

GOODHART’S LAW AND THE GAMING OF UK PUBLIC SPENDING NUMBERS

Christopher Hood and Barbara Piotrowska

Nuffield/ESRC History of UK Public Expenditure Research Project¹

Blavatnik School of Government

University of Oxford

Word count: 6957

¹ We are grateful to the Nuffield Foundation and the ESRC for funding the research on which this paper is based (Grant reference OPD/43172). We acknowledge extremely helpful comments and suggestions from Maia King, Andy King, Gwyn Bevan and David Heald, as well as the participants of the January 2019 National University of Singapore workshop “Gaming in Performance Management Reforms and Its Countermeasures”

Abstract

Goodhart's Law, originally inspired by money-supply indicators, predicts high-consequence administrative numbers tend to be gamed out of meaningfulness. This paper argues that, in addition to the well documented manipulation of aggregate input numbers at the top-levels of decision making and performance indicators used for *output* control at the lower level, meso-level gaming happens also on the *input* side of public expenditure planning and control. A key mechanism through which it operates is creative categorization in the classification of units of spending. Based on a UK study, we explore three questions: can we find evidence for the existence of gaming in public expenditure control, how does the creative categorization work and how material or consequential is it? Using case studies of "protected" spending, public-private partnerships and accounting changes, we show that gaming understood as creative categorization is readily observable in UK public expenditure control and Goodhart's Law effects can indeed be material or consequential, both in scale and in their implications for government accountability. We conclude that creating new spending categories to control public expenditure and limit gaming is a two-edged sword, since it itself creates new opportunities for gaming.

Keywords: public finance management, public spending, gaming, Goodhart's Law, fiscal illusion, PFI, resource accounting, creative categorization

Introduction: three questions about gaming in public spending input controls

The management of public spending, a central process of public management in any political system, involves control through numbers. These numbers are applied at different levels of policy making, including: macro-level (overall spending aggregates), meso-level (spending allocations to individual departments or policy domains), and micro-level (local authorities and satellite agencies). And those spending control numbers can focus on inputs, outputs or throughputs.

For numerical output controls at meso-level, in forms such as performance targets or rankings accompanying budget allocations in fields such as healthcare or education, a substantial literature points to the susceptibility of such controls to gaming behaviour in the sense of tactical manipulation by players, often producing Goodhart's Law effects of making the relevant control numbers less meaningful, to be discussed in the next section.

This paper focuses instead mainly on the meso-level *input*-control side of public financial management, the allocation of resources to different departments or policy areas, to examine three questions. First, can gaming and Goodhart's Law effects be readily detected in that sphere as well? Second, how exactly does gaming produce those Goodhart's Law effects? And third, how substantial, material or consequential are those effects?

There is an existing literature on gaming of input-control numbers in public expenditure at macro level, which focuses on comparative analysis of internationally-reported data on "stock-flow adjustments" as a proxy for gaming. In this paper we focus instead on the meso-level of input control numbers in public financial management, using in-depth qualitative

analysis to answer the three questions we set out earlier. Accordingly, we focus here on a single-country case, that of the UK, and draw on qualitative material in the form of published and official documents, over 100 semi-structured interviews of key players and observers of the system over two decades, and three specific case studies of categorical changes. Our analysis provides proof of existence of gaming and Goodhart's Law effects on the input side of meso-level public spending control in that country, identifies the key mechanisms through which the gaming operated and enables us to assess how material or consequential the effects of such gaming were in the case studies described later. The UK is an important case for such an analysis, since it is a developed democracy that is commonly rated fairly highly in conventional international assessments of the quality of public financial management such as the IMF Fiscal Transparency Tool and in comparative perspective it does not show particularly strong signs of gaming in its stock-flow adjustment numbers. We are certainly not suggesting, however, that gaming of public spending input numbers is by any means a UK-specific phenomenon.

We argue that meso-level input control of public spending is linked to creating categories of spending that are intended to be treated in different ways, and that gaming and the associated Goodhart's Law outcomes are a result of creative use of those categories. Some of the categories designate types of spending considered to be broadly desirable (such as capital expenditure), while others denote items considered less desirable (such as levels of debt or administration costs). Such categories work in a manner similar to output performance targets where, depending on whether any given category denotes spending that is considered desirable or less desirable, there will be incentives for over- or under-reporting the levels of spending within that category. We show how these effects operated in the predicted direction in three case-study examples, show that the consequences of the number gaming were not

trivial either in the amounts of money involved or in their implications for accountability, and in the concluding section we reflect on the implications of these findings for the understanding of input controls in public expenditure.

Literature review

Charles Goodhart's eponymous "law," originally formulated in the mid-1970s in relation to different target measures of the money supply, states: "any observed statistical regularity will start to collapse once pressure is placed on it for control purposes" (Goodhart, 1975/1981, p.116). Others have glossed or re-stated Goodhart's claim, for instance Marilyn Strathern (following Hoskin, 1996): "When a measure becomes a target, it ceases to be a good measure." (Strathern, 1997, p.308). The effect Goodhart and others describe comes about at least in part as a result of a change in incentives for reporting the relevant measures, leading in turn to number gaming in the form of creative reporting.

Number gaming in public administration

The term "gaming" did not appear in Charles Goodhart's original formulation of his "law," and that particular word seems to have come – or come back - into currency in economics, management and public administration in the early 2000s to denote *tactical behavior designed to secure benefits of some kind by getting around rules or exploiting fuzziness in categorizations*, or more succinctly "hitting the target and missing the point" (Bevan and Hood, 2005). But recognition of the phenomenon both pre-dates that particular term and is reflected in other related words or phrases such as "creative accounting," "creative compliance," "spin," "gimmickry," "rule bending," or "fiscal illusion."

The various terms describing different guises of the same phenomenon indicates that, while ubiquitous, the term “gaming” does not have only one universally accepted definition. Rather it seems to be one of those terms that are frequently used but rarely carefully defined. Indeed, as with all matters relating to propriety or appropriateness, who counts what as gaming (as opposed to creative “workarounds,” legitimate fudges or convenient interpretations of rules to advance deserving causes that feature in the literature on rule bending), can be expected to be socially variable. (See, for example, Maynard-Moody & Musheno, 2003; Spreitzer & Sonenshein 2004; DeHart-Davis, 2007; Portillo, 2012; Tummers, Bekkers, Vink, & Musheno, 2015.)

The classic example of the way number gaming is understood in the context of performance targets is a famous if perhaps apocryphal 1950s cartoon in the Soviet magazine “*Krokodil*”, depicting a Soviet nail-making factory producing one single enormous nail to meet its annual production target, because that target was set in tons (Nove, 1980, p.97). That case exemplifies gaming because the outcome is not what the target-setters intended and is evidently contrary to the spirit of the planning system, but it is neither an outright fabrication (the nail is not imaginary) nor an interpretation that is explicitly forbidden.

Goodhart’s Law and gaming in public expenditure

Indeed, Goodhart’s Law is part of a broader literature on tactical interpretations or applications of administrative numbers, some of it preceding Goodhart’s own work. Within the sphere of public expenditure, the two main applications in that gaming literature apply to input control of the very top-level aggregate numbers and output control in the form of

performance targeting at the meso-level, rather than the input controls at meso-level that this study focuses on.

As for analysis of gaming of input numbers at the macro level of public spending, Arnout Wellink suggested over twenty years ago that Goodhart's Law was applicable to the public expenditure domain (1996). And in the wake of the fiscal rules established in the 1997 EU Stability and Growth Pact, a number of comparative studies explored the relationship between the stringency of fiscal rules and gaming of fiscal numbers, in several cases using "stock-flow adjustments" (SFAs) as a proxy for the extent of gaming (or "gimmickry"). SFAs are a standard Eurostat-reported measure computed as annual changes in debt levels minus annual budget deficits, with large negative scores assumed to be indicative of attempts to contain debt without reducing spending, and large and consistently positive scores indicative of efforts to hide spending (Irwin, 2015)². For instance, Jürgen Von Hagen and Guntram Wolff (2006) used SFAs as an indicator of creative accounting adopted by governments to hide their deficits to meet Stability and Growth Pact rules. Vincent Koen and Paul Van den Noord (2005) showed that counting the number of one-off fiscal measures and classification errors by national governments led to the same conclusion. And James Alt, David Lassen and Joachim Wehner (2014) used SFAs both to assess the degree of gaming in the context of the Stability and Growth Pact and to explore other determinants of such gaming.

² Such interpretations arise from the assumption that there is no reason to expect a systematic bias in the relationship between the stock of debt outstanding and the sum of all budget deficits over time. Hence the attraction of taking persistent SFA adjustments and a large difference between debt stocks and accumulated deficits as indicating gaming in some form.

However, an analysis of UK SFAs reveals why they are an imperfect proxy for gaming.

Figure 1, based on available Eurostat data (2006-2017), shows SFA levels for the UK from 2002 to 2016, as compared with the other EU member states over that period. The UK SFA numbers co-vary with the broader European trends and their fifteen-year average is close to zero (0.47) (Figure 1). This observation reinforces the positive judgement of the reliability of UK public finance numbers presented in evaluations, such as the IMF Fiscal Transparency Evaluation.

[Figure 1 near here]

The problem with using SFAs as a proxy for gaming is that deliberate and transparent policy changes have to be distinguished from gaming and the former tended to dominate the UK's SFA scores shown in Figure 1 during the global financial crisis of 2008 and subsequently.

Both UK and EU SFAs rose sharply around the financial crisis, not because of a sudden explosion of gaming behavior as defined earlier, but because the governments concerned chose to bail out their collapsing banks by buying financial assets in the form of bank shares (which added to government debt but not deficit). And the UK's SFAs rose more than those of the EU countries as a whole because the UK had a larger banking sector and its bailout costs were correspondingly higher. Equally, when those bailout costs started to be recouped a few years later, they were reflected in notably negative UK SFA scores. Hence, particularly in a crisis and its aftermath, SFAs are at best a doubtful proxy for gaming. This approach therefore needs to be at least complemented by more qualitative analysis of gaming.

As for the meso-level of public spending control numbers, there is a broader literature on gaming that is mostly restricted to the evaluation of output or outcome controls that are linked in some way to resource allocation. Literature in that vein, mostly exploring the

vulnerability of performance indicators to gaming, includes, for instance, studies of the operation of the target system in the former USSR and other centrally planned economies before the fall of the USSR (such as Berliner, 1957; Nove, 1957; Grossman, 1960; Harrison, 2011); work on responses to wartime target systems for arms production, such as the US aircraft production in the 1940s (for example, Holley, 1989, p. 242); and more recent work on strategic manipulation of high-consequence public service performance indicators in Western developed countries, for example when job-placement agencies or contractors blur the distinction between “new jobs” and “existing jobs” (Stephan 2016) or when hospitals blur the boundary between admission to the emergency room and the commencement of treatment to game waiting-time targets (McDermott 2011; see also Bevan and Hood, 2005; Hood, 2006). Such gaming behavior leads to Goodhart’s Law consequences of making the relevant performance numbers less reliable or less consistent over time.

However, there is less attention in the literature mentioned above to gaming on the input side of public expenditure control at the meso-level, and accordingly this paper focuses on analysis of that area of spending control. As mentioned earlier, we use a country case to show that gaming at the meso-level can be found on the input side of spending control as well as the output side (even in a country where standard quantitative measures in the form of SFA analysis, do not indicate high levels of gaming at the macro level). We also show how the gaming works (through creative use of categories intended to improve spending control) and assess its scale and Goodhart’s Law effects.

Creative categorization: the source of input gaming in public expenditure

Public expenditure control can be approached by targeting outputs or inputs. As we have seen above, targeting outputs and performance measures has been proved to be associated with

gaming. However, the management of public spending inputs at meso-level (hereafter simply “public spending” or “public expenditure”) involves more or less elaborate classification systems, designed to control and shape that spending in various ways. Objectives for spending control at this level include prioritization and allocation of resources (distributing funds among different domains and departments to reflect government priorities) and efficient service delivery (aiming for the best levels of service provision achievable within available resources) (Schick, 1998). To achieve such goals, rules are developed that link to categorical distinctions. Table 1 shows ten examples of commonly-observable examples of such distinctions.

[Table 1 near here]

In this paper, informed by the literature discussed earlier, we consider gaming in public expenditure control as a sub-set of rule-challenging or rule-bending behavior that is:

- not unambiguously illegal
- not unambiguously pursued for personal gain
- focused on the way numbers are calculated rather than other ways of countering rules or conventions.

By “not unambiguously illegal”, we mean strategic behavior that may counter the spirit or original intentions of policies or procedures, but which is either legal or considered at least by some to fall within the “grey” category of “legitimate fiddles” or “honest graft” to be found in most social settings. (“Honest graft” is the famous phrase used by the early twentieth century “Tammany Hall” boss in New York, Senator George Washington Plunkitt (Riordan, 1904/1963).) In common parlance, gaming tends to be distinguished from outright lying or clear fabrication of numbers. It can therefore be considered more like tax avoidance - “creative compliance” (McBarnet and Whelan, 1991) that exploits legal or quasi-legal

loopholes to minimize tax liability - than “tax evasion” in the sense of clear-cut violation, for example by outright lying about income or profits.

By “not unambiguously pursued for personal gain” (another Plunkitt theme), we mean strategic behavior that may well help or protect the individual careers of those involved, but largely arises in the pursuit of organizational or policy interests. That contrasts with the focus of Gerald Mars’ (1982, p.137) classic study of workplace cheating which explored “fiddle proneness” defined as “any work context where a propensity exists for a job to offer regular material rewards that are excluded from formal accounts or which are included under ambiguous or deceptive headings.” Here we are not concerned with such micro-level fiddles but rather with ways of getting more funding for preferred policies or “flattering” public spending numbers to enhance or protect government or organizational reputations.

Finally, we focus on gaming that centers on the way numbers are calculated or presented rather than other kinds of tactical behavior. After all, strategic behavior in public finance occurs in multiple guises, of which finessing numbers is only one. For example, bargaining tactics that concentrate on trying to “bounce” the Treasury or Finance Ministry (such as a spending minister calling the Finance ministry with an urgent political demand for extra funds on the eve of a public holiday, or publicly announcing extra spending for particular policies on TV as a *fait accompli* without prior consultation with the Finance Ministry) might be interpreted as tactics designed to get around normal procedures for considering extra spending proposals as a “gathered field” all at once. But we can distinguish those kinds of bargaining tactics from gaming of the numbers themselves, and it is the latter that we focus on here.

For public expenditure control, the key mechanism in gaming so defined is that of creative categorization, which is a process of placing units of spending in more rather than less favorable categories for the player(s) concerned, following the old British Treasury joke that “there is no fiscal crisis that cannot be dealt with by swingeing classification changes”. The potential for creative categorization in public spending control arise because the handling of public spending numbers ultimately comes down to assigning cases (units of spending) among different categories (types of spending, subject to different rules or treatments) by making qualitative judgements. Because those spending categories are ultimately described in natural language rather than machine code (including the basic distinction between what is and is not to count as “public spending”), the inherent plasticity of such language produces scope for blurring or jumping of categories when it is advantageous to do so. (A similar phenomenon is observable in the case of output performance information: see Moynihan, 2008, p.106).

Blurring the boundaries: evidence of creative categorization in the UK

Having identified creative categorization as a – or the - key mechanism in the gaming of public expenditure input control numbers, we now explore whether and how that mechanism operated in the UK between the early 1990s and the mid-2010s.

Table 2 identifies commonly-alleged forms of creative categorization observable over that period in the UK, corresponding to each of the ten categorical distinctions shown in Table 1. In the right hand column it also gives examples of each of those ten types.

[Table 2 near here]

The following section goes beyond the summary examples shown in Table 2 to develop a theory of how gaming through categorical categorization works for meso-level public spending control and then applies that theory to three specific examples.

Theory of tactical categorization: how number gaming operates in public spending planning and control

To understand how categorization of spending can produce Goodhart's Law outcomes, consider the difference between the true and the reported value of spending within any given category. If there are no incentives to mis-report, these two values are likely to be the same. But if there are implicit or explicit rewards to demonstrate an increase or decrease of spending in any given category, the reported value of that category of spending is likely to be affected. Specifically, if the category in question represents spending that is seen as such as various types of "protected spending," the incentive will be to report a value that is higher than the true one, by putting as much spending as possible into that category. Conversely, categories that identify types of spending considered to be less desirable, such as administration costs or spending funded by debt, will produce corresponding incentives for underreporting.

How do such processes produce Goodhart's Law effects? Where a measure or categorization precedes the introduction of incentives to over- or under-report, the values, of what is reported before and after the incentivization are likely to become incomparable. In those circumstances, any reported change will include both unobservable changes arising from new incentives for under-or over-reporting as well as any alteration in the true value of the spending in that category. Where a spending category is introduced with incentives to under- or over-report from the very beginning, the reported value is likely to be contaminated by the

incentive effect from the start, but even then Goodhart's Law can apply if the strength of the incentives changes (for instance if the screw is tightened over administration costs).

Case studies of gaming: “tuck-unders”, PFIs and Resource Accounting

To examine and demonstrate how such effects operate in practice, we now explore three specific case studies of meso-level movement of unit of spending between control categories in the UK: namely between unprotected versus “protected” or ring-fenced spending, between spending counted as “public spending” and spending that is off the balance sheet, and between cash and non-cash or resource spending.

Blurring protected and unprotected spending: “tuck unders”

The sixth category in Tables 1 and 2 distinguishes spending that is ring-fenced or protected in some way from spending which is considered of lesser priority and thus liable to be restrained or traded off against other claims on the public purse in budgetary negotiations. Most governments draw some such distinction between more and less protected spending domains, and indeed there have been numerous instances of such spending input protection in the UK over recent decades. One example is defense spending, for which NATO sets guidelines in terms of proportion of GDP to be committed to such spending by its member states (currently set at 2 per cent of GDP). Similarly, since 2005 there has been a floor target of 0.7 per cent of GNI to be spent on overseas aid, to match a UN commitment to development spending originally set in 1970 (Townsend, 2010), and indeed that target figure was written into UK law in 2015 (Anderson, 2015). A third example is a major program of spending restraint for deficit reduction announced in a key Spending Review in 2010, in

which three key domains of public spending (overseas aid, health and education) were to be officially “protected” from the cutbacks imposed elsewhere. Given that designation of some types of spending as protected was meant to ensure a “floor” of spending levels, our analysis above would lead us to expect gaming to bring into the “protected” category expenditure that would not otherwise be reported within the relevant spending domains. And the creative categorization associated with this form of gaming is known as “tuck unders” in UK parlance.

The term “tuck under” denotes an unfunded or only partially funded mandate, a situation where a department’s responsibilities are increased without a corresponding increase in its budget (Colebrook, 2015). In fact, the term can apply to two different situations, depending on whether the “protected” category is spending defined in terms of type of spending or department.

First, if a protected category of spending is not unambiguously linked to one department, there is scope for creative classification in increasing what is counted under the category in other departments, without actually increasing the budget of the principal one. An example is the treatment of the UK Department for International Development [DfID] in the UK’s 2015 spending review. A Treasury document (HMT, 2015) declared that “DfID will remain the UK’s primary channel for aid, but to respond to the changing world, more aid will be administered by other government departments, drawing on their complementary skills.” In 2017 the Minister for International Development told the Parliament that over £100m of government spending would be officially classified as Development Assistance and therefore count towards the 0.7 per cent target (Hansard, 1 November 2017, col 30WS). Such

classification changes allowed the UK government to claim it was keeping its 0.7 per cent commitment - but without increasing DfID's budget in line with the growth in GNI.

The NATO guideline that its member states should spend 2 per cent of GDP on defense invites similar strategic classification moves. For example, in 2016 the House of Commons Select Committee on Defence questioned some of the spending items that had been included as "defense spending" in that year, since revisions to the criteria used to calculate its defense budget as reported to NATO had helped the UK to hit the NATO input target (HC Select Committee on Defence, 2016). Among the items newly included in the sums reported to NATO for 2015 were: provision for war pensions; assessed contributions to UN peacekeeping missions; and pensions for retired Ministry of Defence civilian staff (Chalmers, 2015). The move could be considered gaming in so far as all those changes were within NATO guidelines, and consistent with its rules, but nevertheless conveniently "tucked under" types of expenditure not previously included into the category of "NATO defence spending" (Wright, 2016).

A second type of tuck under involves protecting a department's budget but altering its responsibilities. That happened in 2011 when responsibility for funding the BBC World Service was moved from the (unprotected) Foreign Office budget to that of the (protected) DfID budget – a funding switch that seems to have followed suggestions from a parliamentary committee (BBC, 2011). Something similar happened in 2013, when a "Better Care Fund" was created (Colebrook, 2015), which involved a financial contribution from the (protected) Department of Health NHS budget to the funding of social care, hitherto provided under (unprotected) local government funding. So while the Department of Health's overall

NHS budget was protected, its financial responsibilities were increased, meaning that the “protected” budget was diluted by additional service obligations.

The Goodhart’s Law effect of such “tuck unders” is to make it much harder to use official spending statistics to make valid like-for-like comparisons over time. And such comparisons can matter for accountability: if promises of keeping spending on a given category above a certain floor level turn on reported numbers with incomparable baselines and levels, it becomes correspondingly harder to verify if such promises have been kept.

On and off the balance sheet: the Private Finance Initiative (PFI)

UK government fiscal rules in the 1990s and 2000s targeted, among other measures, the level of debt. In terms of our theory sketched out earlier, spending funded by public debt broadly counted as an “undesirable” spending category insofar as the fiscal rules defined a ceiling for that kind of spending. Consequently, we would expect to see a shift in spending away from spending formally funded by debt towards a spending vehicle that did not count towards debt. And that is precisely what can be observed in the form of a proliferation of “Private Finance Initiative” (PFI) projects. Such projects are an example of gaming in that they constituted a new category of spending that kept the capital cost of major projects off the government balance sheet, thereby blurring the categorical distinction between “public spending” and off-balance sheet spending.

Traditionally major public sector capital projects (such as building government offices, schools or hospitals) were carried out in the UK by government borrowing the money in the financial markets (using its ability to raise money at lower interest rates than private firms, with the resultant debt showing on the government balance sheet) and then managing the

project, typically through contracts with one or more firms. From the early 1990s, an alternative mechanism developed, known as the “Private Finance Initiative” (PFI). PFI differed from traditional procurement for capital projects in that under PFI it was private contractors, not government, who borrowed the money on the capital markets, while government or public sector bodies paid for the assets under long-term contracts (typically 25-30 years) by annual payments that covered the debt interest incurred by the contractor, shareholder dividends, asset maintenance and in some cases other services such as cleaning (National Audit Office [NAO], 2018).

The original justification for PFI was that it could deliver major public capital projects more cost-effectively than the traditional procurement approach, which tended to produce large cost overruns and late delivery (typically as a result of specification changes as projects developed) as well as neglect of assets by cutting maintenance spending when money was short (often meaning that costly remedial spending would be needed later). So if the extra cost of private firms borrowing on the capital markets under PFI was outweighed by the avoidance of cost overruns and late delivery (as a result of transferring the risk of cost overruns and delays from government to PFI contractors), PFI could be considered to offer better overall value for money than traditional procurement (HC Treasury Select Committee, 2011).

However, as noted above, PFIs also provided a convenient way to remove spending off balance sheet. Debt raised by PFI contractors did not count towards the national debt for National Accounts purposes and so did not feature in government debt figures (unlike conventional procurement), and PFI capital investment was not recorded as public spending (NAO, 2018). That categorization issue meant that, irrespective of value for money issues, PFI contracting made it easier for government to meet fiscal rules that targeted levels of

public debt, and for individual departments to avoid large upfront capital costs (Mann, 2012). And it is certainly striking that PFI projects increased sharply in the 2000s, with PFI capital investment rising to a peak in 2007/08 at £8.6bn (NAO, 2018). As a result, lengthy debates ensued about whether the burgeoning use of PFI projects represented genuine value for money or was largely driven by the desire to get around debt limits (See, for example, NAO, 2018), thus representing a gaming response that ought to be countered by changing how debt liabilities were reported. For example, in 2011 the House of Commons Treasury Select Committee argued that “PFI should be brought on balance sheet. The Treasury should remove any perverse incentives unrelated to value for money by ensuring that PFI is not used to circumvent departmental budget limits.” (House of Commons Treasury Select Committee, 2011).

The Goodhart’s Law effect of the growth of PFI certainly meant that figures for government liabilities as recorded on the National Accounts became less meaningful as an indication of the total extent of government debt, with up to £33bn of PFI capital liabilities not included in the National Accounts measure of debt in 2016-17 (Office of Budget Responsibility [OBR], 2018).

In fact, as time went on, the spending categories applying to PFI projects came to be changed with new reporting requirements that eventually brought PFI liabilities into public sector net debt estimates (Chesson & Maitland-Smith, 2006; Office of Budget Responsibility, 2018). Along with these changes in reporting requirements, the use of PFI peaked in 2008 and declined after that (reflecting rising costs of private finance after the financial crisis (NAO, 2018), some dramatic PFI collapses and a sharply declining risk appetite among PFI contractors) to the point where the Chancellor (Finance Minister) announced in a budget

speech in 2018 that PFI would not be used for future projects (HM Treasury [HMT], 2018). But in its pre-2008 heyday, PFI represented a notable case of creative categorization.

The cash/resource boundary: accounting in the Ministry of Defense

A third example of strategic classification arose from the introduction of a new accounting framework into UK government in the early 2000s. The traditional cash accounting system reflected in parliamentary Estimates and Appropriation Accounts was overlaid by a new Resource Accounting and Budgeting (RAB) framework introduced in two stages in the early 2000s for all government departments. The shift, first announced in 1993, aimed to enable spending departments to “evaluate cost of using capital and current resources on an equivalent basis”, “strengthen cash management in departments”, improve accountability and generally make government more “business-like” (Heald, 2005; HMT, 1994). RAB had two main elements. One was to require that costs be recorded when resources were committed or consumed rather than when cash was spent. Second, to provide a better measure of the full economic costs, RAB incorporated non-cash costs including depreciation and a cost of capital charge (the opportunity and financing costs of holding capital). The amount that departments could spend in cash under RAB equaled their resource and capital budgets minus those non-cash costs. When RAB was introduced, departments had to bid separately for “non-cash” budgets to pay for those non-cash costs and the system created incentives for departments to vire non-cash into cash spending by reducing their non-cash costs (Committee of Privy Counsellors, 2016). Returning to the theory of creative categorization sketched out earlier, spending classified as cash was typically preferable to departments than spending classified as non-cash, since the former could be used to fund many kinds of activities while the latter could only be used to pay for capital charges or notional costs of depreciation and the like. Given the greater desirability of cash, we would expect to see attempts to reclassify spending

from non-cash to cash wherever that was allowable. And the case of Ministry of Defense and NHS trusts in the 2000s demonstrate that this is indeed what happened in some major areas of spending.

The 2016 Chilcot Report into the UK government's controversial decision to go to war in Iraq in 2003 highlighted the effect of full RAB (first introduced in the 2002 Spending Review) on the financial situation of the Ministry of Defence (MoD). The accounting changes had a big effect on the MoD because it had a large asset base, and those changes also presented the department with a potential opportunity to increase its available cash expenditure by reducing its non-cash costs such as depreciation and the cost of capital. Indeed, the official Treasury "Settlement Letter" accompanying the MoD's spending allocation after the 2002 Spending Review seemed to promise there would be flexibility in the department's ability to spend money saved from decreasing non-cash costs as cash (Committee of Privy Counsellors, 2016, pp. 507-511).

However, the extent to which the MoD chose to use this flexibility, and the extent of its non-cash cost savings came into question, since in early 2003 the MoD announced a plan to transfer £1bn a year from non-cash to cash in 2003/04, 2004/05 and 2005/06. The Treasury questioned whether MOD had overestimated its non-cash costs (making it easier subsequently to identify and claim savings). Some of those savings were said to be due to "windfall gains or creative accounting", for example as the MOD changed the depreciation profiles on its equipment by delays in bringing new equipment into service. Indeed, the Treasury's analysis of the MOD's planned transfer assigned 35 percent of the sum to accounting changes contrived to show non-cash savings (Committee of Privy Counsellors, 2016, pp. 510-511).

A similar situation occurred a few years later when it emerged that in FY 2005/6 NHS trusts (the provider organizations for public health care services) were hugely overspent, to the extent of more than £1.2bn (Palmer, 2006), despite massive and historically unprecedented increases in health budgets in the early 2000s. Part of this overspend resulted from issues such as rapid cost inflation, a sharp decline in health care productivity and failure to anticipate patterns of healthcare demand (Palmer, 2006). But the cash overspend was also partly caused by strategic classification changes and accounting adjustments (NAO and Audit Commission, 2006, p. 34; see also HC Committee of Public Accounts, 2007). And as with the MoD case, there also appear to have been cases of “non-cash” items being vired into “near cash” to reduce reported deficits: a form of viring between accounting categories that resource accounting in principle allowed, but which when exercised on a substantial scale threatened to undermine the government’s then budget-balance rule which was set in cash. The Goodhart’s Law effect in both cases was to blur the boundary between cash and non-cash spending and again to make over-time comparison more challenging.

These three cases indicate some of the different ways in which creative categorization can apply to input control of public spending. The PFI case, involving the largest amount of spending out of the three, stemmed from a new financial instrument intended to correct for persistent problems with conventionally financed public sector capital projects (namely delays and cost overruns). The “tuck under” case represented a response to a trend towards input controls to protect politically favored spending domains from cutbacks, which at the same time provided opportunities for creative classification over exactly what units of spending were to be protected. The third case involved the introduction of an ambitious and relatively sophisticated new accounting framework intended to lead to better decision-making

and less short-termism in public spending, but which for a time offered to the wily (or desperate) the opportunity to blur the distinction between “non-cash” and “near cash” to enable more short-term spending in cash. Both the new framework and the gaming response had short-term Goodhart’s Law effects in limiting the ability to track spending on a consistent basis over time.

Why creative categorization matters: the costs and benefits of proliferating categorical distinctions

The three cases discussed in the previous section are not isolated examples. All of them reflect a pattern of proliferating categorical distinctions introduced into public expenditure management in the UK over recent decades, all intended to improve public spending control but, in many cases, also offering new opportunities for gaming. Table 3 gives some selected examples of such categorical changes.

[Table 3 near here]

The pattern of proliferating categories in the UK reflects attempts to limit what is perceived to be gaming, for example by changing budget carry-over rules to limit wasteful end-of-financial year expenditure surges or late bill payments. However, the attempts to deal with gaming by proliferating or modifying expenditure categories may be a double-edged sword, capable of creating new kinds of gaming as well as countering older types. In similar vein, Timothy Irwin (2015) argues that manipulation involving unrecognized assets and liabilities can be reduced by recognizing a wider variety of assets and liabilities, but on the other hand means that more assets and liabilities can be mis-measured. A possible case in point is the

different ways countries can hide spending liabilities in their public sector accounting systems: countries that exclude state-owned companies and similar bodies from their central public spending control systems can lose significant liabilities in the accounts of such bodies, whereas countries like the UK, which traditionally takes the whole public sector as the unit of control for public spending, are more likely to hide major spending liabilities by categorizing them as private rather than public debt by blurring the distinction between public and private companies (OBR, 2017).

Gaming of public spending input control numbers by creative categorization has some obvious attractions as a workaround in difficult political and financial circumstances. But as we have shown it can also serve to limit or blunt accountability by the Goodhart's Law effect of making comparisons over time difficult or impossible, as well as creating unfunded mandates (as in the "tuck under" cases).

Indeed (as shown in the cases of PFIs and the non-cash to near-cash transfers in the MoD in the early 2000s), creative categorization can in some circumstances affect the fiscal sustainability of an entire country. The PFI case involved a volume of spending large enough to make the difference between hitting and missing official debt limitation rules, and the defence and health switches from non-cash to near-cash also involved sums large enough to make the difference between hitting and missing the then government's "Golden rule" for budget balance.

Conclusion

This paper aimed to add to the limited literature linking Goodhart's Law to the control of meso-level public spending input numbers (as opposed to other high consequence

performance numbers on the output side, such as targets and rankings for service delivery), by answering the three questions raised at the outset: Is there evidence of gaming of public spending input numbers in the UK, a developed democracy well rated for its public financial management? How did such gaming operate? What were its consequences?

First, we showed that evidence of gaming – defined as creative accounting or finessing of administrative numbers that stops somewhere short of clear-cut illegality – could indeed be found on the input as well as output side of public expenditure control at the meso-level in the UK case. Gaming so defined went beyond the finessing of macro-fiscal rules to satisfy or reassure international bodies or financial markets, as discussed by the papers noted earlier on creative compliance with the rules of the EU Stability and Growth Pact.

Second, in answer to the question of how gaming worked in input control numbers, we argued that creative choices over how to assign cases (units of spending) among different categories lay at the heart of such gaming. Depending on whether the categories represented “desirable” or “undesirable” kinds of spending, the amount of spending that belongs to the relevant category will be over- or under-reported. The trio of cases drawn from the UK’s public expenditure control and planning system over recent decades all involved creative categorization as their central mechanism, which in turn had Goodhart’s Law effects on the relevant reported public spending numbers.

Third, in response to the question of how consequential or material were the Goodhart’s Law effects of gaming those public expenditure input numbers, the cases examined here show that numbers at stake in the relabeling of categories of spending were far from trivial, running from the equivalent of at least some hundreds of millions of dollars in the case of the “tuck

unders” to billions in the case of the defence non-cash-to-near-cash transfer, and indeed tens of billions in the case of off-balance sheet debt at the height of the Private Finance Initiative. And there are other cases not examined here (such as the off-balance sheet student loan debt that finances much of the English higher education sector, held by a notionally private company) that arguably run into more than ten billion. Such numbers certainly cannot be dismissed as immaterial or inconsequential, and they present major challenges for public accountability.

Going beyond those three questions, the analysis of the paper suggests that efforts to improve public spending control through new categorizations of spending intended to protect certain types of spending, produce better decision-making over spending, even to limit gaming, can have the unintended result of instigating new forms of gaming and associated Goodhart’s Law effects. Indeed, all of the three cases discussed in the paper stemmed from new categorizations of that type offering new opportunities for gaming by blurring or jumping the new boundaries that had been introduced.

In answering the three questions posed at the outset, this study of course raises other questions that cannot be answered here. For instance, who benefits and who loses from gaming over expenditure numbers and its Goodhart’s Law effects? How can the costs of such gaming be assessed, against any possible benefits of such behavior, notably as a “workaround” that enables flexibility? Who sees what sorts of categorical finessing as acceptable or unacceptable gaming, and what are the formal or informal social mechanisms that serve to limit or reverse such gaming? Can gaming of meso-level public expenditure input numbers of the type analyzed in this paper be compared internationally and if so how? Such questions pose plenty of challenges for future research.

References

Alt, J., Lassen, D. D., & Wehner, J. (2014). It isn't Just About Greece: Domestic Politics, Transparency and Fiscal Gimmickry in Europe. *British Journal of Political Science*, 44 (4), 707-716.

Anderson, M. (2015, March 9). UK Passes Bill to Honour Pledge of 0.7% Foreign Aid Target. *The Guardian*. Retrieved from <https://www.theguardian.com/global-development/2015/mar/09/uk-passes-bill-law-aid-target-percentage-income>

BBC News (2011, November 13). Government grant for BBC World Service Trust. Retrieved from: <https://www.bbc.co.uk/news/entertainment-arts-15711864>

Berliner, J. S. (1957). *Factory and Manager in the USSR*, Cambridge, MA: Harvard University Press

Bevan, G., & Hood, C. (2006). What's Measured is What Matters: Targets and Gaming in Healthcare in England. *Public Administration*, 84(3), 517-38.

Chalmers, M. (2015). *Osborne's Summer Surprise for Defence Guaranteed Real-Terms Spending Increases* (Briefing Paper SDSR 2015: Hard Choices Ahead). London: Royal United Services Institute. Retrieved from https://rusi.org/sites/default/files/201507_bp_osbornes_summer_surprise_for_defence.pdf

Chesson, A., & Maitland-Smith, F. (2006). *Including Finance Lease Liabilities in Public Sector Net Debt: PFI & Other*. Office for National Statistics. Retrieved from

https://web.archive.org/web/20120312224718/http://www.hm-treasury.gov.uk/d/financeleasesarticle_200906.pdf

Colebrook, C. (2015, November 26). The twelve tricks in George Osborne's spending review. *New Statesman*. [online] Retrieved from <https://www.newstatesman.com/politics/economy/2015/11/twelve-tricks-george-osbornes-spending-review>

Committee of Privy Counsellors (2016) *The Report of the Iraq Inquiry: Volume X, Section 13.1 Resources*. HC 265-X. London: The House of Commons. Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/535427/The_Report_of_the_Iraq_Inquiry_-_Volume_X.pdf

DeHart-Davis, L. (2007). The Unbureaucratic Personality. *Public Administration Review*, 67 (5), 892-903.

Eurostat (2006). Stock-flow adjustment in government accounts. *Economy and Finance*(26), Retrieved from <https://ec.europa.eu/eurostat/documents/3433488/5444497/KS-NI-06-026-EN.PDF/ce8d76ca-b9e7-4c53-8948-7da159b94369>

Eurostat (2009) *Stock-flow adjustment (SFA) for the Member States, the euro area and the EU27 for the period 2005-2008, as reported in the October 2009 EDP notification*. Retrieved from https://ec.europa.eu/eurostat/documents/4187653/5779233/STOCK_FLOW_2009-EN.PDF/6a7e3307-f323-4ba8-91ed-1d4e57751fb3?version=1.0

Eurostat (2011). *Stock-flow adjustment (SFA) for the Member States, the euro area and the EU27 for the period 2007-2010, as reported in the April 2011 EDP notification*. Retrieved

from https://ec.europa.eu/eurostat/documents/4187653/5781598/STOCK_FLOW_2011-EN.PDF/551deb30-f7c7-41ce-a70c-63766cfcfe0c?version=1.0

Eurostat (2014) *Stock-flow adjustment (SFA) for the Member States, the euro area and the EU28 for the period 2010-2013, as reported in the October 2014 EDP notification*. Retrieved from

https://ec.europa.eu/eurostat/documents/4187653/5798141/STOCK_FLOW_2014_OCT-EN.PDF/f4a1b352-1f08-48d6-8303-f94f792caeab

Eurostat (2015). *Stock-flow adjustment (SFA) for the Member States, the euro area and the EU28 for the period 2011-2014, as reported in the October 2015 EDP notification*. Retrieved from <https://ec.europa.eu/eurostat/web/products-eurostat-news/-/STOCK-FLOW-2015-OCT>

Eurostat (2017). *Stock-flow adjustment (SFA) for the Member States, the euro area and the EU28 for the period 2013-2016, as reported in the April 2017 EDP notification*. Retrieved from <https://ec.europa.eu/eurostat/documents/4187653/7992709/STOCK-FLOW-2017-APR.pdf>

Goodhart, C. (1981). *Problems of Monetary Management: The U.K. Experience*. In Courakis, A. S. (Ed.). (1981). *Inflation, Depression, and Economic Policy in the West*. Oxford: Alexandrine Press (Reprinted from Goodhart, C. A. E. (1975). "Problems of Monetary Management: The U.K. Experience". *Papers in Monetary Economics*. I. Reserve Bank of Australia)

Grossman, G. (1960). *Soviet Statistics of Physical Output of Industrial Commodities: Their Compilation and Quality*, Princeton, NJ: NBER and Princeton University Press.

Hansard HC Deb (1 November 2017) vol 630, col 30WS. [Electronic version]. Retrieved from <http://bit.ly/2JFI8un>

Harrison, M. (2011). Forging Success: Soviet Managers and Accounting Fraud, 1943 to 1962. *Journal of Comparative Economics*, 39(1), 43-64.

Heald, D. (2005). The Implementation of Resource Accounting in UK Central Government. *Financial Accountability and Management*, 21(2), 163-190.

HM Treasury (1994). *Better Accounting for the Taxpayers' Money: Resource Accounting and Budgeting in Government* (Cm 2626). London: HMSO, Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/272027/2626.pdf

HM Treasury (November 2015). *UK Aid: Tackling Global Challenges in the National Interest* (Cm 9163). London: HMSO. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/478834/ODA_strategy_final_web_0905.pdf

HM Treasury (2018, October 29). *Budget 2018: Philip Hammond's speech*. Retrieved from <https://www.gov.uk/government/speeches/budget-2018-philip-hammonds-speech>

Holley, I. B. Jr (1989). *Buying Aircraft: Matériel Procurement for the Army Air Forces*. United States Army in World War II Special Studies. Washington, DC: Center of Military History, United States Army

Hood, C. (2006). Gaming in Targetworld. *Public Administration Review*, 66 (4), 515-20.

Hoskin, K. (1996) The Awful Idea of Accountability: Inscribing People into the Measurement of Objects. In R. Munro & J. Mouritsen (Eds), *Accountability: Power, Ethos and the Technologies of Managing* (pp. 265-82). London: International Thomson Business Press

House of Commons Committee of Public Accounts (2007, March 20). *Financial Management in the NHS Seventeenth Report of Session 2006–07* (HC 361). London: HMSO. Retrieved from <https://publications.parliament.uk/pa/cm200607/cmselect/cmpubacc/361/361.pdf>

House of Commons Treasury Select Committee (2011, July 18) *Private Finance Initiative* (HC 1146). Retrieved from <https://publications.parliament.uk/pa/cm201012/cmselect/cmtreasy/1146/1146.pdf>

House of Commons Select Committee on Defence (2016, April 12) *Shifting the goalposts? Defence Expenditure and the 2% Pledge* (HC 494). London: HMSO. Retrieved from <https://publications.parliament.uk/pa/cm201516/cmselect/cmdfence/494/494.pdf>

Irwin, T.C. (2015). Defining the government's debt and deficit. *Journal of Economic Surveys*, 29(4), 711-732.

Koen, V. and Van den Noord, P. (2005). *Fiscal Gimmickry in Europe*. OECD Library. Retrieved from <https://www.oecd-library.org./content/paper/237714513517>

Mann, N. (2012, January 11). PFI “Still Being Used to Keep Costs Off Balance Sheet. *Public Finance*. Retrieved from <https://www.publicfinance.co.uk/news/2012/01/pfi-‘still-being-used-keep-costs-balance-sheet’>

Nove, A.(1980). The Soviet Economy: Problems and Prospects. *New Left Review*, I/119 (Jan-Feb 1980): 3-19

Office for Budget Responsibility. (2017). *Economic and fiscal outlook* (Cm 9530) Retrieved from <https://cdn.obr.uk/Nov2017EFOwebversion-2.pdf>

Office for Budget Responsibility. (2018). *Fiscal Sustainability Report*. London: HMSO. Retrieved from <https://cdn.obr.uk/FSR-July-2018-1.pdf>

Palmer, K. (2006). *NHS Reform: Getting back on track*. London: King's Fund

Portillo, S. (2012). The paradox of rules: Rules as resources and constraints. *Administration & Society*, 44(1), 87-108.

Riordan, W. (1963). *Plunkitt of Tammany Hall: A Series of Very Plain Talks on Very Practical Politics*. New York: Dutton. (Original work published 1905)

Schick, A. (1998). A Contemporary Approach to Public expenditure Management. *World Bank Institute*, 68, 2-111.

Spreitzer, G.M., & Sonenshein, S. (2004). Toward the construct definition of positive deviance. *American behavioral scientist*, 47(6), 828-847.

Stephan, G. (2016). Public or Private Job Placement Services – Are Private Ones More Effective? *IZA World of Labor* August 2016. Retrieved from <https://pdfs.semanticscholar.org/028b/3abeeba5166506ca2f5b0afa6f0e36c55da6.pdf>

Strathern, M.(1997). Improving Ratings: Audit in the British University System. *European Review*, 5(3), 305-321.

Townsend, I. (2010) *0.7% of National Income as Aid: Key Issues for the 2010 Parliament*, House of Commons Library Research Paper

Tummers, L. L., Bekkers, V., Vink, E., & Musheno, M. (2015). Coping during public service delivery: A conceptualization and systematic review of the literature. *Journal of Public Administration Research and Theory*, 25(4), 1099-1126.

Von Hagen, J., & and Van den Noord, P. (2006). What do Deficits Tell Us About Debt? Empirical Evidence on Creative Accounting with Fiscal Rules in the EU. *Journal of Banking and Finance*, 30 (12), 3259-3279.

Wright, O. (2016, April 21). MoD Accused of Using “Creative Accounting” to Meet Nato Commitment on Defence Spending. *The Independent*. Retrieved from <https://www.independent.co.uk/news/uk/politics/mod-accused-of-using-creative-accounting-to-meet-nato-commitment-on-defence-spending-a6993526.html>

Wellink, A.H.E.M. (1996). Budgetary Control: Goodhart’s Law in Government Finances? In Kool, C M J, Muysken, J, van Keen, T (Eds.) *Essays on Money, Banking and Regulation* (pp. 69-90), Boston, MA: Springer.

Table 1: Ten Examples of Categorical Distinctions in Public Spending Control

Boundary lines	
‘Public spending’ On balance sheet	Non-public spending or public non-spending (e.g. guarantees) Off balance sheet
Spending consistent with stated fiscal or spending rules e.g. of budget balance or debt quotients	Spending that goes beyond stated fiscal or spending rules
‘Controlled public spending’ (subject to tighter limitations or containment)	Less- or non-controlled spending (subject to looser limitations or containment, e.g. debt interest payments)
‘Committed’ or allocated spending	Unallocated or ‘reserve’ spending
‘Baseline’ spending (Costs of existing policy)	Additional spending Costs of new or changed policies
‘Protected’ or ‘ring fenced’ domain specific spending	‘Unprotected,’ vireable or ‘non-ring fenced’ spending
‘Capital’ spending	Current or resource spending
Spending within a budget period	Spending outside a budget period
‘Administration costs’ ‘Back office’	‘Programme costs’ ‘Front line’
‘Cash’ spending	Non-cash outlays (resource commitments, ‘tax expenditures’ etc.)

Table 2: Examples of ‘Gaming’ by Blurring or Jumping Categorical Distinctions in Public Spending Control

Boundary lines		Examples of creative categorization
‘Public spending’ On balance sheet	Non-public spending or public non-spending Off balance sheet	Taking spending off balance sheet by notionally ‘private’ debt liabilities
Spending consistent with stated fiscal or spending rules e.g. of budget balance or debt quotients	Spending that goes beyond stated fiscal or spending rules	Presentational adjustments in forecasts relevant to fiscal rules to create ‘back pockets’
‘Controlled public spending’ (subject to tighter limitations or containment)	Less- or non-controlled spending (subject to looser limitations or containment)	Classifications that reflect the porous distinction between controlled and less controlled spending
‘Committed’ or allocated spending	Unallocated or ‘reserve’ spending	Spending allocations that are not fully completed
‘Baseline’ spending Costs of existing policy	Additional spending Costs of new or changed policies	Padding the baseline
‘Protected’ or ‘ring fenced’ domain specific spending	‘Unprotected,’ vireable or ‘non-ring fenced’ spending	‘Tuck unders’ i.e. switches between protected and unprotected spending
‘Capital’ spending	Current or resource spending	Viring from capital to resource spending
Spending within a budget period	Spending outside a budget period	Blurring temporal boundaries
‘Administration costs’ ‘Back office’	‘Programme costs’ ‘Front line’	Reclassifying administrative costs as programme costs
‘Cash’ spending	Non-cash outlays (resource, ‘tax expenditures’ etc.)	Viring non-cash into near-cash

Table 3: Categorical Complication or Multiplication: Selected UK Examples 1993-2016

Categorical Distinction	Categorical Change or Multiplication
‘Public spending’ (on balance sheet) versus non-public spending (off balance sheet)	New spending categories relating to private finance arrangements in 1992, 1997-8, 2010s
Spending consistent with stated fiscal etc rules versus spending that goes beyond fiscal rules etc	New fiscal rules in 1993, 1998, 2008, 2010, 2015, 2016
‘Controlled public spending’ versus less or non-controlled spending	Introduction of a new control total in 1992, a revised control total categorization in 1998 (DEL-AME) and many adjustments thereafter
Spending within a budget period versus spending over a longer period	Removal of end-year-flexibility in mid-1990s, restoration of EYF in 1998, replacement of EYF by ‘Budget Exchange’ in 2010
‘Administration costs’ versus programme costs	Major re-categorization in 2004 reflecting a front-line/back office distinction, plus frequent detailed re-categorizations over the period
‘Protected,’ ‘ring fenced’ spending versus unprotected, virable or non-ring-fenced spending Domain specific spending	Input protection targets for aid spending 2005 and 2015; new protected/unprotected categorization 2010.
‘Committed’ or allocated spending versus unallocated or ‘reserve’ spending	Creation of ‘DUPs (Departmental Unallocated Provisions) 1998
‘Cash’ spending versus non-cash	Introduction of RAB1 in 2000, RAB2 2002 and later modifications, publication of WoG accounts 2010

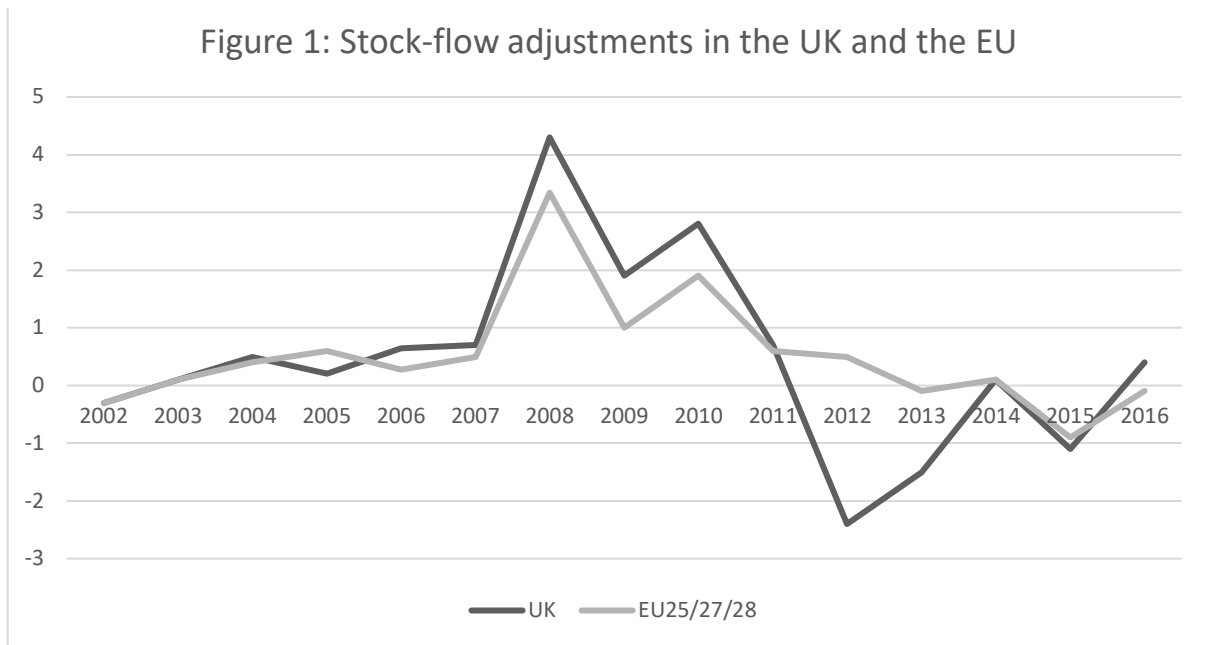


Figure 1. Stock-flow adjustments in the UK and the EU in percent of GDP (reconstructed using data from Eurostat (2006-2017))