Post-mortem of the VP-function?
Meta-regression analyses of economic voting in the United Kingdom – an anthology

Stian Skaar Ludvigsen

PhD Dissertation

Post-mortem of the VP-function?
Meta-regression analyses of economic voting in the United Kingdom – an anthology

Politica

ISBN: 978-87-7335-137-6

Cover: Svend Siune
Print: Juridisk Instituts Trykkeri, Aarhus Universitet
Layout: Annette Andersen

Submitted October 2009
The public defense takes place February 12 2010
Published February 2010

Forlaget Politica
c/o Department of Political Science
Aarhus University
Bartholins Allé 7
DK-8000 Århus C
Denmark

Tel. +45 8942 1253
Fax +45 8613 9839
e-mail: politica@ps.au.dk

Post-mortem of the VP-function?

Meta-regression analyses of economic voting in the United Kingdom – an anthology

Stian Skår Ludvigsen
Deptartment of Political Science
Aarhus University

Abstract

This paper is an overview of my anthology for the PhD degree at the Department of Political Science, Faculty of Social Sciences, Aarhus University. The anthology consists of five papers and this summary.

The general topic of this anthology is studies known as *vote and popularity (VP) function studies*, which estimate the effects of economic outcomes on voting or party popularity, and the motivation for this anthology has been the various concerns that have been voiced over the unstable nature of these studies of economic voting. The dissertation builds the argument that the theoretical and empirical complexity of the VP-function simultaneously necessitates complex models *and* calls for some ways to reduce this complexity when reviewing the literature and taking stock of the accumulated knowledge. The dissertation solves this through meta-regression analysis, where the cases for the empirical analyses are studies of government support in the United Kingdom. In addition to this summary, two papers summarize and build upon the theoretical foundations of the literature, and three papers apply meta-regressions in order to quantitatively synthesize the results from the literature. These three test theory and explain variation in the results from the literature that is due to specification differences and estimate and control for the impact of publication selection bias. These aspects have been seen as problematic in the literature.

The dissertation sets out as a post-mortem, with the assumption that the literature may be drawing its final breath. With the statistical techniques applied here, inconclusive results from the literature are reconciled through controls for publication selection bias and specification differences. The pathology thus turns out to be a diagnosis of symptoms rather than a post-mortem, and the conclusion is that the rumours of the death of the VP-function are greatly exaggerated.

Contents

List of tables

List of figures

Foreword and acknowledgements

Research is a creative discipline. In other words, it is not disciplined. I have had manual jobs where blueprints have told me what to do, and one of the most important requirements for the job was charged batteries for my power-tools. In research, your brain is your power-tool, and you cannot plug it into a socket and charge it over night while you are clocked out. Charging your brain requires long walks in the park, exercise, a good night's sleep, good company, and healthy food. As a PhD-student, these things are often neglected, and the brain suffers. My dear friends, my family, and my girlfriend have also been neglected at times. For that, I am very sorry, but I am also deeply grateful that they have not only allowed me to neglect them, but supported me while doing so. This dissertation would not have been possible had it not been for my dear Idun. Your love and your moral (and financial) support is what have kept me afloat! This dissertation is dedicated to you!

When I moved to Denmark, my brother Knut told me to never mind the three letters that I could hope to put after my name when I would be done, and instead be happy about the fact that I was being paid for doing something I love. Of course, it has not always been love – a few months behind schedule (and consequently, without pay) is an indication of this – but the job has been very dear to me, all things considered, and I hope to keep doing what I love. Therefore, it is not so much the academic title that worries me at the moment, rather the job title now that it is time to find a new venue for my research. I am sorry to leave Aarhus, but I am deeply grateful for the time I have had here, and I hope to collaborate with friends and colleagues from Aarhus in the future.

Peter Nannestad and Martin Paldam opened the academic world for me in several ways: Their article from 1994 was the first I read on the topic of economic voting when I started the work on my master's (candidatus rerum politicarum) thesis in Bergen. I contacted Martin Paldam during my search for literature for that thesis, who not only gave me useful suggestions on literature, but also suggested that I sent a paper proposal for the European Public Choice Society's meeting in Durham in 2005. So I did, and Peter Nannestad was chairing the session I presented at. I felt warmly received by these two and many others, and the EPCS-conference gave me a glimpse of a backstage life of academia that I realized I wanted to be a part of. Aarhus University was therefore an obvious choice for the pursuit of this life. Peter became my primary supervisor, and Martin had a great impact not only on the choice of methodology, but also on the places I have travelled to learn the trade: first Durham, then Aarhus, then Melbourne, and a couple of conferences in Denmark and France.

I did not only want to go where Martin suggested. Two papers have been written during my PhD that is not a part of this anthology. One is a theoretical paper on *trust*, using game theory, and was a result of the three PhD courses I took. This one does simply not fit in with the other articles. It was presented at the Norwegian Political Science Conference in Trondheim and the World Public Choice meeting in Amsterdam, both in 2007. The other paper studied Palestinian terrorism in light of theories on economic voting in Israel – and was in line with my original project proposal – but the data I had received from Israel had some major coding errors, and it took four months to receive re-coded datasets. It has been presented at the World Association for Public Opinion Research meeting in Jerusalem in 2007 as work-in-progress. There has been no time to update this paper with the re-coded data. The result, however, is fortunate: the PhD dissertation before you is much more homogenous than I initially intended it to be.

The Department of Political Science at Aarhus University has proven itself as an excellent venue for research and academic development. I am deeply grateful for the resources and opportunities given to me by my department. It is safe to say that this dissertation would not have been produced had it not been for this department and its staff and environment: an environment that is geared towards a collective effort in producing the best possible PhDs. There are many people here that deserve credit and gratitude for my dissertation: Peter Nannestad – my primary supervisor – for believing in me and following me through; Søren Risbjerg Thomsen for his solid co-supervision; the late Lise Togeby and the rest of the academic management at the Department of Political Science in Aarhus for giving me this great opportunity; Birgit Kanstrup, Inge Rasmussen, and Anne-Grethe Gammelgaard for excellent managerial, financial, and secretarial services; Vivi Mikkelsen, Berit Møller, and Stig Petersen for their friendly IT-support; Annette Andersen for her astonishingly quick and reliable assistance and proof-reading – always provided with a smile; Anne Gry Gudmundsdotter Rønningen for a helpful second opinion on the Norwegian summary; Kirstine Korsager and Casper Borchmann for reliable and much needed research assistance in stressful periods; Karen Prehn, Mette Ahlers Marino, and Gorm Willemoes Jensen for providing the cosiest library services I have ever come across; my wonderful office-mates through the years: Mads Leth Jakobsen, Line Renate Hansen Gustafsson, and Yonathan Schvartzman; Carsten Jensen for providing me with shelter and late-night academic discussions during the last weeks of writing; and my many friends, colleagues, and students.

There are people and institutions outside of the department that share credit for this dissertation as well: Especially Chris Doucouliagos who have been the

most helpful and patient host and friend during my visit to Melbourne and my efforts to learn the trade of meta-analysis; Jon Jay Neufeld for his professional proof-reading of the anthology summary and helpful suggestions on how to improve my message; Øyvind Olufsen for a critical view on the Norwegian summary; other friendly and helpful people (including – but not limited to): Martin Paldam, Tamir Sheafer, Randy Silvers, and Tom Stanley; official and non-official hosts through the years: the departments of Comparative Politics and of Economics at the University of Bergen, of Political Science at the Hebrew University of Jerusalem, and of Economics, Finance and Accounting at Deakin University. Finally, mom and dad, family and friends: your support and encouragement has been intrinsic for my work towards this dissertation!

With all this support, there is only one person to blame for any errors that may have been made: yours truly.

The tragic loss of a dear friend through 26 years, Lars Dugstad Wake, came out of nowhere when I had a mere seven weeks left before my deadline, and at the end of a rather long writer's block. His accident was a reminder of the fragility of life, and of the little time we have on this planet to get things done. My thoughts have been with his dear Nina and Eirik Andreas when writing this summary of my anthology.

A technical note

The papers in the anthology at hand can be quite intimidating to a non-technical or non-specialist audience. I have tried to limit the use of abbreviations and jargon, but there are times when I feel phrases such as *'ceteris paribus'* is more appropriate than *'all things equal'* and *'viz.'* works better than *'namely'.* This should work fine with fellow political scientists, but may make laypersons feel the need for a Latin dictionary. However, a dictionary should be unnecessary if they have already read this. The non-technical audience, on the other hand, may need some guidance. Even those who have some experience with regression analysis may be put off by regression techniques such as *meta-regression with bootstrapped panel-adjusted standard errors.* They should not worry (too much). Those familiar with multiple regression should be able to read this dissertation without problems, and should be comforted by knowing that meta-regression is basically a regression of regression analyses. For those who are not comforted by this, I will recommend two very helpful and quick-read introductions to regression analysis: Lewis-Beck (1983) and Midtbø (2007). The latter is in Norwegian, written by my master's thesis supervisor, while the former is published in *the Sage University Paper series on Quantitative Applications in the Social Sciences*. This series has many helpful brief books

for those who are interested in understanding more of what is going on in this anthology, including (but not limited to) multiple regression (Berry, 1993; Berry & Feldman, 1985), multilevel modelling (Luke 2004), and bootstrapping (Mooney & Duval, 1993). I do not deal with much of the technicalities of – and assumptions behind – the models used in this dissertation, thus I hope to avoid intimidating those with limited expertise within quantitative methodology. For the more technically astute, issues such as panel-adjusted standard errors are dealt with in Petersen (2009), and multilevel models in Rabe-Hesketh & Skrondal (2008) and Wooldridge (2002). A nice Danish introduction to multilevel models is given by Andersen (2007). Technical issues aside, much of this dissertation is theoretical, thus I hope a wide audience will enjoy my work.

Aarhus & Bergen, October 2009

For Idun

1. Introduction

This dissertation introduces *meta-regression analysis* to Nordic and British Political Science. This is a novel, but very useful approach to conducting literature reviews. The anthology essentially constitutes a set of literature reviews in which two papers summarize and build upon the theoretical foundations of the literature at hand, and three papers apply regressions *of* regressions in order to quantitatively synthesize the results from more than thirty-five years of research into government popularity in the United Kingdom.

The general topic of this anthology is studies known as *vote and popularity function (VP-function) studies*, which estimate the effects of economic outcomes on voting or party popularity. The motivation for this dissertation is the various concerns voiced over the unstable nature of these studies of economic voting. As Whiteley (1984: 4) pointed out: "relationships can be made to appear, disappear and reappear merely by adding or subtracting a few observations". The instability is usually expressed by varying results over time and across countries, partly due to varying institutional structure of the cases and partly due to the varying complexity and other specification differences of the competing estimations. The title of the dissertation reflects a concern that the field of VP-function studies has been mortally wounded by this instability, so that there is a risk that research into the VP-function is being put to rest.

Even within countries, "coefficients tend to come and go when, for example, the period analyzed changes" (Paldam, 1997: 347). *Country* is held constant in this project, as it only deals with VP-functions in the United Kingdom, but a huge variation still remains in the British literature. Both of my meta-analyses of the effect of inflation and unemployment on government popularity (see Chapters 8 and 9) have 54 explanatory variables that control for variation in a literature consisting of 39 and 40 studies, respectively. And I did not need to stop at 54. Countless variables could have been added had I coded specifications that only appear once or a few times in the literature.

Needless to say, it is impossible to keep track of these differences and their impact upon the results through a narrative review. And narrative reviews are abundant. The search protocols in the meta-analyses list 28 reviews in addition to three published conference volumes. The literature is extensive, and my meta-analyses merely tap into a fraction – possibly less than ten percent – of the entire literature. This ten percent does not even represent everything that has been published of British VP-functions, but it ought to be as good as all of the literature that has been published with the minimum of information required in order to standardize and compare the results. The studies that have been left out have simply not met the relevant inclusion criteria.

Borrowing a quote from an entirely different literature, there may be "subtle differences in measurement, theoretical questions, and empirical context" (Baumgartner, 1998: 6) that not only blur the view for the narrative reviewer, but actually hinder any reader of the literature from seeing the cumulative knowledge. However, it is not necessary to succumb to the fatalism expressed by Baumgartner (op.cit), that "the development of a cumulative body of evidence [in the study of groups in politics is] an elusive goal"; or by Geller & Singer (1998: 3), "that even if [the studies of international conflict] all point to the same conclusion, we cannot assume theoretical convergence, nor can we assume that they point to differing conclusions even if the statistical results are quite dissimilar". The results from studies of groups in politics and of international conflict can be assessed using meta-regression analysis (MRA), which with its standardization procedures and controls for differences in specification can report the accumulated knowledge. This is quite simply the purpose of this PhD project: to report the accumulated knowledge of the effects of economic outcomes on government popularity in the United Kingdom.

This anthology is the very first to review the VP-function literature using quantitative methods, which is well overdue given the extensive literature. Statistical techniques are absolutely necessary in order to deal with the heterogeneous specifications and results, and I have imported meta-regression analysis from economics to political science. Meta-regression is simply a regression of regressions, where the published effects of the relationship of interest are used as observations of the dependent variable, and the variation in model specification, publication or research practices, data, etc., constitute the explanatory variables. Meta-analysis started in education research (the term was coined by Glass, 1976), then spread to psychology, medicine, biostatistics, land management, etc., to economics.

This dissertation operates on the border between economics and political science, and there are possibly just as many – if not more – economists as political scientists who have published VP-functions. Meta-regressions, on the other hand, have been published by many more economists than political scientists. Within economics, there have been studies such as Bateman & Jones (2003) on the recreational value of British woodlands, Brouwer et al. (1999) on the value of non-market environmental services in American and European wetlands, Doucouliagos (2005) on the impact of economic freedom on economic growth, Doucouliagos & Paldam (2008) on the (lack of) effect of foreign aid on economic growth in developing countries, Longhi et al. (2005) on the effect of immigration on wages, Doucouliagos & Stanley (2009) on minimum wages, and Weichselbaumer & Winter-Ebmer (2005) on the gender

wage gap. Further examples from economics can be found in Stanley (2001: Table 1) and Stanley et al. (2008: 277). Within political science, there are only Imbeau et al. (2001) on the impact of party ideology on government policies, Lau et al. (1999, 2007) on the effects of negative political advertising, Roscoe & Jenkins (2005) on the impact of campaign contributions on roll call voting, Geys (2006) on voter turnout, and Doucouliagos & Ulubaşoğlu (2008) on economic growth and democratization. The latter is the most technically advanced meta-regression published in a political science journal but was authored by two economists, so the review technique remains quite unknown to political scientists. In that sense, this dissertation truly represents a contribution to political science, but it analyses a political economy literature; as such, it is also an important contribution to the expanding meta-regression literature within economics.

My dissertation consists of five papers investigating various aspects of economic voting (none are published):

Two theoretical papers:

1. *'Conditional evaluations in economic voting'* (Ludvigsen, Paper 1), which is a review of the theoretical contributions to the literature. The article is not published nor has it been presented at any conference.
2. *'Party positions, voter preferences, and the cost of ruling'* (Ludvigsen, Paper 2), which is a theoretical discussion of the possible reasons for the cost of ruling. The paper has been presented to the Danish Public Choice Society, Deakin University, and the Nordic Political Science Association in 2008.

Three empirical papers:

3. 'Personal economic expectations and government popularity in the United Kingdom – a meta-analysis' (Ludvigsen, Paper 3). Presented at the Meta-Analysis for Economics Research Workshop in Nancy, France, 18.10.08; the Department of Political Science, University of Aarhus; the Department of Comparative Politics, University of Bergen; and the Department of Economics, University of Bergen. The main conclusion of the paper is that there is a robust – but not very strong – positive correlation between expectations and government popularity, although the published results are inflated by publication bias. There is no difference between Labour and the Conservatives.
4. 'Inflation and government popularity in the United Kingdom – a meta-analysis' (Ludvigsen, Paper 4). Main conclusion: negative correlation with government popularity. No publication bias. Effects are stronger for Labour than for the Conservatives.

5. ‘Unemployment and government popularity in the United Kingdom – a meta-analysis’ (Ludvigsen, Paper 5). Main conclusion: strong negative correlation with government popularity. Possible publication bias. Effects are slightly stronger for Labour than for the Conservatives.

Although the anthology makes a few theoretical contributions, the main contribution of the dissertation is the application of meta-regression analysis to a political science literature together with the empirical contribution that follows. To the best of my knowledge, I am the only political scientist in Europe working with meta-regression and the only political scientist working on the particular method introduced here.[1]

This summary is structured as follows: A run-down of the two theoretical papers is given first, starting with the general theoretical paper in Chapter 2 and continuing with the more specific theoretical paper in Chapter 3. Then, there are three chapters providing more detailed information regarding the motivation and methodology of meta-regression analysis: Chapter 4 discusses meta-analysis as a review method; Chapter 5 discusses problems with publication selection bias; and Chapter 6 gives more details about how to test and control for publication bias. The three meta-analyses are then summarized, with the analysis of personal economic expectations in Chapter 7, the analysis of inflation in Chapter 8, and the analysis of unemployment in Chapter 9. Chapter 10 compares the three analyses and Chapter 11 concludes. A Norwegian summary follows the list of references. The five papers follow in the order they are discussed here, constituting the full PhD dissertation.

1 A Google Scholar search on October 22, 2009, for the terms ‘meta-regression’ AND ‘political science’ did not reveal any meta-regressions published by European political scientists. I am also the only political scientist in the Meta-Analysis of Economic Research Network (MAER-Net), spearheaded by Tom D. Stanley of Hendrix College, Arkansas (http://www.hendrix.edu/maer-network/).

2. Conditional evaluations in economic voting

This paper is the main theoretical paper of my dissertation. It "summarizes the theoretical literature on economic voting and attempts to capture and unify the theoretical complex of evaluations of government performance" (Ludvigsen, Paper 1: 1). The motivation for this paper is not the lack of reviews (the paper lists 33), but rather the fact that most of "these reviews have also given room to empirical analyses, and thus, left pieces of the theoretical complex scattered around" (op.cit). The last comprehensive review was Nannestad & Paldam (1994). Although it has aged well, I argue that it is due time to recapture the theories of economic voting and "clean up the clutter" (Ludvigsen, Paper 1:1). On this basis, "I argue that the theoretical and empirical complexity of economic voting necessitates complex models but will easily exhaust one's degrees of freedom and make comprehensive models nearly impossible to measure" (Ludvigsen, Paper 1: 1). The conclusion that can be drawn from this paper is that a systematic review of the existing empirical contributions offers a possible way forward. In other words, this paper lays the path for the meta-analysis of economic voting.

The paper starts with a generic VP-function, which is also described in the three meta-analyses. This model constitutes the common thread running throughout the entire dissertation. Although it gets repetitive, I find it important to show it here as well (see also Nannestad & Paldam, 1994; Lewis-Beck & Stegmaier, 2000, 2007; Hibbs, 2006):

$$VP_t = \alpha + \sum_{m=1}^{M} \beta_m Econ_{m,t} + \sum_{p=1}^{P} \varphi_p Pol_{p,t} + \gamma Trend_t + \sum_{i=2}^{I} \delta_i Gvt_{i,t} + \varepsilon_t \qquad \textbf{[E1]}$$

In this model, you have the aggregated vote (V) *or* popularity (P) of the government measured over time as a function of various economic outcomes, political events, time trends, and (sometimes) government-specific fixed-effects. Lagged observations of both the dependent and independent variables may also be included. Subscript t indicates observation at time $t = 1, 2, \ldots, T$; the Σ's indicate groups of variables ranging from 1 (2) to a number M, P, or I, respectively; and the subscripts m, p, and i are variable identifiers within these groups. α is the constant, β, φ, γ, and δ are slope parameters, and ε is the error term.

Ludvigsen (Paper 1) deals with the theoretical complexity behind this model, while Ludvigsen (Paper 2) deals with the theories behind the 'Trend' variable in the model. The three meta-analyses deal with three of the most common variables that constitute the economic fraction of the model. The political fraction is typically ad hoc and usually measured by dummy variables, so although a meta-analysis of the most common political variables is

possible, it may not be particularly interesting. One could argue that this is unfortunate, since this is a PhD dissertation in political science, but the 400 or so studies of the VP-function and the 33 or so reviews are proof that the economy is perceived as being vital to political support. Nonetheless, Equation 1 constitutes the common thread running throughout my dissertation.

There are loads of theories attempting to explain parts or all of the mechanisms in the VP-function, and Ludvigsen (Paper 1) deals with five general theories about the overall function, three ideal types of voters, and a number of conditioning moderators. Furthermore, there are eight possible explanations for the 'cost of ruling', which is the deterioration of support seen by so many governments and measured by the trend variable. These eight explanations are dealt with in Ludvigsen (Paper 2).

The full evaluation complex is only dealt with in Ludvigsen (Paper 1), but four of the five general theories linking economic outcomes with political support are repeated in two of the meta-analyses (Ludvigsen, Paper 4, 5). As I repeat several times, testing four theories with one model may seem like shooting fish in a barrel, but the theories are developed or redeveloped *within* the literature in much part due to non-confirmatory findings. I.e., the instability of the VP-function has spurred new theories, which does not necessarily solve the problems, since they add to the problem of acquiring cumulated knowledge. As Baumgartner (1998: 6) has written about the literature on groups in politics: "Diversity of approach must be balanced with some degree of shared theoretical perspective in order to produce a literature endowed with coherence and comparability". Fortunately, the theories *do share* a common heritage and are all tested by the same models. A model developed to test one theory can therefore be used to tests new theories added at a later stage. Thus, shooting fish in a barrel should be allowed with this literature.

The common theoretical ancestor of the political support functions is the *Responsibility Hypothesis*, which was first formulated as a *reward-punishment* hypothesis inspired by the works of V.O. Key (1964, 1966). The Responsibility Hypothesis simply expects voters to reward governments for improved economic outcomes while punishing them for deteriorated economic conditions. This holds regardless of the ideological orientation of the government (or the voters).

Some of the very first VP-function studies already indicated that voters may evaluate and respond to positive outcomes differently than to negative outcomes. Voters were commonly found to punish harder than they reward. This *grievance asymmetry* is dealt with to some extent in Ludvigsen (Paper 1), but more thoroughly discussed in Ludvigsen (Paper 2), since grievances

may be seen as a possible explanation for the cost of ruling. Asymmetries also provided grounds for the first alternative theory to be put forth, viz. the *Clientele Hypothesis.* Here, the asymmetries are slightly different from grievance asymmetries. Rather than asymmetric evaluation of positive versus negative outcomes, the Clientele Hypothesis expects asymmetric responses to different parties.

Three studies are seen to have started the VP-function literature independently of one another (Nannestad & Paldam, 1994): Goodhart & Bhansali (1970) for British popularity functions, Mueller (1970) for American popularity functions, and Kramer (1971) for American vote functions. The Clientele Hypothesis came as early as 1974, when Butler & Stokes (1974: 374) argued that "the consequence of hard times is not the lessening of support for the governing party generally in the country but rather an increase of support for each party in the class whose interests it represents and a decline of support for each party in the opposite class". In other words, voters are expected to be attracted to the party that *cares the most* about an issue when that issue develops negatively. This holds whether the parties are in government or not. I.e., left-leaning parties are expected to see increased support when unemployment rises, whereas right-leaning governments are expected to see increased support when inflation rises. The term 'Clientele Hypothesis' was coined by Rattinger (1981), who later (1991) fused the Clientele Hypothesis with the Responsibility Hypothesis by adding that both left- and right-wing governments may be punished for negative developments, but that right-wing governments are expected to be punished harder for increasing unemployment than left-wing governments, and left-wing governments punished harder for increasing inflation than right-wing governments.

Hibbs, Rivers & Vasilatos advanced a contrary hypothesis in 1982. Formulated as a *Salient Goal Hypothesis* Powell & Whitten (1993: 404) expected voters to "hold right-wing governments to a higher standard on inflation and be less concerned about unemployment" [and vice versa]). In other words, parties failing to meet their salient goals will be punished for letting down their constituents, so left-wing governments should be punished more for increased unemployment, and right-wing governments more for inflation.

Before this, Paldam & Schneider (1980) introduced the *Stability Hypothesis* as an antithesis to the Responsibility Hypothesis, which they saw in light of the structural change in the Danish party system after 1973. An expanded party system allowed room for 'exciting' parties based upon ideology, single issues, charisma or protest. These parties may be "unlikely government participants" (ibid: 152), but voters might afford to vote for them when times are good. When times are bad, however, voters "rally around … the more re-

sponsible parties" (ibid: 153). Hence, the *likely* government participants should see a positive effect of a deteriorating economy, regardless of their left/right orientation. I choose to view this hypothesis as a relaxation of Inglehart's (1971, 1977) postmaterialism thesis, since it allows for value change *within* generations, and I discuss the possibility that the materialist/postmaterialist continuum also has a pre-materialist component, into which I fit Barack Obama's *Guns and Religion Hypothesis:* When times get *really* tough, voters may act against their immediate materialist interests and vote to secure their physical and moral existence.

Finally, the fifth theory was proposed by van der Brug, van der Eijk & Franklin (2007), who argued that large (mainstream) parties could be seen as taking part in policymaking also when they are in opposition, and therefore, may be seen as responsible for policy outcomes alongside the governing parties. If this is the case, then "large opposition parties may well lose support as a consequence of a slowing economy, just as large government parties do" (ibid:57). I refer to this as the *Policy-Player Hypothesis.* The Policy-Player Hypothesis, however, is a theory about opposition parties, and is therefore not included in the meta-analyses that focus on governing parties.

I will return to *how* the relevant theories are tested in the chapter on the meta-analysis of inflation and government popularity (Section 8.3.1, p. 49-50).

Before describing these unconditional theories of economic voting, Ludvigsen (Paper 1) argues that typical specifications of the VP-function lack controls for dynamic voter heterogeneity (changing composition of voters) and after describing the unconditional theories, Paper 1 discusses possible types of voters through Whiteley's (1984) ideal types of voters: the *optimizing voters*, who are fully informed and have clearly defined and stable preferences; the *satisficing voters*, who are partially informed and have limited and variable preferences; and the *chaotic voters*, who have no consistent preferences and do not seek information before casting their votes – if they do so at all. Then there is a section where the so-called *conditional moderators* are discussed. These are *partisan loyalties*, which are deep-rooted ideological predispositions of the voters; *grievance asymmetries*, which I have already mentioned and deal with in greater detail in Ludvigsen (Paper 2); *the clarity of responsibility*, which is the question of *who* is to be blamed or rewarded; the impact of *political systems* and *economic institutions*; the role and accuracy of the *media;* the *cost of ruling* – again dealt with in detail in Ludvigsen (Paper 2); *prospective versus retrospective evaluations; sociotropic versus egotropic evaluations;* and *myopia versus far-sightedness.* Do voters consider the past or the future? Do they consider the wealth of the nation or the size of their own pocket-book? And do they have a short or a long perspective,

both as regards the past as well as the future? All of these questions are discussed – but not resolved – in Ludvigsen (Paper 1). The point I make with this discussion is that students of economic voting should – ideally – take into account the full complexity, including the theoretical micro-foundations of the models, as called for by Paldam (1991), and design models that enable researchers to test all theories and account for all evaluation mechanisms and filters, as well as any random or fixed effects that may apply to the models (e.g., Paldam argued that models should control for the stability of governments and "analyze periodicity more systematically" (1991: 29)).

However, VP-functions are typically based upon relatively few observations and do not have enough degrees of freedom to model the full complexity and control for all variation. I therefore conclude that meta-regressions offer a possible means of traversing the complexity. In other words, the paper sets out the path for the meta-analyses in my anthology.

3. Party positions, voter preferences, and the cost of ruling

The second theoretical paper (Ludvigsen, Paper 2) follows the first theoretical paper quite naturally, but the path it follows is a slight digression from my project. The paper is not necessary for the meta-analyses, but it is necessary in order to fully understand the theoretical complexity of VP-functions and does cover one of the variables in the meta-regressions. It is therefore not a wasted digression. In fact, this paper is where most of my own theoretical contributions to the literature lie.

The cost of ruling – the depreciation in government popularity that takes place for no apparent reason other than the government's willingness to assume responsibility – is a "nuisance to politicians, party strategists, and political analysts" alike (Ludvigsen, Paper 2: 2). Paldam (1991: 19) talks of the "paradox of ruling", where the paradox is seen in a rational expectations framework, since "the average government surely rules exactly as the rational voter expects, so why should the voter punish it by moving his or her vote to the opposition?" (Paldam, 1997: 346-347).

The average reduction in the vote for western governments is about 2.5 percentage points (Nannestad & Paldam, 2002; Narud & Valen, 2008). Obviously, there is considerable variation in the election results, and it is necessary to control for politics, events, economic outcomes, institutions, etc.; nevertheless, the pattern cannot be missed. In time-series models of government support that control for all of this, the trend variable (see Equation 1) should pick up the cost of ruling. I.e., the trend variable will display the deterioration of support as time goes by. I have not recorded the published effects of time for my meta-analyses, and the focus of this paper is theoretical, so I have not collected empirical data on this. In the British case, there are too few studies that have actually included a trend variable, so I cannot estimate a meta-regression model of the trend variable based upon the British data. Nevertheless, the trend variable is an obvious candidate for meta-analysis if one wishes to estimate the average cost of ruling *after* adding controls.

The central motivation of Ludvigsen (Paper 2) is that I think it is important for governments and analysts to identify the causes of this deterioration of support. "Without this information", I argue, "governments risk pursuing the wrong policies in the attempt to counter the effect, while analysts risk making imprecise models of – and inferences from – electoral behaviour" (ibid: 2). Voters may also make more informed decisions if they are aware of their own nature – that they have a proclivity to evaluate the incumbent more negatively than the challenger.

Instead of testing some of the theories, I choose to list them all. They can then be brought together and offer "a useful summary of the theories, thereby stimulating useful further thought, then analysis, on the subject" (to quote one anonymous referee). In other words, I build the foundation upon which others may test the theories (obviously, quite a few have been tested already – Narud & Valen (2008) being the most comprehensive: see Footnote 5 in Ludvigsen (Paper 2)). However, not all of the possible explanations are theoretical; nevertheless, they may be important and are therefore included in my review of theories on the cost of ruling.

The eight possible explanations that I provide are: 1) a Statistical Artefact; 2) the Coalition of the Minorities Theory; 3) the Grievance Asymmetry; 4) the Median Gap Theory; 5) Political Business Cycles; 6) Perpetual Opposition; 7) Asymmetric Cost of Voting; and 8) Self-fulfilling Prophecy. The first and last are a-theoretical. The Coalition of the Minorities Theory is Anthony Downs' (1957); the Grievance Asymmetry is itself seen as a 'complex' (Nannestad & Paldam, 2002), with multiple explanations of its own, some of which I have suggested; the Median Gap Theory has been worked out by Paldam & Skott (1995) and Stevenson (2002), based on Hotelling (1929), Smithies (1941), and Downs (1957); the Political Business Cycle relates to both the Median Gap Theory and the Coalition of the Minorities Theory; I have introduced the role of Perpetual Opposition Parties, based upon Nannestad & Paldam (2002); and I have also introduced the Asymmetric Cost of Voting. I also bring independent perspectives into the Coalition of the Minorities Theory and the Political Business Cycle.

3.1. A Statistical Artefact

Since governments are usually formed by the winners of elections, it is possible that the cost of ruling comes as a statistical artefact of this. Chappell & Veiga (2000: 191, Footnote 5) argued that election winners may have "received a positive error term" and that their next election will more likely have an average error term than yet another positive one. Fiorina & Shepsle (1989: 438) pointed out that "incumbent politicians almost by definition have more supporters than nonsupporters"; thus, there will be more votes to lose than to gain in the next election. It is therefore quite evident that there will be a cost of ruling. I challenge this view in the case of minority coalition governments, which may be formed by a coalition of losers. If the statistical artefact explanation is correct, then the minority coalition should see an increase in support in the next election. This is all easily testable, however, and Narud & Valen (2008) have already tested some hypotheses relating to coalition governments.

3.2. The Coalition of the Minorities

As first suggested by Downs (1957), the theory suggests that it is easier for opposition challengers than incumbents to make inconsistent promises to different groups, thus forming a coalition of these groups against the incumbent. When the "sucker-fraction" (Nannestad & Paldam, 2002: 28) of these groups realizes that the inconsistent promises cannot be fulfilled, it turns against the government it elected; and hence cause the cost of ruling. I suggest that this theory may also explain abstention: "repeatedly inconsistent promises by different parties may lead to increased alienation of the less sophisticated voters" (Ludvigsen, Paper 2: 5).

3.3. The Grievance Asymmetry

The discussion of the Grievance Asymmetry Complex fills the largest part of this paper. A grievance asymmetry occurs when voters punish governments more for legislation or policy outcomes that they disapprove of than they reward for a corresponding positive development. For instance, in the VP-functions, a grievance asymmetry may be that governments are punished more for increasing unemployment than they are rewarded for a corresponding decrease in unemployment. I find five possible explanations for the grievance asymmetry. Three have been suggested by others: *information asymmetry* (gains are more difficult to recognize than losses) (Lau, 1985; Nannestad & Paldam, 1997; Yang & Holzer, 2006); *loss aversion* (voters place greater weight on measurable losses than on prospective gains) (Lau, 1985; Tversky & Kahneman, 1992; Levy, 1997); and principal-agent theory (voters control politicians by punishing disapproved behaviour more than they reward approved behaviour) (Fiorina & Shepsle, 1989; Nannestad & Paldam, 1997).

The two remaining explanations have been forwarded by myself and are based upon a disaggregation of the voters: First, I consider the possibility that voter heterogeneity plays a role in the Grievance Asymmetry Complex, but in a way so that between-group grievance asymmetry is actually caused by within-group *gratification asymmetry.* That is, it is possible that there is no difference in how easy it is to aggravate voters, but that some voters are more easily gratified than others. I illustrate this with an example of taxation as a source of government revenue, arguing that altruists who oppose a tax decrease will respond just as harshly as egoists who oppose a tax increase if the policy they oppose is introduced, but that altruists who support a tax increase may respond less positively than egoists who support a tax decrease if their preferred policy is introduced. This is simply because the egoist may see immediate effects of policy change, while it may take some time for altruists to see the effects of increased taxes (increased redistribution or government

spending). If the two groups are about equal in size, "the aggregated negative response to a tax increase will be stronger than the positive response to a tax decrease. Hence, the existence of a gratification asymmetry between groups adds to the grievance asymmetry at the aggregate" (Ludvigsen, Paper 2: 8).

It is also possible that the gratification asymmetry exists at the aggregate and as a *direct* cause for the cost of ruling: voters may have increasing expectations from one election to the next (Narud & Aardal, 2007), i.e., "increasing expectations may lead voters in one election to be harder to gratify than the same voters in the preceding election, and thus increasing numbers of those voters who seat a new government will vote against this government at the next election" (Ludvigsen, Paper 2: 7). I refer to this as *longitudinal gratification asymmetry.*

The final possible explanation for the grievance asymmetry is the possibility that it does not matter to altruists whether they are directly affected by the macroeconomic indicators – such as inflation, unemployment, real disposable income, economic growth – but that it matters to the egoists. An egoist moving from a secure job to an insecure job will suddenly start caring about unemployment levels and thus start punishing the government for increased unemployment and rewarding for decreased unemployment – until he or she is out of the insecure category. The moment they are secure, egoists ought to stop rewarding improved conditions for others.

A problem with these micro- and meso-level explanations is that they cannot explain relatively stable costs of ruling, since government policies as well as group sizes are likely to vary across time and space.

3.4. The Median Gap Theory

The longitudinal gratification asymmetry can also be contested: In the Median Gap Theory (Paldam & Skott, 1995; Stevenson, 2002), "What is seen as the increasing impatience of voters ... may in fact be an increasing number of centrist voters without any fixed allegiances but with a rational desire to have their concerns addressed" (Ludvigsen, Paper 2: 8). In other words, centrist voters who alternate their vote between two blocs may act according to their rational interest. Their constantly shifting vote may keep the two blocs from drifting away from the median *voter* and towards the parties' own median *members.* This, I argue, is easily testable, especially with a framework similar to the Narud & Valen (2007) study of voters, party elites, and issues.

3.5. Political Business Cycles

Political Business Cycles include both opportunistic and partisan periods (Mueller, 2003). In other words, they occur not only for the purpose of winning elections, but also for the purpose of implementing the parties' respec-

tive policy preferences. Based upon Downs (1957) and Stevenson (2002), I argue that there may be an *asymmetric ideological elasticity* between opposition parties and governing parties, where the leaders of opposition parties have more freedom to propose policy closer to the median voter than the median party member, while the leaders of incumbent parties are expected by the party members to respond to the preferences of the median member. Nannestad & Paldam (2002) have argued that the Median Gap Theory only allows for very small partisan cycles if the median voter is the gravitating force, and I add the case of minority and coalition governments to this: first, minority governments in PR systems must negotiate with the *median MP* throughout its term, while coalition governments should be expected to enter into coalition negotiations with their ideal positions rather than their election promises; thus, the agreements reached by majority coalition governments should lead to relatively stable equilibriums with little room for partisan cycles.

However, loss aversion among members of governing parties may not be an aversion against electoral defeat but against the loss of their hold on policy. If this comes into play, even the slightest partisan move within the PBC may be difficult for the leaders of the incumbent party to counter when the next election nears; thus, ideological immobility within the incumbent party gives room for a more opportunistic opposition. Hence, the asymmetric ideological elasticity, and the median voter moves to the opposition as a consequence.

3.6. Perpetual Opposition Parties

The cost of ruling may be "an effect of opposition policy proposals rather than government policy outcomes" (Ludvigsen, Paper 2: 10). Voters can judge the incumbent upon its record (revealed policies), while opposition parties – especially those with no or only a distant government record – must be judged by its promises (declared policies) (Nanestad & Paldam, 2002). In some situations, you may have parties that never expect to be held accountable for their promises (perpetual opposition parties) but nevertheless attract voters. The cost of ruling will then be a natural consequence of the growth of such parties. Sooner or later, however, this growth will stop, "either because there are no more voters to attract (for instance, if the party becomes too extreme) or because the party *will* be held accountable if it grows large enough or becomes mainstream" (Ludvigsen, Paper 2: 10).

3.7. Asymmetric Cost of Voting

Where the Coalition of the Minorities Theory suggests a positive correlation between abstention and the cost of ruling, my argument about the Asymmet-

ric Cost of Voting implies the possibility of a negative correlation between the two. Following the first theory, voters who are uninspired with the government may either vote for the opposition or stay at home. Following the second theory, voters who are inspired by the opposition will be easier to mobilize than voters who are inspired by the incumbent. Ceteris paribus, anyone making the effort to vote is more likely to support the opposition than the incumbent. This, I argue, is because a vote for the incumbent only has one use: support; whereas a vote for the opposition serves two uses: support for the opposition *and* protest against the incumbent (there is little sense – particularly in multi-party systems – in viewing a vote for the incumbent as a protest against the opposition). Thus, the net cost of voting for the incumbent is higher than voting for the opposition, and opposition voters should therefore be easier to mobilize. "If this is the case, then the asymmetric cost of voting is part of the explanation for the cost of ruling" (Ludvigsen, Paper 2: 11).

3.8. A Self-Fulfilling Prophecy

Finally, the cost of ruling may become a self-fulfilling prophecy (Nannestad & Paldam, 2002). Parties realizing that their time is limited due to the cost of ruling "may be quick to ditch their centre-seeking election promises and pursue their real preferences, especially if they do not need to accommodate the median MP" (Ludvigsen, Paper 2: 12).

3.9. Discussion and conclusion

I provide a couple of suggestions for empirical tests, primarily based upon Narud & Valen's (2007) framework for studying the issue positioning of voters and party elites, and a number of research questions are offered. I do not answer them, however, and I argue that my readers should understand why by this point: the range of theories and the long list of questions necessitate several empirical studies. As well as this catalogue of theories.

4. Meta-analysis as review

We should not trust single samples of empirical results. One knows not to generalize about the world from interviewing one person. Do we know not to generalize about the world from reading one study? A study may argue convincingly for the generalizations made from it, but reading only one study may give a false impression of the population it is drawn from, and another study may arrive at different conclusions. What then? If we have two studies with different results, we may draw the conclusion that the world lies somewhere between the two, possibly near the average. We may attempt to weigh the results according to some quality standard (theoretical fit, methodological soundness, prestige of the author or journal). All things equal, we are either likely to average the results or discard them as *unstable* or *worthless.*

In the field of educational research, Glass (1976: 8) wrote that "we find ourselves in the mildly embarrassing position of knowing less than we have proven". This is a field in which quantitative analyses abound, but I will argue that the quote can be safely applied to most forms of social research: it is a tedious task to keep track of published results – and not least of their incongruities. The most common way of doing so is through narrative literature reviews, where a select number of different ways of arriving at different conclusions are discussed. More elaborate reviews are sometimes conducted in which results are counted and synthesized, and attempts at integrating the results are made – although often through dichotomies: significant vs. non-significant (Glass, 1976). Such reviews are still being conducted – more than 30 years after Glass wrote his critique of non-quantitative reviews of quantitative research.

Narrative reviewers try the best they can – given their tools – to aggregate and conclude from heterogeneous studies. One illustration of this is taken from Geller & Singer (1998), which is the (to date) most comprehensive attempt at reviewing the literature on international conflict:

> Despite the strong scientific norms in favor of reproducibility, the inducements away from reproducibility – and thus, comparability – are often powerful. That means that even though several studies are intended to test the same theoretical model, they often will not. Partly this is data and measurement problem, with individual investigators measuring the same variables in different ways, observing different regions of the world or looking at different historical periods. Furthermore, we can use different research designs, postulating different time lags between predictor and outcome variables, computing moving averages over time spreads of differing lengths, using different transformations to cope with the historical outlier cases, and

> assuming the reciprocal effects of our predictor variables to be additive in some designs and multiplicative in others.
>
> This lack of perfect – or even proximate – similarity from one study to the next means that even if they all point to the same conclusion, we cannot assume theoretical convergence, nor can we assume that they point to differing conclusions even if the statistical results are quite dissimilar (Geller & Singer, 1998: 3)

However, we should not forget that heterogeneous studies are the result of differences in sampling, stratification, estimation, and publication. Meta-regression analysis is able to deal with these differences, and Geller & Singer apparently had the motivation for a statistical meta-analysis but did not apply the tools. Instead, they wished for "greater attention to the canons of reproducibility and comparability" (ibid: 4) and fatalistically stated that they merely had to make do with "rather idiosyncratic approaches to the measurement of our variables, empirical domain, and research design" (op.cit).

It appears as though Geller & Singer ignored the tools of meta-analysis due to what they claimed was a mixed success by Rummel (1985), which was "the best-known effort to explicitly compare and combine the statistical results" (Geller & Singer, 1998: 4) of the field of international conflict.[2] Instead, Geller & Singer aimed to "summarize and synthesize a large proportion of these empirical findings and then to integrate them into as coherent an explanation of modern interstate war as possible" (ibid: 3).

Valuable as such efforts are, they run into several problems, especially if one wishes to arrive at firm conclusions about a given effect. Conventional narrative reviews are prone to the reviewer's subjectivity with regards to which studies are included or emphasized, how the results should be interpreted, what causes the between-study variation and inconsistencies, and which model specifications to favour (Stanley & Jarrell, 1989). Narrative reviewers would be naturally inclined to place weight on studies that are well-designed or published within the 'right' journals. As Glass (1976) pointed out, however, even a study that is poorly designed or published in a non-prestigious journal may arrive at valid conclusions. The obvious way to avoid

2 Rummel (1985) offers an example of an elaborate form of vote counting, where the results of each study are counted, assigned values, and weighed according to the author's evaluation of their relevance and importance. The weighed averages are then t-tested. However, this is a rather simple technique that was criticized as early as 1971 (Light and Smith) (Glass, 1977), because it is still very much dependent upon the author's own biases (which Rummel himself was aware of and tried to correct for). The technique is also not able to measure or control for publication selection bias.

such subjectivity is either to include *everything* or to draw a random sample (Stanley & Jarrell, 1989). Nevertheless, the issue of *publication selection bias* remains: narrative reviewers may ignore or discount statistical results when encountering inconsistent findings and possibly overemphasize the results when the findings are coherent (Leamer & Leonard, 1983; Stanley & Jarrell, 1989). Still, even if the reviewer is able to stifle his or her own biases, authors, editors, and referees may still have preferred significant results over insignificant ones and results confirming theory or expectations over results that contradict them (Hedges, 1992; Stanley, 2005; Stanley & Jarrell, 1989). If results are difficult to explain, they are also going to be difficult to publish.

Thus, Geller & Singer were wrong to discard meta-analysis. Glass (1976) argued that simple summary and *vote-counting* are "too weak [methodologies] for the complexity of the problem … The proper integration of research requires the same statistical methods that are applied in primary data analysis" (ibid: 6). But then again, a "proper integration" may have been premature in 1998, since there have been a number of great advances in the application of meta-regression analysis since then. For instance, the aggregation of estimated effects may result in rather complex error structures (Glass, 1977). A simple OLS at the meta-analytical level is therefore not sufficient, even if the primary data analyses used OLS. In other words, the proper integration of research often requires *more advanced* statistical methods than those applied in primary data analysis. The introduction of formal tests for *publication selection bias* represents another advance. This is the topic for the next chapter.

5. Publication selection bias

> To be acceptable to one's scientific peers, research findings must be original, replicable, significant and relevant to the existing body of theory. Little room is left for the result of a sound research process that does not fulfil all of the above criteria (Lehrer et al., 2007: 52)

The issue of publication selection bias, i.e., the tendency for authors, editors and referees to prefer results that confirm theory over inconclusive or contradictory results, is of utmost importance when conducting any form of literature review – not only quantitative reviews – and a review that does not control for this will itself be biased. This insight dates back to at least the late 1950s and early 60s (Bakan, 1966; McNemar, 1960; Medawar, 1963; Melton, 1962; Sidman, 1960; Sterling, 1959), and the intellectual effort and creativity invested in dealing with this problem is quite impressive (see more on this in the next chapter).

Publication selection basically means that some studies or results are selected for publication, while others are filed away. Hunter & Schmidt (2004) preferred to call the resulting bias an *availability bias.* Other names for this include *file-drawer bias, retrieval bias* and *source bias* (ibid). One way of attempting to overcome this is to try and collect *everything* that has been written, and include dissertations, working papers, conference papers, manuscripts, etc. Rosenthal (1994) called this *the fugitive literature.* I have chosen not to pursue the fugitive literature relating to my work. Instead, I have strictly limited my data collection to studies published by publishing houses or peer-reviewed journals.

Publication bias is not necessarily a Law of Nature, and evidence exists that some literatures are unbiased (Hunter & Schmidt, 2004). My article on inflation in the UK is one of these. Hristos Doucouliagos, whom I visited at Deakin University, is currently working on various meta-meta-analyses and has received data from a number of meta-analysts, primarily within economics. The impression thus far appears to be that most – but not all – literatures are biased (*MAER-net* workshop presentation, Nancy, 18.10.2008; Doucouliagos & Stanley, 2008).

One form of publication bias is the one most often referred to when these biases are explained, viz. that results are published according to their significance levels (see for instance Begg, 1994; Hedges, 1992; Hunter & Schmidt, 2004; Sterling, 1959). This bias is found in Paper 3, and it seems in that case to be a result of the most prolific author often using a general-to-specific approach without reporting the general models. Thus, regressions of models with insignificant variables may never even have made it to a printer.

Another form of publication bias relates to the research process itself in that authors "systematically misrepresent the process by which the conclusions have been reached" (Begg, 1994: 400; Medawar, 1963). We are obviously biased in how we pick our research topics. Working on topics of no interest to us would result in unmotivated research – and would probably not be very interesting to read. However, it would help if social scientists would keep 'lab' journals of their research process (myself included), though it is unlikely that anyone would be interested in going through a huge stack of years of random scribbling when reviewing a 25-page manuscript as long as it is not a matter of life and death or $21,000,000 research grants.[3] Our cherry-picking of topics and data is therefore likely to continue unchecked.

The conclusions a narrative reviewer will draw about the effects in a given literature will normally reflect the aggregate results from the population of studies. The narrative reviewer may reflect little upon the problems of not having accessed all studies – and be even less concerned about the results that have never made it beyond a computer's virtual memory. Even still, the narrative reviewer may be cherry-picking "a couple of dozen studies from the obvious journals" (Glass, 1976: 4), so one in fact ends up with a second-order publication bias. Even with a complete population of studies, a narrative review of the cases I study would lead to erroneous conclusions regarding the relationship between British government popularity and each specific effect, as well as about the interrelationship between the effects and hence about the overall British VP-function (Equation 1). Thus, it is of utmost empirical importance to control for publication bias.

No matter which scientific philosophy one adheres to, one cannot trust a literature that is biased by publication selection. This insight becomes a philosophical topic in and of itself. A student once asked me to explain my work, and the look in his eyes clearly communicated the frustration growing within him: How could he trust Science if its reporting is biased? Had everything he had been told for four years been a lie? Had he wasted years of his life and taken student loans to study abroad for nothing? I told him that no, he should not consider it a waste; rather, it was a lesson in critical reading. Elementary school pupils learn not to trust everything written on the Internet, but they still use it. Older students learn not to trust everything written in a newspaper, but they still read them. But have we ever told our university students not to trust something written in a peer-reviewed journal?

How, then, should results be reported and assessed? There have been numerous recent calls to focus more on insignificant or negative results, to

3 The average size of grants to biomedical research from the Bill and Melinda Gates Foundation in 2003 (Moses et al., 2005).

find new ways of testing theory, and to develop new ways of publishing (Coleman, 2007; Lehrer et al., 2007; Young et al., 2008). As Young et al. point out, it is somewhat paradoxical that authors continue to compete for publication in the limited space of printed journals in the digital age.

One cannot quite argue that the standard for academic publication is paradigmatic (Kuhn, 1962), since we do have several choices in terms of where and how to publish; however, I will argue that there are anomalies that surface – such as the contention made by Young et al. (2008) that more prestigious journals publish less accurate results (see also *The Economist,* 2008)[4] – and lead to calls for a shift in publication standards. Paradigm or not, such a shift – or adjustment – is likely to be upsetting.

Indeed, investigating publication bias is often unwelcome. Journal editors may object to being investigated in this fashion and reject paper submittals out-of-hand. If the meta-analyses are sent for review, the reviewers are likely to be included in the population of studies subject to meta-analysis. It is therefore necessary to tread carefully, especially when being a young meta-analyst. Meta-analyses can be controversial for other reasons as well: washing results for biases and the heterogeneity of specifications and reports may produce evidence that is contrary to (politically correct) conventional wisdom; cf. the debate in Denmark after Doucouliagos & Paldam (2008) found that foreign aid has no effect on economic growth (*Jyllandsposten*, 2008; *Politiken*, 2008; *Kristeligt Dagblad*, 2008).

However, investigating publication bias is not only of concern to those who benefit from such bias (authors who confirm expectations, NGOs, aid recipients, pharmaceutical companies, neo-conservative politicians), but arguably even more so for those at risk of being hurt by publication bias (patients, aid donors, soldiers, students, authors who contradict expectations). There is growing concern with the issue of biased conclusions from medical research – with recent reports in international media (*The Economist*, 2008), and Danish media (*Dagbladet Information,* 2008) – which is possibly the field of study with the gravest implications for human life. We are aware of the likelihood of publication bias, we are concerned about its implications, and we have the tools to control and adjust for it (at least for quantitative studies). I will therefore argue that it is impermissible to avoid the issue of publication bias when attempting to synthesize published results.

My approach to this is through *meta-regression analysis* (MRA). Where *meta-analysis* "refers to the analysis of analyses" (Glass, 1976: 3), meta-

4 This, Young et al. (2008) argue, is due to authors overbidding their results in an artificially scarce, oligopolistic market of prestigious journals in a competition for publication, funding, and status.

regression analysis is a "regression analysis of regression analyses" (Stanley & Jarrell, 1989: 161). Simply put, it is a systematic, quantitative way of conducting reviews of quantitative studies, and thus a way of estimating effects across studies, despite heterogeneous approaches and results within a literature.

Although the possibilities of meta-regression analysis are hinted at by the first author to coin the term *"meta-analysis"* for systematic quantitative research reviews (Glass, 1976; 1977), the term *"meta-regression analysis"* first appears to have been proposed by Stanley & Jarrell (1989). Its most obvious applicability is for research topics with inconclusive or disputed cumulative findings. As I have pointed out, differences in research design, units of analysis, coding, statistical techniques, errors, etc., may all contribute to variance in the results and thus to blurring the conclusions. MRA is able to see through this as long as there are more estimates than differences and as long as the published results are quantifiable, comparable or standardizable, and come with some form of sample statistics. MRA is also useful for topics where authors seem to *agree*. This is because MRA can easily reveal and control for the degree of publication selection bias. As I see it, the introduction of formal tests (and corrections) of publication selection bias constitutes the most important recent advancement of meta-analysis, and the introduction of meta-analysis to Nordic and British Political Science is my contribution.

6. An introduction to meta-regression analysis

The basic steps in a formal test for publication selection bias are presented in Paper 3. I will describe the logic here in a slightly different way, using a hypothetical case of opinion polls as an example.

6.1. How to detect publication bias?

Consider an election between two candidates. On Election Day, the preferences of the electorate are measured at the polling stations. The preference for a candidate P equals the actual election outcome, α, so that

$$P = \alpha \qquad \textbf{[E2]}$$

Since the full population of actual voters has participated, there is no sampling error (ignoring the absentees, who have chosen to not voice their opinion), and if the election was fair and votes counted correctly, there is no measuring error. In an opinion poll, there will be both sampling error and measuring error. If these errors are random, the preference of the full population will equal the aggregate response, α, plus a random error, ε, so that

$$P = \alpha + \varepsilon \qquad \textbf{[E3]}$$

If the errors are random, then sampling error should not matter for the calculation of the average result from opinion polls. However, the errors may not be random, and a non-random bias is likely to be correlated with any sample statistic. A relevant statistic in the case of opinion polls would be sample size. As N increases, so does the precision of the surveys, the random error decreases, and the aggregate response becomes more trustworthy (see Figure 6.1). However, it is necessary to use a sample statistic that is negatively correlated with the precision of the studies in order to assume α as the unbiased aggregate response. If I were to study published empirical results, as I do in the meta-analyses, I could use the standard error of the effects, because the precision decreases as the standard error increases. In the case of opinion polls, I could use the inverse of N in order to have a negatively correlated statistic, so that

$$P = \alpha + \beta \frac{1}{N} + \varepsilon \qquad \textbf{[E4]}$$

Equation 4 is the simple form of the meta-regression model with control for publication bias, where β will be insignificant if the results are randomly scattered around α as the precision drops, as illustrated in Figure 6.2. This simple

MRA is usually modelled with the standard error of the estimate instead of the inverse sample size as the independent variable.

Figure 6.1. Hypothetical values of *P* versus *N*. (no bias, average equals population preference (dashed line))

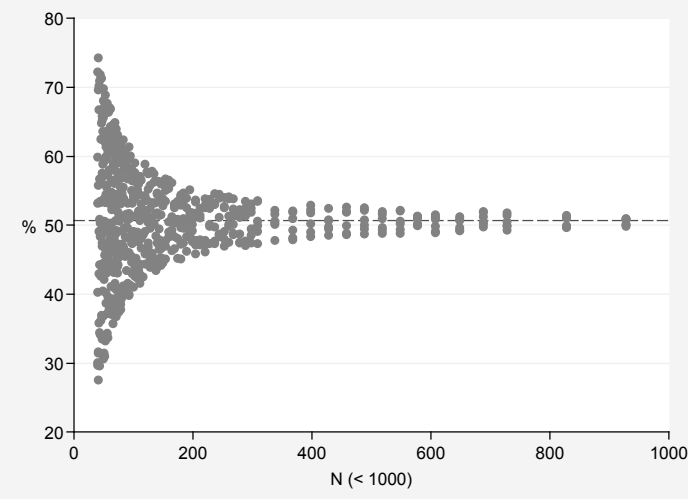
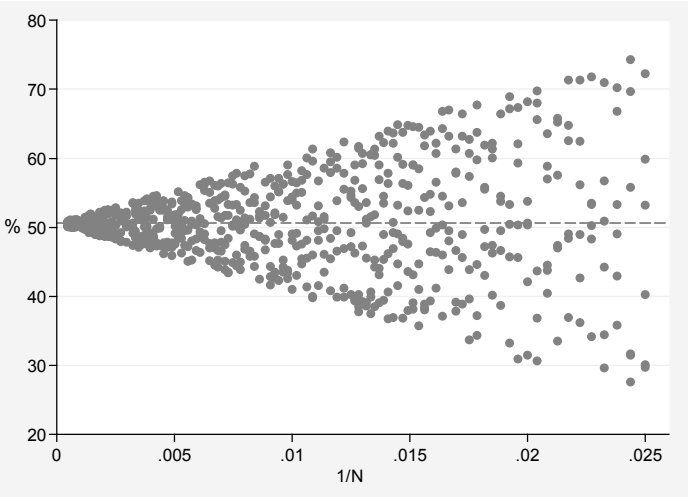


Figure 6.2. Hypothetical values of *P* versus *1/N*. (no bias, average equals population preference (dashed line))



Although the model is simple, the amount of intellectual effort invested in its development is rather astonishing. As I have pointed out, the issue of publication bias stems back at least to the 1960s. The more 'recent' chronology of its evolution goes from Light & Pillemer (1984: introducing the funnel plot of the relationship between effect size and sample size (not the inverse) – cf. Figure 6.1), Stanley & Jarrell (1989: introducing the MRA), Card & Krueger (1995: first introducing sample statistics to the MRA), Görg & Strobl (2001: continuing along the lines of Card & Krueger), to Stanley (2005: setting the standard for the approach I take in my studies). The numerous alternative

statistical procedures developed alongside the MRA can be reviewed in Begg (1994), Hunter & Schmidt (2004), Macaskill et al. (2001), and Stanley (2005), offering testament to the enormous efforts that have addressed this problem – which is dumbfounding in retrospect when considering the simple logic behind the very simple Equation 4, which can be applied to any comparable or standardized statistic.

The relationship between precision and sample statistics will typically be non-linear (Stanley & Doucouliagos, 2007): the increased precision gained from adding 10 more respondents to a sample of 1000 is far from that which is gained when adding 10 more respondents to a sample of 10. Non-linear models will therefore be necessary. Linear regressions on non-linear relationships will otherwise produce incorrect intercepts. Notice that *1/N* not only produces an inverse relationship between precision and sample size, it also transforms a non-linear relationship into a linear one. Both of these points are illustrated by comparing Figure 6.2 with 6.1.

Figures 6.1 and 6.2 are of a hypothetical sample of 690 polls, with N ranging from 40 to 2178,[5] α set at 50.5 pct., and no systematic bias. The average estimate is 50.65 (slightly above 50.5 due to a partially random syntax used to produce these estimates). In other words, the estimates are randomly scattered around 50.65 pct., the random variation decreases reciprocally as the sample size increases (Figure 6.1), and increases linearly as the inverse of N increases (Figure 6.2), and a regression of Equa tion 4 would return a β with zero slope. Both Equations 3 (similar to a simple un-weighted poll-of-polls) and 4 would accurately reflect the population preference.

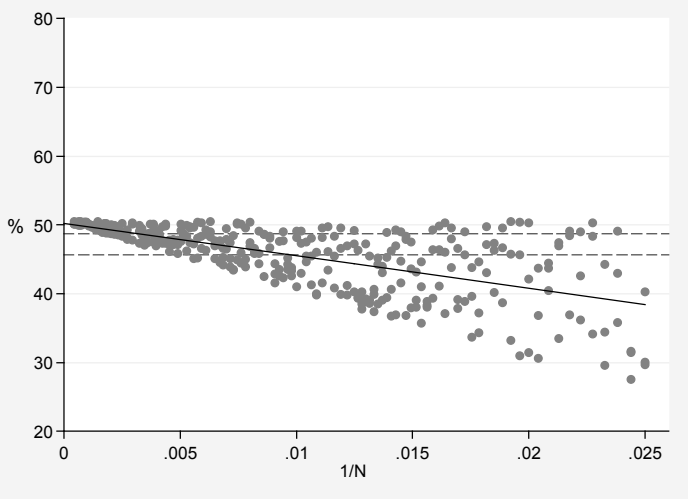
If for some reason there is a systematic bias in the *reporting* of the results so that results over 50.5 pct. are not reported, then the results would appear as in Figure 6.3. A regression of Equation 3 or a simple poll-of-polls (without controlling for sample size) would return a biased constant of 45.7 pct. If this was candidate *P's* vote share of a two-party vote, then even a weighted poll-of-polls (average adjusted for sample size) would predict the wrong winner (P = 48.7 pct.). A regression of Equation 4, however, would reveal a significant effect of sample size, and thus a systematic bias.

The slope of *1/N* will start from an unbiased constant, and Equation 4 would in fact pick the correct winner. A regression of the estimates in Figure 3 gives an α of 50.24, which is slightly below the hypothetical result and has

5 Pollsters never use sample sizes anywhere near as low as 40 (unless you consider an ask-five-people-column in the newspaper a poll), but social scientists frequently do (the recent U.S. election was the 56th), and this is really a story about social scientists, not about pollsters.

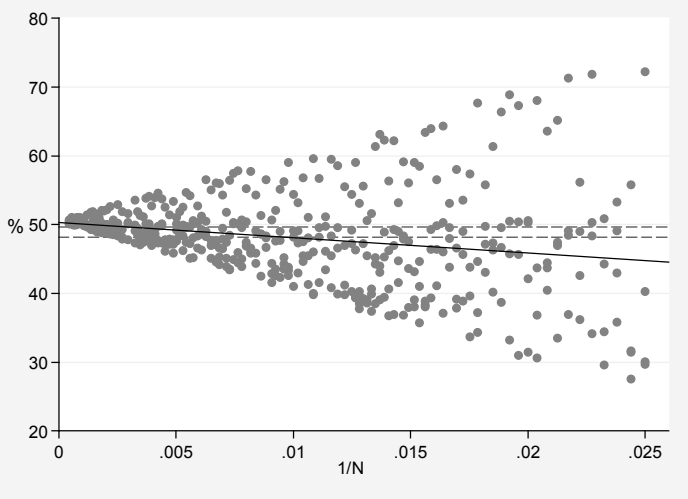
a 95 pct. confidence interval that crosses the 50 pct. threshold; nevertheless, it is a much more precise estimate than what a poll-of-polls could give.

Figure 6.3. Hypothetical values of *P* versus 1/N (heavy bias, average smaller than population preferences)



Notes: upper dashed line is weighted average (48.7%); lower dashed line is unweighted average (45.7%); solid line is OLS regression slope (α = 50.24 [s.e. = 0.35]; β = – 472.43 [s.e. = 29.38])

Figure 6.4. Hypothetical values of *P* versus 1/N (bias, average smaller than population preferences)



Notes: upper dashed line is weighted average (49.6%); lower dashed line is unweighted average (48.2%); solid line is OLS regression slope (α = 50.30 [s.e. = 0.38]; β = – 221.15 [s.e. = 60.06])

An example of a semi-biased poll-of-polls is illustrated in Figure 6.4. In this case, the results are still biased in favour of the losing candidate, but the funnel graph appears to be symmetrical. However, there are three times as many observations below the true population preference as above. Therefore, a poll-of-polls cannot predict the correct winner, but a meta-regression of the poll estimates can (unweighted average = 48.2, weighted average = 49.6, α = 50.3).

These examples have been hypothesized for opinion polls, but they also apply to empirical results published by scientists. The less precision the studies have, the greater the variance in estimated effects, and non-biased literatures would produce estimates on both sides of the genuine effect. If the literature produces systematically biased results, this would be exposed by meta-regressions (notice that the graphical illustration of this differs somewhat from this stylized example in my empirical papers).

Meta-regressions are also convenient when aggregating results from a heterogeneous literature, such as the one described in the quote above from Geller & Singer (1998). Dummy variables controlling for specification differences and other incongruities can be added to Equation 4. Such *moderator variables* can be included as separate terms and then explain variation in the true effect, around α (conventionally known as Z variables), and as dummies interacted with the sample statistic, and then explain variation in publication bias, around β (conventionally known as K variables) (Stanley & Doucouliagos, 2007). This is the *multivariate meta-regression model* (see Equation 5 below for an example).

The bivariate meta-regression model is only able to reveal systematic *reporting* biases for the overall population of estimates. In the case of opinion polls, it would not single out which polling agencies are biased and which are not. The bivariate model would also not be able to reveal whether there is a systematic *response* bias, i.e., where voters systematically lie about their preferences. If they do so to all polling agencies, the multivariate model would not pick it up either, but the multivariate model could be useful if there is some variation in how the agencies safeguard themselves against this problem.

Extending Equation 4, a 'simple' multivariate model controlling for this would be:

$$P_k = \alpha_1 + \sum_{z=2}^{Z=3} \alpha_z Agencydummy_{z,k} + \beta_1 \frac{1}{N_k} + \sum_{l=2}^{L=3} \beta_l \left(\frac{1}{N_k} \times Agencydummy_{l,k} \right) + \varepsilon_k \quad \textbf{[E5]}$$

Subscript k indicates estimate $k = 1, 2, \ldots, K$,[6] the Σ's indicate groups of variables ranging from 2 to a number Z *or* L, respectively (in this case 3 for both), and the subscripts z and l are variable identifiers within these groups. α_1 is the constant, α_z explains variation of support for candidate P around the constant, β_1 is the coefficient of publication bias, β_l explains variation of support around the publication bias, and ε is the error term.

Figure 6.5 shows the same estimates as in Figure 6.2, but hypothesized to come from three different polling agencies. 1.5 pct. of the respondents to each of the polls from one of these agencies have systematically lied about their preference, indicating that they would vote for candidate P when they in fact intended to vote against the candidate. The estimates from this polling agency are indicated with white dots. Because a sub-set of estimates is given a new value, the average of the remaining estimates is not exactly the same as in Figure 6.2.

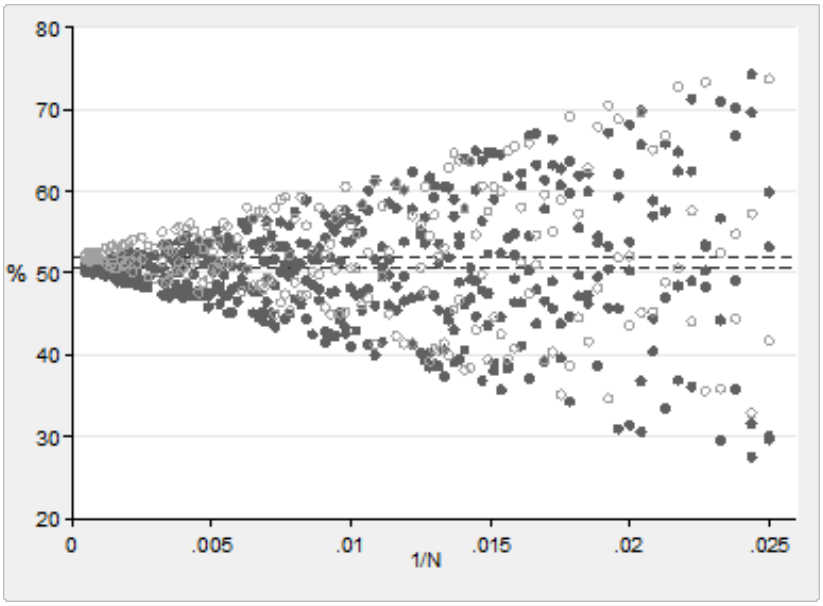
The average support for candidate P for *the two remaining agencies is now* 50.69 pct., while the average for the agency with response-bias was 50.57 pct. before the 1.5 percentage point bias was added, and thus 52.07 pct. afterwards. The two dashed lines in Figure 6.5 indicate these averages. The overall pre-set average is now 51.15 pct., which corresponds with the unweighted average of all estimates in Figure 6.5.

A regression of Equation 4 would return an insignificant β. Not controlling for the different agencies would therefore lead to the conclusion that 51.15 pct. of the population favours candidate P, when the real preference in the population should be randomly scattered around 50.65 pct. A set of general-to-specific regressions of Equation 5, however, washes away this misperception and returns the value of α for the two unbiased agencies at 50.69 (s.e. = 0.33) and a significant coefficient for the biased agency of 1.38 (s.e. = 0.57).[7] The meta-regressions were not able to pick up the full 1.5 percentage point bias because of the small – and insignificant – variations in the original unbiased subsamples.

6 K number of estimates is not to be confused with K-variables. The mix-up is due to an unfortunate case of notation standards that are not entirely compatible when brought together.

7 The general-to-specific approach is necessary because the meta-analyst is interested in the value of the constant – unlike many fields of social science research. Non-significant coefficients create noise around this constant.

Figure 6.5. Hypothetical values of *P* versus 1/N, response-bias in 1/3 of the opinion polls



Notes: white-dotted estimates are from the agency with response-bias; upper dashed line is the average of the biased responses (52.0%); lower dashed line is the average of unbiased responses (50.69%).

I have some problems with cross-level interactions in my papers, so my multivariate models only include moderator variables that explain variation around the true effect (but I still control for bias relating to the sample statistic). These variables are dummies relating to certain study characteristics, such as whether the study was conducted in order to make predictions about future elections, dummies relating to differences in the dependent variables in the original studies, such as whether the estimates are for Labour or the Conservatives – or for both, whether the dependent variable is measured as a voting intention for the governing party or as an evaluation of the governing party's performance, whether the dependent variable is measured in first-order changes or at levels, etc. Then there are dummies for the various ways the estimates of interest are measured and reported, and which control variables are included in the models.

6.2. Some problems, limitations, and possible criticisms

6.2.1. Number of observations

It is possible that there are not enough studies within a field of research, but this can often be corrected for through the inclusion of *all estimates* produced within a study; thus, the units of observations are not the studies, but their estimates. In the case of personal economic expectations and British govern-

ment popularity, I found 99 estimates within 22 studies. However, the question of how many studies are enough for a meta-analysis can be turned on its head: Glass (1977: 362) asked rhetorically, "How many studies can be read and integrated *without* resorting to statistical methods to reveal aggregate findings and relationships? The number is probably very small" (my emphasis).

6.2.2. Quantitative statistics

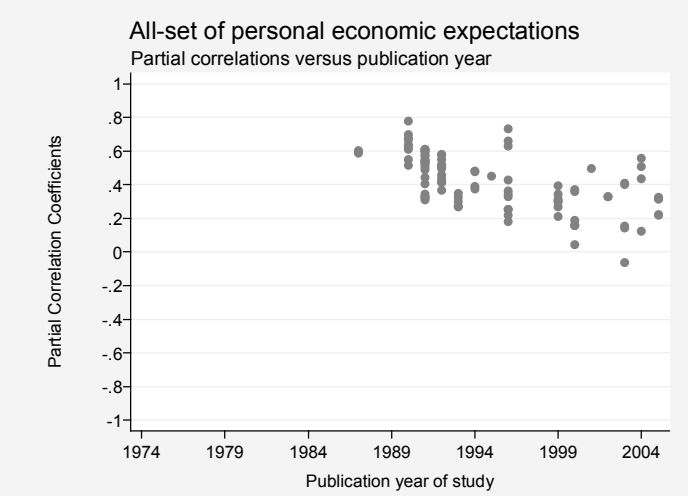
Meta-regression analysis requires that the estimates that are subject to analysis come with sample statistics. This implies quantitative data, and typically data arrived at through regression analyses. MRA also requires that the data can be standardized, which normally entails that already-standardized coefficients, or elasticities, *t*-statistics, or standard errors, are reported and that the degrees of freedom can be calculated. The latter is necessary in order to calculate *partial correlation coefficients* as well as weighted averages. However, although MRA is developed for quantitative studies, publication bias is also likely to exist within qualitative studies. Even if the results from qualitative studies could be standardized, for instance into -1, 0, 1 for conclusions about a hypothesis, the small variation in sample statistics found in these studies make it unlikely that we can apply MRA to them. A table or a funnel plot can nevertheless be constructed on the standardized conclusions and some of the properties of the literature, and a bias could possibly be seen, even if it could not be tested statistically.

6.2.3. Data dependency

The assumption of normally distributed residuals will typically be violated with literature reviews. The introduction of multilevel panel stratification by Brouwer et al. (1999), Rosenberger & Loomis (2000), and Bateman & Jones (2003), was a surprisingly late development within meta-analysis given that the first author to coin the term *"meta-analysis"* for systematic quantitative research reviews (Glass, 1976) recognized the need for clustering or hierarchical modelling, since "the data set to be analyzed will invariably contain complicated patterns of statistical dependence (Glass, 1977: 375). Estimates are simply not independent of each other: first, studies that are of interest to the meta-analyst necessarily report estimates on the same object or population of objects, thus there is a between-study dependence; second, each study may produce several estimates, thus there is a within-study dependence; third, authors may produce several studies or belong to research groups that do, thus there is a within-author/within-group dependence; fourth, there may be cross-level dependencies, i.e., the within-study and between-study errors are not independent of each other. Cross-level interactions will contribute to

this. Fifth, authors may respond to previously published studies, thus there is autoregressive dependence; and finally, authors may respond to methodological developments or the data may be time-dependent, thus there are temporal effects on the results. All-in-all, this creates a highly imbalanced time-series of estimates, where the imbalance lies in the complicated dependencies and in the fact that there are multiple observations of the same phenomenon at the same point in time and that these observations are scattered unevenly across time. The estimates of personal economic expectations in Figure 6.6 illustrate this. There is no getting around the fact that the data are dependent of each other. However, this is an elementary trait of meta-analysis, and I attempt to correct for this through multilevel hierarchical models.

Figure 6.6. Partial correlation coefficients of the personal expectations literature versus publication year



6.2.4. Not original research

Glass (1976: 4) argued that "a good review is the intellectual equivalent of original research", and Rummel (1985: 421-422) argued that "a test against the literature is stronger … than any particular data-based test". I am not so sure about the latter: *one* study may actually find the most representative model of the world, while this will be 'just' another sample in a literature-based test. However, it is not necessarily easy to pick the 'winner' when conducting a narrative review. A simple solution forwarded by Stanley, Jarrell & doucouliagos (2009) is to drop the 90 pct. least precise studies, and only focus on the top tier. However, there is more than just the precision that separates the studies, and I would rather include the information from the less precise studies and apply a multivariate test of the literature with the meta-

analytical techniques with which we can test and correct for publication selection biases and complex data dependencies.

In this sense, even if we assume that the results from data-based tests are not normally distributed, a test against the literature is stronger than any *random* data-based test. Referring to the quote from Glass (1976: 8) in Chapter 4, a meta-analytical review will therefore enable us to know more about what has been proven. Such a review should therefore have stronger implications on a field of research than merely adding another original piece to the puzzle.

This brings us to the empirical analyses of my dissertation.

7. Personal economic expectations and government popularity in the United Kingdom – a meta-analysis

The first empirical paper (Ludvigsen, Paper 3) presents a meta-regression analysis of the impact of personal economic expectations on British government popularity. I have identified 22 studies that have produced a total of 99 estimates of this effect and which could be included in the meta-analysis.

Personal economic expectations (PE) are measured by the aggregate of respondents in surveys indicating the degree to which they think their own household's financial situation will improve or not in the coming year. Using this measure in VP-functions thus provides an estimate of *prospective voting*, i.e., how voters' expectations about the future have an impact upon the popularity of the current government. There are strong theoretical reasons to include such measures in models of government support (see, e.g., Alesina & Rosenthal, 1995). It is not evident to rational choice theorists why voters should punish or reward a government for past outcomes unless the past is used to draw inferences about the future. So why not ask the voters directly what their expectations are?

There are some problems with doing this: Expectations will typically correlate quite strongly with observed macroeconomic outcomes and therefore be problematic to include together with observed outcomes in the same models. Moreover, expectations are likely to be 'noisier' than observed outcomes, as expectations "suffer from severe problems of projection and rationalization" (Hibbs, 2006: 584). Hibbs has therefore argued that "Devising models that bring forward-looking, competency models to macroeconomic data with statistical power poses one of the greatest challenges to future research" (op.cit).

Herein lays the motivation for this paper. A first step, I argue, towards answering this challenge is to synthesize the results from the literature on expectations and assess whether they are coherent despite the projection and rationalization problems. I do not approach these problems from a theoretical angle; instead, I use meta-regression analysis in order to assess whether expectations have an impact across the literature.

I do not synthesize the entire literature on expectations on government popularity. Rather, I take a rather small nibble out of the literature, only examining the case of the United Kingdom and only studying estimates of *personal economic expectations,* as defined above. Nonetheless, this nibble provides 99 estimates, which I see as a sufficient start in the efforts towards responding to Hibbs' challenge.

The paper seeks to answer three questions: 1) Is there a true effect of personal economic expectations on British government popularity? 2) Does publication selection bias the results? And if so, to which degree and in which direction? 3) Can the variation in results be explained? Before answering these questions, I introduce the reader to the models and methodology of meta-regression analysis. I see this as necessary, since very little has been published on the subject within political science. The modelling is therefore somewhat more general than strictly necessary. This may be confusing to the reader, as I end up testing a model which is slightly less intricate than the most extensive general model.

I will not repeat the modelling here, but there are two key aspects of meta-regression analysis which are important to repeat: first, as with ordinary regression analysis, it explains variation in results that is due to specification differences; and second, it estimates and controls for the impact of publication selection bias. Both of these aspects have been seen as problematic in the VP-function literature. This is described in a lengthy quote from Paldam (1991), where the first paragraph describes variation due to specification differences, and the second paragraph describes (part of) the publication selection process:

> First *X* presents an impressive study of the V- or P-function for country *Z*, with a nice theory and – most important – very fine econometric fits: a high R^2, very significant *t*-ratios, and, in addition, some new econometric trick like the ζζ-test from the latest issues of *Esoterica.* Everybody is impressed, until a few years later *Y* demonstrates that, by one little change, *X*'s result collapses. The change may be in the calculation period, the time-series used, or maybe *Y* applies another, even newer, econometric trick. Then *X* manages to get the results back by another little twist, etc. From time to time this causes writers to doubt that there is such a thing as a VP-function – or that the VP-function can survive a real thorough statistical test.
>
> One reason for this predicament is to be found in the sociology of our scientific societies (i.e., economics and political science). The literature is so huge that we can read only a fraction. We all prefer to read something smashing. The publication pressures on the few journals everybody sees are enormous, so articles have to be short to be accepted. All problems with results, qualifications, etc. are therefore normally cut away. People present only the best results obtained after many experiments. Consequently, the results are normally much too good relative to the true model. This is possible due to the flexibility of econometrics: we do have a large tool kit that allows us to work with

> models and improve fits until they become better than the 'true model' (Paldam, 1991: 9-10)[8]

Meta-regression analysis is used to deal with both of these problems. Understandably, Stanley, Doucouliagos & Jarrell (2008) refer to this approach as the "socio-economics of economics research" (from the title). More than just specification and publication differences can be modeled into meta-regressions. The full sociology of research can be coded and measured, such as the gender of the authors, which field of study or research community they belong to, how many years experience they have, etc. Theoretically, even the impact from the authors' caffeine-intake or hours of sleep could be measured if this information was available.

Much of this information is obviously not of interest, and it is not so much the coding of differences which is the problem, but knowing where to draw the line. One obvious limitation is the degrees of freedom available. This particular meta-analysis is the one with the fewest observations, so I have focused on some key specification dif ferences. The variables are discussed in the paper, but the list of variables used in Ludvigsen (Paper 3) is presented in Table 7.1.

The question of publication selection bias is possibly even more important than the specification differences. The latter can be read directly from the studies (although regression analysis is necessary in order to estimate the effects of the differences), while the former may be more difficult to detect without statistical procedures. Reading through a select number of studies – even studies with the exact same specifications – and averaging them as the 'genuine' effect may be seriously incorrect if the results are biased from publication selection.

Meta-regressions can be used to achieve unbiased results. The control for publication selection that I apply rests on the simple assumption that studies with low levels of precision will have more random results than studies with high precision levels. As described in Chapter 6, if there is a non-random publication selection, then a control for the precision of the studies will be significant.

8 The reader should not be surprised, then, to learn that Martin Paldam was the first to introduce me to meta-regression analysis.

Table 7.1. List of variables for paper 3: "Personal economic expectations and government popularity in the UK – a meta-analysis"

Variable	*k*	*Type*	*Values*
Estimate identifiers:			
Study ID	22	Nominal	Unique identification number for each study, used for panel-adjusted standard errors and two-level analysis
Publ. year	99	Date	Publication year of study, used for assessment of residuals
Average data year	99	Date	Average year of observations in the dataset, used for assessment of residuals
Best estimate	21	Dummy	1 if author shows preference for the equation that this estimate belongs to or if subjectively considered by myself to be the best estimate; 0 otherwise. Used for separate models of best estimate.
N	99	Scale	Number of observations used in the model, either as reported or as calculated by myself as based upon other information in the study
df	99	Scale	Degrees of freedom in the model, either as reported or as calculated by myself as based upon other information in the study
Measures of estimate:			
b	99	Scale	Published unstandardised regression coefficient of estimate, used to calculate *t*-value if this is not reported
SE of b	99	Scale	Published or calculated standard error of unstandardized regression coefficient
t	99	Scale	t-value of unstandardized regression coefficient, either as reported or calculated from b/SE
Partial correlation	99	Scale	Partial correlation coefficient, a standardized estimate, *the dependent variable*, calculated as follows: $r_{PCC_k} = \frac{t_k^2}{\sqrt{(t_k^2 + df_k)}} x \frac{t_k}{\lvert t_k \rvert}$
SE of P.C.	99	Scale	Standard error of the partial correlation coefficient, calculated as follows: $\sigma_k = \frac{r_{PCC_k}}{t_k}$
SE^2	99	Scale	Variance of the partial correlation coefficient
Precision effect	99	Scale	Inverse of standard error of partial correlation coefficient (= $1/\sigma_k$)
Study characteristics			
Prediction model	24	Dummy	1 if words such as 'prediction' and 'forecast' are used by the author about the study or parts of the study, either in the title, abstract, or text; 0 otherwise
Book chapter	34	Dummy	1 if estimate comes from a book chapter where; 0 otherwise (= journal article)
Not OLS	22	Dummy	1 if the estimation procedure is not OLS; 0 otherwise

Measure and reporting of dependent variable (in primary studies)			
Variable	*k*	*Type*	*Values*
Labour	13	Dummy	1 if the dependent variable is a measure for the Labour Party only; 0 otherwise
Tory	77	Dummy	1 if the dependent variable is a measure for the Conservatives only; 0 otherwise
Economic management	8	Dummy	1 if the dependent variable is the electorate's evaluation of the government's economic management competence; 0 otherwise
ΔDep	13	Dummy	1 if the dependent variable is measured in first-order changes; 0 otherwise
Poll of Polls	31	Dummy	1 if dependent variable is an average of observations from several agencies; 0 otherwise
Measure and reporting of estimates			
ΔPE	27	Dummy	1 if PE is explicitly measured in first-order changes or as deviation from trend; 0 otherwise
Lagged PE	18	Dummy	1 if PE is measured at lag compared to the dependent variable; 0 otherwise
No t-stats reported	43	Dummy	1 if estimate is reported without t-values; 0 otherwise
Independent variables characteristics:			
Control for lagged dependent	70	Dummy	1 if lagged observations of the dependent variable are included in the model; 0 otherwise
Control for economic management	2	Dummy	1 if a control for the approval ratings of the government's economic performance is included in the model; 0 otherwise
Control for PM/Party approval	9	Dummy	1 if a control for the approval ratings of the prime minister or the party in government is included in the model (if relevant); 0 otherwise
Control for trend	4	Dummy	1 if a linear trend variable is included among the controls; 0 otherwise
Political variables	82	Dummy	1 if events or political variables are included among the controls; 0 otherwise
Multiple lags of PE	7	Dummy	1 if several lags of PE included among the controls; 0 otherwise
Inflation	30	Dummy	1 if a measure of inflation is included in the model; 0 otherwise
Unemployment	51	Dummy	1 if a measure of unemployment is included in the model; 0 otherwise
Other economic variables	55	Dummy	1 if other measures of expectations or other economic variables are included in the model; 0 otherwise

I run both bivariate and multivariate regressions of these data. The bivariate regressions are simply of the standardized effects against the measure of publication bias in order to attain an average effect across the studies which is controlled for bias, but not specification differences. The bivariate regressions show that there is a quite substantial effect after controlling for publication bias but that there also is publication bias in the literature. The weighted average partial correlation of personal expectations on UK government popularity is 0.347, while the effect *after* controlling for publication bias is reduced to 0.25. In other words, the average published effect is nearly 40 pct. stronger than the unbiased average. Nevertheless, an unbiased effect of 0.25 is still quite strong.

I argue that the publication selection bias may be a result of an often-used general-to-specific approach in the literature, where the general models typically go unreported. Only reporting specific models necessarily means that only significant effects are reported, which my meta-regressions testify to. Although the motive of such publication selection is not sinister, it nevertheless gives the readers a biased impression of the results. Drawing a simple average from the literature would then lead to biased conclusions.

The multivariate regressions add specification differences to the model and show that the effects also depend on variation in the specifications. However, the effects from this variation are no greater than I am still able to conclude that there is indeed an effect of personal economic expectations on government popularity in the UK. The default specification of PE (where all moderators are set to nil, including publication bias) gives a partial correlation coefficient of 0.32. The only variable that truly disrupts the effect is the use of economic management evaluations as a control variable in the models. This has a strong, robust, and negative impact upon the effect of PE. I argue that the effect may be due to a strong correlation between economic management evaluations and voting intentions, but it is nevertheless important to be wary of this effect in future studies.

The results from this analysis indicate that it may be worthwhile to include survey data in modelling prospective evaluations of government popularity. I argue that this brings us a step closer to answering Hibbs' challenge. However, it is also very interesting to note that the literature pertaining to personal expectations seems to be quite robust in relation to specification differences, and thus be a response to some of the concerns about the unstable VP-function. Finally, the PE literature also shows that there is no difference in terms of how Labour or the Conservatives have been evaluated. The Responsibility Hypothesis appears to be strengthened by this meta-analysis.

8. Inflation and government popularity in the United Kingdom – a meta-analysis

The second empirical paper (Ludvigsen, Paper 4) is a meta-regression analysis of the impact of inflation on British government popularity. That paper would be too long for publication if I were to include a full variable discussion in it, so the variables are only listed there and the variable discussion is given here instead.

The paper starts with an introduction to the theories behind the VP-function and proceeds with a description of the coding of the dependent variable (the partial correlation of inflation on government popularity) and how publication bias is estimated. The coding of the independent variables is then listed before the results. Of the independent variables, this paper distinguishes between *moderating* and *mediating* variables. *The Handbook of Research Synthesis and Meta-Analysis*, 2nd edition, edited by Cooper, Hedges & Valentine (2009), was published after I had written the first meta-analysis and made me aware of a distinction between these types of variables that is of some theoretical importance but of no practical significance for the analyses. Moderating variables explain variation in the estimates of the meta-analytical dependent variable (inflation) due to operational differences of inflation or the primary dependent variable (government popularity), while mediating variables explain variation in the estimates of inflation due to the other control variables that are included in the primary models. In other words, moderating variables have direct effects upon the estimates of inflation while mediating variables have indirect effects.

The list of independent variables in the paper is *very* long (54 variables). I have only found room to discuss the control for publication bias *before* the analyses. The rest of the variables have been listed in a table with their coding and the mean statistics of the dependent variable for each independent variable. In order to avoid shying away from a variable discussion, I have included the variable discussion in this summary instead. The following is very long and detailed and should be seen as an appendix to Ludvigsen (Paper 4). Readers who want to avoid bogging down in the variable details may jump to Section 8.5 on page 79. However, this discussion – including the footnotes – offers testament to the heterogeneity of the literature, much of which also applies to the two other meta-analyses.

8.1. Moderators of variation in specification of inflation

8.1.1. Lagged observations

Most studies have used a measure of inflation which is observed at the same point in time as the primary dependent variable, these are held as reference category (coded 0 on this moderator), while some studies have used a measure of inflation which is observed at a lag from the primary dependent variable. There are several arguments for doing so. First, economic news may be reported at lags; second, economic news – or the effects of economic events – may take some time to be recognized by the electorate; and third, surveys during any given month may be conducted *before* the events or the release of economic figures. Price & Sanders (1994) used a model with *two* measures of inflation in the same specification; one in the same month as the primary dependent variable and the other lagged two months. They found that "Inflation begins to feed through immediately, while the effect of interest rates is still coming through after a year" (ibid: 302), while Anderson (1995: 94) argued that "one needs to make sure that as many people in the sample as possible have heard news about, or experienced firsthand, the economic conditions of the country." Clarke et al. (1986: 130) also argued that "The assumptions that voters have sociotropic orientations and learn about macroeconomic conditions primarily through the mass media suggest that these variables should operate with a lag of one month".

The studies that report one or more lagged estimates of inflation (at any lag) are Anderson (1995), Bélanger, Lewis-Beck & Nadeau (2005), Clarke, Mishler & Whiteley (1990), Clarke et al. (1986), Hibbing (1987), Lewis-Beck et al. (2004), Mosley (1978), Mughan (1987), Nadeau, Niemi & Amato (1996), Norpoth (1987, 1992), Price & Sanders (1993, 1994), Sanders (1991, 1996, 2000), Sanders & Gavin (2004), Weakliem (1986), and Whiteley (1984, 1986).

The mean statistics and mean difference (MD) tests reveal a significant bivariate difference between lagged and unlagged observations of inflation.[9] However, most studies have identifiable 'best estimates' which authors indicate a preference for or from which they base their conclusions. These best estimates are too few to test in a multivariate setting ($k = 37$) as heterogeneous as this one (54 moderating and mediating variables), but bivariate MD tests indicate no difference between lagged and unlagged best estimates (not

9 Number of estimates in the reference category (unlagged) = 183, average partial correlation coefficients (PC) = -.176 (with standard deviation = .214), average t-value = -1.95 (s.d. = 2.49); number of estimates from lagged observations of inflation = 83, avg PC = -.043 (s.d. = .268), avg t = -0.23 (s.d. = 1.66). Mean difference test for PC (reference – lagged) = -3.96***, MD test for t-values = -6.60*** (critical values: $p < .1$ [*] = 1.645, $p < .05$ [**] = 1.96, $p < .01$ [***] = 2.58).

reported here). The difference in the all-set and the lack of difference in the best-set indicate that authors have landed on appropriate specifications in both cases. Thus, I do not expect a significant coefficient of the moderator of lagged observations after controlling for all specification differences in the multivariate MRA.

8.1.2. Changes

Unto itself, inflation represents a measure of change, as it reflects the change in prices. However, there are two main reasons for also measuring changes in inflation. The first has to do with *stationarity*. As Headrick & Lanoue (1991) correctly pointed out: prices generally rise and rarely decline. In order to attain stationarity, several authors have measured inflation in terms of its first-order changes. A counterargument and alternative approach can be found in Sanders (2005a), who argued that

> de-trending data always carries the risk that, where the trends in different variables are causally related ... the magnitude of statistical relationships can be underestimated. Of course, leaving the trends in carries the opposite risk of overestimating effect magnitudes. However, in the analysis conducted here, because a time term is included in all the functions in order to take account of the 'costs of ruling', any simple linear trend effects are picked up primarily by that time term. This in turn minimises the risk of overestimating the magnitudes of other (non-de-trended) variables (Sanders, 2005a: 177)

The second reason for measuring inflation in changes is theoretical. Voters receive information about macroeconomic conditions in two ways: through their wallets and through information from others. These 'others' are often the media, and the media are more likely to report on changes or deviations from trends than to report on stability or expected developments (Nadeau et al., 2000; Stimson, 1991). It has also been argued that in their first-hand experience of inflation, British voters are more sensitive to unexpected changes than to expected developments (Bélanger et al., 2005; Mughan, 2004; Hibbs & Vasilatos, 1981; Nadeau et al., 2000) and especially to inflation 'crises' (Alt, 1979; Mosley, 1984b).

These reasons – and combinations of them – have lead to several ways in which changes can be measured. The most common measure of change is the use of first-order differences, but other specifications also exist for taking trends, expectations or crises into consideration. The latter specifications are more ad hoc and have therefore been coded together with estimates from first-order changes.

The studies reporting one or more estimates measured from changes are Alt (1979), Borooah & Ploeg (1983), Frey & Schneider (1981, 1982), Headrick

& Lanoue (1991), Hibbs & Vasilatos (1981), Minford & Peel (1982), Mosley (1978, 1984a, b), Norpoth (1992), Pissarides (1980), Sanders (2000, 2005a), Sanders & Gavin (2004), Weakliem (1986), and Whiteley (1984, 1986).

The mean statistics reveal a significant difference between estimates in the reference versus the moderating category; yet again, however, the difference disappears between the best-sets.[10] I therefore expect the moderating variable for changes to inflation to be insignificant in the multivariate MRA, despite theoretical arguments for the opposite.

8.1.3. Nonlinear measures

Some studies have used nonlinear specifications. A moderating variable for this has been coded 1 for estimates that are derived from logarithmic (Whiteley, 1986) or squared (Minford & Peel, 1982) specifications of inflation. No studies used the inverse or other nonlinear measures of inflation.[11] I do not expect this moderator to be significant.

8.1.4. Seasonal adjustment

Eight studies have explicitly mentioned using seasonally adjusted observations of inflation, but most of the studies have not. Pissarides (1980) gave one reason for not using adjusted data, since the typical figures reported by the mass media are unadjusted, but several studies are not very clear as to which specific figures are used. The moderator controlling for seasonally adjusted data has only been coded as present if there is explicit mention of seasonal adjustment or if the models are replications of previous studies where this has been mentioned. The studies that have been coded with this moderator on one or more estimates are Alt (1979), Clarke et al. (1990), Frey & Schneider (1978, 1981, 1982), Hibbs et al. (1982), and Nadeau et al. (1996, 2000).[12] The reliability of this measure is low, and the mean difference tests indicate no difference. I therefore expect this moderator to be insignificant.

---

10 Reference category: k = 116, avg PC = -.171 (s.d. = .312), avg t = -2.00 (s.d. = 3.15); changes in inflation: k = 150, avg PC = -.105 (s.d. = .156), avg t = -0.95 (s.d. = 1.44). MD test for PC (reference – moderator) = -2.10**, MD test for t-values = -3.34***. Statistics for best-sets not reported here.

11 Reference: k = 243, avg PC = -.133 (s.d. = .246), avg t = -1.42 (s.d. = 2.48); nonlinear inflation: k = 23, avg PC = -.147 (s.d. = .146), avg t = -1.27 (s.d. = 1.16). MD test for PC (reference – moderator) = 0.42, MD test for t-values = -0.52.

12 Reference: k = 240, avg PC = -.128 (s.d. = .240), avg t = -1.42 (s.d. = 2.46); seasonally adjusted inflation: k = 26, avg PC = -.189 (s.d. = .221), avg t = -1.34 (s.d. = 1.70). MD test for PC (reference – moderator) = 1.31, MD test for t-values = -0.21.

8.1.5. Perceived

Three studies have used one or more estimates of *perceived* inflation rather than objective indicators. Mosley (1984a, b) was interested in how the media reported economic indicators, and Sanders (2000) was interested in how the respondents themselves perceived inflation. Both found rather strong effects, with an average partial correlation coefficient (PC) for Mosley's six estimates = -.266 (s.d. = .120) and an average PC for Sanders' seven estimates = -.146 (s.d. = .081).[13]

On the other hand, Anderson (1995) – and presumably most others – have followed Kramer (1983) in assuming that

> … it is sensible and acceptable to use objective indicators of economic performance. Kramer argues that people's images of economic performance – while susceptible to occasional errors at the individual level – are typically correct and unbiased in the aggregate, since 'errors in individual perceptions are assumed to be distributed randomly so that aggregate perceptions are accurate and reliable' (Clarke et al. 1992: 54) (Anderson, 1995: 89)

Given both the qualitative difference between objective and subjective indicators and the clear MD test results, I expect that using perceived inflation rather than objective inflation has a significant impact upon the results.

8.1.6. Expected

Expectations seem to be correlated with perceptions (Alt, 1979) and are most likely formed by them. People are likely to infer the past into the future, so the two studies using one or more models with expectations of inflation (Alt, 1979; Minford & Peel, 1982) should resemble the studies that use perceptions. Obviously, this must be tested in a multivariate model, but the simple averages indicate that this may be correct. The average PC for Alt's six estimates is -.231 (s.d. = .103), and the average for Minford & Peel's 15 estimates is -.188 (s.d. = .127).[14] The expectation for the multivariate MRA is the same as with perceptions.

---

13 Reference: k = 253, avg PC = -.131 (s.d. = .243), avg t = -1.34 (s.d. = 2.42); perceived inflation: k = 13, avg PC = -.201 (s.d. = .115), avg t = -2.81 (s.d. = 1.38). MD test for PC (reference – moderator) = 2.00**, MD test for t-values = 3.57***.

14 Reference: k = 245, avg PC = -.129 (s.d. = .245), avg t = -1.42 (s.d. = 2.49); expected inflation: k = 21, avg PC = -.200 (s.d. = .120), avg t = -1.35 (s.d. = 0.87). MD test for PC (reference – moderator) = 2.34**, MD test for t-values = -0.25.

8.1.7. Multiple estimates

Some studies have used several lags or specifications of inflation in the same model. This is a bit challenging to code, and some subjective coding is necessary. I have chosen to use two dummies to code these studies. Estimates are coded as 'principal' for the measure of inflation that is in its simplest form and/or is closest in time to the primary dependent variable, while estimates are coded as 'secondary' for measures of inflation that are in a more complex form and/or are further away in time from the primary dependent variable. Estimates of interacted variables have been dropped since the effects cannot be untangled without re-running the primary regressions. In cases where interacted variables pair with uninteracted variables, the estimates from uninteracted variables are coded as 'principal' while the interacted effects are dropped. Estimates are not coded with either of these moderators if they stem from models with only one measure of inflation.

Studies in which these moderators are coded on one or more estimates are: Borooah & Ploeg (1983), Headrick & Lanoue (1991), Minford & Peel (1982), Price & Sanders (1993, 1994), Sanders (1996, 2000), Sanders & Gavin (2004).

These two moderators are somewhat 'residual categories', so I expect them to be insignificant; nevertheless, I include them in order to control for the possibility that specifications with multiple measures of inflation lead to different results than studies with only one measure of inflation. However, the mean statistics show that the 'secondary' category is significantly different from the reference category and that inflation in the 'secondary' category on average is without effect on government popularity.[15]

8.1.8. Price index

DeHaven (1991) used the level of the consumer price index – instead of changes to it – to measure inflation. The mean statistics indicate that the effects are positively signed.[16] Given the qualitative difference between using the level of prices vs. the change in prices, I expect this specification to lead to different results than the reference specification.

---

15 Reference: k = 199, avg PC = -.151 (s.d. = .260), avg t = -1.65 (s.d. = 2.61); principal: k = 31, avg PC = -.124 (s.d. = .144), avg t = -1.22 (s.d. = 1.43); secondary: k = 36, avg PC = -.048 (s.d. = .150), avg t = -0.26 (s.d. = 1.23). MD test for PC (reference – 'principal') = -0.86, MD test for t-values = -1.35; MD test for PC (reference – 'secondary') = -3.34*** , MD test for t-values = -5.04***

16 Reference: k = 260, avg PC = -.141 (s.d. = .237), avg t = -1.47 (s.d. = 2.40); level of price index: k = 6, avg PC = .146 (s.d. = .020), avg t = 0.94 (s.d. = 0.14). MD test for PC (reference – moderator) = -17.10***, MD test for t-values = -15.11***.

8.1.9. GDP deflator

Two studies have measured inflation by changes in the GDP deflator instead of changes in price indices: Dewan & Dowding (2005) and Pissarides (1980). According to Pissarides (1980: 571-72), the "GDP deflator instead of the retail price index made a small and insignificant difference to the equations' performance", and I do not expect this specification to lead to results that are different from the reference specification.[17]

8.2. Variation in publications and observations

8.2.1. Publication decade

Research and publication practices may vary over time. Study quality, artifacts, and biases may therefore be correlated with publication year (Wood & Eagly, 2009). In addition to controlling for temporal variation in study quality and artifacts, using moderating variables for publication year thus works together with the variance of the partial correlation coefficients in controlling for publication bias. However, a moderating variable for publication year cannot be specified as a continuous variable, since the estimates are pooled and have not been consistently and consecutively published. Instead, I have chosen to use three dummies, one for each decade of publication, with the 1970s as base.

The studies from the 1970s seem to have produced significantly different results than later studies; however, the specifications and the data were also different.[18] Footnote 18 therefore gives no reason for any expectations.

8.2.2. Observation years

Figure 8.1 displays the historical inflation in the United Kingdom for the period covered by this meta-analysis. The time periods covered by the studies may be of importance. Alt (1979) tested the hypothesis that economic condi-

17 Reference: k = 261, avg PC = -.135 (s.d. = .240), avg t = -1.42 (s.d. = 2.42); GDP deflator: k = 5, avg PC = -.091 (s.d. = .114), avg t = -0.97 (s.d. = 1.15). MD test for PC (reference – moderator) = -0.84, MD test for t-values = -0.83.

18 1970s (reference): k = 94, avg PC = -.227 (s.d. = .267), avg t = -2.43 (s.d. = 3.10); 1980s: k = 87, avg PC = -.129 (s.d. = .177), avg t = -1.20 (s.d. = 1.60); 1990s: k = 47, avg PC = -.082 (s.d. = .188), avg t = -0.68 (s.d. = 1.67); 2000s: k = 38, avg PC = -.018 (s.d. = .251), avg t = -0.28 (s.d. = 1.70). MD test for PC (reference – 1980s) = -2.91***, MD test for t-values = -3.40***; MD test for PC (reference – 1990s) = -3.72***, MD test for t-values = -4.34***; MD test for PC (reference – 2000s) = -4.98***, MD test for t-values = -5.30***.

tions only affect government popularity when the economy is 'in crisis'.[19] Headrick & Lanoue (1991: 80) found "that British voters are not consistently attentive to economic fluctuations. When conditions are undergoing only slight change [in absolute as well as relative terms], citizens turn their attention elsewhere", and Sanders (1991) argued in a rather understated manner that the changing economic circumstances in Britain make "universal generalisation about the relationship between popularity and the state of the economy … impractical" (ibid: 258).

Figure 8.1 Inflation in the UK, 1948/6 – 2006/12

Note: Data from Office for National Statistics (2009), series CZBH (annualized percentage change in the retail price index – all items). The uninterrupted line displays the monthly annualized inflation while the dotted line represents the three-year moving average.

Not only have the economic circumstances been changing, so have the political circum stances. Pooling all of the data together without controlling for temporal variation is therefore inadvisable.

Furthermore, publication years are necessarily correlated with observation years, since observations cannot post-date publications. Without control-

19 The three-year moving average in Figure 5.1 follows Alt's (1979) definition of critical rates of inflation. Inflation above this moving average is at crisis levels according to this definition.

ling for observation years, the moderators for publication years will not only pick up variation in study quality, artifacts, and biases, but also temporal variations in the data. In order to avoid this and to separate the effects of different observation years, I have coded the estimates according to which years the primary studies cover. For the same reasons as with publication years, I cannot use a continuous variable to cover observation years. I must use dummy variables instead, and the coding will be a bit rough since one dummy for each year (or month) would consume too many degrees of freedom. I have chosen to use twelve five-year periods: 1947-51, 52-56, 57-61, 62-66, 67-71, 72-76, 77-81, 82-86, 87-91, 92-96, 97-01, 02-06. No observations pre-date 1947 or post-date 2006.

In order to minimize the noise from overlapping observations, I have chosen to code the data accordingly: for each observation year moderator (1947-51 … 2002-06), the value '1' is given if the dataset covers any of the years xxx7 to xxx1 and xxx2 to xxx6, *except* if the dataset ends in a year 7 or 2 or starts in a year 1 or 6. I.e., a dataset with observations beginning in (e.g.) 1956 is given '0' on the 1952-56 moderator and '1' on the 1957-61 moderator. Likewise, a dataset ending in (e.g.) 1957 is given '1' on the 1952-56 moderator and '0' on the 1957-61 moderator. This ensures that the data are coded for their *main* observation periods.[20]

8.2.3. Other estimation procedure than OLS

Studies with one or more estimates (explicitly) derived through procedures other than Ordinary Least Squares regression are Alt (1979), Dewan & Dowding (2005), Hibbs & Vasilatos (1981), Hibbs et al. (1982), and Weakliem (1986).[21]

An expectation of no significance of the moderator variable implicitly assumes that authors apply the correct procedures with their data; however, this is not necessarily the case (see Petersen, 2009).

8.2.4. Reporting standard errors without t-values

The partial correlation coefficients are calculated from the *t*-values of the primary estimates and the degrees of freedom of the primary models. Most estimates are reported with *t*-values, but 86 estimates are reported only with their standard errors (DeHaven, 1991; Dewan & Dowding, 2005; Gavin &

20 Mean statistics for the partial correlations of the twelve observation year moderators (without MD tests) are given in Ludvigsen (paper 4, Table 1).

21 Reference: k = 254, avg PC = -.140 (s.d. = .234), avg t = -1.49 (s.d. = 2.35); number of estimates from non-OLS procedures = 12, avg PC = -.009 (s.d. = .312), avg t = 0.22 (s.d. = 2.84). MD test for PC (reference – moderator) = -1.44, MD test for t-values = -2.04**.

Sanders, 2003; Goodhart & Bhansali, 1970; Hibbs & Vasilatos, 1981; Hibbs *et al.*, 1982; Nadeau et al., 1996, 2000; Norpoth, 1987; Sanders, 2000, 2005a; Sanders & Gavin, 2004). *t*-values can be easily calculated, but the calculation may result in crude estimates in some cases. Consider a small effect size reported with two decimals: If this effect size is 0.02 and the standard error is 0.01, then the same effect size with *four* decimals may be anything between 0.0150 and 0.0249. The standard error may be anything between 0.0050 and 0.0149. Thus, the *t*-value may be anywhere between (0.015/0.0149) = 1.007 and (0.0249/0.005) = 4.98.

It is not fair to assume that authors deliberately round off small values in order to overstate the significance levels – but this cannot be ruled out. On the other hand, it *is* fair to assume that authors who round off small values are likely to understate the significance of their findings. If this is the case, I expect the moderator that controls for reporting standard errors without *t*-statistics to be significant *and with the opposite sign* of the constant (i.e., that the significance levels of the estimates that are reported solely with standard errors have been understated). However, the mean statistics give cause for concern.[22] The difference is not significant for partial correlations, but it is for *t*-values, and with a direction so that *t*-values that are reported are substantially *smaller* than the *t*-values that have been calculated from the standard errors. If significant, a moderator for estimates reported without *t*-values is therefore likely to have the *same* sign as the constant (which is the combined value of all reference categories).

8.2.5. Published in a book vs. published in a journal

The quality of the studies may be of concern. I have not collected working papers, conference contributions or other unpublished work. Most studies have been published in peer-reviewed journals, but some have been published in books, either in books authored by the researcher(s) (Alt, 1979; Anderson, 1995; Borooah & Ploeg, 1983; Mosley, 1984b; Norpoth, 1992) or in a chapter in a book co-edited by one of the researchers (Hibbs & Vasilatos, 1981). The mean statistics provide no indication that there is a bivariate difference between the two publication channels.[23] Thus, there is no reason for

22 Reference (estimates reported with t-values): k = 180, avg PC = -.119 (s.d. = .229), avg t = -0.96 (s.d. = 1.67); estimates reported with standard errors only: k = 86, avg PC = -.165 (s.d. = .257), avg t = -2.35 (s.d. = 3.28). MD test for PC (reference – moderator) = 1.41, MD test for t-values = 3.68***.

23 Reference: k =224, avg PC = -.137 (s.d. = .250), avg t = -1.47 (s.d. = 2.54); published in books: k = 42, avg PC = -.119 (s.d. = .164), avg t = -1.07 (s.d. = 1.35). MD test for PC (reference – moderator) = -0.61, MD test for t-values = -1.50.

immediate concern that the estimates are in some way moderated by the publication channel.

8.3. Variation in measures of the dependent variable in primary studies

According to the responsibility hypothesis, all governments are held accountable for the economy. Grievance asymmetries have been found, however, meaning that governments may be punished more for a negative development than they are rewarded for a corresponding positive development (Monroe, 1984; Nannestad & Paldam, 2002; Stevenson, 2002), and the responsibility hypothesis cannot be tested against its alternatives if both parties are included in all models. Also, regardless of party,

> The student of government popularity in Britain faces an embarrassment of riches when it comes to selecting a measure of such popularity. A short list of alternatives includes the approval of the 'government' as a collective body, the approval of the prime minister personally, or the voting support for the governing party in a hypothetical election next week. In addition, any of these measures have their opposite sides in the form of disapproval or support for opposition parties and leaders (Norpoth, 1992: 136)

This variation must therefore be controlled for.

8.3.1. Conservatives vs. Labour

The primary dependent variable has in one or more models been support for the Conservative Party in Alt (1979), Anderson (1995), Borooah & Ploeg (1983), Clarke et al. (1986, 1990), DeHaven (1991), Goodhart & Bhansali (1970), Headrick & Lanoue (1991), Mosley (1978, 1984a, b), Nadeau et al. (1996), Norpoth (1987, 1992), Sanders (1991, 1996, 2000), for Labour in Anderson (1995), Borooah & Ploeg (1983), Gavin & Sanders (2003), Mosley (1978, 1984a, b), Sanders (2005a), Sanders & Gavin (2004), or for *both* in Alt (1979), Anderson (1995), Borooah & Ploeg (1983), Bélanger et al. (2005), Dewan & Dowding (2005), Frey & Schneider (1978, 1981, 1982), Goodhart & Bhansali (1970), Headrick & Lanoue (1991), Hibbing (1987), Hibbs & Vasilatos (1981), Hibbs et al. (1982), Lanoue & Headrick (1994), Lewis-Beck et al. (2004), Minford & Peel (1982), Mosley (1978, 1984a, b), Mughan (1987), Nadeau et al. (2000), Pissarides (1980), Price & Sanders (1993, 1994), Sanders (2000), Weakliem (1986), and Whiteley (1984, 1986). This latter category is coded as the reference category in the MRA.

Two of the four theories suggest that there is a difference between how Labour is evaluated versus the Conservatives, and all of them when adding asymmetry to the mix. Empirical findings suggest the same: Sanders (2000:

289) found that "although Conservative governments are damaged by rising inflation … the equivalent damage inflicted on Labour governments is almost three times greater"; and Minford & Peel (1982) pointed out that the parties have applied genuinely different policy instruments in dealing with inflation: "Labour governments respond to higher expected inflation with increased budget deficit, Tory governments by reducing deficit" (ibid: 265). Arguably, voters could react differently to economic outcomes under different parties if the parties apply different economic policy instruments.

However, it is impossible to see any indication of this in the mean statistics.[24] Nonetheless, the four hypotheses given in the introduction can be tested by these moderators:

- H_1: the Responsibility Hypothesis suggests that both parties should be hurt by rising inflation.
- H_2: the Clientele Hypothesis suggests that the Conservatives should be rewarded for inflation (or at least punished less than Labour).
- H_3: the Salient Goal Hypothesis suggests that the Conservatives should be punished harder for inflation than Labour.
- H_4: the Stability Hypothesis suggests that voters are attracted to 'responsible' parties, in which case both parties should be rewarded for increasing inflation, especially during crisis periods. However, Sanders (2000) argued that the Labour Party was no credible opposition to the Conservatives between 1982 and 1994, for which reason there should be no effect of inflation whatsoever for the Conservatives during that period.

8.3.2. Prediction model

Some studies have been conducted with the aim to predict or to develop a prediction model for later elections. These models may be more parsimonious than models used to explain observed popularity, but this should be picked up by the other moderating variables. However, the motivation behind the studies may impact the writing or review process in ways that go undetected by the other moderators.

---

24 Both parties (reference): k = 171, avg PC = -.149 (s.d. = .255), avg t = -1.66 (s.d. = 2.65); Conservatives only: k = 72, avg PC = -.100 (s.d. = .190), avg t = -0.99 (s.d. = 1.85); Labour only: k = 23, avg PC = -.130 (s.d. = .245), avg t = -0.84 (s.d. = 1.58). MD test for PC (reference – Conservatives) = -1.68*, MD test for t-values = -2.26**; MD test for PC (reference – Labour) = -0.35, MD test for t-values = -2.12**; MD test for PC (Labour – Conservatives) = -0.55, MD test for t-values = 0.37.

Estimates in studies where words such as 'prediction' or 'forecast' have been used to describe the study or parts of it, either in the title, abstract, text, or models, have been coded with a '1' on this moderator. These studies are: Bélanger et al. (2005), Frey & Schneider (1981, 1982), Lewis-Beck et al. (2004), and Sanders (1991, 1996, 2005a).

Although the mean difference tests are significant,[25] the standard deviations of the prediction estimates are too large to support any expectations about the results from the MRA.

8.3.3. Vote- versus popularity function

I have included studies based upon both election results (vote function studies) and opinion polls (popularity functions). Popularity functions have more observations and typically more control variables due to the more degrees of freedom. The variable for the variance of the partial correlation coefficients and the mediating control variables will therefore pick up most of these differences. However, there are other empirical and theoretical differences between vote functions and popularity functions as well. Whiteley (1984) argued that these two types of studies should be treated differently, because

> The circumstances of the electorate casting their votes after a highly publicized and mobilizing Presidential or General election campaign is clearly very different from that of individuals answering questions about their hypothetical voting behaviour and issue preferences, when an actual election is years away. We should expect mid-term polls to give a less valid measure of actual voting behaviour and issue opinions, than surveys carried out during the election campaign; one is hypothetical whereas the other is concrete. This implies that the poll series contain a lot of measurement error or 'noise' compared with the situation when individuals actually have a choice, at election time. Similarly politics is not a very salient phenomenon in the day-to-day lives of the average voter. Therefore we might expect issue opinions to be particularly influenced by one-off events which receive wide coverage such as a political scandal, or a financial row. This means that the error terms in the popularity function estimating equations are likely to be much more complex than they are assumed to be in existing models (Whiteley, 1984: 5)

It is therefore necessary to control for possible differences between these two types of studies, which may not be covered by the variance or the mediators. Five studies used election results as dependent variables: Bélanger et al.

25 Reference: k = 253, avg PC = -.145 (s.d. = .222), avg t = -1.49 (s.d. = 2.39); prediction models: k = 13, avg PC = .081 (s.d. = .415), avg t = 0.08 (s.d. = 2.29). MD test for PC (reference – moderator) = -1.95*, MD test for t-values = -2.40**.

(2005), Hibbing (1987), Hibbs & Vasilatos (1981), Lewis-Beck et al. (2004), and Mughan (1987).

However, I have my doubts as to whether the empirical and theoretical differences translate into a statistical difference in the multivariate MRA. The mean statistics show a *VERY* large standard deviation of the estimates from vote functions.[26]

8.3.4. Voting intentions, government lead, popularity of the prime minister, and evaluation of the government's economic competence

With the percentage of the actual or intended vote as the reference specification, three moderators control for which specification of popularity has been used.

First, several studies have used the *government lead* over the main opposition party as the dependent variable in *all* specifications: Borooah & Ploeg (1983), Dewan & Dowding (2005), Frey & Schneider (1978, 1981, 1982), Headrick & Lanoue (1991), Hibbs & Vasilatos (1981), Hibbs et al. (1982), Lanoue & Headrick (1994), Minford & Peel (1982), Mosley (1978, 1984a, b), Pissarides (1980), Price & Sanders (1994), and Whiteley (1984, 1986). In addition, Frey & Schneider (1981) used the lead in one of their two specifications, and Goodhart & Bhansali (1970) used the lead in all but three of their 35 models.

The main difference between measures of government lead vs. voting intention is the scaling (hence a reason for using standardized correlation coefficients in the MRA), but there may be other differences as well. One of these is where the difference in popularity between the two main parties is constant because of the third-party vote. In this case, the government lead does not change between two observations. Another is where the popularity of the governing party does not change but the popularity of the main opposition party does. Since elections are zero-sum games, the popularity of the government is not independent of the popularity of the other parties. This dependence translates into different movements in the two types of series. It is therefore necessary to control for this difference in the MRA.

Second, the *evaluation of the government's economic competence* has been used as a dependent variable. Questions can be raised regarding the relationship between aggregate evaluations of the government's economic record and voting intentions. Some studies have modeled economic compe-

26 Reference (popularity functions): k = 257, avg PC = -.137 (s.d. = .208), avg t = -1.45 (s.d. = 2.35); vote functions: k = 9, avg PC = -.038 (s.d. = .702), avg t = -0.18 (s.d. = 3.40). MD test for PC (reference – moderator) = -0.42, MD test for t-values = -1.12.

tence as exogenous to voting intention (these are coded among the mediating variables – more on this later), but three studies have used this measure as the dependent variable. These are Alt (1979), Gavin & Sanders (2003), and Sanders & Gavin (2004).

Questions can also be raised as to whether this measure is exogenous to – and sufficiently distinct from – voting intentions. Sanders (1999) and Sanders & Gavin (2004) demonstrated a strong relationship between economic competence and voting intentions (Pearson's $r = .95$ between January 1991 and March 1997 (Sanders, 1999)). Given this strong correlation together with the few effects estimated with this specification, I do not expect the moderating variable for this to be significant.

This is related to a discussion of whether general support for the government or the popularity of the prime minister is exogenous to voting intentions. Exogeneity has been claimed by Clarke & Lebo (2003), Clarke et al. (1990), Clarke & Stewart (1995), and Clarke, Stewart & Whiteley (1997) against the view of Sanders (1999) and Sanders, Ward & Marsh (1987, 1991). Although correlations between voting intentions and support for the prime minister were found to be .83 (Clarke & Stewart, 1995 (August 1979 to April 1992)), .94 (Clarke et al., 1997 (January 1992 to November 1995)), and .94 (Sanders, 1999 (July 1979 to May 1997)), Clarke et al. (1997) argued that exogeneity tests proved their view, while Sanders (1999) rejected the tests on theoretical grounds.

Some studies have used the evaluations of party or party leader performance as control variables in specifications of voting intentions (more on these mediators later), and three studies have used the *popularity of the prime minister* as the dependent variable in some of their models: Goodhart & Bhansali (1970) and Norpoth (1987, 1992).

These three types of dependent variables are held against the percentage of the vote or the percentage response to variants of the question: "If an election was held tomorrow, which party would you vote for?"

There are marked bivariate differences between estimates of inflation from models of voting intentions compared to other models.[27] The most interesting finding in the mean statistics is the complete lack of average effect

27 Reference (percentage of the vote or intended vote): k = 84, avg PC = -.063 (s.d. = .264), avg t = -0.60 (s.d. = 1.74); government lead: k = 172, avg PC = -.162 (s.d. = .214), avg t = -1.72 (s.d. = 2.48); economic evaluations: k = 4, avg PC = -.257 (s.d. = .332), avg t = -1.55 (s.d. = 2.00); PM popularity: k = 6, avg PC = -.245 (s.d. = .306), avg t = -3.81 (s.d. = 4.51). MD test for PC (reference – lead) = 2.99***, MD test for t-values = 4.15**; MD test for PC (reference – economic evaluations) = 1.15, MD test for t-values = 0.93; MD test for PC (reference – PM popularity) = 1.42, MD test for t-values = 1.73*.

on voting intentions. The number of studies of PM popularity or economic competence is inadequate to show a significant difference here, but the difference is highly significant between voting intentions and governing party lead. This is surprising, but the reason may simply be due to the data used. The average observation year for the voting intention models is 1980, while the average observation year for the lead models is 1966. There are some other differences regarding the data used as well, so the difference may not exist when these differences have been controlled for in the MRA.

8.3.5. First-order changes of the primary dependent variable

Should government popularity be specified by its levels or by changes to it? One answer to this is of course technical: If a pre-whitened model is used to overcome problems of autocorrelation, then the specification will obviously be in first-order changes on both the left- and right-hand sides, but several authors have measured the effect of changes in inflation on the *level* of popularity, so the question is not entirely technical.

Whiteley (1986) gave a theoretical reason for specifying changes on both the left- and right-hand sides:

> We know from a long tradition of research into electoral behavior that voting intentions are influenced by short term factors or issues, and also by long term predispositions which in the Michigan tradition of electoral analysis are subsumed under the heading of party identification. Most popularity function models have ignored long term predispositions altogether (e.g. Goodhart and Bhansali 1970), and as such are misspecified. Such factors are omitted largely because data on party identification is not usually available in the polls (Whiteley, 1986: 46)

Whiteley (1984, 1986) overcame the problem of long-term predispositions by estimating a popularity function of first-order changes with the reasonable assumption that long-term predispositions do not change over brief periods of time.

Other studies with models of first-order changes in the dependent variable are Alt (1979), Clarke et al. (1990), and Sanders & Gavin (2004). There are slight differences in the mean statistics, with a smaller standard deviation than average estimate for estimates from models of first-order changes, but the differences are not significant.[28]

---

28 Reference: k = 243, avg PC = -.132 (s.d. = .247), avg t = -1.40 (s.d. = 2.49); primary dependent in first-order changes: k = 23, avg PC = -.160 (s.d. = .120), avg t = -1.48 (s.d. = 1.03). MD test for PC (reference – moderator) = -0.95, MD test for t-values = -0.30.

8.3.6. Nonlinear measures of the primary dependent variable

Clarke et al. (1990) specified government popularity not only in changes, but also by the natural log. Other studies with nonlinear models of government support are Borooah & Ploeg (1983), Hibbs et al. (1982), and Price & Sanders (1993, 1994). There are notable differences in the mean statistics, but large standard deviations render the differences insignificant.[29]

8.3.7. Source of the primary dependent variable

Most studies have used data from Gallup. Estimates from these studies, together with vote function studies, are held in the reference category, while two moderators control for other primary dependent data. Two studies have compared Gallup data with data from the National Opinion Polls Ltd (Goodhart & Bhansali, 1970, and Minford & Peel, 1982). A moderator is therefore used to control for the estimates produced from NOP data. Most of the studies carried out by David Sanders have used averaged opinion data from several agencies (Price & Sanders, 1993, 1994; Sanders, 1991, 1996, 2005a; Sanders & Gavin, 2004). This is also controlled for by a moderator.

There are some significant differences in the mean statistics,[30] but this may be due to publication years, as the list of studies illustrates.

8.4. Mediating variables

The mediating variables also come in many operationalizations, but controlling for all of this would exhaust all of my degrees of freedom. Instead, I have chosen to simply control for the *presence* of various mediating variables.

8.4.1. Unemployment

Together with inflation, unemployment is the most common control variable in models of government popularity. Some have avoided putting these two variables together due to a strong negative correlation between the two (e.g., Clarke et al., 1986), while others have allowed for the combination due to the contested Phillips curve (e.g., Norpoth, 1992). Norpoth (1987) found that inflation was insignificant when estimated both with and without a control

---

29 Reference: k = 245, avg PC = -.139 (s.d. = .232), avg t = -1.48 (s.d. = 2.34); nonlinear primary dependent: k = 21, avg PC = -.084 (s.d. = .306), avg t = -0.64 (s.d. = 2.98). MD test for PC (reference – moderator) = -0.79, MD test for t-values = -1.25.

30 Reference: k = 227, avg PC = -.129 (s.d. = .243), avg t = -1.38 (s.d. = 2.38); NOP: k = 18, avg PC = -.279 (s.d. = .167), avg t = -2.75 (s.d. = 2.64); poll-of-polls: k = 21, avg PC = -.069 (s.d. = .197), avg t = -0.62 (s.d. = 1.97). MD test for PC (reference – NOP) = 3.54***, MD test for t-values = 2.13**; MD test for PC (reference – poll-of-polls) = -1.30, MD test for t-values = -1.65*.

for unemployment and thus argued that unemployment did not have a mediating impact upon the (lacking) effect of inflation.

Most of the authors have some models with both variables. Due to general-to-specific reporting, however, a few end up with models with only one of the two included in their final (or reported) models. For instance, Lewis-Beck et al. (2004: 282) found after "a series of multivariate tests [that] the unemployment variable, as well as the other economic variables, manifested little if any independent effect", and that "the survivor variable was inflation, which carries an even larger coefficient once supressor effects are removed via multivariate analysis" (sic). Whiteley (1984: 19) arrived at a similar result: "Inflation appears to be more salient than unemployment". Whitely therefore argued that "It appears to make sense for governments to concentrate on inflation, if necessary at the expense of unemployment, since the political cost of doing so appear to be less than the reverse policy" (ibid: 20).

On a similar vein, Sanders (2000) expected a lack of the Phillips curve for Thatcher and Major as a strategic result:

> Thatcherite discourse, in short, sought to neutralise unemployment as a source of political support. Note, however, that no such effort was made with regard to inflation; on the contrary, the defeat of inflation was expressly regarded as key objective of macro-economic policy. These rather different discursive strategems for unemployment and inflation imply two very different predictions for the way that these variables should have related to government support during the Thatcher and Major administrations: governing party support during the Thatcher and post-Thatcher period should have been unrelated to unemployment but negatively correlated with inflation (Sanders, 2000: 278)

However, Sanders found no effect of either for Thatcher and Major.

Four studies have *not* included unemployment and inflation together in any reported models: Bélanger et al. (2005), Clarke et al. (1986), Lewis-Beck et al. (2004), and Sanders (2005a).

If there is indeed a Phillips curve that has an impact upon political support in Britain, then the mediator for unemployment will be significant and with an *opposite* sign of the constant. Should the mediator be significant with the *same* sign as the constant, then there is a *positive* correlation between unemployment and inflation. This indicates support for the responsibility hypothesis (if negative) and the stability hypothesis (if positive) on both infla-

tion *and* unemployment. The mean statistics indicate that there is indeed a mediating effect of unemployment upon the effects of inflation.[31]

8.4.2. Real disposable income or real wages

The third most common economic variable in models of British government popularity is a measure of disposable income, which may obviously correlate with inflation (Alt, 1979). 13 studies have included a measure of real disposable income or real wages in the same models as inflation: Alt (1979), Borooah & Ploeg (1983), Frey & Schneider (1978, 1981, 1982), Goodhart & Bhansali (1970), Hibbing (1987) Hibbs & Vasilatos (1981), Hibbs et al. (1982), Minford & Peel (1982), Mosley (1984b), Sanders (1996), and Weakliem (1986).

Interestingly, the mean statistics show absolutely no difference between the effect size of models with and without an income variable, but they do show a significant difference in the significance level of the inflation estimates.[32] Obviously, models with more control variables may have fewer degrees of freedom, and this is indeed the case.[33]

8.4.3. Personal economic expectations

Sanders et al. (1987) introduced personal economic expectations into models of UK government popularity. This was supported by Clarke et al. (1990) but criticized by Norpoth (1992). Based upon accumulated evidence, Sanders (2005b: 49) claimed that "subjective economic perceptions … were frequently found to be stronger predictors of party support than objective macroeconomic indicators", and Sanders (1996: 205) argued that "what has become increasingly clear over the last decade or so … is that the effects of the real economy on UK voters' political preferences are strongly mediated by voters' economic perceptions".

Eight studies have combined personal economic expectations and inflation in their models: Clarke et al. (1990), Gavin & Sanders (2003), Nadeau et al. (1996, 2000), Sanders (1991, 1996, 2000), and Sanders & Gavin (2004).

---

31 Reference: k = 29, avg PC = -.045 (s.d. = .316), avg t = -0.53 (s.d. = 1.79); control for unemployment: k = 237, avg PC = -.145 (s.d. = .226), avg t = -1.52 (s.d. = 2.44). MD test for PC (reference – mediator) = 1.66*, MD test for t-values = 2.68***.

32 Reference: k = 198, avg PC = -.134 (s.d. = .253), avg t = -1.55 (s.d. = 2.60); control for income: k = 68, avg PC = -.134 (s.d. = .194), avg t = -1.01 (s.d. = 1.61). MD test for PC (reference – mediator) = -0.01, MD test for t-values = -2.02**.

33 Average degrees of freedom in the reference category = 117 (s.d. = 83); avg d.f. in models with control for income = 72 (s.d. = 57). MD test for d.f. (reference – mediator) = 4.87***.

Mean statistics do not indicate that economic expectations have a mediating effect between inflation and government popularity.[34]

8.4.4. GNP, GDP or consumption growth

The first runner-up after 'the big two' among economic variables in vote- and popularity functions is generally a measure of economic growth (typically gross national or domestic product) (Lewis-Beck & Paldam, 2000). This has been less common in models of British government support. Only Booroah & Ploeg (1983), Dewan & Dowding (2005), Norpoth (1992), and Pissarides (1980) have applied this in models in which inflation has been included.[35]

8.4.5. Other economic variables

A number of other economic variables have also been used. Interest rates are the most notable of these other economic variables, as they are "of peculiar importance in Britain because of the large proportion of the population with variable interest rate mortgages on their homes" (Sanders, 1991: 236). Interest rates may be strongly correlated with inflation, however, since both affect buying power (ibid).

Borooah & Ploeg (1983) used exchange rates and tax ratio; Dewan & Dowding (2005) used exchange rates; Gavin & Sanders (2003) used interest rates; Goodhart & Bhansali (1970) used monetary balance and interest rates; Hibbs et al. (1982) used the exchange rate; Minford & Peel (1982) used interest rates; Nadeau et al. (2000) used interest rates and elite economic forecasts; Pissarides (1980) used the exchange rate and tax rate; Price & Sanders (1993, 1994) used the interest rate, as did Sanders (1991); Sanders (1996) used interest and tax rates; and Sanders & Gavin (2004) used interest rates and a measure of the balance in economic news. I have made a residual category of all these models. I do not expect a mediator for this to have any significant effect upon the effect sizes of inflation. It should nevertheless be controlled for, since these models have slightly fewer degrees of freedom and thus lower precision, which may impact the significance levels and therefore the standardized partial correlation coefficients and ultimately also the publication of the results.[36]

---

34 Reference: k = 241, avg PC = -.136 (s.d. = .246), avg t = -1.43 (s.d. = 2.48); control for expectations: k = 25, avg PC = -.116 (s.d. = .154), avg t = -1.19 (s.d. = 1.40). MD test for PC (reference – mediator) = -0.59, MD test for t-values = -0.77.
35 Reference: k = 256, avg PC = -.135 (s.d. = .241), avg t = -1.42 (s.d. = 2.49); control for GDP or GNP: k = 10, avg PC = -.124 (s.d. = .149), avg t = -1.23 (s.d. = 1.42). MD test for PC (reference – mediator) = -0.22, MD test for t-values = -0.41.
36 Reference: k = 230, avg PC = -.144 (s.d. = .235), avg t = -1.52 (s.d. = 2.36); control for 'other' economic variables: k = 36, avg PC = -.071 (s.d. = .256), avg t =

8.4.6. Party- or administration-specific fixed effects

Several models cover more than one party, and often more than one administration. However, each administration may have a unique level of support with them into their governing period (Mueller, 1970), so it may be relevant for authors to control for administration-specific fixed effects. Indeed, very few authors have controlled for this. Of the 30 studies covering more than one party or administration, only eight control for different parties or administrations in their models (Borooah & Ploeg, 1983; Dewan & Dowding, 2005; Goodhart & Bhansali, 1970; Hibbs & Vasilatos, 1981; Hibbs et al., 1982; Minford & Peel, 1982; Sanders, 2000; Whiteley, 1986).

Instead of coding the presence of administration or party effects, which would result in a high number of missing observations, I have coded a mediating variable for studies covering more than one administration or party *without* controlling for party-specific effects (administration-specific effects are covered by the party mediator). This mediator is therefore not a control for specification, but rather for *misspecification*. Again, the mean statistics show no difference in the effect sizes from these models but show a difference in the significance of these models.[37] This may also be due to more degrees of freedom in the models that 'should' include a control for party versus the models that are 'correctly' specified.

8.4.7. Control for PM or party approval

A few studies have controlled for the popularity or approval ratings of the prime minister or the party of the prime minister in their models of government support (Bélanger et al., 2005 (gvt record); Clarke et al., 1986 (leader satisfaction); Gavin & Sanders, 2003 (both); Goodhart & Bhansali, 1970 (PM pop); Lanoue & Headrick, 1994 (PM pop); Lewis-Beck et al., 2004 (gvt record); Nadeau *et al.*, 1996, 2000 (PM pop)).

The discussion as to whether prime minister or party approval ratings are sufficiently distinct from voting intentions was mentioned under the moderators for the primary dependent variables. If one accepts that they are, then another discussion arises: *whether it is the approval ratings of the prime minister or of the governing party that affect voting intentions.* There is also the question of whether to use the percentage of voters who think party leader X

-0.70 (s.d. = 2.57). MD test for PC (reference – mediator) = -1.61, MD test for t-values = -1.81*.

37 Reference: k = 134, avg PC = -.121 (s.d. = .256), avg t = -1.10 (s.d. = 2.14); 'missing' control for party: k = 132, avg PC = -.148 (s.d. = .256), avg t = -1.73 (s.d. = 2.60). MD test for PC (reference – mediator) = 0.84, MD test for t-values = 2.16**.

is best suited for the job as prime minister rather than the approval ratings (Anderson & Ward, 1996; Nadeau et al., 1996). There is no room here to engage in that discussion, and the mediating variable in the MRA simply controls for whether any such measure is used in the primary models.[38]

8.4.8. Control for economic management evaluations

Another possible primary dependent variable that has also been used as a control variable is the evaluation of the economic competence of the respective parties. This measure is rather new, and has been reported by Gallup since 1991 (Sanders, 1996). However, Alt (1979) also included a measure of the government's economic record in one of his models. The studies that have applied this measure as a control variable *together with inflation* in models of government support are Alt (1979), Gavin & Sanders (2003), and Sanders & Gavin (2004). There are only four estimates of inflation that have been produced with this control variable, but it may nevertheless be important, so I include a control for this as a possible mediating variable in my MRA. The average of these four estimates is indeed significantly different from the average estimate in the reference category, with inflation having a *positive* (non-significant) effect on government popularity in the models that include economic management evaluations.[39]

8.4.9. Control for lagged observations of the primary dependent variable

"Government popularity is essentially a first-order autoregressive process" (Alt, 1979: 114), and it is quite easy to find very high explained variance in time series of support for political parties simply by using the lagged dependent as the only independent variable. Data from Norway, for instance, show that the average R^2 for AR(1) models of monthly party support from September 1997 through June 2009 was 0.82.[40]

Using an AR(1) model of government popularity is also "convenient because it means that only current values of the independent economic vari-

---

38 Reference: k = 250, avg PC = -.145 (s.d. = .221), avg t = -1.45 (s.d. = 2.34); control for party or PM approval: k = 16, avg PC = -.034 (s.d. = .407), avg t = -0.73 (s.d. = 3.14). MD test for PC (reference – mediator) = -1.74*, MD test for t-values = -0.90.

39 Reference: k = 262, avg PC = -.138 (s.d. = .238), avg t = -1.44 (s.d. = 2.40); control for management evaluations: k = 4, avg PC = .116 (s.d. = .189), avg t = 0.63 (s.d. = 1.12). MD test for PC (reference – mediator) = -2.66***, MD test for t-values = -3.57***.

40 Using election results for months with elections, poll-of-polls for other months from September 1997 to June 2009 (courtesy of Bernt Aardal) and interpolated data for each July, I obtained a minimum R2 = .49 for the agrarian Centre Party and a maximum R2 = .95 for the Socialist Left Party.

ables need to be included in the model instead of a more complex lagged model formulation" (Anderson, 1995: 93). Other AR( ) models are also coded with this mediator. The following studies have one or more AR( ) models: Anderson (1995), Borooah & Ploeg (1983), DeHaven (1991), Dewan & Dowding (2005), Frey & Schneider (1978, 1981, 1982), Gavin & Sanders (2003), Goodhart & Bhansali (1970), Headrick & Lanoue (1991), Minford & Peel (1982), Nadeau et al. (1996, 2000), Norpoth (1987, 1992), Pissarides (1980), Price & Sanders (1993, 1994), Sanders (1991, 1996, 2000, 2005a), Sanders & Gavin (2004), Weakliem (1986), and Whiteley (1984).

The effect sizes of inflation produced from AR( ) models appear to be different from models that do not control for lags of the dependent variable.[41]

8.4.10. Control for trend

As previously mentioned, Sanders (2005a) controlled for trend in order to avoid de-trending his data. The later studies also had an interest in the trend variable in-and-of-itself, in that it is a measure of the 'cost of ruling' (see, e.g., Nannestad & Paldam, 2002). Authors have been aware of cyclical trends from the very beginning of the literature (Goodhart & Bhansali, 1970) and controlled for such cycles (see the next mediating variable), but only nine studies have used a variable that increases by one for each observation in order to control for trend (and capture the cost of ruling): Alt (1979), Anderson (1995), Borooah & Ploeg (1983), Bélanger et al. (2005), Dewan & Dowding (2005), Goodhart & Bhansali (1970), Hibbs & Vasilatos (1981), Lewis-Beck et al. (2004), and Sanders (2005a).[42]

8.4.11. Control for cycle

"The cycle function … is necessary to evaluate movements in mass political support that are independent of the duration of administrations and the timing of elections" (Hibbs & Vasilatos, 1981: 36). Cyclical effects are operationalized somewhat differently from trends, and can be specified autonomously from the cost of ruling (Frey & Schneider, 1978). The operationalization used by Anderson (1995) illustrates this:

---

41 Reference: k = 156, avg PC = -.161 (s.d. = .279), avg t = -1.77 (s.d. = 2.83); controls for lagged dependent: k = 110, avg PC = -.096 (s.d. = .159), avg t = -0.90 (s.d. = 1.47). MD test for PC (reference – mediator) = -2.43**, MD test for t-values = -3.24***.

42 Reference: k = 218, avg PC = -.120 (s.d. = .214), avg t = -1.08 (s.d. = 1.96); control for trend: k = 48, avg PC = -.199 (s.d. = .322), avg t = -2.91 (s.d. = 3.44). MD test for PC (reference – mediator) = 1.62, MD test for t-values = 3.55***.

> To capture the surge prior to election day, I created a variable called 'back-swing,' which was coded +1, +2, and +3 in the three months before an election, and 0 otherwise. Similarly, I created a variable called 'post-election,' which is coded 6 in the first month after the election and gradually returns to 0 in the first half year afterward and remains at 0 otherwise (Anderson, 1995: 99).

Nearly half the effect sizes are estimated with similar controls, but the control for cycle has been very popular among some of the more prolific authors. The authors who have used a control for cycle in one or several models are Alt (1979), Anderson (1995), Bélanger et al. (2005), Dewan & Dowding (2005), Frey & Schneider (1978, 1981, 1982), Goodhart & Bhansali (1970), Hibbs & Vasilatos (1981), Minford & Peel (1982), Mosley (1978, 1984a, b), and Pissarides (1980).[43]

8.4.12. Control for decay

Regarding time-related effects, Hibbs & Vasilatos (1981), Hibbs et al. (1982), and Norpoth (1987, 1992) have used decay-parameters interacted with events and economic outcomes in order to control for decaying effects through time.[44]

8.4.13. Control for political variables

"It would be controversial, if not downright silly, to claim that the economy moves public opinion regardless of a country's political context. Politics is a crucial *independent and mediating* factor that partially determines the dynamics of support (Eulau and Lewis-Beck 1985)" (Anderson, 1995: 87, *my italics*). "Harold Macmillan famously referred to 'events, dear boy' when asked by an interviewer what most affected his government's electoral fortunes" (Sanders, 2005a: 177). However, most political events can only be controlled for using dummy variables, so it is not the events themselves that have been controlled for, but rather the time at which the events took place.

There is a long list of political variables that have been controlled for in the British VP-functions: the Suez Crisis in 1956 (Price & Sanders, 1993, 1994 – not in reported models); the Profumo Scandal in 1963 (Lanoue & Headrick, 1994); the devaluation of the pound in 1976 (Whiteley, 1986); the 'Winter of Discontent' before Thatcher's rise to power (Anderson, 1995;

---

43 Reference: k = 142, avg PC = -.090 (s.d. = .252), avg t = -0.85 (s.d. = 2.08); control for cycle: k = 124, avg PC = -.185 (s.d. = .213), avg t = -2.05 (s.d. = 2.58). MD test for PC (reference – mediator) = 3.37***, MD test for t-values = 4.16***.

44 Reference: k = 256, avg PC = -.144 (s.d. = .232), avg t = -1.51 (s.d. = 2.34); control for decay: k = 10, avg PC =.111 (s.d. = .283), avg t = 1.14 (s.d. = 2.67). MD test for PC (reference – mediator) = -2.80***, MD test for t-values = -3.10***.

Nadeau et al., 2000; Price & Sanders, 1993, 1994); miners' strikes (Anderson, 1995; Clarke et al., 1986; Lanoue & Headrick, 1994); the 1983 Labour Manifesto (Sanders, 1996); the Falklands War (Anderson, 1995; Clarke et al., 1986, 1990; DeHaven 1991; Dewan & Dowding, 2005; Headrick & Lanoue, 1991; Lanoue & Headrick, 1994; Nadeau et al., 2000; Norpoth, 1987, 1992; Price & Sanders, 1993, 1994; Sanders, 1996, 2000); the bombing of the Conservatives' conference hotel in Brighton in 1984 (Nadeau et al., 1996, 2000; Price & Sanders, 1993, 1994); the Westlands Affair in 1986 (Lanoue & Headrick, 1994); the 'poll tax' controversy in 1990 (none of the included studies); the Gulf War (Dewan & Dowding, 2005; Nadeau et al., 2000), which coincided with the removal of Mrs Thatcher in 1990 (Nadeau et al., 2000; Sanders, 1996, 2000); the exchange-rate mechanism crisis in 1992 (Sanders, 1996 – not in model with inflation); Labour's 'honeymoon' after coming to power in 1997 (Sanders, 2005a); the Kosovo War in 1999 (Gavin & Sanders, 2003); the 'fuel crisis' in 2000 (Sanders, 2005a; Sanders & Gavin, 2004); 9/11 (Sanders, 2005a); and the Iraq war since 2003 (Sanders, 2005a).

Any model that has some control for political variables is coded with a mediator for this. The mean statistics show significant differences, with no effect of inflation in models that control for political events.[45] However, the average publication year for estimates without control for political events is 1980 (average data year = 1966), while the average publication year for estimates with control for political events is 1995 (average data year = 1980).

8.4.14. MA

The final mediator is a control for moving averages. Only two studies have been coded with MA( ) models (Whiteley, 1984, 1986).[46]

8.5. Analysis, results, discussion, and conclusion

The previous section is itself a review of the literature, although it is overly detailed and it covers all of the studies that could be included in the meta-analysis and none of the studies that have been excluded for various reasons (such as cross-sectional studies rather than time-series studies, disaggregated studies, studies of party support also when not in government, and studies that do not provide the necessary statistics). Despite a rather narrow focus,

45 Reference: k = 179, avg PC = -.166 (s.d. = .250), avg t = -1.76 (s.d. = 2.64); control for political events: k = 87, avg PC = -.069 (s.d. = .198), avg t = -0.68 (s.d. = 1.58). MD test for PC (reference – mediator) = -3.44***, MD test for t-values = -4.15***.

46 Reference: k = 257, avg PC = -.134 (s.d. = .242), avg t = -1.39 (s.d. = 2.43); control for MA: k = 9, avg PC = -.148 (s.d. = .079), avg t = -2.09 (s.d. = 0.85). MD test for PC (reference – mediator) = 0.47, MD test for t-values = 2.18**.

this review documents the heterogeneity of the literature, but it is difficult to get an impression of what the genuine effect of inflation is on government popularity. This is where meta-regression has its place.

I run several MRA models, most of which are for robustness checks. Given the two-level nature of the data (estimates within and between studies), I focus on the mixed-effects models. There is one variable I decide to drop due to improbable values, and there are some observations I decide to drop due to outlying values. The variable that I drop is the moderator of estimates that are reported without *t*-values, as this produces a coefficient which I find to return values that are far beyond reason. Dropping this does indeed seem reasonable, because the outlying observations appear to cause some trouble after this drop – indicating collinearity (which is difficult to predetermine with dummy variables). Thus, I also drop the outlying observations of the dependent variable. As the rest of the variables seem to behave, I am confident about the conclusions drawn using my final model.

The results are quite nice. First, there is no publication selection that biases the effects of inflation on British government popularity. Second, the ceteris paribus effect is somewhat smaller (and with the opposite sign) than that of personal economic expectations (Ludvigsen, Paper 3) with a partial correlation coefficient of -0.225. Third, there is little variation that is due to observation period. Most periods are quite stable; however, there seems to be instability due to specification differences. Fourth, both parties suffer from increasing inflation, but Labour has been punished about 50 pct. harder than the Conservative Party. Thus, whereas Ludvigsen (Paper 3) concludes on the Responsibility Hypothesis in regards of personal economic expectations, this paper concludes on the late version (Rattinger, 1991) of the Clientele Hypothesis with regards to inflation.

9. Unemployment and government popularity in the United Kingdom – a meta-analysis

The final paper (Ludvigsen, Paper 5) in essence the same as Paper 4, as it is basically the same analysis but with a different dependent variable. Obviously, there is also some difference in which papers are included in the analysis for the simple reason that some studies have not tested the effect of inflation, while others have not tested the effect of unemployment on British government popularity. The regression analyses are also slightly different, as I have only reported mixed-effects models, I have included bivariate regressions (as in Ludvigsen, Paper 3), some of the variables differ, and I also report weighted multivariate regressions (unlike the two others). However, the introduction and section describing the dependent variable and the test for publication bias are nearly identical to Ludvigsen (Paper 4).

I make no apologies for this: the nature of the data is quite similar, but the analyses must be treated in different papers due to the sheer data load. I will argue that it is both important and interesting to have comparable analyses of inflation and unemployment. The two are seen as 'the big two' within economic voting, and coding the analysis of the one without coding a similar analysis of the other would be a waste. The consequence is the 'same' analysis on two different dependent variables.

This chapter will be short, since there is 'nothing new' leading up to the analysis. I will therefore go straight to the summary of the analysis, results, discussion, and conclusion of Paper 5.

The regression analysis is where the paper differs the most from the two others. The list of moderating and mediating variables is largely the same as with the paper on inflation, but there are naturally differences in how unemployment is specified compared to inflation. Naturally, this is reflected in the analysis. Furthermore, there is a slight difference in the mediating variables, as the exchange rate and interest rate have been separated from the group of 'other' economic variables (there was enough information to do so). I have also chosen to only estimate mixed-effects models. Robustness checks were the motivation for the range of estimation procedures in the two previous papers, but there are a number of data issues in this paper requiring a slightly different approach. I have therefore focused on the estimation procedure which I think is the most appropriate given the data structure, and produce a variety of models with this procedure instead of reporting a variety of procedures for one model.

First, this gives me the liberty to estimate both unweighted and weighted regression models. In the weighted models, both sides of the equation are

divided by the standard error of the partial correlation coefficients. Thus, the models are weighted against a measure of heteroskedasticity. In this case, the weighing does not produce very different results, and although they may seem slightly more difficult to interpret, they are considered more efficient, since more accurate estimates are given more weight and are therefore preferable.

Weighing is discussed in the two other empirical papers but only conducted in bivariate regressions in Paper 3. The reason for not estimating weighted multivariate models in the two other empirical papers is due to the somewhat more complex models. The interpretation may seem more difficult, since the dependent variable in the weighted models is the *t*-value instead of the partial correlation coefficient, while the coefficients retain their property. In other words, they still represent the effect of their variables on the partial correlation coefficient. In order to calculate the effects on the *t*-value, the coefficients must be divided by the standard error of the partial correlation coefficient. In order to 'stay on message', I therefore estimated only the unweighted models in the first two papers.

I found the weighing to be necessary in the third empirical paper, however, as I obtained some strange results on the constant and the control for publication bias. First, the models show a significant publication bias running in the *opposite* direction of the average effect of unemployment on government popularity. In other words, the models show that as the precision of the studies drops, they produce *less* – not more – significant results. This means that the results are biased in the other direction than what would be expected from publication selection. Second, the constant is more than 2.5 times greater in the multivariate model than in the bivariate model, which would typically mean that the constant is massively affected by the moderators and mediators or that it could be a result of an even greater effect of publication bias in the multivariate models as compared to the bivariate models.

I find evidence of the latter when using weighted multivariate models, as the weighing has little impact upon the moderators and mediators but considerable impact on the constant and the publication bias. This picture, however, is confused by running a weighted *bivariate* model. In the weighted bivariate model, both the constant (here, the average impact of unemployment on government popularity, controlled for publication bias and two-level random effects) and the coefficient for publication bias drop in magnitude, and the latter becomes insignificant. In the weighted multivariate model, however, both the constant (here, the ceteris paribus impact of unemployment) and the coefficient for publication bias increase in magnitude (al-

though here again, the latter has a standard error which increases even more).

Ultimately, the weighing did not help much. It did not make things much worse either, however, so I chose to keep the weighted models but drop the control for publication bias. This, I argue, can be done for several reasons: First, there is no reason to expect the literature to be driven by a bias towards *positive* effects of unemployment on government popularity; second, there is no reason to expect that the literature is driven by a bias towards *no* effects of unemployment; third, there seems to be a correlation between observation years and publication bias. This, I argue, is expected. Variables controlling for observation years "may pick up aspects relating to the publication process and thus work together with the variance of the partial correlation coefficients in controlling for publication bias" (Ludvigsen, Paper 5: 23). Each variable by itself is not highly correlated with the variance of the partial correlation coefficient, but the group of variables seems to have an impact when removed from the model altogether. I do not find room to show this in the paper, but there is room to show it here. Table 9.1 is therefore a reproduction of the general Models 4 (full specification of weighted model) and 5 (similar to Model 4, but without control for publication bias) from Ludvigsen (Paper 5), with an additional model in which the observation moderators have been removed from Model 4 – here shown as Model 6 (see Paper 5 for table and variable description). Comparing Models 5 and 6, one sees that the constant is untouched by this exercise, whereas comparing Models 4 and 6 reveals that the variance of the partial correlation coefficient drops to insignificance.

I am satisfied that Model 5 gives a fair representation of the literature, and as I argue in the paper, I will rather drop the control for publication bias and keep the control for observation years than drop the latter (and consequently also drop the former in the stepwise reduction). My conclusions are therefore based upon the reduced Model 5.

It is unnecessary to consider all of the coefficients in the multivariate model to know that the impact of unemployment on British government popularity is unstable. The constant alone tells us this: there is a massive difference between the constants of the bivariate model(s) and the multivariate model(s). This indicates that specification choices seriously impact the results. However, the ceteris paribus specification (where all moderators and mediators are set to 0) is reasonably comparable to the ceteris paribus specifications of the two other meta-analyses, and the constant in this analysis shows that – with this specification – the impact of unemployment on government popularity in the UK is much greater than the impact of personal economic expectations or inflation.

	Table 9.1: Dependent variable: Partial correlation coefficients (*r*)	Model 4 Weighted general REML	Model 5 Weighted general REML	Model 6 Weighted general REML
Bias	Variance of *r*	**7.252 (3.60)**	Dropped	2.806 (1.34)
Specification of unemployment	Raw unemployment rates	0.123 (1.64)	0.126 (1.62)	**0.228 (2.90)**
	Lagged unemployment	-0.037 (1.18)	-0.032 (1.00)	-0.019 (0.52)
	First-order changes	0.034 (0.89)	0.026 (0.66)	0.045 (0.98)
	Nonlinear measure	0.060 (0.93)	0.070 (1.04)	0.125 (1.79)
	Seasonal	0.039 (0.53)	0.070 (0.92)	-0.011 (0.16)
	Perceived	-0.047 (0.49)	-0.040 (0.41)	-0.104 (1.01)
	Expected	0.005 (0.03)	0.067 (0.36)	0.088 (0.41)
	Principal	-0.015 (0.32)	-0.024 (0.49)	-0.011 (0.19)
	Secondary	0.023 (0.46)	0.011 (0.22)	0.016 (0.28)
Variation in publications and observations	Published in 1980s	**0.295 (3.69)**	**0.367 (4.56)**	**0.197 (2.45)**
	Published in 1990s	0.138 (1.37)	**0.224 (2.20)**	0.013 (0.13)
	Published in 2000s	**0.459 (3.24)**	**0.518 (3.53)**	**0.485 (3.72)**
	Data from 1947-51 (Labour)	0.057 (1.42)	0.062 (1.48)	Dropped
	Data from 1952-56 (Tory)	**0.157 (5.06)**	**0.139 (4.43)**	Dropped
	Data from 1957-61 (Tory)	0.041 (1.19)	0.039 (1.10)	Dropped
	Data from 1962-66 (Tory + Labour)	-0.005 (0.14)	-0.023 (0.61)	Dropped
	Data from 1967-71 (Labour + Tory)	0.021 (0.63)	0.006 (0.17)	Dropped
	Data from 1972-76 (Tory + Labour)	**0.118 (3.39)**	**0.129 (3.62)**	Dropped
	Data from 1977-81 (Labour + Tory)	-0.022 (0.70)	-0.028 (0.84)	Dropped
	Data from 1982-86 (Tory)	-0.105 (1.81)	-0.126 (2.11)	Dropped
	Data from 1987-91 (Tory)	**0.123 (2.14)**	0.108 (1.83)	Dropped
	Data from 1992-96 (Tory)	0.096 (1.50)	0.056 (0.86)	Dropped
	Data from 1997-01 (Labour)	-0.172 (0.86)	-0.116 (0.56)	Dropped
	Not OLS	-0.007 (0.14)	-0.004 (0.07)	-0.024 (0.38)
	No t-values	-0.133 (1.49)	-0.148 (1.60)	-0.150 (1.76)
	Book	-0.053 (0.85)	-0.110 (1.76)	-0.005 (0.08)
Specification of the primary dependent variable	Conservatives	0.085 (1.85)	**0.108 (2.31)**	0.087 (1.76)
	Labour	0.064 (1.03)	0.099 (1.57)	-0.024 (0.38)
	Prediction	-0.087 (0.85)	-0.091 (0.87)	0.079 (0.72)
	Vote-function	-0.229 (1.17)	0.033 (0.17)	0.024 (0.12)
	Government lead	-0.072 (1.09)	-0.101 (1.48)	-0.019 (0.26)
	PM approval	-0.020 (0.30)	-0.037 (0.54)	0.033 (0.43)
	Economic competence	-0.196 (1.23)	-0.165 (1.00)	-0.241 (1.44)
	Changes	**0.202 (2.03)**	**0.240 (2.35)**	-0.097 (0.98)
	Nonlinear	-0.176 (1.87)	-0.153 (1.56)	0.191 (1.71)
	Poll-of-polls	0.188 (1.62)	0.187 (1.54)	0.193 (1.57)
	NOP	-0.089 (1.93)	**-0.098 (2.06)**	**-0.126 (2.44)**
Mediators	Inflation	**0.181 (3.32)**	**0.186 (3.34)**	**0.254 (4.18)**
	Personal expectations	0.025 (0.29)	0.011 (0.13)	0.045 (0.48)
	GDP or GNP	-0.052 (0.44)	-0.007 (0.06)	-0.079 (0.79)
	Real wages or disp. income	0.003 (0.07)	0.007 (0.17)	0.058 (1.18)
	Exchange rate	-0.122 (1.47)	-0.116 (1.36)	-0.017 (0.17)
	Interest rate	0.028 (0.34)	0.013 (0.15)	-0.022 (0.23)
	Other economic variables	0.038 (0.51)	0.042 (0.55)	-0.051 (0.60)
	Missing control for party	0.036 (0.92)	0.050 (1.25)	0.069 (1.59)
	Government or PM approval	0.038 (0.88)	0.038 (0.87)	0.073 (1.45)
	Economic competence	-0.074 (0.39)	0.009 (0.04)	-0.096 (0.47)
	Lagged dependent	0.054 (1.71)	0.063 (1.94)	0.066 (1.79)
	Trend	**0.165 (4.53)**	**0.166 (4.42)**	**0.216 (5.20)**
	Cycle	0.026 (0.90)	0.025 (0.84)	0.020 (0.60)
	Decay	0.021 (0.21)	-0.020 (0.19)	-0.115 (1.08)
	Political variables	0.033 (0.69)	0.025 (0.49)	0.045 (0.82)
	Moving averages	**-0.482 (3.09)**	**-0.595 (3.75)**	**-0.341 (2.07)**
	Constant	**-0.928 (8.51)**	**-0.892 (7.96)**	**-0.890 (7.62)**
Test statistics	Groups	40	40	40
	Estimates	274	274	274
	Probability distribution	$\chi^2_{(55)} = 661.8$	$\chi^2_{(54)} = 604.5$	$\chi^2_{(44)} = 430.1$
	p =	0.000	0.000	0.000
	Log likelihood	-557.0	-565.0	-583.9
	Between-study variance	0.952	1.071	0.872
	Residual variance	1.510	1.585	2.316
	Intraclass correlation	0.387	0.403	0.273
	Akaike Information Criterion	1228.1	1242.0	1259.8
	Bayesian Information Criterion	1434.0	1444.3	1426.0
	LR test vs. linear regression	**9.37**	**10.55**	**5.09**
	p =	0.001	0.001	0.012

The multivariate analysis shows that there are some specification choices that make no difference (the specification of unemployment itself), but that the specification of the dependent variable and the choice of control variables has a very serious impact. The choice of time period also has some impact, but this is more modest. Interestingly, there is only a very modest difference between the Conservative Party and Labour. Whereas I conclude in Ludvigsen (Paper 4) that the Clientele Hypothesis appears to be supported in the case of inflation, the case of unemployment seems to support the Responsibility Hypothesis. The case of personal economic expectations did the same, which brings me to the chapter comparing the results.

10. Comparison of the meta-analyses

This chapter will focus on four aspects: 1) publication bias, 2) the effects on government popularity – all things equal, 3) the four main theories, and 4) other findings of (comparative) interest.

10.1. Publication bias

Publication selection bias is expected in most fields of study (Hunter & Schmidt, 2004). Without controlling for publication bias, the narrative reviewer would risk making his or her conclusions based upon a biased – and uncorrected – literature. Fortunately, statistical tools provide the ability to deal with this, and the meta-analyses can thus be conducted with less fear of making biased inferences.

Publication bias is also a concern in the literature on British economic voting. I find a publication bias in Paper 3, none in Paper 4, while I have to haggle with the data in Paper 5 before they confess to having none. The results in Papers 3 and 4 seem to be straightforward, while the results in Paper 5 appear after I am able to make the case that the apparently opposite bias is due to some data artefact.

I argue that the publication bias found in Paper 3 is possibly due to the most prolific author's penchant for only reporting the results from his reduced models. There are only three groups of authors behind the studies included in the meta-analysis of Paper 3, and most of the estimates have been produced by one author, David Sanders, either alone or in co-authorship. Since Sanders often only reports significant findings, these findings will necessarily bias the results. Non-significant findings may be due to specification choices that are never reported, and thus the average result from the studies will be biased in the direction of the specifications that produce significant results. Indeed, I find that this selection bias increases the results by about 35 pct. on average, which is quite drastic.

However, I do not argue that Sanders' motivation for this selection is sinister, merely that his selective reporting has a clear consequence for those of us who wish to evaluate the studies and their findings. Publication bias aside, I find a clear effect of personal economic expectations on government popularity in the UK, so my meta-analysis does indeed support Sanders' findings, although with some moderation.

There is no publication bias as regards inflation and unemployment. The reason why there is no drive to select significant results from these studies may be that the authors have had a genuine curiosity about the effects of these two variables. With more authors searching for these effects, the reporting habits of one author are not as likely to bias the results. Nonetheless, as I

argue in Paper 5 (p. 25), there are reasons to expect a unidirectional bias in the literature:

> … the theoretical base for the VP-function is the Responsibility Hypothesis (H_1), which predicts a negative correlation between increased unemployment and government popularity. The predictions of the Clientele Hypothesis (H_2) and the Salient Goal Hypothesis (H_3) go in different directions for different parties, but they include the possibility of asymmetric punishment, and thus do not necessarily predict genuine rewards. The only theory that predicts an increase in government popularity – for both parties – as a result of increasing unemployment, is the Stability Hypothesis (H_4). However, the only study of British data that mentions the Stability Hypothesis is Anderson (1995). I will therefore argue that there are hardly any theoretical expectations of a positive effect of unemployment on British government popularity.

The same holds for inflation, which implies that one could reasonably expect a negative effect of inflation and unemployment instead, and therefore also expect a publication bias towards such effects. There is none.

10.2. The effects on government popularity – all things equal

As already stated, the constant in the regressions is of much importance, as it represents the default specification of the relationship of interest. I have tried to specify my models so that the constants represent as generic models of government popularity as possible. I call this the *ceteris paribus specification*, meaning that the constant represents the relationship of interest without publication bias, without any control variables, without any lags, first-order changes, nonlinear relationships or other idiosyncrasies. The constant also represents estimates of the relationship of interest for both parties (when in government) and estimates from popularity functions.

Comparing the constants of the preferred models from each of my meta-analyses, the ceteris paribus partial correlation coefficient of the effect of personal economic expectations on British government popularity is + 0.320, of inflation = -0.225, and of unemployment = -0.775. However, the effect of personal economic expectations cannot be directly compared to the two others, because the latter two have been estimated with controls for publication decades in a manner so that estimates from the 1970s are in the reference category. Studies of PE were not conducted in the 1970s. In fact, the average publication year of the studies that included PE was 1995. The comparable effects of unemployment for studies published in the 1990s is estimated to be -0.557, while the effect of unemployment for studies published in the 2000s is estimated to be -0.268. The comparable effects for inflation are -0.225 for the 1990s and + 0.179 for the 2000s.

Regardless of the decade of publication, these meta-analyses leave the impression that British governments have had the most to fear from rising unemployment and the least to fear from rising inflation. This contradicts the argument made by Whiteley (1984: 20) that "It appears to make sense for governments to concentrate on inflation, if necessary at the expense of unemployment, since the political cost of doing so appear to be less than the reverse policy". Whiteley's result may be due to his choice of specification, and specification differences have indeed given VP-functions their reputation of being unstable.

This also becomes evident in this discussion: the results are somewhat unstable across specifications as well as across publication decades. Nonetheless, what I find to produce the least instability is the choice of observation period (this is not controlled for in the paper on PE). There are some years of economic turmoil in British history, which are mirrored in the meta-regressions that have enough degrees of freedom to control for this. However, there are no more than four or five five-year periods producing significantly different effects (although longer time periods are largely covered by the moderators for decade of publication).

The really drastic instability comes, I argue, from some of the modelling choices made, but I will not repeat all that here.

10.3. Theory-testing

Once again, the four theories tested by my meta-analyses are:

H_1: the Responsibility Hypothesis suggests that both parties should be hurt by rising inflation, rising unemployment, and lowered economic expectations.

H_2: the Clientele Hypothesis suggests that

a) the Conservatives should be rewarded for inflation (or at least punished less than Labour).

b) Labour should be rewarded for unemployment (or at least punished less than the Conservatives).

H_3: the Salient Goal Hypothesis suggests that

a) the Conservatives should be punished harder for inflation than Labour.

b) Labour should be punished harder for unemployment than the Conservatives.

H_4: the Stability Hypothesis suggests that voters are attracted to 'responsible' parties, in which case both parties should be rewarded for increasing inflation and unemployment, especially during crisis periods. However, Sanders (2000) argued that the

Labour Party was no credible opposition to the Conservatives between 1982 and 1994. For this reason, there should be no effect of changes in the macroeconomic indicators whatsoever for the Conservatives during that period.

There are no theoretical expectations in the lines of H_2 or H_3 on personal economic expectations. Nonetheless, I thought it important to control for the possibility, but I am relieved that I do not have to grapple with unequal effects between Labour and the Conservatives on this issue.

The multivariate regressions find:

1) No difference in the effect of personal economic expectations between the two parties, both have been rewarded for increasing expectations. This supports $H_{1.}$
2) Both parties have been punished for inflation, but the effect of inflation is about 50 pct. larger in models of Labour support than in models of the Conservatives or of both parties. This supports the moderated version of H_2a.
3) Both parties have been punished for unemployment, but there is a marginally weaker punishment in models of Conservative support than in models of Labour or of both parties. The difference is weakly significant, but only amounts to a 5-15 pct. difference on the constant. I do not think this constitutes enough of a difference to confirm the Salient Goal Hypothesis in this case. Both parties are punished for unemployment, and H_1 is supported.

There is nothing in my meta-regressions suggesting that British parties have been rewarded for negative economic developments. Therefore, H_4 and the un-moderated version of H_2 can be discarded. There is scant evidence of H_3 in the case of unemployment, but I do not find this evidence to be strong enough to actually support H_3. I only find support for an alternative theory to the Responsibility Hypothesis in the case of inflation. However, the alternative theory is merely a moderated version of the Clientele Hypothesis, which accepts that both parties are punished for inflation, but with the expectation that a rightist party is punished less than a leftist party. Seeing this effect in relation to the two other economic indicators, Rattinger's 1991-version of H_2 fits the British case more as a moderated Responsibility Hypothesis than as a moderated Clientele Hypothesis. With this slight moderation included, the Responsibility Hypothesis has been found to hold in the UK.

10.4. Other findings of (comparative) interest

I will not compare all of the results here, since doing so would require excessive repetition of variable discussions, etc. Instead, I will focus on a few of the interesting similarities and differences. Three groups of variables are discussed here: moderators of the specification of the dependent variable in the primary studies, moderators of general model specification, and mediators (control variables in the primary studies).

10.4.1. Moderators of the primary dependent variable

Vote function vs. popularity function: I have discussed the difference between vote functions and popularity functions in Section 8.3.3 (see also Paldam, 1991). The difference is important. First, voters may evaluate governments differently when asked in an opinion poll as compared to Election Day. In the latter, they may prepare themselves for the vote, collecting more information and thus making more informed choices than they do when rung up any given weeknight to answer an opinion poll in the middle of cleaning up after supper, watching TV or putting the children to bed. Moreover, surveys collect the opinion from samples of the population, while elections collect the opinion of the entire population. These are arguments *against* popularity functions, but there are also arguments in favour of them: opinion polls are collected much more frequently than election results. Thus, one may estimate popularity functions with shorter time spans *and* more degrees of freedom.

There are not many studies that are vote functions in the literature on economic voting in Britain. Of the total of 639 estimates collected for my meta-analyses, only 15 come from vote functions: six of unemployment and nine of inflation. The effect, however, is present. The impact of inflation on the vote for the incumbent is more than twice as strong as on the popularity of the incumbent. With outliers, the effect is even stronger. The impact of unemployment on the vote is not robustly different from the impact on the popularity. Initially, the difference is about as strong in absolute terms as for inflation, but this difference vanishes once the control for the precision of the studies is dropped. However, one should not forget that the effect of unemployment is very strong regardless, and the ceteris paribus effect of inflation on the vote still does not compare with the ceteris paribus effect of unemployment on the popularity.

Nevertheless, the results indicate that voters do indeed evaluate governments differently on Election Day than when answering a survey about their preferences for political parties, soft drinks, and glossy magazines. Importantly, the impression from the popularity functions *is not weakened* when

controlling for vote functions. The punishments are indeed accentuated on Election Day as compared to the day-to-day surveys.

Prediction models: Once you have a VP-function, it is tempting to stretch it into the unobserved and attempt to predict future election outcomes: *"Back to the Future"*, as a colleague of mine called it in his blog (Arnesen, 2009), looking back at observed effects and assuming they will reappear in the next election. I thought this might have a moderating impact, as the motivation of the authors may be different when they try to make a model that fits into the future compared to a model explaining the past; however, there is no difference. The coefficient for prediction models appears as significant a few times with the inflation data, but it is not present once I have dropped the outlying observations. There is no effect present in the other two datasets.

10.4.2. Moderators of general model specification

OLS: I also thought it might make a difference whether the models are estimated through *Ordinary Least Squares* or other estimation techniques. It might, but I have not been able to detect such a difference. The moderator I have used is a rough one, where anything that was *not* OLS was coded with this moderator. There is no effect of this on unemployment or personal expectations, but some indication of this on inflation (though not in the final model). The lack of difference between OLS and other estimation techniques (after controlling for other specification choices) is comparable to other studies I am aware have tested for this (Doucouliagos & Paldam, 2008; Doucouliagos & Ulubaşoğlu, 2008). However, Abreu, de Groot & Florax (2005) on economic growth literature had the freedom to test for different non-OLS procedures and found some differences. All of the significant differences, however, were with the same sign and could possibly be covered by a common moderator.

10.4.3. Mediators

Ideally, none of the control variables used in the primary studies should have a mediating impact upon the results. However, this is often not the case – and is not even expected. Some mediators were found to have a significant impact upon the results of the literature, while others – more or less unexpectedly – did not have a significant impact across the literature.

One type of control variable that has *not* had a mediating impact upon the relationships of interest is the group of *political variables*, thoroughly described in the case of inflation in Section 8.4.13. The aggregation might produce an excessively crude estimate, but it is nevertheless interesting to note that controls for political events have not had an impact at the meta-aggregate level on the relationship between the economy and the vote.

The mediating impact of the other economic variables is also interesting. I do not find that the inclusion of unemployment or inflation has had a mediating impact on the partial correlation between personal economic expectations and government popularity. Nor have I found any mediating impact of personal expectations on unemployment or inflation. From this, I then assume that the three may safely be included in the same model. However, the relationship between unemployment and inflation is not as clear-cut. The inclusion of unemployment does not have a mediating impact upon inflation, but inflation has a mediating impact upon the partial correlation between unemployment and government popularity. This is disconcerting.

As I have argued in Section 8.4.1, if the Phillips curve exists and has an impact upon political support in Britain, then the mediator for unemployment (inflation) will be significant and with an *opposite* sign of the constant in the meta-regression of inflation (unemployment). One would expect the relationship to be either existent or non-existent in *both* meta-analyses. As I argue in the paper dealing with inflation (Paper 4: 18), the lack of impact "could put a rest to the Phillips curve in relation to British government popularity", but I am not so certain in the unemployment paper (Paper 5: 25), where I "caution against the inclusion of both unemployment and inflation in the same model without considering their mediating effect on each other".

Other economic variables that I have found to have a mediating impact are *economic growth* (GDP/GNP), which increases the effect of inflation on government popularity, and the dollar to sterling *exchange rate*, which increases the effect of unemployment. That there is multicollinearity between economic growth and inflation and between exchange rates and jobs should hardly come as a surprise.

Other mediators and moderators are also found. See the empirical papers for results and discussions of these.

11. Conclusion

The general topic of this anthology has been studies known as vote and popularity function studies, which estimate the effects of economic outcomes on voting or party popularity, and the motivation for this anthology has been the various concerns that have been voiced over the unstable nature of these studies of economic voting.

This dissertation has shown that there is an extreme theoretical complexity concerning economic voting, with a corresponding empirical heterogeneity. I have built the argument that a) this complexity needs to be taken account of in future models of economic voting, and b) the complexity needs to be reduced in literature reviews in order to understand what we *have* proven. This sends mixed signals, but the one does not preclude the other now that we have statistical review techniques at our disposal.

My entire PhD dissertation consists of a set of literature reviews: Two papers summarize and build upon the theoretical foundations of the VP-function, three papers apply regressions of regressions in order to quantitatively synthesize the literature, and this summary essentially represents a review of my reviews. The three quantitative literature reviews explain variation in the results from the literature that is due to specification differences and estimate and control for the impact of publication selection bias. Both of these aspects have been seen as problematic within the literature.

In addition to summarizing my papers, this summary has provided more detailed information about the motivation and methodology of meta-regression analysis, discussing meta-analysis as a review method, discussing problems with publication selection bias, and illustrated with a simple – hypothetical – example of how to test and control for publication bias.

The main contribution made by the dissertation is the application of meta-regression analysis to political science, but the dissertation also makes a number of theoretical contributions. After laying out the conditional evaluations of economic voting and arguing that the inherent complexity cannot be handled within single empirical studies, I demonstrate how meta-regression offers a possible means of navigating through this complexity, but I also add some more theory to this complexity in my paper on party positions, voter preferences, and the cost of ruling. This is where I have my chief theoretical contribution to the literature, where I further develop several of the possible explanations for the cost of ruling, and argue that underlying between-group gratification asymmetries and voter distributions possibly add to the possible explanations for the grievance asymmetry.

The meta-analyses are of a literature in which there has been some debate over the empirical results, and a central tenet of this anthology has been

that rejecting an entire literature or delegitimizing it as unstable due to some contradicting findings is ill-advised. Differences may be due to variation in research questions, measurement, data, research practices, publication selection, etc., that must be controlled for before making such dramatic calls. Controlling for all of this, it becomes possible to actually see the cumulated knowledge and draw new conclusions from old studies in ways that narrative reviewers cannot.

The issue of publication selection bias has been another important topic. I see controls for publication selection bias as being of utmost importance for any review of empirical results – including narrative results – and I argue that it is impermissible to avoid this issue when synthesizing published results. Nevertheless, this has been missing from all but one of the few meta-analyses that have been produced from political science literatures.

There is some publication bias in the literature on personal economic expectations, and controlling for this has been necessary in order to arrive at a more precise average estimated effect on British government popularity. In terms of inflation and unemployment, I find no publication bias. Nonetheless, the control for it has been important and useful.

Regarding specification differences, I find that there are a number of specification choices that make no difference, some make a little difference, and a few are quite consequential. All things considered, I argue that the instability of estimated effects is caused more by variation in general modelling choices than in the choice of which years to cover.

Regarding theory, I find that both parties have been equally rewarded for increasing personal economic expectations, and (almost equally) punished for increasing unemployment. Thus, both of these findings support the Responsibility Hypothesis (H_1), while the parties have been asymmetrically punished for increasing inflation. All things equal, Labour has been punished about 50 pct. harder for inflation than the Conservatives. Nonetheless, the Conservatives have not escaped punishment, and although this finding unto itself supports the moderated version of the Clientele Hypothesis (H_2), the overall findings suggest that the Responsibility Hypothesis generally holds in the British case.

It turns out – after all – that my dissertation has been less a post-mortem and more a pathological study. The field of study has been dormant due to some fear of diseases, but it should not be put to rest. Rather than a post-mortem, this project may be the revival VP-functions need: the Responsibility Hypothesis has been found to hold in the British case, and the effects of economic outcomes and expectations on British government popularity are quite substantial. The problem lies with the impact on the results from variation in

specifications. The meta-regression analyses applied here estimate and adjust for publication bias, specification differences and errors, and reconcile inconclusive results. The expected result from a post-mortem is that the doctor declares the body not only to be deceased, but also to have died from certain causes. Instead, I have proved that the body of studies is not deceased, but somewhat diseased, and I have cast light on the symptoms of these diseases. Knowing the symptoms, we may take these into consideration in the future and learn to live with our disorders.

Further implications of this dissertation may not only be a revival of the VP-function, but also a realization for other social scientists that they should not be afraid to work in fields of theoretical and empirical complexity. Although their results may be unstable when isolated or reviewed through traditional means, the results of their efforts can be measured through quantitative approaches. This is likely to provide many a cluttered field of research with a way of cumulating their knowledge.

References

Abreu, M., H.L.F. de Groot & R.J.G.M. Florax (2005): "A meta-analysis of β-convergence: The legendary 2%", *Journal of Economic Surveys* 19(3): 389-420.

Alesina A. & H. Rosenthal (1995): *Partisan Politics, Divided Government and the Economy,* Cambridge: Cambridge University Press.

Alt, J.E. (1979): *The Politics of Economic Decline: Economic Management and Political Behaviour in Britain since 1964*, Cambridge: Cambridge University Press.

Andersen, S.C. (2007): "Multilevel-modeller: en introduktion og et eksempel", *Politica* 39(3): 294-316.

Anderson, C.J. (1995): *Blaming the Government: Citizens and the Economy in Five European Democracies*, Armonk, NY: M.E. Sharpe.

Anderson, C.J. & D.S. Ward (1996): "Barometer elections and comparative perspective", *Electoral Studies* 15(4): 447-460.

Arnesen, S. (2009): "Tilbake til fremtiden", http://voxpublica.no/2009/06/tilbake-til-framtiden, accessed 04.10.09.

Bakan, D. (1966): "The test of significance in psychological research", *Psychological Bulletin* 66: 432-437.

Bateman, I.J. & A.P. Jones (2003): "Contrasting conventional with multi-level modeling approaches to meta-analysis: Expectation consistency in U.K. woodland recreation values", *Land Economics* 79(2):235-258.

Baumgartner, F.R. (1998): *Basic Interests: The Importance of Groups in Politics & in Political Science*, Ewing, NJ: Princeton University Press.

Begg, C.B. (1994): "Publication bias", in H. Cooper & L. Hedges (eds.), *The Handbook of Research Synthesis*, New York: Russell Sage Foundation.

Bélanger, E., M.S. Lewis-Beck & R. Nadeau (2005): "A political economy forecast for the 2005 British General Election", *British Journal of Politics and International Relations* 7(2): 191-198.

Berry, W.D. (1993): *Understanding Regression Assumptions* (Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-092), Newbury Park: Sage.

Berry, W.D. & S. Feldman (1985): *Multiple Regression in Practice* (Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-050), Beverly Hills: Sage.

Borooah, V.K. & F. van der Ploeg (1983): *Political Aspects of the Economy*, Cambridge: Cambridge University Press.

Brouwer, R., I.H. Langford, I.J. Bateman & R.K. Turner (1999): "A Meta-Analysis of Wetland Contingent Valuation Studies", *Regional Environmental Change* 1(1): 47-57.

Butler, D. and D. Stokes (1974): *Political change in Britain – The evolution of electoral choice*, 2nd ed., London: Macmillan.

Card, D. & A.B. Krueger (1995): "Time-series minimum-wage studies: a meta-analysis", *American Economic Review* 85: 238-243.

Chappell, H.W. & L.G. Veiga (2000): "Economics and Elections in Western Europe: 1960-1997", *Electoral Studies* 19: 183-197.

Clarke, H.D. & M. Lebo (2003): "Fractional (co)integration and governing party support in Britain", *British Journal of Political Science* 33(2): 283-301.

Clarke, H.D., M.C. Stewart & G. Zuk (1986): "Politics, economics and party popularity in Britain, 1979-1983", *Electoral Studies* 5(2): 123-141.

Clarke, H.D., W. Mishler & P. Whiteley (1990): "Recapturing the Falklands: Models of Conservative popularity, 1979-83", *British Journal of Political Science* 20(1): 63-81.

Clarke, H.D. & M.C. Stewart (1995): "Economic evaluations, prime ministerial approval and governing party support: Rival models reconsidered", *British Journal of Political Science* 25(2): 145-170.

Clarke, H.D., M.C. Stewart & P. Whiteley (1997): "Tory trends: Party identification and the dynamics of Conservative support since 1992", *British Journal of Political Science* 27(2): 299-319.

Clarke, H.D., E. Elliott, W. Mishler, M.C. Stewart, P. Whiteley & G. Zuk (1992): *Controversies in Political Economy: Canada, Great Britain, the United States,* Boulder, CO: Westview Press.

Coleman, S. (2007): "Testing theories with qualitative and quantitative predictions", *European Political Science* 6: 124-133.

Cooper, H., L.V. Hedges & J.C. Valentine (eds.) (2009): *The Handbook of Research Synthesis and Meta-Analysis*, 2nd ed., New York: Russell Sage Foundation.

Dagbladet Information (2008): "Sundhedsforskere fordrejer resultater", p. 12-13, 08.10.08.

DeHaven, M.J. (1991): "Internal and external determinants of foreign policy: West Germany and Great Britain during the two-track missile controversy", *International Studies Quarterly* 35(1): 87-108.

Dewan, T. & K. Dowding (2005): "The corrective effect of ministerial resignations on government popularity", *American Journal of Political Science* 49(1): 46-56.

Doucouliagos, H. (2005): "Publication Bias in the Economic Freedom and Economic Growth Literature", *Journal of Economic Surveys* 19(3): 367-387.

Doucouliagos, H. & M. Paldam (2008): "Aid effectiveness on growth: A meta study", *European Journal of Political Economy* 24(1): 1-24.

Doucouliagos, H. & M.A. Ulubaşoğlu (2008): "Democracy and economic growth: A meta-analysis", *American Journal of Political Science* 52(1): 61-83.

Doucouliagos, H. & T.D. Stanley (2008): "Theory competition and selectivity: Are all economic facts greatly exaggerated?", *Working Paper,* Economics Series SWP 2008/06, Melbourne: Deakin University, School of Accounting, Economics and Finance.

Doucouliagos, H. & T.D. Stanley (2009): "Publication selection bias in minimum-wage research? A meta-regression analysis", *British Journal of Industrial Relations* 47(2): 406-429.

Downs, A. (1957): *An Economic Theory of Democracy,* New York: Harper & Row.

The Economist (2008): "Publish and be wrong", 09.10.08, http://www.economist.com/science/displayStory.cfm?story_ id=12376658, accessed 28.10.08.

Eulau, H. & M.S. Lewis-Beck (eds.) (1985): *Economic Conditions and Electoral Outcomes in the United States and Western Europe*, New York: Agathon Press.

Fiorina M.P. & K. Shepsle (1989): "Is negative voting an artifact?", *American Journal of Political Science* 33: 423-439.

Frey, B.S. & F. Schneider (1978): "A politico-economic model of the United Kingdom", *Economic Journal* 88(350): 243-253.

Frey, B.S. & F. Schneider (1981): "A politico-economic model of the U.K.: New estimates and predictions", *Economic Journal* 91: 737-740 (1982: "Corrigendum: A politico-economic model of the U.K.: New estimates and predictions", *Economic Journal* 92: 410).

Frey, B.S. & F. Schneider (1982): "Politico-economic models in competition with alternative models: Which predicts better?", *European Journal of Political Research* 10: 241-254.

Gavin, N. & D. Sanders (2003): "The Press and its influence on British political attitudes under New Labour", *Political Studies* 51(3): 573-591.

Geller, D.S. & J.D. Singer (1998): *Nations at war: A scientific study of international conflict*, Cambridge: Cambridge University Press.

Geys, B. (2006): "Explaining voter turnout: a review of aggregate-level research", *Electoral Studies* 25: 637-663.

Glass, G.V. (1976): "Primary, Secondary, and Meta-Analysis of Research", *Educational Researcher* 5(10): 3-8.

Glass, G.V. (1977): "Integrating Findings: The Meta-Analysis of Research", *Review of Research in Education* 5: 351-379.

Görg, H. & E. Strobl (2001): "Multinational companies and productivity spillovers: a meta-analysis", *Economic Journal* 111: F723-F740.

Goodhart, C.A.E. & R.J. Bhansali (1970): "Political economy", *Political Studies* 18(1): 43-106.

Headrick, B. & D.J. Lanoue (1991): "Attention, asymmetry, and government popularity in Britain", *Western Political Quarterly* 44(1): 67-86.

Hedges, L.V. (1992): "Modeling publication selection effects in meta-analysis", *Statistical Science* 7 (2): 246-255.

Hibbing, J.R. (1987): "On the issues surrounding economic voting: Looking to the British case for answers", *Comparative Political Studies* 20(1): 3-33.

Hibbs, D.A. (2006): "Voting and the macroeconomy", pp. 565-585 in B.R. Weingast & D. Wittman (eds.), *The Oxford Handbook of Political Economy*, Oxford: Oxford University Press.

Hibbs, D.A., D. Rivers & N. Vasilatos (1982): "On the demand for economic outcomes: macroeconomic performance and mass political support in the United States, Great Britain, and Germany", *Journal of Politics* 43: 426-462.

Hibbs, D.A. & N. Vasilatos (1981): "Macroeconomic performance and mass political support in the United States and Great Britain", pp. 31-47 in D.A. Hibbs, B. Fassbender & R.D. Rivers (eds.), *Contemporary political economy: Studies on*

the interdependence of politics and economics, Amsterdam: North-Holland Publishing Company.
Hotelling, H. (1929): "Stability in competition", *Economic Journal* 39: 41-57.
Hunter, J.E. & F.L. Schmidt (2004): *Methods of Meta-Analysis: Correcting Error Bias in Research Findings*, 2nd ed., Thousand Oaks: Sage Publications.
Imbeau, L.M., F. Pétry & M. Lamari (2001): "Left-right party ideology and government policies: A meta-analysis", *European Journal of Political Research* 40: 1-29.
Inglehart, R. (1971): "The silent revolution in Europe: Intergenerational change in post-industrial societies", *American Political Science Review* 65(4): 991-1017.
Inglehart, R. (1977): *The Silent Revolution: Changing Values and Political Styles among Western Publics*, Princeton: Princeton University Press.
Jyllandsposten (2008): "På englenes side", *Indblik*, p. 1, 13.01.08
Key, V.O. (1964): *Politics, Parties & Pressure Groups,* New York: Thomas Y. Crowell.
Key, V.O. (1966): *The Responsible Electorate: Rationality in Presidential Voting 1936-1960,* Cambridge, MA: Harvard University Press.
Kramer, G.H. (1971): "Short-term fluctuations in U.S. voting behaviour, 1896-1964", *American Political Science Review* 65(1): 131-143.
Kramer, G.H. (1983): "The ecological fallacy revisited: Aggregate- versus individual-level findings on economics and elections, and sociotropic voting", *American Political Science Review* 77(1): 92-111.
Kristeligt Dagblad (2008): "Forskere ser ingen gevinst ved u-landsbistand", 14.01.08, http://www.kristeligt-dagblad.dk/artikel/274742:Udland--Forskere-ser-ingen-gevinst-ved-u-landsbistand, accessed 04.10.2009.
Kuhn, T.S. (1962): *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press.
Lanoue, D.J. & B. Headrick (1994): "Prime ministers, parties, and the public: The dynamics of government popularity in Great Britain", *The Public Opinion Quarterly* 58(2): 191-209.
Lau, R. (1985): "Two explanations for negativity effects in political behavior", *American Journal of Political Science* 29: 119-138.
Lau, R.R., L. Sigelman, C. Heldman & P. Babbitt (1999): "The effects of negative political advertisements: a meta-analytic assessment", *American Political Science Review* 93: 851-875.
Lau, R.R., L. Sigelman & I.B. Rovner (2007): "The effects of negative political campaigns: a meta-analytic reassessment", *The Journal of Politics* 69(4): 1176-1209.
Leamer E. & H. Leonard (1983): "Reporting the fragility of regression estimates", *Review of Economics and Statistics* 65: 306-317.
Lehrer, D., J. Leschke, S. Lhadhimi, A. Vasilu & B. Weiffen (2007): "Negative results in social science", *European Political Science* 6: 51-68.
Levy, J.S. (1997): "Prospect theory, rational choice, and international relations", *International Studies Quarterly* 41: 87-112.
Lewis-Beck, M.S. (1983): *Applied Regression. An Introduction* (Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-022), Beverly Hills: Sage.

Lewis-Beck, M.S., R. Nadeau & E. Bélanger (2004): "General election forecasts in the United Kingdom: A political economy model", *Electoral Studies* 23(2): 279-290.

Lewis-Beck, M.S. & M. Paldam (2000): "Economic voting: An introduction", *Electoral Studies* 19(2/3): 113-121.

Lewis-Beck, M.S. & M. Stegmaier (2000): "Economic determinants of electoral outcomes", *Annual Review of Political Science* 3: 183-219.

Lewis-Beck, M.S. & M. Stegmaier (2007): "Economic models of voting" in R. Dalton and H.-D. Klingemann (eds.), *The Oxford Handbook of Political Behaviour*, Oxford: Oxford University Press.

Light, R.J. & D.B. Pillemer (1984): *Summing Up: The Science of Reviewing Research*, Cambridge, MA: Harvard University Press.

Light, R.J. & P.V. Smith (1971): "Accumulating Evidence: Procedures for Resolving Contradictions Among Different Research Studies", *Harvard Educational Review* 41: 429-471.

Longhi, S., P. Nijkamp & J. Poot (2005): "A Meta-Analytic Assessment of the Effect of Immigration on Wages", *Journal of Economic Surveys* 19(3): 451-477.

Ludvigsen, S.S. (Paper 1): "Conditional evaluations in economic voting", paper for PhD dissertation, Aarhus: Dept. of Political Science, Aarhus University.

Ludvigsen, S.S. (Paper 2): "Party positions, voter preferences, and the cost of ruling", paper for PhD dissertation, Aarhus: Dept. of Political Science, Aarhus University.

Ludvigsen, S.S. (Paper 3): "Personal economic expectations and government popularity in the United Kingdom – a meta-analysis", paper for PhD dissertation, Aarhus: Dept. of Political Science, Aarhus University.

Ludvigsen, S.S. (Paper 4): "Inflation and government popularity in the United Kingdom – a meta-analysis", paper for PhD dissertation, Aarhus: Dept. of Political Science, Aarhus University.

Ludvigsen, S.S. (Paper 5): "Unemployment and government popularity in the United Kingdom – a meta-analysis", paper for PhD dissertation, Aarhus: Dept. of Political Science, Aarhus University.

Luke, D.A. (2004): *Multilevel Modeling* (Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-143), Thousand Oaks: Sage.

Macaskill, P., S.D. Walter & L. Irwig (2001): "A comparison of methods to detect publication bias in meta-analysis", *Statistics in Medicine* 20: 641-654.

McNemar, Q. (1960): "At random: sense and nonsense", *American Psychologist* 15: 295-300.

Medawar, P.B. (1963): "Is the scientific paper a fraud?", *Listener* 70: 377-378.

Melton, A. (1962): "Editorial", *Journal of Experimental Psychology* 64: 553-557.

Midtbø, T. (2007): *Regresjonsanalyse for samfunnsvitere. Med eksempler i SPSS*, Oslo: Universitetsforlaget.

Minford, P. & D. Peel (1982): "The political theory of the business cycle", *European Economic Review* 17: 253-270.

Monroe, K.R. (1984): *Presidential Popularity and the Economy*, New York: Praeger.

Mooney, C.Z. & R.D. Duval (1993): *Bootstrapping. A Nonparametric Approach to Statistical Inference* (Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-095), Newbury Park: Sage.

Moses, H., III, E.R. Dorsey, D.H.M. Matheson & S.O. Their (2005): "Financial anatomy of biomedical re-search", *Journal of the American Medical Association* 294: 1333-1342.

Mosley, P. (1978): "Images of the 'floating voter': Or the 'political business cycle' revisited", *Political Studies* 26: 375-394.

Mosley, P. (1984a): "'Popularity functions' and the role of the Media: A pilot study of the popular press", *British Journal of Political Science* 14(1): 117-129.

Mosley, P. (1984b): *The Making of Economic Policy: Theory and Evidence from Britain and the United States Since 1945*, New York: St. Martin's Press.

Mueller, J.E. (1970): "Presidential popularity from Truman to Johnson", *American Political Science Review* 64: 18-34.

Mueller, D.C. (2003): *Public Choice III*, Cambridge, UK: Cambridge University Press.

Mughan, A. (1987): "The 'hip pocket nerve' and electoral volatility in Australia and Great Britain", *Politics* (later *Australian Journal of Political Science*), 22(2): 66-75.

Mughan, A. (2004): "Forecasting and the outcome of the 2001 British General Election", *Electoral Studies* 23(2): 291-296.

Nadeau, R., R.G. Niemi & T. Amato (1996): "Prospective and comparative or retrospective and individual? Party leaders and party support in Great Britain", *British Journal of Political Science* 26(2): 245-258.

Nadeau, R., R.G. Niemi & T. Amato (2000): "Elite economic forecasts, economic news, mass economic expectations, and voting intentions in Great Britain", *European Journal of Political Research* 38(1): 135-170.

Nannestad, P. & M. Paldam (1994): "The VP-function: A survey of the literature on vote and popularity functions after 25 Years", *Public Choice* 79: 213-245.

Nannestad, P. & M. Paldam (1997): "The grievance asymmetry revisited: A micro study of economic voting in Denmark, 1986-92", *European Journal of Political Economy* 13: 81-99.

Nannestad, P. & M. Paldam (2002): "The cost of ruling – a foundation stone for two theories", in H. Dorussen and M. Taylor (eds.), *Economic Voting*, London: Routledge.

Narud, H.M. & B. Aardal (2007): "Økonomisk stemmegivning i oljefondets skygge", in B. Aardal (ed.), *Norske velgere – En studie av stortingsvalget 2005,* Oslo: N.W. Damm & Søn.

Narud, H.M. & H. Valen (2007): *Demokrati og ansvar: Politisk representasjon i et flerpartisystem,* Oslo: N.W. Damm & Søn.

Narud, H.M. & H. Valen (2008): "Coalition Membership and Electoral Performance in Western Europe", in K. Strøm, W.C. Müller & T. Bergmann (eds.), *Cabinets and Coalition Bargaining. The Democratic Life Cycle in Western Europe,* Oxford: Oxford University Press.

Norpoth, H. (1987): "Guns and butter and government popularity in Britain", *American Political Science Review* 81(3): 949-959.
Norpoth, H. (1992): *Confidence Regained: Economics, Mrs. Thatcher and the British Voter*, Ann Arbor: University of Michigan Press.
Norpoth, H., M.S. Lewis-Beck & J.D. Lafay (1991): *Economics and Politics – The Calculus of Support*, Ann Arbor: The University of Michigan Press.
Office for National Statistics (2009): http://www.statistics.gov.uk/statbase/tsdintro.asp, accessed 29.08.2009.
Paldam, M. (1991): "How robust is the Vote function? A study of seventeen nations over four decades", in Norpoth et al. 1991.
Paldam, M. (1997): "Political business cycles" in D. Mueller (ed.), *Perspectives on Public Choice: A Handbook*, Cambridge: Cambridge University Press.
Paldam, M. & F. Schneider (1980): "The macro-economic aspects of government and opposition popularity in Denmark 1957-78", *Nationaløkonomisk Tidsskrift* 118(2): 149-170.
Paldam, M. & P. Skott (1995): "A rational-voter explanation of the cost of ruling", *Public Choice* 83: 159-172.
Petersen, M.A. (2009): "Estimating standard errors in finance panel data sets: comparing approaches", *The Review of Financial Studies* 22 (1): 435-480.
Pissarides, C.A. (1980): "British government popularity and economic performance", *The Economic Journal* 90: 569-581.
Politiken (2008): "Forskning skyder dansk ulandsbistand i sænk", 13.01.08, http://politiken.dk/udland/article457175.ece, accessed 05.05.08.
Powell, G.B, Jr. & G.D. Whitten (1993): "A cross-national analysis of economic voting: Taking account of the political context", *American Journal of Political Science* 37(2): 391-414.
Price, S. & D. Sanders (1993): "Modeling government popularity in postwar Britain: A methodo-logical example", *American Journal of Political Science* 37(1): 317-334.
Price, S. & D. Sanders (1994): "Economic competence, rational expectations and government popularity in post-war Britain", *Manchester School* 62: 296-312.
Rabe-Hesketh, S., Skrondal, A. (2008): *Multilevel and Longitudinal Modeling Using Stata*, second ed., College Station: Stata Press.
Rattinger, H. (1981): "Unemployment and the 1976 election in Germany: Some findings at the aggregate and the individual level of analysis", in D.A. Hibbs, B. Fassbender & R.D. Rivers (eds.), *Contemporary political economy: Studies on the interdependence of politics and economics*, Amsterdam: North-Holland Publishing Company.
Rattinger, H. (1991): "Unemployment and elections in West Germany", pp. 49-62 in Norpoth et al. (1991).
Roscoe, D.D. & S. Jenkins (2005): "A meta-analysis of campaign contributions' impact on roll call voting", *Social Science Quarterly* 86: 52-68.

Rosenberger, R.S. & J.B. Loomis (2000): "Panel stratification in meta-analysis of environmental and natural resource economic studies", *Journal of Agricultural and Applied Economics* 32(3): 459-470.

Rosenthal, M.C. (1994): "The fugitive literature", in H. Cooper and L. Hedges (eds.), *The Handbook of Research Synthesis*, New York: Russell Sage Foundation.

Rummel, R.J. (1985): "Libertarian Propositions on Violence Within and Between Nations – A Test Against Published Results", *The Journal of Conflict Resolution* 29(3): 419-455.

Sanders, D. (1991): "Government popularity and the next General Election", *The Political Quarterly* 62(2): 235-261.

Sanders, D. (1996): "Economic performance, management competence and the outcome of the next General Election", *Political Studies* 44: 203-231.

Sanders, D. (1999): "Conservative incompetence, Labour responsibility and the feelgood factor: Why the economy failed to save the Conservatives in 1997", *Electoral Studies* 18(2): 251-270.

Sanders, D. (2000): "The real economy and the perceived economy in popularity functions: how much do voters need to know?: A study of British data, 1974-97", *Electoral Studies* 19(2-3): 275-294.

Sanders, D. (2005a): "Popularity function forecasts for the 2005 UK General Election", *The British Journal of Politics and International Relations* 7(2): 174-190.

Sanders, D. (2005b): "The political economy of UK party support, 1997-2004: Forecasts for the 2005 General Election", *Journal of Elections, Public Opinion, and Parties* 15(1): 47-71.

Sanders, D. & N. Gavin (2004): "Television news, economic perceptions and political preferences in Britain, 1997-2001", *The Journal of Politics* 66(4): 1245-1266.

Sanders, D., H. Ward, D. Marsh, with T. Fletcher (1987): "Government popularity and the Falklands War: A reassessment", *British Journal of Political Science* 17(3): 281-313.

Sanders, D. H. Ward & D. Marsh (1991): "Macroeconomics, the Falklands War, and the popularity of the Thatcher Government: A contrary view", pp. 161-182 in Norpoth et al. (1991).

Sidman, M. (1960): *Tactics of Scientific Research*. New York: Basic Books.

Smithies, A. (1941): "Optimum Location in Spatial Competition", *The Journal of Political Economy* 49: 423-439.

Stanley, T.D. (2001): "Wheat from Chaff: Meta-Analysis as Quantitative Literature Review", *The Journal of Economic Perspectives* 15(3): 131-150.

Stanley, T.D. (2005): "Beyond Publication Bias", *Journal of Economic Surveys* 19(3): 309-345.

Stanley, T.D. & S.B. Jarrell (1989): "Meta-regression analysis: A quantitative method of literature surveys", *Journal of Economic Surveys* 3: 54-67.

Stanley, T.D. & H. Doucouliagos (2007): "Identifying and Correcting Publication Selection Bias in the Efficiency-Wage Literature: Heckman Meta-Regression", *Working Paper,* Economics Series 2007-11, Melbourne: Deakin University, School of Accounting, Economics and Finance.

Stanley, T.D., H. Doucouliagos & S.B. Jarrell (2008): "Meta-regression analysis as the socio-economics of economics research", *The Journal of Socio-Economics* 37: 276-292.

Stanley, T.D., S.B. Jarrell & H. Doucouliagos (2009): "Could it be better to discard 90% of the data? A statistical paradox", *Working Paper,* Economics Series SWP 2009/13, Melbourne: Deakin University, School of Accounting, Economics and Finance.

Sterling, T.D. (1959): "Publication decisions and their possible effects on inferences drawn from tests of significance – or vice versa", *Journal of the American Statistical Association* 54: 30-34.

Stevenson, R.T. (2002): "The cost of ruling, cabinet duration, and the "median-gap" model", *Public Choice* 113(1-2): 157-178.

Stimson, J.A. (1991): *Public Opinion in America,* Boulder, CO: Westview.

Tversky, A. & D. Kahneman, D. (1992): "Advances in prospect theory: Cumulative representation of uncertainty", *Journal of Risk and Uncertainty* 5: 297-323.

Weakliem, D.L. (1986): "Economic conditions and government popularity in Great Britain, 1956-1983", *Political Behavior* 8(4): 335-355.

Weichselbaumber D. & R. Winter-Ebmer (2005): "A Meta-Analysis of the International Gender Wage Gap", *Journal of Economic Surveys* 19(3): 479-511.

Whiteley, P. (1984): "Inflation, unemployment and government popularity – dynamic models for the United States, Britain and West Germany", *Electoral Studies* 3(1): 3-24.

Whiteley, P. (1986): "Macroeconomic performance and government popularity in Britain: The short-run dynamics", *European Journal of Political Research* 14: 45-61.

Wood, W. & A.H. Eagly (2009): "Advantages of certainty and uncertainty", pp. 455-472 in Cooper *et al*. (2009).

Wooldridge, J.M. (2002): *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Yang, K. & M. Holzer (2006): "The performance – trust link: Implications for performance measurement', *Public Administration Review* 66: 114-126.

Young, N.S., J.P.A. Ioannidis & O. Al-Ubaydli (2008): "Why current publication practices may distort science", *PLoS Medicine*, 5 (10): e201. doi:10.1371/journal.pmed.0050201.

Executive summary in Norwegian

Kapittel 1

I det første kapittelet presenteres avhandlingens generelle tema, motivasjon, metode og struktur. Det generelle temaet er såkalte *valg- og popularitets-funksjoner* (VP-funksjoner), som er analyser av endringer i regjeringsoppslutning over tid, der økonomiske forklaringer står sentralt. Etter nærmere 40 år og rundt regnet 400 publikasjoner av slike studier står litteraturen fast i et problem som ble observert allerede etter ti år – nemlig at funnene ikke er robuste. Effektene av økonomiske variabler på regjeringsoppslutning synes å variere etter land, tidsrom, kontrollvariabler og modellspesifikasjoner. Heri ligger motivasjonen for avhandlingen min: avhandlingen er et sett med litteraturgjennomganger, der to artikler er tradisjonelle narrative gjennomganger av det teoretiske grunnlaget for modeller av økonomisk stemmegivning, og tre artikler er kvantitative sekundæranalyser av de empiriske funn i britiske VP-funksjoner.

Det er bruken av kvantitativ sekundæranalyse som er avhandlingens sterkeste bidrag til statsvitenskap generelt og til nordisk og britisk statsvitenskap spesielt. Metoden jeg anvender er *metaregresjon*, som i all "enkelhet" er regresjon av regresjoner, og dette brukes til å kontrollere for den innvirkning datagrunnlag, modellspesifikasjoner og kontrollvariabler har hatt på VP-funksjonens manglende robusthet (*land* holdes konstant, ettersom jeg kun analyserer studier av britisk regjeringsoppslutning).

I metaregresjon brukes de publiserte estimatene av de studerte sammenhenger som observasjoner av den avhengige variabelen, mens variasjonen i modellene av disse sammenhengene utgjør de uavhengige variablene. Denne formen for sekundæranalyse startet i undervisningsforskning og bredte seg raskt til psykologi, medisin og bio-statistikk, og senere til landforvaltningsstudier og økonomi. Det er noen av de teknikkene som har blitt utviklet innenfor økonomi som jeg bringer videre til statsvitenskapen. Innenfor statsvitenskapen har det til nå bare vært en håndfull meta-analyser, og svært få metaregresjoner. Jeg bruker metaregresjosmodeller som i statsvitenskapelig sammenheng kun er benyttet én gang tidligere (av to økonomer). På den bakgrunn hevder jeg at min avhandling er et unikt bidrag til statsvitenskapen.

Kapittel 2

Dette kapittelet oppsummerer den grunnleggende teoretiske litteraturgjennomgangen. Det har blitt utviklet fem teorier om økonomisk stemmegivning, hvorav fire er testbare gjennom meta-analysene. De fire teoriene som lar seg teste er:

1) *Ansvarlighetshypotesen* som antar at regjeringer – uavhengig av "farge" – holdes ansvarlige for økonomiske resultater og straffes for negativ utvikling og belønnes for positiv (denne kan også reformuleres som en *kompetansehypotese* der regjeringer holdes ansvarlige for velgernes generelle velferd);

2) *Klientellhypotesen* som forventer at velgere tiltrekkes av partier som har fremst på sin agenda å bekjempe et spesifikt problem som utvikler seg negativt. Det betyr at borgerlige partier forventes å få økt oppslutning (eller mildere straff enn arbeiderpartier) ved økende inflasjon, og at arbeiderpartier forventes å få økt oppslutning (eller mildere straff enn borgerlige partier) ved økende arbeidsledighet;

3) *Håndteringshypotesen* trekker asymmetrien i motsatt retning: i henhold til denne hypotesen vil kjernevelgerne holde "sitt" parti ansvarlig for sine hjertesaker. Partier som ikke er i stand til å håndtere sine hjertesaker vil dermed bli straffet for ikke å svare til velgernes forventninger. Et arbeiderparti (i regjering) vil således straffes hardere enn et borgerlig parti for økende arbeidsledighet, mens et borgerlig parti vil straffes hardere enn et arbeiderparti for økende inflasjon;

4) *Stabilitetshypotesen* forventer at tradisjonelt "ansvarlige" parti tiltrekker velgere i dårlige tider, uavhengig av partienes grunnsyn, og at velgerne i gode tider kan ta seg råd til å stemme på mer "spennende" partier. Stabilitetshypotesen kan derfor ses som en antitese til ansvarlighetshypotesen.

Den femte hypotesen er det nyeste tilskuddet av hypoteser (fra 2007) og den har ikke fått et etablert navn ennå. Jeg velger å kalle den *deltakerhypotesen*, og den kan ses som en antitese til stabilitetshypotesen (men er ikke helt lik med ansvarlighetshypotesen fordi den omhandler opposisjonspartier så vel som regjeringspartier). I henhold til denne hypotesen straffer velgerne parti som kan ses som deltakere i politikkutforming, også de som er med å forhandle om politikk uten å sitte i regjering. I nedgangstider kan velgerne vurdere at store opposisjonspartier også har hatt et ansvar for politikkutformingen, og skal derfor også straffes. Siden dette er en hypotese som skiller seg fra ansvarlighetshypotesen i sitt syn på opposisjonsparti, kan den derfor ikke testes med det analysedesign som jeg har valgt (som kun tester oppslutning for regjeringspartier). Jeg ser for øvrig ikke denne teorien som spesielt sentral.

Etter å ha presentert de generelle teoriene, fortsetter kapittel 2 med å oppsummere og presentere en rekke modererende forhold rundt økonomisk stemmegivning. Det argumenteres her for at VP-funksjonenes manglende stabilitet kan skyldes disse modererende forholdene, og at de generelle teoriene må testes med disse forholdene tatt i betraktning: det være seg

velgernes natur – hvordan de håndterer informasjon og strukturerer sine preferanser, betydningen av partitilhørighet, asymmetriske evalueringer, (u)klare ansvarsforhold, politiske og økonomiske institusjoner, media, regjeringsslitasje, og velgernes ulike perspektiv (fremadskuende versus tilbakeskuende, egoistiske versus altruistiske, og kort versus lang tidshorisont).

Gitt alle disse modererende forholdene, og den typiske mangel på frihetsgrader i tidsserieanalyser av regjeringsoppslutning, argumenteres det med at enkeltstudier vil ha problemer med å modellere inn alle forhold, og det konkluderes med at meta-analyse er en mulig måte å rydde opp i kompleksiteten på. Med andre ord, så legger dette kapittelet og det tilhørende paperet ytterligere motivasjon til de empiriske papers.

Kapittel 3

Dette kapittelet oppsummerer den andre teoretiske litteraturgjennomgangen, som er en gjennomgang av mulige forklaringer på regjeringsslitasje der også flere egne teoretiske bidrag presenteres. Litteraturdiskusjonen samler tråder og trekker noen av dem videre til nye forklaringer på regjeringsslitasje, og som sådan er dette paperet avhandlingens primære teoretiske bidrag.

Regjeringsslitasje er riktignok et lite sidespor i avhandlingen, da det ikke er en sentral del av meta-analysene, men det inngår likevel som en kontrollvariabel i metaregresjonene, og er et viktig fenomen i studiet av regjeringers vekst og fall. Av den grunn føler jeg at paperet utgjør en viktig del av avhandlingen.

Kapittel 3 er en detaljert sammenfatning av det tilhørende paperet, og de åtte mulige forklaringene på regjeringsslitasje som presenteres i paperet er:

1) *en statistisk artefakt* – regjeringsslitasje kan med andre ord skyldes statistiske årsaker fordi regjeringer oftest utgjøres av valgvinnere som enten kan ha fått feiltermene på sin side, og som da med stor sannsynlighet vil ha nøytrale eller negative feiltermer ved neste valg, eller fordi valgvinnere gjerne har flere stemmer å tape enn å vinne ved neste valg, og derfor har lettere for å få redusert oppslutning enn økt oppslutning;

2) *minoritetskoalisjonsteorien* – hvor opposisjonspartier kan ha lettere for å "snekre sammen" en koalisjon av ulike grupper ved å gi inkonsistente valgløfter, og som de blir avslørt på hvis de kommer i regjering. De godtroende eller de som ikke så inkonsistensen vil så straffe den nye regjeringen for dette;

3) *straffeasymmetri* – det kan være lettere for velgere å registrere en forverring enn en forbedring, og dermed belønnes ikke regjeringer for en forbedring like mye som de straffes for en motsvarende forverring. Men det kan også være andre grunner til straffeasymmetri. To øvrige etablerte

forklaringer er at velgerne har tapsaversjon eller at de oppfører seg som prinsipaler der asymmetri inngår i sanksjonsskjemaet. Jeg foreslår to ytterligere forklaringer med utgangspunkt i sammensetningen av velgermassen. For det første kan det hende at det forekommer en *belønningsasymmetri* mellom grupper av velgere, der velgerne straffer like hardt for uønskede utfall, men belønner ulikt (noe jeg illustrerer med et eksempel av egoister som ønsker skattelettelser mot altruister som ønsker skattehevninger – begge gruppene vil reagere umiddelbart på politikk som ikke er i henhold til deres preferanser, mens sistnevnte vil reagere saktere på politikk som er i henhold til preferansene enn førstnevnte). For det andre kan det simpelthen hende at velgere av ulik natur reagerer ulikt på økonomisk politikk som rammer dem personlig. For en altruist skal det strengt tatt ikke bety noe hvorvidt vedkommende selv er arbeidsledig eller ikke, mens en egoist strengt tatt kun vil belønne reduksjoner i arbeidsledigheten inntil vedkommende selv har kommet seg ut av arbeidsledighet og over i en sikker stilling;

4) *Mediantomromsteorien* – medianvelgere kan ha rasjonelle interesser i å alternere deres stemme mellom to blokker fra valg til valg, for dermed å holde blokkene nær sentrum;

5) *Politiske forretningssykluser* – det vil være en avstand mellom medianvelgeren og median*partimedlemmene*. Asymmetrisk ideologisk elastisitet mellom regjeringspartier og opposisjonspartier kan da forklare regjeringsslitasje hvis medlemmer av opposisjonspartier gir partiledelsene mer bevegelsesfrihet i retning av medianvelgerne enn det (mulige) tapsaverse medlemmer av regjeringspartier tillater;

6) *Evigvarende opposisjonsparti* – det kan finnes parti som har få realistiske sjanser til regjeringsdeltakelse, men som likevel tiltrekker seg velgere. Regjeringsslitasjen vil øke i takt med slike partiers vekst, men før eller siden vil veksten stoppe opp, enten fordi partiene har uttømt velgerpotensialet, eller fordi de til syvende og sist utgjør et uunngåelig regjeringsalternativ.

7) *Asymmetriske valgkostnader* – det koster velgerne noe å stemme, i form av tid eller – som Aftenposten kunne melde i 2007 – bulker på parkeringsplassen, og nytteverdien av å stemme er liten. Likevel kan den være større for velgere av opposisjonsparti enn for velgere av regjeringsparti, fordi en opposisjonsstemme *både* er en støtte til det partiet man stemmer på *og* en protest mot den sittende regjeringen, mens en stemme på regjeringen kun i ekstreme tilfeller vil kunne ses som en protest mot opposisjonen. Alt annet likt, vil nettokostnadene ved å stemme på regjeringspartier kunne ses å være høyere enn nettokostnadene ved å stemme på opposisjonen, og dermed vil opposisjonspartier lettere kunne mobilisere velgere enn det regjeringspartier kan, hvilket vil slå ut som regjeringsslitasje.

8) *Selvoppfyllende profeti* – regjeringer som innser at deres dager er talte kan få det travelt med å implementere partienes preferanser heller enn medianvelgernes, og dermed bidra til sin egen regjeringsslitasje.

Kapittel 4

Kapittel 4 er et kort kapittel som kritiserer konvensjonelle narrative litteraturgjennomganger og går noe mer i dybden på motivasjonen for å bruke metaregresjon i stedet. Dess mer kunnskap som produseres, dess vanskeligere kan det være å samle den akkumulerte kunnskapen i narrative litteraturgjennomganger, fordi det finnes et vell av analyseteknikker, modellspesifikasjoner, datautvalg, caseutvalg, teorier, osv som vanskeliggjør sammenlikning i en tradisjonell litteraturgjennomgang. Denne variasjonen kan kontrolleres for i metaregresjoner, så fremt man er i stand til å standardisere sammenhengene av interesse. En annen fordel med kvantitative sekundæranalyser er at de i større grad er i stand til å dekke all publisert kunnskap (eller et genuint tilfeldig utvalg) knyttet til en problemstilling enn hva en narrativ litteraturgjennomgang kan gjøre.

Kapittel 5

Kapittel 5 omhandler et fenomen av sterk vitenskapsfilosofisk og -sosiologisk interesse, nemlig *publiseringsbias*, som er en fordreining av virkeligheten som følge av forskeres, redaktørers, og fagfellers tendens til å støtte funn som helst bekrefter konvensjonelle antakelser (rådende syn) *og* er både spennende og statistisk signifikante. Heri ligger også en kritikk av den konvensjonelle narrative litteraturgjennomgang, fordi en litteraturgjennomgang som ikke er i stand til å kontrollere for variasjon i studienes design og gjennomføring heller ikke kan kontrollere for eventuelle bias i publiseringen av resultater. Faktisk har den tradisjonelle gjennomgang ofte *selv* en tendens til å plukke ut interessante resultater som videreformidles, og dermed forekommer ikke bare en reproduksjon av publiseringsbias i litteraturoversiktene, men også forsterkning av seleksjonseffekten via et annengrads publiseringsbias. Dette er altså nok en kritikk av den tradisjonelle gjennomgangsformen.

Det er åpenbart problematisk at slike seleksjonseffekter produseres, reproduseres og forsterkes, fordi den vitenskapelige produksjon da skaper et fordreid bilde av virkeligheten. Dette kan få særdeles alvorlige konsekvenser, spesielt innenfor medisinsk forskning, men også innenfor de mykere vitenskaper, og det er mitt argument at det vil være utilgivelig å overse dette fenomenet når litteratur skal sammenfattes. Dermed bygges det opp enda et argument for meta-analyse, og slik jeg ser det er introduksjonen av estimat og kontroll av publiseringsbias mitt viktigste bidrag til statsvitenskapen.

Kapittel 6

Dette kapittelet viser ved hjelp av et konstruert datasett hvordan meta-regresjon kan brukes til å estimere og kontrollere for publiseringsbias. Den grunnleggende tankegangen bak dette er at studienes presisjonsnivå skal være uten sammenheng med de publiserte estimatene i en litteratur som er uten bias. Dess lavere presisjonsnivå en modell har, dess større variasjon vil forekomme, og denne variasjonen vil slå ut begge veier med et gjennomsnitt på null i en litteratur som er uten publiseringsbias. Hvis det derimot forekommer ikke-tilfeldig seleksjon av resultater til publisering, vil studienes presisjonsnivå være korrelert med estimatene som publiseres, fordi studier med lavt presisjonsnivå vil publiseres dersom de kommer frem til "korrekt" resultat, men ikke dersom de kommer frem til et "ikke-korrekt" resultat. I kapittelet viser jeg både ved hjelp av ligninger og figurer hvordan man går frem for å avklare dette.

Jeg diskuterer også noen problemer, begrensninger og mulige ankepunkter mot metaregresjon i slutten av dette kapittelet. Utfordringene ligger i at man naturligvis behøver et visst antall publiserte estimat for å kunne bruke kvantitativ analyse på dem, at disse estimatene selv må være produserte via kvantitative analyser, at observasjonene normalt ikke vil være uavhengige av hverandre, og at man kan kritisere sekundæranalyser for ikke å være originalt arbeid.

Kapittel 7

Etter å ha lagt det teoretiske og metodiske grunnlaget i de foregående kapitlene, kommer avhandlingen nå til den første empiriske analysen. Denne er en meta-analyse av effekten av personlige økonomiske forventninger på britisk regjeringsoppslutning. Personlige økonomiske forventinger er et aggregert mål av i hvilken grad surveyrespondenter tror at deres husholdnings økonomiske/finansielle situasjon vil forbedres i løpet av det kommende året. Det er altså både et subjektivt og et prospektivt mål, og det har vært reist noen kritiske spørsmål til bruken av et slikt mål i VP-funksjoner. De teoretiske problemene ved slik bruk griper jeg ikke fatt i her, i stedet forsøker jeg å avklare i hvor sterk grad de publiserte empiriske analysene har kommet frem til at slike forventinger har en effekt på regjeringsoppslutning i Storbritannia. Jeg kommer frem til at det er en rimelig sterk standardisert effekt på tross av en viss påvirkning fra modellspesifikasjoner etc., og til tross for at jeg kan påvise publiseringsbias i litteraturen. Hvis jeg ikke hadde kontrollert for publiseringsbias ville jeg trodd at effekten var ca. 40 prosent sterkere – alt annet like – enn det jeg påviser i min meta-analyse. Jeg konkluderer likevel

med at det synes vel verdt å inkludere personlige økonomiske forventinger i VP-funksjoner også i fremtiden. I tillegg viser meta-analysen at det ikke er en forskjell i hvordan det britiske arbeiderpartiet eller de konservative belønnes for økende forventinger. Dette funnet gir da støtte til ansvarlighetshypotesen.

Kapittel 8

Kapittel 8 oppsummerer meta-analysen av effekten av inflasjon på britisk regjeringsoppslutning. I tillegg kan kapittel 8 ses som en appendiks til det empiriske paperet, da kapittelet inneholder en lengre variabelbeskrivelse som det ikke var plass til i paperet som er skrevet i artikkellengde. Den lange variabelbeskrivelsen indikerer den heterogenitet som finnes i litteraturen, både hva gjelder modellspesifikasjoner *og* resultater. Metaregresjonene av denne variasjonen viser at det også her finnes en klar effekt av inflasjon på regjeringsoppslutning, uten at det har forekommet et systematisk publiseringsbias. Variasjonen i resultater skyldes i hovedsak forskjeller i modellspesifikasjonene, med unntak av ulike observasjonsperioder som synes å ha liten effekt. Til sist viser metaregresjonene at begge partiene som har hatt regjeringsmakt har tapt på inflasjon, men at Arbeiderpartiet har blitt straffet omtrent 50 prosent hardere enn de konservative – alt annet likt. Dermed synes klientellhypotesen isolert sett å bli støttet (i sin modererte versjon).

Kapittel 9

Dette kapittelet oppsummerer meta-analysen av effekten av arbeidsledighet. I det tilhørende paperet ble det plass til variabeldiskusjonen, så dette er ikke tatt med her, men kapittelet inneholder en regresjonsmodell som det ikke var plass til i paperet. Det er stor likhet mellom paperet om arbeidsledighet og paperet om inflasjon. Mange av de modererende variablene i metaregresjonene er like, og litteraturen som analysen bygger på er i stor grad den samme. Jeg bruker derimot litt ulike regresjonsteknikker – der den vesentlige forskjellen ligger i bruken av vektede modeller. Begrunnelsen for det er at jeg i utgangspunktet får et uventet fortegn på kontrollen for publiseringsbias. De opprinnelige modellene viser at dess mindre presisjon studiene har, dess mindre blir de publiserte effektene. Ved hjelp av de vektede modellene (der presisjonen brukes som vekt) får jeg avklart at publiseringsbiaset ikke er til stede.

Resultatene fra metaregresjonene viser for øvrig at de ulike måtene å spesifisere arbeidsledighet på er uten effekt på resultatene, men at spesifikasjonen av den avhengige variabelen og utvalget av kontrollvariabler har en effekt. Alt annet like, er effekten av arbeidsledighet på regjeringsoppslutning

stor, og det er kun en marginal forskjell mellom Arbeiderpartiet og de konservative. Isolert sett konkluderer jeg da med at ansvarlighetshypotesen støttes.

Kapittel 10

Kapittel 10 sammenligner noen av funnene fra de tre empiriske paperne. Det er fire aspekt jeg fokuserer på: publiseringsbias, konstantleddene, teoritestene og noen andre funn av sammenliknbar interesse.

Jeg argumenterer for at publiseringsbiaset som ble funnet i litteraturen om personlige økonomiske forventinger skyldes den mest aktive forskerens tendens til å kun publisere sine reduserte modeller. Disse modellene vil per definisjon kun inneholde signifikante variabler, og dermed får dette konsekvenser når resultatene aggregeres. Jeg mener derfor at kontrollen for publiseringsbias var nyttig og nødvendig, selv om forfatterne av primærlitteraturen ikke nødvendigvis har hatt et uetisk motiv i sin publiseringsseleksjon.

Konstantleddet er av spesiell interesse i metaregresjoner. Det skyldes at konstanten representerer den sammenhengen av interesse der alle variablene er satt til null. I mine metaregresjoner har jeg forsøkt å spesifisere variablene på en slik måte at konstantene blir sammenliknbare mellom analysene. Dette har jeg gjort ved å spesifisere alle de modererende dummyvariablene slik at konstanten representerer den simpleste form for sammenheng av interesse. Altså, at personlige økonomiske forventinger, inflasjon, arbeidsledighet og regjeringsoppslutning er målt lineært, i nivå, uten etterslep, for begge partiene, fra popularitetsfunksjoner, osv. Likevel er ikke personlige økonomiske forventinger fullstendig sammenliknbar med inflasjon og arbeidsledighet, fordi antallet publiserte estimat begrenset frihetsgradene mine til å ha like omfattende multivariate metaregresjoner. På tross av dette, kommer det frem et inntrykk av at britiske regjeringer har hatt mest å frykte fra økende arbeidsledighet, og minst å frykte fra økende inflasjon.

Hva gjelder det teoretiske, så har jeg trukket ulike konklusjoner isolert sett. Meta-analysene av personlige økonomiske forventinger og av arbeidsledighet støtter ansvarlighetshypotesen, mens analysen av inflasjon synes å støtte den modererte klientellhypotesen. Men når man ser denne i sammenheng med de to andre analysene, finner jeg ingen grunn til å forkaste ansvarlighetshypotesen.

De øvrige funnene vedrører modererende variabler som jeg ikke har beskrevet i dette sammendraget på norsk, så jeg vil ikke gå inn i detaljer her.

Kapittel 11

Dette er konklusjonskapittelet, og oppsummerer avhandlingens generelle tema, motivasjon og resultater. Hele avhandlingen kan ses som et sett av

litteraturgjennomganger, der to papers oppsummerer og bygger på det teoretiske grunnlaget for avhandlingen, tre papers bruker metaregresjoner til å kvantitativt sammenfatte deler av den empiriske litteraturen, og denne overordnede sammenfatningen oppsummerer antologien. I tillegg har denne sammenfatningen gitt mer detaljert informasjon om metaregresjonsanalyse, hvilket til syvende og sist er det jeg ser som det viktigste bidraget i denne avhandlingen. De teoretiske diskusjonene og analyseresultatene er i seg selv viktige, men det er også viktig å få importert en nyttig vitenskapssosiologisk metode til statsvitenskapen generelt.

Meta-analysene har vært gjennomført på et forskningsfelt som i primærstudiene har produsert en del ustabile empiriske resultat. Ustabiliteten har stammet fra forskjeller i undersøkelsesopplegg, problemstillinger, data, analyseteknikker, publiseringspraksis, osv. Dette har det vært nødvendig å kontrollere effekten av før man evt. går til det drastiske skritt å hevde at forskningsfeltet gir uklare resultat som ikke kan akkumuleres, og derfor har lite for seg. I tillegg til å være grunnleggende vitenskapssosiologisk, gjør metaregresjon oss i stand til nettopp å akkumulere resultater på tvers av et heterogent forskningsfelt på en måte som tradisjonelle narrative litteraturgjennomganger ikke kan gjøre.

Jeg har også kritisert de tradisjonelle litteraturgjennomganger for deres manglende metode for å kontrollere for publiseringsbias, og jeg har – så vidt meg bekjent – vært første statsviter til å vise hvordan dette kan gjøres ved hjelp av kvantitative sekundæranalyser.

Resultatene fra metaregresjonene viser at ph.d.-prosjektet mitt ikke har vært en obduksjon slik tittelen gjenspeiler, men heller en sykdomsstudie. Det har vært noenlunde stille fra forskningsfeltet de senere år, men jeg mener å kunne vise at meta-analyse er en måte å komme seg videre på, og en metode som gjør at man ikke trenger å være bekymret for å produsere nye (åpenbart) motstridende funn. For det første kan disse motsetninger overkommes når resultatene skal sammenfattes, nettopp ved hjelp av metaregresjon, og for det andre har man nå avdekket hvordan spesifikasjonsforskjeller spiller inn på publiserte resultat, noe man da kan ta høyde for i senere studier.