foresight

Generalization of information, Granger causality and forecasting

Tobias Rötheli,

Article information:

To cite this document:
Tobias Rötheli, (2017) "Generalization of information, Granger causality and forecasting", foresight, Vol. 19 Issue: 6, pp.604-614, https://doi.org/10.1108/FS-06-2017-0017
Permanent link to this document:
https://doi.org/10.1108/FS-06-2017-0017

Generalization of information, Granger causality and forecasting

Tobias Rötheli

Tobias Rötheli is based at the Department of Economics, Universität Erfurt Staatswissen-schaftliche Fakultät, Erfurt, Germany.

Abstract

Purpose – *This paper aims to analyze forecasting problems from the perspective of information extraction. Circumstances are studied under which the forecast of an economic variable from one domain (country, industry, market segment) should rely on information regarding the same type of variable from another domain even if the two variables are not causally linked. It is shown that Granger causality linking variables from different domains is the rule and should be exploited for forecasting.*

Design/methodology/approach – *This paper applies information economics, in particular the study of rational information extraction, to shed light on the debate on causality and forecasting.*

Findings – *It is shown that the rational generalization of information across domains can lead to effects that are hard to square with economic intuition but worth considering for forecasting. Information from one domain is shown to affect that from another domain if there is at least one common factor affecting both domains, which is not (or not yet) observed when a forecast has to be made. The analysis suggests the theoretical possibility that the direction of such effects across domains can be counter-intuitive. In time-series econometrics, such effects will show up in estimated coefficients with the "wrong" sign.*

Practical implications – *This study helps forecasters by indicating a wider set of variables relevant for prediction. The analysis offers a theoretical basis for using lagged values from the type of variable to be forecast but from another domain. For example, when forecasting the bond risk spread in one country, introducing in the time-series model the lagged value of the risk spread from another country is suggested. Two empirical examples illustrate this principle for specifying models for prediction. The application to risk spreads and inflation rates illustrates the principles of the approach suggested here which is widely applicable.*

Originality/value – *The present study builds on a probability theoretic analysis to inform the specification of time-series forecasting models.*

Keywords *Economic forecasting, Information extraction*

Paper type *Research paper*

1. Introduction

Should an investor or a regulator who wants to forecast the value of an economic variable (e.g. a quantity, a price or a risk premium) from one domain (country, industry, market segment) rely on information regarding the same type of variable from another domain even if the two variables are *not* causally linked? Since the key contributions by Granger (1969, 1980), many econometricians have taken the following position on this question: a stable correlation between two variables across time (i.e. Granger causality) serves as an operational concept of cause and effect and is sufficient for the purpose of forecasting[1]. The described approach stands in contrast to a tradition of structural modeling in econometrics that perceives the predictability of economic variables to be the result of the effects of agents' decisions and market forces that play out over time[2]. This tradition uses economic analysis in the specification of forecasting models. This latter perspective motivates forecasters to be aware of both the danger of overfitting and possible spurious correlations and induces them to check whether their specifications and estimation results are in line with economic theory. This restraint in the specification of models tends to make

Received 1 June 2017
Revised 8 August 2017
Accepted 18 August 2017

Detailed comments by an anonymous referee were very helpful. The author would like to thank participants of seminars at the Swiss National Bank, the Albert-Ludwigs University of Freiburg and the SIBR 2016 Conference in Hong Kong for comments.

researchers choose the explanatory variables to come primarily from the same domain (e.g., country or industry) as the variable to be forecast. This even applies to recent contributions such as Bora *et al.* (2016) and Diks and Wolski (2016). As will be shown in this article, such restraint may be mistaken when the focus is on prediction. We will apply information economics, in particular the study of rational information extraction, to make this point.

In particular, the notion of rational generalization of information becomes relevant here[3]. The analysis will show that the mechanism of generalization of information across domains can lead to effects that are hard to square with economic intuition but worth being considered for forecasting. We will detail that information from one domain affects that from another domain if there is at least one common factor affecting both domains, which is not (or not yet) observed when the forecast has to be made[4]. Data on a variable from another domain are thus used because these contain information on a common driving force. Examples of a common driving force could be the current level of general technology, the general level of risk aversion or the state of animal spirits of entrepreneurs. Such variables are, at the very least, difficult to estimate but, more importantly, they are impossible to measure in real time.

In keeping with our general perspective, we can readily also think of domain-specific driving variables that are difficult or impossible to assess in real time. Examples here would encompass variables like the level of domain-specific factor inputs or the level of industry-specific confidence. Note that the problem of unobservable or hidden variables gains in importance the more short-term the variables are to be forecast: forecasting annual data might not be plagued by the problem of unobserved variables. However, with monthly, daily or higher-frequency observations, the issue becomes increasingly important. For the practice of econometric analysis, the situation with unobservable variables indicates the following: instead of estimating structural equations, the forecaster relies on multivariate time-series analysis. We will detail how this practice can benefit from information economics[5].

In what follows, we will use the terms “output” or “outcome” for variables to be forecast. In the formal model outlined in the next section, we analyze cases where the common driving force affects variables in different domains in the *same* direction. For example, rising risk aversion will lead to withdrawals of funds from all risk-taking companies and industries and thus reduces output everywhere. This setup helps to highlight a puzzling form of information generalization: the fact that its effects can appear paradoxical. We speak of a paradoxical effect of information generalization if an increase in output in one industry *diminishes* the probability that output in the other industry will increase. The term paradoxical is warranted because the common driving force affects the two output series in the same direction. With respect to the econometric analysis of time series, paradoxical effects show up in estimated coefficients that appear to have the wrong sign. Our analysis suggests that such coefficient estimates should not automatically be discarded.

The paper is organized as follows: Section 2 outlines the model of two domains in which domain-specific factors and a general factor together stochastically determine domains’ outcomes. The model makes use of the theoretical apparatus of Hirshleifer and Riley (1992). Section 3 provides numerical illustrations, and Section 4 illustrates the resulting effects in a time-series regression framework. Section 5 offers two empirical illustrations, and Section 6 concludes the article.

2. The model

We study a situation with two domains. These could be two countries, industries or market segments. The central element of the model is the assumption that the outcome in any domain is affected by two types of factors at the time of forecast:[6] a general factor (denoted by *G* for general or global) and a domain-specific factor (denoted by *A* and *B*, one for each domain). *A*, *B* and *G* are all unobservable at the time of the forecast. In our setup, these factors can have only one of two values (1 or 0, i.e. high or low) and follow homogeneous Markov chains:

$$\begin{aligned} P(A_{t+1} &= A_t) = \rho \\ P(B_{t+1} &= B_t) = \lambda \\ P(G_{t+1} &= G_t) = \theta \end{aligned} \tag{1}$$

Hence, each factor remains on the level of the last period with a given probability and changes its value with one minus that probability. The combination of the domain-specific factor and the general factor probabilistically determines the outcome in a domain. Hence, there is *no* causality running from the outcome in one domain to the outcome in the other domain. Figure 1 shows the different combinations of factors and the implied probabilities for a high-level (denoted by $\bar{q}$) and a low-level (denoted by $\underline{q}$) outcome. To keep the problem tractable, the number of parameters used to model probabilities is limited to two (i.e. u and v). To give an example, with $A = 0$ and $G = 0$, the probability that q^A equals $\bar{q}$ is u. We assume that $v > u$ and $1 - v > u$ which implies $1 - u > v$ and $1 - v - u > 0$. This means that a switch of any factor from 0 to 1 increases the likelihood of a high-level outcome in a domain.

This setup helps us to clarify an important and intriguing aspect of the generalization of information: the fact that its effects can be normal, neutral or paradoxical. We will speak of a *normal effect* of the generalization of information if the rationally assessed value of a variable in one domain in the next period is positively affected by the current value of the variable in the other domain. In other words, the information that domain *A*'s outcome is high at present makes it more likely that domain *B*'s outcome will be high in the future. The case of a *neutral effect* of the generalization of information describes the case where the level of domain *A*'s outcome does not predict domain *B*'s outcome despite the fact that there is a common driving force affecting the two outcomes. We speak of a *paradoxical effect* if the level of *A*'s outcome negatively affects the expected level of *B*'s outcome. This means the rationally assessed likelihood of *B*'s outcome is lower if *A*'s outcome is high compared to the situation where the latter is low. The term paradoxical is warranted in this case because the common driving force affects the two outcomes in the same direction.

As indicated before, the generalization of information describes a situation in which information from one variable (i.e. the outcome in one domain) from which there is no causal force acting on another variable (i.e. the outcome in the other domain) is relevant for forecasting that other variable. This can be expressed in terms of conditional probabilities:

Figure 1 Probabilities of different outcomes in a domain *j*

$P(q_t^j = \bar{q})$		*G* = 0	*G* = 1
A or *B*	0	u	$1-v$
	1	v	$1-u$

$P(q_t^j = \underline{q})$		*G* = 0	*G* = 1
A or *B*	0	$1-u$	v
	1	$1-v$	u

$$P(q^B_{t+1} | q^B_t) \neq P(q^B_{t+1} | q^A_t \cap q^B_t) \quad (2)$$

The case of a normal effect of the generalization of information describes a situation where the probability of an improvement in a domain's level of outcome (i.e. from $\underline{q}$ to $\bar{q}$) is positively affected by the other domain's level of outcome.

$$P(q^B_{t+1} = \bar{q} | q^A_t = \bar{q} \cap q^B_t = \underline{q}) > P(q^B_{t+1} = \bar{q} | q^B_t = \underline{q}) \quad (3)$$

A paradoxical effect of the generalization of information is given when:

$$P(q^B_{t+1} = \bar{q} | q^A_t = \bar{q} \cap q^B_t = \underline{q}) < P(q^B_{t+1} = \bar{q} | q^B_t = \underline{q}) \quad (4)$$

[7]The intermediate situation is then the case of neutral effects of the generalization of information.

$$P(q^B_{t+1} = \bar{q} | q^A_t = \bar{q} \cap q^B_t = \underline{q}) = P(q^B_{t+1} = \bar{q} | q^B_t = \underline{q}) \quad (5)$$

To compare the two relevant conditional probabilities under a given parameterization, we first need to derive the probability of a change in the level of outcome from t to $t + 1$ in one domain conditional only on its currently realized outcome. To be more specific, we compute the probability that the outcome of domain B is high in the next period conditional on the information that the outcome is low in the current period, i.e. $P(q^B_{t+1} = \bar{q} | q^B_t = \underline{q})$. Given that the factors affecting outcome follow homogeneous Markov chains, all possible combinations of factors have, *a priori*, the same probability of occurring. Hence, if we ask with what probability the combination $q^B_t = \underline{q}$ arises from one specific combination of the Factors B and G, application of the Bayes theorem gives us the answer shown in Figure 2[8].

Several more steps have to be taken, and the derivation of the results is quite involved. First, we need to compute, for every possible combination of factors in one domain, the probabilities with which each of the four possible combinations of factors is realized in the next period. For example, the probability that any given B/G-combination is repeated is $\lambda\theta$. Alternatively, the probability that Factor B switches value and G remains unchanged is $(1 - \lambda)\theta$. Now the four probabilities $P(B_{t+1}, G_{t+1} | q^B_t = \underline{q})$, for the four different B/G-combinations, are the sum of the probabilities of each B/G-combination given $q^B_t = \underline{q}$ (as outlined above) weighted with the probabilities that these combinations lead to B_{t+1}, G_{t+1}. Here is an example:

$$P(\underline{B}_{t+1} \cap \underline{G}_{t+1} | q^B_t = \underline{q}) = \lambda\theta\frac{1-u}{2} + (1-\lambda)\theta\frac{1-v}{2} + \lambda(1-\theta)\frac{v}{2} + (1-\lambda)(1-\theta)\frac{u}{2} \quad (6)$$

In a last step, the probabilities thus computed have to be multiplied with the probabilities $P(q^B_t = \bar{q} | B_t, G_t)$ and added up in order to find $P(q^B_{t+1} = \bar{q} | q^B_t = \underline{q})$. After collecting and rearranging a large number of terms this probability can finally be expressed as:

$$P(q^B_{t+1} = \bar{q} | q^B_t = \underline{q}) = 0.5 - (\lambda - 0.5)(u - v)^2 - (\theta - 0.5)(1 - u - v)^2 \quad (7)$$

Figure 2 Probabilities with which $q^B_t = \underline{q}$ comes from specific combination of factors

		G: 0	G: 1
B	0	$(1-u)/2$	$v/2$
	1	$(1-v)/2$	$u/2$

A sufficient condition for $P(q^B_{t+1} = \bar{q} | q^B_t = \underline{q}) < 0.5$, i.e. a change in q^B is less likely than a no-change is $\lambda > 0.5$ and $\theta > 0.5$.

Next, we derive the conditional probability of an improvement in a domain's level of outcome given that the outcome in the other domain is already high. Following the same sort of calculations as indicated before (Appendix gives details), we find this conditional probability to be:

$$P(q^B_{t+1} = \bar{q} | q^A_t = \bar{q} \cap q^B_t = \underline{q}) = 0.5 - (\lambda - 0.5)\frac{(u - v)^2}{(u + v)(2 - u - v)} \quad (8)$$

Equation (8) suggests the following:

P1. The probability of the outcome of a domain is not affected by the persistence (i.e. θ) of the general factor.

The intuition behind this result is the following: different productivity levels in the two industries indicate a 50-50 chance that G is either high or low. In this situation, the level of persistence of the general factor is not relevant for forecasting. For a discussion of the key insights of this analysis, it is helpful to reformulate (8) to the form:

$$P(q^B_{t+1} = \bar{q} | q^A_t = \bar{q} \cap q^B_t = \underline{q}) = P(q^B_{t+1} = \bar{q} | q^B_t = \underline{q}) - \alpha(\lambda - 0.5) + \beta(\theta - 0.5) \quad (9)$$

with $\alpha = \left[(1 - u - v)^2(u - v)^2 / (u + v)(2 - u - v)\right] > 0$ and $\beta = (1 - u - v)^2 > 0$

Based on (9), the following proposition can be stated and proved:

P2. Even if domain factors and the general factor are positively autocorrelated, effects from the rational generalization of information of the normal, neutral or paradoxical form can result.

Proof: From (9), it follows that with $\lambda > 0.5$ and $\theta > 0.5$, the difference $P(q^B_{t+1} = \bar{q} | q^A_t = \bar{q} \cap q^B_t = \underline{q}) - P(q^B_{t+1} = \bar{q} | q^B_t = \underline{q})$ can be positive, negative or zero.

The intuition here is that the relative importance of the domain-specific factor and the general factor in the two domains (i.e. u and v) determines whether a high outcome in the other domain is attributed to a favorable domain factor there or rather to favorable general conditions. The more it is attributed to favorable general conditions – and the more persistent G is – the more likely a normal effect will result. In contrast, the more the outcome difference in the two domains is attributed to differing domain conditions – and the more persistent and the more important the domain effect of the domain under consideration – the more likely will be a paradoxical effect. In such cases the good times in the other domain strengthen the impression that the bad times in a given domain have a domain-specific cause and are liable to last. This insight can be clarified further by formally analyzing the condition under which the generalization of information has just a neutral effect. Intuition can be built by focusing on the special and polar case where $u = 0$. Based on equation (9), we find that the neutral effect emerges when:

$$v = \frac{2(\theta - 0.5)}{(\lambda - 0.5) + (\theta - 0.5)} \quad (10)$$

If v is above this critical level – i.e. the partial effect of the domain-specific factor is strong enough – then the paradoxical effect results. Figure 3 illustrates this relationship numerically. The table shows the critical level of v for different combinations of θ and λ. It is immediately apparent that the paradoxical effect is only possible if $\lambda > \theta$. Furthermore, the higher the value of θ, the stronger must be the domain-specific factor v to generate a paradoxical effect. The empty cells in the table indicate parameter-combinations for which the paradoxical effect can be excluded.

Figure 3 Critical values of v given combinations of θ- and λ-values

λ \ θ	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00
0.55										
0.60	0.666									
0.65	0.500	0.800								
0.70	0.400	0.666	0.857							
0.75	0.333	0.571	0.750	0.888						
0.80	0.285	0.500	0.666	0.800	0.909					
0.85	0.250	0.444	0.600	0.727	0.833	0.923				
0.90	0.222	0.400	0.545	0.666	0.769	0.857	0.933			
0.95	0.200	0.363	0.500	0.615	0.714	0.800	0.875	0.941		
1.00	0.181	0.333	0.461	0.571	0.666	0.750	0.823	0.888	0.947	

Note: If v is above the threshold value, the paradoxical direction of the generalization of information occurs

3. Stochastic simulations

To illustrate the generalization of information and its various effects, we stochastically simulate two interesting parameterizations of the model. This means that we draw random numbers of the latent variables A, B, G and then let a further round of random draws determine the outcomes for q^A and q^B according to the given probability parameters. For convenience concerning the interpretation of the regression results that follow, we take the two possible levels of the q-variables to be 0 and 1. We analyze two cases. In the first case, we have $\rho = 0.8$, $\lambda = 0.8$, $\theta = 0.8$, $u = 0.1, v = 0.3$. According to equation (9), the generalization of information in this case has the normal effect. The theoretically implied conditional probabilities are $P(q^B_{t+1} = \bar{q} | q^B_t = \underline{q})$=0.380 and $P(q^B_{t+1} = \bar{q} | q^A_t = \bar{q} \cap q^B_t = \underline{q}) = 0.481$. We stochastically draw 1,000,000 sets of observations for A, B and G and compute q^A and q^B values. With these observations, the frequency of a high level of q^B after a low level of q^B in the previous period turns out to be 0.380. Furthermore, the frequency of observing a high level of q^B after periods with both a low level of q^B and a high level of q^A is 0.481. Hence, looking at numbers rounded to the third digit, the simulated frequencies are equal to the analytically derived conditional probabilities.

The second case is the parameter combination $\rho = 0.95, \lambda = 0.95, \theta = 0.55, u = 0.0, v = 0.5$, i.e. highly persistent and influential domain-specific factors together with a weakly persistent general factor. In this case, the generalization of information has a paradoxical effect. The theoretically derived probability values are $P(q^B_{t+1} = \bar{q} | q^B_t = \underline{q}) = 0.375$ and $P(q^B_{t+1} = \bar{q} | q^A_t = \bar{q} \cap q^B_t = \underline{q}) = 0.350$, and the corresponding simulated frequencies (rounded to the third digit) are 0.375 and 0.350. That is, a high value of q^A today reduces the probability that q^B moves from a low value today to a high value tomorrow. Again, rounded to the third digit, the simulated results are equal to the analytically derived results.

4. Regression representations

In this section, we document the correspondence between the frequency format and the intuitively easier-to-grasp time-series regression format. Furthermore, the regression representation leads to the discussion of estimated coefficients that appear to have – compared to economic intuition – the wrong sign[9]. Consider regressing the q^B_t on a constant and its lagged value (i.e. q^B_{t-1})

$$q^B_t = \alpha + \beta q^B_{t-1} \tag{11}$$

The value of the coefficient α gives the probability $q^B_{t+1} = 1$ conditional on observing $q^B_t = 0$ based on this univariate regression. For the data simulated in Section 3 with one million draws of each variable, we find the following regression results for the two cases considered. For the normal case, we have:

$$\begin{aligned} q_t^B &= \underset{(0.006)}{0.380} + \underset{(0.009)}{0.240} q_{t-1}^B, \\ R^2 &= 0.057, \quad D.W. = 2.043, \end{aligned} \tag{12}$$

The estimated α-parameter matches the relevant frequency and the theoretically derived probability documented in Section 3. We now turn to including the information from the other domain. In terms of time-series econometrics, this implies estimating the multivariate regression:

$$q_t^B = \varphi + \mu q_{t-1}^B + \gamma q_{t-1}^A \tag{13}$$

Here, the estimate of the relevant conditional probability $P\,(q_{t+1}^B = \bar{q} | q_t^A = \bar{q} \cap q_t^B = \underline{q})$ is the coefficient sum $\varphi + \gamma$. Consider the outcome of the regression exercise:

$$\begin{aligned} q_t^B &= \underset{(0.000)}{0.332} + \underset{(0.010)}{0.186} q_{t-1}^B + \underset{(0.010)}{0.149} q_{t-1}^A, \\ R^2 &= 0.076, \; D.W. = 2.042 \end{aligned} \tag{14}$$

Here, the sign of the coefficient of q_{t-1}^A is positive, and the sum of coefficient values $\varphi + \gamma$ is 0.481. Again, this value coincides with the frequency reported in Section 3. Consider next the results for the outcome in the paradoxical case. Here the univariate result is:

$$\begin{aligned} q_t^B &= \underset{(0.000)}{0.375} + \underset{(0.001)}{0.249} q_{t-1}^B, \\ R^2 &= 0.062, \; D.W. = 2.075, \end{aligned} \tag{15}$$

and for the multivariate case we have:

$$\begin{aligned} q_t^B &= \underset{(0.000)}{0.389} + \underset{(0.001)}{0.259} q_{t-1}^B - \underset{(0.001)}{0.039} q_{t-1}^A, \\ R^2 &= 0.063, \; D.W. = 2.077 \end{aligned} \tag{16}$$

As to be expected for the paradoxical case in the estimate of equation (16), the γ-coefficient is negative. Note again the equality of the estimated α-coefficients (0.375) and coefficient sum ($0.389-0.039 = 0.350$) with the respective frequencies reported in Section 3. In both the normal and the paradoxical cases, the results document Granger causality between the two variables considered. Hence, taking into account information from another domain tends to improve the forecast[10].

5. Empirical illustrations

In the following, we offer two empirical applications for the effects of the generalization of information. The first application concerns financial data of relatively high-frequency. Concretely, we analyze daily observations of risk spreads of high-yield commercial bonds. While general empirical analyses of bond risk spreads have appeared (Reilly *et al.*, 2010), forecast studies are hard to find[11]. The variable studied is the spread (weighted across ratings) between yields on publicly traded commercial bonds of below investment grade (those rated BB or below) and yields on term-wise comparable government bonds. It measures what investors at any point in time receive for bearing risks (e.g. liquidity and default risks). The data are compiled by Merrill Lynch and are made available by the Federal Reserve Bank of St. Louis. For the analysis, we use data from January 1998 until the end of September 2016. The two economic domains studied here are Euro- and US dollar-denominated bonds. The corresponding variables are termed $S^€$ and $S^\$$. The estimated bivariate time-series model in first differences leads to the following results:[12]

$$\Delta S_t^{€} = -0.0005 + \underset{(0.001)}{} 0.409\, \Delta S_{t-1}^{€} + 0.088 \Delta S_{t-1}^{\$},$$
$$\quad (0.001) \qquad (0.047) \qquad (0.031)$$
$$R^2 = 0.125, \; D.W. = 2.076 \tag{17}$$
$$\Delta S_t^{\$} = 0.0003 + 0.247\, \Delta S_{t-1}^{\$} + 0.059 \Delta S_{t-1}^{€},$$
$$\quad (0.001) \qquad (0.036) \qquad (0.023)$$
$$R^2 = 0.089, \; D.W. = 2.023$$

All relevant coefficients are statistically significant at least at the 5 per cent level of significance[13]. The positive coefficients of the lagged variable from another domain document the effect of generalization of local information: there is no known economic mechanism through which the change in the risk spread in one country would induce a change in the risk spread of another country over time. Instead, a variation in one domain is informative and helps prediction because it is influenced by forces that drive risk spreads everywhere.

The next example concerns a variable where only monthly observations are available. Here, we endeavor to predict inflation rates in different countries. Forecasting studies on inflation rates abound in the literature. But yet again, even studies that consider foreign effects besides, e.g. measures of domestic capacity and inflation expectations, emphasize variables (like oil prices or a commodity price index), where an economically plausible causal link exists (Joshi and Acharya, 2011). We depart from this routine and pick the UK and Japan to illustrate the principles explored in this study. Again, there exists no plausible mechanism through which either British or Japanese inflation would significantly affect inflation in the other country through a causal process. However, it is plausible – in keeping with the analysis outlined – that a combination of country-specific and global forces drives national inflation rates around the globe. Thus, we would expect inflation in one country to help predict inflation in the other country. This is exactly what we find. Here are the results of estimating the bivariate system of monthly inflation rates for the sample 1955 through 2014:[14]

$$Inf_t^{UK} = 0.002 + 0.425\, Inf_{t-1}^{UK} + 0.127\, Inf_{t-1}^{JP},$$
$$\quad (0.000) \qquad (0.059) \qquad (0.037)$$
$$R^2 = 0.377, \; D.W. = 2.205 \tag{18}$$
$$Inf_t^{JP} = 0.002 + 0.184\, Inf_{t-1}^{JP} + 0.144\, Inf_{t-1}^{UK},$$
$$\quad (0.000) \qquad (0.070) \qquad (0.058)$$
$$R^2 = 0.290, \; D.W. = 2.100$$

The statistically significant coefficients of the foreign variables again document the potential for a useful generalization of information across domains. In conclusion, forecasters lacking up-to-date information on driving variables are advised to rely on (lagged) information concerning variables from other domains. As the empirical analysis of bond risk spreads and inflation rates illustrates, data from outside the domain (i.e. from another country) tend to help to improve the forecast. For practical applications, it remains to decide from what domain the explanatory variable should come. Besides data availability, the condition that domains need to be affected by a common general variable can be a helpful guide[15].

6. Conclusions

Forecasters and decision makers should indeed consider relying on variables outside the domain for which they forecast even when these variables do not influence the variable to be predicted through mechanisms suggested by standard economic theory. We study how Granger causality between variables can result because important driving variables are – in principle or at least at the time of the forecast – not observable. Under such circumstances information generalization across domains is rational because variables from another domain can reveal information on common driving forces. Hence, Granger causality linking variables from different domains may be the rule and should be exploited for forecasting.

The theoretical analysis suggests the possibility that the direction of such effects across domains can be counter-intuitive. In time-series analysis, such effects show up in coefficients with the "wrong" sign. This study suggests that such coefficient estimates should not automatically be discarded. However, only exceptional circumstances in terms of the necessary parameter constellation can generate such counter-intuitive effects. For practical econometric work with the purpose of forecasting we find the intuitively plausible direction of the effect in the empirical examples studied here. Although we limit the application to the two examples of bond risk spreads and inflation rates, the approach suggested here is widely applicable. Summing up, this study should help forecasters by indicating a wider set of relevant variables for their prediction models.

Notes

1. More precisely, Granger causality asks for past observations of X (the "causal variable") to be correlated with that part of the current value of Y (the variable to be forecast) that cannot be explained by past observations of Y.
2. The Cowles Commission has been an important force in the development of the structural approach (Christ, 1994). It has to be noted that the focus on identification of structural parameters of behavioral functions serves also purposes other than just forecasting.
3. The generalization of information is related to the concept of information contagion (Calvo and Mendoza, 2000; Kodres and Pritsker, 2002; Garleanu *et al.*, 2015; Chen and Suen, 2016).
4. If the common driving variable is observable at the time of the forecast, there is no reason for the forecaster to rely on information regarding the other domain.
5. More concretely, with unobservable driving variables, the forecaster refrains from quantifying effects of local and global forces on domain variables combined with forecasts of these driving forces. Instead, only the lagged values of the variables of interest serve as the conditioning information.
6. The important point here is that these factors are not yet observed in time for the forecast.
7. For reasons of symmetry we have for the case of the normal effect of a generalization of information $P(q^B_{t+1} = \underline{q} | q^A_t = \underline{q} \cap q^B_t = \bar{q}) > P(q^B_{t+1} = \underline{q} | q^B_t = \bar{q})$ and for the paradoxical effect $P(q^B_{t+1} = \underline{q} | q^A_t = \underline{q} \cap q^B_t = \bar{q}) < P(q^B_{t+1} = \underline{q} | q^B_t = \bar{q})$.
8. Consider as an example: $P(\underline{B}_t, \underline{G}_t | q^B_t = \underline{q}) = [0.25(1 - u)]/[0.25(1 - u) + 0.25(1 - v) + 0.25u + 0.25v] = (1 - u)/2$.
9. Kennedy (2005) and Wang and Wang (2009) discuss other causes for estimates with wrong signs and aspects of how to deal with such findings.
10. The corresponding estimates for variable q^A as a function of lagged values of q^A and q^B lead to symmetrical results.
11. On a slightly different turn Giesecke *et al.* (2011) study bond default rates and find that spreads do not help to predict default rates.
12. In reality the simplification of perfect symmetry across regions used in our model will not hold. Hence, symmetry of estimated coefficients across equations should not be expected. Specification (17) does not claim to represent the best possible specification for forecasting the variables considered here.
13. The reported coefficient estimates for the lagged observations from the other country are only marginally affected when, instead of working with first differences, we estimate a vector error correction model that includes the stationary residual from the cointegration regression $S^{\$}_t = 1.408 + 0.685\, S^{€}_t$. It is noteworthy that the error correction term only appears in a statistically significant way in the explanation of the €-risk spread.
14. Phillips–Perron tests show logs of the price levels of the two countries as separately non-stationary but not cointegrated. Instead, both inflation rates are stationary variables.

15. When considering introducing variables from several domains (e.g. from several other countries), the danger of overfitting, as always, must be considered.

References

Bora, O., Kabir, H.M. and Hepsen, A. (2016), "Returns predictability in emerging housing markets", *Journal of Economic Cooperation and Development*, Vol. 37 No. 1, pp. 101-130.

Calvo, G.A. and Mendoza, E.G. (2000), "Rational contagion and the globalization of securities markets", *Journal of International Economics*, Vol. 51 No. 1, pp. 79-113.

Chen, H. and Suen, W. (2016), "Falling dominoes: a theory of rare events and crisis contagion", *American Economic Journal: Microeconomics*, Vol. 8 No. 1, pp. 228-255.

Christ, C.F. (1994), "The Cowles commission's contributions to econometrics at Chicago, 1939-1955", *Journal of Economic Literature*, Vol. 32 No. 1, pp. 30-59.

Diks, C. and Wolski, M. (2016), "Nonlinear Granger causality: guidelines for multivariate analysis", *Journal of Applied Econometrics*, Vol. 31 No. 7, pp. 1333-1351.

Garleanu, N., Stavros, P. and Jianfeng, Y. (2015), "Financial entanglement: a theory of incomplete integration, leverage, crashes, and contagion", *American Economic Review*, Vol. 105 No. 7, pp. 1979-2010.

Giesecke, K. Longstaff, F.A., Schaefer, S. and Strebulaev, I. (2011), "Corporate bond default risk: a 150-year perspective", *Journal of Financial Economics*, Vol. 102 No. 2, pp. 233-250.

Granger, C.W.J. (1969), "Investigating causal relations by econometric models and cross-spectral methods", *Econometrica*, Vol. 37 No. 3, pp. 424-438.

Granger, C.W.J. (1980), "Testing for causality: a personal viewpoint", *Journal of Economic Dynamics and Control*, Vol. 2 No. 1, pp. 329-352.

Hirshleifer, J. and Riley, J.G. (1992), *The Analytics of Uncertainty and Information*, Cambridge University Press, Cambridge.

Joshi, A.R. and Acharya, D. (2011), "Inflation model for India in the context of open economy", *South Asia Economic Journal*, Vol. 12 No. 1, pp. 39-59.

Kennedy, P.E. (2005), "Oh no! I got the wrong sign! What should I do?", *Journal of Economic Education*, Vol. 36 No. 1, pp. 77-92.

Kodres, L.E. and Pritsker, M. (2002), "A rational expectations model of financial contagion", *Journal of Finance*, Vol. 57 No. 2, pp. 768-799.

Reilly, F.K., Wright, D.J. and Gentry, J.A. (2010), "An analysis of credit risk spreads for high yield bonds", *Review of Quantitative Finance and Accounting*, Vol. 35 No. 2, pp. 179-205.

Wang, P. and Wang, P. (2009), "Does a "correct" parameter estimate tell a better story about foreign exchange market efficiency?", *Journal of International Money and Finance*, Vol. 28 No. 2, pp. 183-197.

Appendix. An outline of the derivation of results

The elements of analysis used here can be found in Hirshleifer and Riley (1992, chapter 5). The first step is the calculation of the probabilities of the outcome $(q_t^A = \bar{q}) \cap (q_t^B = \underline{q})$ given all possible eight combinations of the three factors A, B, and G. Figure A1 shows these joint probabilities.

In a further step, Bayes theorem answers the question with what probability the combination $(q_t^A = \bar{q}) \cap (q_t^B = \underline{q})$ arises from any specific combination of these factors. Take the example of the combination $\underline{A}_t, \underline{B}_t, \underline{G}_t$

$$P\,(\underline{A}_t, \underline{B}_t, \underline{G}_t | q_t^A = \bar{q} \cap q_t^B = \underline{q}) = \frac{u(1-u)}{2u(1-u) + 2u(1-v) + 2v(1-u) + 2v(1-v)} = \frac{u(1-u)}{2(u+v)(2-u-v)} \quad \text{(A1)}$$

In this way, we determine the likelihood of every combination of the values for A, B, and G given that $(q_t^A = \bar{q}) \cap (q_t^B = \underline{q})$. Next, we compute, for every combination of factors, the probabilities with which each of the eight possible combinations of factors are obtained in the next period. For example, the probability that any given combination is repeated is $\rho\lambda\theta$. Alternatively, the probability that Factors B and G remain the same but that A switches value is $(1-\rho)\lambda\theta$. Now the probabilities $P\,(A_{t+1}, B_{t+1}, G_{t+1} | q_t^A = \bar{q} \cap q_t^B = \underline{q})$ are the sum of

Figure AI Joint probabilities for high productivity in domain A and low productivity in domain B $P(q^A = \bar{q} \cap q^B = \underline{q})$

		G	
		0	1
A, B	0 0	$u(1-u)$	$(1-v)v$
	0 1	$u(1-v)$	$(1-v)u$
	1 0	$v(1-u)$	$(1-u)v$
	1 1	$v(1-v)$	$(1-u)u$

the probabilities of each $A/B/G$-combination given $(q_t^A = \bar{q}) \cap (q_t^B = \underline{q})$ weighted with the probabilities that these combinations lead to the various $A_{t+1}, B_{t+1}, G_{t+1}$ combinations. Here is an example:

$$P(\underline{A}_{t+1}, \underline{B}_{t+1}, \underline{G}_{t+1} \mid q_t^A = \bar{q} \cap q_t^B = \underline{q})$$
$$= \sum_X \sum_Y \sum_Z \{P(\underline{A}_{t+1}, \underline{B}_{t+1}, \underline{G}_{t+1} \mid A_t, B_t, G_t) \times P(A_t, B_t, G_t \mid q_t^A = \bar{q} \cap q_t^B = \underline{q})\} \tag{A2}$$

Again we multiply all these probabilities with $P(q_t^A = \bar{q} \cap q_t^B = \bar{q} | A_t, B_t, G_t)$ and $P(q_t^A = \underline{q} \cap q_t^B = \bar{q} | A_t, B_t, G_t)$, respectively. Finally, the resulting terms $P(q_{t+1}^A = \bar{q} \cap q_{t+1}^B = \bar{q} | q_t^A = \bar{q} \cap q_t^B = \underline{q})$ and $P(q_{t+1}^A = \underline{q} \cap q_{t+1}^B = \bar{q} | q_t^A = \bar{q} \cap q_t^B = \underline{q})$ are added up to reach the result for $P(q_{t+1}^B = \bar{q} | q_t^A = \bar{q} \cap q_t^B = \underline{q})$.

Corresponding author

Tobias Rötheli can be contacted at: tobias.roetheli@uni-erfurt.de