

Budget Restraint and Military Expenditures in NATO Countries: A Review of the Literature

Prepared By:

Professor Dane Rowlands, The Norman Paterson School of International Affairs
Carleton University
1125 Colonel By Drive
Ottawa, Ontario
Canada, K1S 5B6
dane.rowlands@carleton.ca

Contractor's Document Number: 2015-01629; 00BD (Costing methodology Development)
CSA: Binyam Solomon

The scientific or technical validity of this Contract Report is entirely the responsibility of the Contractor and the contents do not necessarily have the approval or endorsement of the Department of National Defence of Canada.

Contract Report
DRDC-RDDC-2015-C236
October 2015

- © Her Majesty the Queen in Right of Canada, as represented by the Minister of National Defence, 2015
- © Sa Majesté la Reine (en droit du Canada), telle que représentée par le ministre de la Défense nationale, 2015

**Budget Restraint and Military Expenditures in NATO
Countries: A Review of the Literature**

Dane Rowlands

Working Paper No. 04, October 2015

Budget Restraint and Military Expenditures in NATO Countries: A Review of the Literature

Prepared by:

Professor Dane Rowlands, The Norman Paterson School of International Affairs
Carleton University
1125 Colonel By Drive
Ottawa, Ontario
Canada, K1S 5B6
dane.rowlands@carleton.ca

Acknowledgements:

I would like to thank, without implicating in any errors, Mr. Decky Kipuka Kabongi for excellent research assistance on this project, Dr. Binyam Solomon for guidance and comments, and the participants in the NATO Science and Technology SAS-113 4th Meeting workshop in Ottawa, October 14-15, 2015 for discussion. This report was prepared for the workshop under the project working title of “Report on Future Budget Constraints: Challenges and Opportunities”, and was commissioned by Defence Research and Development Canada, Department of National Defence, Government of Canada.

1. Introduction

The broad purpose of this report is to review the academic and scientific literature on the factors affecting the quantity and quality of expenditures on defence by members of a military alliance. The motivation for the study is the expectation that countries in the North Atlantic Treaty Organization (NATO) will be facing budget constraints that will impinge on their contributions to collective security. In addition there has been an increasing tendency to rely on “coalitions of the willing” as the dominant organizing framework for recent military missions undertaken by several NATO members outside the European theatre. This evolving strategic environment suggests that NATO member countries may face pressures to rebalance military force structure and procurement in order to meet changing priorities. Specifically, some countries may potentially wish to alter the relative emphasis that they place on national (“private”) and alliance (“public”) military objectives. In addition, engagement in relatively more offensive missions out of the traditional NATO theatres of operation may also generate pressure to rebalance military forces accordingly.

This literature review is structured in the following manner. Section 2 will examine the literature on military alliances and identify insights relevant for the current review. Section 3 will examine more specific examinations of the production and supply of military goods, while a fourth section focuses on the demand side. A concluding section will identify the key lessons that emerge from the review.

2. Theories of Alliances

The use of economics as a tool for the analysis of military alliances was pioneered by Olson and Zeckhauser (1966), who examined security alliances as a form of public good. Security, once provided, can be enjoyed by all members without compromising the security of other members nor, presumably, can members be excluded from enjoying it. The key behavioural insight is that as a public good there is the opportunity for “free-riding”, or the ability of some members to acquire the benefit of collective security while not paying their “fair share”. Though predicated on a rather restrictive assumption that collective security is a pure public good, the analytical approach introduced by Olson and Zeckhauser, and subsequently extended by them and several others (for example Murdoch, 1995), led to several key insights that are summarized in Sandler and Hartley (2001: 875-876) as follows:

- (i) Larger and richer members of an alliance will tend to bear a disproportionately larger share of the defence burden than smaller and poorer ones (the “exploitation hypothesis”).
- (ii) Total defence expenditure will tend to be allocated sub-optimally.
- (iii) To overcome allocation problems there needs to be a strong central authority or other means (such as threats or norms) to induce greater cooperation.

- (iv) Since deterrence benefits are non-rival, alliance size need not be restricted except for reasons of transactions costs (coordinating too many members) or to the extent that collective security is rival.
- (v) The benefits of deterrence may be disconnected from defence expenditures due to the possibility of free riding, leading some alliance members to reduce expenditure when (and because) others raise theirs.
- (vi) The extent of sub-optimal expenditures depends on the size of the alliance and the distribution of member size.

Sandler and Hartley (2001: 876), Sandler and Forbes (1980) among many others develop or review the main extension to the standard public-goods analysis: the joint-products model. Under this alternative specification, alliance defence expenditures can yield pure public benefits to the alliance members (e.g. nuclear deterrence), but also private benefits acquired by each member specifically, as well as impurely public benefits associated with limiting damage to a specific member (e.g. anti-ballistic missiles or protective shelters). Many of the behavioural problems predicted by the pure public goods model of alliances (such as free riding and the exploitation hypothesis) are mitigated or eliminated when military expenditures provide significant private advantages. The joint-products model thus became a standard analytical framework.¹ Essentially the joint-products framework (with more private benefits) predicts smaller defence spending burden asymmetries within the alliance than the pure public goods model.²

Estimating the degree of burden asymmetry as a test of the joint-products model is compromised by the absence of precise concepts and measures of country burdens and benefits. While we may be fairly confident that military expenditures (usually expressed as a percentage of GDP) are a reasonable proxy for the burden, the classification of some security related expenditures (intelligence, police, etc.) may be missed in some cases. In addition, security threats can also be dealt with by diplomatic means or possibly development assistance, so there is some ambiguity with respect to what ought to be included as elements of the burden.

¹ Another approach to theoretically classifying alliance expenditures and benefits has been suggested in Gates and Terasawa (2003), though it maintains its focus on the degree of “publicness” or “privateness” of such expenditures. They provide some useful concepts such as internal burden (military expenditure) and external burden (increased threats arising from membership in the alliance) and alliance benefits (the reduced expenditures on the military arising from the spill-ins of security associated with the military expenditures of other alliance members). However they do not operationalize these concepts, nor do they test them empirically.

² There are alternative but similar economic approaches to modelling alliance behaviour. One strand of analysis examines whether alliance behavior is best explained as a Nash-Cournot equilibrium or a Lindahl equilibrium (McGuire and Groth, 1985; Sandler and Murdoch, 1990; McGuire, 1990). These different strategic environments can best be regarded as competitive (or non-cooperative) and cooperative, respectively. Similarly Hilton and Vu (1991) use the Stone-Geary functional form to measure alliance member welfare. They reject the naïve pure public goods model and instead find “competitive behaviour between allies or apparently selfless commitments to taking on more than a ‘fair’ burden of the response to increases in the threat” (from the abstract).

Regarding benefits, standard analyses (as in Sandler and Hartley, 2001: 884-885, Table 1) measure country benefits as a weighted combination of variables such as population, GDP and exposed borders. Consequently, by virtue of its extensive maritime borders, Canada emerges as by far the largest free-rider in NATO. From a Canadian perspective this finding seems rather odd, given that for long periods of time Canadian military forces were committed to European defence and deployed in the European theatre, but not the reverse. One potential source of the imbalance is the conflating of the NATO alliance with other defence agreements such as NORAD for North American defence. A second possible source of asymmetry is more tactical: to what extent is an amphibious assault on Canada a serious security threat compared to a land-based assault on Europe? This apparent anomaly sparked a valuable exchange between Solomon (2004, 2005) and Sandler (2005) regarding the most suitable measures of alliance benefits and the sensitivity of the extant empirical analyses of burden asymmetry. Solomon casts doubt on the robustness of past findings, pointing to both the maritime border measure and the inclusion of Canada (a clear outlier) as being key determinants of past results, throwing some doubt on the prior conclusions about the public-private balance in military expenditures. Though Sandler and Hartley (1999) acknowledge that the measures of burden asymmetry are sensitive to assumption choices that can be manipulated to emphasize one result over another, the standard approaches to measurement have remained largely unchanged. Sandler and Shimizu (2014) return to the question of burden sharing and introduce exposure to terrorism as a fourth determinant of the benefits received from alliance membership, the inclusion of which is also theoretically problematic.³ Despite these theoretical and empirical concerns, an extensive literature did develop that focused on measuring the burden asymmetry within NATO (for examples see Murdoch and Sandler 1982, 1984).

The extensive burden-sharing literature has identified several factors associated with changes in military spending patterns and the extent of intra-NATO spending asymmetry. These studies identify five structural factors that have been linked to changes in alliance member behavior: technology, membership expansion, changing strategic environments, changing military doctrine, and the focus on new missions.⁴ While often identified separately, it is obvious that most of these factors, if not all, are related to one another, often quite closely, directly and causally. For example a new strategic environment may require the identification of new doctrine and NATO policy, which in turn will result in new missions.

Taking these factors in turn, technology is linked to alliance behavior and burden asymmetry initially through theoretical considerations tied to the public and private goods nature of

³ The use of the actual incidence of terrorism in a country as one element of alliance benefits is understandable as an empirical necessity in the absence of available and superior proxies, but its inclusion crudely assumes that the diminution of terrorist threats (the true benefit) is both proportional to actual attack incidence and attributable to NATO. Its inclusion reduces but does not eliminate burden asymmetry measures, and there remains significant lack of cohesiveness within the alliance in terms of the distribution of benefits and burdens, which the authors identify as a potentially serious source of disunity and a significant challenge for NATO.

⁴ A sixth factor, budgetary pressure, is examined in more detail in section 4 below.

military spending.⁵For example, Sandler and Forbes (1980) look at the implications of military technology for the joint-products model, classifying systems as either purely deterrent (e.g. nuclear weapons for retaliation), purely protective (missile defence or shelters), or a mix of both. Multiple use technologies obviously complicate any assessment of “publicness” and “privateness” and hence our ability to specify the model that best reflects alliance behavior. Sandler and Forbes (1980) and Hartley and Sandler (1999), among others, also consider technology as a factor affecting burden sharing, while Gonzalez and Mehay (1990) suggest that the nature of weapons systems, and presumably their greater sophistication and scale requirements, may induce more cooperative behavior by alliance members.

NATO’s expanded membership is also linked to the technology discussion. The expansion of the NATO frontier and the inclusion of poorer members risked the “thinning” of NATO’s military capacity, which has some coherence with Hirshleifer’s concept of “weak-link” technology. Specifically, NATO’s overall security may be compromised if some frontier states are seen as less well defended and subject to being overwhelmed. Hartley and Sandler (1999) note both the likely “thinning” effect caused by NATO expansion, as well as the implications for more complicated decision making ability within NATO. Sandler and Murdoch (2000) also suggest that expanded membership may increase the burden asymmetry, a particular risk as expansion has increased the number of weaker and poorer states in the alliance, often with non-NATO frontiers.

The expansion of NATO is in turn linked to the new strategic environment faced by NATO after the collapse of the Soviet Union. The perceived diminution of the military threat from the East, and now specifically Russia, and the eagerness of many former Warsaw Pact countries to secure their new independence from Moscow through NATO security guarantees, allowed and required the alliance to reconsider its threat environment. It is difficult, however to separate the strategic environment from its direct effect on strategic doctrine. Sandler and Hartley (2001) provide a useful review of the earlier empirical literature examining military expenditures in alliance frameworks. The earlier studies often found evidence of the exploitation hypothesis and the pure public goods model, at least for the period immediately after the Second World War when the public good of nuclear deterrence was the dominant strategic paradigm. The subsequent period of détente and arms control ushered in the strategic doctrine of “flexible response”, which lasted from 1967 to about 1995. During these years military expenditures were seen as more “private” in nature, as predicted by the joint-products model, and with concomitant diminution (but not elimination) of the burden asymmetries (Khanna and Sandler, 1996, 1997; Sandler and Murdoch, 2000). Sandler and Hartley (2001: 886) argue that NATO’s mid-1990s adoption of a “new crisis-management doctrine paved the way for it to assume peacekeeping missions whenever its security interests were in jeopardy”.

⁵ Some of the original thinking about technology and its behavioural impacts emerge from Hirshleifer’s work on conflict, which focused on fundamental problems of how to aggregate the different contributions to security. Hirshleifer (1983), for example, identified the theoretical extremes of “best-shot” technology (where collective defence was a function only or primarily of the military capacity of the strongest contributor) and the “weakest-link” technology (where the strength of the alliance is effectively dependent on the military capacity of its weakest member).

Sandler and Murdoch (1990) and Khanna, Sandler and Shimizu (1998) emphasize that these operations are associated with greater asymmetries of burdens that are to the disadvantage of richer NATO allies. This new strategic policy has arguably lasted for at least a decade and a half, though it may need to be revised to reflect the re-emergence of the traditional East-West fault lines that developed immediately after the Second World War, this time in the form of competition with Russia.

The post-Cold War strategic environment and associated revisions to NATO policies have been manifested in changes in the nature of NATO activity. Relieved from the preoccupation with the Soviet Union and its allies, NATO member countries have increasingly become involved in peacekeeping, peace enforcement, and other related activities. These new missions exhibited many differences in comparison to the era of flexible response, including an increasing focus on conflicts outside of the immediate European theatre (at least after the conflicts associated with the dissolution of Yugoslavia), often by only a subset NATO members, and often in the context of asymmetric warfare. Very early on, Sandler and Forbes (1980) identified such missions as likely to yield private and excludable benefits as well as rivalry in consumption, which challenge to the public goods element of alliances. Sandler and Hartley (2001) point out that while some of these missions may improve global security and thus constitute a non-excludable public good for all alliance (and presumably non-alliance) countries, other missions may be of primary interest to only a few members, and thus be partially if not extensively rival. Khanna, Sandler and Shimizu (1998) suggest that the emphasis on such missions coincided with more disproportionate military expenditures within the alliance due the varying degrees of importance that different allied countries attached to specific missions. In contrast Ringsmose (2010) suggests that these new missions can best be understood by regarding NATO as a “club good” that arises in part from the US role as a “security guarantor of last resort”.⁶ Ringsmose concludes that NATO will undertake more of these non-traditional missions, especially when US interests are engaged and that as a result the public goods-related collective action problems associated will dissipate and alliance cooperation will increase.

The preceding analyses have largely been undertaken within the dominant theoretical framework of the joint-products model. It should also be noted, however, that the data evaluated in the burden-sharing studies are often consistent with multiple theoretical explanations, including those derived from traditional international relations theory. For example Knorr (1985) examines expenditure asymmetry from the perspective of norms of fairness. O Neal and Elrod (1989) interpret unequal burden sharing using hegemonic stability theory.⁷ Similarly Morrow (1991) focuses on the implications for a military alliance of power asymmetries between members, using data to support the prediction that alliances with significant asymmetries are more stable. Palmer emphasizes the importance of domestic politics and the political orientation of parliaments in determining defence expenditures in Europe (1990a), and the dominance of long-term commitment over short-term expediency that

⁶ Club goods are similar to public goods in that they are non-rivalrous, but unlike pure public goods are excludable.

⁷ O Neal and Elrod’s work provoked a debate with Murdoch and Sandler (1991), and while there was no real resolution of differences, some interesting methodological issues emerged.

emerges out of NATO's internal "bargaining". Amara (2007, 2008) finds that while the strategic environment plays some role in behaviour, military expenditures by alliance members primarily reflected their specific political, economic, and military exigencies (including regional circumstances). Consequently she discounts the importance of alliance-related phenomena such as incentives for free-riding or strategic doctrine. Oma (2012) focuses on threats and security, but concludes that all system-level explanations fail to predict spending patterns, and that models must account for the actual policy-making processes of member states, and specifically of their leader's abilities and incentives to affect spending patterns. Oma's analysis is consistent with the broader public policy literature that emphasizes the importance of perceived domestic interests, especially the political or electoral interests and calculations of democratic governments. As these authors illustrate, the empirical evidence is often consistent with, and often unable to distinguish between, several theoretical interpretations. Consequently these different theoretical traditions provide a valuable set of alternative approaches to understanding and analyzing the behavior alliance members.

Overall, the standard literature on military alliances establishes many of the core concerns about how member government behave within an alliance framework. Key insights such as the potential for free-riding and asymmetric burden sharing may appear to be theoretical, but do need to be understood in terms of alliance unity. The literature also points to the importance of strategic policy, geopolitical conditions, and technology as key factors affecting the extent to which member countries consider military expenditures as contributing to their private interests versus the extent to which they reflect their commitment to NATO and its needs. Understanding the reality of policy making with respect to defence policy, and the relative balance of "public" alliance interests and "private" domestic imperatives is important, especially for understanding the implications of budgetary pressures. How budgetary restraint is translated into defence expenditures, however, also depends on the supply conditions for weapons production.

3. The supply of military goods.

The previous section focused on the literature that explores the relationship of spending patterns and behavior as it relates to alliance theory generally. In this section we examine in more detail those papers that analyze issues related more specifically to the supply side of military procurement and spending; we subsequently examine the demand side in more detail. Separating these literatures is a little artificial, as in many instances observations on supply and demand can always be interpreted in the context of alliance theory, and it is often necessary to examine both supply and demand simultaneously when examining the overall performance of the market.

The first paper we examine, Sandler and Hartley (1995), has some elements of supply, demand and market behaviour despite its emphasis on military procurement. They identify several important features of the supply of military goods, specifically:

- (a) The size and structure of the defence industry is heavily influenced by government (which often acts as a monopsony, i.e. sole-buyer), which typically specifies the technical standards for military equipment.
- (b) Defense equipment is costly, especially modern weapons systems that contain high technology components.
- (c) Due to the need for weapons systems to be close to the technological frontier, defence industries have are relatively intensive in research and development (R&D) compared to many other industries.
- (d) Government regulation is a critical determinant of the openness, viability and profitability of the defence industry. Most importantly governments must determine how to structure procurement contracts so as to regulate the profitability of acquisition programs for suppliers.

Sandler and Hartley (1995) illustrate the intimate connection between the government demand side of military procurement and the supply side, since governments make the critical choices about what weapons systems to purchase, and often choose which supplier to use and which type of contract structure to apply. These decisions ultimately help to shape the structure of the market. Not only is the military goods market typically monopsonistic on the demand side, but it is usually characterized by significant market power on the supply side as well. For some weapons systems there may be very few, possibly even just one, supplier. The markets themselves generally exhibit high degrees of uncertainty, imperfect information, barriers to entry and exit, and are supplied by firms that Sandler and Hartley (1995) among others identify as being “non-profit-maximizing”. This latter characteristic is problematic from an economics perspective, but may reflect different technocratic goals, different discount rates, and in many instances are themselves simultaneously producers of military and civilian goods that have cross-market interdependencies. While the degree of competition can be increased by permitting foreign firms to bid for defence contracts, it should be noted that “national security” is often used as a justification for market protection, especially where sensitive technologies are concerned.

It has long been noted in the literature that the presence of a “military-industrial complex” introduces both a market as well as a political distortion. With large contracts and profits at stake, supplier in the market have a strong incentive to lobby for preferential consideration on bids. I would further note that military production often occurs in large plants due to economies of scale, and thus provide a powerful basis for lobbying local politicians to influence defence procurement decisions. Sandler and Hartley (1995) note that interest groups in the military-industrial-policy complex are the major source of waste, and that they constitute a significant obstacle to reforms that would introduce greater efficiencies. They argue that it is not necessarily desirable to reduce the influence of these groups, citing Lee (1991) who noted that “the result may be an increase in the general level of inefficiency in the economy as the composition of government spending becomes more distorted toward other civilian special-interest programs” (Sandler and Hartley, 1995, p. 143).

Many of the findings in Sandler and Hartley (1995) are also identified by Rogerson (1995) in great detail. He also emphasizes the presence of market failures such as private information

with imperfect monitoring, uncertainty, the inability to write and enforce complete long-term contracts in such an environment, the presence of market power by buyers and sellers who seek to maintain bargaining power over other players, the difficulty of measuring R&D quality and performance and, finally, that governments themselves are complex hierarchical institutions with complicated incentives and relationships (as opposed to a single rational actor). In some countries these deficiencies are made worse by the small size of the procurement market, though in turn these governments may be more likely to use offsetting policies that include opening the market to foreign firms, joint ventures, and other means of improving markets.

In particular Rogerson (1995) details the nature and implications of uncertainty, which is present at the design, production, and deployment stages (“internal uncertainty”) as well as well as uncertainty in demand due to changing threats, competing weapons systems, and political forces (“external uncertainty”). Because of these uncertainties it is often difficult or undesirable to write a long-term fixed-price contract, which are often also impossible to enforce and frequently susceptible to renegotiations.

Rogerson (1995) also details the effect of economies of scale on the procurement market. He argues that at the early design stage of a weapons program there may be several firms capable of bidding for a contract, and consequently the U.S. Department of Defense will typically fund two designs to the prototype stage before selecting a winner. The presence of significant economies of scale at the production stage, however, governments must typically be satisfied by only one system and one producer rather than try and introduce some competition.

Finally, the monopsonistic status of government in the procurement market is also examined in some detail by Rogerson (1995). He notes three consequences of how government responds to the high degrees of uncertainty that affects investments in specific machinery or R&D too overcome the reticence of firms to make suitable investments. First governments themselves typically contract to purchase both intermediate and final goods emerging from R&D activity, thereby either paying directly or subsidizing R&D costs. Second, the government frequently purchases specific assets that the weapons suppliers will use. Third, governments often maintain an administrative connection to firms outside of normal contracts, which essentially provides guarantees to the firms about the security of their investments.

The previous papers present the standard model and results for the defence industry from a microeconomic perspective, and highlight some of their important empirical implications and characteristics. Hildebrandt (1999) takes a different approach to understanding supply in the defence sector. He estimates what he calls the “military production function” that relates military inputs into military effectiveness. His approach, grounded in cost-benefit analysis, is aimed more at a wider understanding of military efficiency. Specifically he asks whether national security objectives are being achieved by the efficient use of military assets.

Hildebrandt (1999) examines specific production functions to establish the tradeoffs between achieving certain military outcomes using different military inputs, from which he can then

determine whether the available assets are being used in the most efficient manner possible. He uses three methods in his analysis. He first estimates the “econometric military production function” using regression analysis of data from the Vietnam War. The dependent variable is the estimate of military effectiveness as measured by the difference between the personnel and material that the North Vietnamese were estimated as attempting to move into the South Vietnamese and Cambodian theatres, and what actually arrived after interdiction efforts. The explanatory variables were inputs into interdictions such as specific weapons systems deployed against specific targets (for example fighter sorties targeting trucks and storage areas). Thus, if it is possible to measure or categorize a specific military “output”, and the set of military “inputs” deployed to achieve that output, then the regression analysis can identify the marginal effectiveness of each input in achieving this output.⁸

His second method, the “response-surface military production function” attempts to reproduce research models of large military operations. While there are similarities to the militant production function, this method constructs a predictive model (validated by empirical calibration) that can estimate output levels (sorties per aircraft per day, in Hildebrandt’s example) based on inputs (maintenance personnel) and parts failure rates. This approach allows planners to undertake detailed trade-off analysis to minimize costs of achieving a specific output or efficiency level.

Finally, the technological military production function analysis illustrated by Hildebrandt (1999) uses technological features of the combat environment to establish the tradeoffs between using different forces that vary with respect to quality, quantity, and type. These functions predict combat outcomes by assigning parameters to a model that estimates how military force interactions will be resolved by computing the expected rate of force attrition. These models are sophisticated versions of the simply Lanchester battle equations constructed for combat scenarios in the First World War. With increased computing power, such models have become increasingly sophisticated and now form the basis of the extensive computer simulation modelling of combat used for training by militaries, and indeed by computer games. There are two shortcomings to note, however. First, the parameterization of the models that capture the effectiveness of different force elements and combinations is often speculative, especially for new weapons systems. Therefore basing procurement decisions on the results of these models is highly problematic. Second, there is not much relevant literature that helps us to understand and evaluate how well these models perform in terms of predicting real combat outcomes, and related questions of model structure.

Kirkpatrick (2004) begins his analysis by highlighting the fact that the effectiveness of a military is defined in large part by the capability of rival forces, a point that is linked to Rogerson’s concept of external uncertainty and Hildebrandt’s approach to modelling force effectiveness. In this framework Kirkpatrick argues that the constant pressure for relative advantage increases the demand for new weapons system, driving a rapid pace of technological improvement and

⁸ Skogstad (2014) uses a similar technique to examine the effectiveness of different configurations of North Atlantic convoy escorts in reducing shipping losses.

continuous price increases. He further concludes that these developments will increase the relative share of fixed costs in weapons systems and indeed in the structure of the military. His data review supports his arguments, and points to the consequent disadvantage these trends have for smaller and poorer nations. The increasing costs of sophisticated weapons systems will likely mean that the spending gap between the smaller NATO allies and their larger counterparts, and especially the United States, will continue to increase.

Setter and Tishler (2004) reach similar conclusions after examining the increasing need for, and sophistication of, integrative technologies that connect different military assets for operational purposes. These technologies are advanced, requiring extensive R&D and having a high threshold for minimum efficient scale; only after considerable investment are such technologies likely to generate high returns. Therefore the authors conclude that only a few large militaries should opt for such systems from an efficiency perspective.

Three broad conclusions emerge from this review of the recent literature on the supply of military goods. The first is that there are multiple sources of market failure in the defence sector. These sources of potential inefficiency are present on the supply and demand side, can arise from a variety of pathologies of information and risk, and have technical, economic, and political root causes. Addressing these deficiencies is difficult, since the very nature of security is political, requires high levels of secrecy, and entails considerable risk.

The second theme emphasizes the critical nature of technology and R&D in the defence sector. The extremely competitive nature of security requires constant innovation and technological improvement, significant R&D expenditures, and increasingly sophisticated weapons and advanced capabilities to integrate them. This dimension of the supply problem means that efficient levels of investment, acquisition and deployment can only be obtained for very large military organizations. These technologies have not yet been “scaled down” in a way that makes them accessible to smaller countries in an affordable manner.

Third, there is not a lot of literature or specific analytical modelling that assists decision making with respect to procurement decisions and force structures. The work of Hildebrandt (1999) points us in possible directions, but these approaches do not seem to have been pursued systematically. Consequently we have little analytical basis for choosing one weapons system over another, choosing one structure of force inputs over another, identifying optimal combinations of military inputs or forces, or predicting the overall effectiveness of military capacity vis a vis opponents (especially when there are either new weapons systems employed, or new tactical innovations). This gap remains a potentially serious deficiency.

Some of these problems could theoretically be addressed by a more co-operative international security environment that reduced the need for both secrecy and rapid innovation. The existential nature of security, however, makes such a cooperative arrangement highly unlikely, which is an obvious constraint on addressing some of these supply side pathologies.

A second natural solution to some of the identified difficulties, especially those arising out of economies of scale and high fixed costs, is indeed to be found within an alliance structure. The

cooperative nature of a security alliance such as NATO permits both greater information sharing as well as collaboration on weapons platforms that can allow smaller members to be part of a larger system in an efficient manner. The alliance structure also permits some opening up of the defence procurement market to firms in other NATO member countries. These options for improved effectiveness will be explored in more detail later in this report.

4. The demand for defence expenditures during periods of austerity.

It is difficult to organize the literature according to the categories used in this review. Some studies look at the procurement market, thereby incorporating both supply and demand features. Alliance theory in general incorporates many of the studies focused on the demand side of the market since its theoretical implications deal with the expenditure behavior of alliance members. This section focuses on a few remaining papers of specific significance to the demand for military goods, and changes in that demand.

One strand of the defence demand literature deals with organizational and processes for the budgeting for and tendering of defence procurement. For example Melese, Blandin and O’Keefe (2005) identify the specifics of US government processes for defence spending. As many of these studies are of bureaucratic procedures specific to individual countries, we will not review them here other than to note that these process and management-related studies do exist for some NATO members.

Turning to the more typical economic papers in this area, Murdoch and Sandler (1982, 1984) provide classic studies that examine alliance theory and the joint-products model hypotheses by estimating country defence expenditure functions. It follows in the tradition of the literature reviewed in section 2 of this review, and its results conform to the standard narrative described above. These models and estimating procedures have been applied in other circumstances such as Japanese and American defence spending (Okamura, 1991). In some of these studies the restrictions of alliance-driven hypotheses and interpretations are less pronounced and the findings often emphasize external threats as the key factor driving defence spending (Amara 2007, 2008).

This basic estimation structure is the basis for most demand-related studies. Of specific interest here are the conclusions that they generate related to the budgetary pressures faced by governments, as other structural and strategic factors have been reviewed in section 2 above. A few recent papers examine this question as a consequence of the recent financial crisis and its associated pressure on government expenditure. Hartley and Solomon (2009) try to anticipate the implications of budget cuts for NATO member defence budgets, defence industrial policies, and contributions to NATO’s budget by first reviewing the relevant economic forecasts for NATO member countries. They argue that defence budgets and contributions to NATO will be affected, but probably not severely, as these are determined by a much wider range of factors such as the threat environment and country-specific factors. They also highlight the fact that NATO’s operations in terms of their strategic doctrine and missions can be modified to encourage more “private” benefits that are more resistant to cuts, though as they point out this

shift may be at the expense of more “public” NATO activities. They also hold out some promise that budgetary pressures may help to reinforce the arguments for doing away with wasteful defence industrial programs that favour domestic defence producers over potentially more efficient international suppliers. They further suggest that alliance members may focus more on defence “outputs” rather than “inputs”, and thus force a more serious consideration of alternative input combinations and opportunities for substitutions between different defence elements (e.g. reserves versus regular forces). They also identify additional potential sources of efficiency through closer collaborations between alliance members, for exploiting synergies provided by some weapons systems, and by taking more advantage of specialization and comparative advantage. While not embedded in direct empirical estimations, the paper draws out lessons from past studies to highlight the many sources of inefficiency that arise from the behavior of NATO allies, and raise the idea that the pressure for austerity arising from the global financial crisis should be taken advantage of to try and overcome some of these enduring, wasteful, practices. Since many of these studies appeared fairly quickly after the financial crisis, and thus lacked an empirical base for analysis, it might be a propitious time to address this shortcoming by examining the early responses of NATO countries to recent budget pressures.

Keller (2010) also examines the effects of the economic crisis on NATO budgets, focusing on the European members. Like Hartley and Solomon (2009) it is more an identification of potential policy options than a data-driven examination of actual practice. He similarly identifies options such as pooling, sharing and specialization as ways of raising the efficiency of defence expenditures. He defines pooling as the explicit identification of a joint force to which members will contribute specific components. He argues that pooling is the preliminary manifestation of the same logic that leads to sharing and specialization. Specialization allows NATO allies to focus on their comparative advantages in defence, thereby reaping economies of scale in procurement, training and operations. To be effective, however, allies must then share their military assets to permit the proper combination of capabilities to perform a specific task. Importantly, however, Keller goes on to identify the critical hurdle to such a program of integration: credible commitments to sharing. In Keller’s words: “So while such a specialization would greatly help to save costs, it requires a reliable political arrangement of shared sovereignty, command, and trust that is very tricky to establish” (2010: 113). We will return to this problem at the end of this section.

The first examination of the actual effects of austerity on defence spending is provided by Larrabee et al. (2012). These authors review the planned defence cuts and changes in defence priorities for the United States and several key European NATO members. They first note that the planned cuts are significant, and driven primarily by budgetary pressures rather than any exogenous change in the security environment. They note that while the reduced and redirected US military spending will put pressure on the European NATO members to take on a greater share of military operations in that region, projected cuts will leave them ill-equipped to meet these obligations. Consequently NATO will be hard pressed to meet its primary security obligations, let alone conduct missions further afield with fewer direct benefits. As with several other papers, Larrabee et al. (2012) identify several options for meeting the challenges of

austerity. These include: pooling and sharing (as defined above); leapfrogging (the shifting of resources away from older capabilities and towards new types of capabilities, while cutting expenditures overall); the use of informal ad hoc coalitions (which avoids the need for consensus in NATO but still requires significant interoperability amongst coalitions of the willing and capable); pre-emptive crisis management (to prevent foreseeable crises from requiring subsequent large-scale intervention); and the increased use and formalization of bilateral and plurilateral cooperation agreements (such as the UK-France partnership, but also Germany and the Baltic states). ;

Finally, Sandler and Shimizu (2014) extend their standard empirical analysis to consider the implications of pressures for budget cuts, which they consider to be extensive both in terms of the magnitude of necessary austerity and the number of key alliance members who will face it. They reach many of the same conclusions as Larrabee et al. (2012). The fact that fiscal deficits are particularly problematic for the core European NATO members and the United States is of particular concern in terms of NATO's military capacity. The so called "pivot" of the United States towards Asia, and the concomitant de-emphasis of Europe in American foreign policy priorities, is also identified as a potentially important impediment to NATO's effectiveness. The authors suggest that one policy response is for a two-tiered NATO framework with the United States responsible primarily for North American security, and the UK and France sharing the lead for defence in the European theatre. In this structure they suggest that European states may become more interested in building greater compatibility and complementarity in their military structures, including shared weapons platforms.

The literature examining the demand for defence expenditures under financial restraint, though not large, is fairly unified in its emphasis on where to find efficiencies in expenditure to sustain security at a lower cost. While the search for such efficiencies is always desirable, the global financial crisis adds immediacy and seriousness to the effort. Almost all of the proposed solutions point to standard economic source of efficiency: specialization, comparative advantage, pooling and sharing, and collaboration to promote economies of scale and synergy.

Unfortunately the empirical evidence about the success of these policy options is limited. The absence of a systematic empirical record and associated analysis is a serious gap in our knowledge. There are several missions from which lessons could be drawn, however, including the Balkan wars of the 1990s, Afghanistan, Libya, Somalia, Syria and Iraq. Arguably, however, there are relevant lessons about force cooperation from a long history of joint operations that stretch back to the First World War and beyond. Despite the relevance of these missions we are having to rely primarily on the theoretical identification of policy options rather than those arising from actual field experience.

Indeed the scant anecdotal evidence for some of these cooperative strategies at the tactical level are not encouraging, and relate directly to Keller's identification of the need for credible political commitment. Perhaps as important as political commitment is the need for integration at the command level that allows resource deployment to fit operational imperatives, rather than purely national ones. There are many longstanding grievances surrounding the alleged

asymmetry of treatment of multiple national forces under the command of one nation's general. The deflection or withholding of resources from one national force to the advantage of another is not a new phenomenon, but despite its importance it has received scant attention in the empirical literature regarding its extent or cures. Key questions about the operational prioritization for the deployment of scarce military resources (who will have helicopter support? Where will drones be sent for intelligence gathering?) need to be resolved to the satisfaction of multiple players. One consequence is that excessive specialization may actually be problematic, as it might facilitate actual or perceived asymmetry of treatment. By contrast closer integration may reduce these opportunities, though potentially at the expense of force effectiveness.

In addition to the problem of potential commander bias at the tactical and strategic levels is simply the absence of easy inter-operability. In modern warfare, forces operating with either different rules of engagement or different interpretations of the laws of war may be difficult to integrate. Ultimately one nation's rules will tend to dominate, which effectively subordinates the armed forces of another.

5. Conclusions

Briefly, the literature on defence spending in alliance frameworks generally, and NATO specifically, is fairly well developed. There is a broad understanding of the many forces that shape defence spending generally and within the context of alliances. There are, however, several gaps. For ease of reference the basic findings and gaps are identified in Table 1 below.

The first key gap deals with the supply side of defence spending. Despite the likelihood of significant savings to be had from more open and pooled procurement policies, the practical lessons on how to achieve this greater integration and the consequent benefits of such programs remains relatively underdeveloped. There have been instances of joint weapons system acquisition: these need to be studied in some detail to learn from their successes and failures.

The second key gap is again on the operational side. The theory and policy literature on the demand for defence spending all point to the same sources of efficiency gains. These policies inevitably require closer collaboration amongst alliance members at the strategic and tactical levels. However there is ample evidence that such policies also contain their own pathologies, and these have not been systematically studied or widely acknowledged. While promising, the adoption of more cooperative military structures may well pose serious challenges in terms of national sovereignty, and in terms of operational effectiveness.

Table 1: Key findings and gaps.

Subject area	Key Findings	Gaps or criticisms
Alliance Theory	<ol style="list-style-type: none"> 1. Military expenditures provide public as well as private benefits to alliance members, and the balance of these shapes defence spending patterns and levels of burden asymmetry. 2. While alliance theory is a powerful lens for examining military spending, there are alternative theories that also provide insight and which reflect more closely the realities of policy making. 	<ol style="list-style-type: none"> 1. The estimation of burden asymmetries is compromised by the imprecise definition and operationalization of alliance benefits. 2. There has been only limited efforts at reconciling different theoretical approaches to understanding defence spending
Supply	<ol style="list-style-type: none"> 1. The market for military goods is far from “perfect” in the sense of suffering from demand side failures (monopsony) and supply side failures (imperfect and asymmetric information, limited competition, large economies of scale and barriers to entry and exit). 2. Weapons systems are under very strong pressure to deliver innovation to ensure superiority, and is consequently highly dependent on extensive research and development to deliver complex systems that very few militaries can afford. 3. There is limited understanding of how to identify efficiencies in the choice of military inputs and force structure, especially given the rapid rate of innovation for new weapons systems. 	<ol style="list-style-type: none"> 1. There is a serious gap in the modelling of efficient procurement policies given the extensive market failures in the defence sector. 2. There is a serious gap in our capacity to evaluate the trade-offs between different force structures and military inputs, and the effectiveness of these in terms of delivering victory over opposing forces.
Demand	<ol style="list-style-type: none"> 1. Financial pressures will cause NATO members to adjust defence spending, probably in a negative way though the evidence remains sparse in terms of the current fiscal environment. 	<ol style="list-style-type: none"> 1. There is a need to update models of defence spending under budgetary pressure using more recent data. 2. We need to identify how to take advantage of efficiency

	<p>2. Pressures for more efficient spending may induce governments to compensate through greater pooling, sharing, specialization and other forms of collaboration. However there are serious impediments to pursuing these policy options due to the detailed problems of ensuring proper force integration and balance, especially access to resources that are not owned nationally and thus integrated fully in with a country's military.</p>	<p>enhancing collaborative options in alliance procurement in a manner that both ensures military effectiveness and assures member countries that they will have proper access to the alliance-wide resources that are being coordinated.</p>
--	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

References

1. Amara, Jomana. (2007) 'Evaluating NATO Long Run Defense Budgets Using Unit Root Tests,' *Defense and Peace Economics* 18(2) 157–181.
2. Amara, J. (2008). "NATO Defense Expenditures: Common Goals or Diverging Interests? A Structural Analysis," *Defence Economics*, Vol 19, Issue 6, pp. 449-469.
3. Gates, W. and Terasawa, K. (2003) "Reconsidering publicness in alliance defense expenditures: NATO expansion and burden sharing," *Defence and Peace Economics* Vol. 14(5), pp.369–383.
4. Gonzales, Rodolfo A. and Stephen L. Mehay. (1990). "Burden Sharing in the NATO Alliance: An Empirical Test of Alternative Views," *Public Choice*, Vol.68 (3), pp. 107-16.
5. Hartley, K. and Sandler, T. (1999). 'NATO Burden Sharing: Past and Future,' *Journal of Peace Research*, Vol. 36(6), pp. 665–680.
6. Hartley, K. and B. Solomon. (2009). "NATO and the Financial and Economic Crisis," *NATO Defence College*, Issue 52 (October).
7. Hildebrandt, G.G. (1999). "The Military Production Function," *Defence and Peace Economics*, Vol. 10 (3), pp. 247-72.
8. Hilton, Brian and Anh Vu. 1991. "The McGuire Model and the Economics of the NATO Alliance," *Defence Economics*, Vol. 2 (2), pp. 105-21.
9. Hurley, W.J. (2005) A Clarification of the Concepts of Force Multiplier and Returns to Force Scale, *Defence and Peace Economics*, Vol. 16(6) 463-465.
10. Keller, P. (2010). "Challenges for the Defense Budget after the Economic Crisis: A European View," Adenauer Foundation, Berlin, online at <http://www.kas.de/wf/doc/5177-1442-5-30.pdf>
11. Khanna, J. and Sandler, T. (1996). "NATO Burden Sharing: 1960-1992," *Defense Peace Economics*, Vol.7(2), pp. 115-33.
12. Khanna, J. and Sandler, T. (1997). "Conscription, Peacekeeping, and Foreign Assistance: NATO Burden Sharing in the Post-Cold War Era," *Defence and Peace Economics*, Vol. 8(1), pp. 101-21.

13. Khanna, J., Sandler, T., and Shimizu, H. (1998). "Sharing the Financial Burden for UN and NATO Peacekeeping: 1976-96," *Journal of Conflict Resolution*, Vol. 42 (2), pp. 176-95.
14. Kirkpatrick, David. (2004) "Trends in the Costs of Weapons Systems and the Consequences," *Defence and Peace Economics* 15(3), p. 259- 273.
15. Knorr, K. (1985). "Burden-Sharing in NATO: Aspects of US Policy," *Orbis* 29:3, pp. 517- 36.
16. Larrabee F.S., S. Johnson, J. Gordon IV, P. Wilson, C. Baxter, D, Lai, C. Trenkov-Wermuth. (2012). *NATO and the Challenges of Austerity*, (Arlington, V.A: RAND Corporation).
17. McGuire, M.C. (1990). "Mixed public-private benefit and public-good supply with application to the NATO alliance," *Defence Economics*, Vol. 1(1), pp. 17-35.
18. Melese, F., Blandin, J., and O'Keefe, S. (2005). A New Management Model for Government: Integrating Activity Based Costing (ABC), the Balanced Scorecard (BSC), and Total Quality Management (TQM) with the Planning, Programming and Budgeting System (PPBS)," *International Public Management Review*, Vol. 6(1).
19. Morrow J., D. (1991). "Alliances and Asymmetry: An Alternative to the Capability Aggregation Model of Alliances," *American Journal of Political Science*, Vol. 35(4), pp. 904-13.
20. Murdoch C., J. " Alliances: Theories and Empirics," in Hartley, K. and Sandler, T. (Editors) (1995). *Handbook of Defense Economics*, Vol. I (Amsterdam: Elsevier Science B.V), chapter 5, pp. 89-108.
21. Murdoch J., C. and Sandler, T. (1982). "A Theoretical and Empirical Analysis of NATO," *Journal of Conflict Resolution*, Vol. 26 (2), pp. 237-63.
22. Murdoch, J.C. and Sandler, T. (1991). "NATO burden sharing and the forces of changes: Further observations," *International Studies Quarterly* 33(1), 109-114.
23. 32. Murdoch J., C. and Sandler, T. (1984). "Complementarity, Free Riding, and the Military Expenditures of NATO Allies," *Journal of Public Economics*, Vol. 25 (1-2), pp. 83-101.

24. Okamura, Minuro. (1991). "Estimating the Impact of the Soviet Union's Threat of the United States-Japan Alliance: A Demand System Approach," *Rev. Econ. Statist.* 73:2, pp. 200-207.
25. Olson, M. and Zeckhauser, R. (1966). "An Economic Theory of Alliances," *Review of Economics and Statistics*, Vol. 48 (3), pp. 266-79.
26. Oma M., I. (2012). "Explaining states' burden sharing behavior within NATO," *Cooperation and Conflict*, Vol. 47, Issue 4, pp. 562-573.
27. ONeal, J.R. and M.A., Elrod (1989). "NATO Burden and the Forces of Change," *International Studies*, Vol. 33 (4), pp. 435-456.
28. Palmer, Glenn. 1990a. "Alliance Politics and Issue Areas: Determinants of Defense Spending," *Amer. J. Polit. Sci.* 34:1, pp. 190-211.
29. Palmer, Glenn. 1990b. "Corralling the Free Rider: Deterrence and the Western Alliance," *Int. Stud. Quart.* 34:2, pp. 147-64.
30. Ringsmose, J. (2010). "NATO Burden-Sharing Redux: Continuity and Change after the Cold War," *Contemporary Security Policy*, Vol. 31 (2), p.319-338.
31. Rogerson P., W. "Incentives Models of the Defense Procurement Process," in Hartley, K. and Sandler, T. (Editors) (1995). *Handbook of Defense Economics*, Vol. I (Amsterdam: Elsevier Science B.V), chapter 12, pp. 309-346.
32. Sandler, T. (2005). "NATO Burdens, Benefits, and Borders: Comment," *Defence and Peace Economics*, Vol. 16(4), pp. 317-321
33. Sandler, T. and Forbes, J.F. (1980). "Burden Sharing, Strategy, and the Design of NATO," *Economic Inquiry*, Vol. 18 (3), pp. 425-444.
34. Sandler, T. and Hartley, K. (1995). "Procurement: Theory, Evidence and Policies," Chapter 5 in Sandler, T. and Hartley, K. *The Economics of Defense* (Cambridge: Cambridge University Press), pp. 113-155.
35. Sandler, T. and Hartley, K. (2001). "The Economics of Alliances: The Lessons for Collective Action," *Journal of Economic Literature* (September), Vol. 39 (3), pp. 869-896.
36. Sandler, Todd and James C. Murdoch. 1990. "Nash-Cournot or Lindahl Behavior?: An Empirical Test for the NATO Allies," *Quart. J. Econ.* 105:4, pp. 875-94.

37. Sandler, T. and Murdoch C., J. (2000). "Defense Burdens in the 1990s and Beyond," *Fiscal Studies*. Vol. 21(3), pp. 297-327.
38. Sandler, T. and Shimizu, H. (2014). "NATO Burden Sharing 1999-2010: An Altered Alliance," *Foreign Policy Analysis*, Vol.10 (1), pp. 43-60.
39. Setter, O. and A. Tishler (2004). The Role of Integrative Technologies as a "Force Exponent" on Military Capability. (<http://carecon.org.uk/Conferences/Conf2004/Papers/Setter.pdf>).
40. Skogstad, K. (2014) The Effectiveness of Canada's Navy on Escort Duty. Unpublished mimeo, at <http://www.economichistory.ca/pdfs/2014/skogstad.pdf> accessed 2/October/2015.
41. Solomon, B. (2004). "NATO Burden Sharing Revisited," *Defence and Peace Economics*, Vol. 15 (3), pp. 251-258.
42. Solomon, B. (2005). "NATO Benefits, Burdens, and Borders: Reply," *Defence and Peace Economics*, Vol. 16 (4), pp. 323-326.