead. According to Sims (1980a,b, 1982), non-zero impulse responses between the innovations at some time period ahead imply the presence of a Granger causal relation at the specific horizon. Sims also considers as a measure of Granger type causal priority the innovation accounting (i.e., forecast error variance decompositions) defined as the proportion of the *h*-step-ahead forecast error variance of one variable due to the innovations of an another variable.

[26] The first definition of *h*-step ahead Granger non-causality is due to Lütkepohl (1993).

restrictions on the coefficients of a multivariate linear invertible process. The implications of these general conditions are also investigated, by deriving exact testable restrictions on the parameters of a finite-order VAR model. Their generalization describes a Wiener - Granger causal relationship at a specific horizon, acknowledging all possible indirect effects[27] of the auxiliary series of the system at the previous time periods.

Dufour and Renault's conditions on non-causality between two variables at forecast horizon greater than one, with more than two time series included in the multivariate system, involves examining the statistical significance of multi-linear zero restrictions on the coefficients of the VAR parameters. Therefore, the task of testing such hypotheses using likelihood ratio or Lagrange multiplier tests becomes very challenging given the difficulty of estimating conditional specifications under the null hypothesis. However, the required zero coefficient restrictions can be evaluated by using a Wald test. Lütkepohl and Müller (1994) and Lütkepohl and Burda (1997) propose modified Wald statistics to test the *h*-step ahead noncausality hypothesis. These tests are shown to have a valid asymptotic distribution under the null even when these highly nonlinear zero coefficient restrictions violate the regularity condition of a usual Wald test. Still, the proposed tests yield poor finite sample performance. An alternative test procedure is proposed by Dufour, Pelletier and Renault (2006). Their method requires the estimation of parametric mean regressions denoted as '(*p*,*h*)- autoregressions'. Inference is conducted by testing simple zero coefficient restrictions on the parameters of the '(*p*,*h*)- autoregressions' via an asymptotic chi-square Wald test. The authors also introduce a parametric Monte Carlo procedure to calculate *p*-values, to ensure enhanced finite sample properties. Hill (2007) proposes a sequential Bonferroni type causality test for trivariate processes, while Eichler (2007) evaluates causality relations between multivariate series up to a certain time horizon by means of path diagrams. More recently, Al Sadoon (2010) suggests procedures for long horizon causality testing in subspaces. Dufour and Taamouti (2010) present a linear measure of *h*-period ahead causal priority in the Granger sense between two time series in a multivariate system[28].

---

[27] Dufour and Renault define indirect causality similarly with Hsiao (1983): although one variable, say x, may not predict another variable, say y, directly, it may still have predictive power for variable y through their association with an another variable, say z.

[28] These test procedures are defined in terms of conditional distributions, whose implementation is based on the estimation of mean regression models. Therefore, they may fail to reveal high-order

Our goal is to present a complete analysis of the multiple horizon causal linkages between the volatility of stock returns and industrial production growth rates, when a set of auxiliary processes, such as money supply (M2) growth rates, inflation, and short-term interest rates, is included in the setting. This paper builds on previous research, such as Whitelaw (1994), and Campbell et al. (2001), although it distinguishes itself from the rest of the literature by its explicit focus on the evaluation of the relation between output growth and stock return volatility at both short and long forecast periods in terms of a multivariate system.

First we consider multiple horizon Granger non-causality testing by implementing the econometric procedure of Dufour et al. (2006) on data from four developed economies, namely US, Germany, Japan, and Italy. Following Schwert (1989), Campbell, et al. (2001), we impose no functional form on the evolution of the stock return volatility dynamics. We document a large number of highly significant direct influences from the volatility of stock returns to output growth at both short and long horizons in all four economies. These findings are consistent with the results of Whitelaw (1994), and Campbell et al. (2001), while they also reveal that stock return volatility presents an enhanced ability to predict production growth rates at distant forecast periods. Interestingly, we also find that stock return volatility indirectly causes monthly growth rates of industrial production at long horizons through (i) nominal short-term interest rates in US, and Germany, (ii) money supply growth rates in Japan and (iii) inflation in Italy. These results have not been previously shown in the literature. Moreover, we confirm the finding of Dufour, Pelletier and Renault (2006), Hill (2007), and Dufour and Taamouti (2010) that monetary policy is causal for real activity growth at both short and long horizons.

Second, one may argue that policy makers, investors and portfolio managers are ultimately concerned with the magnitude of the forecasting performance of stock return volatility for output growth, as well as the presence of such predictive relationship. Hence, the intensity of these causality relations is quantified by applying the h-step-ahead causality measure introduced by Dufour and Taamouti (2010). Asymptotic valid confidence intervals are also constructed using a bootstrap

moment associations and nonlinear causal linkages between the time series. Furthermore, Hill (2007) argues that Dufour, Renault, and Pelletier (2006) test procedure fails to provide a clear-cut answer on whether the presence of an exclusive causality relation at horizon h is preceded by neutralized causal effects over the horizons 1 to $h - 1$ (i.e., possible multiple indirect effects may cancel each other out) or by a complete absence of indirect influences between the series over these horizons.

technique presented by the authors, in order to back up the evidence from inference. The outcome of causality measurement largely supports our hypothesis test results. Our main findings suggest that stock return volatility proxy induces significant long horizon spillover effects of large magnitude on output growth in Germany, Japan and Italy. However, in Japan inflation appears to be the leading source of changes at distant periods in real activity growth rates, as well as the stock return volatility. The volatility of the US stock returns appears to strongly cause output growth at distant forecast periods indirectly through money supply growth rates. These strong effects appear in economies with different degrees of market capitalization and proportions of share of household wealth represented by the stocks.

Third, an out-of-sample forecasting exercise using moderate sample sizes shows that combining stock return volatility, short-term interest rates, money supply growth, and inflation in a single regression model yields more accurate long horizon forecasts of output growth than the autoregressive benchmark model. In particular, the use of the particular forecasting relation appears to be beneficial for a wide range of distant forecast periods in accord with our in-sample findings in all four economies. We also find that (static) single indicator specifications based on the causal structure revealed by the hypothesis testing and linear measurement of in-sample predictive power produce superior short-term forecasts of industrial production growth rates relative to the autoregressive (dynamic) model when using the pooling of forecasts across different estimation windows to ensure forecasting stability or a large simulation scheme to offset small sample estimation bias.

The remainder of the paper is organized as follows. In the next section we briefly describe the econometric procedures used in our investigations. Data and empirical results are described in Section 3 and 4, respectively. Section 5 presents an out-of-sample forecasting exercise. Section 6 concludes.

2. Description of Econometric procedures

In this section, we briefly discuss two methodologies for testing and measuring *h*-step-ahead noncausality.

2.1. Testing for *h*-step-ahead Granger non-causality

Multivariate non-causation at different time horizons is tested by using the econometric procedure of Dufour , Pelletier and Renault (2006). The authors present an estimation and inference procedure for testing *h*-step ahead non-causality hypothesis in finite multivariate stationary or nonstationary VAR models. In particular, first they introduce an estimation method denoted as '(*p*,*h*)-autoregressions'. Granger non-causality up to horizon h is tested by evaluating the relevant coefficient restrictions on the (*p*,*h*)- autoregressions parameters.

Let $V_t = [v_{1t},...,v_{mt}]$ be a second order stationary process, where v_{it}, $i = 1,...,m$ are $T \times 1$ vectors. The method of (*p*,*h*)- autoregressions requires that the process V_{t+h} is regressed on $(V_T,...,V_1)$;

$$Z^{(h)} = B^{(h)}P^{(h)} + U^{(h)}, \tag{1}$$

where $Z^{(h)} = (V_{p+h-1},...,V_T)$, $B^{(h)} = (\kappa, \pi_1^{(h)},...,\pi_p^{(h)})$, $P^{(h)} = (P_{(h)p-1}, P_{(h)p},..., P_{(h)T})$, $P_{(h)t} = (1, V_t, V_{t-1},...,V_{t-p+1})'$, $U^{(h)} = (U_{(h)p-1}, U_{(h)p},...,U_{(h)T})$, p is the lag order, κ is an integer, and $U_{(h)t}$ represents the innovations of each (*p*,*h*)- autoregression calculated as $U_{(h)t} = V_{t+h} - B^{(h)}P_{(h)t}$.

Coefficient matrix $B^{(h)}$ is estimated by the method of least squares, $\hat{B}^{(h)} = Z^{(h)}P^{(h)'}\left(P^{(h)}P^{(h)'}\right)^{-1}$. The innovations series of the (*p*,*h*)- autoregressions evolve as a MA(*h*-1) process, and as a consequence, the usual OLS estimator of the variance-covariance matrix of $\hat{B}^{(h)}$ cannot be used. Instead, a heteroskedasticity and autocorrelation consistent (HAC) estimator is preferred, such as the Newey-West (1987) estimator ,

$$\hat{D} = \sum_{k=0}^{l(T)} \left(1 - \left(|k| / l(T)\right)\right) Q, \tag{2}$$

where $Q = \mathrm{cov}\left(P^{(h)}U^{(h)}, P^{(h)}U^{(h)}\right)$, $l(T)$ is the bandwidth parameter, $\lim_{T\to\infty} l(T) = \infty$, and $\lim_{T\to\infty} \left(l(T)/T^{1/4}\right) = 0$. Following Dufour et al. (2006) we set *l* (*T*) = *h* -1. Hence, the variance-covariance matrix of $\hat{B}^{(h)}$ can be consistently estimated as

$$\hat{G} = \Gamma \hat{D} \Gamma', \tag{3}$$

where $\Gamma = \left(P^{(h)} P^{(h)'} / T \right) \otimes I_m$.

The null hypothesis that v_{1t} Granger causes v_{2t} for a specific horizon *h* is defined in terms of zero restrictions on the coefficients of $\hat{B}^{(h)}$:

$$H_o : \left[\hat{\pi}^{(h)}_{21 \cdot r}\right]_{r=1}^{p} = 0 \tag{4}$$

To test the null hypothesis of *h*-step-ahead non-causality, the authors propose an asymptotic chi-square Wald test:

$$W = T\left(C(Y\hat{G}Y')^{-1}C'\right) \xrightarrow{d} \chi^2(p) \tag{5}$$

where $C = Y \cdot \hat{B}^{(h)}$, $Y = \left(0_{p\times 1}, I_p \otimes R', 0_{p\times m}\right) \otimes L'$, while *L* and *R* are two vectors of size $T \times 1$ whose elements are all equal to zero except for a unit value at the position that v_{jt} and v_{it} hold at matrices $Z^{(h)}$ and $P^{(h)}$ respectively. The simulation results of the authors show that inference based on the asymptotic critical values may be misleading due to size distortions. Therefore, their approach for obtaining the *p*-values is to perform a parametric Monte Carlo procedure[29]. Their method can be summarized in two steps:

---

[29] More details on the specific Monte Carlo procedure can be found in Dufour and Jouini (2005).

First, estimates $\hat{B}^{(1)}$ and $\hat{\Sigma}_U = \hat{U}^{(1)}\hat{U}^{(1)\prime} / (T-p)$ are obtained by fitting (p,1)-autoregressions to the series. Similarly, estimates $\hat{B}^{(h)}$ are calculated by fitting (p,h)-autoregressions to the series. Then, the W test is computed as presented in (5).

Second, the estimates $\hat{B}^{(1)}$ are used to calculate the impulse responses $\psi_{(h)}$ for each forecast horizon as defined by

$$\psi_{(h)} = F\hat{B}^{(1)h}F', \tag{6}$$

$$\psi_{(0)} = I_m,$$

where $F = (I_m, 0, ..., 0)$ is a $m \times mp$ matrix.

Then a pseudo-random data set w_t^R consisting of m series is calculated, where t = 1, 2, .., T. The pseudo-data are drawn from a normal distribution with mean zero and variance $\hat{\Sigma}_U$. Next, letting B^R denote a restricted version of $\hat{B}^{(h)}$, i.e., $B^R = (\kappa, \pi_1^{(h)*}, ..., \pi_p^{(h)*})$,

where for each r = 1,...,p we set the following zero coefficient restriction

$[\hat{\pi}_{21}^{(h)}] = 0$

For each forecast horizon h , m-variate time series V_t^R, t = 1,…, T, are constructed based on the data generating mechanism:

$$V_t^R = \kappa + \sum_{r=1}^{p} \pi_r^{(h)*} V_{t-r}^R + \psi_{(h)} w_t^R \tag{7}$$

N replications[30] of V_t^R are generated, and for each sample of simulated data, we apply the same estimation and testing procedure as applied to the actual data. The pseudo test values are denoted as $\tilde{W}$. The decision rule is to reject the null hypothesis of non-causality from v_{1t} to v_{2t} for a specific forecast horizon h at level of significance α if $\hat{p}_N(x) \le \alpha$, where

$$\hat{p}_N(x) = (N+1)^{-1}\left\{1 + \sum_{n=1}^{N} 1(\tilde{W}_n > W)\right\}. \tag{8}$$

[30] The same number of replications is used (i.e., N = 1000) as in Dufour et al (2006)

2.2. **Measurement of *h*-step ahead Granger causality**

Dufour and Taamouti (2010) propose measures for Granger *h*-step ahead causality which evaluate the magnitude of a causality relation between random variables at a specific time horizon *h*. The proposed measures are extensions of Geweke 's (1982) one period ahead causality measures. Let the matrix V_t be partitioned into $V_t = (v_{1t}, v_{2t}, v_{qt})$, where v_{1t}, v_{2t} are two $T \times 1$ vectors and v_{qt} is a $T \times (m-2)$ matrix with auxiliary variables. Assume that V_t evolves as an (*p*,1)- autoregression

$$Z^{(1)} = B^{(1)} P^{(1)} + U^{(1)}, \tag{9}$$

where $Z^{(1)} = (V_p, ..., V_T)$, $B^{(1)} = (\pi_1^{(1)}, ..., \pi_p^{(1)})$, $P^{(1)} = (P_{(1)p-1}, P_{(1)p}, ..., P_{(1)T})$, $P_{(1)t} = (V_t, V_{t-1}, ..., V_{t-p+1})'$, $U^{(1)} = (U_{(1)p-1}, U_{(1)p}, ..., U_{(1)T})$. Matrices $B^{(1)}$ and $U^{(1)}$ are estimated as described in Section 2.1. The variance-covariance matrix of the error term $U^{(1)}$ is estimated as

$$\hat{\Sigma}_U = \hat{U}^{(1)} \hat{U}^{(1)\prime} / (T - p), \tag{10}$$

while the variance-covariance matrix of the forecast error of v_{2t+h} is estimated as

$$\hat{\Sigma}_{U(h)} = \sum_{z=0}^{h-1} R \psi_{(z)} \hat{\Sigma}_U \psi'_{(z)} R', \tag{11}$$

where the impulse responses ψ are calculated as described in (6), and $R = (0, 1, 0, .., 0)$ is a $1 \times m$ vector. Assume that we want to measure the intensity of forecast improvement that variable v_{1t} has for variable v_{2t} at forecast period *h*. Consider the marginal process $V_t^* = (v_{2t}, v_{qt})$ and the dynamics of the process V_t^* be governed by an autoregression of order (*p*,1)

$$Z^{(1)*} = B^{(1)^*} P^{(1)*} + U^{(1)*}, \tag{12}$$

where $Z^{(1)*}, B^{(1)^*}, P^{(1)*}$ and $U^{(1)*}$ are defined as previously. The variance-covariance matrices of the innovations $U^{(1)*}$ and the forecast errors of v_{2t+h} respectively as

$$\hat{\Sigma}_U^* = \hat{U}^{(1)*}\hat{U}^{(1)*'} \Big/ (T-p) \tag{13}$$

$$\hat{\Sigma}_{U(h)}^* = \sum_{z=0}^{h-1} R^* \psi_{(z)}^* \hat{\Sigma}_U^* \psi_{(z)}^{*'} R^{*'}, \tag{14}$$

where impulse responses ψ^* are calculated as described in (6), and $R^* = (1, 0,...,0)$ is a $1 \times (m-1)$ vector. Dufour and Taamouti (Theorem 5.1, p. 48) state that a measure of unidirectional Granger causality at horizon *h* is defined as

$$CM\left[v_{1t} \underset{h}{\rightarrow} v_{2t}\right] = \ln\left\{\det\left(\hat{\Sigma}_{U(h)}^*\right) \Big/ \det\left(\hat{\Sigma}_{U(h)}\right)\right\}. \tag{15}$$

The authors propose a bootstrap procedure to construct confidence intervals for the causality measure at forecast horizon *h*. Their method can be described in 5 steps:

Step 1, (*p*,1)- autoregressions for V_t are estimated by OLS while the estimates $\hat{B}^{(1)} = \left(\hat{\pi}_1^{(1)},...,\hat{\pi}_p^{(1)}\right)$ and the innovations $\hat{U}^{(1)}$ are retained.

Step 2, the causality measure $CM\left[v_{1t} \underset{h}{\rightarrow} v_{2t}\right]$ is estimated using the original data sample.

Step 3, the method of random sampling with replacement is implemented on $\hat{U}^{(1)}$ to construct a matrix of $T \times m$ bootstrap residuals $\tilde{U}^{(1)}$. Then, a bootstrap sample $\tilde{V}_t$ of size $T \times m$ is generated as function of the model

$$\tilde{V}_t = \sum_{r=1}^{p} \hat{\pi}_r^{(1)} \tilde{V}_{t-r} + \tilde{U}^{(1)} \tag{16}$$

based on $\hat{B}^{(1)}$, $\tilde{U}^{(1)}$, while the original data observations $\{V_1,...,V_p\}$ are used as the *p* initial values of $\tilde{V}_t$.

Step 4, steps (1) and (3) are repeated *N* times, and as a consequence, sequences of $\left\{\tilde{V}_{t(n)}\right\}_{n=1}^{N}$ are obtained. For each replication $\tilde{V}_{t(n)}$ we apply the same estimation procedure as applied to the actual data, to calculate the OLS estimates $\left\{\tilde{B}_n^{(1)}\right\}_{n=1}^{N}$. Dufour and Taamouti adopt Kilian 's (1998) approach to achieve approximately unbiased estimates of $B^{(1)}$. The term $\hat{B}^{(1)} - bias$, where $bias = N^{-1}\sum_{n=1}^{N}\tilde{B}_n^{(1)} - \hat{B}^{(1)}$, is used in (15) to generate new bootstrap data and the corresponding estimates of $B^{(1)}$, denoted $\left\{\tilde{B}_n^{(1)0}\right\}_{n=1}^{N}$. Next, a bias-corrected bootstrap estimator $B^{(1)\#} = \tilde{B}^{(1)0} - bias$, where *bias* is estimated as discussed previously, is employed to the calculations of the new bootstrap replications $\left\{\tilde{V}_{t(n)}^{\#}\right\}_{n=1}^{N}$.

Step 5, the corresponding causality measures $\left\{\tilde{CM}_{(n)}\left[v_{1t} \underset{h}{\rightarrow} v_{2t}\right]\right\}_{n=1}^{N}$ are estimated by applying the estimation method on the bootstrap samples $\left\{\tilde{V}_{t(n)}^{\#}\right\}_{n=1}^{N}$.

The authors also apply a bias correction directly to the causality measure itself:

$$\tilde{CM}_{(n)}^{\#}\left[v_{1t} \underset{h}{\rightarrow} v_{2t}\right] = \tilde{CM}_{(n)}\left[v_{1t} \underset{h}{\rightarrow} v_{2t}\right] - \left(N^{-1}\sum_{n=1}^{N}\tilde{CM}_{(n)}\left[v_{1t} \underset{h}{\rightarrow} v_{2t}\right] - CM\left[v_{1t} \underset{h}{\rightarrow} v_{2t}\right]\right). \quad (17)$$

A non-negativity truncation is also imposed on the bootstraped causality measures:

$$\tilde{CM}_{(n)}^{\#}\left[v_{1t} \underset{h}{\rightarrow} v_{2t}\right] = \max\left\{\tilde{CM}_{(n)}^{\#}\left[v_{1t} \underset{h}{\rightarrow} v_{2t}\right], 0\right\}. \quad (18)$$

The empirical analysis is performed using the MATLAB programming language.

3. Data Description

Our investigation focus on four developed countries: US, Germany, Italy, and Japan. The data are monthly seasonally adjusted observations of the aggregate stock price index (dividends are included), nominal short term interest rates, the consumer price index (CPI), the industrial production index, and the money supply (M2). All series are retrieved from Datastream. The sample period spans from January 1973 to September 2011. We focus on the most recent period, because, during this time these four economies have experienced different financial crises, and because also this sample period has been neglected in the literature to date. All time series, including short-term interest rates, are transformed to logarithmic first differences. Real stock returns are computed by subtracting inflation rate from nominal stock returns. CPI series are seasonally adjusted based on the ratio to moving average method.

Estimating univariate autoregressive models for each real stock returns series, and then taking the squared values of the innovations construct a simple proxy for the volatility of stock returns[31]. Implementation of the Akaike information criterion for selecting the number of lags indicates that autoregressions of order 1, 3, 1, and 3 are appropriate for United States, Germany, Japan, and Italy, respectively.

4. Empirical results

Following Dufour et al. (2006), we employ the sequential method proposed by Tiao and Box (1981) to determine the lag order of the (p,1)-autoregressions. We test the hypothesis of P lags against P + 1 lags by means of the likelihood ratio test over a range of values (1, …, 16). Sequential likelihood ratio tests indicate that a VAR (9) model for United States, a VAR (13) model for Germany, a VAR (12) model for Japan, and a VAR (12) model for Italy are sufficient parameterizations.

Tables 1 to 4 present the results of the causality test for US, Germany, Japan and Italy. Each table reports the Wald test of the null hypothesis of non-causality for the specific forecast horizon, and the simulated p-values at significance level 10% (in parentheses). We consider causality testing for the horizons from 1 to 36 months ahead based on the result of Proposition 4.5 of Dufour and Renault (1998). Statistical significant causal relations at level 10% appear in bold. Panel A presents the results when testing for direct causality from a variable to industrial production growth rates for the forecast periods 1 to 36 months ahead. Panel B reports the test results when investigating for possible indirect effects on output growth.

Table 1a demonstrates the results for US. Volatility of stock returns (denoted as V) is causal for industrial production growth rates (denoted as IP) for forecast horizons 8, 13, 16, 18, 19, 20, 21, 29, 30, 31, 32, 33, 34, 35, and 36 at level of significance 10%. Inflation (denoted as IN) does not Granger cause output growth for any horizon, while short-term interest rates (denoted as SI) predict real activity growth only for h = 30. As an instrument of monetary policy, money supply growth rates (denoted as M2) have predictive power for industrial production growth rates at horizons 20, 22, 23, 26, 30, 31, 32, 33, 34, 35 and 36 at nominal level 10%. Surprisingly, we find a large number of long horizon causality results for money

[31] Morana (2009) constructs a similar volatility proxy when analyzing the co-behaviour patterns between exchange rates and different macroeconomic variables.

supply and especially, the stock return volatility proxy. In the majority of the results, the null hypothesis of non-causality is rejected even at level of significance 5%.

Table 1b reports the results, when investigating for the presence of indirect long horizon causal effects. Although stock return volatility predicts short-term interest rates one and two months ahead, it does not predict money supply growth rates at any horizon. Our findings indicate that short-term interest rates have predictive content for money supply growth rates at both short (i.e., h = 1 and 2 months ahead) and long horizons (i.e., h = 20, 23, 24, 25, 26, 27, 28, 29, 30, 32, 33, and 34 months ahead). Money supply growth rates also Granger cause short-term interest rates for short horizons (i.e., 5, 6, and 9 months ahead). Note that our direct test results indicate that short-term interest rates have a limited predictive ability for future output growth. However, short-term interest rates are found to be causal for money supply growth rates at short horizons, which in turn are causal for output growth at long forecast periods. Therefore, short-term interest rates indirectly influence output growth through money supply growth rates. Moreover, the volatility of stock returns causes short-term interest rates, which in turn are found to be causal for real activity growth at distant time periods ahead. Hence, there is evidence that the volatility of stock returns and monetary policy instruments directly and indirectly Granger cause output growth at long horizons. These results are consistent with Dufour et al. (2006), who find that short-term interest rates predict GDP growth rates for a large number of distant forecast periods while nonborrowed reserves indirectly cause future real activity growth through T-Bill yields.

Next, in Table 2a we display the test results for the economy of Japan. Our findings reveal co-behavior patterns between stock return volatility and output growth at 14, 18, 19, 20 and 21 months ahead at level of significance 10%. Inflation is found to predict industrial production growth rates at forecast periods 1, 20, and 21 months ahead. Short-term interest rates cause output growth for the horizon 22, while money supply growth rates for the horizons 1 and 3 at level 10%. When compared with the US test results, we find fewer direct causality results.

Table 2b reports the causality test results when evaluating for indirect influences on Japanese industrial production growth rates. First, the volatility of stock returns does not Granger cause short-term interest rates over any horizon. However, stock return volatility is causal for money supply growth rates for both short (over the range of one month ahead to five months ahead) and long horizons (over the range of 22 months ahead to 35 months ahead) at level 10%. Long horizon causality is found from stock return volatility to inflation for the forecast periods h = 10, 19, 20, 21, 22, 24, 30, 31, 32, 33, 34, 5, and 36 months at level 10%. We find that the null hypothesis of non-causality from short-term interest rates to money supply growth rates is rejected only for 26 and 27 months ahead at level 10%. Our results also indicate that interest rates predict inflation for the forecast periods 9, 10, 27, and 28 months ahead. Money supply growth rates do not have predictive ability for short-term interest rates. However, changes in the growth rates of money supply anticipate changes in inflation over the range from 20 months ahead to 35 months ahead at level 10%. The following indirect causality relations are revealed. First, money supply growth rates cause inflation at short horizons, which in turn cause output growth at long horizons. Thus, there is indirect long horizon causality from money supply growth rates to output growth through inflation. Second, changes in the volatility of stock returns anticipate changes in inflation at short horizons (i.e., 10 months ahead), while inflation causes output growth at forecast period 22. Hence, changes of the stock return volatility yields an indirect influence on real activity growth at distant periods ahead through inflation. Therefore, we find empirical evidence that fluctuations in stock market signal inflationary pressures which affect the real economy.

Table 3a presents the test results for Germany. Again, the proxy for stock return volatility Granger causes production growth rates for a large number of short and long forecast horizons. Significant causality results at level 10% are found for the forecasting periods 3, 7, 8, 9, 10, 21, 22,23, 28, 30, 31, 32, and 33. The forecasting power of the monetary policy appears to be remarkable. Interest rates predict production growth rates over the range from 9 months ahead to 36 months ahead, while money supply growth rates predict production growth rates for almost all the forecasting horizons considered at level 10%. Our results also indicate that there is short horizon causality from inflation to the specific measure of output growth.

Observing the test results for Germany demonstrated in Table 3b, we find that the stock return volatility proxy does not Granger cause money supply growth rates for any horizon, still it does predict short-term interest rates for 13,14, 15, 16, 17, and 18 months ahead, respectively. Money supply growth rates are found to predict interest rates for short horizons (from 1 to 8 months ahead) and long horizons (from 15 to approximately 36 months ahead). Our findings indicate that no other significant causality results exist. Hence, we find indirect causality from stock market volatility to industrial production growth rates through short-term interest rates.

We present the causality test results for Italy in Table 4a. A smaller number of causality relations are revealed when compared to those of the other three economies. Changes in the stock return volatility anticipate changes in real activity growth only for 19 and 20 months ahead. We find that there is only short horizon causality from short –term interest rates to output growth (i.e., at forecast periods 1, 3 and 4 months ahead). As far as it concerns the other instrument of monetary policy, money supply growth rates do not predict real activity growth for any forecast horizon. On the other hand, inflation is found to cause industrial production growth rates for 14, 15, 16 and 20 months ahead.

In Table 4b we investigate for causality chains between the variables in Italy. Volatility of stock returns causes inflation from 1 month ahead to 33 months ahead. The volatility proxy also predicts money supply growth rates over the ranges 17-23, 25-29, and 33-35 months ahead respectively. Although short-term interest rates do not cause stock return volatility at any horizon, they do predict inflation at short horizons (i.e., for h = 2 and 8 months ahead). Our results also indicate that money supply growth has significant predictive ability for inflation for both short (1 to 9 months ahead) and long forecast periods (over the range 22 months ahead and 36 months ahead). Volatility of stock returns causes inflation for short horizons (and long horizons as well), which in turn causes output growth. Thus, we have indirect causality from stock returns volatility to output growth. By using a similar reasoning, we find that there is indirect long horizon causality from money supply to output growth measure through inflation.

Panels 1 to 4 present the causality measures over the range from 1 month-ahead to 36 months-ahead for US, Germany, Japan, and Italy, respectively. Each plot displays the OLS causality measure and the bootstrap percentile bounds at significance level 5%. 10000 replications are used for the calculation of the bootstrap confidence intervals.

In the US (Panel 1), stock return volatility Granger causes output growth 3 months-ahead. The impact of the former to the latter appears to be strong in the economic sense. Causality from money supply growth rates to output growth is statistical different from zero after 7 months. The predictive power of the monetary policy instrument declines gradually after 13 months. On the other hand, nominal short-term interest rates do not Granger cause real activity over any horizon. Inflation has the largest impact on output growth (approximately 0.03 for the first 6 months); from h = 10 inflation is not causal for real activity measure. Causality from short-term interest rates to stock return volatility or money supply growth rates is not statistical significant different from zero over any horizon. Then again, inflation, money supply growth, and the volatility proxy are causal for interest rates. The first two variables have significant predictive power for the first 8 to 10 months. Moreover, volatility of stock returns appear to cause short-term interest rates for all 36 horizons. Stock return volatility also affects M2 growth rates 1 month ahead and after 15 months; this influence is economically strong at both short and long horizons. Hence, there is indirect causality from the volatility of real stock returns to real activity growth. Money supply also is a significant in-sample predictor of stock return volatility over the range from 1 month-ahead to approximately 36 months-ahead. Note that money supply has a stronger influence on the volatility proxy than the latter to the former. Inflation is also found to strongly cause stock return volatility at both short and long horizons; the OLS causality measure increases at longer forecast periods.

Panel 2 reports the causality measure results of Germany. Stock return volatility is found to cause industrial production growth rates strongly over the range from 1 month-ahead to 14 months-ahead. Monetary policy instruments, money supply growth and interest rates present a limited in-sample predictive ability for future real activity growth; the first predict up to 4 months ahead while the second 1 month ahead. Moreover, inflation has a small impact on output growth (up to 4 months ahead).

The results for Japan are presented in Panel 3. The proxy for stock return volatility appears to be causal for monthly growth rates of industrial production at short and long horizons. In particular, the causality measures are significantly different from zero after 16 months and sizeable from an economic viewpoint (the values range from 0.01 to 0.06). Another interesting result is that short-term interest rates do not Granger cause industrial production growth rates over any horizon. The measures are approximately zero and statistical insignificant for all forecast horizons. Then again, money supply growth rates have a large (from an economic viewpoint) significant impact on real activity growth up to 4 months ahead. The causality measures decline gradually until they take zero values. Inflation strongly causes real activity growth

(0.1 the first five months). Our results also show that there is strong causality from money supply growth to inflation and the stock return volatility proxy for a wide range of long horizons, but not for short-term interest rates. In particular, we find that M2 has a large impact on stock return volatility data for all horizons; the causality measures are high as 0.2. The volatility proxy is also strongly causal after 20 months for M2 growth rates. Short-term interest rates fail to cause money supply growth rates and inflation rate. Our results also show that inflation is a significant predictor of the volatility proxy after 17 months.

Panel 4 displays the causality measures between the financial, macroeconomic and monetary variables of Italy. We observe that stock return volatility data is the only significant predictor of future output growth. The causality measures are significantly different than zero after 13 months. Causality from inflation, money supply growth, and short-term interest rates to output growth are zero and statistical insignificant. Furthermore, we find that stock return volatility Granger causes money supply growth, short-term interest rates and inflation at long horizons. Note that volatility has a large impact to inflation (the measure equals approximately to 0.04) after 11 months. Monetary policy does not affect stock return volatility; M2 does not have a significant impact on the volatility of stock returns, while the causality measures from short-term interest rates to the volatility proxy are marginally statistically different from zero after 16 months. On the other hand, inflation has high predictive power for stock return volatility over the range from 1 month-ahead to 36 months-ahead.

Consensus views by academics and market practitioners state that central banks should formulate monetary policy by taking into account actual or predicted inflation and output gap but not stock market volatility. They argue that systematic responses to stock return volatility may induce destabilizing effects on output growth. Bernanke and Gertler (1999) show that including stock market volatility in a policy formulation process may improve macroeconomic performance only when changes in the variance of stock returns signal possible inflationary or deflationary pressures. However, it is almost impossible to disentangle the sources of stock return variability. On the other hand, Cecchetti et al. (2000) emphasize that central banks should follow preventive policies against developments in stock return volatility in order to bound the increase of imbalances in real economy. Their analysis show that there are sizeable gains when central banks determine optimal interest rate by minimizing a weighted average of the output gap and inflation fluctuations while incorporating information about stock return variation. Similarly, Filardo (2004) argues that monetary policymakers should systematically react to stock return volatility, and more importantly to asset price bubbles. Our finding that variability in stock returns produces strong effects on both future inflation and production growth rates in all four economies implies that central banks should incorporate systematically information about stock prices in their policy making. Stock return volatility tends to have a large impact in economies with different degrees of market capitalization and proportions of share of household wealth represented by the stocks. Moreover, it appears that these distortions operate on spending and aggregate demand on the short-run, but they also affect aggregate supply through capital formation on the long-run. These results raise some concern about whether there are actions monetary authorities could take to diminish the likelihood of the undesirable effects of stock return volatility.

5. Out-of-sample forecasting evaluation

Our goal in this section is to show that an out-of-sample forecasting exercise using moderate sample sizes yields the same answers with our in-sample empirical investigations. We also construct single indicator forecasting models for output growth based on the causal lag structure information revealed by our in-sample analysis. Our findings indicate that our models generate more accurate and stable predictions of output growth than the autoregressive benchmark model.

5.1. *Forecasting industrial production growth rates h-steps ahead*

The iterated multistep approach consists of first estimating a dynamic model for the monthly growth rates, and then using the chain rule to compute *h*-step-ahead forecasts of the series. In particular, our forecasting multivariate regression model is

$$y_t = \alpha + \sum_{k=1}^{P} \beta_k y_{t-k} + \sum_{i=1}^{4} \sum_{k=1}^{p} \gamma_{ik} x_{it-k} + \varepsilon_t \qquad (18)$$

where y_t represent the growth rates of industrial production (IP_t), and x_{it} denote the indicators $(V_t, IN_t, SI_t, M2_t)$. The parameters $\alpha, \beta_k, \gamma_{ik}, k = 1,..., p, i = 1,...,4$ are estimated by OLS. By iterating forward one-period ahead h times we are able to compute the forecasts recursively,

$$\tilde{y}_{t+h} = \hat{\alpha} + \sum_{k=1}^{P} \hat{\beta}_k \tilde{y}_{t+h-k} + \sum_{i=1}^{4} \sum_{k=1}^{p} \hat{\gamma}_{ik} \tilde{x}_{it+h-k} \quad (19)$$

based only on values of the series up to the date on which the forecast is made. Long horizon iterated forecasts of the industrial production growth rates are computed as

$$y_{t+h} = y_t^0 + \sum_{j=1}^{h} \tilde{y}_{t+j} \quad (20)$$

where y_t^0 is the log of the industrial production level at time *t*.

The forecast performance of these regressions is compared to that of an autoregressive bencmark model,

$$y_t = \alpha + \sum_{k=1}^{P} \beta_k y_{t-k} + \varepsilon_t \ . \quad (21)$$

To compute the forecasts, the models are estimated, lag lengths are selected -based on the Akaike information Criterion (AIC) with maximum lag order set to five lags– using observations from date 1 through date *q*, where *q* is length of the estimation window. Moving forward by one month, the models are reestimated (and information criteria computed) using data from date 2 through date *q* + 1. This sequence of actions is repeated *T* –*q* - *h* times across the sample. Hence, sequences of *h*-step-ahead forecasts of the growth in industrial production are formed, and as a consequence the corresponding forecast errors, allowing us to compute the root mean squared forecast error (RMSFE), the Theil inequality criterion (Theil), the bias (bias) and variance (Var) components of the Theil inequality decomposition. We choose to set *q* = 0.75, 0.80, 0.60, and 0.65 for US, Germany, Japan and Italy respectively, so that we form series of forecast errors with minimum length 150 to 170 observations.

We should note that our approach for constructing the stock return volatility proxy (V_t) is to recursively estimate an AR(*p*) model (lag length is selected automatically with AIC between lags one and 12) for the real stock returns and to retain the squared innovations as we move the estimation window through the sample.

5.1a. Pseudo-out-of-sample forecasting results

Table 5 demonstrates the forecasting performance of our model in an out-of-sample forecast exercise for US growth rates of industrial production. The RMSFEs indicate that the AR model produces superior forecasts of output growth when compared to our model. On the other hand, Theil coefficient results present a different picture. For h = 5 to 23 Theils show that our model outperforms the benchmark model. For forecast horizons larger than 23, our model has a poor performance when compared to the AR model. Note that the causality measure from M2 to output growth was found to be large and statistical significant over the range from 5 months ahead and 20 months ahead approximately (after 20 months causal influence was found to be significant but of small magnitude) .The bias appears to increase in a higher rate in our model than the AR model when h increases. The variance component of the Theil decomposition is found to decrease in a higher rate in our model than the AR model when h increases.

Table 6 presents the forecasting performance of our model in Germany for horizons one to 36 months ahead. Our forecasting model is found to be beneficial for horizons 8 to 19 months ahead when using RMSFE as a measure of forecasting accuracy. As in US, for forecast periods larger than 19 months the RMSFEs of our model are larger than those of the AR model. These results are in accord with our in-sample findings where statistical significant comovements of large magnitude between stock return volatility and output growth were reported for forecast periods 1 month ahead to 14 months ahead. Theil results indicate that our model has better predictive power than the benchmark model over the range from h = 5 months ahead to h = 28 months ahead. The bias increases as h increases; on the other hand, the variance decreases as h grows.

Table 7 compares the forecasting accuracy of our model to the AR benchmark when implemented to predict japanese monthly industrial production growth rates. There are sizeable gains when using our model to forecast production growth rates at both short and long monthly periods ahead. RMSFEs indicate that our specification is more accurate than the AR benchmark model for almost all forecast horizons. More importantly, greater accuracy is achieved for long horizons. For instance, for h = 5,6, 35 and 36 our model yields RMSFEs 0.086, 0.096, 0.081 and 0.078 respectively, while the AR model yields RMSFEs 0,097, 0,111, 0,461, and 0,472 respectively. Our model is favorably compared to the AR benchmark at long horizons when using Theil coefficient to measure forecasting accuracy. For horizons larger than 21 the proposed specification achieves a superior performance over the AR, especially at very long horizons. Our model appears to produce less biased forecasts of output growth than the benchmark specification for a wide range of forecast periods. On the other hand, forecast bias of the former is more volatile.

The out-of-sample predictive evaluation of both specifications for Italy are demonstrated in Table 8. RMSFEs over the range from 8 months-ahead to 29 months-ahead show that our model outperforms the AR model. We find gains to the use of our model at long horizons; for instance our model and the AR yield RMSFEs 0.078 and 0.094 at h = 20 respectively. Theil coefficients confirm that our model produces superior long horizon forecasts of output growth. In particular, our specification yields more accurate forecasts than the AR model over the range from 11 months-ahead to 33 months-ahead. Forecasts produced by the proposed model appear slightly more biased than those of the AR. Both specifications produce slightly biased predictions of production growth rates of similar size.

5.2. *Forecasting output growth one-step ahead based on long horizon causality relations*

It is of interest to consider short-term forecasts of industrial production growth rates by incorporating in the forecasting regressions the long horizon causal structures between the economic time series found in our in-sample analysis. Thus, our interest is in the out-of-sample predictive power of single-indicator models that employ h lags of x_{it} to forecast real activity growth y_t one month-ahead, i.e.,

$$y_t = \gamma_0 + \sum_h \gamma_{ih} x_{it-h} + \varepsilon_t \tag{23}$$

where h is specified for each indicator x_{it} according to the specific lag structure our in-sample analysis indicates that is appropriate. Two indicators are used, stock return volatility and money supply growth rates for the forecast regression of each economy (in Italy the short-term interest rates are used instead of M2 growth rates). The lag structures implied by the causality relations are

US $\quad V_t : h := \{30,..,36\},\ M2_t : h := \{30,..,36\},$ (24)

Germany $\quad V_t : h := \{28,30,..,33\},\ M2_t : h := \{25,..,32,36\},$ (25)

Japan $\quad V_t : h := \{21,..,27\},\ M2_t : h := \{1,2,3\},$ (26)

Italy $V_t : h := \{16,..,23\}$, $SI_t : h := \{1,3,4\}$. (27)

The forecasting performance of these regressions is compared to that of an autoregressive benchmark model. We select the lag order of the benchmark model at each estimation window using the Akaike Information criterion (AIC), with maximum lag order set to five lags. We consider two estimation windows, equal to 50%*T* and 70%*T*.

Clark and McCracken (2005) show that lack of parameter constancy of the regression model used for forecasting economic series, has a substantial impact on its out-of-sample performance. Recent econometric advances in forecasting economic series present methods, whose implementation does not require testing for heterogeneity of a stochastic process or parameter instability of a regression model. Pesaran and Timmermann (2007) show that averaging forecasts obtained from in-sample model estimation windows of different lengths ensures a satisfactory finite sample performance, especially in the presence of neglected structural breaks of small size. There is evidence that single predictor models with varying in-sample windows outperform the pooling of forecasts from a large set of single-indicator models; see Assenmacher-Wesche and Pesaran (2008), Pesaran, Schuermann and Smith (2009), and Schrimpf and Wang (2009). As well as the standard rolling window approach, we implement forecast combination across observation windows of different lengths, as proposed by Pesaran and Timmerman (2007) and Pesaran and Pick (2009). The starting point of the in-sample window is changed removing one observation, while the forecasts are calculated based on the parameters of the predicting model estimated on these observation windows. Then these forecasts which correspond to the different starting point data windows are averaged. The forecast averages of the rolling window are used for the computation of the root mean square forecast error (hereafter denoted as RMSFE), Theil inequality criterion, bias and variance decomposition based on the Theil inequality.

For each observation window, we use the sequential procedure of Bai and Perron (1998,2003) to estimate multiple structural breaks on output growth dynamics, and then equal the starting points of each sub-window to these break points. Four breakpoints are considered within each rolling window (either 50%T or 70%T). The size and location of the multiple change-points can be consistently estimated by this procedure.

Since we compare the forecast performance of a static model incorporating a long lag structure (single-predictor models) to that of a dynamic model (autoregressive benchmark model), we consider a bootstrap technique to offset the small sample bias in the coefficient estimates of the former, and as consequence, in the forecast accuracy of these regressions. Hence, random sampling with replacement is used on the original growth/return observations of each window so that we generate a large sample. Once a bootstrap sample is computed (set approximately equal to 8 times the original estimation window), we use it to estimate the parameters of each single-predictor and the benchmark model, and therefore, to calculate the forecasts. Forecast averaging over estimation windows of different lengths is also implemented on the bootstrap samples. Our approach again estimates multiple breaks, using the procedure proposed by Bai and Perron, and then equalizes the starting point of each sub-window to the location the break-points.

Table 9 reports the forecast results of the single-indicator and the benchmark models. In particular, the results for three forecasting models are presented. The models are: an AR(p) model, denoted as benchmark model, with maximum $p = 5$ and lag selection based on AIC criterion; a volatility model using all lagged volatility proxy terms as shown in equation (23); a M2 model using all lagged money supply growth rate terms as shown in equation (23) (for Italy short-term interest rates are prefered over M2 growth rates). The lag structure h used in the specification of last two models are presented in (24):(27). Three modeling schemes are displayed. The schemes are: a simple modeling scheme, where the three forecasting models as described previously are implemented on rolling windows of length 50%T and 70%T; a modeling scheme where the forecasts of each model are averaged over different estimation sub-windows; a bootstrap scheme, where the forecasts of each forecasting model are averaged over different bootstrap data sub-windows. Table 9 a, b, c and d records the results for US, Germany, Japan and Italy, respectively.

Consider first the exercise using the simple modeling scheme. The AR benchmark specification outperforms the other two competing models in terms of RMSFEs and Theil coefficients. For instance, in US when a rolling window of 70%T observations is employed, Theils of the volatility and M2 indicator models are 0.7637 and 0.7416 respectively, while the AR model yields 0.6520. We find no gains to the use of single-indicator models based on stock return volatility and money supply growth rates for the other three economies. When 70% of the full sample is used, volatility, M2 and AR models yield Theils 0.9407, 0.9298, and 0.7627 respectively in Germany; 0.9363, 0.9646, and 0.7142 respectively in Japan; 0.9511, 0.9468, and 0.7146 respectively in Italy. This result is robust to the choice of the estimation window for all four economies. However, when forecasting stability is considered by using forecast combinations across different data windows, both static models generate more accurate predictions than the AR dynamic model. US volatility proxy based model yields Theil coefficients 0.6711 and 0.7528 when rolling windows of size 50%T and 70%T are considered, respectively. On the other hand, the AR model yields 0.9864 and 0.9942 respectively. We obtain similar results for the other three economies.

For the forecasting exercise based on the bootstrap technique, the results are qualitatively similar and further strengthen the evidence that our approach produces superior short-term forecasts of industrial production growth. Both single indicator models yield more accurate predictions than the benchmark model; for instance, Theil coefficients of the US volatility and M2 specifications are 0.7382 and 0.8046, 0.6962 and 0.7606, for estimation windows 50%T and 70%T respectively. Then again, Theil coefficients of the AR model are 0.9154 and 0.9993 respectively. Note that the bias of both indicator based models is reduced when the bootstrap technique is used in all four economies except Italy.

6. Conclusions

Our interest is in whether stock return volatility anticipates changes of output growth at distant time periods. In particular, we investigate for the presence of long horizon causation from stock return volatility to the growth rates of industrial production, in terms of indirect influences with monetary policy instruments and inflation in four economies, namely US, Germany, Japan and Italy. Multiple horizon non-causality testing is performed by implementing the test procedure proposed by Dufour et al. (2006) on the former relation, with short-term interest rates, money supply growth rates (M2), and inflation as auxiliary variables.

Our results reveal a large number of significant direct causalities from the volatility of stock returns to output growth at both short and long horizons in all four economies. Monetary policy instruments, such as short-term interest rates and M2 growth rates, also found causality for output growth in the majority of the economies. Interestingly, our results indicate that stock return volatility indirectly predicts monthly growth rates of industrial production at long horizons, in particular through the nominal short-term interest rates (US, and Germany), the money supply growth rates (Japan) and the inflation (Italy).

We are also interested in measuring the degree of forecast improvement that arise from each horizon specific causality relation. In order to do so, we implement the causality measure proposed by Dufour and Taamouti (2010) on the data for different time horizons. A bootstrap technique is used to calculate confidence intervals for the causality measure. Multiple horizon causalities of significant size from stock return volatility to output growth are found in Germany, Japan and Italy. In US the causality measures indicate significant indirect long horizon causation through the money supply growth rates.

Finally, a forecasting exercise reveals that the combination of stock return volatility, short-term interest rates, money supply growth, and inflation in a single regression model generates more accurate forecasts of output growth than the autoregressive model in the long term. Our forecasting results validate the causality linkages between stock market volatility, monetary policy and output growth obtained in our empirical investigations. We also show that single indicator specifications based on the lag structure revealed by our in-sample analysis fares well at predicting short term industrial production growth rates relative to the benchmark model when pooling the forecasts across estimation windows of different lengths or using a large simulation scheme to offset small sample estimation bias.

Our finding that stock return volatility has an impact of significant size on both future expected inflation and production growth rates in all four economies implies that central banks should incorporate systematically information about stock return variation in their monetary policy rules. Stock return volatility causes strongly production growth rates of economies with different degrees of market capitalization and alternate proportions of share of household wealth represented by the stocks. Moreover, it appears that these distortions operate on both the short and the long-run. These results raise some concern about whether there are actions monetary authorities could take to enhance long horizon macroeconomic performance and prevent financial distresses. Cecchetti, et al. (2000) show that it is preferable to develop monetary policy rules that minimize future deviations of inflation and output gap based on stock return volatility rather than determining target interest rates using exclusively expectations of future inflation at some fixed time horizon. They also demonstrate that monetary policy making which focus on preventing the formation of asset price bubbles instead of puncturing them may ensure an overall macroeconomic stability.

Appendix of Chapter 4

Table1a. Causality test results for forecast horizons 1 to 36 months ahead in US

h	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
V→IP	14,378	11,836	8,029	10,129	11,227	10,290	11,355	**20,386**	17,311	12,604	13,012	12,436
	(0,183)	(0,364)	(0,619)	(0,483)	(0,429)	(0,489)	(0,438)	**(0,078)**	(0,179)	(0,360)	(0,386)	(0,476)
SI→IP	13,595	7,529	5,499	5,027	5,836	6,160	7,049	7,139	4,472	6,596	8,285	7,436
	(0,248)	(0,685)	(0,853)	(0,899)	(0,815)	(0,808)	(0,765)	(0,750)	(0,912)	(0,804)	(0,707)	(0,804)
IN→IP	12,989	6,779	10,230	7,810	9,453	9,842	8,749	11,785	11,122	12,500	11,439	9,573
	(0,284)	(0,750)	(0,476)	(0,698)	(0,519)	(0,552)	(0,603)	(0,453)	(0,500)	(0,414)	(0,517)	(0,625)
M2→IP	8,125	7,896	6,918	6,211	6,374	7,324	5,750	4,930	5,565	7,372	7,564	9,920
	(0,634)	(0,612)	(0,776)	(0,819)	(0,806)	(0,737)	(0,856)	(0,920)	(0,873)	(0,780)	(0,754)	(0,603)

h	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
V→IP	**23,469**	12,884	21,834	**34,254**	16,660	**28,610**	**31,935**	**34,946**	**39,792**	17,698	18,036	18,854
	(0,071)	(0,444)	(0,155)	**(0,019)**	(0,356)	**(0,045)**	**(0,050)**	**(0,032)**	**(0,026)**	(0,317)	(0,308)	(0,325)
SI→IP	12,081	16,834	18,512	18,526	17,831	15,375	13,129	10,344	14,077	14,467	13,052	17,833
	(0,494)	(0,222)	(0,209)	(0,203)	(0,252)	(0,362)	(0,474)	(0,688)	(0,474)	(0,446)	(0,524)	(0,343)
IN→IP	18,909	12,147	13,722	22,602	17,975	12,777	11,176	12,521	13,215	10,524	12,418	12,443
	(0,177)	(0,494)	(0,399)	(0,112)	(0,254)	(0,537)	(0,586)	(0,524)	(0,489)	(0,681)	(0,612)	(0,612)
M2→IP	15,394	14,521	15,471	21,150	14,836	11,834	8,231	**29,281**	25,763	**37,618**	**37,773**	24,278
	(0,267)	(0,371)	(0,330)	(0,168)	(0,366)	(0,619)	(0,813)	**(0,075)**	(0,101)	**(0,034)**	**(0,034)**	(0,168)

h	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
V→IP	16,973	14,777	21,169	22,344	**33,216**	**36,260**	**36,562**	**37,865**	**50,000**	**76,344**	**113,989**	**94,502**
	(0,379)	(0,498)	(0,295)	(0,250)	**(0,075)**	**(0,084)**	**(0,078)**	**(0,075)**	**(0,028)**	**(0,002)**	**(0,006)**	**(0,009)**
SI→IP	22,659	25,873	27,310	23,358	20,300	**41,822**	25,293	31,515	31,868	23,054	21,374	32,478
	(0,194)	(0,185)	(0,127)	(0,213)	(0,338)	**(0,052)**	(0,203)	(0,153)	(0,131)	(0,341)	(0,360)	(0,175)
IN→IP	21,455	19,711	22,443	19,643	20,595	25,441	17,551	13,753	13,206	9,702	16,651	17,570
	(0,291)	(0,278)	(0,244)	(0,343)	(0,343)	(0,209)	(0,483)	(0,593)	(0,629)	(0,845)	(0,550)	(0,532)
M2→IP	21,086	**37,067**	24,245	25,560	22,277	**41,320**	**33,702**	**53,978**	**57,360**	**75,356**	**66,952**	**78,350**
	(0,231)	**(0,058)**	(0,205)	(0,164)	(0,237)	**(0,032)**	**(0,075)**	**(0,024)**	**(0,017)**	**(0,002)**	**(0,011)**	**(0,006)**

Note: this table reports the Dufour et al. (2006) test results for forecast horizons (*h*) 1 month ahead to 36 months ahead at nominal level of significance 10% . Simulated *p*-values are in parentheses. Rejection of the null hypothesis of non-causality is in bold. V, SI, IN, M2 and IP stand for volatility of stock returns, short-term nominal interest rates, inflation, M2 growth rate, and industrial production growth rate, respectively.

Table1b. Causality test results for forecast horizons 1 to 36 months ahead in US

h	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
M2→SI	15,139	13,414	15,558	16,43	**18,382**	**25,317**	18,808	18,372	**20,602**	16,184	15,337	10,028
	(0,163)	(0,242)	(0,172)	(0,131)	**(0,098)**	**(0,026)**	(0,114)	(0,123)	**(0,098)**	(0,240)	(0,275)	(0,617)
V→M2	14,300	5,572	3,105	3,041	3,155	6,061	4,336	7,554	7,561	8,142	10,702	9,599
	(0,202)	(0,839)	(0,965)	(0,968)	(0,965)	(0,821)	(0,919)	(0,711)	(0,722)	(0,687)	(0,509)	(0,632)
V→SI	**17,768**	**24,288**	13,077	9,013	5,581	6,615	11,447	10,898	10,757	9,450	7,406	12,784
	(0,089)	**(0,017)**	(0,259)	(0,557)	(0,846)	(0,762)	(0,435)	(0,480)	(0,477)	(0,624)	(0,755)	(0,390)
SI→M2	**49,872**	**23,243**	13,800	12,163	14,572	12,892	9,286	6,823	11,249	15,351	14,957	14,126
	(0,001)	**(0,022)**	(0,227)	(0,290)	(0,197)	(0,267)	(0,558)	(0,800)	(0,452)	(0,251)	(0,264)	(0,323)

h	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
M2→SI	15,170	15,607	14,986	21,073	15,792	15,688	24,766	20,617	13,002	12,075	15,824	15,154
	(0,316)	(0,336)	(0,369)	(0,149)	(0,351)	(0,382)	(0,117)	(0,229)	(0,528)	(0,595)	(0,432)	(0,467)
V→M2	7,971	9,191	7,019	6,030	10,471	14,209	10,339	12,002	10,648	14,973	19,276	24,421
	(0,734)	(0,671)	(0,824)	(0,867)	(0,625)	(0,412)	(0,665)	(0,536)	(0,664)	(0,449)	(0,274)	(0,142)
V→SI	11,293	11,205	8,003	16,889	6,321	7,341	8,256	6,319	6,760	9,468	7,503	5,136
	(0,528)	(0,541)	(0,768)	(0,258)	(0,883)	(0,821)	(0,779)	(0,899)	(0,863)	(0,761)	(0,849)	(0,950)
SI→M2	17,851	19,824	13,162	13,884	3,010	15,148	10,300	**30,899**	25,586	22,219	**32,997**	**37,996**
	(0,194)	(0,150)	(0,394)	(0,407)	(0,990)	(0,371)	(0,616)	**(0,038)**	(0,106)	(0,175)	**(0,057)**	**(0,025)**

h	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
M2→SI	M2→SI	13,727	10,328	12,979	20,057	14,011	9,191	12,371	15,676	17,657	18,690	18,488
	(0,542)	(0,549)	(0,757)	(0,607)	(0,327)	(0,556)	(0,849)	(0,691)	(0,569)	(0,497)	(0,479)	(0,494)
V→M2	V→M2	18,203	26,849	24,047	23,868	24,754	9,855	14,007	19,371	16,013	23,188	31,020
	(0,325)	(0,330)	(0,136)	(0,201)	(0,214)	(0,185)	(0,787)	(0,594)	(0,426)	(0,542)	(0,296)	(0,195)
V→SI	4,884	4,765	5,452	7,561	7,402	7,873	10,129	8,998	10,395	10,120	9,738	12,432
	(0,961)	(0,971)	(0,936)	(0,884)	(0,882)	(0,858)	(0,760)	(0,849)	(0,768)	(0,809)	(0,820)	(0,745)
SI→M2	**37,868**	**42,993**	**33,709**	**52,015**	**48,773**	**60,332**	29,878	**38,054**	**36,493**	**46,487**	28,011	15,644
	(0,031)	**(0,025)**	**(0,060)**	**(0,015)**	**(0,020)**	**(0,009)**	(0,125)	**(0,068)**	**(0,089)**	**(0,029)**	(0,190)	(0,565)

Note: this table reports the Dufour et al. (2006) test results for forecast horizons (*h*) 1 month ahead to 36 months ahead at nominal level of significance 10% . Simulated *p*-values are in parentheses. Rejection of the null hypothesis of non-causality is in bold. V, SI, IN, M2 and IP stand for volatility of stock returns, short-term nominal interest rates, inflation, M2 growth rate, and industrial production growth rate, respectively.

Table 2a. Causality test results for the forecast horizons 1 to 36 months ahead in Japan

h	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
V→IP	19,294	14,118	12,250	12,099	10,718	14,496	12,907	9,893	10,807	10,763	22,768	31,411
	(0,318)	(0,613)	(0,741)	(0,744)	(0,874)	(0,677)	(0,803)	(0,920)	(0,887)	(0,903)	(0,475)	(0,318)
SI→IP	24,323	22,229	24,474	20,788	19,447	13,117	14,044	14,236	17,016	18,453	16,802	12,182
	(0,150)	(0,233)	(0,188)	(0,339)	(0,420)	(0,758)	(0,743)	(0,754)	(0,650)	(0,610)	(0,692)	(0,864)
IN→IP	**30,478**	18,723	14,412	9,639	9,817	8,895	10,257	10,437	17,543	14,622	11,255	10,012
	(0,053)	(0,396)	(0,633)	(0,857)	(0,869)	(0,933)	(0,880)	(0,891)	(0,609)	(0,755)	(0,900)	(0,933)
M2→IP	**28,942**	24,060	**28,649**	19,044	18,457	17,077	18,488	14,178	17,555	12,083	16,762	19,232
	(0,071)	(0,162)	**(0,095)**	(0,392)	(0,480)	(0,560)	(0,528)	(0,752)	(0,622)	(0,850)	(0,710)	(0,590)

h	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
V→IP	32,851	**51,104**	49,234	50,865	47,323	**70,649**	**79,274**	**75,469**	**98,198**	48,411	44,623	52,153
	(0,283)	**(0,065)**	(0,110)	(0,109)	(0,165)	**(0,056)**	**(0,042)**	**(0,064)**	**(0,026)**	(0,249)	(0,353)	(0,284)
SI→IP	7,317	31,587	26,967	40,639	28,720	40,258	40,062	35,618	39,276	**89,971**	59,628	63,408
	(0,986)	(0,278)	(0,459)	(0,211)	(0,440)	(0,260)	(0,304)	(0,373)	(0,356)	**(0,028)**	(0,159)	(0,133)
IN→IP	15,137	14,022	30,198	29,973	38,764	44,099	43,540	**58,669**	**67,370**	28,592	21,201	14,374
	(0,806)	(0,828)	(0,374)	(0,388)	(0,228)	(0,202)	(0,235)	**(0,091)**	**(0,083)**	(0,618)	(0,788)	(0,932)
M2→IP	24,037	10,754	12,993	14,765	11,379	18,877	15,524	18,235	23,360	29,054	17,995	49,739
	(0,464)	(0,936)	(0,870)	(0,840)	(0,928)	(0,766)	(0,861)	(0,823)	(0,698)	(0,624)	(0,871)	(0,321)

h	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
V→IP	34,544	20,268	19,968	19,642	32,443	27,288	57,894	50,067	88,879	63,361	53,869	51,341
	(0,567)	(0,877)	(0,908)	(0,901)	(0,694)	(0,822)	(0,384)	(0,529)	(0,203)	(0,396)	(0,563)	(0,594)
SI→IP	43,557	39,267	33,569	34,172	37,221	34,452	29,385	27,157	60,185	40,757	33,182	32,565
	(0,393)	(0,512)	(0,623)	(0,673)	(0,614)	(0,676)	(0,803)	(0,818)	(0,410)	(0,667)	(0,812)	(0,832)
IN→IP	9,552	21,675	23,155	23,193	40,625	24,963	21,926	20,068	32,291	82,553	80,068	70,948
	(0,990)	(0,849)	(0,828)	(0,836)	(0,563)	(0,846)	(0,887)	(0,927)	(0,777)	(0,213)	(0,277)	(0,383)
M2→IP	60,358	46,375	17,040	27,248	23,617	18,915	25,406	20,545	24,097	29,760	46,603	55,223
	(0,204)	(0,373)	(0,923)	(0,782)	(0,835)	(0,926)	(0,849)	(0,923)	(0,880)	(0,802)	(0,612)	(0,537)

Note: this table reports the Dufour et al. (2006) test results for forecast horizons (*h*) 1 month ahead to 36 months ahead at nominal level of significance 10% . Simulated *p*-values are in parentheses. Rejection of the null hypothesis of non-causality is in bold. V, SI, IN, M2 and IP stand for volatility of stock returns, short-term nominal interest rates, inflation, M2 growth rate, and industrial production growth rate, respectively.

Table 2b. Causality test results for the forecast horizons 1 to 36 months ahead in Japan

h	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
M2→SI	7,004	5,561	5,397	5,166	10,437	10,105	12,209	14,870	13,687	11,081	9,780	9,110
	(0,949)	(0,968)	(0,978)	(0,983)	(0,835)	(0,858)	(0,786)	(0,681)	(0,758)	(0,856)	(0,923)	(0,947)
V→M2	**36,376**	**48,175**	**61,968**	**50,522**	**31,358**	16,580	15,795	22,885	25,333	21,455	25,756	32,656
	(0,016)	**(0,009)**	**(0,001)**	**(0,006)**	**(0,079)**	(0,531)	(0,605)	(0,341)	(0,293)	(0,439)	(0,319)	(0,186)
V→SI	12,661	9,733	7,281	6,359	4,936	6,069	5,520	5,934	6,257	5,306	3,956	5,929
	(0,679)	(0,821)	(0,923)	(0,959)	(0,992)	(0,978)	(0,991)	(0,977)	(0,978)	(0,993)	(0,998)	(0,990)
SI→M2	17,070	13,574	15,592	11,217	20,089	31,186	29,673	24,828	18,667	15,305	23,274	40,916
	(0,400)	(0,664)	(0,557)	(0,787)	(0,365)	(0,116)	(0,138)	(0,269)	(0,498)	(0,689)	(0,384)	(0,088)
V→IN	8,485	8,183	12,338	21,296	21,855	23,699	20,668	18,961	35,257	**38,126**	27,696	31,717
	(0,888)	(0,915)	(0,717)	(0,318)	(0,312)	(0,294)	(0,451)	(0,522)	(0,118)	**(0,091)**	(0,324)	(0,250)
SI→IN	19,917	21,196	25,474	20,791	20,205	23,278	23,939	18,997	**53,049**	**55,640**	28,701	37,095
	(0,312)	(0,251)	(0,149)	(0,331)	(0,402)	(0,291)	(0,308)	(0,508)	**(0,015)**	**(0,016)**	(0,278)	(0,148)
M2→IN	17,619	15,153	11,209	13,521	16,915	15,141	20,159	21,609	15,111	20,862	22,949	38,093
	(0,391)	(0,549)	(0,818)	(0,670)	(0,516)	(0,664)	(0,463)	(0,463)	(0,715)	(0,546)	(0,502)	(0,164)

h	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
M2→SI	7,255	9,929	11,040	16,152	18,852	17,322	12,954	16,956	23,615	23,898	34,194	22,450
	(0,983)	(0,932)	(0,938)	(0,825)	(0,746)	(0,802)	(0,918)	(0,851)	(0,700)	(0,682)	(0,500)	(0,781)
V→M2	21,332	18,005	18,869	26,196	17,226	27,806	25,297	36,471	40,588	**86,622**	**84,885**	**91,175**
	(0,559)	(0,673)	(0,669)	(0,455)	(0,772)	(0,454)	(0,539)	(0,336)	(0,299)	**(0,024)**	**(0,037)**	**(0,031)**
V→SI	8,513	9,897	10,050	13,724	22,299	13,441	16,645	17,141	18,031	13,022	6,910	6,875
	(0,960)	(0,942)	(0,940)	(0,846)	(0,579)	(0,893)	(0,806)	(0,793)	(0,810)	(0,921)	(0,991)	(0,993)
SI→M2	32,750	10,237	11,465	37,173	35,291	23,834	22,646	26,877	26,364	23,602	60,477	37,021
	(0,218)	(0,920)	(0,907)	(0,198)	(0,286)	(0,574)	(0,643)	(0,575)	(0,573)	(0,670)	(0,134)	(0,411)
V→IN	37,667	28,558	13,472	15,432	16,240	23,309	**64,221**	**83,729**	**102,321**	**115,436**	65,944	**117,593**
	(0,179)	(0,390)	(0,861)	(0,839)	(0,837)	(0,674)	**(0,074)**	**(0,032)**	**(0,015)**	**(0,013)**	(0,144)	**(0,013)**
SI→IN	37,879	37,606	21,542	17,300	13,464	6,789	7,206	21,716	30,671	35,639	45,443	68,205
	(0,184)	(0,214)	(0,652)	(0,775)	(0,877)	(0,991)	(0,991)	(0,709)	(0,523)	(0,483)	(0,336)	(0,141)
M2→IN	45,765	45,914	33,632	39,546	55,241	59,958	51,298	**70,608**	**80,460**	**75,114**	72,221	56,160
	(0,106)	(0,119)	(0,343)	(0,238)	(0,103)	(0,110)	(0,169)	**(0,089)**	**(0,054)**	**(0,086)**	(0,121)	(0,259)

h	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
M2→SI	33,809	18,437	19,646	19,146	15,266	14,508	18,621	20,518	19,040	22,400	24,630	15,500
	(0,556)	(0,870)	(0,882)	(0,893)	(0,952)	(0,960)	(0,929)	(0,915)	(0,931)	(0,914)	(0,904)	(0,972)
V→M2	**75,568**	50,610	53,125	66,963	79,446	**110,209**	63,865	**113,004**	**108,716**	**189,152**	**144,212**	106,789
	(0,074)	(0,259)	(0,304)	(0,191)	(0,119)	**(0,051)**	(0,253)	**(0,061)**	**(0,065)**	**(0,011)**	**(0,035)**	(0,122)
V→SI	15,448	14,340	23,895	9,415	16,653	32,576	23,601	22,920	22,166	19,063	30,993	32,701
	(0,919)	(0,929)	(0,770)	(0,990)	(0,933)	(0,688)	(0,848)	(0,871)	(0,898)	(0,935)	(0,798)	(0,780)
SI→M2	40,556	**76,042**	**78,904**	35,968	46,082	52,941	26,111	39,538	55,186	73,521	55,356	61,961
	(0,398)	**(0,090)**	**(0,097)**	(0,585)	(0,384)	(0,346)	(0,804)	(0,560)	(0,382)	(0,261)	(0,446)	(0,398)
V→IN	72,905	77,372	63,208	48,196	83,506	**109,075**	**178,650**	**209,740**	**155,553**	**126,066**	**159,605**	**128,485**
	(0,134)	(0,123)	(0,248)	(0,447)	(0,138)	**(0,073)**	**(0,015)**	**(0,002)**	**(0,023)**	**(0,073)**	**(0,033)**	**(0,088)**
SI→IN	49,684	48,764	**88,652**	**101,234**	61,764	68,357	55,682	70,253	58,033	65,484	86,906	41,118
	(0,347)	(0,362)	**(0,092)**	**(0,076)**	(0,283)	(0,273)	(0,459)	(0,278)	(0,417)	(0,380)	(0,240)	(0,707)
M2→IN	54,530	58,166	71,581	48,266	**117,863**	**164,109**	**214,533**	**196,578**	**180,683**	**235,857**	**183,121**	109,987
	(0,301)	(0,267)	(0,193)	(0,484)	**(0,052)**	**(0,020)**	**(0,010)**	**(0,013)**	**(0,019)**	**(0,006)**	**(0,019)**	(0,181)

Note: this table reports the Dufour et al. (2006) test results for forecast horizons (*h*) 1 month ahead to 36 months ahead at nominal level of significance 10% . Simulated *p*-values are in parentheses. Rejection of the null hypothesis of non-causality is in bold. V, SI, IN, M2 and IP stand for volatility of stock returns, short-term nominal interest rates, inflation, M2 growth rate, and industrial production growth rate, respectively.

Table 3a. Causality test results for the forecast horizons 1 to 36 months ahead in Germany

h	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
V→IP	20,943	22,748	**25,799**	21,380	23,403	23,017	**28,586**	**35,518**	**33,861**	**32,196**	14,343	16,603
	(0,251)	(0,139)	**(0,088)**	(0,232)	(0,178)	(0,230)	**(0,086)**	**(0,028)**	**(0,090)**	**(0,094)**	(0,693)	(0,599)
SI→IP	12,150	8,154	15,604	15,531	11,412	20,601	19,435	23,339	**33,538**	**29,333**	**45,173**	**38,655**
	(0,470)	(0,790)	(0,324)	(0,298)	(0,554)	(0,155)	(0,202)	(0,122)	**(0,021)**	**(0,056)**	**(0,013)**	**(0,021)**
IN→IP	**29,920**	24,897	**28,684**	19,773	19,372	25,919	25,975	**37,829**	**39,627**	**47,334**	26,616	**36,853**
	(0,034)	(0,103)	**(0,060)**	(0,303)	(0,318)	(0,146)	(0,163)	**(0,041)**	**(0,030)**	**(0,009)**	(0,202)	**(0,092)**
M2→IP	**53,197**	**60,529**	**74,610**	**64,235**	**59,409**	**54,514**	**52,442**	**66,580**	**61,113**	**45,250**	**49,592**	**113,699**
	(0,001)	**(0,001)**	**(0,001)**	**(0,001)**	**(0,001)**	**(0,001)**	**(0,002)**	**(0,002)**	**(0,004)**	**(0,018)**	**(0,016)**	**(0,001)**

h	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
V→IP	19,096	17,720	17,171	23,051	26,680	39,289	43,760	37,152	**45,102**	**54,399**	**55,486**	35,271
	(0,547)	(0,609)	(0,648)	(0,397)	(0,352)	(0,127)	(0,107)	(0,185)	**(0,088)**	**(0,056)**	**(0,047)**	(0,262)
SI→IP	25,805	23,403	27,360	25,752	17,531	**35,641**	23,088	15,565	23,740	22,867	29,349	35,257
	(0,124)	(0,185)	(0,137)	(0,174)	(0,470)	**(0,073)**	(0,322)	(0,562)	(0,324)	(0,354)	(0,206)	(0,144)
IN→IP	**44,224**	28,659	30,213	36,950	36,087	29,571	21,976	11,039	13,559	21,996	24,646	22,589
	(0,039)	(0,232)	(0,221)	(0,116)	(0,139)	(0,294)	(0,521)	(0,940)	(0,884)	(0,620)	(0,543)	(0,631)
M2→IP	**134,278**	**120,453**	**124,705**	**129,345**	**149,234**	**181,370**	**206,442**	**164,436**	**195,839**	**278,081**	**336,938**	**302,350**
	(0,001)	**(0,001)**	**(0,001)**	**(0,001)**	**(0,001)**	**(0,001)**	**(0,001)**	**(0,001)**	**(0,001)**	**(0,001)**	**(0,001)**	**(0,001)**

h	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
V→IP	53,020	42,069	26,027	**56,280**	40,484	**71,664**	**77,799**	**65,163**	**68,311**	51,654	54,995	56,640
	(0,109)	(0,180)	(0,567)	**(0,077)**	(0,305)	**(0,049)**	**(0,039)**	**(0,075)**	**(0,071)**	(0,206)	(0,225)	(0,204)
SI→IP	36,915	37,190	**52,086**	**47,537**	**50,574**	31,244	**59,287**	**59,436**	47,653	**62,067**	**52,569**	**63,433**
	(0,127)	(0,122)	**(0,039)**	**(0,062)**	**(0,060)**	(0,303)	**(0,030)**	**(0,043)**	(0,101)	**(0,041)**	**(0,092)**	**(0,045)**
IN→IP	21,841	20,397	24,641	24,030	22,328	29,632	38,279	40,767	56,368	43,993	22,515	27,801
	(0,691)	(0,747)	(0,661)	(0,680)	(0,710)	(0,536)	(0,378)	(0,369)	(0,200)	(0,354)	(0,815)	(0,719)
M2→IP	**145,389**	**240,420**	**302,396**	**315,934**	**268,432**	**155,161**	**158,216**	**64,245**	37,996	36,419	47,597	**103,553**
	(0,002)	**(0,001)**	**(0,001)**	**(0,001)**	**(0,001)**	**(0,003)**	**(0,001)**	**(0,086)**	(0,431)	(0,468)	(0,290)	**(0,023)**

Note: this table reports the Dufour et al. (2006) test results for forecast horizons (*h*) 1 month ahead to 36 months ahead at nominal level of significance 10% . Simulated *p*-values are in parentheses. Rejection of the null hypothesis of non-causality is in bold. V, SI, IN, M2 and IP stand for volatility of stock returns, short-term nominal interest rates, inflation, M2 growth rate, and industrial production growth rate, respectively.

Table 3b. Causality test results for the forecast horizons 1 to 36 months ahead in Germany

h	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
M2→SI	**33,256**	**37,967**	**54,205**	**43,504**	**33,543**	**39,165**	**29,182**	**34,169**	22,125	24,482	24,114	27,337
	(0,010)	**(0,009)**	**(0,001)**	**(0,003)**	**(0,025)**	**(0,013)**	**(0,063)**	**(0,048)**	(0,266)	(0,238)	(0,266)	(0,177)
V→M2	7,450	8,348	9,177	8,584	7,731	7,003	6,800	8,949	11,793	17,372	25,051	12,113
	(0,936)	(0,911)	(0,890)	(0,922)	(0,932)	(0,973)	(0,959)	(0,921)	(0,839)	(0,569)	(0,278)	(0,842)
V→SI	16,902	15,164	14,772	15,212	19,140	17,896	19,603	29,764	25,122	25,764	25,228	22,049
	(0,383)	(0,469)	(0,516)	(0,489)	(0,314)	(0,389)	(0,311)	(0,083)	(0,166)	(0,212)	(0,243)	(0,378)
SI→M2	15,512	12,549	12,668	14,449	11,630	10,442	12,667	11,038	17,339	14,495	14,255	11,072
	(0,465)	(0,704)	(0,689)	(0,616)	(0,795)	(0,849)	(0,774)	(0,858)	(0,546)	(0,719)	(0,747)	(0,904)
IN→M2	7,971	8,493	11,149	7,755	7,455	9,707	8,669	8,006	9,012	7,496	18,649	18,127
	(0,927)	(0,908)	(0,791)	(0,940)	(0,953)	(0,874)	(0,927)	(0,947)	(0,933)	(0,969)	(0,540)	(0,577)

h	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
M2→SI	29,286	31,381	**43,083**	**53,009**	**63,372**	**60,802**	**60,126**	**68,506**	**48,298**	43,583	**84,478**	**85,776**
	(0,180)	(0,151)	**(0,042)**	**(0,015)**	**(0,003)**	**(0,008)**	**(0,012)**	**(0,004)**	**(0,069)**	(0,110)	**(0,006)**	**(0,005)**
V→M2	7,062	7,815	12,363	10,948	14,656	11,378	11,520	12,529	25,342	31,186	32,590	46,052
	(0,978)	(0,974)	(0,872)	(0,922)	(0,809)	(0,911)	(0,922)	(0,906)	(0,484)	(0,344)	(0,346)	(0,154)
V→SI	**37,599**	**48,435**	**56,864**	**68,786**	**69,042**	**39,597**	35,506	39,972	35,981	30,691	27,270	18,168
	(0,074)	**(0,022)**	**(0,010)**	**(0,005)**	**(0,005)**	**(0,093)**	(0,163)	(0,113)	(0,180)	(0,315)	(0,418)	(0,741)
SI→M2	14,850	10,671	11,301	9,456	11,818	16,093	18,920	14,834	12,395	10,147	13,259	16,366
	(0,742)	(0,920)	(0,892)	(0,957)	(0,899)	(0,791)	(0,649)	(0,825)	(0,910)	(0,962)	(0,887)	(0,827)
IN→M2	17,124	17,599	11,787	11,319	10,791	18,445	18,262	17,768	20,404	27,309	34,470	34,253
	(0,622)	(0,668)	(0,894)	(0,923)	(0,941)	(0,690)	(0,700)	(0,752)	(0,719)	(0,467)	(0,300)	(0,307)

h	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
M2→SI	**51,064**	**54,226**	**55,119**	**104,101**	**89,300**	**101,766**	**93,029**	**96,520**	**82,264**	51,914	**98,197**	**112,874**
	(0,091)	**(0,090)**	**(0,081)**	**(0,004)**	**(0,011)**	**(0,007)**	**(0,013)**	**(0,016)**	**(0,030)**	(0,196)	**(0,020)**	**(0,014)**
V→M2	31,345	35,176	29,208	35,733	26,421	35,988	43,362	39,050	54,656	33,110	22,392	19,430
	(0,448)	(0,326)	(0,497)	(0,397)	(0,623)	(0,416)	(0,306)	(0,405)	(0,197)	(0,538)	(0,820)	(0,877)
V→SI	25,903	19,368	23,736	19,413	24,725	22,986	27,882	28,894	22,854	29,326	36,275	22,403
	(0,503)	(0,709)	(0,603)	(0,770)	(0,601)	(0,693)	(0,540)	(0,570)	(0,721)	(0,579)	(0,440)	(0,809)
SI→M2	10,470	12,643	12,401	11,676	13,610	10,890	16,564	19,663	23,198	40,248	34,778	18,317
	(0,964)	(0,933)	(0,945)	(0,943)	(0,919)	(0,974)	(0,889)	(0,837)	(0,776)	(0,425)	(0,553)	(0,905)
IN→M2	13,045	15,469	13,101	16,871	15,037	15,798	18,906	30,093	36,004	30,029	17,358	34,662
	(0,930)	(0,872)	(0,935)	(0,888)	(0,917)	(0,923)	(0,830)	(0,603)	(0,498)	(0,661)	(0,896)	(0,594)

Note: this table reports the Dufour et al. (2006) test results for forecast horizons (*h*) 1 month ahead to 36 months ahead at nominal level of significance 10% . Simulated *p*-values are in parentheses. Rejection of the null hypothesis of non-causality is in bold. V, SI, IN, M2 and IP stand for volatility of stock returns, short-term nominal interest rates, inflation, M2 growth rate, and industrial production growth rate, respectively.

Table 4a. Causality test results for the forecast horizons 1 to 36 months ahead in Italy

h	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
V→IP	16,647	11,978	9,909	11,557	11,439	13,343	13,988	16,390	18,757	22,007	26,589	32,281
	(0,404)	(0,621)	(0,773)	(0,708)	(0,740)	(0,601)	(0,610)	(0,495)	(0,435)	(0,307)	(0,188)	(0,122)
SI→IP	**34,236**	20,990	**24,568**	**25,741**	23,397	20,417	18,296	14,496	20,735	20,421	20,953	15,350
	(0,013)	(0,175)	**(0,097)**	**(0,089)**	(0,154)	(0,257)	(0,394)	(0,596)	(0,308)	(0,383)	(0,339)	(0,612)
IN→IP	10,392	12,086	14,226	11,948	15,898	18,827	21,599	27,091	22,018	22,009	25,214	32,567
	(0,781)	(0,628)	(0,482)	(0,665)	(0,489)	(0,364)	(0,287)	(0,143)	(0,323)	(0,347)	(0,233)	(0,144)
M2→IP	12,183	15,270	19,887	18,774	11,893	10,436	8,381	8,741	7,469	20,309	28,753	23,517
	(0,661)	(0,420)	(0,233)	(0,302)	(0,691)	(0,814)	(0,905)	(0,912)	(0,940)	(0,370)	(0,174)	(0,312)

h	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
V→IP	19,626	22,654	22,565	23,112	29,573	28,256	**51,790**	**46,474**	32,508	30,190	28,372	22,144
	(0,460)	(0,382)	(0,392)	(0,397)	(0,276)	(0,313)	**(0,057)**	**(0,083)**	(0,264)	(0,373)	(0,422)	(0,614)
SI→IP	12,749	17,984	15,129	14,561	21,969	16,326	16,297	16,246	13,156	10,158	8,793	12,855
	(0,784)	(0,540)	(0,709)	(0,766)	(0,446)	(0,722)	(0,728)	(0,755)	(0,877)	(0,942)	(0,962)	(0,911)
IN→IP	34,338	**41,828**	**44,718**	**50,253**	37,418	38,935	40,489	**44,462**	45,428	36,687	31,805	31,808
	(0,119)	**(0,053)**	**(0,045)**	**(0,033)**	(0,107)	(0,140)	(0,113)	**(0,099)**	(0,107)	(0,227)	(0,331)	(0,370)
M2→IP	17,984	14,798	14,068	10,876	10,622	13,414	12,554	13,349	18,855	15,592	24,933	17,052
	(0,527)	(0,695)	(0,740)	(0,895)	(0,896)	(0,837)	(0,841)	(0,843)	(0,673)	(0,820)	(0,503)	(0,770)

h	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
V→IP	20,291	39,639	41,313	46,754	39,346	46,765	45,688	33,871	20,925	18,347	15,924	18,860
	(0,705)	(0,293)	(0,272)	(0,214)	(0,350)	(0,253)	(0,272)	(0,512)	(0,813)	(0,870)	(0,922)	(0,890)
SI→IP	17,593	19,297	17,048	18,594	19,289	18,415	24,757	26,269	44,743	53,702	51,273	41,436
	(0,768)	(0,754)	(0,808)	(0,776)	(0,800)	(0,824)	(0,687)	(0,659)	(0,326)	(0,251)	(0,277)	(0,429)
IN→IP	41,679	52,233	43,049	35,156	46,997	48,592	31,316	31,078	18,822	30,538	19,017	29,228
	(0,205)	(0,133)	(0,210)	(0,363)	(0,211)	(0,227)	(0,525)	(0,552)	(0,854)	(0,624)	(0,890)	(0,700)
M2→IP	32,996	23,021	22,042	16,574	16,580	16,091	17,391	24,219	45,328	72,217	18,074	24,370
	(0,359)	(0,676)	(0,688)	(0,825)	(0,863)	(0,894)	(0,861)	(0,740)	(0,309)	(0,102)	(0,906)	(0,766)

Note: this table reports the Dufour et al. (2006) test results for forecast horizons (*h*) 1 month ahead to 36 months ahead at nominal level of significance 10% . Simulated *p*-values are in parentheses. Rejection of the null hypothesis of non-causality is in bold. V, SI, IN, M2 and IP stand for volatility of stock returns, short-term nominal interest rates, inflation, M2 growth rate, and industrial production growth rate, respectively.

Table 4b. Causality test results for the forecast horizons 1 to 36 months ahead in Italy

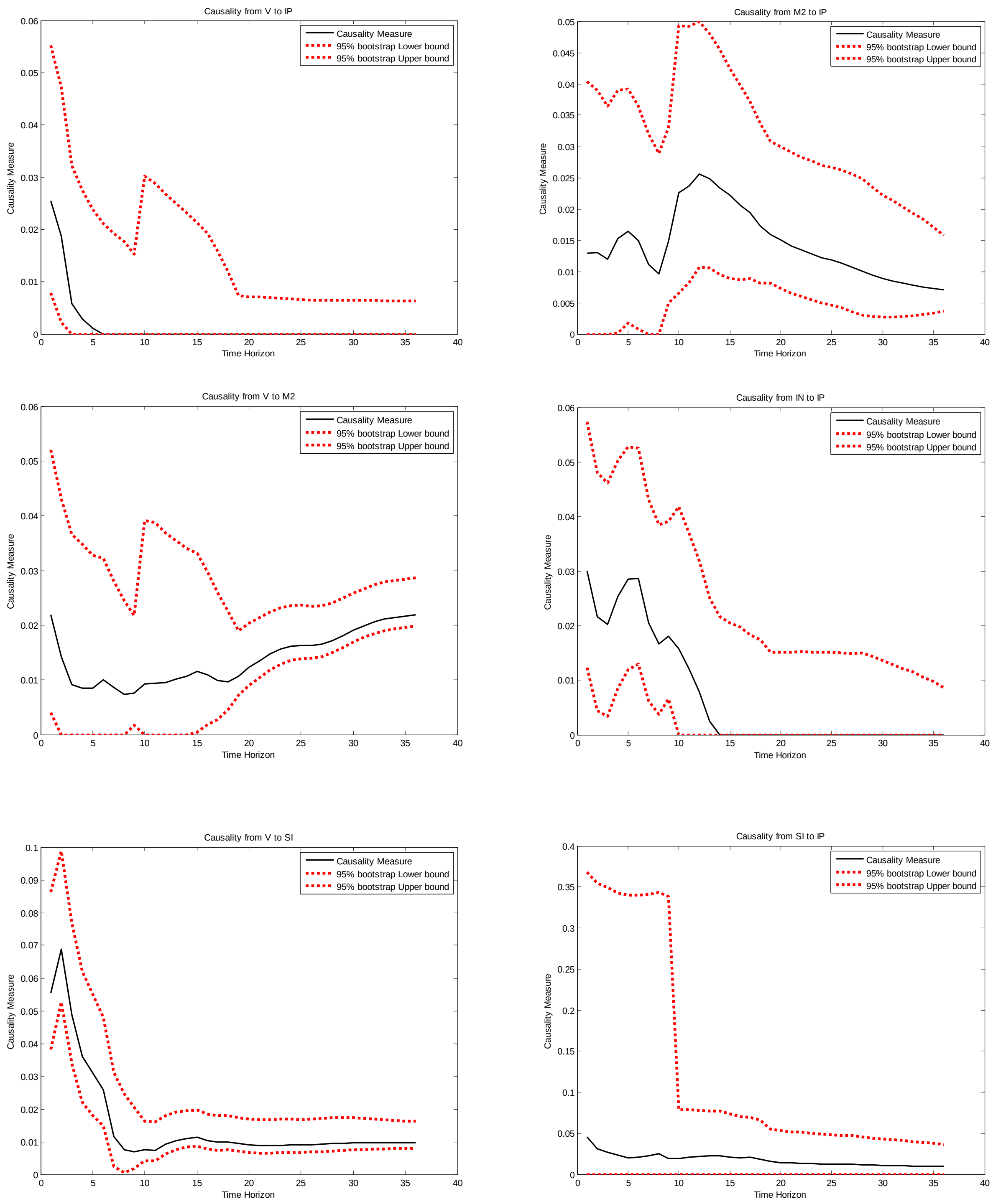
h	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
SI→V	14,376	15,604	17,522	12,576	10,641	10,280	8,320	8,642	12,695	15,567	15,156	14,441
	(0,513)	(0,469)	(0,377)	(0,670)	(0,834)	(0,837)	(0,923)	(0,920)	(0,788)	(0,607)	(0,705)	(0,747)
V→IN	**28,249**	**28,654**	**30,157**	**31,654**	**34,884**	**39,537**	**36,877**	**38,316**	**45,205**	**50,243**	32,283	**41,563**
	(0,047)	**(0,054)**	**(0,053)**	**(0,043)**	**(0,028)**	**(0,019)**	**(0,024)**	**(0,030)**	**(0,007)**	**(0,011)**	(0,110)	**(0,046)**
M2→IN	**55,648**	**45,897**	**49,177**	**42,680**	**43,457**	**53,865**	**52,360**	**52,946**	**52,549**	27,302	**36,765**	**38,319**
	(0,001)	**(0,001)**	**(0,002)**	**(0,005)**	**(0,007)**	**(0,002)**	**(0,003)**	**(0,003)**	**(0,007)**	(0,148)	**(0,072)**	**(0,076)**
V→M2	13,808	16,162	16,696	23,244	16,675	19,087	20,131	22,348	21,850	31,151	26,661	12,298
	(0,560)	(0,424)	(0,400)	(0,179)	(0,474)	(0,397)	(0,372)	(0,297)	(0,328)	(0,126)	(0,236)	(0,827)
SI→IN	24,021	**28,077**	23,995	22,931	23,050	29,701	24,275	**34,274**	23,353	20,079	12,349	16,581
	(0,108)	**(0,056)**	(0,128)	(0,155)	(0,157)	(0,065)	(0,198)	**(0,047)**	(0,270)	(0,421)	(0,800)	(0,623)

h	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
SI→V	16,358	21,905	30,070	8,231	9,653	8,670	6,985	10,252	12,096	6,779	8,984	5,410
	(0,674)	(0,448)	(0,222)	(0,971)	(0,933)	(0,965)	(0,984)	(0,956)	(0,917)	(0,991)	(0,980)	(0,995)
V→IN	**29,251**	**43,150**	**76,238**	**77,026**	**75,669**	**43,668**	**58,670**	**94,643**	**85,593**	**89,001**	**68,749**	**79,577**
	(0,193)	**(0,051)**	**(0,004)**	**(0,004)**	**(0,006)**	**(0,099)**	**(0,032)**	**(0,004)**	**(0,006)**	**(0,004)**	**(0,024)**	**(0,018)**
M2→IN	26,210	28,156	26,542	32,921	46,828	36,088	35,057	43,348	44,678	**65,694**	**60,200**	48,416
	(0,351)	(0,298)	(0,382)	(0,241)	(0,074)	(0,193)	(0,261)	(0,125)	(0,133)	**(0,040)**	**(0,077)**	(0,157)
V→M2	13,703	20,115	30,395	32,598	**88,057**	**67,019**	**73,341**	**70,394**	**73,082**	**77,614**	**71,764**	39,678
	(0,783)	(0,551)	(0,231)	(0,217)	**(0,003)**	**(0,019)**	**(0,009)**	**(0,022)**	**(0,016)**	**(0,021)**	**(0,026)**	(0,255)
SI→IN	20,542	29,928	26,138	25,129	21,495	15,499	21,724	31,547	26,165	18,256	25,076	15,405
	(0,478)	(0,211)	(0,332)	(0,381)	(0,538)	(0,754)	(0,569)	(0,315)	(0,486)	(0,714)	(0,526)	(0,857)

h	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
SI→V	11,055	25,926	19,340	30,538	43,252	43,123	29,421	25,370	23,448	54,360	42,018	51,200
	(0,948)	(0,614)	(0,816)	(0,516)	(0,316)	(0,347)	(0,616)	(0,700)	(0,794)	(0,282)	(0,471)	(0,341)
V→IN	**92,187**	**161,746**	**98,158**	**105,759**	**81,628**	**84,737**	**108,043**	**82,085**	**79,382**	61,658	55,107	46,194
	(0,013)	**(0,003)**	**(0,011)**	**(0,013)**	**(0,041)**	**(0,045)**	**(0,021)**	**(0,055)**	**(0,078)**	(0,181)	(0,239)	(0,394)
M2→IN	**63,571**	**68,033**	**64,814**	**104,137**	**127,539**	**120,533**	**132,517**	53,743	76,005	66,464	67,979	**101,334**
	(0,097)	**(0,075)**	**(0,085)**	**(0,026)**	**(0,010)**	**(0,013)**	**(0,012)**	(0,263)	(0,105)	(0,167)	(0,187)	**(0,063)**
V→M2	**75,921**	**86,542**	**74,262**	**161,164**	**137,440**	69,829	58,979	61,404	**97,723**	**148,357**	**120,401**	52,478
	(0,035)	**(0,014)**	**(0,039)**	**(0,003)**	**(0,004)**	(0,101)	(0,178)	(0,166)	**(0,037)**	**(0,006)**	**(0,030)**	(0,341)
SI→IN	10,431	17,241	21,927	21,669	21,527	48,681	42,869	48,014	44,893	37,474	65,315	70,069
	(0,956)	(0,844)	(0,742)	(0,722)	(0,758)	(0,263)	(0,330)	(0,284)	(0,347)	(0,488)	(0,184)	(0,134)

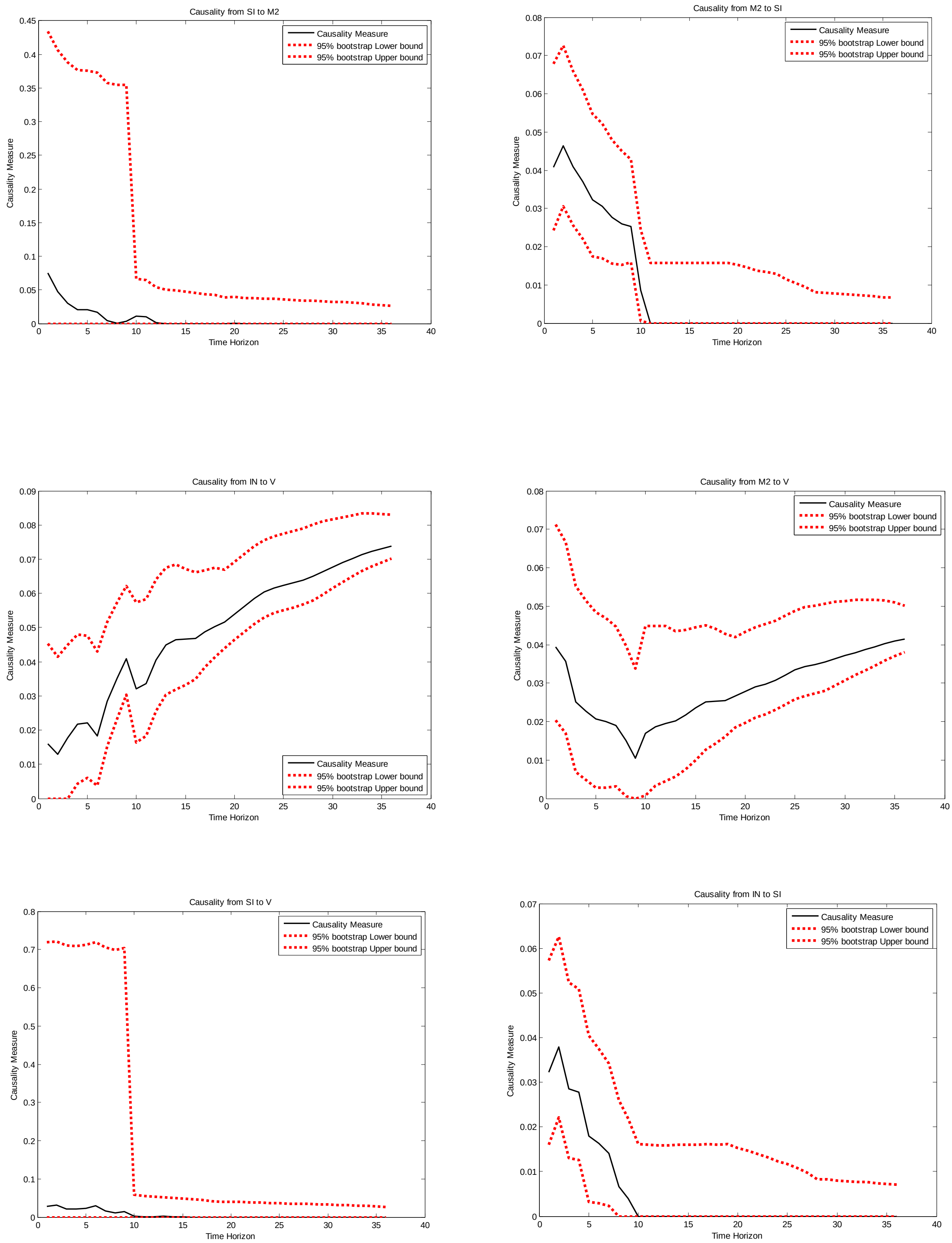
Note: this table reports the Dufour et al. (2006) test results for forecast horizons (*h*) 1 month ahead to 36 months ahead at nominal level of significance 10% . Simulated *p*-values are in parentheses. Rejection of the null hypothesis of non-causality is in bold. V, SI, IN, M2 and IP stand for volatility of stock returns, short-term nominal interest rates, inflation, M2 growth rate, and industrial production growth rate, respectively.

Panel 1a: Granger causality measures from 1 month-ahead to 36 months ahead between the volatility of stock returns, nominal short-term interest rates, inflation, money supply growth rates and industrial production growth rates in US



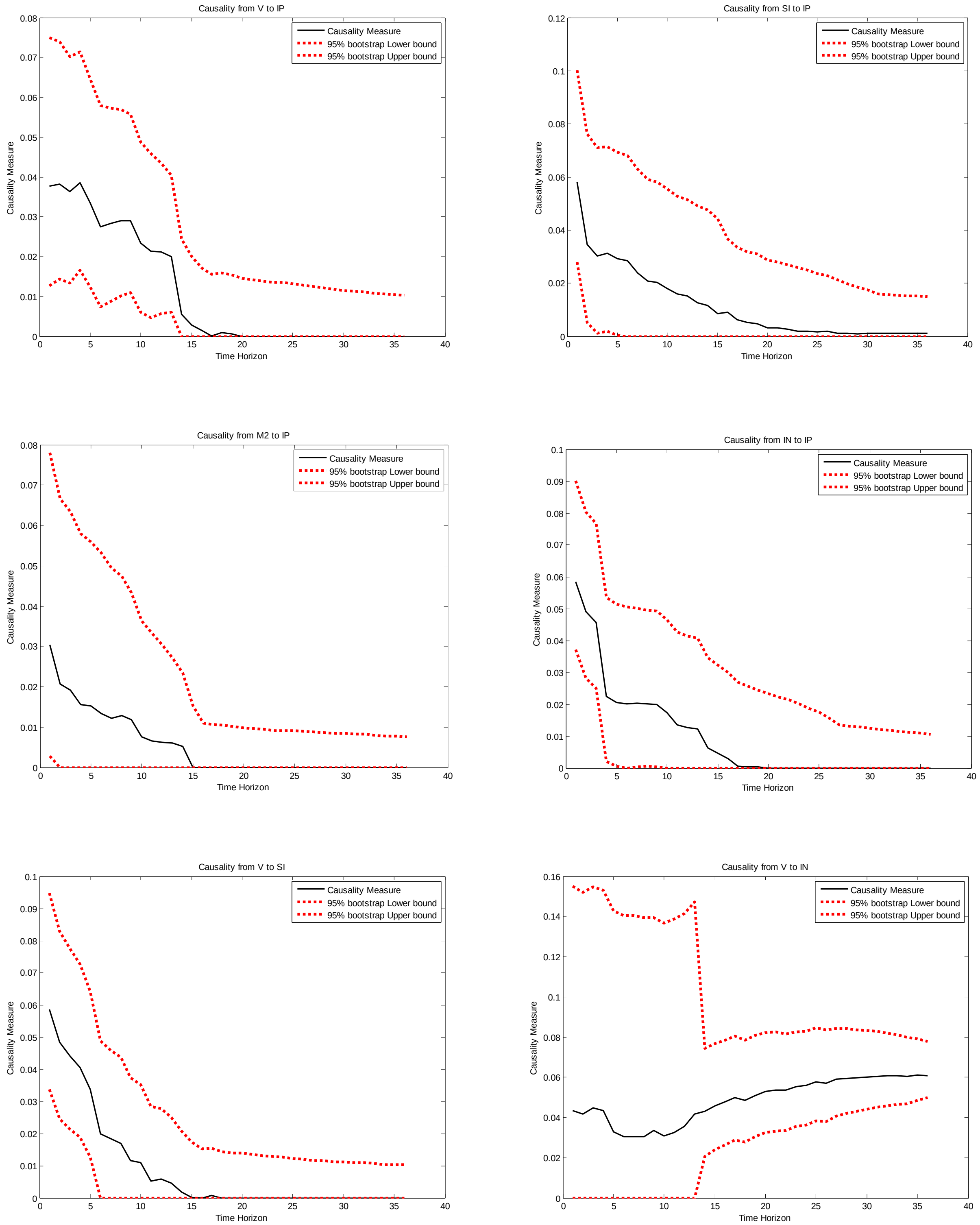
Note: V, SI, IN, M2 and IP stand for volatility of stock returns, short-term nominal interest rates, inflation, M2 growth rates, and industrial production growth rates, respectively.

Panel 1b: Granger causality measures from 1 month-ahead to 36 months ahead between the volatility of stock returns, nominal short-term interest rates, inflation, money supply growth rates and industrial production growth rates in US



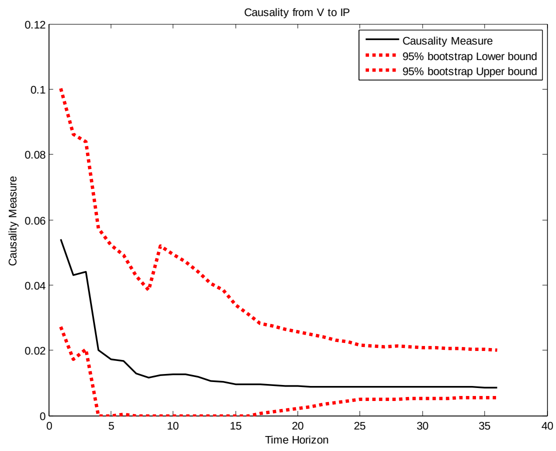
Note: V, SI, IN, M2 and IP stand for volatility of stock returns, short-term nominal interest rates, inflation, M2 growth rates, and industrial production growth rates, respectively.

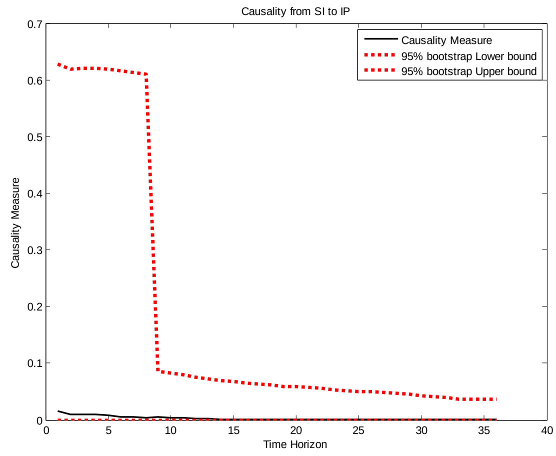
Panel 2: Granger causality measures from 1 month-ahead to 36 months ahead between the volatility of stock returns, nominal short-term interest rates, inflation, money supply growth rates and industrial production growth rates in Germany

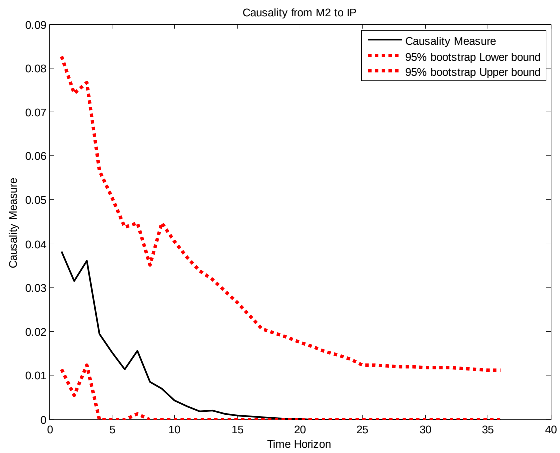


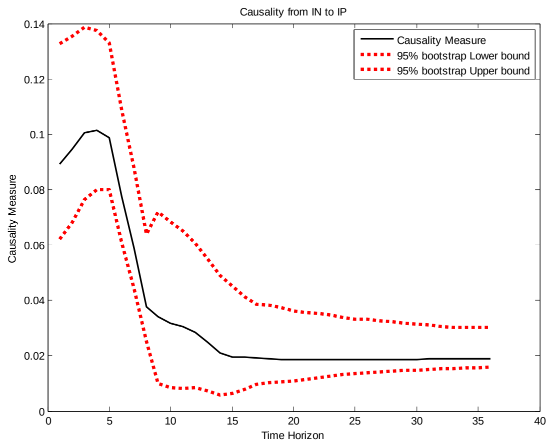
Note: V, SI, IN, M2 and IP stand for volatility of stock returns, short-term nominal interest rates, inflation, M2 growth rates, and industrial production growth rates, respectively.

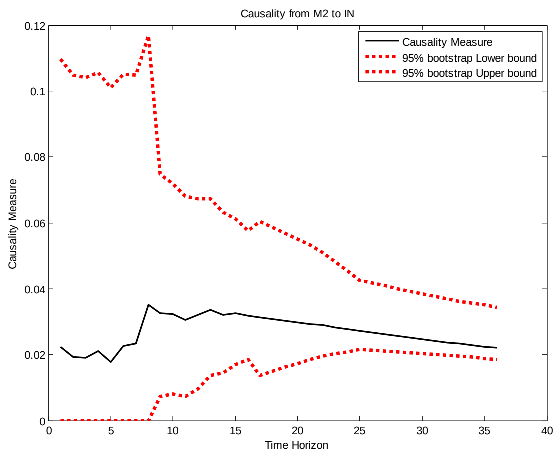
Panel 3a: Granger causality measures from 1 month-ahead to 36 months ahead between the volatility of stock returns, nominal short-term interest rates, inflation, money supply growth rates and industrial production growth rates in Japan.

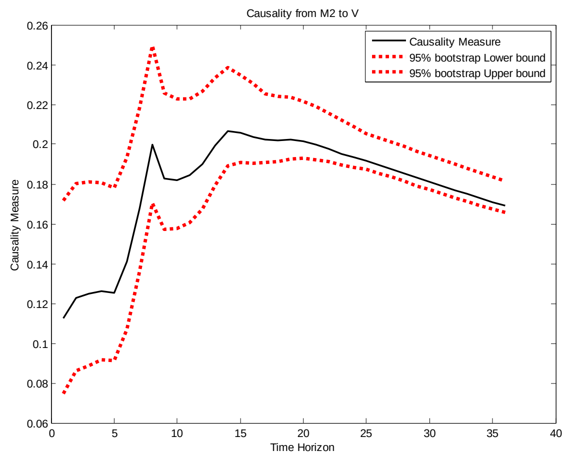






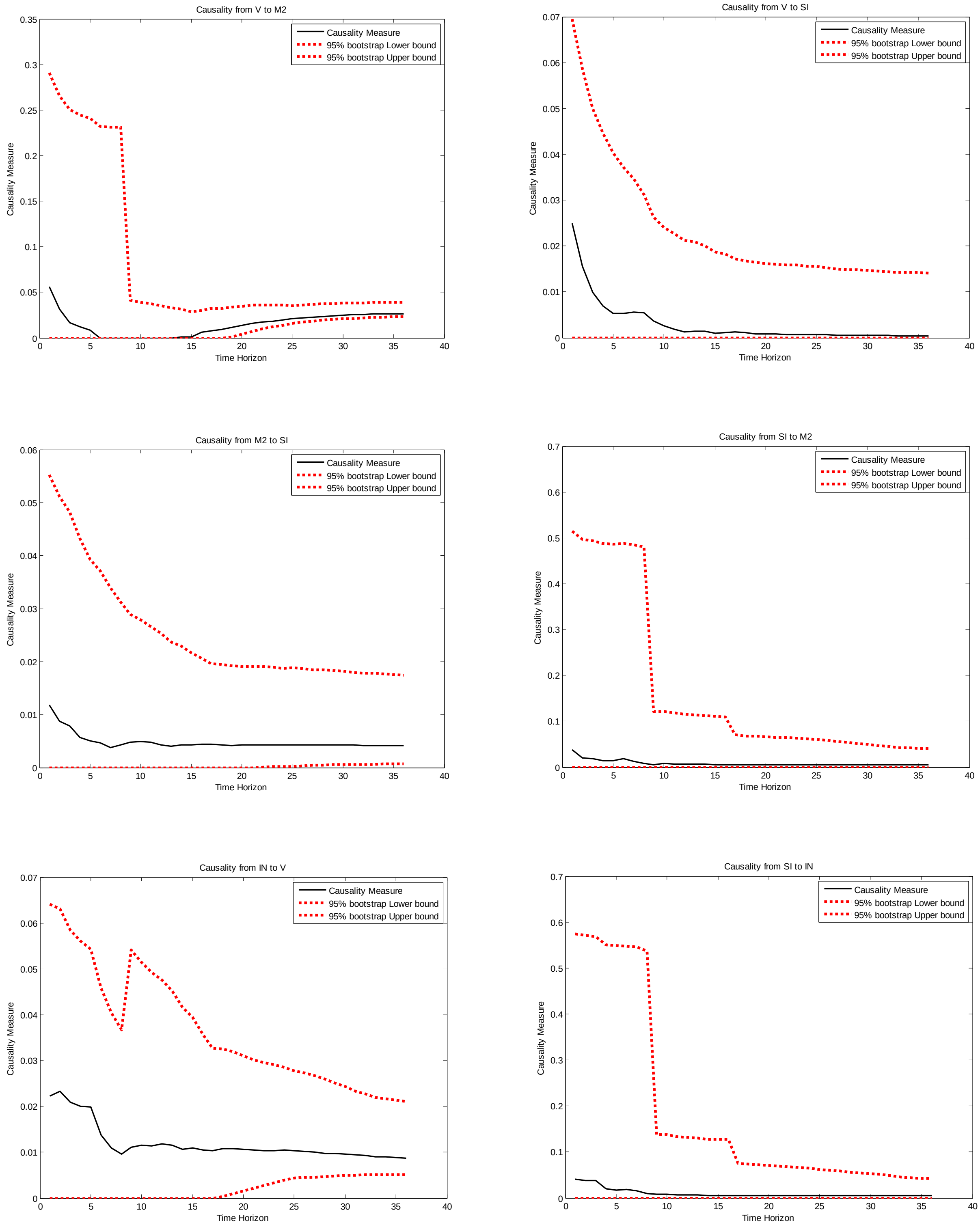






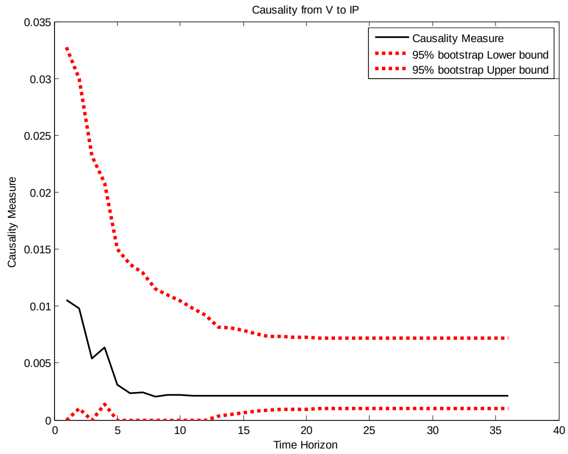
Note: V, SI, IN, M2 and IP stand for volatility of stock returns, short-term nominal interest rates, inflation, M2 growth rates, and industrial production growth rates, respectively.

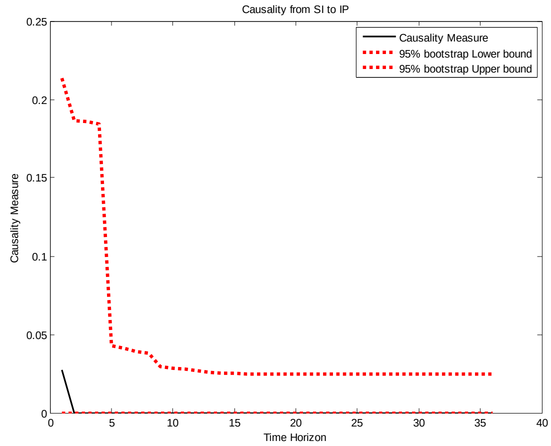
Panel 3b: Granger causality measures from 1 month-ahead to 36 months ahead between the volatility of stock returns, nominal short-term interest rates, inflation, money supply growth rates and industrial production growth rates in Japan.

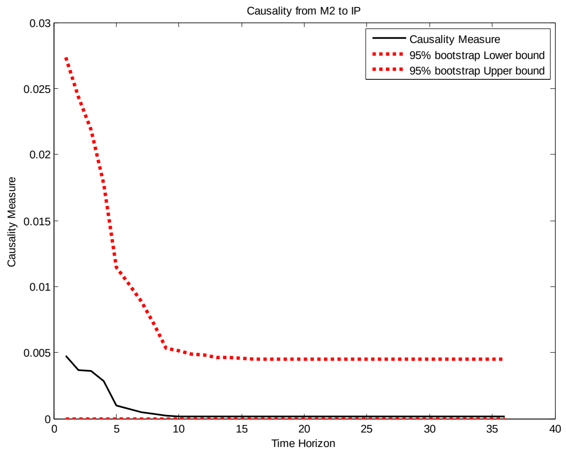


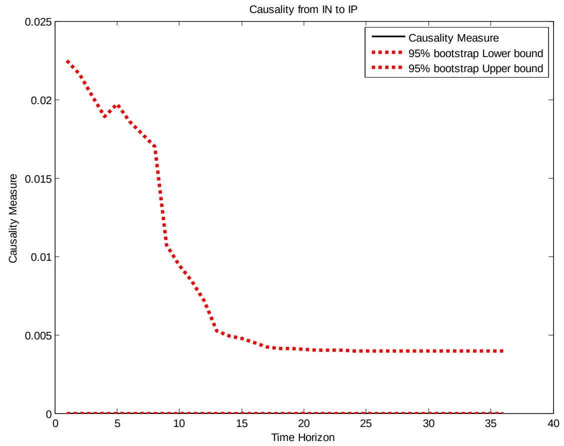
Note: V, SI, IN, M2 and IP stand for volatility of stock returns, short-term nominal interest rates, inflation, M2 growth rates, and industrial production growth rates, respectively.

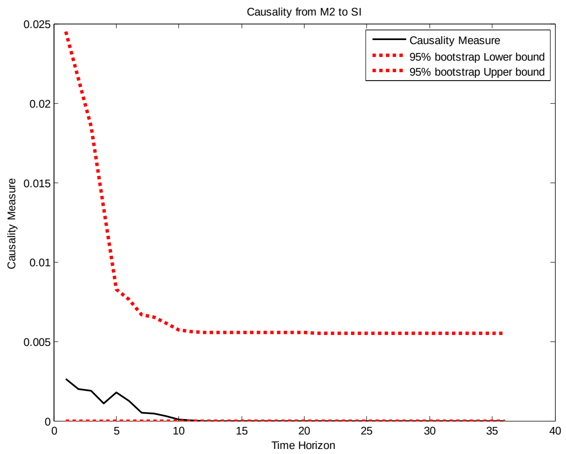
Panel 4a: Granger causality measures from 1 month-ahead to 36 months ahead between the volatility of stock returns, nominal short-term interest rates, inflation, money supply growth rates and industrial production growth rates in Italy.

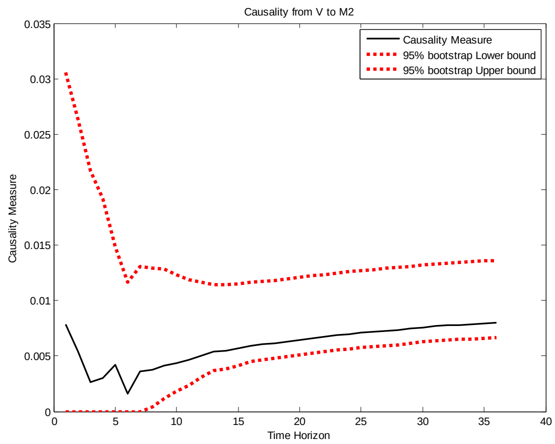






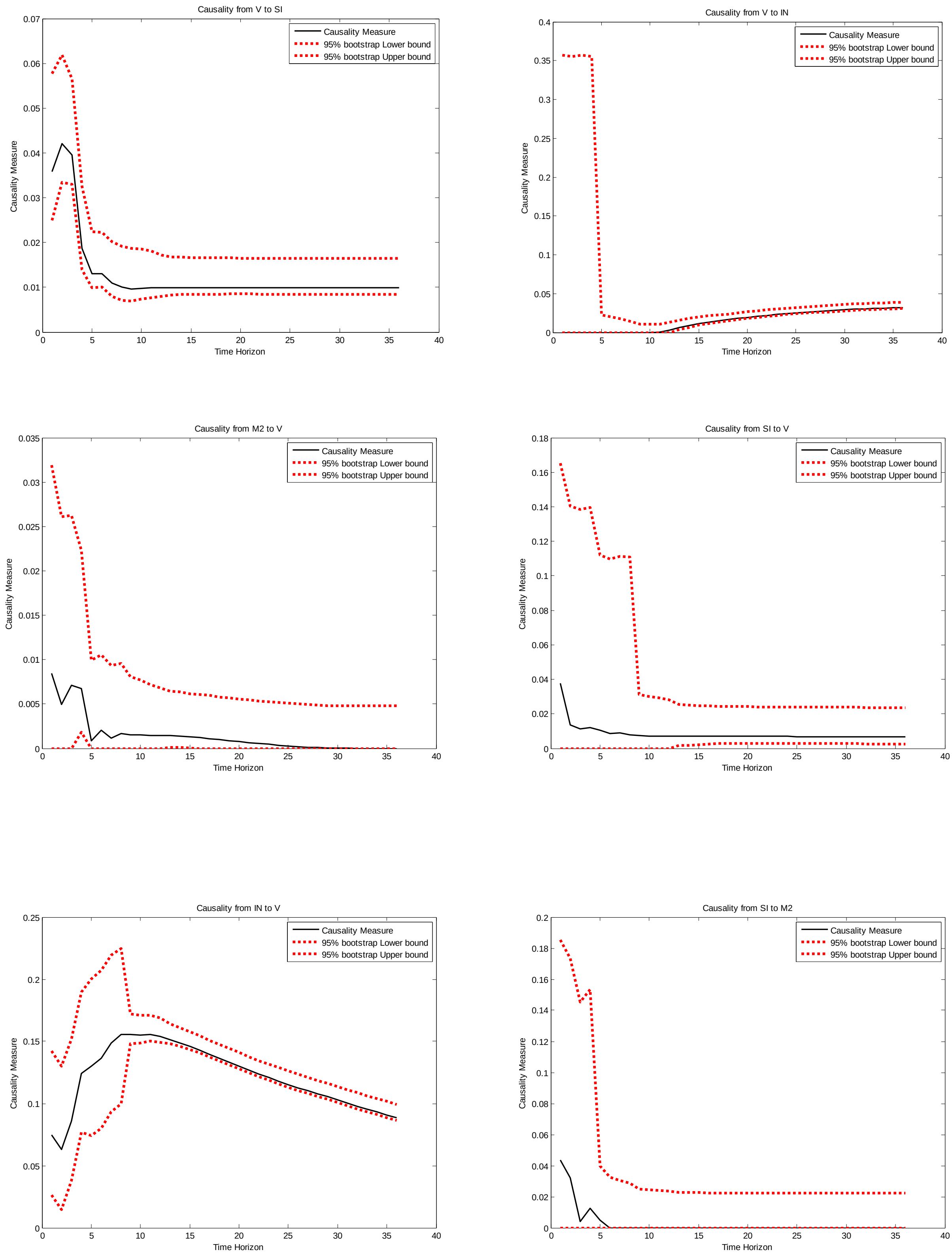






Note: V, SI, IN, M2 and IP stand for volatility of stock returns, short-term nominal interest rates, inflation, M2 growth rates, and industrial production growth rates, respectively.

Panel 4b: Granger causality measures from 1 month-ahead to 36 months ahead between the volatility of stock returns, nominal short-term interest rates, inflation, money supply growth rates and industrial production growth rates in Italy.



Note: V, SI, IN, M2 and IP stand for volatility of stock returns, short-term nominal interest rates, inflation, M2 growth rates, and industrial production growth rates, respectively.

Table 5. Measuring the forecasting ability of various monthly indicators in US

h	RMSFE		Theil		Bias		Var	
	model	AR	model	AR	model	AR	model	AR
1	0,010	0,009	0,623	0,616	0,020	0,000	0,874	0,885
2	0,013	0,012	0,546	0,523	0,088	0,007	0,686	0,743
3	0,018	0,016	0,567	0,549	0,138	0,018	0,505	0,543
4	0,023	0,021	0,592	0,582	0,171	0,025	0,427	0,460
5	0,028	0,025	0,609	0,610	0,199	0,031	0,383	0,419
6	0,034	0,030	0,641	0,653	0,213	0,037	0,340	0,377
7	0,040	0,036	0,670	0,691	0,227	0,042	0,305	0,346
8	0,046	0,041	0,687	0,715	0,244	0,049	0,285	0,333
9	0,052	0,045	0,701	0,736	0,270	0,061	0,267	0,321
10	0,057	0,049	0,714	0,755	0,296	0,074	0,249	0,308
11	0,062	0,053	0,723	0,769	0,326	0,088	0,234	0,298
12	0,067	0,056	0,726	0,775	0,362	0,106	0,221	0,292
13	0,070	0,059	0,728	0,776	0,405	0,129	0,207	0,286
14	0,074	0,061	0,729	0,775	0,458	0,161	0,189	0,275
15	0,077	0,063	0,727	0,771	0,504	0,189	0,178	0,270
16	0,081	0,064	0,722	0,764	0,541	0,212	0,171	0,273
17	0,084	0,066	0,716	0,756	0,566	0,225	0,167	0,276
18	0,087	0,068	0,710	0,749	0,587	0,234	0,162	0,279
19	0,089	0,068	0,700	0,735	0,609	0,241	0,155	0,279
20	0,090	0,068	0,692	0,722	0,627	0,247	0,145	0,276
21	0,092	0,068	0,682	0,707	0,651	0,255	0,134	0,271
22	0,093	0,067	0,674	0,691	0,675	0,264	0,120	0,261
23	0,093	0,066	0,664	0,670	0,702	0,277	0,105	0,250
24	0,095	0,065	0,658	0,655	0,724	0,286	0,090	0,232
25	0,096	0,064	0,654	0,643	0,743	0,296	0,074	0,211
26	0,097	0,063	0,652	0,631	0,760	0,305	0,059	0,187
27	0,098	0,062	0,649	0,619	0,778	0,316	0,044	0,158
28	0,100	0,061	0,648	0,605	0,795	0,331	0,031	0,132
29	0,101	0,060	0,645	0,591	0,811	0,346	0,022	0,111
30	0,103	0,059	0,645	0,579	0,823	0,361	0,015	0,093
31	0,105	0,059	0,646	0,569	0,830	0,375	0,010	0,076
32	0,108	0,060	0,647	0,564	0,833	0,383	0,007	0,065
33	0,111	0,060	0,649	0,561	0,837	0,393	0,005	0,056
34	0,114	0,061	0,653	0,558	0,838	0,405	0,004	0,052
35	0,118	0,062	0,656	0,556	0,841	0,421	0,003	0,051
36	0,122	0,064	0,662	0,558	0,845	0,438	0,003	0,053

Note: this table reports the root mean square forecast error (RMSFE), the Theil criterion (Theil), the bias (bias) and the variance (Var) components of the Theil mean square forecast error decomposition for thirty six forecasting horizons (*h*) based on the forecasting model

$$y_t = \alpha + \sum_{k=1}^{P} \beta_k y_{t-k} + \sum_{i=1}^{4} \sum_{k=1}^{p} \gamma_{ik} x_{it-k} + \varepsilon_t$$

where $y_t = IP_t$, and x_{it} are the predictors, i.e., $x_{it} = V_t, IN_t, SI_t, M2_t$. IP_t, V_t, IN_t, SI_t and $M2_t$ stand for industrial production growth rates, volatility of stock returns, short-term nominal interest rates, inflation, and M2 growth rates respectively. The iterated approach is used to construct the *h*-step ahead forecasts of monthly industrial production growth rates.

Table 6. Measuring the forecasting ability of various monthly indicators in Germany

h	RMSFE		Theil		Bias		Var	
	model	AR	model	AR	model	AR	model	AR
1	0,028	0,027	0,942	0,914	0,000	0,004	0,318	0,340
2	0,030	0,027	0,829	0,738	0,000	0,006	0,586	0,694
3	0,037	0,033	0,816	0,756	0,001	0,003	0,703	0,771
4	0,043	0,041	0,808	0,796	0,003	0,002	0,711	0,717
5	0,046	0,046	0,785	0,804	0,006	0,001	0,749	0,714
6	0,051	0,051	0,775	0,812	0,010	0,000	0,766	0,718
7	0,055	0,055	0,769	0,819	0,014	0,000	0,764	0,719
8	0,058	0,059	0,754	0,815	0,022	0,001	0,776	0,731
9	0,062	0,062	0,746	0,816	0,030	0,002	0,776	0,734
10	0,065	0,065	0,739	0,817	0,042	0,005	0,767	0,734
11	0,067	0,068	0,732	0,818	0,057	0,009	0,755	0,728
12	0,070	0,071	0,726	0,817	0,074	0,015	0,741	0,724
13	0,072	0,073	0,722	0,818	0,092	0,021	0,726	0,717
14	0,074	0,075	0,717	0,820	0,114	0,030	0,704	0,701
15	0,075	0,076	0,712	0,819	0,145	0,043	0,675	0,686
16	0,075	0,077	0,707	0,819	0,177	0,057	0,644	0,666
17	0,077	0,078	0,704	0,818	0,196	0,065	0,619	0,654
18	0,079	0,079	0,702	0,814	0,200	0,065	0,595	0,648
19	0,079	0,079	0,694	0,801	0,199	0,061	0,576	0,649
20	0,080	0,078	0,686	0,787	0,196	0,057	0,553	0,652
21	0,080	0,077	0,677	0,773	0,194	0,052	0,525	0,647
22	0,079	0,075	0,666	0,753	0,192	0,047	0,496	0,645
23	0,079	0,073	0,657	0,735	0,188	0,041	0,464	0,640
24	0,079	0,072	0,650	0,719	0,183	0,036	0,432	0,631
25	0,078	0,070	0,641	0,698	0,180	0,032	0,400	0,628
26	0,079	0,069	0,634	0,679	0,177	0,028	0,372	0,627
27	0,079	0,067	0,629	0,660	0,174	0,025	0,342	0,624
28	0,079	0,065	0,621	0,635	0,172	0,022	0,313	0,628
29	0,079	0,064	0,615	0,614	0,171	0,019	0,286	0,628
30	0,080	0,062	0,611	0,595	0,169	0,017	0,258	0,623
31	0,080	0,060	0,605	0,568	0,169	0,015	0,225	0,620
32	0,081	0,058	0,603	0,549	0,168	0,013	0,198	0,612
33	0,081	0,056	0,601	0,529	0,169	0,013	0,173	0,605
34	0,083	0,056	0,602	0,519	0,172	0,013	0,161	0,608
35	0,085	0,057	0,603	0,511	0,177	0,016	0,155	0,621
36	0,087	0,057	0,606	0,504	0,185	0,020	0,148	0,632

Note: this table reports the root mean square forecast error (RMSFE), the Theil criterion (Theil), the bias (bias) and the variance (Var) components of the Theil mean square forecast error decomposition for thirty six forecasting horizons (*h*) based on the forecasting model

$$y_t = \alpha + \sum_{k=1}^{P} \beta_k y_{t-k} + \sum_{i=1}^{4} \sum_{k=1}^{p} \gamma_{ik} x_{it-k} + \varepsilon_t$$

where $y_t = IP_t$, and x_{it} are the predictors, i.e., $x_{it} = V_t, IN_t, SI_t, M2_t$. IP_t, V_t, IN_t, SI_t and $M2_t$ stand for industrial production growth rates, volatility of stock returns, short-term nominal interest rates, inflation, and M2 growth rates respectively. The iterated approach is used to construct the *h*-step ahead forecasts of monthly industrial production growth rates.

Table 7. Measuring the forecasting ability of various monthly indicators in Japan

h	RMSFE		Theil		Bias		Var	
	model	AR	model	AR	model	AR	model	AR
1	0,034	0,030	0,608	0,524	0,003	0,003	0,751	0,778
2	0,047	0,047	0,612	0,565	0,005	0,005	0,531	0,325
3	0,064	0,067	0,698	0,636	0,002	0,004	0,351	0,133
4	0,076	0,082	0,729	0,653	0,002	0,005	0,286	0,059
5	0,086	0,097	0,748	0,673	0,002	0,008	0,268	0,028
6	0,096	0,111	0,760	0,682	0,003	0,010	0,259	0,010
7	0,104	0,126	0,774	0,697	0,004	0,013	0,250	0,002
8	0,111	0,141	0,785	0,712	0,005	0,015	0,240	0,000
9	0,119	0,155	0,798	0,720	0,004	0,015	0,234	0,003
10	0,125	0,169	0,809	0,729	0,003	0,015	0,229	0,009
11	0,130	0,181	0,811	0,731	0,002	0,015	0,225	0,019
12	0,134	0,193	0,815	0,737	0,001	0,014	0,217	0,033
13	0,136	0,203	0,816	0,738	0,000	0,013	0,207	0,052
14	0,135	0,212	0,808	0,739	0,000	0,011	0,198	0,077
15	0,134	0,222	0,799	0,744	0,001	0,009	0,183	0,108
16	0,131	0,231	0,801	0,754	0,005	0,006	0,152	0,152
17	0,132	0,242	0,802	0,763	0,008	0,005	0,138	0,183
18	0,133	0,254	0,802	0,770	0,008	0,006	0,131	0,206
19	0,133	0,266	0,797	0,776	0,007	0,007	0,126	0,229
20	0,133	0,278	0,791	0,782	0,005	0,009	0,119	0,255
21	0,131	0,289	0,780	0,788	0,004	0,011	0,113	0,281
22	0,130	0,300	0,769	0,791	0,002	0,014	0,105	0,311
23	0,128	0,310	0,758	0,796	0,001	0,016	0,096	0,339
24	0,126	0,323	0,749	0,805	0,000	0,019	0,084	0,366
25	0,122	0,336	0,733	0,815	0,000	0,021	0,072	0,393
26	0,118	0,346	0,712	0,819	0,000	0,024	0,062	0,424
27	0,116	0,357	0,699	0,825	0,002	0,026	0,050	0,454
28	0,114	0,370	0,696	0,835	0,004	0,028	0,035	0,481
29	0,112	0,383	0,690	0,846	0,006	0,031	0,021	0,509
30	0,109	0,395	0,684	0,856	0,011	0,033	0,009	0,541
31	0,105	0,409	0,672	0,869	0,017	0,035	0,001	0,576
32	0,099	0,422	0,661	0,885	0,027	0,038	0,005	0,619
33	0,092	0,436	0,637	0,907	0,044	0,040	0,043	0,663
34	0,085	0,448	0,603	0,914	0,060	0,041	0,099	0,701
35	0,081	0,461	0,574	0,918	0,072	0,042	0,148	0,721
36	0,078	0,472	0,551	0,917	0,076	0,042	0,177	0,735

Note: this table reports the root mean square forecast error (RMSFE), the Theil criterion (Theil), the bias (bias) and the variance (Var) components of the Theil mean square forecast error decomposition for thirty six forecasting horizons (*h*) based on the forecasting model

$$y_t = \alpha + \sum_{k=1}^{P} \beta_k y_{t-k} + \sum_{i=1}^{4} \sum_{k=1}^{p} \gamma_{ik} x_{it-k} + \varepsilon_t$$

where $y_t = IP_t$, and x_{it} are the predictors, i.e., $x_{it} = V_t, IN_t, SI_t, M2_t$. IP_t, V_t, IN_t, SI_t and $M2_t$ stand for industrial production growth rates, volatility of stock returns, short-term nominal interest rates, inflation, and M2 growth rates respectively. The iterated approach is used to construct the *h*-step ahead forecasts of monthly industrial production growth rates.

Table 8. Measuring the forecasting ability of various monthly indicators in Italy

h	RMSFE		Theil		Bias		Var	
	model	AR	model	AR	model	AR	model	AR
1	0,027	0,025	0,923	0,849	0,032	0,020	0,335	0,352
2	0,030	0,024	0,835	0,629	0,054	0,044	0,547	0,625
3	0,037	0,030	0,851	0,642	0,065	0,050	0,587	0,606
4	0,043	0,037	0,844	0,672	0,076	0,054	0,570	0,537
5	0,049	0,044	0,835	0,699	0,085	0,055	0,569	0,479
6	0,055	0,052	0,836	0,740	0,089	0,053	0,555	0,437
7	0,061	0,060	0,837	0,775	0,095	0,054	0,531	0,399
8	0,067	0,066	0,838	0,792	0,100	0,057	0,509	0,387
9	0,071	0,071	0,828	0,794	0,109	0,063	0,496	0,385
10	0,075	0,076	0,817	0,804	0,121	0,069	0,485	0,378
11	0,078	0,080	0,803	0,808	0,137	0,077	0,472	0,370
12	0,079	0,082	0,782	0,807	0,160	0,090	0,463	0,364
13	0,080	0,084	0,763	0,804	0,188	0,106	0,450	0,359
14	0,080	0,086	0,738	0,798	0,223	0,124	0,442	0,352
15	0,080	0,088	0,710	0,797	0,268	0,144	0,435	0,340
16	0,079	0,089	0,683	0,791	0,318	0,167	0,429	0,334
17	0,079	0,091	0,660	0,794	0,351	0,174	0,429	0,328
18	0,080	0,094	0,649	0,801	0,356	0,170	0,417	0,321
19	0,080	0,095	0,636	0,804	0,361	0,165	0,399	0,312
20	0,078	0,094	0,621	0,801	0,362	0,158	0,373	0,300
21	0,078	0,093	0,617	0,797	0,355	0,153	0,329	0,285
22	0,077	0,091	0,609	0,792	0,351	0,149	0,291	0,272
23	0,076	0,090	0,603	0,790	0,343	0,142	0,251	0,254
24	0,075	0,088	0,599	0,785	0,335	0,137	0,210	0,237
25	0,075	0,086	0,602	0,785	0,321	0,131	0,159	0,210
26	0,074	0,084	0,605	0,781	0,307	0,126	0,116	0,187
27	0,074	0,081	0,610	0,773	0,296	0,125	0,072	0,157
28	0,074	0,078	0,622	0,772	0,278	0,120	0,039	0,128
29	0,073	0,074	0,634	0,763	0,265	0,119	0,011	0,091
30	0,074	0,070	0,653	0,759	0,247	0,117	0,000	0,060
31	0,074	0,066	0,683	0,754	0,226	0,116	0,007	0,027
32	0,075	0,063	0,723	0,770	0,206	0,112	0,036	0,004
33	0,077	0,061	0,771	0,791	0,185	0,107	0,081	0,002
34	0,080	0,059	0,821	0,805	0,167	0,106	0,124	0,016
35	0,083	0,059	0,845	0,808	0,157	0,106	0,144	0,027
36	0,086	0,057	0,871	0,791	0,150	0,113	0,164	0,043

Note: this table reports the root mean square forecast error (RMSFE), the Theil criterion (Theil), the bias (bias) and the variance (Var) components of the Theil mean square forecast error decomposition for thirty six forecasting horizons (*h*) based on the forecasting model

$$y_t = \alpha + \sum_{k=1}^{P} \beta_k y_{t-k} + \sum_{i=1}^{4} \sum_{k=1}^{p} \gamma_{ik} x_{it-k} + \varepsilon_t$$

where $y_t = IP_t$, and x_{it} are the predictors, i.e., $x_{it} = V_t, IN_t, SI_t, M2_t$. IP_t, V_t, IN_t, SI_t and $M2_t$ stand for industrial production growth rates, volatility of stock returns, short-term nominal interest rates, inflation, and M2 growth rates respectively. The iterated approach is used to construct the *h*-step ahead forecasts of monthly industrial production growth rates.

Table 9. Forecasting one-month ahead industrial production growth rates using information from the in-sample significant causality relations

	Single indicator models											
	Volatility				Benchmark				MPI			
	rmsfe	Theil	bias	Var	rmsfe	Theil	bias	Var	rmsfe	Theil	bias	Var
Simple model												
US												
0.5T	0.0069	0.6972	0.0227	0.7826	0.0065	0.6293	0.000	0.3498	0.0071	0.6790	0.0435	0.6012
0.7T	0.0078	0.7637	0.1059	0.8200	0.0070	0.6520	0.0100	0.3634	0.0080	0.7416	0.1415	0.6745
Germany												
0.5T	0.0152	0.9319	0.0000	0.9110	0.0151	0.6944	0.0004	0.3264	0.0152	0.9275	0.0001	0.9483
0.7T	0.0162	0.9407	0.0005	0.9723	0.0177	0.7627	0.0001	0.2781	0.0162	0.9298	0.0001	0.9843
Japan												
0.5T	0.0252	0.9249	0.0013	0.8143	0.0268	0.6591	0.000	0.1146	0.0252	0.9213	0.000	0.7948
0.7T	0.0301	0.9363	0.0022	0.8863	0.0336	0.7142	0.0005	0.1516	0.0302	0.9646	0.0002	0.9417
Italy												
0.5T	0.0153	0.9532	0.000	0.8830	0.0151	0.7022	0.0120	0.3425	0.0152	0.9287	0.0130	0.9059
0.7T	0.0180	0.9511	0.0043	0.8924	0.0183	0.7146	0.0260	0.3226	0.0181	0.9468	0.0188	0.9341
Pooling of forecasts from estimation windows of different lengths												
US												
0.5T	0.0068	0.6711	0.0305	0.6669	0.0069	0.9864	0.0608	0.9322	0.0076	0.6505	0.0856	0.3008
0.7T	0.0079	0.7528	0.1113	0.6410	0.0074	0.9942	0.0000	1.0021	0.0086	0.7061	0.2212	0.3923
Germany												
0.5T	0.0152	0.9026	0.0000	0.8518	0.0152	0.9928	0.0040	0.9887	0.0156	0.8878	0.0000	0.6881
0.7T	0.0162	0.9104	0.0001	0.9030	0.0162	0.9952	0.0062	0.9937	0.0164	0.9406	0.0013	0.8653
Japan												
0.5T	0.0261	0.8396	0.0033	0.5251	0.0247	0.9654	0.0000	0.9692	0.0264	0.8424	0.0005	0.4963
0.7T	0.0309	0.8698	0.0012	0.6378	0.0298	0.9747	0.0002	0.9827	0.0312	0.8680	0.0024	0.6222
Italy												
0.5T	0.0154	0.8938	0.0004	0.7504	0.0151	0.9951	0.0020	0.9954	0.0152	0.9205	0.0064	0.8494
0.7T	0.0181	0.9079	0.0004	0.8178	0.0179	0.9978	0.0068	0.9985	0.0180	0.9532	0.0092	0.9021
Bootstrap model												
US												
0.5T	0.0068	0.7382	0.0037	0.8647	0.0070	0.9154	0.0626	0.7371	0.0070	0.6962	0.0307	0.7085
0.7T	0.0076	0.8046	0.0582	0.8974	0.0074	0.9993	0.0009	1.0062	0.0079	0.7606	0.1146	0.7957
Germany												
0.5T	0.0152	0.9236	0.0000	0.9427	0.0152	0.9997	0.0041	0.9999	0.0153	0.9149	0.0005	0.8965
0.7T	0.0162	0.9248	0.0000	0.9656	0.0162	0.9998	0.0064	1.0007	0.0162	0.9229	0.0000	0.9586
Japan												
0.5T	0.0250	0.9686	0.0001	0.9331	0.0249	0.9998	0.0000	1.0062	0.0251	0.9585	0.0000	0.8918
0.7T	0.0301	0.9761	0.0008	0.9611	0.0300	0.9999	0.0003	1.0106	0.0301	0.9773	0.0001	0.9515
Italy												
0.5T	0.0152	0.9380	0.0091	0.9013	0.0151	0.9997	0.0020	1.0028	0.0152	0.9325	0.0104	0.8894
0.7T	0.0181	0.9560	0.0144	0.9158	0.0179	0.9999	0.0068	1.0017	0.0180	0.9514	0.0154	0.9156

Notes: Single indicator models are estimated with stock return volatility (denoted as volatility), and monetary policy indicators (denoted as MPI). Money supply growth rates are used as MPI indicator for US, Germany and Japan, while short-term interest rates for Italy. The benchmark model is an AR of IP growth rates where lag order is selected based on the AIC criterion. Lag lengths used in the specification of the single indicator models are presented in (24):(27). Estimation windows of lengths 0.5T and 0.7T are used (T is the length of the full sample). In the first panel, forecasts are based on a single estimation window. In the second panel, forecasts from estimation windows of different lengths are averaged. In the third panel, the models are estimated using a large simulation technique.

Chapter 5

Summary and Conclusions

Nowadays, Granger causality tests are standard tools to investigate causal relationships between financial and economic time series. Econometric advances in the field have shown that the causal relationship between two variables is not invariant to the integration and cointegration properties of the processes nor the relevant information that is available and included in the analysis. Hence, various notions of Granger non-causality are developed in the context of linear bivariate or multivariate stationary or nonstationary discrete time processes. Several of these causality concepts are reviewed in this thesis. Their extended concept is contrasted to the standard Granger causality concept. A wide range of causality tests have been used to investigate the independence between the second moments of the time series. There is currently much interest in testing causality-in-variance by policy makers, portfolio managers, and academic researchers.

This thesis consists of three chapters. Chapter 1 is directed towards testing Granger non-causality-in-variance. The size and power properties of general Granger causality-in-variance test procedures are evaluated by means of extensive Monte Carlo simulations. In particular, we focus our attention on four testing procedures: the Likelihood Ratio (LR) tests in the framework of a GARCH-BEKK(1,1) model as employed by Comte and Lieberman (2000); Cheung and Ng's sample cross-correlation (hereafter denoted as CCF) based S test (1996); the semiparametric CCF Q tests proposed by Hong (2001); and the Lagrange Multiplier (LM) test of Hafner and Herwartz (2006).

The finite sample size of these tests is evaluated under alternative models regarding the degrees of persistence of the volatility processes. The simulation of power is set up in a way which allows to analyze the performance of these tests for bivariate processes where causality is present at distant time periods and others where causality is present at short time periods.

Our results show that Comte and Lieberman's LR as well as and Hafner and Herwartz's LM tests suffer from severe size distortions, while they demonstrate very low power, under long horizon causality alternatives. Both cross correlation tests are reasonably well sized. However, Hong's Q test demonstrates less sensitivity to arbitrary choices of the weighting scheme and alternative volatility dynamics, when compared to Cheung and Ng's S test. Furthermore, cross correlations tests are favorably compared to LR and LM tests in terms of empirical power under a sequence of local alternatives.

Therefore, the finite-sample performance of the kernel Q tests is proved to be clearly superior to the other methods currently popular in the literature. Moreover, the kernel CCF approach is convenient in practice because it only requires estimating univariate parametric specifications for each separate time series. This approach applies a weighted scheme on the sample cross-correlations. However, efficient implementation of the method requires the proper selection of the amount of local averaging imposed to the cross-correlations. For a kernel based test procedure, this is controlled by the parameter denoted as the bandwidth. To date, how best to choose the bandwidth parameter used in the calculations of the kernel based CCF tests remains unclear.

Our simulation results demonstrate that the choice of (non-uniform) kernel function has no impact on the finite sample properties of the kernel type tests. On the other hand, arbitrary selections of the bandwidth parameter have a significant effect on the power properties of the tests in finite samples. In particular, we find that the horizon of causality determines how power depends on the choice of bandwidth parameter. Choosing a small bandwidth may result in inferential biases since possible long horizon causalities will be ignored, while selecting a large bandwidth will come at cost of loosing empirical power under the presence of a short-horizon causality. Therefore, it is very useful for the researcher to have a data-driven bandwidth selector that estimates the appropriate amount of smoothing implemented on the sample cross-correlations.

Motivated by these findings, we introduce three simple methods for automatic bandwidth selection used in Hong's *Q* test calculations. The first approach detects optimal bandwidths those that maximize the power function of the Q tests. The second approach estimates optimal bandwidth as a change-point in the distributional dynamics of the Q tests. The third estimator selects the bandwidth which minimizes an estimate of the integrated squared error in the context of a kernel regression. A simulation study illustrates the gains from using these three bandwidth selection procedures.

We apply the test procedures to the classic question of whether changes in the aggregate stock return volatility anticipate changes of industrial production growth volatility in four economies, namely US, UK, Japan and Italy. Previous emprical evidence is inconclusive. Implementation of the LM, LR as well as the S and Q tests using an arbitrary lag and bandwidth selection respectively, yield mixed results. On the other hand, under optimal bandwidth selection the Q tests show that there is unidirectional causality from stock return volatility to industrial production growth volatility in US, UK and Japan at level 5%. Moreover, the large bandwidth estimates indicate the presence of long horizon causality relations. We find limited evidence of causality from output growth volatility to stock return volatility.

A simple efficient test procedure for second-order causation that does not impose an explicit functional form on the evolution of the second order dynamics has yet to be established. Chapter 2 adresses the issue of testing Granger non-causality-in-variance in an unconditional context. A modified general concept of Granger second-order non-causality for a bivariate covariance stationary process is introduced, which in contrast to the standard causality testing frameworks does not require the estimation of a parametric specification for the second moment dynamics of the processes.

Unidirectional second-order noncausality between two variables is characterized in terms of the cross-correlation function between the absolute values of innovations obtained from ARMA processes. Therefore, a simple cross-correlation based test is all that is required to test non-causality between two stationary time series at a fixed lag order. In particular, to investigate the relationship between the second moments of two stochastic processes under consideration, our method uses the sample cross-correlation at a specific lag between the absolute values of the residuals resulting from fitting univariate ARMA specifications to each separate series. Using some assumptions, under the null hypothesis that the cross- correlation at a particular lag *k* is zero, we show that the sample cross-cross-correlation at lag order *k* between these volatility proxies appropriately rescaled converges asymptotically to the standard normal distribution.

Our test of second-order non-causality can only be performed for a specific lag order at a time. An approach for performing joint hypothesis testing is the implementation of our test sequentially at multiple lag time periods. Bonferroni -type bounds are employed to control the overall test size. Bonferroni size adjustments are known to yield conservative test procedures. Nevertheless, the use of the adjusted Bonferroni inequality based comparison procedure developed by Rom (1990) ensures high finite sample power.

The finite sample size and power of our test relative to the standard causality tests previously reviewed are investigated by means of Monte Carlo simulations. It is shown that our test has excellent finite sample properties under a series of local alternatives. In particular, simulation evidence suggests that out test has practical performance second to none in the existing literature because its enhanced size and power properties are robust with respect to the implemented lag truncation while it holds for different sample sizes. An empirical illustration on the causal relationship between stock return volatility and output growth volatility highlights its practical importance.

Chapter 3 adresses the issue whether there is multiple long horizon causation from stock return volatility to output growth in terms of causal chains involving monetary policy indicators. In particular, we are interested in investigating for the precise prediction horizon at which fluctuations in the volatility of the agreggate stock returns foretell changes in real activity growth patterns within a framework of linear vector autoregressive discrete time processes, where monetary policy indicators enter the multivariate system as auxiliary variables.

Multiple horizon non-causality is tested by implementing the test procedure proposed by Dufour, Pelletier and Renault (2006) on data from four economies, namely US, Germany, Japan and Italy. Their test procedure requires the estimation by least squares of long horizon parametric autoregressions. Their approach is convenient because highly nonlinear restrictions on the coefficients of the vector autoregressions are imposed under the null. By estimating these parametric formulations, one has only to use a standard Wald test to evaluate simple linear zero coefficient restrictions on the parameters of these vector autoregressions. A Monte Carlo method to calculate the *p*-value of the Wald test ensures a satisfactory finite sample performance of the test procedure.

Our results reveal a large number of highly significant direct and indirect causalities from stock return volatility to output growth at both short and long horizons in all four economies; the latter occur through the nominal short-term interest rates in US, and Germany; the money supply growth rates in Japan; and the inflation in Italy. Our evidence also confirms earlier findings in the literature that monetary policy fares well at predicting changes of the real activity at both short and long term. The degree of forecast improvement that arise from each causality relation is also considered using the causality measure proposed by Dufour and Taamouti (2010). Causalities of significant size from stock return volatility to output growth are found in Germany, Japan and Italy, while in US evidence suggests significant indirect long horizon causation through the money supply growth rates. A pseudo out-of-sample forecasting evaluation also shows how conditioning on such information yields better output growth predictions at both short and long forecast periods.

Bibliography

Al-Sadoon, M. M., 2010. Causality Along Subspaces: Inference. Working paper.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Ebens, H. 2001. The Distribution of Realised Stock return Volatility. Journal of Financial Economics, Vol. 61, 43-76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P. 2001. The Distribution of Realized Exchange Rate Volatility. Journal of the American Statistical Association 96, 42-55.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P. 2003. Modeling and Forecasting Realized Volatility. Econometrica Vol. 71, 579-625.

Andreou, E., Ghysels, E., 2002. Detecting multiple breaks in Financial Market Volatility Dynamics. Journal of Applied Econometrics, Vol. 17, No. 5, 579-600.

Ang, A., Bekaert, G., 2007. Stock return predictability: Is it there? Review of Financial Studies 20, 651-707.

Assenmacher-Wesche, Katrin, and M. Hashem Pesaran, M. H., 2008. Forecasting the Swiss economy using VECX* models: An exercise in forecast combination across models and observation windows. National Institute Economic Review 203, 91-108.

Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66, 47-78.

Bai, J., Perron, P., 2003. Computation and analysis of multiple structural change models. Journal of Applied Econometrics 18, 1-22.

Baillie, R., Bollerslev, T., 1990. Intra-day and inter-market volatility in foreign exchange rates. Review of Economic Studies 58, 565–585.

Baillie, R., T., Bollerslev, T., Mikkelsen, H., O., 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 74, 3-30.

Barro, R., 1990. The stock market and investment, Review of financial Studies 3, 115-131.

Bauwens, L., Laurent, S., Rombouts, J. V. K., 2006. Multivariate GARCH models: a survey. Journal of Applied Econometrics 21, 79-109.

Beltratti, A., Morana, C., 2006. Breaks and persistency: macroeconomic causes of stock market volatility. Journal of Econometrics 131, 151-177.

Ben Hariz, S. Wylie, J.J., Zhang, Q., 2007. Optimal rate of convergence for nonparametric change-point estimators for nonstationary sequences Ann. Statist. 35, 1802-1826.

Berman, S. M., 1964. Limits theorems for the maximum term in stationary sequences. The Annals of Mathematical Statistics Vol. 35 N. 2, 502-516.

Bernanke, B., Gertler, M., 1999. Monetary policy and asset price volatility. Federal Reserve Bank of Kansas City Economic Review 84,17-51.

Billio, M., Pelizzon, L., 2003. Volatility and shocks spillover before and after EMU in European stock markets. Journal of Multinational Financial Management 13, 323–340.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics 31, 307–327.

Bollerslev, T., Mikkelsen, H.O., 1996. Modeling and pricing long memory in stock market volatility. Journal of Econometrics 73, 151–184.

Boudjellaba, H., Dufour, J. - M., Roy, R., 1992. Testing causality between two vectors in multivariate autoregressive moving average models. Journal of the American Statistical Association Vol. 67 Issue 420, 1082-1090.

Bowman, A., 1984. An alternative method of cross-validation for the smoothing of density estimates. Biometrika 71, 353-360.

Box, G. E. P., Pierce, D. A., 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. Journal of the American Statistical Association 65, 1509-1526.

Breidt, F.J., Crato, N., de Lima, P., 1998. The detection and estimation of long memory in stochastic volatility. Journal of Econometrics 83, 325–348.

Breitung, J., Candelon, B., 2006. Testing for short- and long-run causality: a frequency-domain approach. Journal of Econometrics 132, 363-378.

Bruneau, C., Jondeau, E., 1999. Long-run causality, with application to international links between long-term interest rates. Oxford Bulletin of Economics and Statistics Vol. 61 issue 4, 545-568.

Campbell, J. Y., Shiller, R. J., 1988. Interpreting Cointegrated Models. Journal of Economic Dynamics and Control Vol. 12, 505-522.

Campbell, J.,Y., Lettau, M., Malkiel, B., G., Xu, Y., 2001. Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. Journal of Finance 56, 1-43.

Campbell, J.Y., Shiller, R.J., 1988. Stock prices, earnings, and expected dividends. Journal of Finance 43, 661–676.

Caporale, G.M., Pittis, N., Spagnolo, N., 2002. Testing for causality-in-variance: An application to the East Asian markets. International Journal of Finance and Economics 7, 235–245.

Caporin, M., 2007. Variance (Non) Causality in Multivariate GARCH. Econometric Reviews 26, 1-24.

Carlstein, E., 1988. Nonparametric change-point estimation. Ann. Statist. 16. 188-197.

Cecchetti, S., Genberg, H., Lipsky, J., Wadhwan, S., 2000. Asset prices and central bank policy. International Center for Monetary and Banking Studies, Brookings Institution Press, London.

Chao, J., Corradi, V., Swanson, N.R., 2001. Out-of-sample tests for Granger causality. Macroeconomic Dynamics 5, 598-620.

Chen, A.S., Cheng, L.Y., Cheng, K.F 2009. Intrinsic bubbles and Granger causality in the S&P 500: Evidence from long-term data. Journal of Banking and Finance, 33, 2275-2281.

Chen, N., 1991. Financial investment opportunities and the macroeconomy. Journal of Finance 46, 529- 554.

Cheung, Y.M., Ng, L.K., 1996. A causality-in-variance test and its application to financial market prices. Journal of Econometrics 72, 33-48.

Cheung, Y.W., Ng, L.K., 1990. The dynamics of S&P 500 index and S&P 500 futures intraday price volatilities. Review of Futures Markets 9, 458–486.

Chow, K.V., Denning, K.C., 1993. A simple multiple variance ratio test. Journal of Econometrics 58, 385–401.

Clark, T. E., McCracken, M. W., 2005. The power of tests of predictive ability in the presence of structural breaks. Journal of Econometrics 124, 1-31.

Comte, F. , Lieberman, O., 2003. Asymptotic theory for multivariate GARCH processes. Journal of Multivariate Analysis 84, 61–84.

Comte, F., Lieberman, O.,2000. Second order noncausality in multivariate GARCH processes. Journal of Time Series Analysis 21, 535–557.

den Haan W.,J., Levin, A., 1997. A Practitioner's guide to robust covariance matrix estimation. Handbook of Statistics Vol.15, Rao CR, Maddala GS editions, 291-341.

Dickey, D.A., Fuller, W.A, 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74, 427-431.

Diebold, F. X., Yilmaz, K., 2008. Macroeconomic Volatility and Stock Market Volatility, World-Wide. PIER Working Paper No 08-031

Ding, Z. , Granger, C. W. J., Engle, R., F. 1993. A long memory property of stock market returns and a new model. Journal of Empirical Finance 1, 83-106.

Ding, Zhuanxin, Engle, R. F., 1994. Large scale conditional covariance matrix modelling, estimation and testing. University of California mimeograph, San Diego.

Dufour , J. M., Pelletier, D., Renault, E., 2006. Short run and long run causality in time series: inference. Journal of Econometrics 132, 337–362.

Dufour , J. M., Renault, E., 1998. Short run and long run causality in time series: theory. Econometrica 66, 1099-1125.

Dufour, J. M., Jouini, T., 2005. Finite-sample simulation-based inference in VAR models with applications to order selection and causality testing. Technical report.

Dufour, J. M., Taamouti, A., 2010. Short and long run causality measures: theory and inference. Journal of Econometrics 154, 42-58.

Dufour, J. M., Tessier, D., 1993. On the relationship between impulse response analysis, innovation accounting and Granger causality. Economics Letters 42, 327-333.

Dumbgen, L., 1991. The asymptotic behavior of some nonparametric change-point estimators. Ann. Statist. 19, 1471-1495.

Eibens, H., 1999. Realized stock volatility, mimeo. The Johns Hopkins University, Department of Economics.

Eichler, M., 2007. Granger causality and path diagrams for multivariate time series. Journal of Econometrics 137, 334-353.

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation, Econometrica 50, 987–1008.

Engle, R.F., Ito, T., Lin, W.L., 1990. Meteor shower or heat waves? heteroskedastic intra-daily volatility in the foreign exchange market. Econometrica 59, 524–542.

Engle, R.F., Kroner, K.F., 1995. Multivariate simultaneous generalized ARCH. Econometric Theory 11, 122–150.

Estrella, A. and Mishkin, F., 1996. Predicting US recessions: financial variables as leading indicators. Federal Reserve Bank of New York Research paper No. 9609.

Fama, E. , 1990. Stock returns, expected returns, and real activity. Journal of Finance 45, 1089 - 1108.

Fama, E. F., 1981. Stock returns, real activity, inflation, and money. American Economic Review 71, 545-565.

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

Faroque, A., Veloce, W., 1990. Causality and the structure of Canada's balance of payments: Evidence from time series. Empirical Economics 1.5, 267-283.

Filardo, A., 2004. Monetary policy and asset price bubble: calibrating the monetary policy trade-offs. BIS Working paper No. 155, Basel.

French, K. R., Schwert, G. W., Stambaugh, R. F., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3-29.

Gagnon, L., Karolyi, G. A., 2006. Price and volatility transmission across borders. Financial Markets, Institutions & Intruments 15, 107-158.

Geske, R., Roll, R., 1983. The monetary and fiscal linkage between stock returns and inflation. Journal of Finance 38, 7-33.

Geweke, J., 1982. Measurement of linear dependence and feedback between multiple time series. Journal of the American Statistical Association 77, 304-313.

Geweke, J., 1984b. Measures of conditional linear dependence and feedback between time series. Journal of the American Statistical Association 79, 907-915.

Geweke, J., Meese, R., Dent, W., 1983. Comparing alternative tests of causality in temporal systems: analytic results and experimental evidence. Journal of Econometrics Vol. 21 Issue 2, 161–194.

Geweke, J.. 1984a. Inference and causality in economic time series models, in: Z. Griliches and M.D. Intriligator eds., Handbook of econometrics, Vol. II (North-Holland. Amsterdam) 1102-1144.

Granger, C. W. J., 1969. Investigating causal relations by econometric models and cross spectral methods. Econometrica 37, 424–438.

Granger, C.W. J.,1969. Some recent developments in a concept of causality. Journal of Econometrics 39, 199-211.

Granger, C.W.J., 1980. Testing for causality: A personal view. Journal of Economic Dynamics and Control 2, 329–352.

Granger, C.W.J., Hyung, N., 1999. Occasional structural breaks and long memory, mimeo. University of California, San Diego, Department of Economics.

Granger, C.W.J., Robins, P.P., Engle, R.F., 1986. Wholesale and retail prices: bivariate time-series modeling with forecastable error variances. In: Belsley, D.A., Kuh, E. (Eds.),Model Reliability, MIT Press, Cambridge, pp. 1–17.

Hafner, C. M., Herwartz, H., 2004. Testing for Causality in Variance using Multivariate GARCH Models. Working paper .

Hafner, C. M., Herwartz, H., 2006. A Lagrange multiplier test for causality in variance. Economic Letters 93, 137-141.

Hamao, Y., Masulis, R., Ng, V., 1990. Correlations in Price Changes and Volatility Across International Stock Markets. Review of Financial Studies 3, 281–307.

Hannan, E. J., 1970. Multiple time series. New York: John Wiley & Sons Inc.

Harvey, C. R., 1989. Forecasts of economic growth from the bond and stock markets. Financial Analysts Journal 45, 38-45.

Harvey, C.R., Siddique, A., 1999. Autoregressive conditional skewness. Journal of Financial and Quantitative Analysis 34 , 465-487.

Haugh, L.D., 1976. Checking the independence of two covariance-stationary time series: A univariate residual correlation approach. Journal of the American Statistical Association 71, 378–385.

Henry, M., Payne, R., 1997. An investigation of long range dependence in intra-day foreign exchange rate volatility, LSE Financial Market Group, Discussion Paper n. 264.

Hill, J., 2007. Efficient tests of long-run causation in trivariate VAR processes with a rolling window study of the money-income relationship. Journal of Applied Econometrics 22, 747-765.

Hochberg, Y., 1988. A sharper Bonferroni procedure for multiple tests of significance. Biometrika 75, 800-802.

Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6, 65-70.

Hommel, G., 1988. A stagewise rejective multiple test procedure based on a modified Bonferroni test. Biometrika 75, 383-386.

Hong, Y., 2001. A test for volatility spillover with application to exchange rates. Journal of Econometrics 103, 183–224.

Hong, Y., Liu, Y., Wang, S., 2009. Granger causality in risk and detection of extreme risk spillover between financial markets. Journal of Econometrics 150, 271-287.

Hoover, K. D. 2001. Causality in Macroeconomics. Cambridge: Cambridge University Press.

Hosoya, Y., 1991. The decomposition and measurement of the interdependence between second-order stationary processes. Probability Theory and Related Fields 88, 429–444.

Hsiao, C., 1982. Autoregressive modeling and causal ordering of economic variables. Journal of Economic Dynamics and Control 4, 243-259.

Hu , Mei-Yuan Chen, R. C.W. Fok, Huang,1997. Causality in volatility and volatility spillover effects between US, Japan and four equity markets in the South China Growth Triangular. International Financial markets institutions & money 7, 351-367.

James, C., Koreisha, S., Partch, M., 1985. A VARMA analysis of the causal relations among stock returns, real output, and nominal interest rates. Journal of Finance 40, 1375-1384.

Johansen, S., 1988. Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control 12, 231-254.

Kaul, G., 1987. Stock returns and inflation: The role of the monetary sector. Journal of Financial Economics 18, 253-276.

Kilian, L., 1998. Small-sample confidence intervals for impulse response functions. Review of Economics and Statistics 80, 218-230.

Kim, S. W., Rogers, J. H., 1995. International stock price spillovers and market liberalization: Evidence from Korea, Japan, and the United States. Journal of Empirical Finance 2, 117-133.

Kleibergen, F., Paap, R., 2006. Generalized reduced rank tests using the singular value decomposition. Journal of Econometrics 133, 97-126.

Kokoszka, P., Leipus, R., 1998. Change-point in the mean of dependent observations. Statistics and Probability Letters 40, 385-393.

Kokoszka, P., Leipus, R.,2000. Change-point estimation in ARCH models. Bernoulli 6, 1-28.

Koutmos, G., Booth, J., 1995. Asymmetric Volatility Transmission in International Markets. Journal of International Money and Finance 14,747–762.

Kurozumi, E. , 2005. The rank of a sub-matrix of cointegration. Econometric Theory 21, 299–325.

Kyereme, S.S., 1991. Exchange rate, price, and output inter-relationships in Ghana: Evidence from vector autoregressions. Applied Economics 23, 1801-1810.

Lee, B., 1992. Causal relations among stock returns, interest rates, real activity, and inflation. Journal of Finance 47, 1591- 1603.

Lemmens , A., Croux ,C., Dekimpe, M. G., 2008. Measuring and testing Granger causality over the spectrum: an application to European production expectation surveys. International Journal of Forecasting 24, 414-431.

Li, W. K., McLeod, 1980. Distribution of the residual autocorrelations in multivariate ARMA time series models. Journal of Royal Statistical Association Ser. B. 43, 231 – 239.

Lin, W.L., Engle, R.F., Ito, T., 1994. Do bulls and bears move across borders? International transmission of stock returns and volatility. Review of Financial Studies 7, 507–538.

Lundbergh, S., Teräsvirta, T.,2002. Evaluating GARCH models. Journal of Econometrics 110 , 417–435.

Lütkepohl, H., 1993. Introduction to multiple time series analysis. Springer- Verlag, second edition, Berlin.

Lütkepohl, H., Burda, M. M., 1997. Modified Wald tests under nonregular conditions. Journal of Econometrics 78, 315-332.

Lütkepohl, H., Müller, M. M., 1994. Testing for multi-step causality in time series. Discussion Paper.

Mandelker, G., Tandon, K., 1985. Common stock returns, real activity, money, and inflation: some international evidence. Journal of International Money and Finance 4, 267-286.

Marcellino, M., Stock, J.H., Watson, M. W., 2003. Macroeconomic forecasting in the Euro area: country specific versus area-wide information. European Economic Review 47, 1-18.

Marcellino, M., Stock, J.H., Watson, M. W., 2006. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. Journal of Econometrics 135, 499–526.

McLeod, A., I., Li, W., K., 1983. Diagnostic checking ARMA time series models using squared-residual autocorrelations. Journal of Time Series Analysis 4 Vol. 4, 269-273.

Merton , R. C., 1973. An Intertemporal Capital Asset Pricing Model. Econometrica 41, 867-887.

McMillin, W.D., 1988. Money growth volatility and the macroeconomy. Journal of Money, Credit and Banking 20, 319-335.

Mikosch, T., Starica, C., 1998. Change of structure in financial time series, long range dependence and the GARCH model, mimeo. University of Groningen, Department of Mathematics.

Miller, R.G., 1966. Simultaneous statistical inference. McGraw-Hill, New York, NY.

Morana, C., 2009. On the macroeconomic causes of exchange rate volatility. International Journal of Forecasting 25, 328-350.

Morana, C., Beltratti, A., 2004. Structural change and long-range dependence in volatility of exchange rates: either, neither or both? Journal of Empirical Finance 11, 629-658.

Morgenstern, O., 1959. International Financial Transactions and the Business Cycle. Princeton University Press, Princeton.

Mosconi, R., Giannini, C., 1992. Non-causality in cointegrated systems: representation estimation and testing. Oxford Bulletin of Economics and Statistics 54, 399-417.

Newey, W. K., West, K. D., 1987. Hypothesis testing with efficient method of moments estimators. International Economic Review 28, 777-787.

Pantelidis, T., Pittis, N., 2004. Testing for Granger causality in variance in the presence of causality in mean. Economic Letters 85, 201-207.

Pesaran, M. H. , Schuermann, T., Smith, L. V., 2009. Forecasting economic and financial variables with global VARs. International Journal of Forecasting 25, 642-675.

Pesaran, M. H., Pick, A., 2011. Forecast combination across estimation windows. Journal of Business & Economic Statistics 29, 307-318.

Pesaran, M. H., Timmermann, 2007. Selection of estimation window in the presence of breaks. Journal of Econometrics 137, 134-161.

Phillips, P.C., Perron, P., 1988. Testing for a unit root in time series regression. Biometrika 75, 335-346.

Pierce, D. A., 1979. R-squared measures for time series. Journal of the American Statistical Association, 74, 901–910.

Pierce, D.A., Haugh, L.D., 1977. Causality in temporal systems. Journal of Econometrics 5, 265-293.

Robin, J. - M., Smith, R. J., 2000. Tests of rank. Econometric Theory 16, 151-175.

Rodrigues, P. M. M., Rubia, A., 2007. Testing for causality in variance under nonstationarity in variance. Economics Letters 97, 133-137.

Rom, D. M., 1990. A sequentially rejective test procedure based on a modified Bonferroni inequality. Biometrika 77, 663-665.

Rudemo, M., 1982. Empirical choice of histograms and kernel density estimators. Scandinavian Journal of Statistics 9, 65-78.

Ryden, T., Terasvirta, T., Asbrink, S., 1998. Stylized facts of daily return series and the hidden Markov model. Journal of Applied Econometrics 13, 217–244.

Schrimpf, A., Wang, Q., 2010 .A reappraisal of the leading indicator properties of the yield curve under structural instability.' International Journal of Forecasting 26, 836–857.

Schwert, G. W., 1989. Why does Stock Market Volatility Change Over Time? Journal of Finance 44, 1115-1153.

Schwert, G. W., 1990. Stock returns and real activity: a century of evidence. Journal of Finance 45, 1237-1257.

Shiller, R., 1989. Market Volatility, Cambridge, MIT Press.

Sims, C. A., Stock, J. H., Watson, M. W., 1990. Inference in linear time series models with some unit roots. Econometrica 58 (1), 13-144.

Sims, C.A., 1972. Money, income and causality. American Economic Review 62, 540-552.

Sims, C.A., 1980. Comparison of interwar and postwar business cycles: monetarism reconsidered. American Economic Review 70, 250-257.

Sims, C.A., 1980. Macroeconomics and reality. Econometrica 48, 1-48.

Sims, C.A., 1982. Policy analysis with econometric models. Brooking Papers on Economic Activity 1, 107-164.

Stam, A., DeLorme, C. D., Finkenstadt, B., 1991. Cross national money-income causality for the floating exchange rate period. Journal of Macroeconomics 13, 207-237.

Stock, J.H., Watson, M. W., 2003. Forecasting output and inflation: the role of asset prices. Journal of Economic Literature 41, 788-829.

Stock, J.H., Watson, M. W., 2004. Combination Forecasts of output growth in a seven-country data set. Journal of Economic Literature 23, 405-430.

Taamouti, A., Bouezmarni, T., El Ghouch, A., 2013. Nonparametric estimation and inference for Granger causality measures. Working paper.

Taylor, S., 1986. Modelling financial time series. New York, John Wiley & Sons.

Tegene, A., 1991. Trade flows, relative prices, and effective exchange rates: A VAR on Ethiopian data. Applied Economics 23, 1369-1376.

Tiao, G. C., Box, G. E. P., 1981. Modeling multiple time series with applications. Journal of the American Statistical Association 76, 802-816.

Timmermann, A., 2006. Forecast combinations. In Granger, C., elliot, G., and Timmermann, A., editors. Handbook of Economic Forecasting. North-Holland, Amsterdam, 135–196.

Tjostheim, D., 1981. Granger-Causality in multiple time series. Journal of Econometrics 17, 157-176.

Toda, H.Y., Phillips, P.C.B., 1993. Vector autoregressions and causality. Econometrica 61, 1367-1393.

Van Dijk, D., Osborn, D. R., Sensier, M.,2005. Testing for causality in variance in the presence of breaks. Economics Letters 89, 193–199.

Vilasuso, J., 2001. Causality tests and conditional heteroscedasticity: Monte Carlo evidence. Journal of Econometrics 101, 25-35.

White, H., 1984. Asymptotic theory for econometricians. Academic Press.

Whitelaw, R. F., 1994. Time variations and covariations in the expectation and volatility of stock market returns. Journal of Finance 49 (2), 515-541.

Wiener, N., 1956. The theory of prediction. Beckenback, E. F. (ed.). The Theory of Prediction. McGraw-Hill, New York (Chapter 8).

Wong, H., Li, W., K., 1996. Distribution of the Cross-Correlations of Squared Residuals in ARIMA Models. The Canadian Journal of Statistics 24, 489- 502.

Yamada, H., Toda, H.Y., 1998. Inference in possibly integrated vector autoregressive models: some finite sample evidence. Journal of Econometrics 86, 55–95.

Yamamoto, T., Kurozumi, E., 2006. Tests for long-run Granger non-causality in cointegrated systems. Journal of Time Series Analysis 27, 703-723.

Yao, F., Hosoya, Y., 2000. Inference on one-way effect and evidence in Japanese macroeconomic data. Journal of Econometrics 98, 225−255.

Zivot, E. and Andrews, D. W. K., 1992. Further evidence on the Great Crash, the Oil-Price shock, and the unit-root hypothesis. Journal of Business & Economic Statistics 10, 251-267.