PHD SCHOOL BOCCONI UNIVERSITY
PHD IN ECONOMICS

Essays in Empirical Political Economy and Macroeconometrics

Candidate:
Paolo Bianchi
Matr. 1003870

Commission:
Guido Tabellini
Carlo Ambrogio Favero
Tommaso Nannicini

Ciclo XX

Essays in Empirical Political Economy and Macroeconometrics

Paolo Bianchi

January 30th, 2009

Contents

Preface

I would like to thank my parents and my brother Michele, for their love and support; my wife, Yuan for her encouragement, support and patience. Without their help this degree would not have been possible.

I would like to thank my committee members for their encouragement and support. Thanks to Prof. Calro Favero for his suggestions and supervision regarding the macroeconometrics chapter. Thanks to Prof. Tommaso Nannicini for his help with the political economy chapters. I thank Prof. Guido Tabellini for being an excellent supervisor. I have greatly enjoyed working with someone I admire so much.

Introduction

This thesis contains chapters which deal with very different issues. The while the first two chapters can be framed in the vast field of political economy, the last chapter compare two Bayesian approaches to assess economic models within the Vector Autoregression framework, a benchmark in most empirical macroeconomics. I study the institutional feature of legislature size and its impact on the size of government using data from Italian municipalities.

Part I

Empirical Political Economy

Chapter 1

Council Size Effect in Italian Municipalities: Results from a Natural Experiment

1.1 Introduction

The debate on the proper size of legislative assemblies often underlines the so-called `common pool' problem: the larger the number of represented interests in the assembly, the larger the propensity to spend in projects financed with resources drawn from the community at large. The common pool argument was formalized in [22] and a wide variety of empirical studies found support for this intuition.[1] Despite the positive relation between the size of legislatures and the size of government looked one of the most corroborated stylized facts in political economy, recent contributions, both theoretical, [21], and empirical, added further insight and casted doubt on the directions of the causal effect of legislature size on government size.

The first formal argument why the number of policy makers might affect spending choices was provided by [22]. The argument is that legislators will try to benefit their constituents at the expense of the general community through pork barrel spending and other distributive policies and, since each legislator will internalize all the benefits from spending but only a fraction of the costs, this would give rise to excessive government spending. There are a number of empirical studies that find support for this claim but it is questionable whether they have identified a causal relationship since they have not convincingly addressed the endogeneity of legislature size. This is discussed further below.[2]

Legislature size is also of perennial interest to policy makers. For example, one of the earliest discussions of legislature size appears in the Federalist Papers.[3] The debate about the appropriate size is still going on today in many countries. In England, for example, there is a current research project on council size and democracy on behalf of the Electoral Commission, which is an independent body set up by the UK Parliament. The outcome of this research is intended to provide the Boundary Committee for England a robust basis for what might be the appropriate council size. Clearly, one should also take into account any potential effects of council size on the budget when considering the appropriate size of a legislative body.

The key contribution of this paper is to estimate the causal effect of

[1] The following studies have found a positive association between legislature size and the size of government: [2], [6], [8, 9] and [16]

[2] The following studies have found a positive association between legislature size and the size of government: [2], [6], [8, 9] and [16]

[3] In the Federalist papers it is argued that the number of representatives must be large enough to possess knowledge of the interests of numerous constituents and make collusion against the public interest difficult, but small enough to avoid the 'confusion and intemperance of a multitude.'

council size on government spending. The challenge of estimating a causal effect is, of course, to find some credible exogenous source of variation in legislature size. Generally, it is very difficult to find such variation since it is the policymakers themselves that decide on both size and policy which makes legislature size endogenous to policy. Noetheless, the national laws that regulat the council-size in Italian local governments provide an unusually credible source of exogenous variation in size. In Italy the council size is a *deterministic* function of the population size, i.e. the council size must be 13, 17, 21, 31, 41, 47, 51, or 61 depending on the population size in the locality. The law thus creates discontinuities in the council size, which provides the opportunity to implement a regression-discontinuity (RD) design.

The RD analysis in this paper will, however differ from the standard one where the comparison is made with respect to the outcomes of different subjects whose value of an underlying targeting variable is 'just below' and 'just above' a fixed threshold (e.g., [10], [11]). A *between-subjects* RD design is the appropriate method when there are many subjects close to the threshold as in many RD applications in labor economics which uses very large micro data sets. However, when the subjects are political entities (e.g., countries, states, cities, and localities) there will typically only be a few observations around the threshold which give rise to a problem of discreteness in the treatment-determining covariate.[4] In this case, the treatment effect is not identified without assuming a parametric functional form for the model relating the outcomes of interest to the treatment-determining variable, as discussed by [12].

In this paper, I argue that a *within-subject* RD design might be preferable to a between-subjects RD design in terms of moth efficiency and bias when the covariate that determines treatment is highly discrete. The idea behind the within subject RD approach is to compare the outcome for the same subject 'just before' and 'just after' the policy change since it seems likely that relationship between the treatment-determining variable and the outcome will be approximately time-invariant. In other words, comparing the same subject under two different treatment conditions (e.g., an Italian municipality that changed its council size from 13 to 17) not only removes much variability due to the different characteristics of the subjects but also reduces the problem of functional form misspecification since a within-subject RD design effectively controls for *any time-invariant* functional form relationship between the treatment-determining variable and the outcomes of

[4]Clearly when there are only observations from a single subject, i.e., time series data, it is not even possible to use a between-subject RD design but one can still do a within-subject RD design.

interest. Another attractive feature of using a within subject RD design is that one can address issues concerning the estimation of dynamic causal effects. Studying dynamic policy-making is an important issue in the political economics literature.[5]

Using a within-subject RD design, the results from this paper show that there is a *negative* relationship between the council size and the size of government. The estimated council-size effect is also economically large. The estimate council-size in Italy suggest that increasing the council size with 4 council members would lead to a 2.3 percent reduction in spending. Thus, the finding of a negative council-size effect in two different settings bolsters claims of external validity. Moreover, since the negative council size effect is found at multiple treatment threshold this also lends additional credibility to that the findings can be generalized.

The result of a negative relationship between council size and government size is strongly at odds with the conventional wisdom based on the model by [22] and the previous empirical work supporting it. One potential reason for the conflicting findings is that predictions from the [22] model are not applicable since Italy have proportional representation systems with multimember districts while the model is based on plurality rule and single member districts. However, all of the previous studies have used data with multimember districts and they still find a positive relationship.[6] For example, in [2] study more than 85 percent of the US cities have at large electoral systems, in which candidates for office are elected from the entire jurisdiction.

A second potential reason for the conflicting results may therefore be that previous studies have failed to establish a *causal* relationship since they have not properly addressed the endogeneity of council size. In fact, in this paper I also find a positive *statistical association* between council size and size of government when I do not take into account the endogeneity of council size suggesting that previous work is plagued by endogeneity problems.

How do we explain the negative relationship between council size and government size? One explanation is provided by [21] who show that the prediction of a positive relationship from the model by [22] is not robust. Specifically they show that the relationship between legislature size and government spending can also be negative. Thus, from a theoretical point of view the direction of the relationship between legislature size and government size is not as clear cut as the conventional wisdom would suggest.

Nonetheless, the explanations provided by Weingast et al. and Primo and

[5] There is recent theoretical work which explicitly study dynamic issues in legislative policy making, see e.g., [4]

[6] Most countries in the cross-country studies by studies by [6] and [16]

Snyder are ultimately unsatisfactory from a game theoretical point of view since the legislature is assumed to adopt a norm of universalism, i.e. each legislator chooses the amount of public spending that he would like for his district. These desired spending levels are then passed unanimously by the legislature. This postulated behavior of the legislators is hard to make sense of since at the time of voting for the omnibus bill, all the legislators would be better off agreeing to reduce all their spending levels. For this reason, recent theoretical work on legislative policy-making is instead based on a legislative bargaining model by [3] but this work do not explicitly investigate the link between legislature size and economic policy choices.

The results from this paper show that more theoretical work is needed on the relationship between legislature size and spending choices. This work has to take into account that the negative relationship is found in a specific political setting: Italy has a closed list PR system with strong political polarization and where the decision in the council is taken by simple majority. The negative effect are at work for council sizes in the range 13-21.

This paper contributes to a number of different literatures. First, it is related to the literature that investigates how political institutions shape economic policies. For example, models by [17], [13], and [14] compare how different electoral rules lead to different fiscal policies. Economic effects of other dimensions of political institutions have also been studied. [18] compare how different forms of government (parliamentary versus presidential) lead to different fiscal policy outcomes. [19, 20] create a comprehensive data set on political institutions and empirically investigate how different constitutional arrangements shape fiscal policies. While the empirical results from the research program of comparative political economy are very interesting, it faces very challenging identification problems as discussed by [19], and [1]. In fact, Acemoglu ‘questions whether this research has successfully uncovered causal effects.’

1.2 Institutional Details and Data

This section describes the local governments in Italy with a specific focus on the national council size laws that provide the source of variation used to estimate the effect of legislature size on government size. It also presents the data used in the empirical analysis.

1.2.1 Italian local governments

As of 2006 there are 8101 municipalities (*Comuni*) in Italy[7], they are part of a four level government system, the other three levels are the central government, the 20 regions (*Regioni*), and the 103 provinces (*Province*).

Each municipality has a mayor (*Sindaco*, a cabinet (*Giunta*), and a municipality council (*Consiglio Comunale*). The municipality council and the mayor are elected directly by the population, while the executive cabinet is appointed by the mayor. The council is the municipality legislative authority. It is responsible for monitoring the implementation of the political program by the elected mayor and the cabinet. The main responsibilities of Italian municipalities concern services to people, services to the community, local economic development, and the use of territory. Typical services provided are water supply, waste disposal, public local transport, elementary education etc.

A national law prescribes a specific number of council members in relation to the legal population (*Popolazione Legale*) of the municipality. The legal population is determined by the last official census. Every 10 years in Italy, the census is independently carried out by the national statistics authority, ISTAT[8]. The last two censuses were in 1991 and 2001. The data collected by the 2001 census has been declared the legal population of Italy and of its municipalities by the ordinary supplement to the April 7th, 2003 n. 54 Italian Official Gazette (*Gazzetta Ufficiale della Republica Italiana*). After these figures were published, every newly elected municipality council had its size determined by the 2001 census population. Most of the municipalities will keep their council size unchanged while others will be forced by law to increase or decrease the number of elected councilors. Since the population size data are produced independently by the Italian statistical institute, there is no possibility of manipulation by the local governments. I use this variation to identify the causal effect of the size of the local legislative body on the size of local government. The council size law is displayed in Table 1.3 and it states that if a municipality's population is less or equal to 3,000 the council must consist of 13 members; if the population is larger than 3,000 but less or equal to 10,000 the law states that council size must be 17, etc.

1.2.2 Data and Sample Selection

To measure the size of the local government, on the one hand, I use the current expenditures per capita which is a widely used measure; on the other

[7]See `http://www.interno.it/` for information about local governments in Italy
[8]See `http://www.istat.it`

hand, I focus on the per capita revenues from taxes and tariffs that are decided by the local government. Municipality budget data are available at the Italian Ministry of Internal Affairs website and cover the years from 1998 to 2006. Information is available at the aggregated level as well as disaggregated by destination (*Intervento*): personnel, consumable goods, services, use of third party goods, extraordinary expenditures, interests, taxes and amortization. The last two census population and the end-of-the-year population are available at ISTAT[9]. Also the proportion of people under 15 and above 65 years of age are elaborations from ISTAT data.

Information about legislatures' characteristics, such as the year of election, education of mayor and council members, can be found in the archive of local and regional elected administrators (*Anagrafe Amministratori Locali e Regionali*) , which is kept and updated by the Central Directorate of Electoral Services at the Italian Ministry of Internal Affairs[10]. The archive also includes some individual information as the date of birth, gender, electoral list, and profession, etc.

In selecting the municipalities employed in the empirical analysis it is important to take into account two divergent requirements. On the one hand, I need to include a sufficiently large number of observations crossing the threshold in order to increase statistical power and to meaningfully appeal to asymptotic results for inference; on the other hand, I need a control group with good matching properties to comply with the conditional mean assumption. Therefore, to identify the effect of the different council size, I focus on municipalities with the population size around the threshold, as recorded by the 2001 census. Italian law allows for 7 thresholds: at the population sizes of 3,001, 10,001, 30,001, 100,001, 250,001, 500,001 and 1,000,001. I choose to restrict the population interval around each of the discontinuity points to ± 5 percent i.e., those local governments with a population size (reported by the census) in the set of intervals $\{$ [2850, 3150], [9500, 10500], [28500, 31500], [95000, 105000], [237500, 262500], [475000, 525000], [950000, 1050000]$\}$. The information about this sample is provided in Table 1.4. For example, Table 1.4 reveals that there were only 10 municipalities in the range [28500, 30000] and 18 in the range [30001, 31500]. Among them 8 municipalities increased the council size and only 1 whose council size decreased. It is possible to increase the number of observations by widening the interval around the threshold, but this would produce also a lower matching quality between the treated and the control group.

Interesting enough, Table 1.4 shows that there are very few municipali-

[9]See `http://www.istat.it`

[10]See `http://amministratori.interno.it/index_amminist_cit.html`

ties in the intervals around the 4 highest thresholds, among these only one municipality experienced a change in the number of council members in the interval around the 100,001 threshold. Clearly these data will not provide any remarkable information, therefore I decide to ignore it and proceed into the analysis using the first three thresholds.

The dataset structure and the population-based council size law allow for two different approaches to identify the effect of the number of legislature members on the local government size. The first approach crucially relies on the assumption of random sorting of the municipalities on either side of the thresholds established by the law. Assuming random assignment, the treatment effect is locally identified exploiting the variation between the observations with a different number of council members that are in a neighborhood of the thresholds. The second approach takes advantage of the longitudinal structure of the dataset and therefore the possibility to observe both before and after those municipalities which changed the number of council members. In this case the treatment effect is identified exploiting the variation within the observations that modifies the legislature size. In this case the crucial assumption is slightly different, and amounts to exogenous change in the forcing covariate for the sample of treated municipalities. Moreover, given the possible presence of common time trends among the municipalities, it is necessary to use a good reference group to difference out such unobserved components.

1.3 Evaluation Approach and Results

In this section I present the two econometric approaches used to estimate the effects of the council size on the local government size. I start with the design exploiting the between variation, that compares groups' conditional means evaluated near the thresholds defined by the population based rule setting the number of council members in the Italian local governments. Then I present the results obtained with the econometric approach that uses the variation within the municipalities before and after the legislature size change induced by the law. I postpone the discussion of the results and the ... of robustness checks to the next section.

1.3.1 Empirical Approach - Regression Discontinuity Approach

In the first part of the empirical analysis, I pool together the observations over time of all the municipalities, then I identify the effect of the change in the council size on the current expenditures and on the revenues from taxes and tariffs for the Italian local governments. The identification obtains from the discontinuity between regression functions at the cutoff point, LP^c.

This approach is known in the treatment evaluation literature as Regression Discontinuity Design (RDD), see **??** and **??**. The method amounts to the estimation of the boundary points on the opposite sides of the cut-off, see **??**. The randomness of the sorting of the observations on either side of the cut-off is a crucial assumption for the identification of the effect.

In the following empirical analysis I employ a split (third-grade) polynomial approximation on either sides of the cut-off, using a symmetric intervals of the the forcing variable as performed in **??** and **??**. For each of the first three thresholds, the sample is restricted to those municipalities in the a interval around the cutoff, $LP_{it} \in [LP_k^c - bw_k, LP_k^c + bw_k]$, for $k \in \{1, 2, 3\}$ and $LP_k^c \in \{3001, 10001, 30001\}$. Where bw_k, the bandwidths, are selected using cross-validation, in the spirit of **??**. In the cross-validation I allows for maximum symmetric bandwidth equal to 1000, 2500 and 10000 respectively. In practice, the following model is estimated separately for each interval:

$$Y_{kit} = \gamma_0^k + \gamma_1^k \widetilde{LP}_{kit} + \gamma_2^k \widetilde{LP}_{kit}^2 + \gamma_3^k \widetilde{LP}_{kit}^3 + I_{kit}(\delta_0^k + \delta_1^k \widetilde{LP}_{kit} + \delta_2^k \widetilde{LP}_{kit}^2 + \delta_3^k \widetilde{LP}_{kit}^3) + \zeta_{kit} \tag{1.1}$$

where Y_{kit} denotes the outcome variable of interest, e.g. log per capita expenditures, I_{kit} is an indicator that the legal population is on or above the cutoff $LP_{kit} > LP_k^c$, while $\widetilde{LP}_{kit} = LP_{kit} - LP_k^c$ is the difference between the legal population of the observation i, at time t, with respect to the relative threshold LP_k^c. The treatment effect at the threshold is estimated by the coefficient δ_0^k, $k \in 1, 2, 3$. In this context RDD loses some statistical power given that the effect estimates are obtained running three different regressions, each including only the municipalities in the relevant interval. RDD exploit variation between observations to identify the parameter of interest. Another possible drawback is the lost in efficiency against approaches that uses information from all three discontinuities, given that the effect is linear in the number of council members.

Moreover, a further concern for the identification of the council size effect is due to the contemporaneous presence of other policy changes at the three population size cut-offs. For instance, at the $3,001$ threshold there is a 50% increase in the wage of mayors. For other legislative thresholds see Table 1 in **??**. One further assumption is needed in order to interpret the results of RDD approach as the effect of the policy under examination, namely that the effect of council size change is of first order, as well as that (locally) there are no strong interactions between policies.

1.3.2 Empirical Approach - Panel Model

In the second part of the analysis I exploit the longitudinal nature of the data, and try to identify the council size effect with panel data techniques. The dataset for the panel analysis is made up of around 300 observations (municipalities), considering the selection described in the previous section. Moreover, I split the data into two periods, the first extends from 2001 to 2006 and includes all the available periods following the official release of the 2001 census data. The remaining three years period, from 1998 to 2000, is employed to set up a falsification test to assess the validity of the identification strategy. The aim of the falsification test it to show that there is no spurious effect for those municipalities that will subsequently modify council size. In practice, the presence of a significant effect before the actual change in the council size will cast doubts on the nature of the identified effect.

Considering the variation in the census population as exogenous, the induced change in the council size of the municipalities around the thresholds can be considered as a random treatment assignment. Therefore, it is possible to exploit the information in the variation *within* the observations to estimate the desired effect. The econometric model is defined as

$$y_{it} = \alpha CouncilSize_{it-1} + x'_{it}\beta + \psi_t + \phi_i + \varepsilon_{it} \tag{1.2}$$

where the dependent variable is either the log of per capita current expenditures or the log of the current expenditures as a share of income. In the sample under inspection, the variable *CouncilSize* assumes values from 12, for those municipalities with legal population less than 3001, to (30) for legal population larger than 30000, as for all other legislature variables, the values are relative to the previous year. An exception is the dummy variable for the election year, that captures the possible political cycle influence on spending. The share of population below 15 and the share of population above 65 are included as time-varying control variables in all specifications, these are thought to be important determinant of municipalities spending, in particular for education, social, and health services. All estimated models include year fixed effects, ψ_t, and locality fixed effects, ϕ_i. The coefficient of interest is α that measures the impact of increasing the council size on the government size. Under the assumption $E(\varepsilon_{it}|CouncilSize_{it-1}, \phi_i, \psi_t) = 0$, it is possible to consistently estimate the causal effect using a fixed effect estimator, since $CouncilSize_{it-1}$ varies both between and within the observations.

Importantly the council size law can however, only induce a council-size change in combination with the election year subsequent to the census, since a local government's council size is based on its population size at a specific date, for the specificity of this paper, on the 21st of October 2001. This could imply that the population size of a locality after the census can cross the threshold without triggering a change in council size. Moreover, current year expenditures are constrained by provisional budgets approved by the end of the previous year, therefore the influence of a newly elected council will be from the year after the election.

Dynamic Panel

The outcome variables under examination can be characterized by relevant persistence along the time dimension, since our identification strategy could be . In order to

The procedure requires heteroskedasticity of the observation specific errors and strict exogeneity of the other regressors.

Falsification Test

To implement a falsification test, I repeat the analysis of the benchmark model in Table 1.7 using observations from the first three years of the dataset from 1998 to 2000, before the last census occurred. The estimation includes all the municipalities whose 2001 Census population is within the ±5% interval around the thresholds. Using the information in the data set is possible to assign a pseudo-treatment to those local government that will actually experience council size change in the first election after the census. These results are reported in Columns 2 and 4 of Table **??** where, to ease comparison of the results, the baseline results are reported again in Columns 1 and 3.

1.3.3 Results

In this section I present the results on the relationship between the number of council members and expenditures size obtained by the cross-sectional and longitudinal approaches outlined in the previous discussion.

Regression Discontinuity Approach

In this section I present the evidence provided by the regression discontinuity design on whether the number of council members affects the local government size. The analysis is limited to the three lowest thresholds in order to have some observations near the cut-off and on both sides. Table 1.6 shows that for the 3001 threshold both the estimated effect on expenditures and revenues is large and negative, 7.4 percentage and 28 percentage decrease respectively

Panel Model Approach

Longitudinal nature of the information Table 1.7 shows the results of parametric OLS estimates on a panel with fixed and time effects (all the displayed standard errors are obtained taking into consideration possible heteroskedasticity among the municipalities). Columns 1-3 show the results from using *log(per capita spending)* as a measure of government size, while Columns 4-6 show the results from using *log(spending as a share of income)*.

The first specification includes election year dummy, the share of population below 15 years of age and the share of population of the elderly (65+) as control variables. Subsequent specifications play more attention on the following factors: (i) economies of scale effects of population and (ii) the presence of region specific time varying effects correlated with both council size and current expenditures. I discuss each in turn below.

Columns 1 and 4 show the results from the baseline specification. For example, using *log(per capita spending)* as the measure of government size, the estimate is -0.0406 (Column 1). This means that when the council size increases with 4 members the government size *decreases* with 4.06 percent, the presence of an additional council member contributes to a reduction in the municipalities current expenditures by around 1 percent. The estimated effect is also statistically different from zero at the 1% level. This result does not change if we instead used *log(spending as a share of income)* as the measure of government size, the estimates of council-size effect displayed in Column 4 is -0.0374 (implying a contraction of 3.74 percent). This finding contrasts with the results from previous empirical studies, with the exception of a recent work of Petterson-Lidbom (2008). It is important to note that the panel estimates control for municipality specific features that are time invariant as well as common time trends, in addition I include in the baseline specification time varying factors that can influence political decisions, relaxing the conditional mean independence assumption. The election year dummy can (partially) capture political cycle movements in the government size, while the (log) share of population of the young and of the elderly can influence spending in education and in the social program sectors respectively. The estimate of the election year dummy coefficient is negative but not significantly different from zero, also the estimates for the demographic variables do not greatly influence the expenditures, possibly this last result is due to a low within-municipality variability in the period under analysis and their first order influence being captured by the inclusion of fixed effects.

Economies of scale and regional trends. To control for possible economies of scale, I include a second order polynomial in the log of (lagged) population, results are very similar to the previous ones as can be seen by comparing the estimate in Column 1 with the estimates in Column 2 for log(spending per capita), an the estimate in Column 4 with the estimate in Column 5 for log(spending as a share of income), especially in the first case the coefficient is almost unchanged and its statistical significance is at the 1% level. Another important determinant of local administrations expenditures can be regional trends in fiscal variables, therefore an important check is to assess the robustness of the results of the baseline specification when region specific time dummies are included as controls, the estimates are reported in Column 3 and 6. The F test for the joint statistical significance of these effects rejects the null hypothesis at the 1% level. The estimated coefficients are still close to the ones of the baseline model, nonetheless, now they are statistically significant only at 5% level and 10% respectively. Acknowledging this reduction in statistical significance, I do not consider of great concern this fact because the abatement of variation caused by the inclusion of a large number of control variables can reduce the power of these tests.

Robustness checks - controlling for observables. In this section I show the robustness of the previously presented results to inclusion in the regression of further controls possibly correlated with both the council size and the expenditures in local governments, thus tackling possible omitted variables bias. I start by adding observable characteristics of the education of mayors and of council members. For the firsts, the indicators if the mayor highest education achievement is a bachelor degree or the high school diploma, while for the seconds, the percentage of council members with a bachelor and those with a high school diploma. The results are reported in Table 1.8, Column 1 and 3 for the two different measures of local government size, in both cases the estimated coefficient is negative and is significant a the 5% level for the former while the p-value of the latter is a borderline 10.3% . Controlling for political cycle effects, including dummy variables for the year before an election and for the year after the election, does not affect greatly the estimates of the baseline model, maintaining also the level of significance for the council size coefficient respectively at the 1% and 5% level.

Sensitivity analysis - matching quality concerns. An issue of firs order importance in estimation of treatment effects is the quality of matching (or counterfactual), in other words treated individuals should be compared with a similar set of untreated observations. In this section I assess the robustness of the benchmark results to different choices of the set of municipalities included in the analysis.

First, I estimate the benchmark model including municipalities from different intervals around the 10,001 legal population threshold then including municipalities which experienced a similar legal population growth (reduction) but that did not cross the threshold. In Table 1.9 I present the results for the baseline model using the following legal population intervals: { [9,000-11,000]; [9,200-10,800]; [9,400-10,600]; [9,600-10,400]}. To start, have a look at the results when the dependent variable is the (log) per capita expenditures, these are reported in Panel A of the table, here it is possible to note that the estimated effect of legislature size keeps negative and strongly significant. Nonetheless the estimated absolute impact of the council size change is a decreasing function of the interval width, it ranges from -0.0514 to -0.0312, by using municipalities whose legal population is between 9,600 and 10,400 or between 9,000 and 11,000, respectively. A similar trend is observed in Panel B of the same table, where the dependent variable is the (log) expenditures as share of income, in this case there is also a visible deterioration in the level of statistical significance of the estimated coefficients. Now, Table 1.10 reports summary statistics for the population change of the municipalities included in the benchmark analysis, in particular the population change ranges from -879 to -482 and from 157 to 2712 for the municipalities whose council size decreased and increased respectively. In this robustness analysis I include four set of municipalities in the control group:

1. municipalities experiencing a population increase between 150 and 2000 and reached in the 2001 Census a population between 9500 and 10000;
2. municipalities experiencing a population increase between 150 and 2000 and that started with a population between 10001 and 10500 recorded in the 1991 Census;
3. municipalities experiencing a population reduction between 400 and 1000 and reached in the 2001 Census a population between 10001 and 10500;
4. municipalities experiencing a population reduction between 400 and 1000 and that started with a population between 9500 and 10000 recorded in the 1991 Census.

The results of this analysis are reported in Table 1.11 that induced the change in the number of legislators in the municipalities, i.e. focus our attention to the legal population growth between the last two official census. Using municipalities with similar population growth

Dynamic Panel Analysis. A potential problem with this specification is that the within-groups estimator is biased and inconsistent in the presence of a lagged dependent variable in a short panel ([15]). Thus, I show the estimates both from the within-groups and from the between estimator, that tend to be biased in the opposite direction, see [5], and discuss results based on the Corrected Least Squares Dummy Variable estimator implemented by [7] in Stata.

1.4 Disaggregated expenditure regressions

The empirical evidence seems to clearly support that larger councils are associated with lower spending. Since the reason for this counterintuitive result finds unsatisfactory answers in theoretical work, I think is beneficial to further explore the effects on the different components of the municipal expenditures. In this section I try to cast some light on the causes of the legislature size effect by estimating separate regressions for disaggregated spending.

I start studying the behavior expenditure categories defined by the services they ultimately provide then I consider expenditures chapters defined by the goods or services they buy. In the first analysis the focus is on (1) general administration, (2) education, (3) territory related interventions, (4) social and health, and (5) all other expenditures. The first four categories of spending account for about 76 percent of all spending for all Italian municipalities, while in the sample they account for about 80 percent. The aim of this analysis is to gain some insight on how the reduction in the aggregate spending due to the increased council size is performed in practice. Detailed description of the categories of spending are reported in Panel A of Table **??**, while summary statistics are reported in Table **??**, where I put in evidence the figures for the the all Italian municipalities, for the municipalities whose population size of 2001 Census was between 9,500 and 10,500, as well as for beginning of and end of period. We can note that the overall expenditure for all Italian municipalities has increased by 2.8 percent (in nominal terms) between the years 2003 and 2006 while the municipalities in the sample experience and increase of 4.7 percent. Among the categories under study the general administration chapter saw a similar rise in the population and in the sample, while the other three categories all experienced greater expansion (or less reduction in the case of the “territory”) in the sample.

In the latter part the focus is (1) personnel, (2) services, (3) transfers, and (4) all other expenditures. The first three categories of spending account for about 83.3 percent of all spending for all Italian municipalities, while in the sample they account for about 81.4 percent. Detailed description of these chapters of spending are reported in Panel A of Table 1.1, while summary statistics are reported in Panel B of Table 1.1, where I put in evidence the figures for the the all Italian municipalities, for the municipalities whose population size of 2001 Census was between 9,500 and 10,500, excluding the municipalities in the autonomous and special statute regions. Among the categories under study in the whole population the personnel expenditures saw a rise (in nominal terms) of 5 percent between the years 2003 and 2006, while in the restricted sample the growth was of 7.4 percent, in the same period, the expenditure for services kept almost unmodified for both the whole sample and the restricted one, while the expenditure for transfers experienced a sharp growth in the restricted sample 34.2 percent not paralleled in the whole population (8.4 percent). Finally the residual category experienced a steady decline from 2003, in both samples decreasing around 10 percent.

The results of the baseline regressions are reported in Table 1.2, as for the aggregated data municipality and year effects are included as well as election year dummy and the log of proportion of the population that is below 15 years and above 65 years of age. The most robust result is that the size of the council has a positive and statistically significant effect on services expenditures. The council size has also a negative effect on the personnel category that is significant at the 10 percent level. The effect on transfers is positive but not statistically different from zero, also the effect on the residual category is not significant.

Another way to assess the causes of the council size effect is to study the determinants of expenditures disaggregated according to the functions

Bibliography

[1] Daron Acemoglu. Constitutions, politics, and economics: A review essay on persson and tabellini's the economic effects of constitutions. *Journal of Economic Literature*, 43(4):1025–1048, December 2005.

[2] Reza Baqir. Districting and government overspending. *Journal of Political Economy*, 110(6):1318–1354, December 2002.

[3] D. Baron and J Ferejohn. Bargaining in legislatures. *American Political Science Review*, 83:1181–1206, 1989.

[4] Marco Battaglini and Stephen Coate. Inefficiency in legislative policy-making: A dynamic analysis. *American Economic Review*, 97:118–149, 2007.

[5] Stephen Bond. Dynamic panel data models: a guide to micro data methods and practice. *Portuguese Economic Journal*, 2002.

[6] John Charles Bradbury and W. Mark Crain. Legislative organization and government spending: cross-country evidence. *Journal of Public Economics*, 82(3):309–325, December 2001.

[7] Giovanni S.F. Bruno. Approximating the bias of the lsdv estimator for dynamic unbalanced panel data models. *Economics Letters*, 87(3):361–366, June 2005.

[8] Thomas W Gilligan and John G Matsusaka. Deviations from constituent interests: The role of legislative structure and political parties in the states. *Economic Inquiry*, 33(3):383–401, July 1995.

[9] Thomas W Gilligan and John G Matsusaka. Evidence from state and local governments in the first half of the 20th century. *National Tax Journal*, 54(1):57–82, March 2001.

[10] Jinyong Hahn, Petra Todd, and Wilbert Van der Klaauw. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–09, January 2001.

[11] David S. Lee. Randomized experiments from non-random selection in u.s. house elections. *Journal of Econometrics*, 142(2):675–697, February 2008.

Table 1.1: Summary Statistics

Panel A. Descriptions

Personnel	Wages of employee, extra-hours payments, welfare system payments, etc.
Services	Acquisitions of services for the operational management, e.g. refectory, training courses, missions
Transfers	Money transfers to families, other institutions as well as to municipal companies
All Other Expenditures	Includes use of third parties goods (e.g. rents, software licenses, etc.), interest payments, taxes, amortizations.

Panel B. Summary Statistics

Expenditure Categories – All Italian Municipalities

	Personnel		Services		Transfers		Others	
Year	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
2003	1,817,721	17,000,000	2,266,703	21,800,000	564,823	5,015,758	992,721	8,653,057
2004	1,829,713	17,000,000	2,276,303	21,400,000	590,868	5,109,675	953,623	8,642,889
2005	1,875,408	18,100,000	2,331,450	20,800,000	622,108	5,289,799	958,870	8,842,699
2006	1,909,106	17,800,000	2,286,699	20,400,000	612,189	4,183,939	905,585	8,650,939

Expenditure Categories – Analysis Sample

	Personnel		Services		Transfers		Others	
Year	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
2003	1,984,521	745,604	2,615,412	1,149,566	575,734	340,983	1,179,802	752,758
2004	2,023,786	739,199	2,624,080	1,128,434	632,796	376,521	1,131,353	586,080
2005	2,079,409	788,730	2,766,393	1,157,875	688,333	396,127	1,125,175	571,317
2006	2,131,667	786,775	2,658,228	1,357,026	772,695	481,669	1,045,343	531,517

Table 1.2: Regressions of Expediture Interventions

Variable	Personnel	Services	Transfers	All Other Interventions
$change10k_{i,t}$	-0.0145*	-0.0841***	0.0917	-0.0337
	(0.0087)	(0.0226)	(0.0658)	(0.0346)
Election Year	-0.0061	0.00296	-0.0387	0.0141
	(0.0067)	(0.0147)	(0.0373)	(0.0202)
$\log(populationunder15)_{t-1}$	0.00666	0.0567	-0.117	0.0134
	(0.0166)	(0.0712)	(0.282)	(0.0658)
$\log(populationover65)_{t-1}$	0.0332**	0.0448	-0.226	0.0371
	(0.0146)	(0.0775)	(0.323)	(0.0536)
N. Observations	423	423	423	423
adj. R-sq	0.094	0.145	0.099	0.241
N. Municipalities	106	106	106	106

[12] David S. Lee and David Card. Regression discontinuity inference with specification error. *Journal of Econometrics*, 142(2):655–674, February 2008.

[13] Alessandro Lizzeri and Nicola Persico. The provision of public goods under alternative electoral incentives. *American Economic Review*, 91(1):225–239, March 2001.

[14] Gian Maria Milesi-Ferretti, Roberto Perotti, and Massimo Rostagno. Electoral systems and public spending. *The Quarterly Journal of Economics*, 117(2):609–657, May 2002.

[15] Stephen J Nickell. Biases in dynamic models with fixed effects. *Econometrica*, 49(6):1417–26, November 1981.

[16] Roberto Perotti and Yianos Kontopoulos. Fragmented fiscal policy. *Journal of Public Economics*, 86(2):191–222, November 2002.

[17] Torsten Persson and Guido Tabellini. The size and scope of government:: Comparative politics with rational politicians. *European Economic Review*, 43(4-6):699–735, April 1999.

[18] Torsten Persson and Guido Tabellini. *Political Economics. Explaining Economic Policy.* MIT Press, 2000.

[19] Torsten Persson and Guido Tabellini. *The Economic Effects of Constitutions.* MIT Press, 2003.

Table 1.3: Council-size law: Italian local government

Population size	Number of council members
0-3.000	13
3.001-10.000	17
10.001-30.000	21
30.001-100.000	31
100.001-250.000	41
250.001-500.000	47
500.001-1.000.000	51
1.000.001-	61

[20] Torsten Persson and Guido Tabellini. Constitutional rules and fiscal policy outcomes. *American Economic Review*, 94(1):25–45, March 2004.

[21] David M. Primo and James M. Jr Snyder. Distributive politics and the law of 1/n. *Journal of Politics*, 70(2):477–486, April 2008.

[22] Barry R Weingast, Kenneth A Shepsle, and Christopher Johnsen. The political economy of benefits and costs: A neoclassical approach to distributive politics. *Journal of Political Economy*, 89(4):642–64, August 1981.

Table 1.4: Number of localities within a ±5% interval from thresholds in 2001 and law induced changes in council size (within 2006)

Threshold	Number of localities below the threshold	Number of localities above the threshold	Number of localities from above to below the threshold	from
3,001	82	135	23	
10,001	40	69	5	
30,001	10	18	1	
100,001	4	2	1	
250,001	0	1	0	
500,001	0	0	0	
1,000,001	0	1	0	

Table 1.5: Total law induced changes in council size

Threshold	Number of localities from above to below the threshold	Number of localities from below to above the threshold
3,001	84	131
10,001	12	79
30,001	6	12
100,001	4	0
250,001	0	1
500,001	0	0
1,000,001	0	0

Table 1.6: Results from regression discontinuity - council size effect for Italian municipalities

	log(spending per capita)			log(revenues per capita)		
Threshold	3,001	10,001	30,001	3,001	10,001	30,001
Effect	-0.0736	0.130	0.252	-0.280*	0.0257	0.298
	(0.107)	(0.0867)	(0.170)	(0.147)	(0.133)	(0.268)
bw	300	1100	4500	450	1500	4500
Observations	3651	2697	756	5272	2627	718

Notes: the effects are relative to an increase of 4 council members for the 3,001 and 10,001 thresholds, while it is relative to an increase of 10 council members for the 30,001 threshold. Estimation method: split polynomial approximation on an interval around the threshold, see Equation (1.1). *bw*: optimal symmetric bandwidths are chosen with cross-validation methods. Clustered standard errors at the local government level in parentheses.

Figure 1.1: Local government size around the threshold

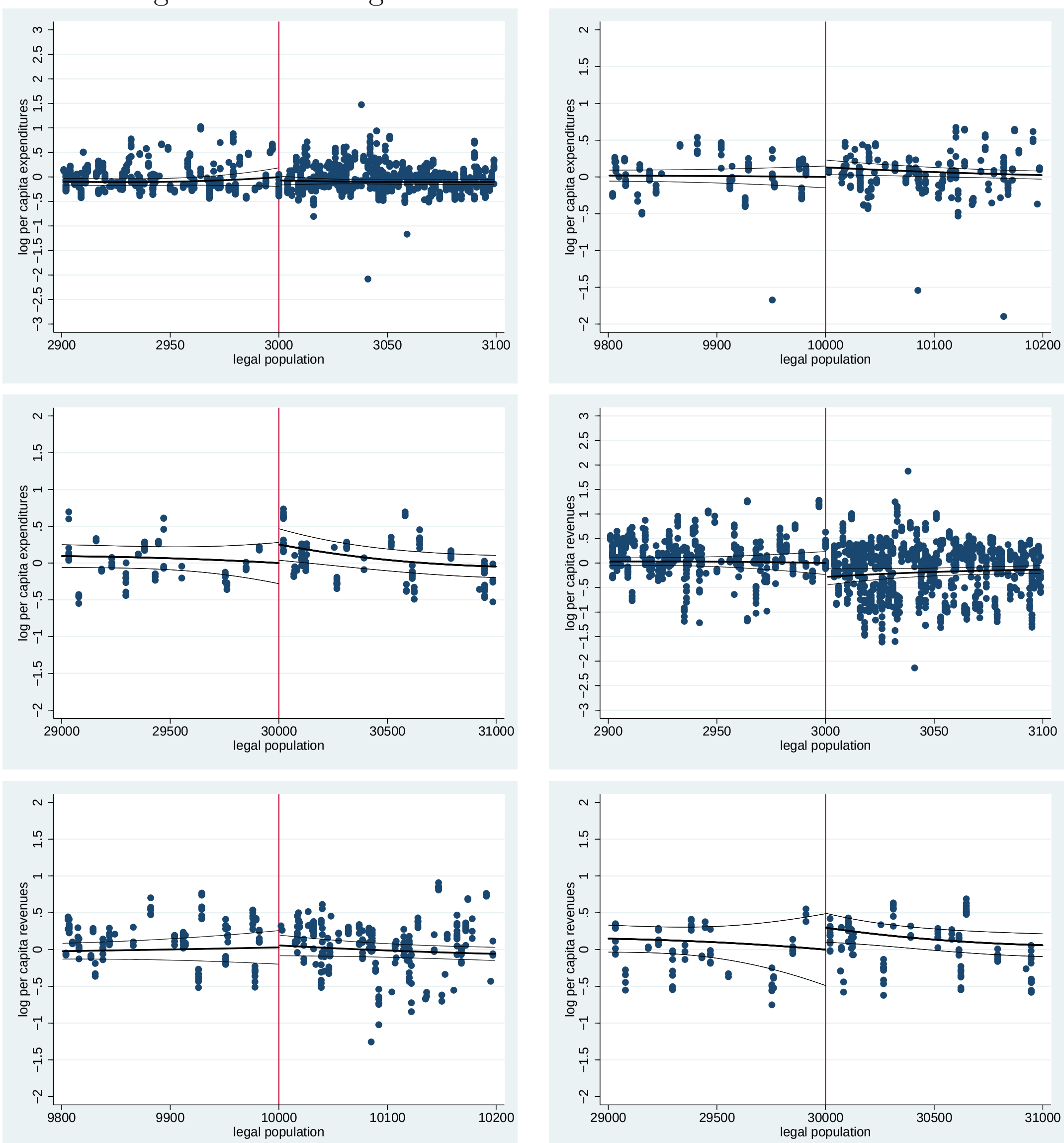


Table 1.7: Results from panel data approach - council size effect for Italian municipalities

	log(current expenditures per capita)			log(taxtariffs revenues per capita)		
	(1)	(2)	(3)	(4)	(5)	(6)
$CouncSize_{i,t-1}$	-0.00595***	-0.00304	-0.00441*	-0.00285	-0.00109	-0.00173
	(0.00205)	(0.00195)	(0.00226)	(0.00402)	(0.00392)	(0.00346)
$electionyear_{i,t}$	-0.00849	-0.00853	-0.00199	-0.00682	-0.00745	-0.00202
	(0.00710)	(0.00714)	(0.00815)	(0.0108)	(0.0109)	(0.0105)
$\log(populationunder15)_{t-1}$	-0.0464	0.00976	0.0280	-0.0894	-0.0555	0.00283
	(0.0441)	(0.0374)	(0.0397)	(0.0746)	(0.0698)	(0.0689)
$\log(populationover65)_{t-1}$	-0.0636*	-0.0562	-0.0418	-0.0882	-0.0780	-0.0647
	(0.0376)	(0.0376)	(0.0405)	(0.0551)	(0.0508)	(0.0534)
$\log(population_{i,t-1})$		-1.199***			-0.921***	
		(0.232)			(0.264)	
$\log(population)^2_{i,t-1}$		-0.308***			-0.305***	
		(0.104)			(0.109)	
Region × Year effect [prob ≤ F]			[0.00]			[0.00]
Observations	1928	1928	1917	1952	1952	1941
Number of municipalities	325	325	325	329	329	329

Table 1.8: Results from panel data approach - council size effect for Italian municipalities - robustness checks

	log(curr. expenditures p.c.)		log(taxtariffs revenues p.c.)	
	(1)	(2)	(3)	(4)
$CouncSize_{i,t-1}$	-0.00465**	-0.00575***	-0.00196	-0.00341
	(0.00226)	(0.00210)	(0.00388)	(0.00408)
$electionyear_{i,t}$	-0.00595	-0.00963	-0.00733	-0.00724
	(0.00758)	(0.00712)	(0.0111)	(0.0105)
$\log(populationunder15)_{t-1}$	-0.0323	-0.0420	-0.0754	-0.0887
	(0.0425)	(0.0443)	(0.0749)	(0.0745)
$\log(populationover65)_{t-1}$	-0.0486	-0.0610	-0.0814	-0.0892
	(0.0372)	(0.0379)	(0.0548)	(0.0550)
Mayor bachelor graduate$_{i,t-1}$	0.0199		-0.0294	
	(0.0162)		(0.0189)	
Mayor high school$_{i,t-1}$	0.0251		-0.0252	
	(0.0191)		(0.0197)	
(% council members ba. grad.)$_{i,t-1}$	0.0347		0.0767	
	(0.0422)		(0.0506)	
(% council members h. school)$_{i,t-1}$	0.0489		0.106**	
	(0.0312)		(0.0517)	
$pre-electionyear_{i,t}$		-0.0108		-0.0425**
		(0.0155)		(0.0178)
$post-electionyear_{i,t}$		0.00205		0.0445*
		(0.0173)		(0.0267)
Observations	1673	1928	1856	1952
Number of municipalities	282	325	316	329

Table 1.9: Results from panel data approach - council size effect for Italian municipalities - legal population interval sensitivity

	log(current expenditures per capita)			log(taxtariffs revenues per capita)		
	(1)	(2)	(3)	(4)	(5)	(6)
	[2,700-3,300] ∪ [9,000-11,000] ∪ [27,000-33,000]	[2,800-3,200] ∪ [9,250-10,750] ∪ [28,000-32,000]	[2,900-3,100] ∪ [9,750-10,250] ∪ [29,000-31,000]	[2,700-3,300] ∪ [9,000-11,000] ∪ [27,000-33,000]	[2,800-3,200] ∪ [9,250-10,750] ∪ [28,000-32,000]	[2,900-3,100] ∪ [9,750-10,250] ∪ [29,000-31,000]
$CouncSize_{i,t-1}$	-0.00503***	-0.00387**	-0.00453*	-0.00312	-0.00333	-0.00184
	(0.00189)	(0.00193)	(0.00274)	(0.00288)	(0.00337)	(0.00553)
$electionyear_{i,t}$	-0.00869*	-0.00703	-0.00543	-0.00601	-0.00799	0.00715
	(0.00451)	(0.00537)	(0.00990)	(0.00694)	(0.00852)	(0.0138)
$\log(populationunder15)_{t-1}$	-0.0983*	-0.0545	-0.00209	-0.182**	-0.126*	-0.0542
	(0.0509)	(0.0452)	(0.0398)	(0.0806)	(0.0756)	(0.0770)
$\log(populationover65)_{t-1}$	-0.0602*	-0.0489	-0.0139	-0.0859	-0.0726	-0.0687
	(0.0350)	(0.0353)	(0.0412)	(0.0578)	(0.0558)	(0.0657)
Observations	3521	2599	1246	3557	2634	1262
Number of municipalities	595	438	210	601	444	213

Table 1.10: Summary statistics - legal population change for municipalities in benchmark analysis

	log(Municipalities around 3,001 threshold - 2001 Census)				
Δpopulation	Obs.	Mean	Std. Dev.	Min	Max
no change in council size	57	239.6	302.9	-341	1034
reduction in council size	14	-158.3	72.8	-341	-63
increased council size	43	369.2	225.1	96	1034
	log(Municipalities around 10,001 threshold - 2001 Census)				
Δpopulation	Obs.	Mean	Std. Dev.	Min	Max
no change in council size	25	827.4	698.7	-562	2712
reduction in council size	2	-522	56.5	-562	-482
increased council size	23	944.7	593.7	157	2712
	log(Municipalities around 30,001 threshold - 2001 Census)				
Δpopulation	Obs.	Mean	Std. Dev.	Min	Max
no change in council size	5	1676.2	1569.4	-866	3406
reduction in council size	1	-866	.	-866	-866
increased council size	4	2311.7	768.8	1621	340

Table 1.11: Results from panel data approach - council size effect for Italian municipalities - population change sensitivity

	log(current expenditures per capita)			log(taxtariffs revenues per		
$CouncSize_{i,t-1}$	-0.00288			-0.000211		
	(0.00213)			(0.00390)		
$change3k_{i,t}$		0.00715			0.0251	
		(0.0111)			(0.0195)	
$change10k_{i,t}$		-0.0606***			-0.0159	
		(0.0218)			(0.0424)	
$change30k_{i,t}$		-0.0864***			-0.0290	
		(0.0233)			(0.0785)	
$changeMinus_{i,t}$			-0.0116***			-0.
			(0.00343)			(0.
$changePlus_{i,t}$			-0.00218			0.
			(0.00273)			(0.
$electionyear_{i,t}$	-0.00571	-0.00653	-0.00576	-0.00940	-0.00172	-0.0
	(0.00483)	(0.00444)	(0.00445)	(0.00753)	(0.00710)	(0.
$\log(populationunder15)_{t-1}$	-0.292***	-0.288***	-0.295***	-0.314**	-0.322**	-0.
	(0.0887)	(0.0875)	(0.0876)	(0.134)	(0.133)	(0
$\log(populationover65)_{t-1}$	-0.128**	-0.132**	-0.136***	-0.149*	-0.157**	-0.
	(0.0518)	(0.0518)	(0.0516)	(0.0766)	(0.0757)	(0
Observations	3247	3309	3309	3298	3365	3
Number of municipalities	556	556	556	563	563	

Table 1.12: Results from panel data approach - council size effect for Italian municipalities - population change sensitivity

	log(current expenditures per capita)			log(taxtariffs revenues per capita)		
$LaggedDep.Variable$	0.157	0.879***	0.341***	0.0785	0.860***	0.320
	(0.123)	(0.0442)	(0.045)	(0.0691)	(0.0297)	(0.0405)
$CouncSize_{i,t-1}$	-0.00533**	-0.000257	-0.0082**	-0.00307	-0.00135*	-0.00363
	(0.00208)	(0.000798)	(0.0033)	(0.00389)	(0.000752)	(0.0041)
$electionyear_{i,t}$	-0.0101	-0.00894	-0.0026	-0.0104	-0.0243**	-0.0149
	(0.00731)	(0.00789)	(0.0089)	(0.0111)	(0.0117)	(0.0155)
$\log(populationunder15)_{t-1}$	-0.0444	-0.0394	-0.001	-0.0945	-0.265***	-0.112
	(0.0403)	(0.0337)	(0.046)	(0.0720)	(0.0667)	(0.107)
$\log(populationover65)_{t-1}$	-0.0580*	0.0495***	-0.021	-0.0841	-0.124***	-0.0965
	(0.0350)	(0.0150)	(0.064)	(0.0525)	(0.0374)	(0.099)
Observations	1866	1866	1866	1902	1902	1902
Number of municipalities	313	313	313	319	319	319

Chapter 2

Military Expenditure,threats and political regime

2.1 Introduction

This work studies the impact of military expenditure on political regime. The importance of the issue rests in the crucial role played by the army in characterizing the formation and survival of political regimes and, more broadly, of nations. On one hand, high expenditure in military security can secure countries against the attack of external enemies, benefiting all the citizens. On the other hand, the army can be used by autocratic rulers to suppress masses and prevent/deter democratic transitions. Analysis of the empirical regularities on the relationship between political regimes and military expenditures shows that authoritarian regimes tend to have larger military budgets. While external threats from other states have a negative influence on the level of democratic institutions in a country. Last but not least this work relates to the literature in international relations that study the empirical regularity of the *democratic peace*. The political science literature has paid more attention on the issue, but there are noteworthy recent works by economists. Among the former there are [4], [6], [9, 10], [3], and [5], for the latter [8], [2] and [7]. While [4], [6], [9, 10] arrive at the conclusion that militarization have a positive influence on economic growth and social development. [3] performs a comparative politics exercise on Latin American countries, he shows that, in the region, militarization has had negative consequences on democracy. [5] explore the relationship between democratization and international conditions. Their analysis concedes that external threats from other states tend to decrease democracy. [8] present model of the extension of suffrage in early 19th century in Europe, as a consequence of the military threat posed by Napoleon's *grand armè*. [7] estimate the influence of bilateral and multilateral trade on the probability of conflicts. They employ a similar approach on modeling the probability of interstate conflict. [2] study non-linearities and omitted variable biases in the relationship between economic growth, militarization and external threats.

2.2 Empirical Analysis

The starting conjecture of the analysis is the following: The impact of military expenditure on political regime is a non-linear function of the effective militarized threat posed by foreign countries and other external forces. military expenditure without threats would reduce the democratic level of institutions, while military expenditure in the presence of sufficiently large threats favors more democratic regimes.

The empirical analysis is performed in two steps. First of all I construct a measure of external military threat from the likelihood that an individual state will engage in a violent dispute with any other state. Then panel data techniques are employed to explore the relationship between the democracy score of a country with its military expenditure, taking into account the level of external threat.

2.2.1 Data description

The data come from various sources and span the period between 1960 and 2000, with varying degree of coverage. For the first part of the analysis the main dependent variable denoting the external military threat come from the Armed Conflict Dataset of the International Peace Research Institute (PRIO) and the Uppsala University. In particular the dyadic dataset of the Uppsala Armed Conflict projects is used. The dyadic dataset contains one line for each year for each pair of countries, a dyad, that coexist in a given year. I present results from two samples, based on data availability, both samples span from 1960 to 2000, the number of dyads in the analysis varies as follows.

Sample	Small	Extended
Year	# dyads	# dyads
1960	ca. 1,800	ca. 6,000
1970	ca. 3,700	ca. 6,500
1980	ca. 5,000	ca. 7,000
1990	ca. 5,600	ca. 7,000
2000	ca. 7,900	ca. 8,000

The main variables, for our investigation, reported by the Armed Conflict project, mark dyads composed by opposing countries in interstate or internationalised internal conflicts. As reported in the dataset codebook, an armed conflict is defined as "...a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths" The variable *conflict* is coded "1" for a dyad-year if either of the variables denoting the dyad as being on opposing side in a interstate or in an internationalised internal conflict is positive, *conflict* is coded "0" otherwise.

As the measure defining the political regime, I employ a widely used "continuous" indicator from the POLITY IV project *www.cidcm.umd.edu/inscr/polity*, the variable *polity2* reported by the POLITY IV dataset is a score for democracy, it takes values from +10 *stronglydemocratic* to -10 *stronglyautocratic*.

Using the the percentage of the national GDP spent on defense I try to operationalize the level of militarization of a country. Where, as defined by [3] "(m)ilitarization(. . .) is the expansion or relative size of some integral part, scope, or mission of the armed forces and may be observed in the size of the budget, the number of soldiers, and the training, equipping, war-readiness, and institutionalization of the armed forces". The data on the level of military expenditures are taken by the Correlates of War (COW) project, that makes available (cow2.la.psu.edu) a very large array of datasets concerning armed conflicts but also country characteristics over the last century, data on GDP is from Penn World Table version 6.1, both series are reported at constant prices before computing the following variable:

$$milexshare_{it} = \frac{militaryexpenditure_{it}}{GDP_{it}} \tag{2.1}$$

Data on income per capita growth rates are taken from Maddison (2000), population data are taken from Maddison (2000) for the smaller sample and from the most recent PWT v. 6.2 for the extended sample. Bilateral measures used in the estimation of the probability of conflict comes from various sources, while detailed alliance memberships comes from the relative dataset in COW v 3.03.

that makes available (at http://cow2.la.psu.edu/) a very large array of datasets concerning armed conflicts but also country characteristics over the last century.

2.2.2 Econometric specification

In constructing the external threat variable I estimate a non-linear "gravity model" (see Frankel, 1997 and Rose, 2000, 2004) for dyadic-year data, where the dependent variable is *conflict* (as defined above). In the main specification, I separate the influence on the probability of conflicts between "gravity" variables and "balance of power" ones. The model of interest is the following:

$$conflict_{ijt} = \{ \begin{matrix} 1 & if & conflict^*_{ijt} > 0 \\ 0 & & otherwise \end{matrix} \tag{2.2}$$

The latent variable $conflict^*_{ijt}$ is modeled according to the following linear models

$$conflict^*_{ijt} = \alpha + \beta' gravity_{ijt} + \gamma' balance_{ijt} + \delta x_{ijt} + \varepsilon_{ijt} \tag{2.3}$$

Where ε_{ijt} has the usual logistic distribution. Among the *gravity* variables there are the sum of the area of the two countries (in logarithm), the sum of the population (one period lag, in logarithm), the logarithm of bilateral distance, the number of countries in the dyad that are landlocked and that are island, dummy variables indicating respectively whether the two countries have a common language, whether the two have been colonized by the same third country and both for common membership in a currency union as well as if the form a currency union de facto. *Balance* of power variables are the absolute value of the logarithm of the two areas ration, and the same function for the lagged populations. In two robustness check I first add a dummy which take the value of one if the two countries are reported as belonging to at least one common alliance in a given year, according to the COW database on Alliance (V 3.03), then I add two variables taking into account the number of years that each country of the dyad spent without recorded war with any country, the two variables included are the minimum and the maximum of the two values respectively.

The results from the previous analysis allow to compute the probability of conflict between two countries in a given year.

Table 2.1: Non-linear gravity model, logit estimation results

	Dependent variable conflict (0, 1)		
Specification	Simple	Allies	Peace years
Sum ln areas	0.18	0.23	0.14
	(0.072)	(0.075)	(0.074)
BoP ln areas	0.25	0.287	0.214
	(0.098)	(0.099)	(0.099)
Sum ln population (t-1)	0.32	0.28	0.22
	(0.058)	(0.057)	(0.059)
BoP ln population (t-1)	-0.38	-0.43	-0.46
	(0.12)	(0.117)	(0.119)
Border	2.02	2.12	2.06
	(0.40)	(0.406)	(0.408)
Common language	1.51	1.68	0.99
	(0.25)	(0.250)	(0.260)
Common colony	-0.49	-0.52	-0.18
	(0.37)	(0.386)	(0.401)
Ln distance	-0.91	-1.01	-0.97
	(0.226)	(0.230)	(0.230)
Allies		-0.798	
		(0.249)	
Peace years min			-2.47
			(0.228)
Peace years max			0.015
			(0.011)
Constant	-16.68	-15.87	-9.83
	(1.688)	(1.681)	(1.796)
Observations	198689	198689	197011

Standard errors in parentheses

$$\widehat{\Pr}\left(conflict_{ijt} = 1\right) = \Lambda\left(\widehat{\alpha} + \widehat{\beta}' gravity_{ijt} + \widehat{\gamma}' gravity_{ijt} + \widehat{\delta}' x_{ijt}\right) \tag{2.4}$$

The estimated *threat* is the sum of the probability of conflict for each country in a given year. The sum is over all the other countries in the world.

$$threat_{it} = \sum_{j \neq i} \widehat{\Pr}\left(conflict_{ijt} = 1\right) \tag{2.5}$$

The estimated *threat* measure is used in a panel model together with military expenditure to assess the impact on political regime, the conjecture is that the relationship is a non-linear function. In particular, military expenditure without threat would hamper democracy, but military expenditure in presence of sufficiently large threats is positively related to democracy. Denoting democracy by *dem*, military expenditure by *milex* and country's effective threat by *threat*, the above conjecture can be expressed as

$$\frac{\partial dem}{\partial milex} = a_1 + a_2 threat, \quad a_1 < 0 \text{ and } \quad a_2 > 0 \tag{2.6}$$

$$\frac{\partial dem}{\partial threat} = b_1 + b_2 milex, \quad b_1 < 0 \text{ and } \quad b_2 > 0 \tag{2.7}$$

The model estimated is

$$dem_{it} = a_1 milex_{it-1} + b_1 threat_{it-1} + a_2 (milex_{it-1})(threat_{it-1}) + \gamma' W_{it} + \nu_{it} \tag{2.8}$$

where W_{it} is a set of control variables and fixed effects. In all specifications I include country fixed effects that controls for unobserved country specific features, e.g. culture and religion (at least those are constant over time). The aim is to exploit within country variation. Moreover, I include year fixed effects interacted with continents dummies in order to take care of continent-specific shocks on the dependent variable. Identifying assumption needs that lagged military expenditure and lagged threat measure are uncorrelated with country-specific and time varying shocks. Possible failure is whether errors in the model are auto-correlated, say, of first order. Jointly with the fact that shifts in political regimes contemporaneously affect the military expenditure. Indeed, we need to perform a full battery of mis-specification tests.

The following regressions employ measures of threat obtained on a more balanced and complete sample in the first step. This to limit the problems of selection in the observations. Countries for which data are not available in the "small"

Table 2.2: Panel estimates - small sample

Threat specification		Simple	
		dep. var. *polity2*	
polity2 (t-1)		0.87	0.92
		(0.008)	(0.016)
		(0.014)	(0.025)
milex (t-1)	-0.44	-0.09	-0.08
	(0.065)	(0.032)	(0.032)
	(0.288)	(0.047)	(0.041)
threat (t-1)	-17.64	-4.66	-4.76
	(5.69)	(2.83)	(2.83)
	(15.76)	(3.12)	(3.25)
Interaction (MilEx * CflRisk) (t-1)	2.78	1.18	1.16
	(1.268)	(0.633)	(0.633)
	(3.737)	(0.715)	(0.636)
Contintents * Polity2 (t-1)			[YES]
			[0.00]
			[0.00]
Observations	3986	3984	3984
Number of countries	137	137	137
Adjusted R-squared	0.24	0.81	0.81

Regressions include country fixed effects, year fixed effects and continents $*$ year fixed effects. In parentheses, above: standard errors in parentheses; below: robust standard errors in parentheses;

Table 2.3: Panel estimates - small sample - cont.

Threat specification		Allies	
		dep. var. *polity2*	
polity2 (t-1)		0.87	0.92
		(0.008)	(0.016)
milex (t-1)	-0.46	-0.088	-0.08
	(0.062)	(0.031)	(0.031)
threat (t-1)	-17.64	-4.66	-4.76
	(5.69)	(2.83)	(2.83)
	(15.76)	(3.12)	(3.25)
Interaction (MilEx * CflRisk) (t-1)	3.08	1.10	1.09
	(1.127)	(0.566)	(0.566)
Contintents * Polity2 (t-1)			[YES]
			[0.00]
Observations	3986	3984	3984
Number of countries	137	137	137
Adjusted R-squared	0.24	0.81	0.81

Regressions include country fixed effects, year fixed effects and continents $*$ year fixed effects. Standard errors in parentheses;

Table 2.4: Panel estimates - small sample - cont.

Threat specification	Peace years dep. var. *polity2*		
polity2 (t-1)		0.87	0.92
		(0.008)	(0.016)
milex (t-1)	-0.46	-0.088	-0.08
	(0.062)	(0.031)	(0.031)
threat (t-1)	-17.64	-4.66	-4.76
	(5.69)	(2.83)	(2.83)
	(15.76)	(3.12)	(3.25)
Interaction (MilEx * CflRisk) (t-1)	3.08	1.10	1.09
	(1.127)	(0.566)	(0.566)
Contintents * Polity2 (t-1)			[YES]
			[0.00]
Observations	3986	3984	3984
Number of countries	137	137	137
Adjusted R-squared	0.24	0.81	0.81

Regressions include country fixed effects, year fixed effects and continents $*$ year fixed effects. Standard errors in parentheses;

Table 2.5: Panel estimates - extended sample

Threat var. specification	Simple	Simple 3yrs	Simple 5yrs
		dep. var. *polity2*	
polity2 (t-1)	1.06	1.06	1.06
	(0.065)	(0.065)	(0.0649)
polity2 (t-2)	-0.16	-0.16	-0.16
	(0.045)	(0.045)	(0.045)
milex (t-1)	-0.192	-0.19	-0.188
	(0.077)	(0.074)	(0.074)
threat (t-1)	0.305	0.32	0.29
	(0.35)	(0.31)	(0.31)
Interaction (MilEx * CflRisk) (t-1)	0.19	0.115	0.097
	(0.217)	(0.061)	(0.063)
Contintents * Polity2 (t-1)	[YES]	[YES]	[YES]
	[0.00]	[0.00]	[0.00]
Observations	4002	4002	4002
Number of countries	131	131	131
Adjusted R-squared	0.82	0.82	0.82

Bootstrapped robust standard errors in parentheses

Table 2.6: Panel estimates - richer dynamics and economic crises - extended sample

Threat var. specification	Simple		Simple 3yrs dep. var. *polity2*		Simple 5yrs	
polity2 (t-1)	1.02	1.00	1.02	1.00	1.02	1.00
	(0.04)	(0.049)	(0.039)	(0.048)	(0.039)	(0.048)
Polity2 (t-2)	-0.11	-0.11	-0.11	-0.11	-0.11	-0.11
	(0.025)	(0.029)	(0.025)	(0.029)	(0.0257)	(0.028)
Economic crises (t-1)	0.38	0.33	0.38	0.34	0.38	0.33
Threshold: -4 for 2 years	(0.17)	(0.208)	(0.17)	(0.21)	(0.17)	(0.20)
Military expenditure - GDP share(t-1)	-0.17	-0.145	-0.18	-0.15	-0.17	-0.15
	(0.08)	(0.09)	(0.077)	(0.09)	(0.07)	(0.089)
Conflict risk (t-1)	0.28	0.75	0.26	0.70	0.23	0.68
	(0.35)	(0.42)	(0.32)	(0.37)	(0.32)	(0.376)
Interaction (MilEx * CflRisk) (t-1)	0.12	0.085	0.10	0.085	0.086	0.079
	(0.22)	(0.27)	(0.06)	(0.07)	(0.063)	(0.078)
Sample if war==0	No	Yes	No	Yes	No	Yes
Observations	3791	3343	3791	3343	3791	3343
Number of countries	125	125	125	125	125	125
Adjusted R-squared	0.82	0.83	0.82	0.83	0.82	0.83

Robust standard errors in parentheses

sample may have poor institutions. Results for two specifications of the logit regressions are reported, simple and controlling for allied pairs of countries. To each specification we added time fixed effects (3 years and 5 years dummies).

The following regressions control for economic crises. The analysis in [1] provides some evidence in favours of those theories that "emphasize economic crises as events destabilising (...) regimes, and leading to regime transitions". We define an economic crisis as a sudden, sharp and sustained decline in growth relative to two years ago. More specifically, there is an economic crisis at time if the two-years average growth rate of GDP per capita is less than a certain threshold. I choose the threshold as 4 percent (Tables 7 and 9). We report results also for the threshold at 3 percent for three-years averages (Tables 8 and 10). Each specification is tested on both the complete sample and the sample restricted to years of peace.

2.3 Conclusions

This work provide some evidence that the basic conjecture regarding the non-linear relation between military spending and political regimes in a panel of countries. We control for internal destabilizing factors as economic crises and still we find some supportive evidence. When we exclude in the estimation sample the observation that are at war, still we find positive coefficient for the interaction term, even if not always significant, while the influence of economic crises washes out. The evidence is provided by within country variation. The empirical investigation evidence that both military expenditure is harmful to democratic development, while the effect of external threat has to be further explored. Moreover we find support for the conjecture that countries that increase military expenditure in presence of high external threat tend to improve their democratic institutions. We obtain these results by estimating a measure of external threat as the probability of conflict between the country and any other state in the world. Then we use a fixed effect panel estimator to control for time invariant country-specific features

Bibliography

[1] Daron Acemoglu, Simon Johnson, James A. Robinson, and Pierre Yared. Income and democracy. *American Economic Review*, 98(3):808–42, June 2008.

[2] Joshua Aizenman and Reuven Glick. Military expenditure, threats, and growth. *Journal of International Trade & Economic Development*, 15(2):129–155, June 2006.

[3] Kirk Bowman. *Militarization, Democracy, and Development: the perils of praetorianism in Latin America*. University Park, PA: Penn State University Press, 2002.

[4] Bullock and Firebaugh. Guns and butter? the effect of military, economic and social development in the third world. *Journal of Political and Military Sociology*, 1990.

[5] Michael Colaresi and William Thompson. The democracy-economic development nexus: Does the outside world matter? *Comparative Political Studies*, 2003.

[6] EL Kick, R Nasser, B Davis, and L Bean. Militarization and infant mortality in the third world. *Journal of Political and Military Sociology*, 1990.

[7] Philippe Martin, Thierry Mayer, and Mathias Thoenig. Make trade not war? *Review of Economic Studies*, 75(3):865–900, 07 2008.

[8] Davide Ticchi and Andrea Vindigni. War and endogenous democracy. IZA Discussion Papers 3397, Institute for the Study of Labor (IZA), March 2008.

[9] Erich Weede. Rent seeking, military participation, and economic performance in ldcs. *Journal of Conflict Resolution*, 1986.

[10] Erich Weede. The impact of military participation on economic growth and income inequality: some new evidence. *Journal of political and military sociology*, 1993.

Part II

Macroeconometrics

Chapter 3

VAR Based Model Evaluation in Macroeconometrics : a Comparison

3.1 Introduction

This paper compares two approaches for Vector AutoRegression (VAR) based model evaluation developed in the Bayesian analysis of macroeconomic time series. We can informally call these two approaches as the limited information and the full information Bayesian model evaluation. The first approach starts from a statistical model that well describes the data, then it derives the theoretical restrictions imposed on the coefficients of the statistical model by the economic theory. In the next step, a prior distribution is imposed on the coefficients of the statistical model. The priors are conceived so to allow to control for the *tightness* of the theoretical implications. As a term of comparison, an alternative theoretical model is developed. The restrictions of this second model bear no economic content, but justified by its ability to improve statistical model fit. The two theories are compared in terms of Bayes Factors (see [6]) for different values of the tightness parameter. In practice, the tightness parameter value that delivers the highest Bayes factor is compared to the analogous tightness imposed on the alternative theory, if the ratio of the two is sufficiently small (and the relative Bayes Factor is enough high), the evidence is in favor of the economic theory.

The second approach starts from the economic model and notes that the solution is a restricted version of a particular statistical model. In the case of DSGE models the solution is (approximated by) a restricted VAR. Then, an hybrid model is estimated by means of a mixed strategy estimation, where the actual data are supplemented with observations "generated" by the theoretical model. The (relative) amount of artificial data is controlled by an hyper-parameter, called λ, that is a direct measure of the weight given to the DSGE model. For $\lambda = 0$ we have the unrestricted VAR, for $\lambda = \infty$ we have the restricted VAR implied by the DSGE model. The value of λ associated with the highest marginal likelihood for the data, denoted as $\widehat{\lambda}$, delivers the so (mixed) model of interest. The evaluation of the economic theory is based on $\widehat{\lambda}$. For small values the theoretical model is considered useless in describing the data and therefore the model cannot be considered "true", while large values lead to opposite conclusions.

The first approach is in the spirit of [5] and [3], and developed in [2] and [1], where present value implied restrictions are exploited. The second approach is at the center of a recent literature on the estimation and evaluation of General Equilibrium (GE) model, starting from the works of [7] and especially [4]. This approach evaluate how well a certain set of over-identifying restrictions implied by the a GE model are supported by the data, conditional on the fact that the general statistical model used for comparison is straightforwardly obtained by relaxing the model subject to evaluation. This approach disregard the critique in Spanos (1990) that, in the words of Favero (2006), "...(a)ny identified structure that is estimated without checking that the underlying statistical model is a good description of the data is bound to fail if the statistical model is not valid". The critique can be summarized in acknowledging that statistical identification is distinct from structural identification and should be consistently pursued in any model evaluation approach.

3.2 Model evaluation - general elements

Denote with θ the parameters of the theoretical model, with y and z the endogenous and exogenous variables of the model, respectively. The solution to a fully specified model consists in expressing the endogenous variables as a function of the exogenous variables and the parameters. Closed form solutions do not exist in general for the class of DSGE models used by the profession, therefore one need to calculate approximate solutions, for instance by log-linearizing the model around the steady state conditions, obtaining the correspondent linear rational expectation (LRE) system. To solve the LRE model one can apply a range of available solution methods, e.g. Sims (2002). The solution comes in the form of a state space model

$$x_t = A_{x0}(\theta) + A_{xx}(\theta)x_{t-1} + A_{xz}(\theta)z_t \tag{3.1}$$

$$y_t = A_{y0}(\theta) + A_{yx}(\theta)x_t + A_{yz}(\theta)z_t \tag{3.2}$$

where x_t includes the endogenous states, y_t the other endogenous variables and z_t the exogenous shocks. Here $A_{ll'}(\theta)\ l, l' = \{y, x, z, 0\}$ are matrices of coefficients that depends on the structural parameters of the model. Note that typically the solution implies cross-equation restrictions, one element of θ can determine the value of coefficients in more than one equation.

3.2.1 Expectation Hypothesis of the Term Structure

In the rest of the paper all the empirical analyses, both on real data and on simulated ones, will be concerned with a single economic model, the Expectation Hypothesis of the term structure of interest rates (EH). The EH predicts that the expected value of holding for one period a multi-period bond is given by the interest rate on the relative one-period bond. In formulas we have

$$E_t(p_{t+1,T} - p_{t,T}) = r_{t,t+1} \tag{3.3}$$

Where $p_{t,T}$ is the (log) price at time t of a bond with maturity at T. Equivalently, the EH imposes that the yields to maturity of bonds with maturity T at time t is the average of the expected yields of the bonds from time t to time $T-1$ with maturity one period ahead.

$$r_{t,T} = \frac{1}{T-t}\sum_{i=0}^{T-1} E_t r_{t+i,t+i+1} \tag{3.4}$$

At this point it is important to note that the EH does not specify the process for the formation of the expectation by the market on the future evolution of the short-term rate. This model is intentionally considered for its simplicity and flexible formulation. In the following, we focus on the formulation of the expectation hypothesis popularized by Cambpell and Shiller (1991) expressing the yields spread as function of the expected changes in the short-term interest rate. Subtracting $r_{t,t+1}$ from both sides and rearranging one get

$$S_{t,T} = \sum_{i=1}^{T-1}(1 - i/(T - t + 1))E_t\Delta r_{t+i,t+i+1} + \tau + \eta_t \tag{3.5}$$

Since we deal with quarterly data, where the short-term interest rates are annualized three-months federal fund rate and we investigate the expectation hypothesis for a one-year T-bill interest rate, it is convenient to simplify the notation and rewrite the above general formula for our specific case. We denote the short-term and the long-term with r_t and R_t, respectively and use $S(t)$ to identify their spread. The relationship for the two pure-discount bonds is given by

$$R_t = \frac{1}{4}\sum_{i=0}^{3} E_t r_{t+i} \tag{3.6}$$

We go further with respect to the usual formulation of the EH and we consider a indeed looser version allowing for a term premium constituted by constant plus a martingale difference process, i.e. $\tau + \eta_t$,

$$R_t = \frac{1}{4}\sum_{i=0}^{3} E_t r_{t+i} + \tau + \eta_t \tag{3.7}$$

Finally, our long-short spread is given by

$$S_t = \sum_{i=1}^{3}(1 - i/4)E_t\Delta r_{t+i} + \tau + \eta_t \tag{3.8}$$

Consider a linear approximation for the solution of the EH for the term structure, this can be expressed as a linear rational expectation (LRE) model, whose solution has a representation in state-space form, as in the system (3.1) and (3.2). Where the states, x_t, is a $2p$-variate vector including both the change in short term rate ($\Delta r_t = r_t - r_{t-1}$), the long-short spread ($S_t = R_t - r_t$) and enough lags as prescribed by the equation describing the dynamics of the change in the short term rate, i.e. $x_t = [\Delta r_t, S_t, \ldots, \Delta r_{t-p+1} S_{t-p+1}]'$. The first row in the matrices $A_{x0}(\theta)$,

$A_{xx}(\theta)$ and $A_{xz}(\theta)$ includes the coefficients of the equation for the conditional expectations of the market on the change in the short term rate. For the second row, the EH maps the previous parameters into the coefficients of the dynamics of the long-short spread. To complete the model one need to specify the measurement equation that link the unobserved states to the observable data. In our analysis the matrices $A_{yl}(\theta)$, $l = 0, x, z$, are trivial, since the states are observed without error and in particular the matrix $A_{yx}(\theta)$ selects the first two elements of x_t, therefore $y_t = [\Delta r_t, S_t]'$.

This shows that we can work with a VAR in companion form, without losing information on the system.

$$Y_t = \mu + \mathrm{A}Y_{t-1} + \epsilon_t \tag{3.9}$$

where the companion matrix is given by

$$\mathrm{A} = \begin{bmatrix} \alpha' & \beta' \\ I_p & 0_{p\times p} \\ \gamma' & \delta' \\ 0_{p\times p} & I_p \end{bmatrix}$$

where the coefficients are given by

$$\begin{array}{cc} \alpha = [\alpha_1, \ldots, \alpha_p]' & \beta = [\beta_1, \ldots, \beta_p]' \\ \gamma = [\gamma_1, \ldots, \gamma_p]' & \delta = [\delta_1, \ldots, \delta_p]' \end{array}$$

$$\begin{aligned} \Delta r_t &= \sum_{j=1}^{p} [\alpha_j \Delta r_{t-j} + \beta_j S_{t-j}] + \varepsilon_{1t} \\ S_t &= \sum_{j=1}^{p} [\gamma_j \Delta r_{t-j} + \delta_j S_{t-j}] + \varepsilon_{2t} \end{aligned}$$

$$Y_t = \begin{bmatrix} \Delta r_t & \ldots & r_{t-p+1} & S_t & \ldots & S_{t-p+1} \end{bmatrix}'$$

Provided the VAR is stable, the EH imposes the following restrictions on the coefficients of the VAR:

$$4h_S\left(I_n - A\right)^2 - h_{\Delta r}\left(3A - 4A^2 + A^5\right) = 0 \tag{3.10}$$

where $h_{\Delta r}$ and h_S are the 1st and the $(p+1)$-th row of an identity matrix of dimension $2p$:

$$h_{\Delta r} = \begin{bmatrix} 1 \; 0_{p-1} & 0_p \end{bmatrix}$$
$$h_S = \begin{bmatrix} 0_p & 1 \; 0_{p-1} \end{bmatrix}$$

Call $g(\alpha, \beta, \gamma, \delta) = 0$ the restrictions implied above. We have in total $2p$ restrictions.

3.3 Model evaluation frameworks

3.3.1 Full information classical statistical framework

A first possible approach to evaluate macroeconometric models is to follow the classical statistical framework. Maximum likelihood estimation of the parameters in θ can be achieved by standard Kalman filter applied to the system (3.1) and (3.2), once the distribution of the exogenous shocks are assumed to be Gaussian. The maximum likelihood estimator of θ is defined by

$$\widehat{\theta}_{ML} = \underset{\theta \in \Theta}{\arg\max} \quad L(y|\theta) \tag{3.11}$$

where $L(y|\theta)$, the likelihood of the data, is obtained as a byproduct of filtering. The maximized value of the likelihood function is $L(y|\widehat{\theta}_{ML})$.

To perform a test of the validity of the theoretical model, one tests the restrictions imposed by the solution in (3.1) and (3.2) on a less restricted statistical representation for the data. The benchmark statistical model considered for the data, y_t, $t = 1, \ldots, T$ with $y_{-p+1}, \ldots, y_0$ hold fixed, is an unrestricted VAR .

$$y_t = \mu + \Pi_1 y_{t-1} + \cdots + \Pi_p y_{t-p} + \varepsilon_t \tag{3.12}$$

Assume the data be normally distributed, then collecting the constant and the parameters on the lagged coefficients in Π, and denoting the variance covariance matrix of the one-step ahead forecast errors of the VAR by Σ, one can write down the sample likelihood

$$L(y|\Pi, \Sigma) = (2\pi)^{-\frac{Tn}{2}} |\Sigma|^{-\frac{T}{2}} exp\{-\tfrac{1}{2} \textstyle\sum_{t=1}^{T} (y_t - X_t\Pi)'\Sigma^{-1}(y_t - X_t\Pi)\} \tag{3.13}$$

The maximized value of the likelihood function for the VAR is

$$L(y|\widehat{\Pi}, \widehat{\Sigma}) = (2\pi)^{-\frac{Tn}{2}} \left|\widehat{\Sigma}\right|^{-\frac{T}{2}} exp\{-\frac{Tn}{2}\} \tag{3.14}$$

where $\widehat{\Pi}$ coincides with OLS estimates of the coefficients and $\widehat{\Sigma}$ is the covariance matrix of the estimated residuals. The validity of the restrictions are tested using the likelihood ratio (LR) statistic for the null that the theoretical model is true

$$LR \;=\; -2\log\left(\frac{L(\widehat{\Pi},\widehat{\Sigma})}{L(\widehat{\theta}_{ML})}\right) \quad \underset{H_0}{\longrightarrow} \quad \chi^2_{\#\text{ restrictions}} \tag{3.15}$$

3.3.2 Limited information Bayesian framework

The second approach is used in Carriero (2005) and developed from an idea of Ingram and Whiteman (1994). It is Bayesian in the sense that combines prior information derived by the theoretical model and the information from the data through the likelihood function. Moreover the approach is of limited information nature in the sense that uses only the information obtained by a set of model implications, for instance a set of restrictions on the coefficients of the VAR representation for the data. The formulation in (3.1) and (3.2) imply a mapping between the structural parameters, θ, and the parameters of the statistical model, $\psi = \psi(\theta)$, i.e. we can rewrite the restrictions in terms of the mapping between structural parameters and parameters of the statistical model

$$g(\psi(\theta)) = 0 \tag{3.16}$$

We introduce prior information on the structural parameters, to keep analytical tractability our prior distribution is Gaussian with known mean and variance covariance matrix:

$$\theta \;\sim\; \mathcal{N}(\theta_0, \Sigma_0) \tag{3.17}$$

Using first-order Taylor expansion around the mean of the prior distribution, one get the following approximated distribution for the implied parameters of the statistical model

$$\psi(\theta) \;\sim\; \mathcal{N}\left(\psi(\theta_0), \Sigma_{\psi_0}\right) \tag{3.18}$$

where $\Sigma_{\psi_0} = \frac{\mathrm{d}\psi}{\theta'}\Sigma_0\frac{\mathrm{d}\psi'}{\theta}$. Then one incorporates the information derived by the restrictions in (3.16), by considering the Jacobian of $g(\psi(\theta))$ evaluated at $\psi(\theta_0)$

$$G = \left.\frac{\mathrm{d}g}{\mathrm{d}\psi'}\right|_{\psi=\psi(\theta_0)} \tag{3.19}$$

and approximating the restrictions around the prior mean of the structural parameters

$$G(\psi(\theta) - \psi(\theta_0)) = 0 \tag{3.20}$$

if the model is true then the coefficients of the statistical model will satisfy the restrictions with a small discrepancy. This fact suggests that we can assess the validity of the model by introducing a prior on (3.20) as

$$G(\psi(\theta) - \psi(\theta_0)) \quad \sim \quad \mathcal{N}(0, \Sigma_{\mathcal{M}} = \sigma I) \tag{3.21}$$

where σ is the a priori tightness of the restrictions. One more step is needed to elicit the prior on ψ, which has a normal distribution. Combining the prior with the likelihood of the data one can compute the posterior probability of the model, conditional on a given value of σ. For the evaluation, a more general competing theory is needed, for instance this second theory could impose no restrictions on the coefficients of the statistical model. The theories are compared using the Bayes factor, which is a summary of the evidence provided by the data in favor of one theory opposed to the other. The value of σ that maximizes the Bayes factor of the two model, denoted $\widehat{\sigma}$, together with value of the Bayes factor itself, provides some information about model fit.

3.3.3 Full information Bayesian framework

A possible way of combining the likelihood of the data in (3.13) with prior information deduced from a theoretical model, is to add observations generated from the model. Consider a sample of λT such artificial observations, y^*, the likelihood of the extended sample is given by

$$L(y^*, y|\Pi, \Sigma, \lambda) = L(y^*|\Pi, \Sigma, \lambda)L(y|\Pi, \Sigma) \tag{3.22}$$

Following this observation, Del Negro and Schorfheide (2004) implements a Bayesian estimation method by hierarchically defining a prior on the parameters of the DSGE , that maps a prior on the parameters of the VAR representation for the DSGE model. Such derived prior assigns positive probability to the state of the world where the DSGE is false.

The prior has mode $\Pi(\theta)$, $\Sigma(\theta)$, and is denoted as:

$$P(\Pi, \Sigma|\theta, \lambda) \tag{3.23}$$

the parameter λ controls the tightness of the prior believes around the mode. For small values of λ the prior provides little or no information on Π and Σ, for λ that goes to infinity the DSGE is considered the true model, i.e. the prior concentrate the probability mass over $\Pi(\theta)$ and $\Sigma(\theta)$.

The marginal likelihood of the data, conditional on the priors and on a value for λ is denoted $\mathcal{L}(y,\lambda)$, which is defined as follows:

$$\mathcal{L}(y,\lambda) = \int_{\theta}\int_{(\Pi,\Sigma)} L(y|\Pi,\Sigma)P(\Pi,\Sigma|\theta,\lambda)P(\theta)\mathrm{d}(\Pi,\Sigma)\mathrm{d}\theta. \tag{3.24}$$

The value of λ that maximizes the marginal likelihood of the data, denoted $\widehat{\lambda}$, is the key output of the procedure, and provides some information about model fit.

In principle, evaluating $\mathcal{L}(y,\lambda)$ requires solving a massive numerical integration problem. To avoid this, it is possible to employ a posterior simulator and integrating via monte carlo. Using $P(\Pi,\Sigma|\theta,\lambda)$ conjugate with the normal likelihood. That is, $P(\Pi,\Sigma|\theta,\lambda)$ is the product of the inverse Wishart density for Σ and the multivariate normal density for Π conditional on Σ. With this specification of the prior the integral,

$$\int_{(\Pi,\Sigma)} L(y|\Pi,\Sigma)P(\Pi,\Sigma|\theta,\lambda)P(\theta)\mathrm{d}(\Pi,\Sigma), \tag{3.25}$$

can be evaluated analytically for given values of θ, λ. Dramatically reducing the dimension of the integration problem for evaluating $\mathcal{L}(y,\lambda)$. Nonetheless, the maximization problem

$$\widehat{\lambda} = \underset{\lambda\geq 0}{\operatorname{argmax}}\mathcal{L}(y,\lambda) \tag{3.26}$$

is limited to a restricted grid of values.

For a specific value of λ the mode of the posterior distribution of Π and Σ is the so called DSGEVAR model. So, a side product of the calculations is a best hybrid parametrization. If that parametrization is 'far' from the DSGE model (i.e. $\widehat{\lambda}$ is small) this is indication that the DSGE model fits poorly. If the hybrid parametrization corresponds closely to that implied by the DSGE model (i.e. $\widehat{\lambda}$ is large) this is an indicator of good fit.

3.4 Model Evaluation - Application to EH

In this section we present the application of the three model evaluation approaches to the valuation of the EH of the term structure.

3.4.1 Classical framework

The first approach used in the evaluation of the LRE is the full information Maximum Likelihood method. We cast the bivariate system for the change in short

term rate and the spread in state-space form. The dynamic of the changes in the short term rate imposes exact restrictions on the equation for the spread[1].
The model is the following: The equation for the change in short term rate is left unrestricted, including p autoregressive terms and lags of the spread:

$$\Delta r_t = \mu_1 + \sum_{i=1}^{p} \alpha_i \Delta r_{t-i} + \sum_{i=1}^{p} \beta_i S_{t-i} + \varepsilon_t$$

While the spread, according to the expectation hypothesis, is a linear combination of future change in short term, plus a constant term premium and an idiosyncratic component from a martingale difference process

$$S_t = \sum_{i=1}^{3} (1 - i/4) E_t \Delta r_{t+i} + \tau + \eta_t$$

Where $Var\,(\varepsilon) = \sigma_\varepsilon^2$, $Var\,(\eta) = \sigma_\eta^2$ and $Cov\,(\varepsilon, \eta) = 0$

We can represent the system in state-space form as in (3.1) and (3.2) Where $z_t = [\varepsilon_t \quad \eta_t]'$.

In our case the states are both the change in short term interest rate and the spread

$$x_t = \begin{bmatrix} \Delta r_t & S_t & \dots & \Delta r_{t-p+1} & S_{t-p+1} \end{bmatrix}' \tag{3.27}$$

$$A_{x0} = \begin{bmatrix} \mu_1 \\ \mu_2(\mu_1, \alpha, \beta) + \tau \end{bmatrix} \tag{3.28}$$

$$A_{xx} = \begin{bmatrix} \alpha' & \beta' \\ \gamma'(\alpha, \beta) & \delta'(\alpha, \beta) \\ I_{2p-2} & 0_{2p-2,2} \end{bmatrix} \tag{3.29}$$

$$A_{xz} = I_2 \tag{3.30}$$

$$\Sigma_z = \begin{bmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & \sigma_\eta^2 \end{bmatrix} \tag{3.31}$$

The measurement equation is defined for $y_t = [\Delta r_t, S_t]'$

$$A_{y0} = 0_{2,1} \tag{3.32}$$

$$A_{yx} = \begin{bmatrix} I_2 & 0_{2,2p-2} \end{bmatrix} \tag{3.33}$$

$$A_{yz} = 0_{2,1} \tag{3.34}$$

The likelihood function of a state space model can be expressed in terms of the one-step ahead forecast errors, conditional on the initial observations, and of

[1] To solve the LRE model we use the package *gensys*, Sims (2002)

their recursive variance, both of which can be obtained with the Kalman filter. Therefore, given some initial parameter values, the Kalman filter can be used to recursively construct the likelihood function.
Maximum Likelihood estimates of the parameter vector $\widehat{\theta}_{ML}$ are obtained implementing the Kalman filter for the state-space model previously described. In the following table we report the value of the log-likelihood for the unrestricted VAR and for the (restricted) state-space model and the relative likelihood ratio tests.

Table 3.1: Maximum Likelihood - Likelihood ratio test

Model	$\log L(\widehat{\Pi}, \widehat{\Sigma})$	$\log L(\widehat{\theta}_{ML})$	LR	p-value
VAR(1)	-65.36	-89.01	47.28	$5e$-011
VAR(2)	-61.84	-72.60	21.52	0.0003
VAR(3)	-60.94	-71.21	20.55	0.0022
VAR(6)	-60.55	-70.92	20.63	0.0020

The null hypothesis for the validity of the EH restrictions is strongly rejected in all four models.

3.4.2 Limited Information Bayesian

A useful formulation of the VAR is The system in SUR form is

$$y = \Xi\psi + \epsilon \tag{3.35}$$

where

$$y = \begin{bmatrix}\Delta r_1 & \dots & \Delta r_T & S_1 & \dots & S_T\end{bmatrix}' \tag{3.36}$$

$$\Xi = [I_2 \otimes X] \tag{3.37}$$

$$X = \begin{bmatrix} 1 & \Delta r_p & \dots & \Delta r_1 & S_p & \dots & S_1 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & \Delta r_{t-1} & \dots & \Delta r_{t-p+1} & S_{t-1} & \dots & S_{t-p+1} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & \Delta r_{T-1} & \dots & \Delta r_{T-p+1} & S_{T-1} & \dots & S_{T-p+1} \end{bmatrix} \tag{3.38}$$

The parameters are given by

$$\psi = \begin{bmatrix}\mu_1 & \alpha' & \beta' & \mu_2 & \gamma' & \delta'\end{bmatrix}' \tag{3.39}$$

The errors are Multivariate Normally distributed

$$\varepsilon \sim \mathcal{N}\left(0, \quad \Sigma_0 \otimes I_2\right) \tag{3.40}$$

The prior on the parameters for the unrestricted model can be described as

$$\psi \sim \mathcal{N}\left(\psi_0, \quad \Sigma_{\psi_0}\right) \tag{3.41}$$

In testing a weaker version of the Expectation Hypothesis, we consider a subset of restrictions imposed on $\vartheta = (\alpha', \beta', \gamma', \delta')'$ can be described by the vector function $g\left(\vartheta_{EH}\right) = 0$.

Leaving unrestricted the dynamics of the change in the short term rate, and allowing for a constant term premium in the long-term rate, the rational expectation model imposes $2p$ cross-equation restrictions, described in (3.10).

We can approximate the restrictions with a first order Taylor expansion as described in (3.20), the Jacobian of the constraint w.r.t. the parameters in ϑ is

$$\begin{aligned}\frac{\mathrm{d}G'(\theta)}{\mathrm{d}(\theta')} &= -4\left(I_{2p} \otimes h_S'\right)\{2I_{4p^2} - \left[(A' \otimes I_{2p}) + I_{2p} \otimes A\right]\}K_{2p,2p}\frac{\mathrm{d}vec(A')}{\mathrm{d}(\theta')} \\ &\quad -(I_{2p} \otimes h_{\Delta r}')\{3I_{4p^2} - 4\left[(A' \otimes I_{2p}) + I_{2p} \otimes A\right] \\ &\quad + \left[\sum_{i=0}^{4}\left(A'\right)^{4-i} \otimes A^i\right]\}K_{2p,2p}\frac{\mathrm{d}vec\left(A'\right)}{\mathrm{d}(\theta')}\end{aligned} \tag{3.42}$$

We can add uncertainty around the constraints implied by the Expectation Hypothesis specifying a prior distribution on the parameters involved as

$$\vartheta \sim \mathcal{N}\left(\vartheta_{EH0}, \quad \Sigma_{EH0}\right) \tag{3.43}$$

Moreover, using the expansion (3.20), a linear combinations of the parameters provide further information that can be written as:

$$g\left(\vartheta\right) \sim \mathcal{N}\left(g\left(\vartheta_{EH0}\right), \quad \Sigma_0\right) \tag{3.44}$$

We can alternatively write:

$$\vartheta_{EH} \sim \mathcal{N}\left(\ \vartheta_0, \quad \Sigma_{\vartheta_0}\ \right) \tag{3.45}$$

Since linear combination of Normal random variables are still Normal

$$G\left(\vartheta_{EH0}\right)'\vartheta_{EH} \sim \mathcal{N}\left(G\left(\vartheta_{EH0}\right)'\vartheta_0 \quad , \quad G\left(\vartheta_{EH0}\right)'\Sigma_{\vartheta_0}G\left(\vartheta_{EH0}\right)\right) \tag{3.46}$$

The two set of restrictions are the same, therefore we have that

$$g\left(\vartheta_{EH0}\right) = G\left(\vartheta_{EH0}\right)'\vartheta_0 \tag{3.47}$$

$$\Sigma_0 = G\left(\vartheta_{EH0}\right)'\Sigma_{\vartheta_0}G\left(\vartheta_{EH0}\right) \tag{3.48}$$

Note that the restrictions are a system of $2p$ equations in $4p$ unknowns, therefore the Jacobian $G(\theta_{EH0})'$ is a $2p \times 4p$ matrix, hence we cannot use its inverse to derive the parameters in (3.44) in terms of the parameters in (3.45)

We can set a diffuse prior on the whole set of coefficients (ψ):

$$\psi \;\sim\; \mathcal{N}\left(\; \psi_0, \;\; \Sigma_{\psi_0} \;\right) \tag{3.49}$$

Setting $\Sigma_{\vartheta_0} = \sigma_0 I_{2p}$ and $\Sigma_0 = \sigma I_{4p+2}$, we can combine the prior information as

$$H\psi \;\sim\; \mathcal{N}\left(\begin{bmatrix} \psi_0 \\ g(\vartheta_0) \end{bmatrix}, \; \begin{bmatrix} \Sigma_{\psi_0} & 0_{2p+2\times 2p} \\ 0_{2p\times 2p+2} & \Sigma_0 \end{bmatrix} \right) \tag{3.50}$$

Now we can derive the prior for the whole set of parameters implied by prior information from Expectation Hypothesis restrictions and the diffuse prior:

$$\vartheta_0 = H^{-1} g(\vartheta_{EH0}) \tag{3.51}$$

$$\Sigma_{\vartheta_0} = H^{-1}\Sigma_0 \left(H'\right)^{-1} \tag{3.52}$$

The results of limited information Bayesian evaluation are reported in Figures 3.1 and 3.2, the maximized log-Bayes factors for the 4 models are ,7.5, 5.5, 7.1 and 11.5, respectively, which are obtained for values of σ of about 0.45, 1.2, 1.5, 2. While a value of about 20 is used for σ_0, this value has been chosen in order to have a maximum distance between loose prior bayesian estimates and the OLS estimates less then 0.01. This gives a ratio of the two parameters ranging from 0.022 to 0.1, rather close to zero. The evidence speaks in support of the validity of the EH model, but we should be aware of the values of the Bayes factors that are not large enough to support the theory in a stronger way.

Figure 3.1: Bayes Factor for different tightness - $\Delta r_t \sim VAR(1)$

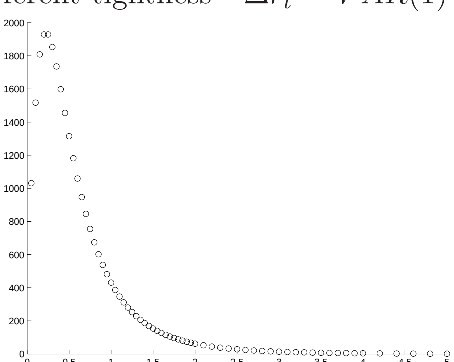


Figure 3.2: Bayes Factor for different tightness - $\Delta r_t \sim VAR(2)$

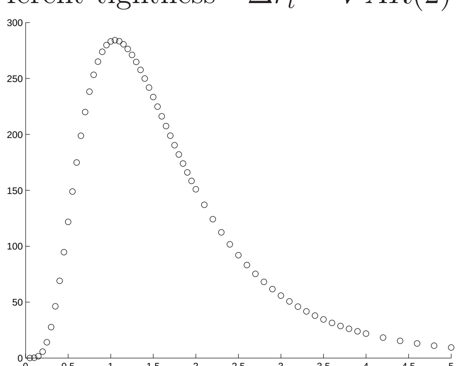


Figure 3.3: Bayes Factor for different tightness - $\Delta r_t \sim VAR(3)$

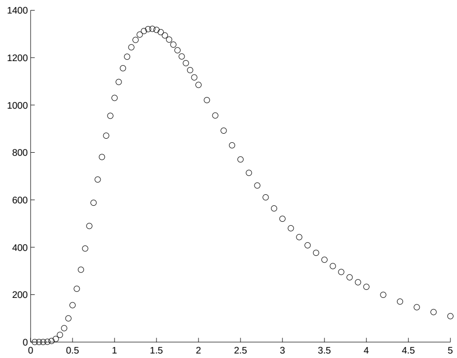
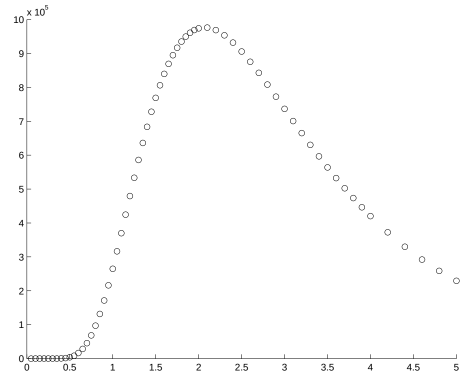


Figure 3.4: Bayes Factor for different tightness - $\Delta r_t \sim VAR(6)$



3.4.3 Full Information Bayesian

In a series of papers Del Negro and Schorfheide (2004, and 2006) and Del Negro, Schorfheide Smets and Wouters (2007) introduce a framework for the evaluation of DSGE models. The procedure makes use of an hybrid model that is a mixture of an unrestricted VAR for the data and the VAR implied by the DSGE model. This mixture is parametrized by an hyperparameter called λ. For $\lambda = 0$ the hybrid model reduces to the unrestricted VAR for $\lambda = \infty$ it reduces to the DSGE model. The value of λ that maximizes the marginal likelihood provides an overall assessment of the validity of the DSGE model restrictions. The chosen benchmark to evaluate this model is the unrestricted VAR derived from the solved DSGE model.

Our analysis consider the Expectation Hypothesis model introduced above. One can obtain the Equation (3.6) as linearized solution of a rational expectation model. We follow Del Negro and Schorfheide (2004) to solve the Linear Rational Expectation model by applying Sims' algorithm (2002) and by ruling out indeterminacy, to get transition equation as in (3.2).

The model solution delivers the dynamics of deviations from the steady state. To obtain the dynamics of change in the short rate and the spread we combine the transition equation with a measurement equation in (3.1). In EH model, there are two measurement equations. The change in short-term rate is obtained directly from the states of the LRE model, while for the spread equation we add a constant term premium and a martingale difference process component:

$$\Delta r_t = \Delta \widetilde{r}_t(+\mu_1) \tag{3.53}$$

$$S_t = \widetilde{S}_t(+\mu_1) + \tau + \eta_t \tag{3.54}$$

Taking together the transition equation and the measurement equation the EH model parameters are stacked into the vector:

$$\theta = [\alpha_1, \beta_1, \ldots, \alpha_p, \beta_p, \mu, \tau, \sigma_\varepsilon, \sigma_\eta]'$$

We consider quarterly data, where the interest rates, r_t and R_t, are annualized three-months federal fund rate and one-year T-bill interest rate.

As it is clear from the transition and measurement equation, the EH model imposes tight restrictions across the parameters of the moving average (MA) representation for the change of short-term rate and the spread. Given that the MA representation can be closely approximated by a finite order VAR representation, to evaluate the EH model one assesses the validity of the restrictions imposed by such model on an unrestricted VAR representation for the two series of interest.

Likelihood Function

Consider an unrestricted VAR representation for the vector of relevant variables as in (3.12), this is a less restrictive representation than the one implied by the EH model,as the vector of parameter of EH model, θ, is of much lower dimension than the VAR parameter vector.

The likelihood function, assuming that in (3.12) the one step ahead forecast errors ε_t have a multivariate normal distribution $N(0, \Sigma_\varepsilon)$ conditional on past observations of y_t. Considering (3.12) in matrix representation:

$$y = X\Pi + \varepsilon \tag{3.55}$$

where y is a $(T \times 2)$ matrix with rows y_t', X is a $(T \times k)$ matrix ($k = 1 + 2p, p =$number of lags) with rows $X_t'' = [1, y_{t-1}', ..., y_{t-p}']$, ε is a $(T \times 2)$ matrix with rows ε_t' and Π is a $(k \times 2) = [\mu, \Pi_1, ..., \Pi_p]'$.

The likelihood function given by (3.13), can be rewritten as

$$p(y|\Pi, \Sigma_\varepsilon) \propto |\Sigma_\varepsilon|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2} tr[\Sigma_\varepsilon^{-1} \left(y'y - \Pi'X'y - y'X\Pi + \Pi'X'X\Pi\right)]\right\} \tag{3.56}$$

conditional on observations $y_{1-p}, ..., y_0$.

Prior distribution

To construct the prior distribution for the relevant inference suppose that the actual observations are augmented with $T^* = \lambda T$ artificial observations (y^*, X^*) generated from the EH model based on the parameter vector θ. The likelihood function for artificial observations is:

$$p(y^*(\theta)|\Pi, \Sigma_\varepsilon) \quad \propto$$

$$|\Sigma_\varepsilon|^{-\frac{\lambda T}{2}} \exp\left\{-\frac{1}{2} tr[\Sigma_\varepsilon^{-1}\left(y^{*\prime}y^* - \Pi^{*\prime}X^{*\prime}y^* - y^{*\prime}X^*\Pi^* + \Pi^{*\prime}X^{*\prime}X^*\Pi^*\right)]\right\} \quad (3.57)$$

The concept of using dummy observations to introduce prior information comes from noting that combining the likelihood for the sample of artificial and actual observations one obtains:

$$p(y^*(\theta), y|\Pi, \Sigma_\varepsilon) = p(y^*(\theta)|\Pi, \Sigma_\varepsilon)p(y|\Pi, \Sigma_\varepsilon) \quad (3.58)$$

In the above expression, $p(y^*(\theta)|\Pi, \Sigma_\varepsilon)$ can be interpreted as a prior density for Π and Σ_u.

To remove stochastic variation in the prior (3.57) one replace the nonstandardized sample moments $y^{*\prime}y^*, X^{*\prime}y^*$, $y^{*\prime}X^*$ and $X^{*\prime}X^*$ by their expected values. Considering the EH model, the vector y_t is covariance stationary and the expected values of the sample moments are given by the scaled population moments $\lambda T\Gamma^*_{yy}(\theta), \lambda T\Gamma^*_{XX}(\theta)$ and $\lambda T\Gamma^*_{yX}(\theta)$, for example, $\Gamma^*_{yy}(\theta) = E_\theta[y_t y_t']$.

Using population moments implies that (3.57) can be treated as $p(\theta|\Pi, \Sigma_u)$, then by adding an initial improper prior and provided that $\lambda \geq \frac{k+2}{T}$ and Γ^*_{XX} is invertible, we have a proper non-degenerate prior.

Define the functions:

$$\begin{aligned}
\Pi^*(\theta) &= \Gamma^{*-1}_{XX}(\theta)\Gamma^*_{Xy}(\theta) \\
\Sigma^*_\varepsilon(\theta) &= \Gamma^*_{yy}(\theta) - \Gamma^*_{yX}(\theta)\Gamma^{*-1}_{XX}(\theta)\Gamma^*_{Xy}(\theta)
\end{aligned}$$

Conditional on θ the prior distribution of the VAR parameters and the distributions are:

$$\begin{aligned}
\Sigma_\varepsilon|\theta &\sim \mathcal{IW}(\lambda T\Sigma^*_\varepsilon(\theta), \lambda T - k, 2) \\
\Pi|\Sigma_\varepsilon, \theta &\sim \mathcal{N}(\Pi^*(\theta), \Sigma_\varepsilon \otimes (\lambda T\Gamma^*_{XX}(\theta))^{-1})
\end{aligned}$$

and then the specification of the prior is completed with a distribution of the model parameters $p(\theta)$. Overall the prior takes the following hierarchical structure:

$$p(\Pi, \Sigma_\varepsilon, \theta) = p(\Pi, \Sigma_\varepsilon|\theta)p(\theta)$$

The prior is designed to assign probability mass outside the subspace traced out by $\Pi^*(\theta), \Sigma^*_\varepsilon(\theta)$.These two functions are identified by assuming that the data are generated from a DSGE model with parameters θ, to then choose among all possible p-th order VARs the one with coefficient matrix $\Pi^*(\theta)$ that minimizes the one-step ahead quadratic forecast loss. The corresponding forecast error covariance matrix is given by $\Sigma^*_\varepsilon(\theta)$.

They suppose that data are generated from a DSGE model with parameters θ. Among the p-th order VARs the one with the coefficient matrix $\Pi^*(\theta)$ minimizes the one-step ahead quadratic forecast error loss and the corresponding forecast error covariance matrix is given by $\Sigma_\varepsilon^*(\theta)$.

Probability mass is distributed around $\Pi^*(\theta)$ by using the covariance matrix $\Sigma_\varepsilon \otimes (\lambda T \Gamma_{XX}^*(\theta))^{-1}$.

The prior distributions for the structural parameters for the 4 different analyzed models, $p(\theta)$, are specified in Appendix in Table 2. We calculate that approximately 2.5% of the prior lies in the indeterminacy region of the parameter space; in fact, the prior is truncated in order to restrict it to the determinacy region of the EH model.

Posterior distribution

The posterior distribution is better described by factorizing it into the posterior density of the VAR parameters given the EH model parameters and the marginal posterior density of the EH model parameters:

$$p(\Pi, \Sigma_\varepsilon, \theta | y) = p(\Pi, \Sigma_\varepsilon | y, \theta) p(\theta | y) \tag{3.59}$$

To construct $p(\Pi, \Sigma_\varepsilon | y, \theta)$ define:

$$\begin{aligned} \widehat{\Pi}(\theta) &= \left(\lambda T \Gamma_{XX}^*(\theta) + X'X\right)^{-1} \left(\lambda T \Gamma_{Xy}^*(\theta) + X'y\right) \\ \widehat{\Sigma}_\varepsilon(\theta) &= \frac{1}{(\lambda+1)T} \left(\lambda T \Gamma_{yy}^*(\theta) + y'y\right) - \\ &\quad - \frac{1}{(\lambda+1)T} \left(\lambda T \Gamma_{yX}^*(\theta) + y'X\right) \left(\lambda T \Gamma_{XX}^*(\theta) + X'X\right)^{-1} \times \\ &\quad \left(\lambda T \Gamma_{Xy}^*(\theta) + X'y\right) \end{aligned}$$

$\widehat{\mathbf{\Pi}}(\theta), \widehat{\Sigma}_\varepsilon(\theta)$ are the maximum likelihood estimates based on artificial sample and actual sample, where λ determines the length of the artificial sample. Conditional on θ the EH model prior and the likelihood function are conjugate, so the posterior distribution of Π, Σ_ε is also of the inverted Wishart-Normal form:

$$\begin{aligned} \Sigma_\varepsilon | \theta, y &\sim \mathcal{IW}((\lambda+1) T \widehat{\Sigma}_\varepsilon(\theta), (\lambda+1)T - k, n) \\ \Pi | \Sigma_\varepsilon, \theta, y &\sim \mathcal{N}(\widehat{\Pi}(\theta), \Sigma_\varepsilon \otimes (\lambda T \Gamma_{XX}^*(\theta) + X'X)^{-1}) \end{aligned}$$

The marginal posterior density for θ and the selection of λ, from a finite grid, are obtained with MCMC techniques based on the above computations. The following figures shows the computations for the marginal likelihood for different values of the hyper-parameter λ. For all four different versions of the EH, the value of $\widehat{\lambda}$ is given by 0.5. On one hand, the value of $\widehat{\lambda}$ tell us that the EH, in the version presented in this work, is not completely useless in helping explaining the data. On the other hand, in all four figures we can note that after reaching the peak at $\widehat{\lambda}$, the marginal likelihood decreases steeply, this cast some doubt on the actual adequacy of the theoretical model.

Figure 3.5: Marginal likelihood - λ grid - $\Delta r_t \sim VAR(1)$

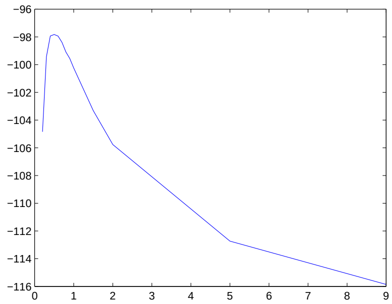


Figure 3.6: Marginal likelihood - λ grid - $\Delta r_t \sim VAR(2)$

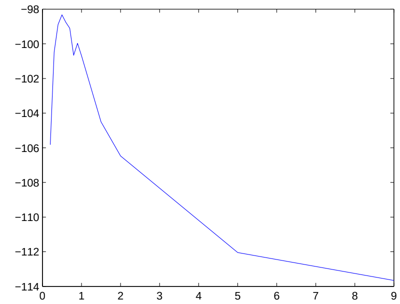


Figure 3.7: Marginal likelihood - λ grid - $\Delta r_t \sim VAR(3)$

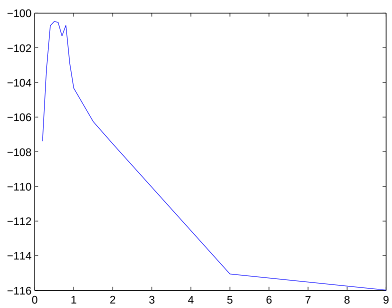
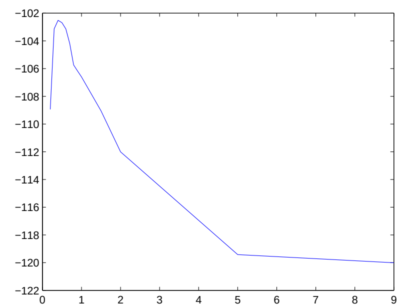


Figure 3.8: Marginal likelihood - λ grid - $\Delta r_t \sim VAR(6)$



3.5 Conclusions

This paper compares two approaches for Vector AutoRegression (VAR) based model evaluation developed in the Bayesian analysis of macroeconomic time series. The first approach starts from a statistical model that well describes the data, then it derives the theoretical restrictions imposed on the coefficients of the statistical model by the economic theory. In the next step, a prior distribution is imposed on the coefficients of the statistical model. The priors are conceived so to allow to control for the *tightness* of the theoretical implications. As a term of comparison, an alternative theoretical model is developed. The restrictions of this second model bear no economic content, but justified by its ability to improve statistical model fit. The two theories are compared in terms of Bayes Factors (see [6]) for different values of the tightness parameter. In practice, the tightness parameter value that delivers the highest Bayes factor is compared to the analogous tightness imposed on the alternative theory, if the ratio of the two is sufficiently small (and the relative Bayes Factor is enough high), the evidence is in favor of the economic theory. With the help of some Monte Carlo Experiment, the comparison between the two methods highlighted that they behave similarly when the model under validation is the true model, but when the empirical model fails to acknowledge the true process generating the market expectations, the simple version of the expectation hypothesis under examination finds support from the limited information Bayesian approach, while the full information Bayesian approach performs relatively poorly. This could be due to lack of statistical identification hampering critically the full information method.

To be added

3.6 Appendix - derivations

3.6.1 Expectation Hypothesis restrictions

We assuming we can represent the dynamics of system as a VAR

$$\mathbf{z}_t = A\mathbf{z}_{t-1} + \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t \quad \text{s.t.} \quad E\left(\boldsymbol{\varepsilon_t}|I_{t-1}\right) = 0$$

$$E_t\mathbf{z}_{t+i} = A^i\mathbf{z}_t + \left(I + A + \cdots + A^{i-1}\right)\mu + E_t\boldsymbol{\varepsilon}_{t+i}$$

The expectation hypothesis states that the spread forecasts changes in short-term interest rate, can be written as

$$
\begin{aligned}
h_S \mathbf{z}_t &= h_{\Delta r} \frac{1}{T} \sum_{i=1}^{T-1} (T-i) E_t \mathbf{z}_{t+i} \\
h_S \mathbf{z}_t &= h_{\Delta r} \frac{1}{T} \sum_{i=1}^{T-1} (T-i) \left[A^i \mathbf{z}_t + \left(I + A + \cdots + A^{i-1} \right) \boldsymbol{\mu} \right] \\
&= h_{\Delta r} \{ \frac{1}{4} \sum_{i=1}^{3} \left[A^i \mathbf{z}_t + \left(I + A + \cdots + A^{i-1} \right) \boldsymbol{\mu} \right] \\
&+ \frac{1}{4} \sum_{i=1}^{2} \left[A^i \mathbf{z}_t + \left(I + A + \cdots + A^{i-1} \right) \boldsymbol{\mu} \right] + \frac{1}{4} \left[A \mathbf{z}_t + \boldsymbol{\mu} \right] \} \\
&= h_{\Delta r} \{ \frac{1}{4} \left[A \mathbf{z}_t + \boldsymbol{\mu} + A^2 \mathbf{z}_t + (I + A) \boldsymbol{\mu} + A^3 \mathbf{z}_t + (I + A + A^2) \boldsymbol{\mu} \right] \\
&+ \frac{1}{4} \left[A \mathbf{z}_t + \boldsymbol{\mu} + A^2 \mathbf{z}_t + (I + A) \boldsymbol{\mu} \right] + \frac{1}{4} \left[A \mathbf{z}_t + \boldsymbol{\mu} \right] \} \\
&= h_{\Delta r} \frac{1}{4} \{ \left[(A + A^2 + A^3) + (A + A^2) + A \right] \mathbf{z}_t \\
&+ \left[I + (I + A) + (I + A + A^2) + I + (I + A) + I \right] \boldsymbol{\mu} \} \\
&= h_{\Delta r} \frac{1}{4} \{ A \left[(I - A^3)(I - A)^{-1} + (I - A^2)(I - A)^{-1} + (I - A)(I - A)^{-1} \right] \mathbf{z}_t \\
&+ \left[3(I - A)(I - A)^{-1} + 2(I - A^2)(I - A)^{-1} + (I - A^3)(I - A)^{-1} \right] \boldsymbol{\mu} \} \\
&= h_{\Delta r} \frac{1}{4} \left[(3A - 4A^2 + A^5)(I - A)^{-2} \mathbf{z}_t + (4I - 5A + A^5)(I - A)^{-2} \boldsymbol{\mu} \right]
\end{aligned}
$$

Neglecting the constant term

$$
4 h_S \left(I_n - A \right)^2 - h_{\Delta r} \left(3A - 4A^2 + A^5 \right) = 0
$$

3.6.2 Jacobian of restrictions

Denote with

$$
G' = \frac{\mathrm{d} g(\cdot)}{\mathrm{d}(\alpha', \beta', \gamma', \delta')} \tag{3.60}
$$

$$
4 \frac{\mathrm{d} vec \left(h_S' (I - A)^2 \right)}{\mathrm{d}(\alpha', \beta', \gamma', \delta')} = -4 \left(I_{2p} \otimes h_S' \right) \{ 2 I_{4p^2} - \left[(A' \otimes I_{2p}) + I_{2p} \otimes A \right] \} \tag{3.61}
$$

$$
K_{2p,2p} \frac{\mathrm{d} vec \left(A' \right)}{\mathrm{d}(\alpha', \beta', \gamma', \delta')} \tag{3.62}
$$

Where the commutation matrix $K_{2p,2p}$ is such that

$$
vec(A) = K_{2p,2p} vec(A') \tag{3.63}
$$

and

$$\frac{\mathrm{d}vec(A)}{\mathrm{d}(\alpha',\beta',\gamma',\delta')} = K_{2p,2p}\frac{\mathrm{d}vec(A')}{\mathrm{d}(\alpha',\beta',\gamma',\delta')} \tag{3.64}$$

$$\frac{\mathrm{d}vec(A')}{\mathrm{d}(\alpha',\beta',\gamma',\delta')} = \begin{bmatrix} I_{2p} & 0_{2p\times 2p} \\ 0_{2(p-1)\times 2p} & 0_{2(p-1)\times 2p} \\ 0_{2p\times 2p} & I_{2p} \\ 0_{2(p-1)\times 2p} & 0_{2(p-1)\times 2p} \end{bmatrix} \tag{3.65}$$

$$\begin{aligned} \frac{\mathrm{d}vec(h_{\Delta r}(3A-4A^2+A^5))}{\mathrm{d}\theta'} &= \left(I_{2p}\otimes h'_{\Delta r}\right)\{2I_{4p^2}-4[(A'\otimes I_{2p})+I_{2p}\otimes A] \\ &+ [\sum_{i=0}^{4}(A')^{4-i}\otimes A^i]\}K_{2p,2p}\frac{\mathrm{d}vec(A')}{\mathrm{d}\theta'} \end{aligned}$$

Summing up the two terms we get expression (3.42)

3.7 Appendix - Solution of LRE model

We solve the Linear Rational Expectation model by using the method proposed by Sims (2002), necessary step is to cast the model in the following form:

$$\Gamma_0\widetilde{\mathbf{Y}}_t = \Gamma_1\widetilde{\mathbf{Y}}_{t-1} + C + \Psi\mathbf{Z}_t + \Pi\zeta_t \tag{3.66}$$

$$\widetilde{\mathbf{Y}}_t = \begin{bmatrix} \Delta\widetilde{r}_t \\ \widetilde{S}_t \\ \vdots \\ \Delta\widetilde{r}_{t-p+1} \\ \widetilde{S}_{t-p+1} \\ E_t\Delta\widetilde{r}_{t+1} \\ E_t\Delta\widetilde{r}_{t+2} \end{bmatrix} \quad \mathbf{Z}_t = \begin{bmatrix} \varepsilon_t \end{bmatrix} \quad \zeta_t = \begin{bmatrix} \zeta_t^0 = \Delta\widetilde{r}_t - E_{t-1}\Delta\widetilde{r}_t \\ \zeta_t^1 = E_t\Delta\widetilde{r}_{t+1} - E_{t-1}\Delta\widetilde{r}_{t+1} \\ \zeta_t^2 = E_t\Delta\widetilde{r}_{t+2} - E_{t-1}\Delta\widetilde{r}_{t+2} \end{bmatrix}$$

$$\Gamma_0 = \begin{bmatrix} 1 & 0 & 0_{1,2p-2} & 0 & 0 \\ \frac{3}{4} & 0 & 0_{1,2p-2} & \frac{1}{2} & \frac{1}{4} \\ 0_{2p-2,1} & 0_{2p-2,1} & I_{2p-2} & 0_{2p-2,1} & 0_{2p-2,1} \\ 1 & 0 & 0_{1,2p-2} & 0 & 0 \\ 0 & 0 & 0_{1,2p-2} & 1 & 0 \end{bmatrix}$$

$$\Gamma_1 = \begin{bmatrix} \alpha_1 & \beta_1 & \{\alpha_j \quad \beta_j\}_{j=2:p} & 0 & 0 \\ 0 & 1 & 0_{1,2p-2} & 0 & 0 \\ I_{2p-2} & 0_{2p-2,1} & 0_{2p-2,1} & 0_{2p-2,1} & 0_{2p-2,1} \\ 0 & 0 & 0_{1,2p-2} & 1 & 0 \\ 0 & 0 & 0_{1,2p-2} & 0 & 1 \end{bmatrix}$$

$$\Psi = \begin{bmatrix} 1 \\ 0 \\ 0_{2p-2,1} \\ 0 \\ 0 \end{bmatrix} \qquad \Pi = \begin{bmatrix} 0 & 0 & 0 \\ \frac{3}{4} & \frac{1}{2} & \frac{1}{4} \\ 0_{2p-2,1} & 0_{2p-2,1} & 0_{2p-2,1} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

3.8 Appendix 3 - Full information Bayesian MCMC

TABLE 2: Prior Distribution for DSGE Model Parameters

NAME	RANGE	DENSITY	STARTING VALUE	MEAN	SD
Model DSGE-VAR$(\lambda, 1)$					
α_1	$\mathbb{R}$	*Normal*	−0.075	−0.075	0.750
β_1	$\mathbb{R}$	*Normal*	0.200	0.200	0.500
μ	$\mathbb{R}$	*Normal*	0.000	0.000	1.000
τ	$\mathbb{R}$	*Normal*	0.000	0.000	1.000
σ_ε	$\mathbb{R}^+$	*Inv.Gamma*	0.500	0.500	0.500
σ_η	$\mathbb{R}^+$	*Inv.Gamma*	0.500	0.500	0.500
Model DSGE-VAR$(\lambda, 2)$					
α_1	$\mathbb{R}$	*Normal*	0.000	0.000	0.750
β_1	$\mathbb{R}$	*Normal*	0.100	0.100	0.650
α_2	$\mathbb{R}$	*Normal*	−0.750	−0.750	1.250
β_2	$\mathbb{R}$	*Normal*	0.000	0.000	0.650
μ	$\mathbb{R}$	*Normal*	0.000	0.000	1.000
τ	$\mathbb{R}$	*Normal*	0.000	0.000	1.000
σ_ε	$\mathbb{R}^+$	*Inv.Gamma*	0.500	0.500	0.500
σ_η	$\mathbb{R}^+$	*Inv.Gamma*	0.500	0.500	0.500
Model DSGE-VAR$(\lambda, 3)$					
α_1	$\mathbb{R}$	*Normal*	0.100	0.100	0.750
β_1	$\mathbb{R}$	*Normal*	0.100	0.100	0.650
α_2	$\mathbb{R}$	*Normal*	−0.800	−0.800	1.250
β_2	$\mathbb{R}$	*Normal*	−0.100	−0.100	0.700
α_3	$\mathbb{R}$	*Normal*	−0.150	−0.150	0.700
β_3	$\mathbb{R}$	*Normal*	0.100	0.100	0.450
μ	$\mathbb{R}$	*Normal*	0.000	0.000	1.000
τ	$\mathbb{R}$	*Normal*	0.000	0.000	1.000
σ_ε	$\mathbb{R}^+$	*Inv.Gamma*	0.500	0.500	0.500
σ_η	$\mathbb{R}^+$	*Inv.Gamma*	0.500	0.500	0.500
Model DSGE-VAR$(\lambda, 6)$					
α_1	$\mathbb{R}$	*Normal*	0.250	0.250	0.650
β_1	$\mathbb{R}$	*Normal*	−0.050	−0.050	0.600
α_2	$\mathbb{R}$	*Normal*	−0.800	−0.800	1.450
β_2	$\mathbb{R}$	*Normal*	−0.150	−0.150	0.650
α_3	$\mathbb{R}$	*Normal*	−0.250	−0.250	0.750
β_3	$\mathbb{R}$	*Normal*	0.050	0.050	0.500
α_4	$\mathbb{R}$	*Normal*	−0.200	−0.200	0.750
β_4	$\mathbb{R}$	*Normal*	−0.050	−0.050	0.450
α_5	$\mathbb{R}$	*Normal*	−0.200	−0.200	0.650
β_5	$\mathbb{R}$	*Normal*	−0.100	−0.100	0.500
α_6	$\mathbb{R}$	*Normal*	−0.050	−0.050	0.850
β_6	$\mathbb{R}$	*Normal*	0.000	0.000	0.450
μ	$\mathbb{R}$	*Normal*	0.000	0.000	1.000
τ	$\mathbb{R}$	*Normal*	0.000	0.000	1.000
σ_ε	$\mathbb{R}^+$	*Inv.Gamma*	0.500	0.500	0.500
σ_η	$\mathbb{R}^+$	*Inv.Gamma*	0.500	0.500	0.500

Bibliography

[1] Andrea Carriero. Explaining us-uk interest rate differentials: A reassessment of the uncovered interest rate parity in a bayesian framework. *Oxford Bulletin of Economics and Statistics*, 68(s1):879–899, December 2006.

[2] Andrea Carriero. A bayesian framework for the expectations hypothesis. how to extract additional information from the term structure of interest rates. Working Papers 591, Queen Mary, University of London, Department of Economics, 2007.

[3] D.N. DeJong, B.F. Ingram, and C.H. Whiteman. A bayesian approach to dynamic macroeconomics. *Journal of Econometrics*, 98:203–223, 2000.

[4] Marco Del Negro, Frank Schorfheide, Frank Smets, and Rafael Wouters. On the fit of new keynesian models. *Journal of Business & Economic Statistics*, 25:123–143, April 2007.

[5] Beth F. Ingram and Charles H. Whiteman. Supplanting the 'minnesota' prior: Forecasting macroeconomic time series using real business cycle model priors. *Journal of Monetary Economics*, 34(3):497–510, December 1994.

[6] R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90:773–795, 1995.

[7] Marco Del Negro and Frank Schorfheide. Priors from general equilibrium models for vars. *International Economic Review*, 45(2):643–673, 05 2004.