Military Expenditure and Granger Causality: A Critical Review[1]

J. Paul Dunne,
University of West of England

Ron P. Smith,
Birkbeck College,
University of London

June 2010

Abstract

A large literature has used tests for Granger (1969) non-causality, GNC, to examine the interaction of military spending with the economy. Such tests answer a specific though quite limited question: can one reject the null hypothesis that one variable does not help predict another? If this null is rejected, there is said to be Granger causality, GC. Although the limitations of GNC tests are well known, they are often not emphasised in the applied literature and so may be forgotten. This paper considers the econometric and methodological issues involved and illustrates them with data for the US and other countries. There are three main issues. First, the tests may not be informative about the substantive issue, the interaction of military expenditure and the economy. The difficulty is that Granger causality, incremental predictability, does not correspond to the usual notion of economic causality. To determine the relationship of the two notions of causality requires an identified structural model. Second, the tests are very sensitive to specification. GNC testing is usually done in the context of a vector autoregression, VAR, and the test results are sensitive to the variables and deterministic terms included in the VAR, lag length, sample or observation window used, treatment of integration and cointegration and level of significance. Statistical criteria may not be very informative about these choices. Third, since the parameters are not structural, the test results may not be stable over different time periods or different countries.

Keywords: Military Spending; economic growth; causality; VAR
JEL Classification: H56; C01; O40

[1] We are grateful to Vincenzo Bove, Elisa Cavartorta and Hashem Pesaran for comments on an earlier version.

1. Introduction.

A large literature has used Granger (1969) causality, GC, tests to examine the interaction of military spending with the economy. The null hypothesis in these tests is Granger non-causality, GNC: that one variable does not help predict another. For instance, Dunne and Smith (1990), in the first issue of this journal, could not reject the hypothesis of GNC in both directions between the share of military expenditure and unemployment in a sample of OECD countries. Tang et al. (2009) also test for GNC using global panel data and find no evidence for GC from unemployment to military expenditure, but some evidence for GC from military expenditure to unemployment for non-OECD countries. There are also many studies testing GNC between military expenditure and other economic variables, particularly output; early examples being Joerding (1986), Kinsella (1990), Chowdhury (1991), Chen (1993) with more recent examples being Abu-Bader and Abu-Qarn (2003) and Kollias et al. (2004, 2007), who also take account of the possibility of cointegration.

GNC tests are a useful statistical technique to answer a specific though quite limited question: does one variable help predict another? Although the limitations of GNC tests are known, they are often not emphasised in the applied literature and so may be forgotten. Therefore it seems useful to survey the application of the technique and its limitations, using data from the US and other countries for illustration. We emphasise the short-run time series issues, the longer-run cross section issues in the military expenditure growth relationship are discussed in Dunne et al. (2005) and some of the wider econometric issues in Dunne and Smith (2007).

The attraction of the GC approach is that it treats all the variables, military and economic, as endogenous; allows effects in any direction; and uses an ‘atheoretical’ reduced form, vector autoregression, VAR, framework. This has the advantage over a structural model in that it does not require identification. The identification problem is apparent in the contrast between the literature on the economic effects of military expenditure, which treats military expenditure as exogenous and GDP as endogenous, and the literature on the demand for military expenditure which treats military expenditure as endogenous and GDP as exogenous. To identify both relationships requires having some exogenous strategic or political variables (such as measures of threat) that shift military expenditure but not GDP and some economic variables that shift GDP but not military expenditure. It may be difficult to find such variables.

However, the advantages of the reduced form approach come at some costs, including the difficulty of interpreting the results; the sensitivity of the tests to the specification of the VAR and the fact that since the estimates on which the tests are based are not structural, they may not be stable[2].

Section 2 discusses the wider theoretical issues in the interactions between the economy and military expenditure. The subsequent sections discuss a series of issues in the use of the concept of Granger causality: definition and testing, integration and cointegration, measuring the size and direction of the effect, instantaneous causality and the number of variables to include in the VAR. These issues are illustrated with data, primarily from the US, but in section 9 other countries.

2. Interactions between military expenditure and the economy

The standard economic account of how a nation determines military expenditure emphasises perceptions of: the threats to its security; its ability to pay, usually measured by GDP; and the opportunity costs of military expenditures. These perceptions are mediated by domestic political and bureaucratic institutions, including, perhaps, a military industrial complex. The economic effects of military expenditure may include short-run Keynesian demand effects and longer-run supply effects, such as displaced investment. Within standard economic models, these economic effects of military spending are usually regarded as being quite small and so generally ignored, except for the case of financing major wars, Smith (2009).

Historically, the effect of major wars means that military expenditures have shown much larger variations than any other category of government finance, and these are usually considered as exogenous and have been used to measure the effects of fiscal policy, e.g. by Hall (2009) for the US. This variation is clear in the US, where the share of National Defence Expenditures in GDP was less than two per cent of GDP during the inter-war period; then rose with the war, peaking at around 50% of GDP in 1943 and 1944. With the end of World War II, the share fell sharply to around seven per cent, rising again to almost 15 per cent in 1953, with the Korean War.

[2] Micro-econometrics uses a quite different concept of causality from GC, based on comparison of different potential outcomes: two hypothetical states of the world one with the potential cause, usually labelled the treatment, and one without. Heckman (2008) or Angrist and Pischke (2009) provide discussions of this concept and Lechner (2006) discusses the relationship between them, but we confine ourselves to GC.

Subsequently the share trended downwards, jumping upwards in the late 1960s with the Vietnam War; reaching a peak of 10 per cent in 1967. The share then resumed its downward trend till 1979, falling to 5.7 per cent. With the Soviet invasion of Afghanistan, the election of President Reagan and worsening relations with the Soviet Union; the share rose, peaking in 1986 at 7.8 per cent. As the Cold War thawed and then ended, the share fell; reaching a low of 3.8 per cent in 2000. The Global War on Terror, after 2001, increased the share to just over 5 per cent in 2008. By US post-war historical standards, this is still quite low; military expenditure had accounted for over five per cent of US output in every year from 1941 to 1994.

The account above emphasises strategic factors, particularly wars, in shifting the share of military expenditure; but there is a long tradition of explaining military spending not by its strategic function but by its economic function: it is necessary to maintain growth and profitability. Part of the context for this explanation is the high unemployment of the inter-war period. The slump was widely interpreted as an inability of capitalism to generate enough effective demand, consumption or investment, to maintain full employment. Many forecast that World War II would be followed by a slump similar to that following World War I. This did not happen; the period from the end of World War II until the crises of the 1970s was one of low unemployment that, in retrospect, was labelled a golden age of capitalism, Glyn, (2006).

Some argued that military expenditure was the source of the extra effective demand that stopped capitalism sinking into depression; since the US and UK devoted a much higher share of output to the military than their previous peacetime norms. The most influential exposition of this view was Baran and Sweezy (1966). This argument, sometimes labelled military Keynesianism, was developed by various other authors, particularly with reference to the US. They suggested that military expenditure was used to offset the tendency to stagnation and unemployment and adjusted to stabilise the economy and thus was a blessing for capitalism, rather than a burden. Cypher (2007) suggests that there has been a shift from military Keynesianism to “global-neoliberal militarism”, in which the economic benefit of military expenditure comes through rather different channels.

There are a variety of problems with the military Keynesian argument. It is not clear that either Marxist or Keynesian theory actually predicts such under-consumption tendencies. The strategic explanations, rooted in war and the communist

threat, seem a better explanation of military expenditures than economic justifications. It is relatively straightforward to tell a strategic story to explain the share of military expenditure in the US, as was done above. It is very difficult to tell an economic story. Although World War II, the Korean Wars and the peak of the Vietnam wars were periods of relatively full employment in the US, the strong downward trend in the share of military expenditure is not marked by any corresponding upward trend in unemployment. The extent of the communist threat to the US is a matter of dispute, but it was certainly perceived as real by US decision-makers. While economic factors were certainly important at a micro level (weapons projects and base locations) they seem less so at a macro level. Military expenditure would be a very bad fiscal regulator because of the lags before it comes into effect: it takes too long to plan and implement to be an effective stabiliser, Smith & Dunne (1994).

Many countries with low military expenditure, in particular Germany and Japan, showed lower unemployment and faster growth than the US and UK, though it could be argued that they benefited from the spillovers from UK and US military Keynesianism. There are other explanations for the golden age and why it came to an end in the 1970s. When the Cold War ended, the UK and US cut their military expenditures substantially and rather than sinking into unemployment both grew rapidly; benefiting from the peace dividend. The cuts in military expenditure reduced government deficits, which allowed lower interest rates, boosting investment in the technology boom of the 1990s.

Granger causality methods appear to avoid the complexity of the underlying theoretical arguments, by simply considering bivariate relations between military spending and an economic variable, like growth or unemployment. The alternative is to specify an underlying structural model, recent Keynesian examples are Pieroni et al. (2008) and Atesoglu (2009).

Clearly, there are a range of possible interactions. Keynesian demand side explanations might suggest a positive effect of military expenditure on output; supply side displacement of factors of production might suggest a negative effect. Ability to pay arguments might suggest a positive effect of output on military expenditure; while military Keynesian effects to stabilise output might suggest a negative effect: if output drops military expenditure is increased to compensate, D'Agostino et al. (2010). Below we consider whether tests for GNC can shed any light on these interactions.

3. Definition and tests

Suppose that we are interested in the relationship between a measure of military expenditure, e.g. the logarithm of real military expenditure, m_t, and an economic variable y_t, say the logarithm of output, but it could be unemployment or investment; and we have data $t = 1, 2, ..., T$. A variable, say m_t, is said to be Granger causal, GC, for another variable, y_t, if knowing past m_{t-i}, in addition to other available information including past y_{t-i}, helps to explain current y_t; or equivalently, current m_t helps to predict future y_{t+i}. There are a number of points to note about this definition. Firstly GC is defined relative to a particular information set, to which the potential predictor is added.[3] Secondly, this is GC in expected value, one could also have GC in higher moments, for instance the variance of y_t might be predicted. Thirdly, there is the complication that it is possible that m_t may not help predict y_{t+1} one period ahead, but may help predict y_{t+h} several periods ahead, because it works through indirect effects. We ignore this issue which is discussed by Dufour and Renault (1998).

GC measures incremental predictability, relative to an information set, not causality in the usual sense: weather forecasts are GC for the weather, but few regard them as causing the weather. The use of the abbreviation GC for Granger Cause is to emphasise that this is not causation in the usual sense. The fallacy of inferring causation from temporal sequence is known as *post hoc, ergo propter hoc*, 'after it therefore because of it', which is the title of a famous article, Tobin (1970), criticising the use of timing to establish causality. This timing problem is particularly severe in economics because of the role of expectations and intentionality. For instance, the permanent income theory, which says consumption is determined by expected future income, implies that changes in consumption are not predicted by past income, while consumption does predict income. So according to the theory consumption is GC for income, while income is GNC for consumption, the reverse of the economic causality.

[3] Granger (1969, p428) originally defined it that "Y_t is causing X_t if we are better able to predict X_t using all available information than if information apart from Y_t had been used." But this is not operational, since one cannot specify all available information.

Granger (1969) did not provide a test for GC, the standard test was provided by Sims (1972), so it is sometimes called Granger-Sims causality[4]. The general framework is provided by a vector autoregression VAR

$$Y_t = a_o d_t + \sum_{i=1}^{k} A_i Y_{t-i} + u_t \quad \text{or} \quad A(L)Y_t = a_o d_t + u_t \qquad (1A)$$

Where Y_t is an $n \times 1$ vector of endogenous variables; d_t is a vector of deterministic elements; u_t a vector of errors with expected value zero and covariance matrix $E(u_t u_t') = \Sigma$; and $A(L) = I - A_1 L - ... - A_k L^k$ is a polynomial in the lag operator. For illustration assume $n = 2, k = 2$, and that log output and log military expenditure, y_t, m_t, are represented as a second order VAR with trend

$$\begin{aligned} y_t &= a^1_{10} + a^2_{10} t + a^1_{11} y_{t-1} + a^1_{12} m_{t-1} + a^2_{11} y_{t-2} + a^2_{12} m_{t-2} + u_{1t} \\ m_t &= a^1_{20} + a^2_{20} t + a^1_{21} y_{t-1} + a^1_{22} m_{t-1} + a^2_{21} y_{t-2} + a^2_{22} m_{t-2} + u_{2t} \end{aligned} \qquad (1B)$$

The elements of the variance covariance matrix Σ are:

$E(u^2_{1t}) = \sigma_1^2; E(u^2_{2t}) = \sigma_2^2; E(u_{1t} u_{2t}) = \sigma_{12}$.

And the null hypotheses are:

(a) that m_t is GNC for y_t, $H^0_{21} : a^1_{21} = a^2_{21} = 0$ and

(b) that y_t that is GNC for m_t $H^0_{12} : a^1_{12} = a^2_{12} = 0$.

The alternative hypotheses are $H^1_{21} : a^1_{21} \neq 0; a^2_{21} \neq 0$ and $H^1_{12} : a^1_{12} \neq 0; a^2_{12} \neq 0$.

Depending on the results of these tests, there are four outcomes with GC running: in both directions (both hypotheses (a) and (b) are rejected), one direction, the other direction, or neither direction. The hypotheses can be tested with standard F statistics on each equation, though these are only valid (i) asymptotically because of the lagged dependent variables in the equations and (ii) if the variables are stationary, I(0). Both small samples and I(1) variables will cause the p values associated with the usual tests to be smaller than the true values.

Since GC is defined relative to a particular information set, the result of the test is conditional on the specification of the model (e.g. it assumes a particular set of variables in the VAR and that the data are generated by a VAR of finite order rather than say a VARMA model). We have assumed the model is linear, Karagianni and

[4] There are other test procedures, e.g. using the cross-spectrum, but these have not been widely used in defence economics, an exception is Gerace (2002). These face similar difficulties to the VAR approach on which we concentrate.

Pempetzoglu (2009) consider non-linear GC of military expenditures in Turkey. The tests can be sensitive to the inclusion of trends and other deterministic elements, such as seasonal dummies, and the choice of lag length. Information criteria, like the AIC or SBC are often used to choose the lag length, but different criteria may give different indications. There is a crucial trade-off, which appears here and elsewhere. Adding extra lags reduces the probability of misspecification and thus bias; but also increases the standard errors, reducing the power of the test. The power is the probability that one will reject the null (GNC) when it is false, so low power means that one is less likely to find significant Granger causality.

There are issues in the choice of measures of military expenditure. In the literature levels, growth rates and shares in GDP of military expenditure have been used and real military expenditure may be calculated using some general price index like the GDP deflator or a military specific price index. In some circumstances, this does not matter. Suppose y_t is the logarithm of GDP and m_t is log military expenditure (either both real or both nominal), so $sm_t = m_t - y_t$ is the log share. We can reparameterise the VAR in (1), by rearranging the right hand side of the first equation, then subtracting the first equation from the second and rearranging the right hand side of the result to give:

$$y_t = a^1_{10} + a^2_{10}t + (a^1_{11} + a^1_{12})y_{t-1} + a^1_{12}sm_{t-1} + (a^2_{11} + a^2_{12})y_{t-2} + a^2_{12}sm_{t-2} + u_{1t}$$
$$sm_t = (a^1_{20} - a^1_{10}) + (a^2_{20} - a^2_{10})t + [(a^1_{21} - a^1_{11}) + (a^1_{22} - a^1_{12})]y_{t-1} + (a^1_{22} - a^1_{12})sm_{t-1}$$
$$+[(a^2_{21} - a^2_{11}) + (a^2_{22} - a^2_{12})]y_{t-2} + (a^2_{22} - a^2_{12})sm_{t-2} + (u_{2t} - u_{1t})$$

(2A)

So we can write a statistically identical system to (1) as:

$$y_t = b^1_{10} + b^2_{10}t + b^1_{11}y_{t-1} + b^1_{12}sm_{t-1} + b^2_{11}y_{t-2} + b^2_{12}sm_{t-2} + u_{1t}$$
$$sm_t = b^1_{20} + b^2_{20}t + b^1_{21}y_{t-1} + b^1_{22}sm_{t-1} + b^2_{21}y_{t-2} + b^2_{22}sm_{t-2} + (u_{2t} - u_{1t})$$

(2B)

Where the b^k_{ij} are functions of the a^k_{ij}. Which is the more useful parameterisation depends on what it is to be used for; one cannot test between them, since they are observationally equivalent. However, the test of y_t being GNC for sm_t tests a different hypothesis from that for y_t being GNC for m_t, because a different combination of the a^k_{ij} are set to zero; though the test for sm_t and m_t being GNC for y_t are equivalent, as can be seen from the first row of (2A). Pesaran and Smith

(1998, section 4) give a more general treatment of such linear transformations of variables in VARS.

As discussed below, one can also convert (1) into a growth rate model through the use of a vector error correction model, VECM. However, the comparison between models using the levels and logarithms of the variables is more complicated, since they are not nested. A related measurement issue is that one might think that there is endogeneity because GDP includes military expenditure as a component and one might also want to use the non-military component of GDP instead of GDP.

If a set of variables are integrated of order one, I(1), stationary after being differenced once, as seems common for economic variables, and there exist linear combinations of them which are I(0), stationary; they are said to cointegrate. If there is cointegration, there must be GC in at least one direction; some feedback which stops the I(1) variables diverging. In dealing with I(1) data, it is convenient to rewrite the VAR, equation (1), in vector error correction model, VECM, form

$$\Delta Y_t = a_o d_t + \Pi Y_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta Y_{t-i} + u_t \qquad \text{(3A)}$$

or for this specific case

$$\Delta y_t = a^1{}_{10} + a^2{}_{10} t + \pi_{11} y_{t-1} + \pi_{12} m_{t-1} + \gamma_{11} \Delta y_{t-1} + \gamma_{12} \Delta m_{t-1} + u_{1t}$$
$$\Delta m_t = a^1{}_{20} + a^2{}_{20} t + \pi_{21} y_{t-1} + \pi_{22} m_{t-1} + \gamma_{21} \Delta y_{t-1} + \gamma_{22} \Delta m_{t-1} + u_{2t}$$

(3B)

This is just a reparameterisation, the intercepts and residuals of the VECM and VAR will be identical and the VECM parameters are just transformations of the VAR parameters, e.g. $\pi_{11} = a^1{}_{11} + a^2{}_{11} - 1$. If the n variables are I(1) and cointegrate, so that there are $r < n$ linear combinations that are I(0), this implies restrictions on the VECM of the form $\Pi = \alpha\beta$ where α is an $n \times r$ matrix and β a $r \times n$ matrix. The n variables are then determined by r cointegrating variables, which are I(0), and $n - r$ stochastic trends, which are I(1). If the variables are I(1) and do not cointegrate, $r = 0$ and all $\pi_{ij} = 0$, giving a first difference model; i.e. in growth rates if the variables are logarithms. If $r = n$ all the variables are I(0).

The dynamic characteristics of the VAR are determined by the roots of its determinantal equation. A VAR with n variables and k lags will have nk roots. It is

stable if all the inverse roots lie within the unit circle, complex roots will give cyclical behaviour and each stochastic trend will give a root on the unit circle.

In imposing the cointegration restrictions, there is a long-run identification problem, since for any non-singular $r \times r$ matrix P then $\alpha\beta = (\alpha P)(P^{-1}\beta) = \alpha^*\beta^*$ so β and β^* are observationally equivalent. To identify α and β one needs to specify P, i.e. provide r^2 just-identifying restrictions, r on each cointegrating vector. Just-identifying restrictions are not testable. Thus with two variables and one cointegrating vector, $n = 2, r = 1$, one restriction on the cointegrating vector is needed. This can be provided by a normalisation restriction. Normalised on y_t, the I(0) linear combination is $z_t = y_t - \theta m_t - \delta t$, where θ is the long run effect of m_t on y_t and the trend has been restricted to lie inside the cointegrating vector.

Imposing these 2 cross-equation restrictions gives:

$$\begin{aligned} \Delta y_t &= a_{10} + \alpha_1(y_{t-1} - \theta m_{t-1} - \delta(t-1)) + \gamma_{11}\Delta y_{t-1} + \gamma_{12}\Delta m_{t-1} + v_{1t} \\ \Delta m_t &= a_{20} + \alpha_2(y_{t-1} - \theta m_{t-1} - \delta(t-1)) + \gamma_{21}\Delta y_{t-1} + \gamma_{22}\Delta m_{t-1} + v_{2t} \end{aligned} \qquad (4)$$

Where the α_i are feedback or adjustment coefficients which measure how the deviations from equilibrium $(z_t = y_t - \theta m_t - \delta t)$ feed back onto each variable. In this case one could have short-run GC from m_t to y_t, because $\gamma_{12} \neq 0$; yet no long-run causality because $\alpha_1 = 0$. If $\alpha_1 = 0$ (there is no long-run feedback of the disequilibrium on y_t), and $\alpha_2 \neq 0$, (m_t does adjust to the disequilibrium); one would probably want to normalise on m_t, since military expenditure seems to be determined by output. Notice there is no statistical way to determine the correct normalisation.

4. The US case

The issues discussed above will be illustrated on US data. Model (1), a second order bivariate VAR with trend, was estimated on US data for two samples. The shorter sample, 1950-2009, allows comparison with equations using unemployment below; the longer sample, 1932-2009, includes part of the Great Depression and the whole of World War II. The log of real GDP is denoted y_t and m_t is the log of real national defence expenditure converted to constant prices by the GDP deflator. Unusually, the

information criteria agree on the choice of 2 lags for both samples, and the trend is significant in the y_t equation.

We will present the results of the tests in terms of p values, which roughly give the probability that the null hypothesis, GNC, is true; so low p values indicate Granger causality. Table 1 gives p values for the GNC tests in each direction for various models and samples. In model 1 (2 lags, with trend, m_t and y_t), given in the first row, for 1950-2009, GNC of m_t with respect to y_t, would be rejected at the 10% level, with p=0.0959, but not at the 5% level; while GNC of y_t with respect to m_t would be rejected at the 5% level, p=0.0201, but not at the 1% level. For model 1, with the 1932-2009 sample, GNC of m_t with respect to y_t would be rejected at the 5% level, and of y_t with respect to m_t at the 10% level. So for both samples using the 10% level one would conclude that there is GC in both directions; using the 1% level in neither direction; at the 5% level the answer would differ between samples.

Table 1. Bivariate VAR, US real military expenditure and GDP

	p values for null of GNC	1950-2009		1932-2009	
		mGNCy	yGNCm	mGNCy	yGNCm
1	2 lags trend	0.0959	0.0201	0.0166	0.0537
2	2 lags no trend	0.0746	0.0002	0.0905	0.0075
3	1 lag, trend	0.0342	0.4959	0.0179	0.8380
4	3 lags, trend	0.1684	0.0346	0.0240	0.0939
5	Using LSM	0.0959	0.0045	0.0166	0.0107
6	Using SM	0.0487	0.0001	0.4131	0.0110
7	Using LYNM	0.0440	0.0215	0.0004	0.0002

Model 2, in the second row removes the trend; model 3, keeps the trend and uses one lag; model 4 keeps the trend and uses three lags. Model 5 replaces log real military expenditure by the log share with 2 lags and trend and as this is the reparameterisation discussed above, the p value for sm_t being GNC for y_t is identical to that for model 1. Model 6 replaces log real military expenditure by the share, a change in functional form. Model 7 uses log real military expenditure but replaces log GDP by the log of GDP minus military expenditure.

The p values are clearly sensitive to sample, specification, measurement, and significance level; though there is no clear pattern in how these factors change the p

values. While one might expect that using non-military GDP rather than GDP would reduce the predictive power of military expenditure, in fact it increases it markedly for the 1932-2009 period. There is clearly some evidence for predictability. There are 28 tests in total: two directions, two samples, seven models. Of these, in all but 4 cases GNC is rejected at the 10% level. However, the evidence is rather marginal since GNC is only rejected at the 1% level in 6 cases.

Turning to cointegration, for the system using m_t and y_t, over 1950-2009, the inverse roots are: 0.9; $0.43 \pm 0.31i$; and 0.29. This indicates the system is stable, no stochastic trends, but the largest inverse root of 0.9 may be biased down and may not be significantly less than unity. For the 1932-2009 sample the inverse roots are $0.79 \pm 0.25i$ and $0.46 \pm 0.38i$. Over 1950-2009, both the Johansen trace and eigenvalue tests at the 5% level suggest $r = 2$ with no trend, which implies both variables are I(0). However, when a trend is included, Johansen tests suggest $r = 1$, which we choose and we normalise the long-run relationship on log GDP, to give

$$\begin{array}{llll} y_t = & -0.431 \; m_t & +0.038 \; t & +z_t \\ & (0.093) & (0.002) & \end{array}$$

It is not clear how one would interpret this long-run relationship, however it was normalised and it is not plausible to take it at face value: a one percent increase in military expenditure reduces GDP by almost half a percent in the long-run. The VECM is:

$$\begin{array}{llllllll} \Delta y_t = & -0.107 & z_{t-1} & +0.110 & \Delta y_{t-1} & -0.015 & \Delta m_{t-1} & +v_{1t} \\ & (0.030) & & (0.125) & & (0.029) & & \end{array}$$

$SER = 0.022, R^2 = 0.193$

$$\begin{array}{llllllll} \Delta m_t = & -0.486 & z_{t-1} & +1.339 & \Delta y_{t-1} & +0.322 & \Delta m_{t-1} & +v_{2t} \\ & (0.106) & & (0.440) & & (0.101) & & \end{array}$$

$SER = 0.078, R^2 = 0.438$

Both the adjustment coefficients have the expected sign and are significant; both the lagged changes in GDP and military expenditure have significant effects on military expenditure, but neither of the lagged changes have significant effects on GDP. Since it is difficult to interpret the cointegrating relationship it might be better to assume $r = 2$ and treat both variables as trend stationary, but this is a matter of judgement.

5. Size of the effects

Although it is common just to report the results of the test, as was done in table 1 above; knowing that there is GC is of little interest in itself without knowing the sign. It matters whether higher military expenditure predicts higher or lower output, but since the test has a two sided alternative, the test statistics or p values are uninformative about this. Measuring the sign of the effect is not straightforward. For instance, m_t may be GC for y_t, with both a^1_{12} and a^2_{12} significantly different from zero, but of opposite sign and roughly equal size. Then there is a short-run effect but no long-run effect, since $a^1_{12} + a^2_{12} = 0$.

The standard way to analyse the effect of a shock to one variable on another in a VAR is by calculating the impulse response functions, IRFs, which measure the effect of a shock at period t on future values of the variables. IRFs are calculated from the moving average representation of the VAR:

$$Y_t = A(L)^{-1} a_o d_t + A(L)^{-1} u_t = A(L)^{-1} a_o d_t + \sum_{i=0}^{\infty} C_i u_{t-i}$$

where the C_i can be obtained recursively from the A_i. Orthogonalised or Cholesky IRFs, require the model to be identified and assume a recursive structure, implemented by the ordering of the variables. Unless the reduced form contemporaneous covariance matrix Σ is diagonal, the impulse response functions will not be invariant to ordering.

The alternative generalised IRFs, GIRFs, do not require identification, but cannot be given a structural interpretation. The generalised impulse response function measures of the effect of a one standard error shock to particular error, au_t, on a variable, bY_{t+h}, h periods ahead, where a and b are a selection vectors. They are then given by,

$$g(h) = E(b'Y_{t+h} / \xi_t = \sqrt{a'\Sigma a}, I_{t-1}) - E(b'Y_{t+h} / I_{t-1})$$
$$g(h) = b'C_h \Sigma a / (\sqrt{a'\Sigma a}); h = 0,1,2,...$$

If the variables are stationary, $g(h)$ will tend to zero as h gets large.

The GIRFs, for the unrestricted VAR using the share of military expenditure in GDP, 2 lags, with trend 1950-2009, are given in Figure 1 below. The own responses are strongly positive, the responses to the other variable are both negative,

though only significant for some periods. The correlation between the errors $r = \sigma_{12} / (\sigma_1 \sigma_2) = -0.306$ is quite large, so the orthogonalised IRFs are sensitive to the ordering. Figure 2 gives the GIRF for the same specification estimated over the longer period 1931-2009. The impulse response functions are quite different, showing significant positive, not negative responses. This is because over the longer period, the contemporaneous correlation of the errors is positive, $r = 0.428$, rather than negative as it was over the shorter period. Since this correlation is not structural, one would not expect it to be constant over time.

Fig 1 GIRF 1950-2009, share of military expenditure and log GDP

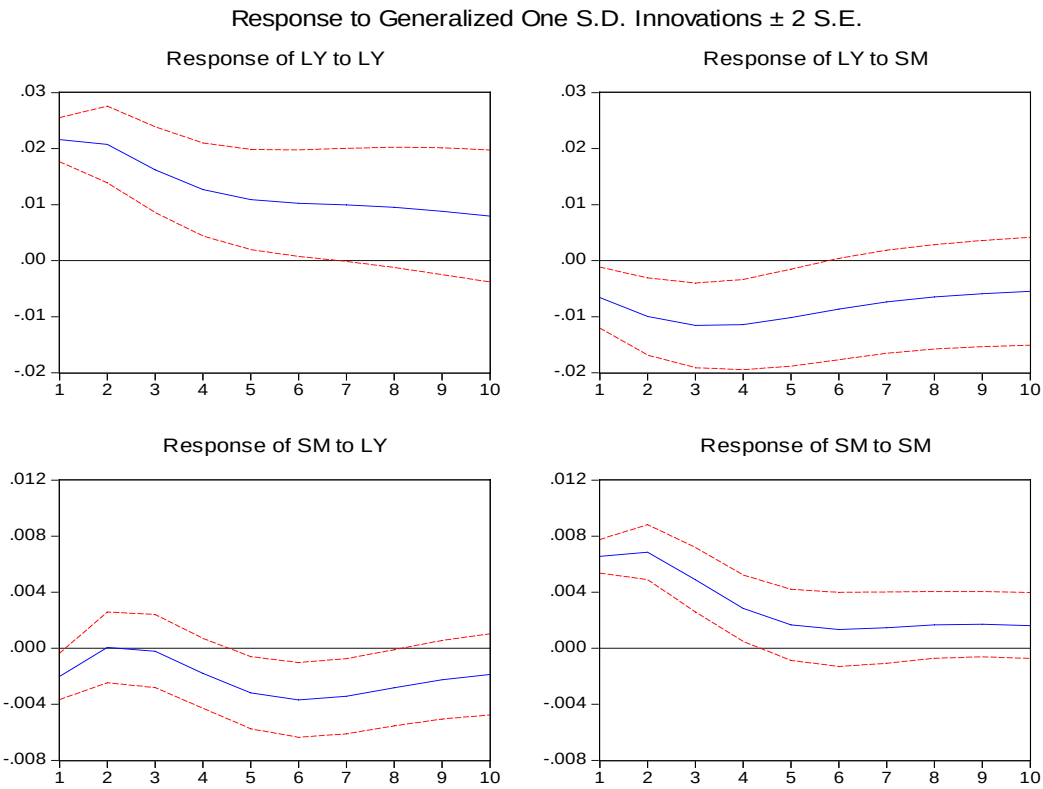
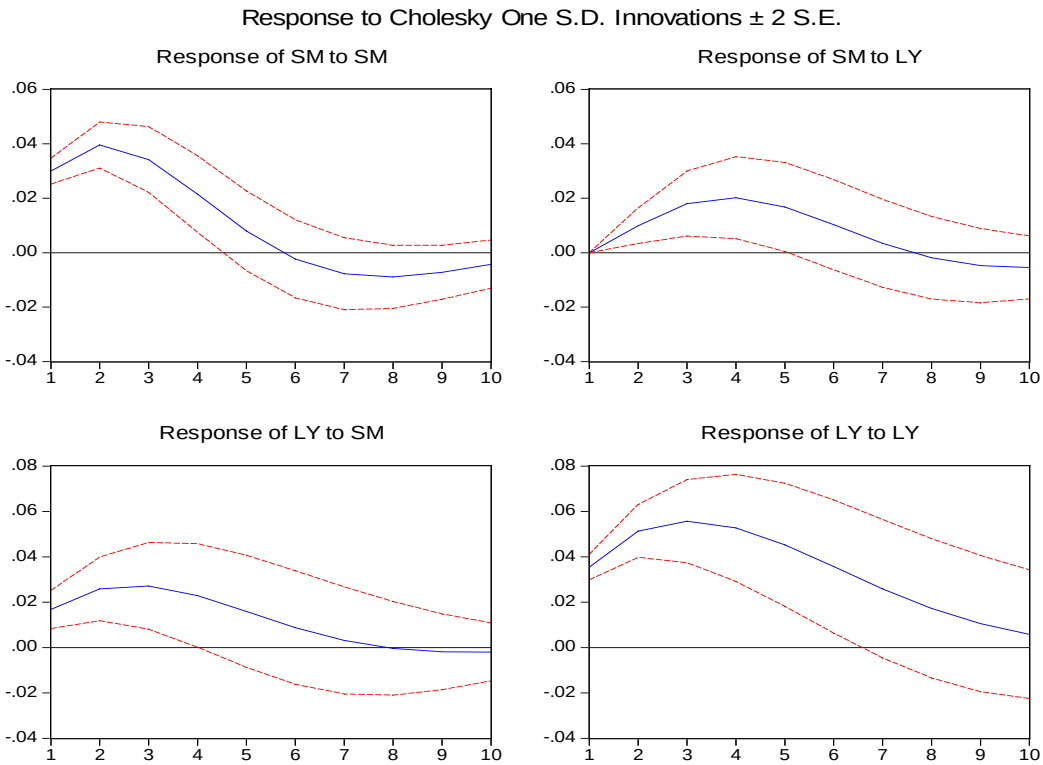


Fig 2 GIRF 1931-2009 share of military expenditure and log GDP



6. **Instantaneous causality**

Granger (1969) discussed instantaneous causality, correlations between the variables within the same period, but was clear that his definition did not cover this, since GC refers only to prediction of later periods. The instantaneous effect is measured by the covariance, $\sigma_{12} = E(u_{1t}u_{2t})$, which as seen above is positive for the longer sample and negative for the shorter sample.

To establish instantaneous causality one must identify the simultaneous relationship. This short-run identification problem is distinct from the long-run identification associated with the cointegrating variables. One can write the VAR in structural form

$$B_0 Y_t = b_o d_t + \sum_{i=1}^{k} B_i Y_{t-i} + e_t$$

with $E(e_t e_t') = \Omega$. The reduced form VAR parameters in (1A) are given by $A_i = B_0^{-1} B_i$; $\Sigma = B_0^{-1} \Omega B_0^{-1}{}'$. Short-run identification requires n^2 restrictions to identify the contemporaneous interactions B_0. The recursive structure used by orthogonalised impulse response functions provides the n^2 restrictions by the assumptions that B_0 is triangular, for a suitable ordering of the variables, and Ω diagonal.

Suppose that $\alpha_2 = 0$ in (4), (which is not the case in our example) then military expenditure is weakly exogenous and we can rewrite the first equation of (4) to condition on Δm_t, giving the traditional single equation error correction model (ECM) determining output:

$$\Delta y_t = \alpha_0 + \beta_0 \Delta m_t + \alpha_1 (y_{t-1} - \theta m_{t-1} - \delta t) + \beta_{11} \Delta y_{t-1} + \beta_{12} \Delta m_{t-1} + e_t$$

Where α_1 is the speed of adjustment; θ the long-run coefficient; and $\beta_1 = \sigma_{12} / \sigma_2{}^2$ measures the short run effect of m_t on y_t, which could be of a different sign to the long-run effect θ. The conditions required to reduce the underlying VAR to an ECM and to test for a long-run relationship, irrespective of whether the data are I(0) or I(1), are discussed in Pesaran, Shin and Smith (2001).

Expectations further complicate the issue, particularly if one variable is a policy instrument. An effective policy instrument will be GNC to the variable it controls, since the variable being controlled will follow its desired path as the policy instrument offsets the shocks. This is one (of many) possible explanation for the result of Tang et al. (2009) who find some evidence for GC from military expenditure to unemployment for non-OECD countries but not for OECD countries. If OECD countries had more effective policy processes they could offset the effects of military expenditure shocks on unemployment, removing the correlation. Of course, the explanation may be because non-OECD military expenditures are more labour intensive as Tang et al. suggest.

7. Additional variables

It is common to test for GC using a bivariate VAR in military expenditure and another focus variable, such as output as above. But this is subject to the criticism that it must be misspecified. Other variables than output, such as the threat, determine military expenditure, and other variables than military expenditure, such as investment, determine output. There is then an issue of the information set used to judge the incremental predictability. The only other information than past m_t used to predict y_t in the VAR above is the two lags of y_t. Since GC tests can be very sensitive to the number of variables included in the VAR, one might want to add all relevant variables but the number of parameters estimated quickly grows large. If there are n variables in a VAR, with k lags; each equation has nk parameters, plus the number of deterministic elements. Large VARs can have very poor statistical properties and GC tests will lose power.

There is a substantial literature on possible ways of reducing the dimensionality of VARs. One route is to apply Bayesian shrinkage to the parameters of a large VAR as in Banbura et al. (2010). Another route is to approximate the possible omitted variables by a few factors estimated from a large data set, which are then used in a factor augmented VAR, FAVAR. Gupta et al. (2010) examine the effect of defence spending on US output using a FAVAR. They converted their data into growth rates to ensure stationarity, and while one can estimate cointegrating

FAVARs, there is the problem that the factors tend to be based on economic variables, while the crucial omitted variables may be strategic.

Table 2 illustrates the sensitivity of the results to the set of economic variables used. Call log GDP, LY, and the share of military expenditure, SM. We add the share of investment in GDP, SI and the unemployment rate, U. Output and unemployment are linked through 'Okun's Law'; investment and output through the accelerator; and there has been considerable interest in the effects of military expenditure on both investment and unemployment. While there are lots of other possible economic variables one could add, such as inflation or interest rates, as noted above what is really lacking is a threat variable to explain the movements in military expenditure.

Table 2 gives the p values for tests of GNC for alternative combinations of the variables. Model 1 is an VAR2 with trend in the four variables, estimated over 1950-2009. The largest inverse root is 0.96, and it proved difficult to identify sensible cointegrating vectors, so an unrestricted VAR is used for the tests.

Table 2: GNC tests for VAR2 with trend 1950-2009

p values for sets of the 4 variables,

	GNC	FROM		
TO	SM	SI	U	LY
Model 1				
SM	-	0.0038	0.7377	0.0163
SI	0.0001	-	0.0000	0.5236
U	0.0721	0.0878	-	0.4712
LY	0.0003	0.0076	0.0000	-
Model 2				
SM	-	-	0.8448	0.0284
U	0.3014	-	-	0.0406
LY	0.0170	-	0.0000	-
Model 3				
SM	-	0.0036	-	0.0189
SI	0.0012	-	-	0.6902
LY	0.0106	0.1255	-	-
Model 4				
SM	-	-	-	0.0001
LY	0.0487	-	-	-
Model 5				
SM	-	0.0000	-	-
SI	0.0010	-	-	-
Model 6				
SM	-	-	0.0083	-
SU	0.5817	-	-	-

The test results for model (1) indicate that at any reasonable significance level (a) unemployment is GNC for the share of military expenditure and (b) log output is GNC for the share of investment and unemployment. The other 9 tests indicate possible GC at significance levels below 10%. Model 2 drops the share of investment, model 3 the unemployment rate, models 4, 5 and 6 are the three bivariate relationships of each economic variable with the share of military expenditure (model 4 is the same as reported in Table 1). There are marked differences between models.

Unemployment is clearly not GC for the military share, when output is included (models 1 and 2) but is strongly GC in the bivariate model 6. The military share is GC for unemployment at the 10% level when output and investment are included, but not when they are dropped. The fact that there is significant GC in both directions between the military and investment shares seems robust across models. In the investment share equation of model 1, SM_{t-1} has a coefficient (t statistic) of -0.80 (-3.7) and SM_{t-2} 0.42 (2.1); in the military share equation SI_{t-1} has a coefficient (t statistic) of 0.42 (3.22) and SI_{t-2} -0.25 (-1.89), the correlation between the errors in the investment and military share equations is -0.61.

These estimates confirm that conclusions on GC can be very sensitive to the number of variables included in the VAR, as well as to lag length, treatment of deterministics, and assumptions about order of integration and cointegration. Given the large number of possible specifications and the danger of data-mining, searching for results in accord with one's beliefs; there is an issue about how results should be reported. We have been able to report the degree of specification sensitivity in much more detail than is usually possible in the literature.

8. Other countries

We now consider more countries, to see whether there is any internationally consistent pattern of GC. Table 3 gives tests for six countries (Italy, Japan, Netherlands, Sweden, UK, US) for the period 1960-2006, using the four variable VAR with trend. The lag order is chosen by the Schwarz Bayesian Information Criterion, which chooses one lag for all countries except the US, where two lags are chosen.

Table 3. Granger Causality p values 1960-2006

To	From			
	SM	SI	U	LY
Italy				
SM	-	0.8936	0.8100	0.0758
SI	0.4012	-	0.3568	0.8117
U	0.3721	**0.0191**	-	**0.0436**
LY	0.0988	0.3814	0.2117	-
Japan				
SM	-	0.2253	0.9263	0.0839
SI	0.2446	-	0.3144	0.7826
U	0.3851	0.6335	-	0.6320
LY	0.7926	0.0696	0.2725	-
Netherlands				
SM	-	0.3937	0.1538	**0.0160**
SI	0.4586	-	0.6660	0.2075
U	**0.0091**	0.4406	-	0.1489
LY	0.1178	0.3364	**0.0053**	-
Sweden				
SM	-	0.8838	0.9076	0.9724
SI	0.1776	-	0.8446	0.9437
U	0.3524	0.8908	-	0.1112
LY	0.1866	0.5442	**0.0191**	-
UK				
SM		0.6239	0.6875	0.6219
SI	0.4836	-	**0.0001**	**0.0000**
U	0.3040	**0.0001**	-	**0.0001**
LY	0.9384	**0.0041**	0.1183	-
US				
SM	-	**0.0227**	0.4861	0.1869
SI	0.9275	-	**0.0231**	0.1879
U	0.2387	0.4508	-	**0.0002**
LY	0.7464	0.7589	0.1567	-

The Table is read in the usual way, so for Italy the only evidence of GC at the 5% level is from the share of investment to unemployment (p=0.019). There is little evidence of GC involving the share of military expenditure, of the 36 tests involving military expenditure only 3 are significant at the 5% level. In the Netherlands the share is GC for unemployment and output is GC for the share. In the US investment is GC for the share. Of the 36 tests not involving military expenditure, ten are significant. Of course, with different samples or specifications of the VAR one might have got different answers.

Above the countries were treated individually, but much of the recent work on military expenditures and GC has treated the multiple countries as a panel (e. g. Tang et al., 2009). Testing for GC in panels raise a range of further econometric issues.

These include the specification of the appropriate null and alternative hypotheses and the choice between the large number of possible estimators for dynamic, heterogeneous panels with cross-section dependence and potentially non-stationary data. Some of these issues are discussed in Smith and Fuertes (2010).

9. Conclusion

The nature of the interaction of military expenditure with the economy has been the subject of extensive investigation and tests for Granger non-causality, in the context of a VAR, have been a popular tool in this investigation. This paper has provided a critical review of the available techniques and illustrated the issues with data for the US and some other OECD countries. The tests were seen to be sensitive to: how many variables are included in the VAR, lag lengths, treatment of deterministic elements, the sample or observation window used, the treatment of integration and cointegration and the significance level used.

Statistical measures may not be informative about these choices in the sense that the likelihood function is relatively flat and it may not be obvious how to trade off the benefit of better fit and the cost of added uncertainty that results from estimating extra parameters. Since the parameters are not structural they may not be stable over different time periods or different countries, and this was the case with the empirical results here.

It is important to recognise that Granger causality test statistics are uninformative about the size or direction of the predicted effects and Granger causality measures incremental predictability not economic causality. To determine how Granger causality relates to economic causality requires an identified structural model, and different, observationally equivalent, just identifying assumptions may give very different causal pictures. While identification is difficult, some orientation of the research effort to try to develop more structural models would seem to be potentially more fruitful than less theoretical statistical approaches. One obstacle to developing more structural models is providing measures of the political and strategic determinants of military expenditures, such as threats.

References

Abu-Bader, S. and Abu-Qarn, A. S. (2003), 'Government expenditures, military spending and economic growth: causality evidence from Egypt, Israel, and Syria', *Journal of Policy Modeling* **25**(6-7), 567-583.

Angrist J.D. and J-S Pischke (2009) *Mostly harmless econometrics: an empiricist's companion*, Princeton University Press.

Atesoglu, H. S. (2009) Defence Spending and Aggregate Output in the United States. *Defence Economics* 20(1), 21-26.

Banbura, M., D. Giannone and L. Reichlin (2010) Large Bayesian Vector Autoregressions, *Journal of Applied Econometrics*, 25(1) 71-92.

Baran P. and P. Sweezy (1966) *Monopoly Capital*, Monthly Review Press.

Chen, C-H. (1993), 'Causality between Defence Spending and Economic Growth: The Case of Mainland China', *Journal of Economic Studies*, 20 (6), pp.37-44

Chowdhury, A. (1991), 'A Causal Analysis of Defence Spending and Economic Growth', *Journal of Conflict Resolution*, 35(1), pp.80-97

Cypher, J.M. (2007) From Military Keynesianism to Global-Neoliberal Militarism, *Monthly Review*, June.

D'Agostino, G., J.P. Dunne and L. Pieroni (2010) "Assessing the effects of military expenditures on economic growth." Forthcoming chapter in the *'Oxford Handbook of the Economics of Peace and Conflict'*, edited by Stegios Skaperdas and Michelle Garfinkel, Oxford University Press.

Dufour, J-M, and E. Renault (1998) Short run and long run causality in time series: theory *Econometrica* 66(5) 1099-1125.

Dunne, J.P. and R.P. Smith (1990) Military Expenditure and Unemployment in the OECD *Defence Economics* 1(1), 57-73.

Dunne JP and RP Smith (2007) The Econometrics of Military Arms Races Ch 28 p913-940 of *Handbook of Defense Economics, Vol 2: Defence in a Globalised World*, ed T. Sandler and K. Hartley.

Dunne JP, Smith RP and Willenbockel D (2005) Models of military expenditure and growth: a critical review, *Defence and Peace Economics*, 16, 449-461.

Georgiou, G., Kapopoulos, P. and Lazaretou, S. (1996), 'Modelling Greek-Turkish Rivalry: An Empirical Investigation of Defence Spending Dynamics', *Journal of Peace Research*, 33 (2), pp.229-239

Gerace, M. P. (2002), 'US Military Expenditures and Economic Growth: Some Evidence from Spectral Methods', *Defence and Peace Economics* **13**(1), 1-11.

Glyn, A (2006) *Capitalism Unleashed*, Oxford Univeristy Press.

Granger, C.W.J. (1969) Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, 37 (3) 424-438.

Gupta, R., A. Kabundi and E. Ziramba (2010) The effect of defense spending on US output: a factor augmented vector autorgression (FAVAR) approach, *Defence and Peace Economics*, 21(2) 135-148.

Hall, R.E. (2009) How much does GDP rise if the government buys more output, *NBER working paper* 15496.

Heckman, J. (2008) Econometric Causality, *International Statistical Review*, 76(1), 1-27, 2008.

Joerding, W. (1986) 'Economic Growth and Defence spending: Granger causality', *Journal of Development Economics*, 21, 35-40

Karagianni S. and M. Pemetzoglu (2009) Defence Spending and Economic growth in Turkey: a linear and non-linear Granger causality approach. *Defence and Peace Economics*, 20(2) 139-148.

Kinsella, D. (1990), "Defence Spending and Economic Performance in the United States: A Causal Analysis", *Defence Economics*, Vol. 1, pp. 295-309.

Kolias C. G. Manolas and S-M Paleologou (2004) Defence expenditure and economic growth in the European Union: a causality analysis *Journal of Policy Modelling* 26, 553-569.

Kollias, C., Mylonidis, N. and Paleologou, S.-M. (2007), 'A Panel Data Analysis Of The Nexus Between Defence Spending And Growth In The European Union', *Defence and Peace Economics* **18**(1), 75-85.

Lechner, M. (2006) The relationship of different concepts of causality in econometrics, *Discussion paper 2006-15, Universitat St Gallen.*

Pesaran, M.H. and Smith R.P. (1998) Structural Analysis of Cointegrating VARs, *Journal of Economic Surveys*, 12 (5) 471-505

Pesaran, M.H. Y. Shin and R.J. Smith (2001) Bounds testing approaches to the analysis of level relationships, *Journal of Applied Econometrics*, 16, 289-326.

Pieroni, L. G. D'Agostino and M. Lorusso (2008) Can we declare military Keynesianism dead? *Journal of Policy Modelling*, 30 (5) 675-691.

Sims, C 1972, Money Income and Causality, *American Economic Review*, 62, 540-555.

Smith, R.P. (2009) *Military Economics*, Macmillan-Palgrave.

Smith, R.P. and JP Dunne (1994) "Is Military Spending a Burden?: A Marxo-marginalist Response to Pivetti", *Cambridge Journal of Economics*, Vol 18, 1994, pp 515-21.

Smith R.P. and A-M Fuertes (2010) Panel Time Series, cemmap lecture notes; available at www.ems.bbk.ac.uk/faculty/smith/index_html.

Tang, J.H. C-C. Lai and E.S. Lin (2009) Military Expenditure and Unemployment Rates: Granger Causality Tests using Global Panel Data *Defence and Peace Economics* 20(4) 253-268.

Tobin J. (1970) Money and income: post hoc ergo propter hoc? *Quarterly Journal of Economics*, 84(2) 301-317.