HAL
open science

Do fiscal variables affect fiscal expectations? Experiments with real world and lab data

Michele Bernasconi, Oliver Kirchkamp, Paolo Paruolo

To cite this version:

Michele Bernasconi, Oliver Kirchkamp, Paolo Paruolo. Do fiscal variables affect fiscal expectations? Experiments with real world and lab data. Journal of Economic Behavior and Organization, 2009, 70 (1-2), pp.253. 10.1016/j.jebo.2008.11.002 . hal-00688186

HAL Id: hal-00688186

https://hal.science/hal-00688186v1

Submitted on 17 Apr 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

Title: Do fiscal variables affect fiscal expectations? Experiments with real world and lab data

Authors: Michele Bernasconi, Oliver Kirchkamp, Paolo Paruolo

PII: S0167-2681(08)00220-5
DOI: doi:10.1016/j.jebo.2008.11.002
Reference: JEBO 2280

To appear in: *Journal of Economic Behavior & Organization*

Received date: 5-9-2005
Revised date: 7-2-2008
Accepted date: 10-11-2008

Please cite this article as: Bernasconi, M., Kirchkamp, O., Paruolo, P., Do fiscal variables affect fiscal expectations? Experiments with real world and lab data, *Journal of Economic Behavior and Organization* (2008), doi:10.1016/j.jebo.2008.11.002

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Do fiscal variables affect fiscal expectations? Experiments with real world and lab data*

Michele Bernasconi‡ Oliver Kirchkamp† Paolo Paruolo‡

February 7, 2008

Abstract

We generate observable expectations about fiscal variables through laboratory experiments using real world data from several European countries as stimuli. We compare a VAR model of expectations for data that is presented in a fiscal frame with one for neutrally presented data. We test the validity of the setup and find that participants understand the meaning of the fiscal variables, but also that their ability to perceive the correct characteristics of fiscal policy is limited. Expectations are consistent neither with rational nor with purely adaptive expectations, but instead follow an augmented-adaptive scheme.

Keywords: Experiments, fiscal policy, expectations, panel data.

JEL classification: C91, D89, E62, H31.

*Financial support from Deutsche Forschungsgemeinschaft through Sonderforschungsbereich 504 and partial financial support from Italian MIUR grants ex 60% is gratefully acknowledged. We thank the associate editor and two anonymous referees for valuable comments and insights which helped us when we prepared the revision of the paper.

†Corresponding autor: Oliver Kirchkamp, Friedrich-Schiller-University Jena, School of Economics and Business Administration, Carl-Zeiss-Str. 3, 07743 Jena, Germany. Phone: +49-3641-943241, Fax: +49-3641-943242, oliver@kirchkamp.de

‡Department of Economics, University of Insubria, Via Ravasi 2, I-21100 Varese, Italy. michele.bernasconi@uninsubria.it, paolo.paruolo@uninsubria.it

Accepted Manuscript

ACCEPTED MANUSCRIPT

1 Introduction

Expectations on fiscal variables are crucial to understand the effect of fiscal policy on the private sector. Little, however, is known about the way people actually form expectations on fiscal variables. While many models are based on the hypothesis of rational expectations, expectations themselves are not easily observable. As a result, empirical tests of the rational expectations hypothesis are often indirect, involving tests of predicted relations between observables, such as relationships between fiscal variables and components of output.

A problem in this approach is that factors that are not part of the model may affect expectations and choices in real economies. This makes the identification of the effects of expectations difficult because "economists cannot observe all the data that economic agents do" (Seater 1993, p. 164).

This limitation is also relevant for expectation measures derived from opinion surveys[1]. Moreover surveys suffer from lack of economic incentives to reveal true opinions, so that for various reasons respondents "may express judgments that are different from the ones they choose to act upon" (Pesaran 1987, p. 209).

An alternative is to measure expectations in a controlled experiment. Expectations experiments have been conducted since as early as the sixties in a wide range of economic contexts[2]. Many of these experiments ask participants to form expectations for artificial time series that are based on an abstract model. The advantage of using artifical time series is that the data generating process is known to the experimenter and can immediately be used as a benchmark. It is possible to find out in a straightforward way under which conditions participants form which kinds of expectations for a

[1]For example, surveys conducted by Grun (1991) and Allers, de Haan, and de Kam (1998) found widespread evidence of misinformation on the conduct of government fiscal policy.

[2]Such experiments include expectations for prices (Fisher 1962 and Schmalensee 1976), expectations about artificially constructed time series (Hey 1994), expectations in normal-form games (e.g., Costa-Gomes and Crawford 2001), about the provision of public-goods (Offerman, Sonnemans, and Schram 1996, and Croson 2000), expectations in the cobweb economy (Hommes, Sonnemans, Tuinistra, and van de Velden 2007), on monetary policies (e.g. Marimon and Sunder 1993 and Bernasconi and Kirchkamp 2000) or even fiscal policy (Swenson 1997).

Accepted Manuscript

ACCEPTED MANUSCRIPT

given abstract model. This convenience comes at a cost: Within this framework one can study expectations only for the given abstract model. First, one has to assume that participants in an experiment understand the formal description of the abstract model, and second, one has to assume that decision makers use similar procedures when they form expectations for different, and not perfectly known, processes in the real world. To bridge the gap between abstract games in the laboratory and economic reality, experimenters can calibrate a specific model with parameters from the real world. Experiments on tax compliance (Alm and McKee 2004) or on decisions over monetary policy (Lombardelli, Talbot, and Proudman 2002, Blinder and Morgan 2005) follow this approach.

In this paper we present a laboratory experiment that is not based on an abstract model with parameters from the real world but on real time series. Compared with experiments where participants forecast abstract time series, the use of real world time series has advantages and disadvantages. An advantage is that we can observe expectations in an economic context with real economic data. If the process of expectation formation depends on the economic context or on the type of the data, then the expectations we can measure are closer to the ones that are relevant for real economic decisions. A disadvantage is that with real world time series, it is no longer obvious what rational expectations are. We no longer have the clean benchmark abstract time series would provide. Instead, we have to construct a benchmark with the help of an econometric model. To do that, we follow a standard approach and assume that the DGP can be described as a VAR. The estimated VAR of the field data will constitute our benchmark solution.

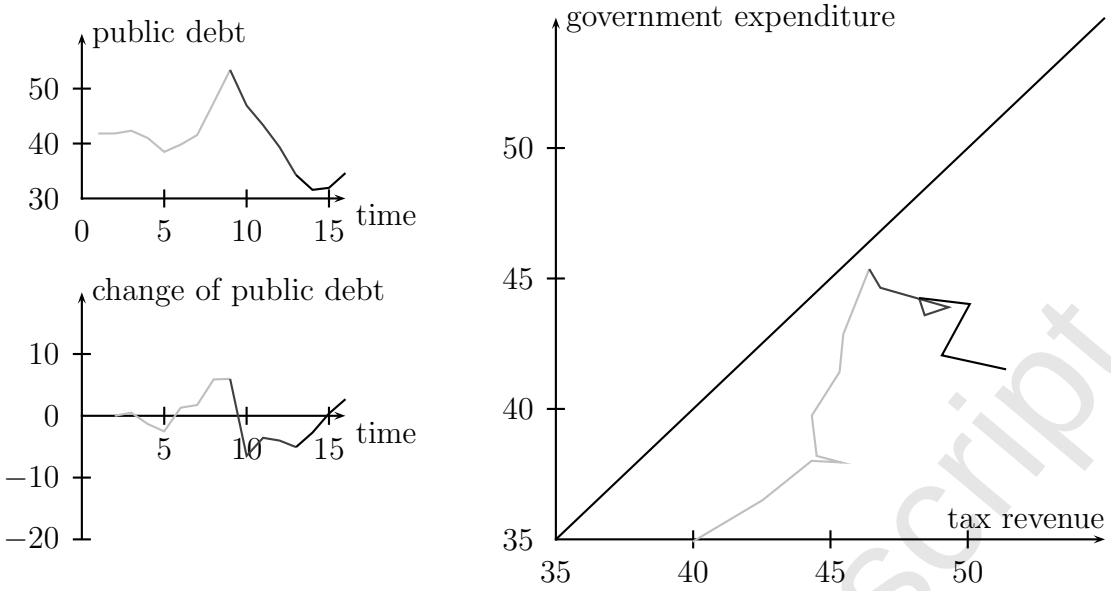
We use annual time series on taxes and public expenditures from 15 OECD countries. Given the novelty of the approach we carefully check the internal validity of the experiment. We also study the impact of learning, fatigue, gender, and field of study of the participants.

To understand and to structure the expectation formation process better we assume that the joint DGP of stimuli and expectations data can be described as a VAR. We find that participants fail to perceive the fiscal properties from the field, thus violating the 'rational expectation hypothesis'. Expectations are rather in line with 'augmented-adaptive models', introduced in the early eighties by various authors (see Pesaran).

The paper is organised as follows. Section 2 describes the setup of the ex-

Accepted Manuscript

ACCEPTED MANUSCRIPT



Values are given as percentage of GDP. Gray lines indicate past, black lines indicate recent years.

FIGURE 1: T_t and G_t treatment

Accepted Manuscript

periment. Section 3 provides descriptive statistics of the results and discusses internal validity. Section 4.1 develops the econometric approach to analyse both the stimulus and the expectations data. Inference results are presented in Section 4.2. The last section summarises the external significance of the approach and of the findings.

2 Experimental setup

In the experiment participants are exposed to graphic representations of time series of annual fiscal variables taken from various European countries. Time is denoted t, stimuli are gross total taxes T_t, total public expenditure G_t (including interest payments), public debt B_t, and change in the debt level $\Delta B_t = B_t - B_{t-1}$, all expressed as yearly percentage of GDP. We focus on the relationship between taxes and expenditure, denoted by the vector $x_t := (T_t, G_t)'$. Here and in the following Δ is the time difference.

Participants know that they are facing European countries, but they do not know which specific country and which period the series refer to. Participants are paid in each period according to the precision of their expecta-

ACCEPTED MANUSCRIPT

tions. The better their expectations, the higher is their income. The payment scheme can be motivated as a utility that a hypothetical agent derives from consumption over the current and the previous period if this agent uses the expectations of our participant and maximises the utility function:

$$u_t = \prod_{i=t-1}^{t} \left(\frac{3}{4} C_i + \frac{1}{4} G_i \right) \tag{1}$$

subject to the budget constraint

$$\sum_{i=t-1}^{t} \underbrace{(1 - C_i - T_i)}_{\text{savings}} \cdot (1+r)^i = 0 \qquad \text{with } r = 0.1 \tag{2}$$

(where 1 is the normalised total GDP). Agents receive initial information on the first seven values of stimuli, which, for most countries, was the period from 1970 to 1976. Let $t-1$ be the last available year and $X_{t-1} := (x_1, \ldots, x_{t-1})'$ the available information; then forecasts are made repeatedly for each subsequent year t until the end of the time series is reached in 1998. Participants would then start with another, randomly selected country.

We compare three experimental treatments. In a baseline treatment participants forecast both T_t and G_t (see figure 1). Forecasts for the next period are made by clicking with the mouse directly into the diagrams.

To control whether participants understand the economic context of the baseline treatment we also run a treatment where this context is removed. The time series are the same as in the baseline treatment, but variables have no economic significance. They are simply called A, B, and C. We call this treatment the 'neutral' treatment.

A main aspect of this paper is to study how participants use information about one variable (e.g. T_t) to forecast another (e.g. G_t). A natural implementation would be to present two variables T_t and G_t and to ask participants to forecast both. This is what we do in the above two treatments. To check whether such a simultaneous forecast is too difficult, we also run a control treatment where participants forecast T_t only[3].

[3]In all three treatments subjects had information for past taxes and public expenditure. The display for the T_t only treatment was similar to that of the T_t and G_t treatment in

ACCEPTED MANUSCRIPT

Forecasts are called $T_t^{E_i}$ and $G_t^{E_i}$, where i indicates agent i and E stands for expectation. The forecasts of agent i that refer to time t are denoted $y_{i,t} := (T_t^{E_i}, G_t^{E_i})'$ in the T_t and G_t treatment and $y_{i,t} := T_t^{E_i}$ in the T_t only treatment.

The time series of the stimuli (see figure 1) are updated every period after forecasts are made, so subjects learn about realizations of the stimuli as the economy moves on. Participants also get feedback about their wage, which reflects the accuracy of their forecast[4].

Openly providing all payoff relevant information in economic experiments is standard and essential. Hence, we explain in the instructions (see appendix) the above payment scheme in details with examples and also with a short technical note. Subjects were not obliged to look at the technical note, but were told that they could consult it and check everything if they wished to control the experimenter.

Stimulus data were from 15 European countries: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, and UK[5]. For the majority of countries the sample period of stimulus data was 1970-98; a few exceptions (Denmark, France, Greece, Ireland, and UK) are due to limits in the availability of the fiscal time series. For all countries, expectations started after the seventh year of the stimulus (which was 1977 for most countries). A graph of the stimulus data for the different countries is shown in figure 2, with the basic statistics

figure 1, with the only difference that the values of taxes and public expenditure were both displayed on the y-axis, with the period of the observations (not the year, which remained unknown to subjects) on the x-axis.

[4]To reward participants, the computer constructs the optimal consumption level C_{t-1} based on a subjects' forecasts $y_{i,t}$ for year t and given eq. (1) and (2). When in the next period $x_t := (T_t, G_t)'$ becomes available, we use equations (2) and (1) to calculate C_t and the utility u_t. The participant's per minute wage is $w = 0.66 \cdot (u_t/u_t^*)^\eta$ where $\eta = 12000$ in the T_t and G_t treatment and 15000 in the T_t only treatment. u_t^* is the utility the participant would obtain with forecasting the true values. This transformation from utilities into wages is monotonic and, hence, does not affect the maximisation problem of the individual, but creates steeper incentives to make good forecasts. Participants are paid this wage up to two minutes for each forecast. Parameters of the payment scheme have been developed in pilot studies previous to the experiment.

[5]All stimulus data used in the experiment were taken from the database "Fiscal Positions and Business Cycles" (OECD 2000).

ACCEPTED MANUSCRIPT

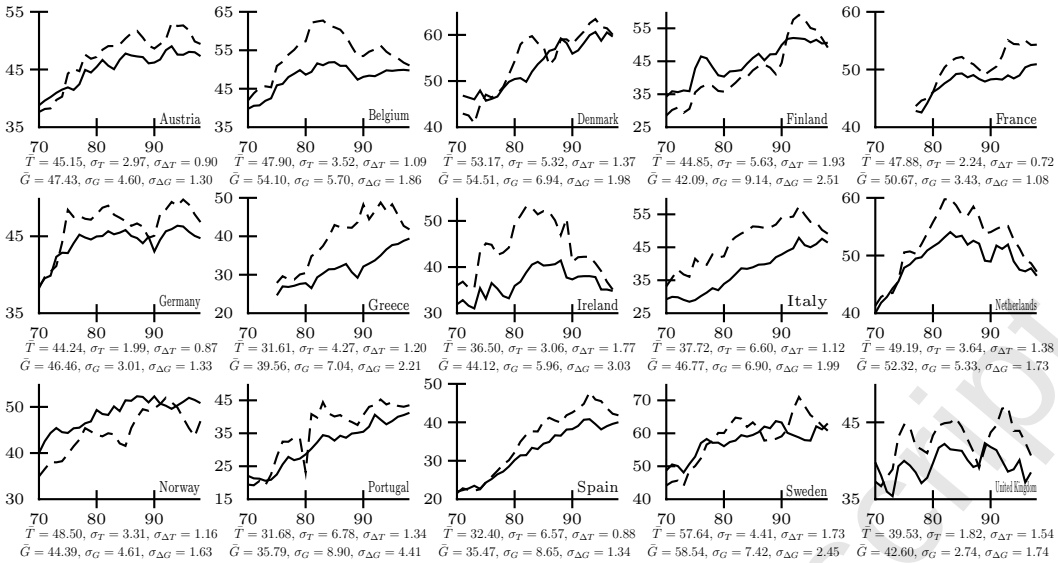


FIGURE 2: Stimulus data (T—, G - -)

of the two series (means, $\overline{G}$ and $\overline{T}$; standard deviations, σ_T and σ_G; and standard deviations of the rates of change, $\sigma_{\Delta T}$ and $\sigma_{\Delta G}$).

In total, 170 subjects participated in the experiment. Each subject could participate only in one treatment: 76 took part in the baseline T_t and G_t treatment, 69 in the neutral treatment, and 25 in the T_t only treatment.

Each participant made forecasts for a sequence of countries. Sequences were determined randomly and independently for all participants. Table 1 summarises the number of participants playing the various countries in the three treatments.

The experiment was run at the experimental laboratory of the SFB 504 in Mannheim. Participants were students at the university: 116 were from economics or business (we will call them economists), 38 from political science, law and languages (in the following we will call these students in other humanities), and 12 from natural sciences. 130 were male and 40 were female. Sessions were conducted individually and lasted for about 2 hours. Participants made, on average, 136 forecasts (between 28 and 309) and completed on average one forecast every 39 seconds.

Accepted Manuscript

ACCEPTED MANUSCRIPT

Country	Sample period	number of participants / country		
		T_t and G_t	T_t only	neutral
Austria	1970-1998	34	14	28
Belgium	1970-1998	52	20	48
Denmark	1971-1995	43	14	43
Finland	1970-1998	33	12	24
France	1977-1998	29	10	31
Germany	1970-1998	14	13	5
Greece	1975-1998	26	13	22
Ireland	1970-1995	45	21	44
Italy	1970-1998	35	16	29
Netherlands	1970-1998	32	9	23
Norway	1970-1998	32	14	26
Portugal	1970-1998	25	13	23
Spain	1970-1998	30	11	24
Sweden	1970-1998	53	16	41
United Kingdom	1970-1995	27	15	19
Total		510	211	430

Since participants made forecasts for more than one country, the total number is not the sum of the number of participants in all countries.

Table 1: Countries played by participants across treatments

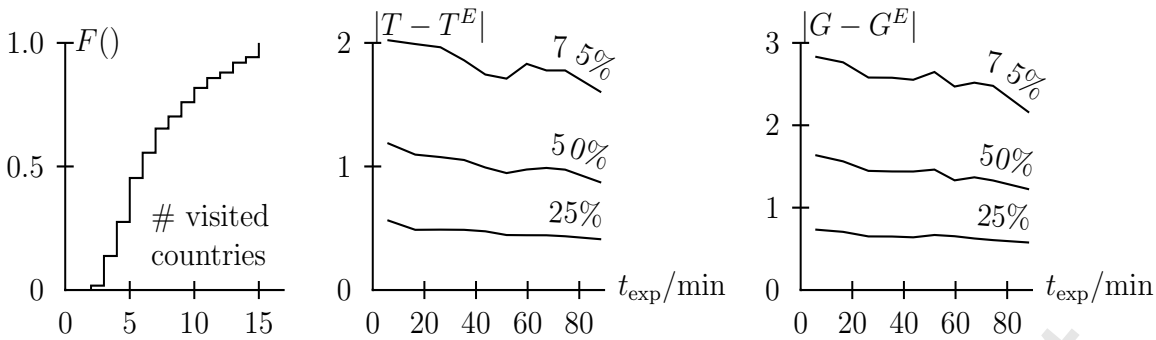
3 Descriptive analysis of the evidence and internal validity

Laboratory experiments are often inspired by the real world and often use in their setup parameters that are based on field data. Using not only parameters, but datasets from the field in the laboratory is novel and, we believe, innovative, but poses various questions. A first one is about the internal validity of the experimental setup.

Learning and Fatigue The longer an experiment lasts, the more time participants have to learn and to understand their task. However, during a long experiment participants might have difficulties in concentrating; thus the quality of observations might eventually decrease. Figure 3 shows the development of absolute forecast errors over time. We see that forecast errors decrease over the course of the experiment. The effect of learning seems to dominate fatigue all the time. Though the change over time is rather small,

Accepted Manuscript

ACCEPTED MANUSCRIPT



The left graph shows the cumulative frequency of the number of countries our participants visit during the experiment; the middle and the right graph show quantiles for forecast errors during the experiment. Quantiles are taken over ten bands with the same number of observations. The horizontal axis show the total time in minutes since the start of the experiment.

FIGURE 3: Countries, learning and fatigue during the experiment

the decreasing trend is significant[6].

Decision time The left graph in figure 4 shows the cumulative frequency of time subjects take for individual decisions. The mean decision time (over all decisions) is about 35 seconds for each forecast; the median (over all participants) is 25 seconds. While risk loving participants might have a preference to use the maximal forecast time of 2 minutes, our participants seem to prefer to share their risk over different forecasts. The two remaining graphs in figure 4 show the relation between forecast errors and decision time. The relationship is, if at all, weak[7]. Harder decisions (characterised by their larger forecast errors) seem to take only slightly more time.

Participants' characteristics The left part of figure 5 shows the cumulative distribution of median absolute forecasts errors $|T - T^E|$ and $|G - G^E|$ per participant for male and female participants. The difference is small and not significant[8]. The centre part of figure 5 shows the cumulative distribu-

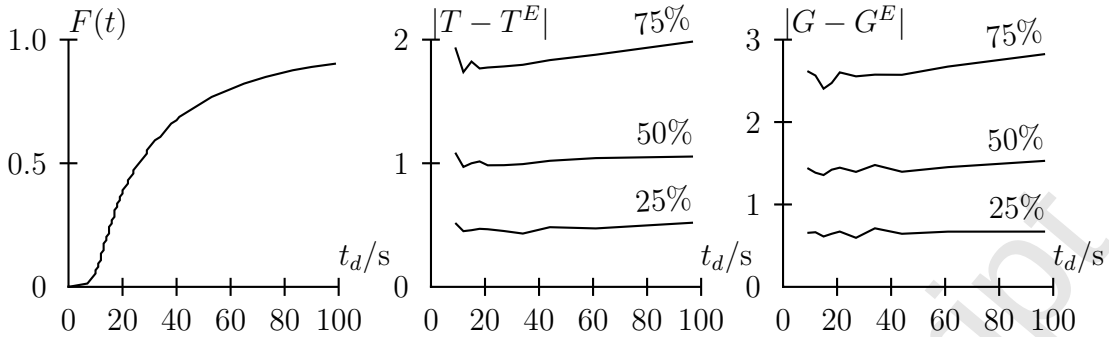
[6] Comparing average errors (per participant) in the first and second half of the experiment with a pairwise Wilcoxon test rejects no or an increasing trend with a p-value of 0.0003 for T and 0.0015 for G. A linear robust model that allows for correlated errors within the same participant rejects with a p-value $< 10^{-7}$.

[7] Separating decision times at the median and comparing average errors as in footnote 6 shows only an insignificant change. A linear robust model yields a p-value of 0.098 for T and 0.077 for G.

[8] Comparing median absolute errors with a Wilcoxon rank sum test yields a p-value of 0.8024 for T and of 0.1235 for G. A t-test yields a p-value of 0.8583 for T and of 0.2841 for

Accepted Manuscript

ACCEPTED MANUSCRIPT



Quantiles are taken over ten bands with the same number of observations.

FIGURE 4: Distribution of decision time in seconds and quantiles of forecast errors over the decision time

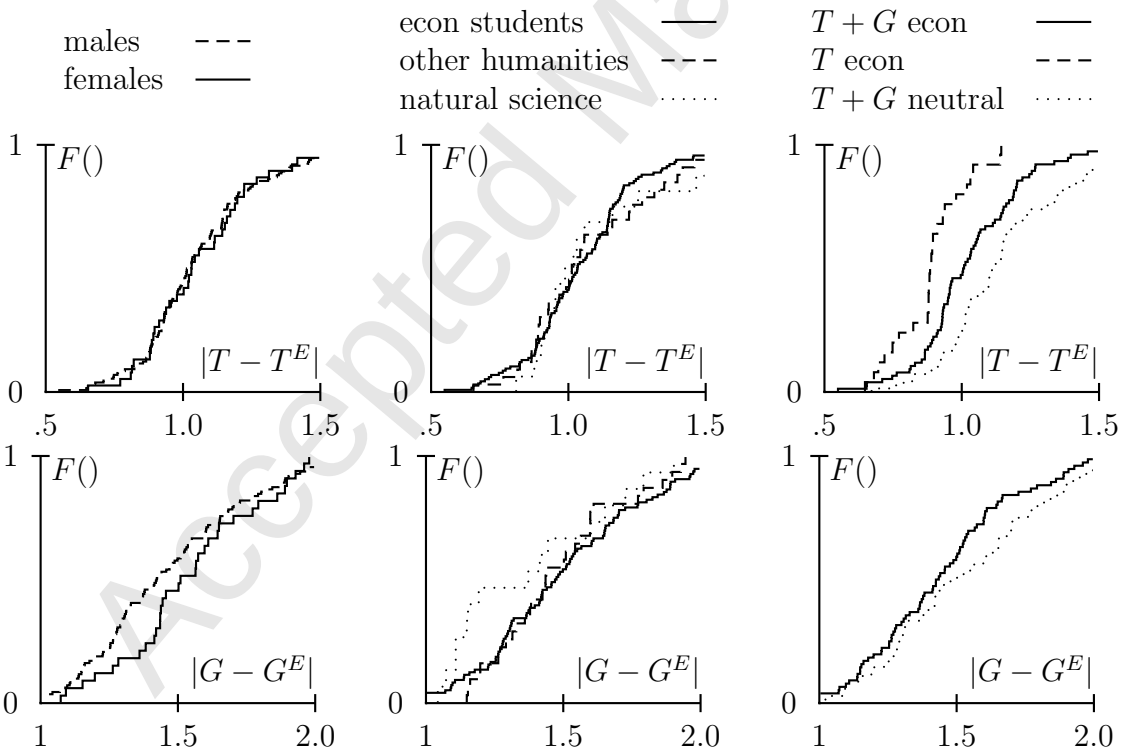


FIGURE 5: Distribution of average individual forecast errors by characteristics of participants and by treatment

ACCEPTED MANUSCRIPT

tion of median absolute forecasts errors per participant for different fields. Students from natural sciences seem to make slightly better forecasts for G than the other students. However, these are only 12 out of 170 participants, and the differenc between the different fields is small and not signficant both for G and for T[9].

Economic context To check whether participants understand the fiscal frame, we compare the T_t and G_t treatment with the neutral treatment. If participants understand and make use of the economic context, we should observe smaller forecast errors in the T_t and G_t treatment than in the neutral treatment. The right diagram in figure 5 shows cumulative distributions of forecast errors for these treatments. We see that, as expected, forecast errors are larger in the neutral treatment; adding economic context contributes significantly to the understanding of the experiment[10]. We also see that a simpler task that requires forecasts only for T makes forecasts more precise for this variable.

Summary regression Figures 3, 4, and 5 should give a rough impression of the determinants of sizes of forecast errors. To control for characteristics of participants better and the random allocation of countries to participants, table 2 presents estimates from linear regressions that explain the squared forecast error as a function of various dummies. Model 1 includes a dummy for the type of treatment (neutral / economic context) and a dummy for the forecasted variables (T-only / T and G). Model 2 adds a dummy for gender (female is the reference group) and two dummies for the subject of studies (economics is the reference group). Model 3, finally, adds dummies for each country.

Only two effects are significant. Both were already visible in figure 5. First, reducing the number of variables to forecast in the T_t only treatment

G. Using mean squared errors instead of medians does not change the result considerably.

[9]Comparing median absolute errors with an F-test yields a p-value of 0.7549 for T and 0.3243 for G; a non-parametric Kruskal-Wallis test yields a p-value of 0.997 for T and 0.3110 for G.

[10]Comparing median absolute errors with a non-parametric Wilcoxon rank sum test yields a p-value of 0.0000075 for T and 0.09715 for G. A t-test yields a p-value of 0.00001 for T and 0.0488 for G.

Accepted Manuscript

ACCEPTED MANUSCRIPT

	$(T-T^E)^2$			$(G-G^E)^2$		
	model 1 $\hat{\beta}/(\hat{\sigma})$	model 2 $\hat{\beta}/(\hat{\sigma})$	model 3 $\hat{\beta}/(\hat{\sigma})$	model 1 $\hat{\beta}/(\hat{\sigma})$	model 2 $\hat{\beta}/(\hat{\sigma})$	model 3 $\hat{\beta}/(\hat{\sigma})$
neutral	0.560* (0.219)	0.604** (0.214)	0.561** (0.207)	1.071* (0.472)	1.360** (0.469)	1.054** (0.346)
T_t only treament	-0.651*** (0.134)	-0.608*** (0.138)	-0.498*** (0.121)			
male		-0.080 (0.169)	-0.055 (0.156)		-0.950 (0.543)	-0.377 (0.388)
other humanities		-0.175 (0.181)	-0.094 (0.166)		-0.251 (0.627)	0.129 (0.453)
natural sciences		0.382 (0.480)	0.502 (0.458)		-1.106 (0.693)	0.028 (0.591)
country dummies			not shown			not shown
constant	2.777*** (0.106)	2.906*** (0.325)	1.117*** (0.299)	6.369*** (0.311)	8.005*** (1.053)	2.151* (0.827)

Standard deviations are shown in parentheses and are calculated using a robust regression which takes into account correlations of observations for each participant. *, **, and *** denote significance at 5%, 1%, and 0.1% level. Similar results can be obtained also for median absolute errors.

Table 2: Estimation of squared forecast errors

also reduces the error. Second, withholding information about the economic context increases the error. The latter shows that participants understand the economic context of our experiment, at least when we present variables with their economic meaning.

External significance Another important question about the experimental setup concerns its external significance. Consumers in the field form expectations on fiscal policy not only on the basis of spending and taxes, but also through the composition of the government, the date of the next election, economic development, foreign policy, and natural disasters, to name only a few. In the experimental laboratory we are able to neglect all these factors and study T and G in isolation. Any systematic change of expectations that we observe in the lab must be due to a change in T or G.

The choice of factors that we study here, T and G, is standard in the empirical literature which studies fiscal conduct of countries. Below we refer to an important stream of literature that, since at least the mid-1980s, has used cointegration techniques and the econometrics of Granger causality to

Accepted Manuscript

ACCEPTED MANUSCRIPT

study the dynamics of the public budget intertemporal constraint[11]. In the next section we develop an empirical model of stimuli and of expectation data that coherently ties the present experiment to such econometric practice. Through it we discuss the meaning and implication of general expectation schemes (rational, adaptive, augmented adaptive), we consider and estimate properties of field data, and we check how these features are perceived by subjects in the experiment.

4 An econometric analysis of the field-lab data: expectations

In this section we want to study the model of expectation formation used by participants and the perceived relationships among fiscal variables. We emphasize again that, unlike previous experiments of expectational models that are based on an artificial data generating process, we are using here data from the field. Using an artificial data generating process has the advantage of providing a clear benchmark solution. A disadvantage is that it might be difficult to explain such a process to participants in an experiment. Another problem is that one observes forecasts for abstract time series when one wants to describe forecasting behaviour for real time series. Using data from the field may make it more likely that the forecasting rules participants use in the experiment are close to the rules they would use otherwise with field data. Using data from the field also means that we lose the obvious benchmark solution. With timeseries that are based on a known abstract model, it would be clear what rational expectations have to be, but what are rational expectations when we use real world data? We now no longer know the benchmark solution; instead we have to estimate one. We acknowledge that this is far from perfect since econometric techniques used to ascertain properties of real data are often less than perfect. Our approach is based on standard techniques used since at least the mid-1980s by economists and econometricians to study the dynamics of fiscal policies. We assume that the data generation processes for both the field and the lab can be taken to be a Vector Autoregressive Process (VAR). We first review the approach

[11]See section 4.1 below for various references to the literature.

ACCEPTED MANUSCRIPT

and adapt it to this experiment[12]; then we present the evidence based on its estimation.

4.1 The VAR model of the field-lab data

The VAR for the field data (stimuli) The field model concerns the stimuli data, that is the vector $x_t := (T_t, G_t)'$, treated for each country separately. A standard VAR representation of the field systems is given by

$$\Delta x_t = \mu_x + \alpha_x \beta'_{1x} x_{t-1} + \sum_{\ell=1}^{k-1} \Gamma_{\ell,xx} \Delta x_{t-\ell} + \epsilon_{xt} \,. \tag{3}$$

The expression takes the stimuli data $x_t := (T_t, G_t)'$ as integrated of order 1, I(1); μ_x is a vector of constants; $\beta'_{1x} x_{t-1}$ is an error correction term (cointegrating relation CI) expressing the long run disequilibrium between taxes and expenditure in the field; $\alpha_x = (\alpha_T, \alpha_G)'$ is the associated vector of the adjustment coefficients; $\Gamma_{\ell,xx}$ are the matrices of the coefficients of partial adjustments; and ϵ_{xt} are error terms i.i.d. $N(0, \Omega_{xx})$.

Versions of this basic model have been analysed extensively in the literature (see, e.g. Trehan and Walsh 1991, Ahmed and Rogers 1995). An issue considered is sustainability of fiscal policy which can be checked by way of cointegration analysis between taxes and public expenditure. In short, cointegration tests are based on the idea that solvency requires the budget deficit to be stationary. Even when T_t and G_t are non-stationary, there may be linear combinations of the form $T_t + \gamma G_t + \rho$ (where γ and ρ are constants) which are stationary. In this case, T_t and G_t are said to be cointegrated of order 1, with cointegrating vector (CI) $(1, \gamma)$. Stationarity of the budget then further requires $\gamma = -1$ (homogeneity condition).

A second empirical issue analysed within equation (3) concerns the direction of causality between taxes and public expenditure[13]. In particular, when the series in $x_t := (T_t, G_t)'$ are cointegrated, at least one of T_t and G_t adjusts to disequilibrium with respect to the long run relation. Four cases

[12]See for example Johansen (1995) for a modern treatment of the literature on VAR. See our technical report (Bernasconi, Kirchkamp, and Paruolo 2004) for a more detailed exposition of how the field and lab VARs can be consistently derived in the present experiment.

[13]See Garcia and Henin (1999) and Payne (1998) for reviews and recent empirical evidence on this classical theme of public finance.

Accepted Manuscript

ACCEPTED MANUSCRIPT

may then emerge: that taxes cause spending, that spending causes taxes, bidirectional causality, and lack of causality. These cases can be represented as a vector $\alpha_x = (\alpha_T, \alpha_G)'$ of the forms $(0, *)'$, $(*, 0)'$, $(*, *)'$ and $(0, 0)'$ (with $*$ indicating a non-zero coefficient), thus providing tests of Granger long-run causality.[14] Short run-causality can be similarly checked by looking for significant off-diagonal coefficients in the matrices $\Gamma_{\ell,xx}$.

The VAR for the lab data (expectations) For the lab model, let m indicate the number of individuals, $i = 1, \ldots, m$, in a treatment. Recall that expectations $y_{i,t} = (T_t^{E_i}, G_t^{E_i})'$ in the T_t and G_t treatment and $y_{i,t} = T_t^{E_i}$ in the T_t only treatment. We now illustrate the model for $y_{i,t} := (T_t^{E_i}, G_t^{E_i})'$, which is applied with the obvious modifications to the T_t only treatment.

For individual i the VAR has the following form:

$$\begin{pmatrix} \Delta T_t^{E_i} \\ \Delta G_t^{E_i} \end{pmatrix} = \mu_i + \alpha_i \begin{pmatrix} T_{t-1} + \widehat{\gamma} G_{t-1} \\ T_{t-1}^{E_i} - T_{t-1} \\ G_{t-1}^{E_i} - G_{t-1} \end{pmatrix} + \sum_{\ell=1}^{k-1} (\Gamma_{\ell,ii}, \Gamma_{\ell,ix}) \begin{pmatrix} \Delta y_{i,t-\ell} \\ \Delta x_{t-\ell} \end{pmatrix} + \widehat{\varepsilon}_{yt} \,. \tag{4}$$

Equation (4) includes a vector of constant terms $\mu_i = (\mu_{iT^E}, \mu_{iG^E})'$; a matrix (2×3) of adjustment coefficients α_i, decomposed as $\alpha_i := (\alpha_{i1}, \alpha_{i2})$, where $\alpha_{i1} = (\alpha_{iT^E}, \alpha_{iG^E})'$ shows how the expectation variables may react to the disequilibrium vector estimated in the field, $T_{t-1} + \widehat{\gamma} G_{t-1}$ (also labeled $_{\text{ECM}}TG$); while $\text{vec}(\alpha_{i2}')' = (\alpha_{iT^E(T^ET)}, \alpha_{iT^E(G^EG)}, \alpha_{iG^E(T^ET)}, \alpha_{iG^E(G^EG)})$ includes the responses to two further possible CI relations given by the errors in expectations $(T_{t-1}^{E_i} - T_{t-1}, G_{t-1}^{E_i} - G_{t-1})'$ (also called $(_{\text{ECM}}T^ET, {}_{\text{ECM}}G^EG)'$); $\Gamma_{\ell,ix}$ captures causality from the field to the lab, and $\Gamma_{\ell,ii}$ causality from past expectations to present ones.

We should note that for participants who perceive the fiscal properties from the field the lab system (4) collapses to the specification of the marginal field system (3). We use the field system (3) as one reference point and say that participants whose expectations follow (3) have *rational expectations*. We will also investigate whether and how participants pick up the actual cointegration by comparing $\alpha_{i1} = (\alpha_{iT^E}, \alpha_{iG^E})'$ from the lab with $\alpha_x = (\alpha_T, \alpha_G)'$ from the field. Also of interest is the inference on the vector of constant

[14] Lack of causality (i.e. $\alpha_x = (0, 0)'$) is incompatible with the presence of cointegration in (3) (see Johansen 1995).

ACCEPTED MANUSCRIPT

terms μ_i, indicating subjects' perception of trends in the data.

One alternative class of expectation formation processes is the adaptive scheme, as originated in the 1950s by the works of different scholars (see Pesaran 1987 for references). We emphasize the bivariate nature that can take here the adaptive schemes. In particular, under *univariate* or *purely adaptive* expectations, past values of the forecasted variable x_t^1 and of its forecast $y_{i,t}^1$ enter the expectation process for $y_{i,t}^1$, where a superscript 1 indicates one stimulus variable and the corresponding forecast. If other variables enter in the estimated equation for $\Delta y_{i,t}^1$, this is evidence against a purely adaptive scheme and in favour of a more general class of models known as *augmented-adaptive* (see again Pesaran 1987 for the many variations nested within this definition)[15]. In augmented-adaptive schemes one can also address questions of perceived causality between taxes and public expenditure[16].

The individual lab sub-systems (4) may be estimated one at the time or jointly. Joint estimation under some homogeneity restrictions allows us to exploit the panel dimension of the data to increase efficiency. In the empirical analysis we will first estimate for each country a model of a representative-agent's expectations. This model we will compare with the corresponding country field model. In this comparison we will focus on the expectations obtained from the experiment conducted under the fiscal frame. Then we will compare results from different treatments, including experiments conducted under the neutral frame. We will also reconsider the homogeneity condition and we will present estimations at the individual level.

[15]In most previous expectation experiments, subjects had to forecast univariate time series (as for example in Schmalensee 1976 or Hey 1994) or reduced-univariate schemes arising from augmented-adaptive forms (as in experiments of price expectations in classical cobweb economies, for example Hommes, Sonnemans, Tuinistra, and van de Velden 2007 or on inflation expectations in overlapping generation economies; see Duffy 1998 for a review). They generally found support for adaptive expectations (though not necessarily of the first order, see, e.g. Hey).

[16]In particular, by inspection of the parameters in (4) one should note the following: the α_{i1} coefficients and the off-diagonal elements in the $\Gamma_{\ell,ix}$ matrices determine the direction of causality from the field to the lab, while the α_{i2} coefficients and the off-diagonal elements in the $\Gamma_{\ell,ii}$ matrices regulate the ones from the past expectations on present expectations.

ACCEPTED MANUSCRIPT

Country	Cointegrating Vector (β'_{1x}, ρ)	Test of $\widehat{\gamma} = -1$ homogeneity condition
Austria	$(1, -0.739, -10.385)$	-13.23^{***}
Finland	$(1, -0.565, -21.436)$	-28.29^{***}
Germany	$(1, -0.572, -17.863)$	-4.32^{**}
Italy	$(1, -0.892, 0)$	-1.72
Netherlands	$(1, -0.606, -17.630)$	-13.51^{***}
Norway	$(1, -1.051, 0)$	-3.52^{**}
Portugal	$(1, -1.177, 13.667)$	-2.09^{*}
Sweden	$(1, -0.950, 0)$	-2.13^{*}
UK	$(1, -0.922, 0)$	-9.18^{***}

Belgium, Denmark, France, Greece, Ireland, and Spain are not shown in the table since they show no cointegration. *, **, *** denote 10%, 5%, 1% significance levels.

TABLE 3: Cointegrating vectors estimated for stimulus data

4.2 Evidence

4.2.1 Comparison between the field and the lab VARs

Table 3 reports the results of the cointegration analysis on the stimulus data[17]. For 9 countries (Austria, Finland, Germany, Italy, Netherlands, Norway, Portugal, Sweden, and UK), we find that taxes and public expenditures are cointegrated; for 6 (Belgium, Denmark, France, Greece, Ireland, and Spain) we find that they are not. Among the former and consistently with the evidence reported in the literature (see, e.g., Manasse 1996), the condition for stationarity of the budget $\widehat{\gamma} = -1$ is rejected for most countries. The estimated CI vectors have the form (β'_{1x}, ρ) with $\beta'_{1x} x_t = T_t + \gamma G_t$ and $\mu_x = \alpha_x \cdot \rho$, which exclude linear trends in the data (see Johansen 1995). This feature of field data is intuitive[18] and indeed passed for all countries in

[17]We follow a standard procedure to test for cointegration in system (3). We start determining the lags order k of the field VAR (which was two for most countries) and then use the Johansen procedure to test for the rank r of the system (that is, the number of CI vectors). We use the LR trace test for $H_0 : r = 0$ versus $H_1 : r = 1$ and exclude the case of a stationary system $r = 2$, where both taxes and expenditure are stationary in levels.

[18]Since x_t contains the ratio of taxes and expenditure to total GDP, we expect the system $x_t := (T_t, G_t)'$ not to contain a linear trend.

Accepted Manuscript

ACCEPTED MANUSCRIPT

the sample. Below we will check how subjects reflect this feature of the real data in their expectations.

Table 4 summarises the main findings of the inference on both the field and the lab systems in the two experimental treatments under the fiscal frame. All parameters reported in the table are significant at least at a 5% level[19]. In considering the results, recall that in the T_t only treatment, in which agents forecast taxes, expectations on $G_t^{E_i}$ are not available.

The leftmost part of the table reports, for the field and for the lab, the coefficients of responses to the field error correction term ${}_{\text{ECM}}TG = T_t + \widehat{\gamma} G_t$. Consider first the field evidence (the first column in the table). Among the nine countries for which cointegration between taxes and expenditure was found, expenditure is long-run adjusting to taxes in four cases (Austria, Norway, Sweden and UK); taxes are adjusting to expenditure in four others (Finland, Germany, Italy and Netherlands); and in one country (Portugal) adjustment is bidirectional.

Results from the experiments show that subjects do not follow the CI characteristics of the field data. In both experimental treatments the lab responses α_{i1} to ${}_{\text{ECM}}TG$ are often not significantly different from zero, and even when they are, α_{i1} are in any case quite small. On the contrary, the vectors of constants estimated in the lab show that, while subjects do not follow the CI characteristics of the real-data, they perceive for most countries a linear trend which is not found by the field VARs.

In the rightmost part of table 4 we show the results of tests on Granger causality in the field and compare them with tests of perceived causality in the lab. Results speak strongly against subjects correctly perceiving the properties of field stimuli data. In the T_t and G_t treatment only in one case (Portugal) the causality matches the causality from the field data; in the T_t only treatment subjects perceive a good deal of causality from expenditure to taxes, but in many cases it is not consistent with the causality in the field.

[19]Detailed parameter estimates, together with standard errors and level of statistical significance, are available in Bernasconi, Kirchkamp, and Paruolo (2004). All models presented here have been selected performing a computerised strategy, which started to drop coefficients with higher p-values. We have checked that the selected models are robust to alternative procedures.

Accepted Manuscript

	Vectors of responses to $_{\text{ECM}}TG$: $T_t + \widehat{\gamma} G_t$			Vectors of lab constants $\mu_i = (\mu_{iT^E} : \mu_{iG^E})$		Vectors of responses to $(_{\text{ECM}}T^E T,\ _{\text{ECM}}G^E G)'$: $(T_t^{E_i} - T_t, G_t^{E_i} - G_t)'$		Direction of short run causality inferred from $\Gamma_{l,xx}$ and $\Gamma_{\ell,ix}$		
	Field α_x'	T_t and G_t treatment α_{i1}'	T_t only treatment α_{i1}'	T_t and G_t treatment μ_i'	T_t only treatment μ_i'	T_t and G_t treatment α_{i2}	T_t only treatment α_{i2}	Field	T_t and G_t treatment	T_t only treatment
Austria	(0,1.308)	(0,0)	(0,*)	(0.264,0.146)	(0.155,*)	((-0.921,0), (0.084,-0.695))	((-0.863,*) ,*)	$T \leftrightarrow G$		$T \leftarrow G$
Belgium				(0.285,0.138)	(0.160,*)	((-0.838,0), (0,-0.656))	((-0.810,*) ,*)	$T \leftarrow G$	$T \leftrightarrow G$	$T \leftarrow G$
Denmark				(0.273,0.172)	(0,*)	((-0.885,0.138), (0.096,-0.622))	((-0.872,*) ,*)		$T \leftrightarrow G$	$T \leftarrow G$
Finland	(-0.953,0)	(-0.145,0)	(-0.164,*)	(0,0)	(0,*)	((-0.923,0.171), (0,-0.512))	((-0.825,*) ,*)	$T \leftarrow G$	$T \leftrightarrow G$	$T \leftarrow G$
France				(0.175,0)	(0,*)	((-0.709,-0.124), (0,-0.849))	((-0.655,*) ,*)	$T \leftarrow G$	$T \leftrightarrow G$	$T \leftarrow G$
Germany	(-1.007,0)	(0,0)	(0.165,*)	(0.199,0.329)	(0.110,*)	((-0.842,0), (0,-0.695))	((-0.961,*) ,*)		$T \rightarrow G$	$T \leftarrow G$
Greece				(0.606,0.199)	(0.581,*)	((-0.915,0), (0,-0.717))	((-1.058,*) ,*)		$T \leftarrow G$	$T \leftarrow G$
Ireland				(0.435,0.561)	(0.322,*)	((-0.739,0), (-0.141,-0.799))	((-0.676,*) ,*)		$T \rightarrow G$	$T \leftarrow G$
Italy	(-0.155,0)	(-0.072,-0.084)	(-0.034,*)	(0,0)	(0,*)	((-0.771,0.074), (0,-0.744))	((-0.864,*) ,*)		$T \rightarrow G$	
Netherlands	(-0.623,0)	(0,0)	(0,*)	(0.280,0.301)	(0.324,*)	((-0.736,0), (0,-0.800))	((-0.809,*) ,*)		$T \leftarrow G$	$T \leftarrow G$
Norway	(0,0.308)	(0,0.083)	(0,*)	(0,0.314)	(0,*)	((-0.732,0), (0.166,-0.751))	((-0.731,*) ,*)	$T \rightarrow G$	$T \rightarrow G$	$T \leftarrow G$
Portugal	(0.290,0.670)	(0,0.175)	(0,*)	(0.711,0)	(0.175,*)	((-0.744,0), (0.240,-0.633))	((-0.758,*) ,*)	$T \leftrightarrow G$	$T \leftrightarrow G$	$T \leftarrow G$
Spain				(0.276,0.141)	(0.202,*)	((-0.732,0), (0,-0.657))	((-0.548,*) ,*)		$T \rightarrow G$	$T \leftarrow G$
Sweden	(0,0.280)	(-0.034,0.086)	(0,*)	(0.239,0.138)	(0.112,*)	((-0.692,0), (0.111,-0.743))	((-0.670,*) ,*)	$T \rightarrow G$	$T \leftarrow G$	$T \leftarrow G$
United-Kingdom	(0,0.435)	(0,0)	(0,*)	(0.249,0.176)	(0.172,*)	((-0.883,0), (0.175,-0.780))	((-0.675,*) ,*)		$T \rightarrow G$	$T \leftarrow G$

TABLE 4: Summary of inference results

Accepted Manuscript

ACCEPTED MANUSCRIPT

4.2.2 Expectation schemes

The comparisons between the field and the lab VARs clearly reject the hypothesis of subjects following rational expectations. Looking at the middle part of table 4 we find instead support for adaptive expectations. Estimates of the vector α_{i2} describe how individuals react in the long run to errors in their expectations $(T^{E_i}_{t-1} - T_{t-1}, G^{E_i}_{t-1} - G_{t-1})'$. In the T_t and G_t treatment the coefficients for ${}_{\text{ECM}}T^E T$ in the equation for $\Delta T^{E_i}_t$ and the coefficients on ${}_{\text{ECM}}G^E G$ in the equation for $\Delta G^{E_i}_t$ are positive and close to 1. The same holds in the T_t only treatment for the coefficients for ${}_{\text{ECM}}T^E T$ in the equation for $\Delta T^{E_i}_t$. In both experimental treatments the estimated parameters are typically less than 1. (This is also confirmed by formal t-tests conducted on the parameters.) Conversely, in the T_t and G_t treatment, the off-diagonal diagonal of α_{i2} are in most cases equal to or very close to zero. In other words, subjects do not adjust expectations for taxes to errors in the expectations on expenditure, and vice versa.

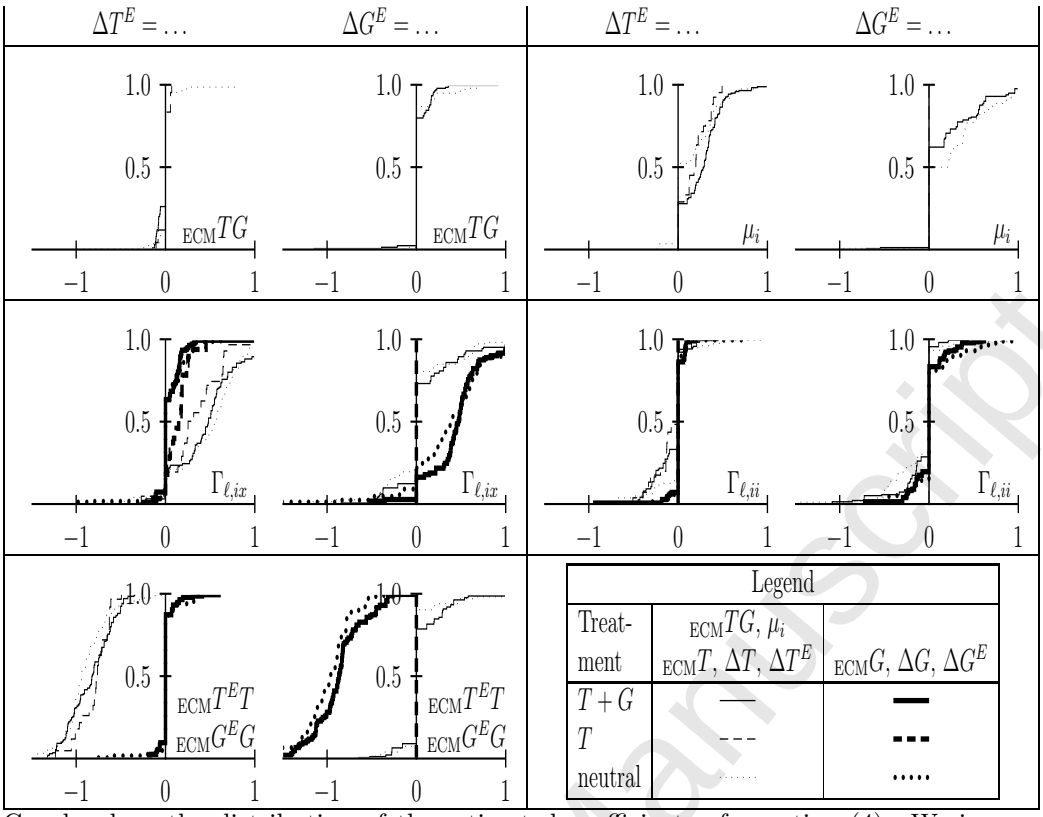
To obtain further evidence on the process subjects use to form expectations, we estimate equation (4) separately for each individual[20]. Figure 6 presents distributions of the coefficients for the two fiscal treatments as well as for the neutral treatment. Results confirm what we learned from the aggregate model. Coefficients of responses to the field error correction term ${}_{\text{ECM}}TG$ are zero or close to zero for all treatments and for both equations. Positive liner trends (μ_i) in the data are perceived by several participants. The coefficients of matrices $\Gamma_{\ell,ix}$ summed over lags show that, in both equations, the diagonal coefficients (hence the own-effects of $\Delta T_{t-\ell}$ on $\Delta T^{E_i}_t$ and of $\Delta G_{t-\ell}$ on $\Delta G^{E_i}_t$) are larger than the off-diagonal terms, but the latter don't necessarily vanish. The coefficients for $\Gamma_{\ell,ii}$, summed-up over lags, refer to the short run effects of past expectations on present expectations and show a moderate tendency of subjects to smooth out peaks in past expectations. The distributions of the adjustment coefficients to the error terms $({}_{\text{ECM}}T^E T, {}_{\text{ECM}}G^E G)'$ confirm that in all treatments subjects adjust to past forecast errors.

As a whole, we take the above results to support the hypothesis of subjects following adaptive expectations, but not quite a purely univariate model. While the results suggest that participants assume some interaction between

[20] To select the model we use the same procedure as in the aggregate model (footnote 19).

Accepted Manuscript

ACCEPTED MANUSCRIPT



Graphs show the distribution of the estimated coefficients of equation (4). We impose the restriction that each individual uses the same model for all countries.

FIGURE 6: Results of estimating equation (4) for each individual

fiscal variables, we can conclude from table 4 that this interaction changes over countries[21]. We take this to be evidence in favor of models of augmented-adaptive expectations.

In section 3 we have compared different treatment frames. In our baseline frame the variables shown on the sceen have their original economic labels. In the alternative 'neutral' frame these labels are just letters from A to D. We suspected that forecasts in the 'neutral' frame are less precise and, indeed, confirmed this hypothesis in figure 5 and in table 2. In the same section we compared different participant groups. Several of our participants are students in economics or business administration. We found, however, that forecasts of these participants are not better than those of other participants.

Can one use the estimates of equation (4) in order to learn more about why economic labels help participants to make forecasts? If participants from

[21]This can be confirmed be estimating equation (4) separately for individuals and countries.

ACCEPTED MANUSCRIPT

data comparison	T_t and G_t treatment frame: neutral ↔ economic labels						T_t and G_t treatment and economic frame participants from economics ↔ other fields					
dep. var.	ΔT^E			ΔG^E			ΔT^E			ΔG^E		
coefficient	econ	neutral	p	econ	neutral	p	other	econ	p	other	econ	p
${}_{\text{ECM}}TG$	-0.021	-0.008	0.543	0.026	0.052	0.390	-0.021	-0.021	1.000	0.030	0.025	0.879
μ_i	0.268	0.199	0.109	0.170	0.271	0.136	0.274	0.266	0.888	0.198	0.159	0.642
$\Delta T : \Gamma_{\ell,ix}$	0.395	0.442	0.478	0.082	0.077	0.961	0.500	0.350	0.068	0.144	0.055	0.324
$\Delta G : \Gamma_{\ell,ix}$	0.069	0.036	0.315	0.425	0.306	0.114	0.057	0.074	0.630	0.582	0.357	0.013
$\Delta T : \Gamma_{\ell,ii}$	-0.049	-0.002	0.149	-0.063	-0.091	0.561	-0.076	-0.037	0.375	-0.095	-0.050	0.335
$\Delta G : \Gamma_{\ell,ii}$	-0.009	0.008	0.352	0.018	0.062	0.226	-0.008	-0.009	0.987	-0.044	0.045	0.052
$\Delta T :_{\text{ECM}} \cdot^{E} \cdot$	-0.881	-0.918	0.437	0.038	0.062	0.623	-0.805	-0.914	0.098	0.101	0.010	0.083
$\Delta G :_{\text{ECM}} \cdot^{E} \cdot$	0.016	-0.005	0.319	-0.910	-1.021	0.035	0.028	0.011	0.504	-0.813	-0.953	0.057

The table reports average estimated coefficients of equation (4). The left part compares the neutral and the economic frame. The right part compares the average coefficients for participants with a background in economics with those from other disciplines. Next to the two means (for 'other' and 'econ') the table reports a p-value for a Welch two sample t-test.

Table 5: Impact of treatment frame and participants on equation (4)

economics are not systematically better in their forecasts, are they at least systematically different? Unfortunately, the answer to both questions is No.

Table 5 compares average coefficients of equation (4) for the treatment with neutral and economic labels and for the participants with different backgrounds. Most differences between coefficients are not significant.

5 Conclusions

Experimental economics has grown substantially over the last two or three decades, as it is now a well-acknowledged method through which decision theorists, game theorists and microeconomists have tested and refined theoretical models in their respective fields of interest.

Relatively few experiments have been conducted in the field of macroeconomics. The reason, probably, is that macroeconomists, dealing with real world questions to a much greater degree than other economists, are convicted that laboratory experiments cannot really answer such type of questions. The approach pursued in this paper follows the idea that using real world data as stimuli for subjects in the experiments might be a step in understanding which variables and forces are more relevant when individuals take actions of macroeconomic relevance.

Accepted Manuscript

ACCEPTED MANUSCRIPT

We propose our experiment as an alternative to experiments that are based on entirely artificial time series or time series whose parameters are based on field data. In our experiment participants receive as stimuli real world data on fiscal variables and form expectations on the basis of that information. Such an extension has advantages as well as disadvantages. Constructing a rational reference point is harder than with artificial time series. However, we have seen in our experiment that using economically meaningful data is important as it affects expectations of participants in our experiment.

References

Ahmed, S., Rogers, J. H., 1995. Government Budget Deficits and Trade Deficits: Are Present Value Constraints Satisfied in Long-Term Data? Journal of Monetary Economics 36, 351–74.

Allers, M., de Haan, J., de Kam, F., 1998. Using Survey Data to Test for Ricardian Equivalence. Public Finance Review 26, 565–82.

Alm, J., McKee, M., 2004. Tax Compliance as a Coordination Game. Journal of Economic Behavior and Organization 54(3), 297–312.

Bernasconi, M., Kirchkamp, O., 2000. Why do monetary policies matter? An experimental study of saving and inflation in an overlapping generations model. Journal of Monetary Economics 46, 315–343.

Bernasconi, M., Kirchkamp, O., Paruolo, P., 2004. Do fiscal variables affect fiscal expectations? Experiments with real world and lab data. Discussion Paper 04-27, University of Mannheim, SFB 504, http://www.kirchkamp.de/pdf/04-27.pdf.

Blinder, A. S., Morgan, J., 2005. Are two heads better than one? Monetary policy by committee. Journal of Money, Credit and Banking 37(5), 789–811.

Costa-Gomes, M., Crawford, V. P., 2001. Cognition and Behavior in Normal-Form Games: An Experimental Study. Econometrica 69(5), 1193–1235.

Accepted Manuscript

ACCEPTED MANUSCRIPT

Croson, R., 2000. Thinking like a game theorist: factors affecting the frequency of equilibrium play. Journal of Economic Behavior and Organization 41(3), 299–314.

Duffy, J., 1998. Monetary Theory in the Laboratory. Federal Reserve Bank of St. Louis Review 80, 9–26.

Fisher, F. M., 1962. A Priori Information and Time Series Analysis. Amsterdam: North-Holland.

Garcia, S., Henin, P.-Y., 1999. Balancing Budget through Tax Increases or Expenditure Cuts: Is It Neutral? Economic Modelling 16, 591–612.

Grun, D., 1991. What people know and what economists think they know: surveys on Ricardian equivalence. Australian Economic Papers 30, 1–9.

Hey, J. D., 1994. Expectations formations: rational or adaptive or. . . . Journal of Economic Behavior and Organization 25, 329–349.

Hommes, C., Sonnemans, J., Tuinistra, J., van de Velden, H., 2007. Learning in Cobweb Experiments. Macroeconomic Dynamics 11, 8–33.

Johansen, S., 1995. Likelihood-based Inference in Cointegrated Vector Autoregressive Models, Advanced Texts in Econometrics. Oxford and New York: Oxford University Press.

Lombardelli, C., Talbot, J., Proudman, J., 2002. Committees versus Individuals: An Experimental Analysis of Monetary Policy Decision-Making. Bank of England Quarterly Bulletin 42(3), 262–273.

Manasse, P., 1996. Are Taxes Too Low? Journal of Economic Dynamics and Control 20, 1263–88.

Marimon, R., Sunder, S., 1993. Indeterminacy of Equilibria in a Hyperinflationary World: Experimental Evidence. Econometrica 61, 1073–1107.

OECD, 2000. Fiscal Positions and Business Cycles, CD-ROM Statistical Compendium. Paris: OECD.

Offerman, T., Sonnemans, J., Schram, A., 1996. Value Orientations, Expectations and Voluntary Contributions in Public Goods. Economic Journal 106(437), 817–845.

Accepted Manuscript

ACCEPTED MANUSCRIPT

Payne, J. E., 1998. The Tax-Spend Debate: Time Series Evidence from State Budgets. Public Choice 95, 307–20.

Pesaran, M. H., 1987. The limits to rational expectations. Oxford and New York: Blackwell.

Schmalensee, R., 1976. An experimental study of expectation formation. Econometrica 44, 17–41.

Seater, J. J., 1993. Ricardian Equivalence. Journal of Economic Literature 31, 142–90.

Swenson, C. W., 1997. Rational expectation and tax policy: experimental market evidence. Journal of Economic Behavior and Organization 32, 433–455.

Trehan, B., Walsh, C. E., 1991. Testing Intertemporal Budget Constraints: Theory and Applications to U.S. Federal Budget and Current Account Deficits. Journal of Money, Credit, and Banking 23, 206–23.

A Instructions to the experiment

The experiment was conducted in German. In this section you find a translation of the instructions for the economic frame. The instructions for the neutral frame are almost indentical, except that we dropped any reference to the economic context:

1. Please read the instructions carefully. Only if you have understood them well you can successfully participate in the experiment and gain money.

2. **Thereafter** fill in the questionnaire at the screen.

Welcome to the strategy experiment

Welcome to the strategy experiment

This strategy experiment is financed by the University of Mannheim and the German research council.

The instructions are simple, and if you pay attention to them and decide carefully, you will win a considerable amount of money, which is disbursed to you at the end of the game.

The payment is dependent on your success. In the experiment you forecast the development of public expenditures and taxation in several European

Accepted Manuscript

ACCEPTED MANUSCRIPT

countries. For that purpose there are past data about budget debt, annual change of budget debt, government expenditure and taxes made available for you. Dependent on the quality of your forecast you receive a payment for each period.

Please note that we do not have any interest in paying less money than you are entitled to receive. We must return all the money which we do not disburse to the German research council.

Please note that we will not deceive you in this experiment. Everything you read in these instructions is correct. You may take this for granted, but actually there are occasionally experiments in psychology, where participants of experiments are deceived about parts of the experiment. This is not the case in economic experiments like this. In the beginning we explain exactly the rules to you, and we will also adhere to them.

Rules

You will play several rounds in turns. In each round it is your task to forecast the development of two variables. These variables refer to the development of government expenditure and taxes in several European states between 1950 and 2000. Which states you play in each case will be specified randomly and is not made known to you. These data are shown graphically.

Top of the screen

[[the figure shown here is similar to figure 1, except that only the first seven periods of UK are shown]]

On the left you can see the development of the budget debt and annual change of budget debt, each in per cent of the gross national product. The horizontal axis shows time in years. You may use this data to obtain a reference point how government expenditure and taxes will change in the future. Current periods are shown in black, past periods are shown in gray.

On the right you can see the government expenditure and taxes, again as percentage share of the gross national product. The vertical axis shows government expenditure, the horizontal axis shows taxes. Above the red diagonal government expenditure is higher than taxes; below the red diagonal government expenditure is lower than taxes. Past periods are shown in a lighter shade of gray than current periods.

Partial representation of the past development You can present yourself also only one part of the past periods to get a better overview.

In order to do so click on the diagrams budget debt and annual change of budget debt. In these diagrams the area to the right of your click is covered with a black rectangle. Also in the diagram of government expenditure and taxes the covered periods are not shown. Each click into the black area of

Accepted Manuscript

ACCEPTED MANUSCRIPT

the diagrams for budget debt and annual change of budget debt uncovers one period after another. A click into the blue area uncovers all periods.

[[the figure shown here is similar to figure 1, except that only the first five periods of UK are shown. In the two graphs on the left of the screen the area for periods six and seven is black. The area for period eight is blue.]]

Forecasts In order to make a forecast about the development of government expenditure and taxes, click onto the white area. Your forecast is shown in blue.

[[the figure shown here is similar to figure 1, except that only the first seven periods of UK are shown. These first seven periods are extended by a short blue line, i.e. an example forecast for one period.]]

If you are content with your forecast, please confirm it by clicking on confirm forecast. If you want to correct your forecast, please click on delete forecast.

Payment Given your forecast the computer determines a consumption decision which would be optimal for a person who lives in the period. From your consumption-decisions you derive a certain utility. This utility is compared with the utility you would have obtained if you had forecasted the true future development of taxes and government expenditure.

You receive a wage of 0.45€ per minute for a correct forecast. Worse forecasts result in smaller wages.

It is worth to spend some time to make a good forecast. Example: You need 2 minutes in order to make a very good forecast and therefore receive wages of 0.45€ per minute. Your income in the 2 minutes is thus 0.90€.

Another person, who makes forecasts for e.g. 4 periods in these 2 minutes, which are not so good, may only receive a wage of 0.10€ per minute for each forecast. The income of this person in the 2 minutes is thus only 0.20€.

You should settle your forecast within 2 minutes. If you need more time for a forecast, you are paid only for the first 2 minutes.

A warning on the left side will remind you, as soon as you need more than 2 minutes.

Furthermore, you see a list with the income from your past forecasts on the left side.

Duration of the experiment The experiment takes 90 minutes, regardless whether you made many or few forecasts in this time. That requires, however, that on average you spend at least 20 seconds for each forecast. If you spend less time for each forecast, you are finished with the experiment sooner, but earn fewer money, accordingly.

Should you have any questions, you now have the opportunity to ask them. In addition, you can ask questions at any time during the experiment.

Accepted Manuscript

ACCEPTED MANUSCRIPT

Appendix to the instructions

To determine your payoff we use the following model. **It is not necessary to understand this model to participate successfully in the experiment.** The model is shown only in case you want to control us.

In two subsequent periods you consume c_0 and c_1 and pay taxes t_0, t_1. You save the remaining part:

$$s_i = 1 - c_i - t_i \tag{5}$$

Your total income in each period is $Y = 1$ (note that all values are relative to the gross domestic product Y).

We call government expenditure g_i. Then your utility in two subsequent periods is

$$u = \prod_{i=0}^{1} (\gamma c_i + (1-\gamma) g_i) \tag{6}$$

In your case $\gamma = 0.75$.

Your budget restriction is

$$\sum_{i=0}^{1} s_i \cdot (1+r)^i = 0 \tag{7}$$

with an interest rate $r = 0.1$.

Based on your forecast for t_1 and g_1 we determine your optimal consumption c_0.

In the next period t_1 and g_1 are realised. Your actual consumption c_1, and, hence, your utility u, follows from the budget restriction. This utility is compared with the utility u^* that you could have obtained with the correct forecast for t_1 and g_1. Your wage is $(u/u^*)^\eta$. In your case $\eta = 12000$. This normalisation does not change your utility maximisation problem.

Accepted Manuscript